Molecular determinants of host adaptation in the plant pathogen Ralstonia solanacearum

Rekha Gopalan Nair

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Rekha GOPALAN NAIR

Le 19 novembre 2020

Déterminants moléculaires de l'adaptation à l'hôte chez la bactérie phytopathogène Ralstonia solanacearum

Ecole doctorale : **SEVAB - Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingenieries**

Spécialité : **Interactions plantes-microorganismes**

Unité de recherche :
LIPME - Laboratoire des Interactions Plantes-Microbes-Environnement

Thèse dirigée par
Alice GUIDOT

Jury
Mme Céline LAVIRE, Rapporteure
M. Lionel MOULIN, Rapporteur
Mme Sophie GAUDRIAULT, Rapporteure
M. Stéphane GENIN, Examinateur
M. Julien BRILLARD, Examinateur
M. Matthieu ARLAT, Examinateur
Mme Alice GUIDOT, Directrice de thèse

Table of Contents

Acknowledgements 5

List of figures 6

List of tables 8

List of abbreviations 9

Chapter 1. General introduction 12

1.1 Evolution and adaptation of bacterial plant pathogens 12

1.1.1 Why is it important to study the evolution and adaptation of bacterial plant pathogens? 12

1.1.2 How could bacterial pathogens evolve and adapt to new host and new environments? 16

1.1.3. Experimental evolution to study pathogen adaptation 22

1.2 Role of epigenetic modification in bacterial evolution and adaptation 24

1.2.1. Significance of DNA methylation in bacteria 24

1.2.2. Different form of DNA methylation in bacteria 26

1.2.3 Hemimethylation and its role 27

1.2.3. Role of DNA methylation in bacterial virulence 28

1.2.4. Methods to detect DNA methylation 31

1.3. A model organism: the *Ralstonia solanacearum* species complex (RSSC) 32

1.3.1. Extensive host range and worldwide distribution 34

1.3.2. A species complex 37

1.3.3. The RSSC life cycle 39

1.3.4. The RSSC genome 39

1.3.5. Pathogenicity determinant and regulation 40
1.3.6 Control measures of bacterial wilt 46
1.4 GMI1000: The study model 48
1.4.1 Experimental evolution of GMI1000 48
1.4.2 Enhanced *in planta* fitness explained by genomic polymorphisms 49
1.5 Objectives 51
Chapter 2. Adaptation of *Ralstonia solanacearum* species complex to a tolerant plant: An overview at the molecular level 53
2.1 Brief introduction 53
2.2 Genetic bases of adaptation of GMI1000 to tomato Hawaii 7996 plant 54
2.2.1 Article 54
Convergent rewiring of the virulence regulatory network promotes adaptation of *Ralstonia solanacearum* on resistant tomato 54
2.2.2 Journal 54
Molecular Biology and Evolution 54
2.2.3 Summary 54
Chapter 3. Two-fold profiling of the experimentally evolved clones: genomic and transcriptomic level 56
3.1. Introduction 56
3.2 Materials and methods 60
3.2.1 Selection of clones 60
3.2.2 Bacterial growth conditions 60
3.2.3 Genomic DNA preparation 60
3.2.4 Detection of genomic polymorphisms 61
3.2.5 Ribosome depletion and RNA sequencing 61

3.3 Results 61
3.3.1 Genomic sequencing analysis divulged between 0 and 2 polymorphisms per clone.. 61
3.3.2 Transcriptomic analyses of the clones obtained from Zebrina and Bean 64
3.4 Discussion 85
Chapter 4. Methylome analyses of the evolved clones 88
4.1 Brief introduction 88
4.1 Material and methods 92
4.1.1 Bacterial strains and growth conditions 92
4.1.2 Single-molecule, real-time (SMRT) sequencing 92
4.1.3 Analysis of methylation profile of GTWWAC 94
4.2 Results 95
4.2.1. SMRT sequencing technology for the analysis of GTWWAC methylation 95
4.2.2. Detection of differential methylation marks between the ancestral clone and the evolved clones 96
4.2.3 Fully methylated differential motifs were observed only on the evolved clones...... 103
4.2.4 Hemimethylation 105
4.2.5 DEGs vs DMRs 109
4.3 Discussion 111
Chapter 5. Conclusion and Perspectives 121
References 126
Annexure 142
Summary 179

Acknowledgements

My sincere thanks to the Labex Tulip PhD grant “Young Scientists for the Future” for giving me the opportunity to work at a renowned laboratory such as INRAE. From the first interview with Dr. Alice Guidot and Dr. Stephane Genin for this PhD position until the end of this PhD journey, they have been quite helpful and extremely considerate, and for that, I am very grateful. I cannot thank enough Alice for helping me find accommodation and navigate in a new country where English is sparsely spoken outside of the lab. My sincere gratitude for making me feel welcome in the team to all members of the team (present and past) and especially the ones with whom I shared the office over the past years. Thanks to them, for all the discussions we had from science to culture and politics! I am extremely glad to have met some nice colleagues from various teams turned treasured friends for life who made this journey even more exciting. Finally yet importantly, I am forever indebted to my family for everything that I am today!

List of figures

Figure 1 The Zigzag model of plant immunity 13
Figure 2 Chemical structure of the various forms of DNA methylation 25
Figure 3 Bisulfite sequencing 29
Figure 4 Single-molecule, real-time (SMRT) sequencing 29
Figure 5 Distribution of Ralstonia solanacearum species complex 33
Figure 6 Phylogenetic analysis of RSSC based on egl sequence 35
Figure 7 Phylogenetic reclassification of RSSC strains 36
Figure 8 Life cycle of Ralstonia solanacearum 38
Figure 9 Major virulence regulatory networks in RSSC strains 42
Figure 10 Experimental evolution of Ralstonia solanacearum 47
Figure 11 Whole genome sequencing of the evolved clones 50
Figure 12 Downregulated genes in Zebrina clones 65
Figure 13 Upregulated genes in Zebrina clones 65
Figure 14 Common downregulated genes in the clones evolved on Bean 68
Figure 15 Common downregulated genes in five clones evolved of Bean................................. 69
Figure 16 Common upregulated genes in the clones evolved on Bean 71
Figure 17 Common upregulated genes in five clones evolved of Bean..................................... 72
Figure 18 Comparison of all DEGs from Hawaii, Zebrina and Bean clones 79
Figure 19 Functional distribution of host specific DEGs in Zebrina and Bean 80
Figure 20 Comparison of downregulated DEGs from the evolved clones (Hawaii, Zebrina, and Bean) to the downregulated gens of hrpB regulon 81
Figure 21 Comparison of DEGs from the evolved clones (Hawaii, Zebrina, and Bean) to the efpR regulon 83
Figure 22 Barplot of the number of differentially methylated motifs between the ancestral and the Hawaii evolved clones 97
Figure 23 Barplot of the number of differentially methylated motifs between the ancestral and the Zebrina evolved clones 98

Figure 24 Barplot of the number of differentially methylated motifs between the ancestral and the Bean evolved clones 99
Figure 25 Circos plot of the distribution of differential methylation among the experimentally evolved clones in comparison to the ancestral clone 108
Figure 26 Number of differentially methylated regions (DMRs) between evolved clones and the ancestral clones. 110
Figure 27 Correlation between the DEGs and DMRs.. 112
Figure 28 Differentially methylated motif at RSc2612 on R. solanacearum 116
Figure 29 Differentially methylated motif upstream RSc0102-03 on R. solanacearum............. 116
Figure 30 Differentially methylated motif upstream RSp1329-30 on R. solanacearum 117
Figure 31 Differentially methylated motif at RSc3393 on R. solanacearum 118
Figure 32 Differentially methylated motif upstream RSp1674-75 on R. solanacearum 119

List of tables

Table 1 Mtases and their motifs of GMI1000 as identified by REBASE and SMRT 30
Table 2 Phenotypic interaction between plant accessions and Ralstonia solanacearum. 45
Table 3 Characteristics of the analyzed clones evolved on Eggplant Zebrina 57
Table 4 Characteristics of the analyzed clones evolved on Bean 58
Table 5 Statistical analyses of the mean CI obtained from the clones of various experimental host 59
Table 6 Gradient logFC (log fold change) and FDR (pvalue adjusted) for all the studied evolved clones 63
Table 7 List of genes downregulated in all the studied Bean clones 70
Table 8 List of genes upregulated in all the studied Bean clones 73
Table 9 List of downregulated genes specific to the studied Bean clones except c1 74
Table 10 List of upregulated genes specific to the studied Bean clones except c1 75
Table 11 Common downregulated genes of HrpB cluster shared among the downregulated genes of all Zebrina, Bean and Hawaii clones 82
Table 12 List of Hawaii tomato evolved clones for methylome analysis 89
Table 13 List of Zebrina evolved clones for methylome analysis 90
Table 14 List of Bean evolved clones for methylome analysis 91
Table 15 List of differentially (fully) methylated positions in all the evolved clones 100
Table 16 List of differentially type 1-hemimethylated positions (methylated ancestral clone) 101
Table 17 List of differentially type 2-hemimethylated positions (methylated evolved clone) .. 102
Table 18 Statistical analysis of differentially fully methylated positions between clones evolved on various host plants 104

List of abbreviations

3-PAME	3-Hydroxy palmitic acid methyl ester
4mC	N-4-methylcytosine
5mC	N-5-methylcytosine
6mA	N-6-methyladenine
ACUR	Alternative codon usage region
BCA	Bacterial control agent
BS-seq	Bisulfite sequencing
BW	Bacterial wilt
CcrM	Cell cycle regulated methyltransferase
CFU	colony forming units
CI	Competitive index
Dam	DNA adenine methyltransferase
DEGs	Differentially expressed genes
DMRs	Differentially methylated regions
dpi	days post inoculation
EE	Experimental evolution
EEL	Exchangeable effector locus
EPEC	Enteropathogenic *Escherichia coli*
EPS	Exopolysaccharides
ETI	Effector triggered immunity
ETS	Effector triggered susceptibility
EV	efpR variant
FDR	False discovery rate
GMI1000	**G**enetique des **M**icroorganismes **I**NRA
GO	Geno ontology
HGT	Horizontal gene transfer

HR	Hypersensitive response
IPD	Interpulse duration
IS	Insertion sequence
logFC	log fold-change
LRR	Leucine rich repeat
MAMP	Microbe associated molecular pattern
mg/L	milligram/liter
mM	millimolar
MSRE	Methylation sensitive restriction enzyme
MTase	Methyltransferase
NB-LRR	Nucleotide binding leucine rich repeat
OD	Optical density
PAMP	Pathogen associated molecular pattern
Pap	Pyelonephritis-associated pilus
PC	Phenotypic conversion
PTI	Pathogen triggered immunity
QTL	Quantitative trait loci
RNAseq	RNA sequencing
RSM	Restriction, specificity, methylation gene
RSSC	*Ralstonia solanacearum* species complex
RT-qPCR	Quantitative Reverse transcription PCR
SCV	Small colony variant
SMRT	Single molecule real-time sequencing
SNP	single nucleotide polymorphisms
SPE	serial passage experiment
T2SS	Type 2-secretion system
T3E	Type 3 effector
T3SS	Type 3-secretion system

TCA	Tricarboxylic acid cycle
WGBS	Whole genome bisulfite sequencing

Chapter 1. General introduction

1.1 Evolution and adaptation of bacterial plant pathogens

1.1.1 Why is it important to study the evolution and adaptation of bacterial plant pathogens?

Bacterial plant pathogens are able to induce diseases in an increasing number of plants all over the world (Strange and Scott, 2005). These phytopathogenic bacteria affect crops globally on a large-scale and have a negative impact on agriculture because of their significant economic losses and environmental ramification. Therefore, it is imperative to develop methods to manage plant pathogens for global food security (Martins *et al.*, 2018).

In 2012, a survey was conducted to nominate the top 10 bacterial pathogens based on their scientific and economic importance (Mansfield *et al.*, 2012). The survey consisted of 458 votes from the scientific community worldwide that led to the formation of a top 10 bacterial plant pathogen list. The list with their ranking are as follows: (1) *Pseudomonas syringae*; (2) *Ralstonia solanacearum*; (3) *Agrobacterium tumefaciens*; (4) *Xanthomonas oryzae* pv. *oryzae*; (5) *Xanthomonas campestris* pathovars; (6) *Xanthomonas axonopodis* pathovars; (7) *Erwinia amylovora*; (8) *Xylella fastidiosa*; (9) *Dickeya (dadantii and solani)*; (10) *Pectobacterium carotovorum* (and *Pectobacterium atrosepticum*). Each of these ten bacterial plant pathogens have significant agronomic importance.

In nature, there is a continuous battle between the pathogen and the plant species resulting in an exceptional coevolution. Upon the pathogen attack, the host plant deploys defense mechanisms to resist the pathogen invasion. However, the resistance mechanisms imposed by the host plant can be overcome by the pathogens using various adaptive mechanisms and so the battle continues. A four phased zigzag model proposed by Jones and Dangl explains

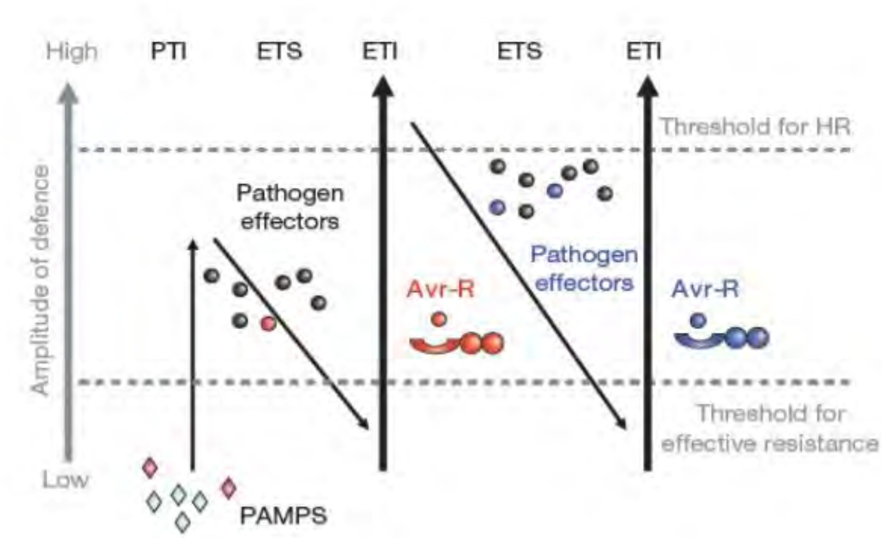


Figure 1 The Zigzag model of plant immunity

The model describes the various stages of the coevolution between the plant and pathogen. There are 4 phases in total and are shown by arrow marks. PAMP – Pathogen associated molecular pattern; PTI – pathogen triggered immunity; ETS – effector triggered susceptibility; ETI – effector triggered immunity. From Jones and Dangl, 2006.

the plant immune system (Figure 1). The plants detect the PAMPs/MAMPs (pathogen/ microbial associated patterns) by PRRs (pattern recognition receptors) activating the PTI (pathogen triggered immunity) that can hinder further colonization of the bacteria in phase 1. In phase 2, successful pathogens releases effectors that interferes with PTI, resulting in ETS (effector-triggered susceptibility). In phase 3, one of the plant NB-LRR (nucleotide binding – leucine rich repeat) proteins recognizes one of the effectors resulting in ETI (effector-triggered immunity). ETI is an intensified version of PTI, usually resulting in disease resistance and hypersensitive cell death response (HR) at the infection site. In phase 4, the pathogen is driven by natural selection to circumvent ETI by either losing the recognized effectors or by gaining new effectors through horizontal gene transfer that can suppress ETI. Natural selection of the plant NB-LRRs can result in new specificities that can recognize the acquired effectors, resulting in ETI again (Jones and Dangl, 2006).

The use of resistant cultivars is a favored disease management strategy as it could be highly effective with negligible deleterious effect to the environment. The evolution of virulence may be pivotal for the emergence and re-emergence of pathogens as host switching, host range expansion and overcoming host resistance compromise the control strategies. The evolution of pathogenicity governs the durability of resistant cultivars (Mcdonald and Linde, 2002; Sacristán and García-Arenal, 2008). Resistant crop breeding which uses the molecular breeding and genetic engineering approach to transfer the resistance genes or QTLs (Quantitative trait loci) against pathogens is one of the better and most effective, environmental friendly approach to counter microbial diseases as opposed to the use of pesticides (Akhon and Machray, 2009; Gust et al., 2010). However, the plant resistance can be transitory as the pathogens are evolving continuously. For example, in the bacterial pathogen *P. syringae* pathovars, the phytotoxins and effectors from the type III secretion system (T3SS) not only aid in evoking the disease but also overcome/suppress host resistance. Phytotoxins such as coronatine, syringolin A, syringomycin and syringopeptin are produced by many pathovars of *P. syringae* during pathogenesis (Ichinose *et al.*, 2013). The phytotoxins coronatine and syringolin A weakens the plant stomatal immunity, whereas syringomycin and syringopeptin targets the plant cell membrane and induces pore formation. This leads to increased permeability and rapid K^+ efflux from the plant cells resulting in the acidification of the cytoplasm and thereby plant necrosis. The T3Es (Type 3 effectors) subdues the PAMP and effector-triggered immunity by inhibiting the signaling pathways that causes plant defense response such as HR (hypersensitive response) and cell death (Bender *et al.*, 1999). In *P. syringa*e pv. *tomato* (*Pto*) strains, that causes bacterial speck in tomato, host resistance was overcome by mutations in one of the T3E, *hopM1* gene and an additional mutation in *fliC* gene that results in reduced recognition by the tomato immune system (Cai *et al.*, 2011). In addition, pathogen evolution can also enlarge their host range. For instance, interspecific homologous recombination led the pathogenic bacterium *X. fastidiosa,* to infect two new hosts namely mulberry and blueberry in the Americas, while the ancestral *X. fastidiosa* subsp. *multiplex* was unable to infect the same host. Sequence analysis from the isolates revealed that a single ancestral homologous recombination event has given rise to *X. fastidiosa* subsp. *morus* and a recombinant *X. fastidiosa* subsp. *multiplex* group that infects multiple hosts in addition to blueberry (novel host) (Nunney *et al.*, 2014; Bonneaud *et al.*, 2019). Epidemiological evidence for the emergence of a new pathogenic variant of *R. solanacearum* in Martinique is another example for host expansion of the pathogens (Wicker *et al.*, 2009). The IIB/4NPB strains are an emerging population of *R. solanacearum* with a striking ability to acquire new hosts. Until 2002, these strains were specific to anthurium and cucurbits, but a year later the isolates were pathogenic to solanaceous crops such as tomato and they were more aggressive than the other *ralstonia* strains that causes bacterial wilt in tomato. This particular strain was rapidly spread throughout the island and has infected several wild species and weeds (Wicker *et al.*, 2007, 2009). The isolated strain from the infected tomato were associated with bananas as former crop on the field. A similar observation were also made on wilted watermelons and zucchinis in the field with an earlier banana cultivation (Wicker *et al.*, 2009).

To cope with pathogen evolution and adaptation, it is crucial to have plants with a durable disease resistance. Durable disease resistance is defined as resistance that remains effective while the 'resistant' cultivar is grown on an environment prone to the disease for extensive period (Johnson, 1983). In order to achieve durable resistance in a plant, it is important to break the boom-and-bust cycle of major-gene resistance (R-genes). In a boom-and-bust cycle, a resistant cultivar with a single major resistance gene is introduced in the field and if it is effective, the cultivar is planted over a large geographical area referred to as the 'boom'. The local pathogen population adapts to the resistance gene by evolving a new population that could overcome the resistance gene referred to as the 'bust'. Selection pressure imposed by the pathogen population leads to the breakdown of resistance, which is achieved, by mutation and/or recombination (Mcdonald and Linde, 2002; Strange and Scott, 2005). It has been hypothesized that the evolutionary potential of a pathogen is the reflection of its population genetic structure. Pathogen populations with high evolutionary potential are most likely to overcome resistance than those pathogens with low evolutionary potential (Mcdonald and Linde, 2002). It is therefore important to understand the fundamental mechanisms and determinisms that govern the pathogen adaptation to their hosts, which is still rudimentary (Mansfield *et al.*, 2012).

1.1.2 How could bacterial pathogens evolve and adapt to new host and new environments?

Pathogens are in a relentless state of flux and evolve continuously (Strange and Scott, 2005). Adaptation to different environments and potential new hosts could drive evolution (Wicker *et al.*, 2012). Bacterial pathogen evolution is driven by selection that leads to genetic and/or epigenetic modification of their genome.

1.1.2.1. The importance of genetic modifications in pathogen evolution

Sexual reproduction is an important mechanism by which genetic variation is established in the alleles. However, for organisms like bacteria that lack the sexual cycle, genetic variation occurs through other mechanisms. In bacteria, genetic variations occur either vertically by mutations or genomic polymorphisms over a number of generations or horizontally by horizontal gene transfer (HGT) of foreign genetic material ranging from as small as 1kb to more than 100kb DNA fragments (Brussow *et al.*, 2004).

a. Vertical evolution

The various means employed by the bacteria to thrive in the environment and host through vertical evolution includes:

- Point mutation (SNPs)
- Amplifications (such as duplication)
- Insertions/Deletions (InDels)
- Large Genomic rearrangements

Single nucleotide polymorphisms (SNPs) or point mutations can have a remarkable effect on the biology of all organisms if the SNP is located on structural genes. SNPs in pathogens can be adaptive leading to niche expansion. A number of SNPs have been identified in a variety of bacterial pathogens that are associated with the pathogenicity-enhancing or 'pathoadaptive' mutation category. For example, SNPs in the fimbrial adhesion gene, *fimH* of *Salmonella typhimurium* has major differences in host colonization properties. Single amino acid replacement caused by SNPs at two position, Gly61Ala and Phe118Ser in the *fimH* loci significantly enhanced cell adhesion and biofilm formation (Weissman *et al.*, 2003).

Duplication of many stress related genes in *Escherichia coli* favors adaptation to heat stress. In particular, duplication of the transcription regulator, *evgA*, permits *E. coli* to withstand over 50°C (Kondrashov, 2012).

Pathogenicity could also be gained by loss of genes. An example of evolution by gene deletion comprises *Shigella* spp., the causal agent of shigellosis disease in human beings, which developed pathogenicity by the loss of *E. coli* specific gene *cadA*. Maurelli and co-workers demonstrated that *Shigella* spp. was evolved from the commensal *E. coli* by not just acquiring the virulence plasmid but also by the gene deletion of *cadA* responsible for the inhibition of enterotoxin activity (Maurelli *et al.*, 1998).

Another mechanism by which variation among the adapted strains can be observed are by genomic rearrangements. Genomic rearrangements mediated by recombination of homologous sequences such as transposons, insertion sequence (IS) and mobile genetic elements plays an important role in the bacterial pathogenicity and virulence (Bartoli *et al.*, 2016). For example, in *Pseudomonas syringae,* the novel EEL (exchangeable effector locus) harbors effector genes containing fragments associated with insertion sequences and mobile genetic elements which are responsible for pathogenicity and fitness in host plants (Alfano *et al.*, 2000).

b. Horizontal evolution

The ability to lose and gain genetic material while conserving the core genome could be a key reason why bacterial pathogens are able to exploit a range of different hosts rapidly and simultaneously (Bonneaud *et al.*, 2019). Species barriers do not restrict HGT and the acquired DNA can encode intact metabolic pathways, complex surface structures or virulence factors (Preston *et al.*, 1998; Brussow *et al.*, 2004; Groisman and Casadesús, 2005; Wion and Casadesús, 2006). There are three main mechanisms that mediate HGT (Thomas and Nielsen, 2005) and they are as follows:

- Transformation
- Conjugation
- Transduction

Natural transformation allows the uptake of free DNA in competent bacteria. This can create mutants with pathogenicity islands or gene clusters encoding virulence factors making them more virulent/aggressive or avirulent strains become virulent. For example, the 35kb region in enteropathogenic *E. coli* (EPEC) enables invasion and survival in mammalian cells whereas the commensal *E. coli* that lacks the region are nonpathogenic (Marcus *et al.*, 2000). Recombination allows the transfer of genetic material within the bacterial populations of same ecological niche or geographic location. This enables rapid adaptation to the hosts as seen with *X. fastidiosa* in Mulberry that is associated with a large recombination event (Nunney *et al.*, 2014; Bonneaud *et al.*, 2019). *R. solanacearum* is also capable of HGT owing to its ability to develop a physiological state of competence that allows exchange of large DNA fragments by natural transformation (Coupat *et al.*, 2008; Guidot *et al.*, 2009). This exceptional property of HGT could be the reason behind the new pathogenic variants of *R. solanacearum* observed in the Martinique islands

(Wicker *et al.*, 2007, 2009). Studies conducted on various phylotypes of *R. solanacearum* reveals the evolutionary potential of HGT on virulence of the pathogen. It was corroborated that the pathogen could take up to 80 kb DNA fragments via natural transformation within the phylotypes (Coupat-Goutaland *et al.*, 2011). The study also found that weakly aggressive strains could acquire a more aggressive phenotype after a subsequent HGT among distantly related strains. The experiment was performed on Psi07 strain of *R. solanacearum* and the transformant was Psi07 with RSc2123 – RSc2155 region (containing Type III effectors) from GMI1000 strain. The transformant became more aggressive on the tomato plant compared to the wild type strain Psi07 (Coupat-Goutaland *et al.*, 2011).

Conjugation is mediated between adjacently located bacteria where the mobile genetic elements are transferred by pili structures. The surging prevalence of antibiotic resistance in many pathogenic bacteria is an outcome of evolution and selective pressure because of extensive use of antibiotics in medicine, animal feeding and agriculture (Grohmann *et al.*, 2003). An earlier example of dissemination of antibiotic resistance via conjugation was observed in *Mycobacteria* and *Streptomycetes* for tetracycline (Davies, 1994).

Transduction is favored by phage mediated HGT, where the bacterial DNA of the host genome is transferred in the bacteriophage head resulting in functional phages, which is then delivered to a suitable bacteria. The absence of phage DNA makes it harmless to the recipient bacteria and thereby the foreign material is injected into the genome (Brussow *et al*., 2004). Generalized transduction has been observed in many pathogenic bacteria including *S. typhimurium* (Schicklmaier *et al.*, 1995) and *Yersinia* strains (Hertwig *et al.*, 1999).

1.1.2.2. Role of phenotypic heterogeneity in pathogen adaptation

Phenotypic switching or phenotypic heterogeneity is another adaptive strategy observed in many bacterial pathogens, where the phenotype of some of the bacterial cells in a clonal population varies from rest of the bacterial cells. This can help adaptation of bacterial pathogens to a new ecological niche through two main beneficial strategies, the bet hedging and the division of labor strategies. Bet hedging enables the pathogen to persist in fluctuating environments, while the division of labor aids fitness by enabling the subpopulation to perform different functions simultaneously (Arnoldini *et al.*, 2014; Weigel and Dersch, 2018). An example of phenotypic heterogeneity by bet hedging strategy is the formation of 'persister' cells. These cells grow slowly or behave dormant but are difficult to destroy (Balaban, 2004). The small colony variant (SCV) is a distinct phenotype exhibited by many pathogenic bacteria that are associated with persistence in the host and can be less susceptible to antibiotics (Proctor *et al.*, 2006; Weigel and Dersch, 2018). Examples of persister cells have been observed in many pathogenic bacteria including *Staphylococcus aureus*, *S. epidermidis* and *Pseudomonas aeruginosa* (Proctor *et al.*, 2006). A heterogeneous population on biofilm formation promote division of labor strategy and are involved in many acute and chronic infections such as those caused by *Vibrio cholerae* and *P. aeruginosa.* Formation of biofilm protects the bacteria against environmental stress and confers resistance to antibiotics (Drenkard and Ausubel, 2002; Weigel and Dersch, 2018).

An example that uses both the strategies is the bistable expression of the virulence gene *hilD*, which belongs to type 3 secretion system 1 (*ttss-1*) in *Salmonella Typhimurium* populations. The bistable expression of the *ttss-1* generates two subpopulations, one is resistant to antibiotics, the second is more effective for host colonization (Arnoldini *et al.*, 2014). One other example of bet hedging is the bistable expression of the global virulence regulator, RovA, in the enteric pathogen *Yersinia pseudotuberculosis*. This heterogeneous expression of *rovA* is essential for virulence during the infection as it promotes both pre-adaption of some cells in the fluctuating environment of the intestinal epithelium during the initial stages of infection and adaptation to the intestinal tract during the later stages of infection (Nuss *et al.*, 2016).

Phenotypic heterogeneity has been ascertained experimentally to be an evolvable selective trait with a fitness benefit (Bódi *et al.*, 2017). Some of this heterogeneous population can be easily visible as colony variation (Van Der Woude and Bäumler, 2004). Other phenotypic variations include antibiotic resistance, growth rates, motility and biofilm formation. This phenomenon can be induced by phase variation, stochastic gene expression and genetic or epigenetic modifications (Ackermann *et al.*, 2008; van Gestel and Nowak, 2016; Weigel and Dersch, 2018).

Well-established examples of phenotypic heterogeneity in the bacterial plant pathogen *R. solanacearum* includes *phcA* (Poussier *et al.*, 2003) and *efpR* mutant (Perrier *et al.*, 2016, 2019). Heterogeneity in *R. solanacearum* is caused by spontaneous mutations in the global regulators PhcA and EfpR. The global virulence regulatory gene, PhcA positively and negatively controls the expression of various genes responsible for pathogenicity and metabolic functions (Clough *et al.*, 1997; Yoshimochi *et al.*, 2009). *R. solanacearum* transitions from the saprophytic form to the parasitic form using PhcA quorum sensing and regulatory system (Peyraud *et al.*, 2016). Successive studies found that the PhcA regulatory system mediates a compromise between optimal growth in the early stages (at low cell density during plant colonization) and the production of virulence factors in the later stages of infection (at high cell density during disease development) (Yoshimochi *et al.*, 2009; Khokhani *et al.*, 2017). It has been reported that some cells of the pathogen population can undergo spontaneous mutations in the *phcA* gene resulting in non-pathogenic variants but exhibiting high metabolic abilities even in high cell density (Poussier *et al.*, 2003; Peyraud *et al.*, 2016) . Some of these *phcA* mutants were reversed to the pathogenic form *in planta* (Poussier *et al.*, 2003). Mutations in *efpR* of *R. solanacearum* causes the isogenic population to produce two different colonies namely the S variant ('smooth' colony, similar to WT) and the EV variant (rough colony, 'efpR variant'). Here the pathogen uses bet hedging strategy to generate two subpopulations, one with better host colonizing properties (S phenotype) and the other adapted to diverse environments with wide metabolic capacity (EV phenotype) (Perrier *et al.*, 2016, 2019).

1.1.2.3. Role of epigenetic modification in pathogen adaptation?

Yet another mechanism by which the bacterial pathogen can adapt is through epigenetic modifications. The concept of 'epigenetics' was first introduced in 1942 by the British embryologist C. H. Waddington to describe the influence of environment on the development of specific traits. However, the physical nature of the gene and their role in heredity was unknown. Modern concept of epigenetics refer to the non-genetic modifications that enable stable alterations in the gene expression and are passed onto next generations (Al Akeel, 2013). The recent advances in technology fostered a renewed interest in bacterial epigenetics and their significance in pathogen adaptation, which are briefed later in this chapter *(section 1.2)*.

1.1.3. Experimental evolution to study pathogen adaptation

Experimental evolution (EE) is a powerful technique to observe the evolutionary dynamics within/between populations and to perceive the molecular mechanisms of adaptation of an organism to a new environment (Elena and Lenski, 2003; Kawecki *et al.*, 2012; Guidot *et al.*, 2014). The experiments range from simple growth experiments that uses Erlenmeyer flasks, agar plates and chemostat to evolution in complex environments like the pathogen host. Lenski and co-workers first visualized the concept of EE in 1988 by growing 12 populations of *E. coli* strain REL606 in minimal media with limiting glucose (Lenski *et al.*, 1991). In this long-term experimental evolution, the strains were propagated for over three decades resulting in more than 72,000 generations to date and the population continues to adapt. A major discovery within the population was the evolution of the citrate utilization, which is significant as *E.coli* cannot utilize citrate under oxidizing conditions (Blount *et al*., 2008).

The evolution experiments in the laboratory setup are now being carried out on various organisms like virus, bacteria, fungi, yeast, plants, insects and vertebrates to further understand the genetic variance that drives adaptation. The use of experimental evolution approach with bacterial pathogen has helped to better understand the virulence evolution and molecular mechanisms for host adaptation. For example, evolution experiment on opportunistic pathogen *Burkholderia cenocepacia* of cystic fibrosis was selected for biofilm formation on a plastic bead suspended in minimal media in a test tube for approximately 1500 generations. The evolved clones had multiple mutations affecting pathways including tricarboxylic acid cycle (TCA) enzymes, polysaccharide production, global transcription and iron scavenging that are associated with adaptation to extended biofilm selection (Traverse *et al.*, 2013). Several studies have also taken place in a setup that closely resembles the natural condition. The environment is still of course controlled and they usually combine multiple stresses such as antibiotics or carbon sources changing over time. Lindsey and colleagues propagated hundreds of *E. coli* populations to evolve in increasing concentrations of the antibiotic rifampicin and successfully identified mutations in *rpoB* gene that confers resistance to rifampicin (Lindsey *et al.*, 2013).

Many microbial evolution experiments has been conducted on more complex framework such as the eukaryotic host to understand pathogen adaptation or the influence of the host in bacterial evolution. The host species include animal or plant systems such as worms (King *et al.*, 2016), mice (Tso *et al.*, 2018), mimosa (Marchetti *et al.*, 2010) and maize (Quesada *et al.*, 2016). An experimental evolution of the plant pathogen *Xanthomonas citri* subsp. *xcc* validated the hypothesis that a resistant host imposes stronger pathogen selection than the susceptible host. The evolution experiment of Xcc strain that causes citrus canker disease was performed on both resistant (kumquat) and susceptible host (grapefruit). The resistant kumquat usually exhibited late HR while the grapefruit was highly susceptible to the Xcc strain. After 55 passages of inoculation of the Xcc strain into kumquat and grapefruit (three lineages each), clones from two lineages from the resistant kumquat manifested loss of HR and the loss was stable (Trivedi and Wang, 2014). There were 28 nonsynonymous mutations observed in these two resistant lineages, of which 11 mutations were on Type III effectors and are generally involved in virulence and pathogenicity related functions. Parallel mutations were observed in these two lineages in multiple T3Es such as *avrXacE1* and *pthA4,* OrfT (cointegrate resolution protein T) and XpsF (general secretion pathway protein F) (Trivedi and Wang, 2014). Another study attempted to convert the plant pathogen, *R. solanacearum* into a legume symbiont using experimental evolution. Here, the symbiotic plasmid of *Cupriavidus taiwanensis* was transferred to a *R. solanacearum* strain before experimental evolution on *Mimosa* plants. Nitrogen fixation was not achieved (yet) but the experiment showed adaptation for nodulation and plant cell infection (Marchetti *et al.*, 2010, 2014). Another experimental evolution with *R. solanacearum* aimed to identify the genetic bases of adaptation of the pathogen to multiple host plants. This experiment successfully identified adaptive mutations in a new transcription regulator, EfpR, that enhanced fitness *in planta* (Guidot *et al.*, 2014) (*detailed in section 1.3.1*).

1.2 Role of epigenetic modification in bacterial evolution and adaptation

1.2.1. Significance of DNA methylation in bacteria

Methylation is the mechanism by which a methyl group is added to the DNA by a methyltransferase (MTase), which changes the activity of DNA but not the sequence and it is one of the most common form of post replicative modification in microbial genomes. It is well established that DNA methylation is present in almost all bacteria since the putative DNA MTases were found in over 95% of the ~29,000 publically available bacterial genomes sequenced to date (http://tools.neb.com/genomes/). Interestingly there is a large number of MTases in prokaryotes in comparison to eukaryotes (Roberts et al., 2014). Blow and coworkers recently identified 620 MTases in the genomes of diverse bacterial and archaeal species (Blow *et al.*, 2016).

The importance of DNA methylation of bacteria is well studied in model organisms like *E. coli* or *Caulobacter crescentus* where the methylation of adenine residues is catalyzed by Dam (DNA adenine methylase) and CcrM (Cell cycle regulated MTase family) MTases, respectively (Casadesús and Low 2006). The role of DNA methylation has been reported for various phenotypes, such as cell division, biofilm formation or virulence mostly in human and animal bacterial pathogens (Casadesús and Low 2006; Fang et al. 2012; Shell et al. 2013). However, the role of DNA methylation in host-pathogen interaction remains largely unexplored for most of the bacterial plant pathogens.

The methylation pattern of the *E. coli* GATC sites are relatively stable except for a few subset that varies depending on the growth phase and the carbon source (Hale *et al.*, 1994). However, in an entomopathogenic bacterium *Photorhabdus luminescens* TT01, the methylome profile remained unchanged over the course of growth phase. Methylome analyses by SMRT sequencing performed on the *P. luminescens* bacterial cells harvested during mid-exponential phase (OD=0.3-0.4), late exponential phase (OD=0.9), stationary phase (OD=1.5) and late stationary phase (OD>3) revealed that the methylation rate remained unchanged during the different growth phase. The corresponding MTase encoding genes were expressed in all the tested growth conditions (Payelleville *et al.*, 2018). Another study performed on a metal reducing

Figure 2 Chemical structure of the various forms of DNA methylation

The methylated base is highlighted in red bacterium, *Shewanella oneidensis* MR-1 also did not observe major changes in the methylation state when grown in aerobic rich medium or anaerobic minimal medium (Bendall *et al.*, 2013).

1.2.2. Different form of DNA methylation in bacteria

DNA methylation is catalyzed by DNA methyltransferases (MTases). The MTases are classified into two groups, (1) the MTases that act with restriction enzymes (a component of Restriction-Modification system) which protects the genomic DNA from foreign material and (2) the solitary MTases (or orphan MTases) that act without restriction enzymes, which are solely involved in the cellular functions. Examples of orphan MTases include Dam of *E. coli* and in other ɣ-Proteobacteria and CcrM of *Caulobacter cresentus* and in other α-Proteobacteria (Low, Weyand and Mahan, 2001; Wion and Casadesús, 2006). Two classes of MTases catalyze base modification in the DNA: (i) the exocyclic MTases that methylates the exocyclic amino nitrogen yielding 6mA (6-methyladenine) and 4mC (4-methylcytosine) modifications and (ii) the endocyclic MTases that methylates the pyrimidine ring carbon yielding, 5mC (5-methylcytosine) modification. The chemical structure highlighting the modified base is given in Figure 2.

DNA methylation can influence DNA–protein interactions and thereby gene expression, especially when the methylated state of the binding site is important for the interacting proteins such as RNA polymerases or transcription factors (Van Der Woude, Braaten and Low, 1996; Jeltsch, 2002; Casadesus and Low, 2006; Low and Casadesús, 2008; Jones, 2012; Sánchez-Romero, Cota and Casadesús, 2015). In mammals, methylation of the cytosine residue (5mC) in the promotor region was linked to the repression of downstream gene transcription whereas methylation on the gene body has been positively correlated to the gene expression (Jones, 2012). However, the same correlation has not yet been proven in prokaryotes.

DNA MTases target a specific DNA motif for DNA methylation. For example, Dam methylates the adenine of the G**A**<u>**T**</u>C motif. The study that identified 620 MTases from diverse bacterial and archaeal species included 217 bacterial species. They found that 6mA modifications were predominant (75%) while 4mC and 5mC modifications accounted for 20% and 5%

respectively (Blow *et al.*, 2016). Orphan MTases are widely distributed in prokaryotes and they are likely to be evolutionarily more conserved than the RM MTases. Examples of highly represented orphan MTase families other than Dam and CcrM among bacteria includes the gene that methylates the motif G**A**N**T**C of *Methylobacterium* sp., RA**AT**TY of Spirochaetaceae, which is also present in *Campylobacter jejuni*, **T**TA**A** of Arthrobacter (Blow *et al.*, 2016) and the conserved G**T**WW**A**C motif across *Burkholderiaceae* (Erill *et al.*, 2017).

1.2.3 Hemimethylation and its role

Active demethylation has not yet been observed in bacteria and therefore the non-methylation of the motifs is generated as a result of competition between the DNA specific binding protein and the DNA MTase (Sánchez-Romero *et al.*, 2015). Demethylation only occurs during DNA replication. Hemimethylation is a phenomenon that also occurs during DNA replication where only one of the strand (template) harbors the methylated base. DNA hemimethylation has many prominent roles in the biology of bacteria from the occurrence/timing of DNA replication, DNA repair, to the timing of transposition and conjugal transfer of plasmids, which are all sensitive to the hemimethylated state of the specific DNA region (Casadesus and Low, 2006). Interestingly, hemimethylated state of the DNA can be either extended, transient or heritable (stable) (Wion and Casadesús, 2006).

For example in *E. coli* during DNA replication, the protein SeqA binds to the hemimethylated sites near the *oriC* that sequesters the replication initiation. Additionally, the replication can also be delayed when SeqA transiently blocks the DnaA protein synthesis that is important for the initiation of replication by binding to the hemimethylated sites in the *dnaA* promotor. Another example is the recognition of the mismatched base pairs for DNA repair that occurs during replication. Here again the mismatched base pairs are discriminated by the absence of methylation in the newly synthesized strand. This transient hemimethylation in the new strand acts as a signal for strand differentiation by the methyl directed mismatch repair protein MutH that cuts the nonmethylated strand. Thereby, only the methylated DNA template strand will be used as the template for the following DNA synthesis. In a similar manner, DNA hemimethylation can also repress transposition of insertion sequences such as IS10 and IS50 and bacterial transposons such as Tn5 and Tn10 (Low *et al.*, 2001; Casadesus and Low, 2006; Wion and Casadesús, 2006).

All these hemimethylated sites were considered transient and not heritable. However, heritable hemimethylation patterns were observed in at least 50 sites of *E.coli* 'G**A**<u>**T**</u>C' (Wion and Casadesús, 2006). Heritable hemimethylation is involved in the regulation of phase variation of Pap (pyelonephritis-associated pili) and Agn43 (Antigen 43) of *E. coli*, where the hemimethylation is a required step for switching ON/OFF. (Casadesus and Low, 2006; Wion and Casadesús, 2006).

1.2.3. Role of DNA methylation in bacterial virulence

5mC modification of eukaryotes is known to have functional roles in gene expression, chromatin organization and genome maintenance. However, the role of 5mC modification in prokaryotes remains rudimentary. In many groups of Proteobacteria, 6mA modification has an impact in the cell regulatory events including bacterial virulence. Dam regulates a number of virulence genes. In *E. coli,* Dam controls the expression of pilus operons that plays a prominent role in virulence of urinary tract infections. For instance, the Lrp regulator of *E. coli* binds to G**A**<u>**T**</u>C sequences in the promoter region of the *pap* operon. Because of the higher affinity of Lrp for non-methylated G**A**<u>**T**</u>C sites, a competition between the binding of Dam and Lrp occurs, resulting in two sub-populations, the one that expresses Pap pilus and the other does not (Blyn *et al.*, 1990; Casadesus and Low, 2006). Alteration in the levels of Dam methylation in *Salmonella, Vibrio choloerae* and *Yersinia pseudotuberculosis* attenuates virulence in animal models (Julio *et al.*, 2001). Additionally, over production of *dam* inhibits or alters virulence in a number of organisms including *Y. pseudotuberculosis, V. cholerae, Salmonella enterica* and *Photorhabdus luminescens* (Julio *et al.*, 2001, 2002; Payelleville *et al.*, 2017). Dam mutants are also shown to have impaired virulence in *Klebsiella pneumonia* (Mehling et al., 2007) and *Y. pestis* (Robinson *et al*., 2005). Another (6mA) MTase CamA identified on *Clostridioides difficile* (formerly called Clostridium) has

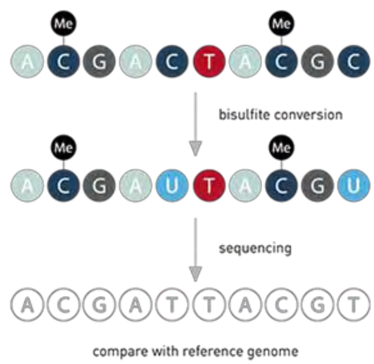


Figure 3 Bisulfite sequencing

The DNA sample treated with bisulfite converts unmethylated cytosine to uracil. Comparison of the test DNA sequence with the reference genome reveals the position of cytosine methylation.

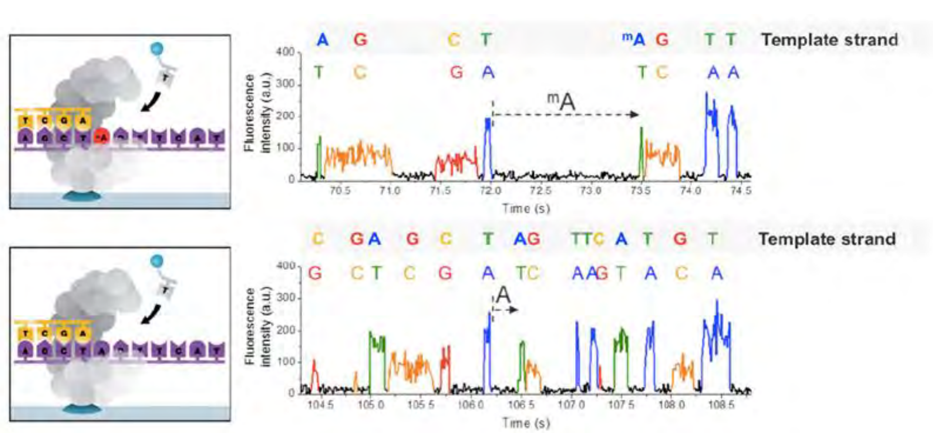


Figure 4 Single-molecule, real-time (SMRT) sequencing

The time required for the incorporation of a nucleotide for the reference genome and the sample genome determines presence of methylated base at any given position.

Gene				MTase Protein			
ID	**gene Name**	**Localization**	**GC%**	**Type**	**protein Name**	**Modification**	**Target site**
RSc0844	***rsoGIM***	Chromosome	68.7	orphan	**M.RsoGI**	-	-
RSc0845	***rsoGIIM***	Chromosome	66.8	orphan	**M.RsoGII**	-	-
RSc0869	***rsoGIIIM***	Chromosome	63.5	orphan	**M.RsoGIII**	m5C	
RSc1982	***rsoGIVM***	Chromosome	67.9	orphan	**M.RsoGIV**	m6A	GTWWAC (Erill et al., 2017)
RSc3396	***rsoGVM***	Chromosome	70.1	RM	**M.RsoGV**	m6A	-
RSc3438	***rsoGVIM***	Chromosome	50.1	orphan	**M.RsoGVI**	m5C	YGCCGGCR
RSp0570-571	***rsoGVIIM***	Megaplasmid	66.7	orphan	**M.RsoGVII**	-	-

Table 1 Mtases and their motifs of GMI1000 as identified by REBASE and SMRT

RM – MTase associated with restriction modification system

W – A or T; Y – C or T; K – G or T; N – any base; R – A or G been found to be involved in the sporulation and biofilm formation (Oliveira and Fang, 2020; Oliveira *et al.*, 2020). In a Gram-positive bacteria *Streptococcus pyogenes*, deletion of Restriction, Specificity, and Methylation gene subunits (ΔRSM strain) responsible for 6mA modification resulted in altered virulence gene expression. The study found that the presence of 6mA modifications influenced the expression of master transcriptional regulator (Mga) that controls various virulence genes, surface adhesions and immune-evasion factors of *S. pyogenes* and the deletion of RSM resulted in increased host immune response (Nye *et al.*, 2020). Natural transformation capacity of *Helicobacter pylori* is crucial in genome diversity, evolutionary potential and virulence regulation. The loss of (the only) 4mC modifications (M2.HpyAII) resulted in reduced transformation capacity, differential expression of virulence genes and reduced ability to induce inflammation of the human cell line (Kumar *et al.*, 2018).

1.2.4. Methods to detect DNA methylation

Over the years, a number of diagnostic tools were developed to detect the various forms of DNA methylation. Historically, whole genome bisulfite sequencing (WGBS) or bisulfite sequencing (BS-Seq) is the most frequently used technique to detect the 5mC modifications. This method involves treatment of the genomic DNA with sodium bisulfite. The treatment deaminates cytosine to uracil followed by sequencing where the uracil (nonmethylated cytosine) is converted to thymidine, while the methylated cytosines are not deaminated and are read as cytosines. The position of the methylated cytosine is identified by comparing the sequences of treated and untreated sequences (Figure 3) (Beck and Rakyan, 2008; Lister and Ecker, 2009; Li and Tollefsbol, 2011).

Single molecule real-time (SMRT) DNA sequencing is an emerging sequencing technique that allows the determination of the genome sequence as well as the methylation profile. The technique monitors the activity of a single DNA polymerase in real time while catalyzing the incorporation of fluorescently labelled nucleotides complementary to the template DNA strand. A pulse of fluorescence (whose color identifies each nucleotide) detects incorporation of the nucleotide. In addition to monitoring the fluorescence pulse for each nucleotide incorporation, the device measures the time between each pulse known as the interpulse duration (IPD). The principle behind the SMRT sequencing for methylome analysis is that the IPD is statistically longer when the template contains a methylated base and has distinct kinetic signatures for 4mC and 6mA modifications. As a result, the methylated nucleotides can be identified by the difference in the IPD ratio between the template (methylated) DNA and the PCR amplified (unmethylated) DNA at any given position (Figure 4). However, this approach is not sensitive to 5mC modifications (Flusberg *et al.*, 2010; Davis, Chao and Waldor, 2013; Blow *et al.*, 2016).

SMRT sequencing of the reference strain GMI1000 detected 3 motifs, the G**T**WW**A**C, CCCAKNAVCR and YGCCGGCRY motifs targeted by the MTases RSc1982 (*rsoGIVM*), while remaining motifs putatively correspond to RSc0869 (*rsoGIIIM*) and RSc3438 (*rsoGVIM*) respectively as given in Table 1 (Erill *et al.*, 2017).

1.3. A model organism: the *Ralstonia solanacearum* species complex (RSSC)

The multihost vascular plant pathogen, *R. solanacearum* has a strikingly wide host range of more than 250 plant species making this pathogen rank among the top 10 plant pathogenic bacteria. It causes brown rot disease in potato, bacterial wilt (BW) in *Solanaceae* family and many other crops including some ornamentals and Moko disease in banana (Genin and Boucher, 2002a; Peeters, *et al.*, 2013). *R. solanacearum* is regarded as a 'species complex' owing to its significant genetic diversity within the species. A species complex defines a cluster of closely related isolates whose individual members may represent more than one species (Genin, 2010; Peeters *et al.*, 2013). This extensive phytopathogen inhabits and curbs the production of diverse crop species including both monocots and dicots in tropical, subtropical and some warm temperate regions of Europe (Castillo and Greenberg, 2007). A root infecting, vascular colonizing pathogen with an extensive host range that rapidly evolves and exhaustively investigated makes

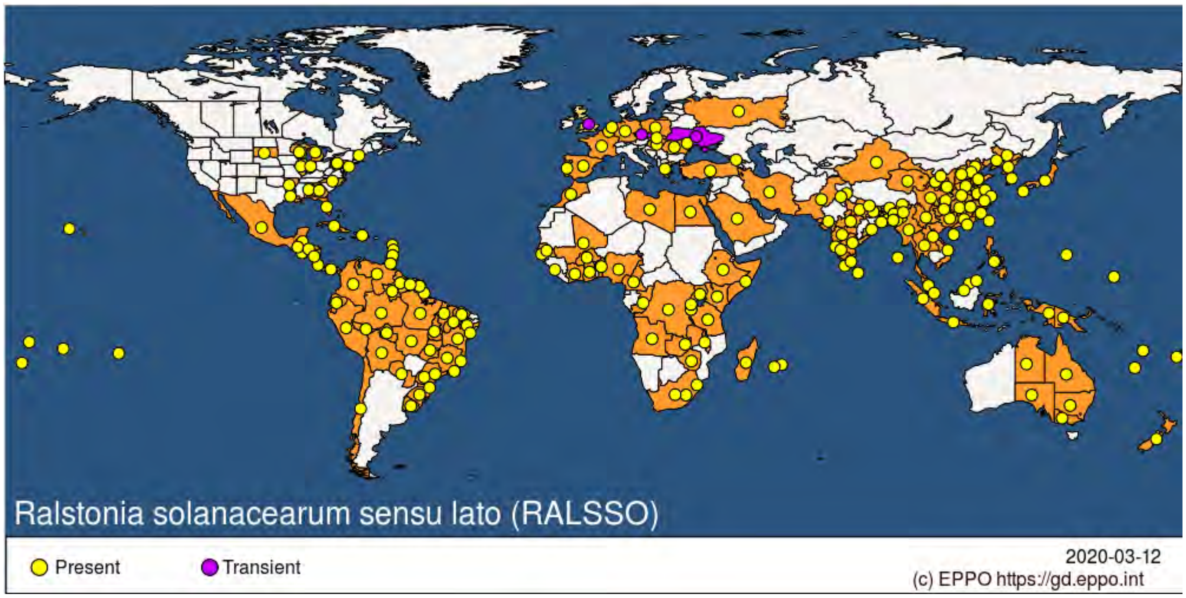


Figure 5 Distribution of Ralstonia solanacearum species complex

Figure obtained from European Mediterranean Plant Protection organization - EPPO Global Database (https://gd.eppo.int/taxon/RALSSO/distribution). Accessed on 20/03/2020 the *R. solanacearum* species complex (RSSC) an interesting model to study bacterial pathogenicity.

1.3.1. Extensive host range and worldwide distribution

The host range of *R. solanacearum* not only includes major agricultural crops like tomato, potato, eggplant, groundnut, ginger, capsicum, banana but also model plants like *Arabidopsis thaliana, Nicotiana benthamiana* and *Medicago truncatula* and ornamental plants like *Pelargonium* spp. (Genin and Boucher, 2002b; Castillo and Greenberg, 2007; Genin, 2010) and quite recently in *Rosa* spp. (Tjou-Tam-sin *et al.*, 2017). Annually, the economic losses in potato accounts up to one billion US dollars (Elphinstone, 2005; Mansfield *et al.*, 2012). However, the direct yield losses vary depending on various factors like the host, cultivar, climate, soil type, cropping pattern and the strain. The yield loss due to bacterial wilt can vary from 0 to 91% in tomato, 33 to 90% in potato, 10 to 30% in tobacco, 80 to 100% in banana and up to 20% in groundnut (Yuliar et al. 2015; Sharma and Singh 2019).

In particular, RSSC is a versatile organism as it has the ability to rapidly evolve and adapt to various host plants, which is supported by field observations reporting the emergence of strains more aggressive and able to colonize novel hosts. The best-known example is the emergence of phylotype II sequevar 4 strains, which are not pathogenic on banana (IIB-4NPB) as opposed to its associated members of the group (Moko disease causing cluster), but are pathogenic to solanaceous crops and cucurbits in Martinique (Wicker *et al.*, 2007). The phylotype IIB-4NPB strains were also isolated from Brazil (Cellier *et al.*, 2012). This was followed by the emergence of strains in unusual hosts like anthurium (*Anthurium andreanum*) in 1999; cantaloupe (*Cucumis melo*. L) in 2001; cucumber (*Cucumis sativus* L.), pumpkin (*Cucurbita moschata* Ls.), zucchini (*Cucurbita pepo* L.) in 2002; and watermelon (*Citrullus lunatus*) in 2003 (Wicker *et al.*, 2007, 2009). Quite recently, RSSC outbreaks were also reported in various parts of Europe that includes Bulgaria, France, Germany, Italy, Netherlands, Poland, Portugal and Spain in 2017 and 2018 on *Solanum tuberosum*, *Rosa* spp. and *Solanum lycopersicon*

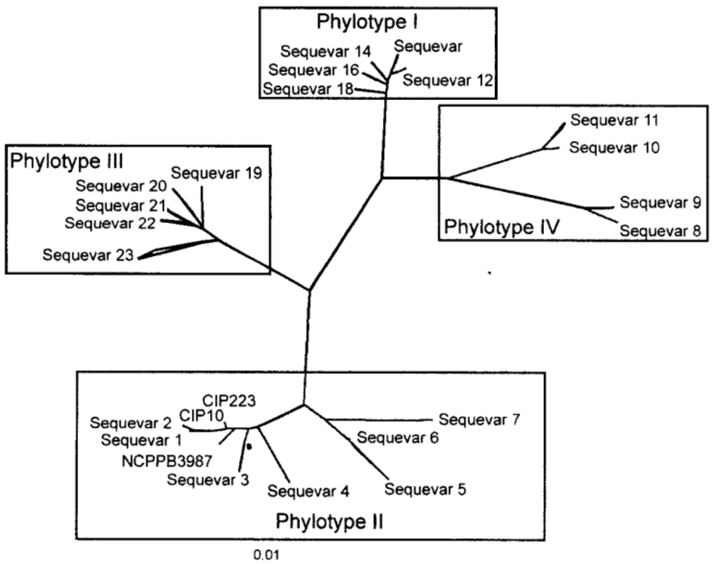


Figure 6 Phylogenetic analysis of RSSC based on egl sequence

Phylogeny of RSSC genomes based on the partial egl sequence analysis dividing into phylotypes and sequevars (Fegan and Prior, 2005).

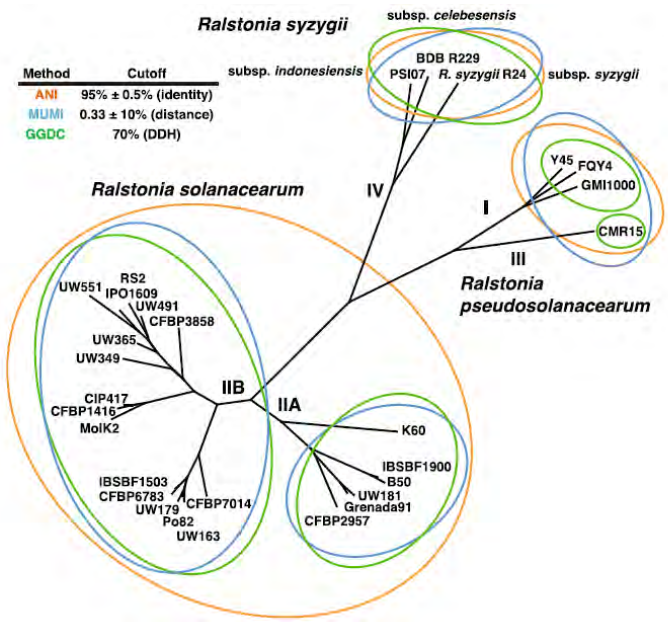


Figure 7 Phylogenetic reclassification of RSSC strains

The revised classification of RSSC genomes. The three subdivisions are as follows: (i) *Ralstonia pseudosolanacearum* comprising the strains from phylotype I and III; (ii) *Ralstonia solanacearum* comprising strains from phylotype IIA and IIB and (iii) *Ralstonia syzygiii* comprising the subsp. *indonesiensis*, subsp. *celebesensis* and subsp. *syzygii* (Prior et al., 2016)

(EFSA PLH Panel. 2019). The current distribution of the bacterial wilt worldwide is presented in Figure 5 (EPPO Global Database, 2020). In addition to the highly adaptive nature of RSSC, the anticipated increase in temperature due to global warming favors bacterial wilt distribution across the world. This could be a serious threat to many host and non-host plants (Lopes and Rossato, 2018).

1.3.2. A species complex

In the past two decades, classification of RSSC has undergone many changes. Formerly, it was classified into five races and six biovars based on the host range and biochemical properties. The classification later evolved into phylotype-sequevar system (Figure 6). This hierarchical classification was based on the partial sequence analyses of endoglucanase (*egl), mutS,* and *hrpB* genes or ITS (intergenic spacer) region and comparative genome hybridization (Prior and Fegan, 2005; Guidot et al., 2007). Phylogenetic analysis of RSSC based on the sequence data unveiled four phylotypes. Each phylotype corroborates with the geographical origin of the strains but not to the host range. There are subgroups within each phylotypes called sequevars, based on the *egl* gene sequence, that are clusters of isolates with similar pathogenicity or common geographical origin (Prior and Fegan, 2005). Phylotype I consists of strains primarily from Asia, phylotype II from America with two monophyletic subpopulations (IIa and IIb), phylotype III from Africa and phylotype IV from Indonesia, Australia and Japan. RSSC also includes *Ralstonia syzgii* that causes Sumatra disease of clove trees in Indonesia and *Ralstonia celebensis* (banana blood disease bacteria (BDB)) which belongs to phylotype IV (Prior and Fegan, 2005; Castillo and Greenberg, 2007; Genin, 2010; Cellier *et al.*, 2012; Mansfield *et al.*, 2012; Wicker *et al.*, 2012; Peeters *et al.*, 2013).

Advancement in molecular tools and the availability of whole genome sequences of many RSSC strains suggested evolutionary divergence among and within each phylotypes. In order to improve the RSSC taxonomy, a polyphasic taxonomic approach was proposed combining the available phenotypic and genotypic data. The proposed classification divides RSSC strains into

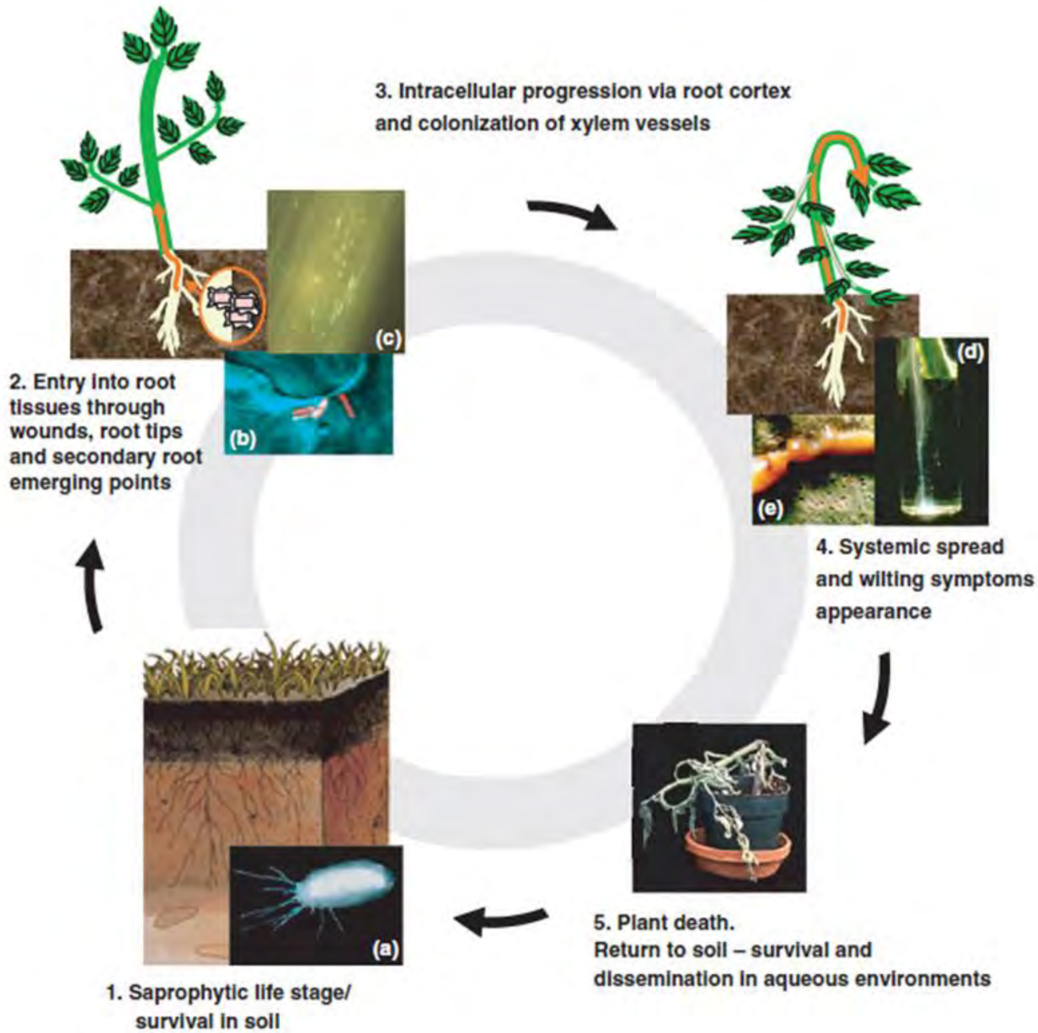


Figure 8 Life cycle of Ralstonia solanacearum

R. solanacearum can survive for long periods in the soil without a true host. Bacteria enters the plant via the roots and invade the vascular system thereby causing wilting symptoms, which is favored by rapid multiplication in the xylem vessels and excessive production of exopolysaccharides that block the xylem flow leading to plant death. (a) Transmission electron microscopic view of GMI1000 (b) Confocal view of bacteria (in red) attached to the plant surface (c) Green fluorescent protein expressing bacteria visualized on the surface of tomato root (d) Oozing bacteria from the infected plant stem to the water (e) Bacteria ooze from the infected plant stem (Genin, 2010) three genospecies based on the phylogenetic analyses of ITS region, partial *egl* sequences and DNA hybridizations. They are as follows: (1) ***Ralstonia solanacearum*** comprising only phylotype II strains; (2) ***Ralstonia syzygii*** comprising only phylotype IV strains and (3) ***R. pseudosolanacearum*** comprising phylotype I and phylotype III strains (Safni *et al.*, 2014). This taxonomic revision was later validated by a combination of genomic and proteomic approaches (Prior *et al.*, 2016) (Figure 7).

1.3.3. The RSSC life cycle

R. solanacearum is a soil-borne Gram-negative bacterium that alternates between two states: as a saprophyte in soil/water and as a pathogen inside the plants Figure 8. The bacteria can survive in the soil for many years and spread through water, rhizosphere contact and farming (Genin and Boucher, 2004; Genin, 2010; Song *et al.*, 2018; Castillo and Agathos, 2019). The pathogen enters plant roots, invades the xylem vessels and spreads rapidly to aerial parts of the plant through the vascular system. Within a few days of infection, the bacteria attains high population by extensive colonization that leads to vascular clogging causing wilting symptoms and ultimately plant death (Genin, 2010; Peeters, Guidot, *et al.*, 2013).

1.3.4. The RSSC genome

The notable feature of RSSC is its high level of diversity, which could be explained by the fact that it possesses a state of natural competence to integrate genetic material after horizontal gene transfer (Bertolla *et al.*, 1999; Guidot *et al.*, 2009). At the genomic level, *R. solanacearum* strains consists of two circular replicons called the chromosome (around 3.7 Mb) and the megaplasmid (around 2.1 Mb) with an average of 67% G+C content. GMI1000 has a natural competence that favors *in vitro* transformation by HGT (Bertolla *et al.*, 1999). The high number of mobile genetic elements and ACURS (Alternative Codon Usage Regions) in both replicons exemplifies the complexity and the plasticity of the genome. ACURS account for a considerable portion of the genome (7%) and they are commonly associated with genes of lower G+C content and probable pathogenicity genes (Salanoubat et al. 2002; Guidot et al. 2007).

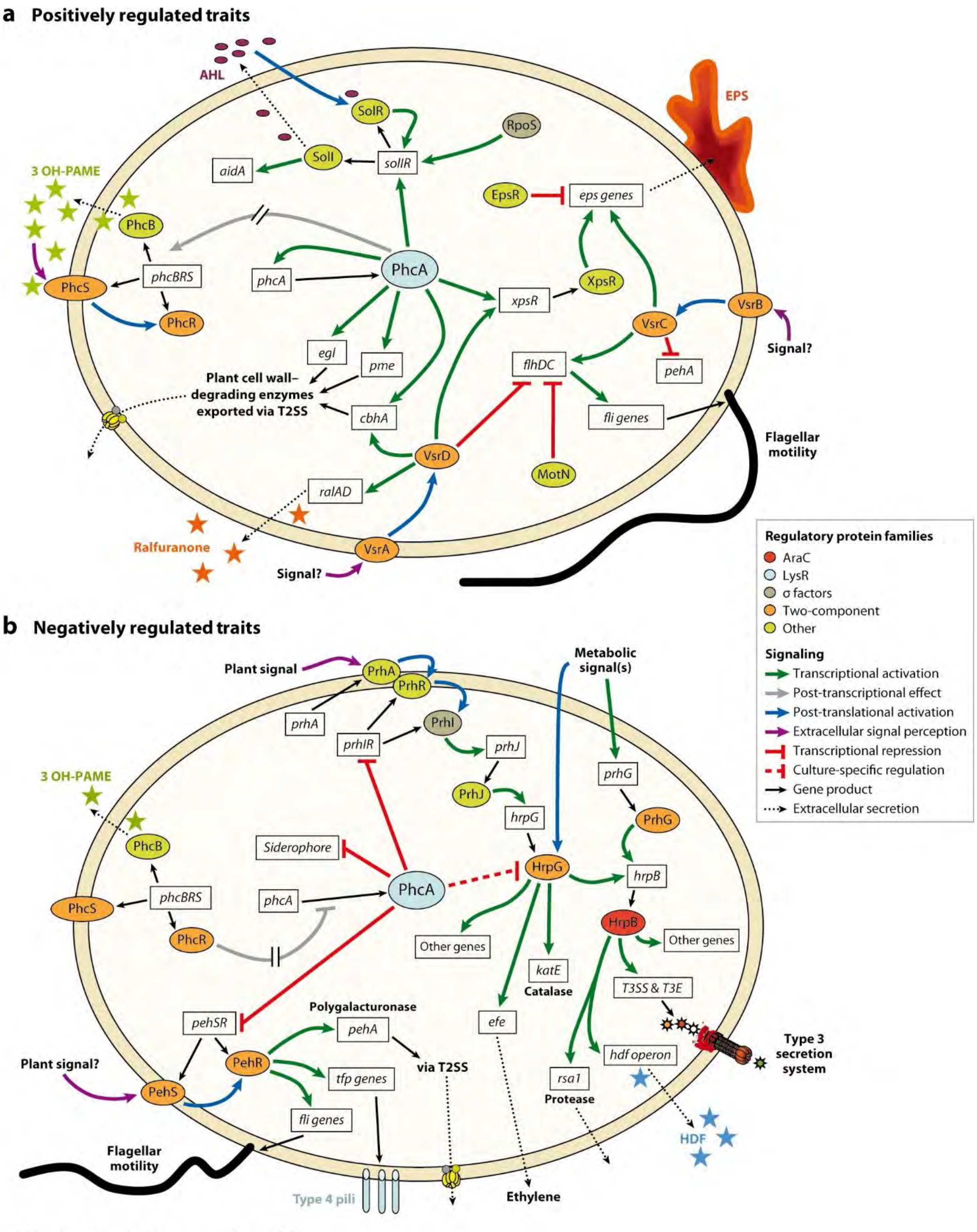
The chromosome encodes all the fundamental mechanisms essential for the bacterial survival such as the genes required for DNA replication, repair and cell division, transcription and translation, but it also contains genes involved in virulence. The megaplasmid contains genes that contribute to the overall fitness and adaptation of the bacteria to various hosts/environmental conditions, but it also contains housekeeping genes. For example, several metabolically essential genes and key virulence genes are present in both the chromosome and the megaplasmid (Genin and Boucher, 2002a; Castillo and Greenberg, 2007; Genin and Denny, 2012; Castillo and Agathos, 2019).

1.3.5. Pathogenicity determinant and regulation

The most important virulence factors of RSSC includes exopolysaccharide biosynthesis (EPS), type III secretion system (T3SS), type II secretion system (T2SS), flagellin and motility related genes (Genin and Denny, 2012).

RSSC produces large amounts of EPS *in planta* as well as on plates. This promotes rapid colonization of the plant stems that lead to completely wilted plants. Mutants lacking *eps* seldom cause wilting symptoms even when introduced directly on the plant stem (Schell, 2000).

The T3SS is a major pathogenicity determinant of RSSC. It is common to many plant and animal pathogenic bacteria. T3SS is a syringe like membrane that injects type 3 effector (T3E) proteins into the plant cell causing infection in susceptible plants or hypersensitive response (HR) in resistant plants (Peeters *et al.*, 2013; Wu *et al.*, 2015). The T3SS machinery is known to encode at least 70 effector proteins in GMI1000 (Peeters *et al.*, 2013). Deletion of a single effector does not show any major effect, which contemplates the synergetic and redundant characteristics of type 3 effectors in RSSC. However, there are some exceptions as seen with GALA7 effector. *gala7* belongs to a seven-gene family that encodes F-box and leucine rich repeat (LRR) domain T3Es. Disruption of this T3SS virulence factor affects the pathogenicity of RSSC strain GMI1000 in



AR Genin S, Denny TP. 2012.
Annu. Rev. Phytopathol. 50:67–89

Figure 9 Major virulence regulatory networks in RSSC strains

The model mainly describes the genes and traits controlled both positively and negatively by the transcriptional regulator, PhcA. The network and the genes that are positively regulated by PhcA, VsrA and VsrD are given in a. The primary regulation of T3SS and other virulence-associated genes are depicted in b. Broken lines represent probable indirect regulation. Lines ending with crossbar denote repression and lines with arrowhead denote activation. EPS – exopolysaccharide; HDF – hrpB-dependent diffusible factor (3-hydroxy-oxindole); T2SS – type II secretion system; T3SS – type III secretion system. (Genin and Denny, 2012)

EfpR is the only regulator that is not yet in this network but we know that its regulon is close to the PhcA regulon according to transcriptomic analyses (Perrier *et al.,* 2016, 2019)

Medicago truncatula. This implies that GALA7 effector has a direct host specific role in the colonization of *M. truncatula*. Interestingly, the virulence of single GALA mutants did not affect the susceptible *Arabidopsis* and tomato plants (Angot *et al.*, 2006).

Similar to T3SS, T2SS of RSSC is another salient pathogenicity determinant, which contributes to the systemic infection of the pathogen in the xylem vessels and in the virulence. The T2SS secrete approximately 30 substrates, which include several cell wall degrading enzymes and other proteins such as endoglucanase (Egl), PehA, PehB, PehC, Pme that are all important for colonization and wilt symptoms (Poueymiro and Genin, 2009; Genin and Denny, 2012; Cianciotto and White, 2017).

Motility also contributes significantly to the pathogen colonization and disease. Flagellin and motility related genes that are shown to be involved in swimming/twitching motility, biofilm formation and root attachment, contributes to virulence on tomato (Peeters *et al.*, 2013). It has also been shown that motility is needed for the pathogen in the early stages of infection i.e. for locating and invading the roots; but are non-essential in the later stages of pathogenesis (Genin and Denny, 2012).

RSSC pathogenesis is regulated by a complex regulatory network. The major regulators of this network are PhcA, VsrA, VsrD, HrpB, HrpG and EfpR (Figure 9). The global virulence regulator PhcA has a central role in this network as it controls directly or indirectly many virulence genes. Levels of functional PhcA are regulated in response to bacterial cell density by a quorum sensing (QS) mechanism and allows transitions from saprophytic to parasitic forms of the pathogen (Peyraud *et al.*, 2016). PhcA is a LysR family transcription regulator that differentially regulates many genes (around 1500 genes) (Perrier *et al.*, 2018). The functions that are regulated by PhcA include EPS synthesis, plant cell wall degrading enzymes, bacterial motility, T3SS and type 6 secretion system (T6SS) (Yoshimochi *et al.*, 2009; Khokhani *et al.*, 2017; Perrier *et al.*, 2018) (Figure 9).

VsrA - VsrD is a two-component response regulator that controls multiple traits and strongly plays a role in the colonization of RSSC in tomato stems. VsrAD also negatively regulates swimming and twitching motility (Genin and Denny, 2012; Peeters, Guidot, *et al.*, 2013).

Yet another major regulatory system in RSSC is the HrpB regulon that mainly controls the various pathogenicity related genes needed to repress the plant defense responses in order to survive the initial stages of infection (Genin, 2010). HrpB, an AraC family transcription regulator affects the expression of more than 180 genes in minimal medium (that mimics the plant xylem). Importantly, HrpB positively regulates most of the T3SS genes and T3E. HrpG is an OmpR family response regulator that positively regulates the *hrpB* expression. It is known that HrpG facilitates adaptation of the bacteria inside the hosts. Also studies show that *hrpB* and *hrpG* mutants are non-pathogenic (Genin and Denny, 2012). Other important proteins secreted by Hrp secretion system includes PopA, PopB and PopC of which PopA is shown to have HR like response when infiltrated in plants at high concentrations, whereas PopB and PopC are involved in the transport of the nutrients from plants or the effectors to plants (Genin and Boucher, 2002b).

One other major virulence regulator is EfpR, which acts as global catabolic repressor in RSSC that controls several metabolic pathways. The transcriptome of *efpR* deleted mutant shows 1031 differentially expressed genes (combined results from Perrier and Capela's article) (in the annexure) (Capela *et al.*, 2017; Perrier *et al.*, 2018). Deletion of *efpR* gene enhances the fitness of the bacteria during growth *in planta* and generates phenotypic heterogeneity. Some of the key genes that were differentially expressed includes the virulence genes such as T3Es, Hrp proteins, EPS synthesis, flagellar proteins, metabolic functions, motility and chemotaxis related genes, thus largely overlapping the PhcA regulon (Capela *et al.*, 2017; Perrier *et al.*, 2018).

efpR mutants produce two distinct type of colonies namely the type S (smooth variant, identical to the WT) and type EV (*efpR* variant) (Perrier *et al.*, 2019). There is a reduced EPS production in *efpR* – EV types and thereby reduced virulence. Swimming and twitching motility are also enhanced in the *efpR* variants (Perrier *et al.*, 2016; Capela *et al.*, 2017). Furthermore, the *efpR* variants can better catabolize a number of carbon substrates like L-Glutamate, L-Proline, L-Histidine and GABA when compared to the wild type GMI1000. Additionally, the growth rates were also significantly higher with these substrates in comparison to GMI1000 (Perrier *et al.*, 2016, 2019).

R. solanacearum strains, Core-Rs2 (phylotype-sequevar)[b]

Line[a]	GMI1000 I-18 W	GMI1000 I-18 CI	GMI1000 I-18 P	CMR134 I-13 W	CMR134 I-13 CI	CMR134 I-13 P	CFBP3059 III-23 W	CFBP3059 III-23 CI	CFBP3059 III-23 P	CMR32 III-29 W	CMR32 III-29 CI	CMR32 III-29 P	CFBP2957 IIA-36 W	CFBP2957 IIA-36 CI	CFBP2957 IIA-36 P	CMR39 IIA-41 W	CMR39 IIA-41 CI	CMR39 IIA-41 P
T1	0.0	26.7	2	60.0	60.0	4	63.3	93.3	4	13.3	13.3	1	26.7	36.7	2	26.7	56.7	3.1
T2	90.0	93.3	5	33.3	53.3	3.1	26.7	76.7	3.2	33.3	33.3	2	13.3	36.7	2	50.0	90.0	4
T3	66.7	76.7	4	16.7	70.0	3.2	3.3	100.0	3.2	3.3	3.3	1	3.3	33.3	2	10.0	83.3	3.2
T4	73.3	83.3	4	43.3	73.3	4	66.7	96.7	4	26.7	26.7	2	0.0	16.7	1	6.7	33.3	2
T5	0.0	6.7	1	0.0	10.0	1	20.0	96.7	3.2	6.7	6.7	1	3.3	30.0	2	0.0	16.7	1
T6	0.0	6.7	1	6.7	40.0	2	80.0	96.7	5	13.3	13.3	1	30.0	46.7	3.1	3.3	46.7	2
T7	0.0	46.7	2	10.0	30.0	2	53.3	100.0	4	13.3	13.3	1	0.0	33.3	2	36.7	73.3	3.2
T8	0.0	10.0	1	3.3	13.3	1	70.0	100.0	4	33.3	33.3	2	0.0	40.0	2	0.0	36.7	2
T9	0.0	26.7	2	13.3	40.0	2	3.3	100.0	3.2	6.7	6.7	1	16.7	43.3	2	0.0	43.3	2
T10	90.0	96.7	5	96.7	96.7	5	100.0	100.0	5	23.3	70.0	3.2	86.7	100.0	5	96.7	100.0	5
Path		type T-2			type T-2			type T-3			type T-1			type T-1			type T-2	
E1	**0.0**	**0.0**	1	6.7	6.7	1	**0.0**	**0.0**	1	**0.0**	**0.0**	1	**0.0**	**0.0**	1	3.3	3.3	1
E2	0.0	16.7	1	0.0	6.7	1	6.7	23.3	2	**0.0**	**0.0**	1	**0.0**	**0.0**	1	6.7	10.0	1
E3	0.0	30.0	2	3.3	6.7	1	6.7	40.0	2	3.3	3.3	1	0.0	3.3	1	0.0	3.3	1
E4	0.0	10.0	1	0.0	3.3	1	0.0	6.7	1	**0.0**	**0.0**	1	**0.0**	**0.0**	1	0.0	33.3	2
E5	13.3	26.7	2	36.7	46.7	3.1	10.0	40.0	2	6.7	13.3	1	**0.0**	**0.0**	1	16.7	16.7	2
E6	43.3	43.3	3.1	13.3	23.3	2	60.0	93.3	4	**0.0**	**0.0**	1	3.3	26.7	2	3.3	10.0	1
E7	20.0	63.3	3.2	63.3	66.7	4	83.3	86.7	5	6.7	6.7	1	3.3	3.3	1	26.7	56.7	3.1
E8	90.0	93.3	5	70.0	70.0	4	100.0	100.0	5	43.3	43.3	3.1	66.7	80.0	4	73.3	73.3	4
E9	0.0	73.3	3.2	20.0	26.7	2	0.0	13.3	1	10.0	10.0	1	**0.0**	**0.0**	1	**0.0**	**0.0**	1
E10	60.0	66.7	4	100.0	100.0	5	100.0	100.0	5	46.7	56.7	4	30.0	40.0	2	66.7	73.3	4
Path		type E-2			type E-2			type E-2			type E-1			type E-1			type E-1	
P1	0.0	96.7	3.2	46.7	83.3	4	70.0	100.0	4	3.3	40.0	2	0.0	43.3	2	0.0	33.3	2
P2	0.0	90.0	3.2	20.0	80.0	3.2	60.0	100.0	4	40.0	83.3	4	0.0	23.3	2	0.0	73.3	3.2
P3	33.3	90.0	3.2	86.7	100.0	5	90.0	100.0	5	0.0	6.7	1	0.0	6.7	1	6.7	23.3	2
P4	0.0	90.0	3.2	26.7	33.3	2	46.7	63.3	4	3.3	3.3	1	0.0	16.7	1	50.0	56.7	4
P5	0.0	80.0	3.2	10.0	13.3	1	0.0	50.0	2	0.0	23.3	2	0.0	6.7	1	10.0	33.3	2
P6	0.0	50.0	2	0.0	53.3	3.2	6.7	73.3	3.2	0.0	23.3	2	**0.0**	**0.0**	1	**0.0**	**0.0**	1
P7	0.0	90.0	3.2	26.7	46.7	3.1	13.3	63.3	3.2	0.0	10.0	1	0.0	3.3	1	16.7	20.0	2
P8	0.0	96.7	3.2	16.7	46.7	2	16.7	73.3	3.2	**0.0**	**0.0**	1	0.0	10.0	1	0.0	3.3	1
P9	0.0	80.0	3.2	3.3	26.7	2	3.3	80.0	3.2	0.0	20.0	2	0.0	6.7	1	**0.0**	**0.0**	1
P10	63.3	83.3	4	90.0	93.3	5	93.3	100.0	5	13.3	33.3	2	0.0	33.3	2	26.7	33.3	2
Path		type P-2			type P-2			type P-2			type P-1			type P-1			type P-1	
Prof		profile-a			Profile-a			profile-b			profile-c			profile-c			profile-c	

Line[a]	PSS366 I-15 W	PSS366 I-15 CI	PSS366 I-15 P	PSS358 I-15 W	PSS358 I-15 CI	PSS358 I-15 P	PSS4 I-15 W	PSS4 I-15 CI	PSS4 I-15 P	CMR15 III-29 W	CMR15 III-29 CI	CMR15 III-29 P	CMR34 IIB-1 W	CMR34 IIB-1 CI	CMR34 IIB-1 P	CFBP6783 IIB-4 NPB W	CFBP6783 IIB-4 NPB CI	CFBP6783 IIB-4 NPB P
T1	100.0	100.0	5	50.0	60.0	4	90.0	93.3	5	80.0	83.3	4	96.7	100.0	5	100.0	100.0	5
T2	100.0	100.0	5	70.0	70.0	4	100.0	100.0	5	93.3	96.7	5	100.0	100.0	5	100.0	100.0	5
T3	83.3	86.7	5	80.0	83.3	4	90.0	96.7	5	96.7	100.0	5	90.0	100.0	5	100.0	100.0	5
T4	93.3	100.0	5	63.3	70.0	4	86.7	90.0	5	86.7	93.3	5	96.7	100.0	5	100.0	100.0	5
T5	93.3	100.0	5	10.0	40.0	2	6.7	16.7	2	70.0	86.7	4	96.7	100.0	5	100.0	100.0	5
T6	73.3	83.3	4	33.3	73.3	4	33.3	46.7	3.1	73.3	96.7	4	96.7	100.0	5	100.0	100.0	5
T7	50.0	50.0	4	23.3	33.3	2	66.7	73.3	4	93.3	93.3	5	100.0	100.0	5	100.0	100.0	5
T8	30.0	40.0	2	36.7	50.0	3.1	13.3	16.7	2	100.0	100.0	5	100.0	100.0	5	100.0	100.0	5
T9	46.7	46.7	4	56.7	60.0	4	70.0	80.0	4	100.0	100.0	5	90.0	100.0	5	96.7	100.0	5
T10	100.0	100.0	5	100.0	100.0	5	100.0	100.0	5	90.0	96.7	5	100.0	100.0	5	100.0	100.0	5
Path		type T-4			type T-3			type T-4			type T-5			type T-5			type T-5	
E1	16.7	33.3	2	13.3	20.0	2	10.0	23.3	2	26.7	36.7	2	6.7	10.0	1	63.3	63.3	4
E2	60.0	63.3	4	16.7	20.0	2	53.3	56.7	4	36.7	56.7	3.1	3.3	76.7	3.2	**0.0**	**0.0**	1
E3	80.0	83.3	4	76.7	86.7	4	23.3	36.7	2	13.3	20.0	2	0.0	46.7	2	66.7	66.7	4
E4	3.3	3.3	1	6.7	13.3	1	13.3	16.7	2	70.0	80.0	4	6.7	36.7	2	26.7	33.3	2
E5	76.7	83.3	4	53.3	63.3	4	46.7	46.7	4	76.7	80.0	4	26.7	80.0	3.2	23.3	33.3	2
E6	0.0	3.3	1	0.0	23.3	2	100.0	100.0	5	96.7	96.7	5	93.3	100.0	5	56.7	63.3	4
E7	50.0	80.0	4	63.3	63.3	4	96.7	100.0	5	76.7	86.7	4	36.7	60.0	3.1	36.7	50.0	3.1
E8	70.0	76.7	4	93.3	93.3	5	100.0	100.0	5	96.7	96.7	5	100.0	100.0	5	66.7	70.0	4
E9	16.7	16.7	2	6.7	30.0	2	26.7	36.7	2	43.3	70.0	4	0.0	10.0	1	63.3	73.3	4
E10	70.0	76.7	4	90.0	90.0	5	96.7	96.7	5	96.7	96.7	5	96.7	96.7	5	90.0	90.0	5
Path		type E-3			type E-3			type E-6			type E-6			type E-5			type E-4	
P1	6.7	73.3	3.2	3.3	6.7	1	40.0	83.3	4	0.0	30.0	2	0.0	93.3	3.2	100.0	100.0	5
P2	3.3	23.3	2	6.7	13.3	1	73.3	80.0	4	0.0	66.7	3.2	6.7	70.0	3.2	53.3	63.3	4
P3	6.7	30.0	2	0.0	6.7	1	70.0	83.3	4	0.0	96.7	3.2	0.0	76.7	3.2	86.7	96.7	5
P4	10.0	16.7	2	**0.0**	**0.0**	1	10.0	10.0	1	43.3	76.7	4	23.3	73.3	3.2	100.0	100.0	5
P5	0.0	6.7	1	**0.0**	**0.0**	1	23.3	30.0	2	0.0	13.3	1	3.3	33.3	2	66.7	73.3	4
P6	0.0	13.3	1	**0.0**	**0.0**	1	3.3	6.7	1	0.0	13.3	1	0.0	30.0	2	63.3	66.7	4
P7	**0.0**	**0.0**	1	**0.0**	**0.0**	1	6.7	16.7	2	0.0	13.3	1	0.0	20.0	2	96.7	100.0	5
P8	3.3	40.0	2	**0.0**	**0.0**	1	6.7	16.7	2	**0.0**	**0.0**	1	3.3	56.7	3.2	90.0	96.7	5
P9	0.0	3.3	1	**0.0**	**0.0**	1	3.3	10.0	1	**0.0**	**0.0**	1	0.0	23.3	2	100.0	100.0	5
P10	26.7	73.3	3.2	0.0	16.7	1	53.3	60.0	4	0.0	66.7	3.2	20.0	80.0	3.2	100.0	100.0	5
Path		type P-1			type P-1			type P-2			type P-1			type P-2			type P-3	
Prof		profile-d			profile-d			profile-e			profile-e			profile-e			profile-f	

[a] Species and lines. T = pathotype on tomato, E = pathotype on eggplant, and P = pathotype on pepper. Path = pathotypes and Prof = pathoprofile based on virulence to Core-TEP.
[b] W = wilted (%), CI = colonization index (%), and P = phenotype, identified by fuzzy analysis clustering and pathoprofiles and pathotypes by agglomerative hierarchical nesting classification an average linkage method. 1 = highly resistant, 2 = moderately resistant, 3.1 = partially resistant, 3.2 = latent infection, 4 = moderately susceptible, 5 = highly susceptible. Number 0.0 in bold signaled the complete incompatible interaction.

Table 2 Phenotypic interaction between plant accessions and Ralstonia solanacearum.

T1-T10 – tomato accessions (T5 – Tomato Hawaii 7996); E1-E10 – eggplant accessions and P1-P10 – pepper accessions. W – wilted %; CI – colonization index %; P – phenotype (1- highly resistant, 2 – moderately resistant, 3.1 – partially resistant, 3.2 – latent infection, 4 – moderately susceptible, 5 – highly susceptible) (Leabeau et al., 2010)

1.3.6 Control measures of bacterial wilt

There are various strategies used to control bacterial wilt of RSSC in the field. This includes physical, chemical, biological as well as cultural practices. The ideal option is the use of healthy seeds and clean irrigation water free of RSSC; however, it is difficult to avoid the dissemination of the pathogen as it can survive for years without a susceptible host. Biological control agents (BCAs) are effective in terms of potentially self-sustaining and disease suppression for longer term in an environmental friendly manner. However, BCAs are unreliable due to their inconsistent colonization and performance in the field (Huet, 2014; Yuliar *et al*., 2015).

The use of resistant cultivars is therefore a favored strategy to control the bacterial wilt disease. In tomato, a reference and widely used resistant cultivar is Hawaii 7996 tomato cultivar. This cultivar was shown to have high level of resistance to several RSSC strains (Lebeau *et al.*, 2010). Lebeau and associates conducted an experiment to determine the bacterial wilt resistance in ten variety of tomato, eggplant and pepper plants against a total of ten strains belonging to diverse phylotypes (except phylotype IV). The strains chosen were GMI1000, IPO1609, CMR15, PS107, CFBP2957, PSS4, JT516, CFBP6784, CFBP6783 and CIR02-080. The tomato accessions included the reference line Hawaii 7996 (=T5) (Table 2). The virulence test of GMI1000 on tomato Hawaii 7996 showed a wilting percentage of 0% and colonization index of 6.7%. None of the accession of tomato, pepper and eggplant showed complete resistance against all strains. Nevertheless, a wide spectrum of high resistance was observed in two accessions of tomato including Hawaii 7996 (=T5), three accessions of pepper and three accessions of eggplant. Overall, bacterial wilt resistance was high in eggplant and pepper, but moderate in tomato (Table 2). In tomato, the only accession to provide complete resistance to more than two strains was Hawaii 7996. In addition, clustering analysis based on the phenotypic scores of the virulence ranked Hawaii 7996 as the most resistant (Lebeau *et al.*, 2010). This strong resistance capacity of Hawaii 7996 makes it an interesting host to identify the evolutionary potential of *R. solanacearum* to overcome plant resistance during experimental evolution.

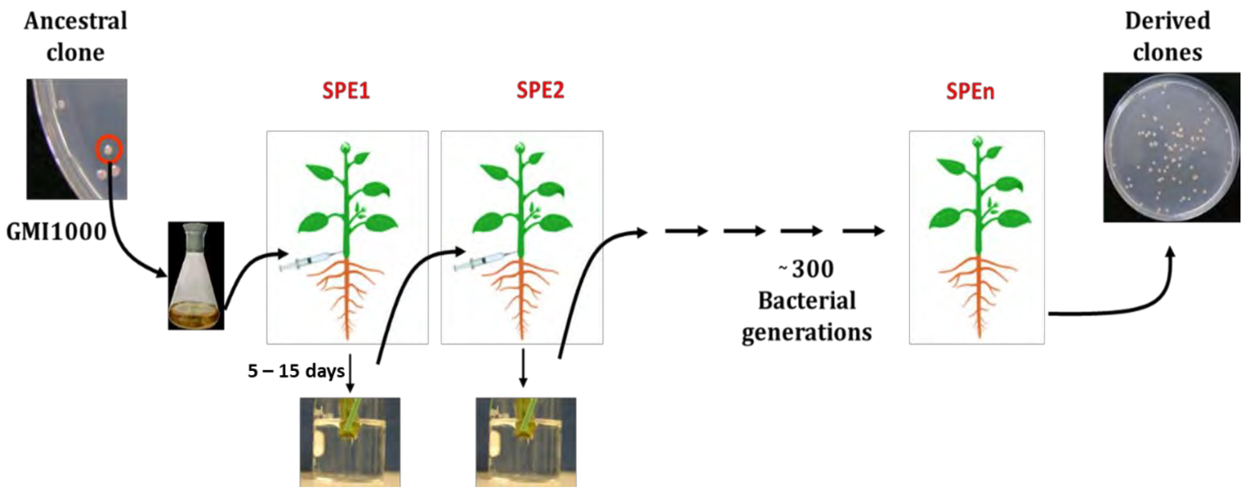


Figure 10 Experimental evolution of Ralstonia solanacearum

Serial passage experiment (SPE) of a single colony of GMI1000 for 300 bacterial generations on various host plants. The bacteria was recovered from each passage after 5 – 15 days post inoculation (5 dpi – susceptible host; 15 dpi – resistant host). The derived clones were measured for *in planta* fitness and the clones with enhanced fitness were sequenced by whole genome sequencing and compared to the ancestral GMI1000.

(Adapted from Guidot et al. 2014)

1.4 GMI1000: The study model

The GMI1000 (**G**enetique des **M**icroorganismes **I**NRA) strain of *R. solanacearum* was among the first plant pathogen to have its whole genome sequenced. The strain was isolated in 1960 from tomato in French Guyana (Salanoubat *et al.*, 2002). GMI1000 belongs to phylotype I that is known to have a high evolutionary potential and can infect several plant species (Genin 2010; Genin and Denny 2012; Wicker et al. 2012). This strain was selected as a reference strain in this study.

1.4.1 Experimental evolution of GMI1000

Before my arrival, an experimental evolution approach combined with whole genome sequencing was opted by Guidot and co-workers at LIPM, INRA with the GMI1000 strain. The goal was to investigate the genetic basis of adaptation of the pathogen to various host plants comprising both original (susceptible plant) and distant hosts (tolerant plant).

The experimental evolution was conducted on eight different host plants both original and distant to GMI1000 *R. solanacearum*. Three susceptible host plant species: Tomato (*Solanum lycopersicum* var. Super Marmande), Eggplant MM61 (*S. melongena* var. Zebrina) and Pelargonium (*Geranium sanguineum* var. Maverick Ecarlate); three tolerant variety: Tomato Hawaii 7996 (*S. lycopersicum*), Cabbage (*Brassica oleracea* var. Bartolo) and Bean (*Phaseolus vulgaris* var. Blanc Precoce); and two resistant host species Eggplant MM134 (S. aethiopicum Aculeatum) and Melon Védrantais (*Cucumis melo).*

Serial passage experiments (SPE) were performed by inoculating a single clone of the reference strain GMI1000 in various host plants and serial passages from one plant to another of the same species in order to maintain the pathogen population for over 300 bacterial generations into the same plant species (Figure 10). For each plant species, five biological replicates were conducted thus, generating five parallel lineages/populations of clones derived from the GMI1000 strain (Guidot *et al.*, 2014).

1.4.2 Enhanced *in planta* fitness explained by genomic polymorphisms

Among the 125 investigated experimentally derived clones, 80% of the clones showed a fitness gain in their experimental host (Tomato var. Marmande, Eggplant var. Zebrina, Cabbage var. Bartolo, or Bean var. Blanc Précoce) compared to the GMI1000 ancestral clone. Whole genome sequencing (Illumina sequencing) of 50 adapted clones revealed an average of only 2.5 genomic polymorphisms per evolved clone compared to the ancestor, and between 0 and 9 genomic polymorphism per clone (Figure 11, Guidot, unpublished data). This experiment also demonstrated the importance of the *efpR* regulatory gene in adaptation to the host plants. Six different SNPs in *efpR* were observed in six independent populations evolved on bean, tomato Marmande, eggplant Zebrina and melon that conferred fitness in plant (Guidot *et al.*, 2014; Perrier *et al.*, 2016). The evolved clones carrying these SNPs in the *efpR* gene behave like the *efpR* deleted mutant thus demonstrating that these SNPs were loss of function mutations (Perrier *et al.*, 2016). Fascinatingly, no polymorphisms were detected in seven of the 50 analyzed clones, which includes two of the clones evolved on Hawaii 7996 tomato. The hypothesis was that their fitness gain could be explained by non-genetic modifications such as DNA methylation.

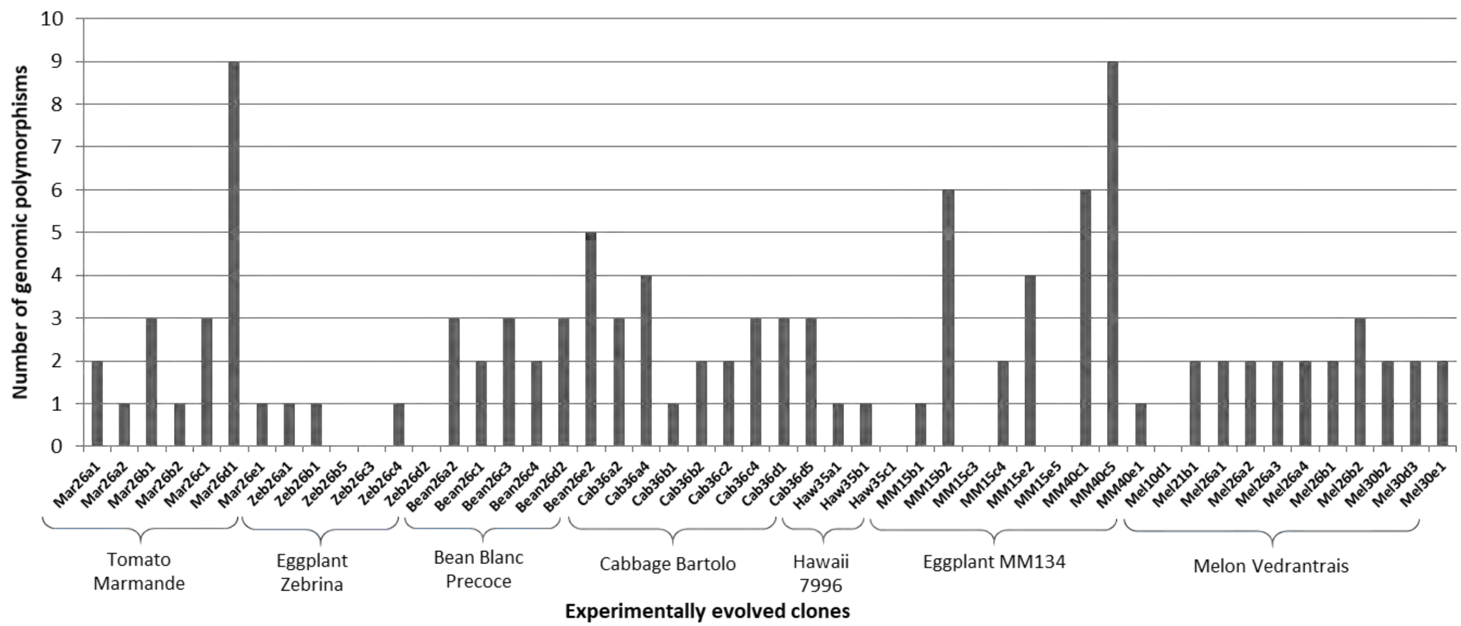


Figure 11 Whole genome sequencing of the evolved clones

Comparison of 50 genomic sequences of evolved clones. 0 to 9 polymorphisms were detected per evolved clone in comparison to the ancestral GMI1000 clone. The number of genomic polymorphisms per clone is given in X-axis and the corresponding evolved clone with its experimental host in given in Y-axis. (Unpublished data, Guidot)

1.5 Objectives

The global objective of my thesis was to determine the genetic and epigenetic modifications that confer adaptation of *R. solanacearum* to various host plants. Using an experimental evolution approach, clones were generated from GMI1000 by serial passage experiments (SPE) during 300 bacterial generations in various host plants (Tomato var. Marmande, Eggplant var. Zebrina, Cabbage var. Bartolo, Bean var. Blanc Précoce) (Guidot *et al.*, 2014) and in Tomato var. Hawaii 7996 (Gopalan-Nair et al., submitted).

Fitness gain was not previously measured for the clones evolved in Hawaii 7996. Therefore, in the first part of my thesis project, the fitness of the clones evolved in Hawaii 7996 was calculated using a competitive index (CI) experiment as previously described (Macho *et al.*, 2010; Guidot *et al.*, 2014). Based on the CI experiment results, 10 of the selected evolved clones were analyzed for genomic polymorphisms and transcriptomic variations in comparison to the ancestral clone. The impact of these genomic polymorphisms were then tested by two methods: gene deletion and mutation insertion, to test for the gain of function or loss of function hypotheses. The methylome diversification was then investigated in 28 experimentally evolved clones, derived from the GMI1000 strain after SPE during 300 generations in tomato Hawaii 7996, Eggplant Zebrina and Bean. This analysis was conducted by using the SMRT (single molecule real time) technology of PacBio sequencing. The transcriptomic profiles of these 28 evolved clones were also investigated in order to connect any differential methylation marks in gene promoters with gene expression.

The project was divided into three parts:

1. Adaptation of *RSSC* to the tolerant tomato plant Hawaii 7996: An overview at the molecular level

2. Genomic and transcriptomic variations in clones experimentally evolved on three different host plants

3. Methylome variation in experimentally evolved clones

Chapter 2. Adaptation of *Ralstonia solanacearum* species complex to a tolerant plant: An overview at the molecular level

2.1 Brief introduction

The first part of my project was to investigate the molecular bases of adaptation of RSSC to a tolerant tomato plant 'Hawaii 7996'. The pathogen causes bacterial wilt (BW) and can infect more than 250 plant species belonging to 54 botanical families (Genin and Boucher, 2002b). The use of resistant cultivar is the most effective and economical choice to control bacterial wilt. However, RSSC is a fast evolving bacterial pathogen that has the ability to overcome plant resistance and can colonize novel hosts owing to its genome plasticity (Guidot *et al.*, 2009; Wicker *et al.*, 2009; Peeters, Guidot, *et al.*, 2013). Hawaii 7996 is a reference resistant cultivar in tomato known to have stable resistance against RSSC. Many studies were conducted to identify the quantitative trait loci (QTLs) associated with resistance against BW. Two major QTLs namely Bwr-6 and Bwr-12 were found to be associated with stable resistance against BW and are found to limit the internal bacterial multiplication in the stem (Carmeille *et al.*, 2006; Lebeau *et al.*, 2010; Wang *et al.*, 2013).

Our laboratory was interested in knowing if RSSC can overcome Hawaii resistance. In addition, if yes by which molecular mechanisms (genetics and/or epigenetics) the pathogen is adapted. Therefore, (prior to my arrival) an evolution experiment of GMI1000 was conducted on Hawaii 7996 plant (unpublished). This experiment was conducted by serial passage experiments (SPE) of the GMI1000 strain on Hawaii 7996 during 300 bacterial generations. The number of SPE required to reach 300 generations was higher (35 SPE) for tomato Hawaii in comparison to the other host plants such as tomato Marmande or eggplant Zebrina, which only required 26 SPE

(Guidot *et al.*, 2014). Five biological replicates were conducted thus generating five parallel lineages (populations) of clones experimentally derived from GMI1000.

A first step of my thesis consisted of phenotypic analyses of 25 derived clones (five clones randomly isolated from each of the five lineages) in comparison to the ancestral GMI1000. The genome of ten evolved clones were then sequenced to identify the genomic polymorphisms that occurred during the experimental evolution. The mutations were constructed in the GMI1000 strain in order to investigate their impact in adaptation to Hawaii. We also performed an RNAseq analysis to characterize the transcriptome of these ten clones in order to identify the differentially expressed genes between the ancestral clone and the clones adapted to Hawaii.

The article submitted in the MBE journal discusses the evolution experiment of GMI1000, phenotypic analyses of the evolved clones along with their genomic and transcriptomic profiles in comparison to the ancestral clone.

2.2 Genetic bases of adaptation of GMI1000 to tomato Hawaii 7996 plant

2.2.1 Article

Convergent rewiring of the virulence regulatory network promotes adaptation of *Ralstonia solanacearum* on resistant tomato

2.2.2 Journal

Molecular Biology and Evolution

2.2.3 Summary

The Gram-negative bacteria RSSC that causes bacterial wilt has widely been used as a model system to study bacterial pathogenicity. The ability of this pathogen to adapt to many host plants and counteract plant resistance is supported by field observations reporting strains more aggressive and able to colonize novel hosts. The global objective of this paper was to estimate the adaptive potential of *Ralstonia solanacearum* to overcome the quantitative resistance of the tomato cultivar Hawaii 7996. An experimental evolution strategy of the RSSC strain GMI1000 on Hawaii 7996 was used by serial passage experiment (SPE) for 300 bacterial generations. Overall, 300 generations of experimental evolution were not sufficient for the bacteria to induce disease in Hawaii 7996. However, phenotypic analyses showed that 24 of 25 tested derived clones had better fitness *in planta* compared to the ancestral clone. The whole genome sequencing of ten evolved clones revealed five different mutations but only 0 to 2 mutations per clone. The mutations reconstructed in the ancestral clone, showed that all five individual mutations were beneficial. *In vitro* transcriptome profiles of these evolved clones were analyzed by RNAseq technology in comparison to the transcriptome profile of the ancestral clone. Some of the key genes that were differentially expressed in comparison to the ancestral clone include the T3Es, Hrp genes, *efpR, prhP,* eps genes, flagellar, and motility related genes. Interestingly, largely overlapping gene expression profiles suggested a convergence towards a global rewiring of the virulence regulatory network in the majority of the clones. Several differentially expressed genes were also detected in the evolved clones with no mutation, thus suggesting a role of non-genetic/epigenetic modifications in adaptation.

<< article >>

Convergent Rewiring of the Virulence Regulatory Network Promotes Adaptation of *Ralstonia solanacearum* on Resistant Tomato

Rekha Gopalan-Nair,[1] Marie-Françoise Jardinaud,[1] Ludovic Legrand,[1] David Landry,[1] Xavier Barlet,[1] Céline Lopez-Roques,[2] Céline Vandecasteele,[2] Olivier Bouchez,[2] Stéphane Genin,*[,1] and Alice Guidot*[,1]

[1]LIPME, Université de Toulouse, INRAE, CNRS, Castanet-Tolosan, France

[2]GeT-PlaGe, Genotoul, INRAE, US 1426, Castanet-Tolosan, France

***Corresponding authors:** E-mails: stephane.genin@inrae.fr; alice.guidot@inrae.fr.

Associate editor: Julian Echave

Abstract

The evolutionary and adaptive potential of a pathogen is a key determinant for successful host colonization and proliferation but remains poorly known for most of the pathogens. Here, we used experimental evolution combined with phenotyping, genomics, and transcriptomics to estimate the adaptive potential of the bacterial plant pathogen *Ralstonia solanacearum* to overcome the quantitative resistance of the tomato cultivar Hawaii 7996. After serial passaging over 300 generations, we observed pathogen adaptation to within-plant environment of the resistant cultivar but no plant resistance breakdown. Genomic sequence analysis of the adapted clones revealed few genetic alterations, but we provide evidence that all but one were gain of function mutations. Transcriptomic analyses revealed that even if different adaptive events occurred in independently evolved clones, there is convergence toward a global rewiring of the virulence regulatory network as evidenced by largely overlapping gene expression profiles. A subset of four transcription regulators, including HrpB, the activator of the type 3 secretion system regulon and EfpR, a global regulator of virulence and metabolic functions, emerged as key nodes of this regulatory network that are frequently targeted to redirect the pathogen's physiology and improve its fitness in adverse conditions. Significant transcriptomic variations were also detected in evolved clones showing no genomic polymorphism, suggesting that epigenetic modifications regulate expression of some of the virulence network components and play a major role in adaptation as well.

***Key words:* experimental evolution, adaptive potential, fitness measure, genomic polymorphisms, transcriptomic variation, bacterial plant pathogen.**

Introduction

Plant–pathogen interaction is in constant evolution through an arms race of pathogen attack and plant defense (Jones and Dangl 2006). The evolutionary and adaptive potential of a pathogen is a crucial determinant for successful host colonization and proliferation. Characterizing the adaptive potential of a pathogen is therefore important to guide strategies for durable resistance breeding.

In bacterial plant pathogens, the evolutionary potential depends on three main factors: 1) their potential for gene flow between geographically separated populations, 2) gene exchange between individuals through horizontal gene transfers, and 3) genomic (or epigenomic) modifications (McDonald and Linde 2002). Several types of genomic modifications could occur and have a severe impact on the bacterial phenotype such as single nucleotide polymorphisms (SNPs), insertions, inversions, deletions, translocations, mobile element insertions, duplications, or large genomic rearrangements. These genomic modifications are specifically important for bacterial pathogen evolution because bacterial pathogens exist as large populations in their host thus enhancing the probability to new mutants with higher fitness to multiply within the population before the mutation is lost through genetic drift.

An elegant approach to investigate the adaptive potential of bacteria through genomic modifications is to study its evolution in real time by conducting an experimental evolution. In this approach, adaptation of the experimentally evolved isolates occur by natural selection in the experimental environment by competing bacteria from the later generation against the ancestral strain (Ebert 1998; Lenski 2017). Whole-genome sequencing of the experimentally evolved isolates and comparison with the ancestral strain genomic sequence allows the detection of all genomic modifications that appeared during experimental evolution (Tenaillon et al. 2012; Barrick and Lenski 2013). This approach has been used to investigate the genomic bases of adaptation of bacterial plant pathogen to their host plant (Guidot et al. 2014; Trivedi and Wang 2014; Meaden and Koskella 2017).

Variable adaptive strategies could occur according to the genetic background conferring resistance of the host plant against pathogens. Plant disease resistance is usually divided into qualitative and quantitative resistance (Poland et al. 2009). Qualitative resistance is controlled by major resistance (R) gene(s), whereas quantitative resistance involves multiple genes or quantitative resistance loci (QRL) with small to moderate effects (Corwin and Kliebenstein 2017). R gene usually confers complete resistance to a specific pathogen inducing a hypersensitive cell death response (HR) at the infection site. However, pathogens can rapidly overcome this resistance through mutations in effectors recognized by the R genes or through acquisition of new effectors by horizontal gene transfers (Jones and Dangl 2006). An experimental evolution conducted with the bacterial plant pathogen *Xanthomonas citri* subsp. *citri* on a resistant host plant inducing an HR experimentally demonstrated that the pathogen rapidly overcome the plant resistance through mutations biased toward type 3 secretion system (T3SS) effector genes (Trivedi and Wang 2014). Pathogens also evolve to overcome plant quantitative resistance. However, this evolution is more difficult to detect and is better characterized as a process of "erosion" rather than a process of breakdown (McDonald and Linde 2002). In this study, we used an experimental evolution approach to characterize the adaptive potential of a strain from the *Ralstonia solanacearum* species complex (RSSC) to overcome the quantitative resistance of tomato.

Strains of the RSSC are responsible for the bacterial wilt disease on more than 250 plant species including economically important crops (Peeters et al. 2013). RSSC strains are recognized as one of the most lethal plant bacterial pathogens with a worldwide geographical distribution (Hayward 1991; Mansfield et al. 2012). Bacteria survive in soil for many years and spread through water, rhizosphere contact, and farming (Genin and Boucher 2004; Genin 2010; Song et al. 2018). They enter the plant through the roots, invades the xylem vessels, and spreads rapidly to aerial parts of the plant through the vascular system (Genin 2010). Within a few days of infection, the bacteria reach high population levels by extensive colonization (up to 10^{10} colony-forming units per gram of fresh weight) that leads to vascular clogging causing wilting symptoms and ultimately plant death (Peeters et al. 2013). The potential of this pathogen to evolve and adapt to numerous host plants substantiates the field observations of the emergence of strains that colonize new hosts (Hayward 1991; Wicker et al. 2007; Wicker et al. 2009). Various strategies are used to control the bacterial wilt disease such as crop rotation, chemical and biological controls but the use of resistant cultivars remains the most effective control strategy (Lebeau et al. 2011). However, resistance breakdown is continuously observed in the field and breeders have to face the problem of resistance durability against this pathogen.

In tomato, the reference resistant cultivar against the bacterial wilt disease is the cultivar Hawaii 7996 which has the most stable source of resistance against different RSSC strains in the field (Lebeau et al. 2011; Wang et al. 2013). Resistance of Hawaii 7996 to bacterial wilt relies on polygenic traits (Thoquet et al. 1996; Carmeille et al. 2006) and is expressed in both root and shoot tissues (Planas-Marquès et al. 2020). Several QRL controlling RSSC bacterial wilt have been identified (Thoquet et al. 1996; Carmeille et al. 2006; Wang et al. 2013). The products and functions of these QRL are still unknown but they appear to confine the bacteria to the primary xylem vessels, even when large amounts of bacteria are injected into the stem (McGarvey et al. 1999; Nakaho et al. 2004). Structure of the rhizosphere microbiome was also shown to be a key parameter of Hawaii 7996 resistance (Kwak et al. 2018). Here, we conducted an evolution experiment with the RSSC strain GMI1000 by serial passages during 300 bacterial generations on the resistant tomato cultivar Hawaii 7996. Five independent lineages of derived clones were generated and phenotypic analyses were conducted on 25 derived clones to measure their fitness gain. The appearance of bacterial wilt symptoms on Hawaii 7996 was followed at each passage. The genome and transcriptome of ten evolved clones were characterized and compared to that of the ancestral clone. The contribution of the genomic modifications to the enhanced fitness in Hawaii 7996 was then functionally investigated.

Results

Experimental Evolution of GMI1000 Strain on the Resistant Tomato Hawaii 7996

Experimental evolution of RSSC strain GMI1000 was performed by serial passage experiment (SPE) into the stem of the tomato cultivar Hawaii 7996. SPEs were conducted by inoculating the bacteria directly into the plant stem in order to control the number of colony-forming unit (CFU) transferred from one plant to the other and to maintain a homogeneous selective environment during the course of the experiment. Bacteria were recovered from the plant stem 15 days postinoculation by natural diffusion into water and immediately reinoculated into the stem of a healthy tomato Hawaii 7996 plant. Five biological SPE replicates were conducted in parallel thus generating five populations (named "A," "B," "C," "D," and "E") of clones derived from the same GMI1000 ancestral clone.

The average estimated number of bacterial generations at each SPE was 9 ± 3 (mean $\pm$ SD). A total number of 35 SPE was then necessary to reach at least 300 bacterial generations into Hawaii 7996 which corresponded to 525 days of experimental evolution. The low number of generations obtained by this calculation is undoubtedly underestimated, not taking into account a probable mortality of some of the bacteria in the plant. During the course of the experimental evolution, no wilting symptom were detected and the in planta growth rate of the GMI1000 strain did not increase by remaining around 10^8 CFU/g of fresh weight after 15 days of infection (supplementary fig. S1, Supplementary Material online). The inoculated populations reached densities above the quorum-sensing threshold (estimated around 10^7 CFU/ml; Flavier et al. 1997; Peyraud et al. 2016) which is sufficient to fully induce virulence gene expression.

Obtaining Experimentally Derived Clones Showing Fitness Gain

At the end of the experimental evolution, we compared the fitness of the derived clones to the fitness of the ancestral clone into the stem of Hawaii 7996 by conducting competition experiments. In these experiments, the ancestral clone (tagged with a gentamycin resistance cassette) was coinoculated with a derived clone at the same proportion into the stem of a same plant. A competitive index (CI) was calculated (see Materials and Methods) and used as a fitness estimator. The CI was determined for 25 derived clones representing five clones randomly isolated from each of the five independent populations obtained after 35 SPEs. These clones were named "Haw35" followed by a letter corresponding to the population they originate from, and a number (e.g., "Haw35a1" is clone number 1 from population A obtained after 35 SPEs on Hawaii 7996; Haw35a1 could also be named "a1" to simplify). As a control, the ancestral clone was coinoculated with the gentamycin-resistant variant of the ancestral clone.

The CI obtained for the ancestral clone and each of the 25 tested derived clones are shown in figure 1. During this CI experiments, the cell densities for both the tested clone and the gentamycin-resistant variant were around 10^9 CFU/g of fresh weight after 15 days of infection (supplementary table S1, Supplementary Material online). The CI for the ancestral GMI1000 clone was not significantly different from one thus demonstrating that the gentamycin-resistant cassette did not affect the in planta fitness of the bacteria, the wild-type (WT) strain, and its gentamycin-resistant derivative having the same fitness into Hawaii 7996. The CI values obtained for the 25 tested derived clones were all (but one) significantly superior to one thus demonstrating that all these derived clones (but the c3 clone) have a better fitness than their ancestral clone into the stem of Hawaii 7996 (fig. 1).

The mean CI for these experimentally evolved clones was 4.96 ± 0.36 (mean $\pm$ SE). Comparison of the mean CI for each of the five populations of experimentally evolved clones revealed a significantly better adaptation to Hawaii 7996 for clones of populations A and B than for other clones, the mean CI for these two populations being 5.46 ± 0.75 and 9.80 ± 1.35 (mean $\pm$ SE), respectively, whereas the mean CI for the populations C, D, and E were 3.28 ± 0.52, 3.50 ± 0.52, and 3.19 ± 0.41, respectively (table 1). This more rapid adaptation observed for A and B clones could be due either to a more rapid accumulation of adaptive mutations or to the fixation of different adaptive mutations than in the other populations.

Genomic Resequencing Reveals That Fittest Bacteria Carry Few or No Genetic Changes

In order to identify the genomic polymorphisms associated to fitness gain during experimental evolution, the whole genomes of ten adapted clones (two per population) were sequenced using both Illumina and Pacbio sequencing technologies. The Illumina sequencing technology was used for the detection of SNPs and small Insertion–Deletion (InDels). The PacBio sequencing technology was used for the detection of large genomic rearrangements.

Comparison of the GMI1000 ancestral clone and the evolved clones genomic sequences revealed between zero and two genomic polymorphisms per clone (table 2). Rather surprisingly, despite the use of two efficient and complementary sequencing technologies, no genomic polymorphism could be detected in five of the ten adapted clones studied, including one clone from population A (the Haw35a4 clone) as well as the four clones from populations C and D. One mutation was detected in the clones Haw35a1,

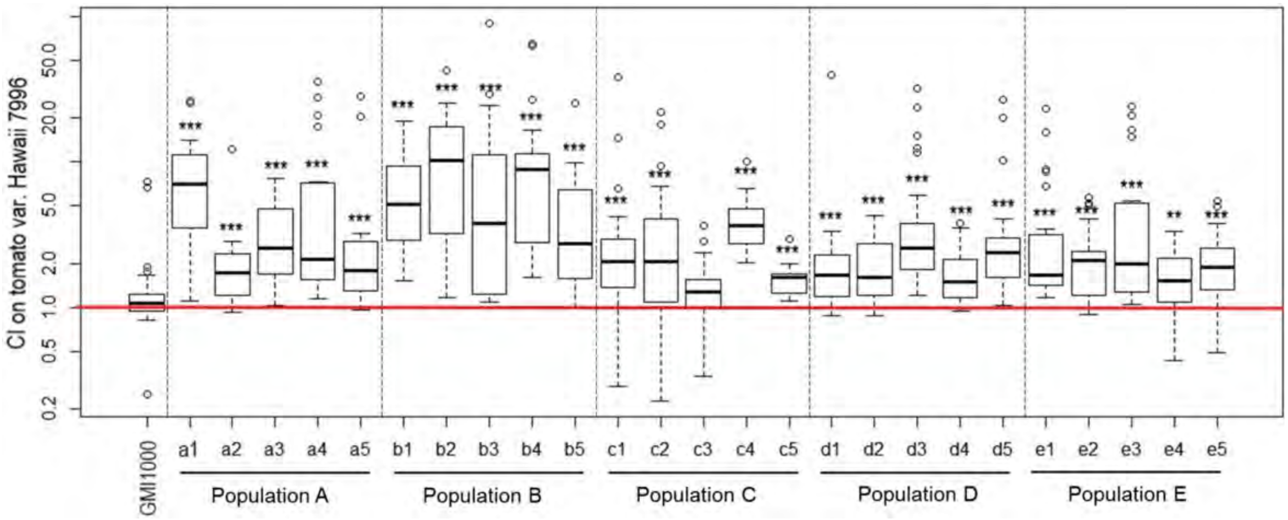


FIG. 1. Box plot of the CI values of the control ancestral GMI1000 clone and the derived clones on tomato Hawaii 7996. All the strains were competed by GRS540 (GMI1000 carrying a Gm resistance cassette). The CI was measured for 25 of the clones derived from GMI1000 after experimental evolution during ~300 bacterial generations into the stem of the Hawaii 7996. These 25 clones represent five clones randomly isolated from each five independent populations generated by SPE (populations A, B, C, D, and E). A minimum of ten replicates were performed for each clone. Statistical analyses were performed using Wilcoxon test (***$P \leq 0.001$). The Y axis gives exponential values of the CI. Extreme individual values (outliers) are represented by empty dots. The black bar inside the box plot indicates the median CI value.

Table 1. Comparison of the Mean CI Values between Populations of Evolved Clones.

	Population A Mean CI = 5.46 SE = 0.75	Population B Mean CI = 9.80 SE = 1.35	Population C Mean CI = 3.28 SE = 0.52	Population D Mean CI = 3.50 SE = 0.52	Population E Mean CI = 3.19 SE = 0.41
Population B	1.19E−4				
Population C	0.02289	5.69E−10			
Population D	0.0424	4.71E−11	0.6972		
Population E	0.01017	2.44E−11	0.7894	0.4157	

Note.—Mean CI values were compared using a Wilcoxon test and the table gives the obtained *P* value. The CI was measured for a total of five clones per population. Significant differences between mean CI are highlighted in gray.

Table 2. List of Genomic Polymorphisms between the Ancestral GMI1000 Clone and the Experimentally Evolved Clones.

Gene			Clones Evolved on Hawaii7996									
ID	Name	Description	a1	a4	b1	b4	c1	c2	d3	d5	e1	e3
Rsc3094		Hypothetical protein										G486A R162R
Rsp0048	*soxA1*	Sarcosine oxidase alpha subunit	T1915C C639R									
Rsp0309	*prhP*	Phenolic acid decarboxylase regulator (PadR)-like				ISrso9 −6						
Rsp1136		ISrso18 Transposase protein									C-218A	C-218A
Rsp1574		Transcription regulator protein			G283T V95L	G283T V95L						

Note.—For each gene, the genomic polymorphism is indicated. For SNPs, both the nucleotide modification and the protein modification are indicated with the original nucleotide or amino acid, the position and the new nucleotide or amino acid. ISrso9 indicates an IS insertion and the number indicates the position of the insertion. The "-" indicates a mutation upstream the start codon of the gene.

Haw35b1, and Haw35e1, and two mutations were identified in the clones Haw35b4 and Haw35e3 (table 2).

The detected genomic polymorphisms were either SNPs or Insertion sequence (IS) insertion. Four SNPs were detected, two were nonsynonymous mutations, one was a synonymous mutation, and one mutation was detected in an intergenic region. The first nonsynonymous mutation occurred in one of the two investigated clones from population A (the Haw35a1 clone) and affected the RSp0048 gene encoding the SoxA1 protein (Sarcosine oxidase alpha subunit) (table 2). The second nonsynonymous mutation was identified in the two investigated clones from population B (Haw35b1 and Haw35b4) and affected the RSp1574 gene encoding a transcription regulator of unknown function (table 2). The synonymous mutation was detected in one of the two investigated clones from population E (Haw35e3) in the RSc3094 gene encoding a hypothetical protein (table 2). The mutation in the intergenic region was identified in the two investigated clones from population E (Haw35e1 and Haw35e3) and occurred between the RSp1136 gene encoding an IS transposase protein and the RSp1137 gene encoding a transmembrane protein (table 2). The IS insertion was detected six nucleotides before the start codon (and so presumably in the ribosomal binding site) of the RSp0309 gene encoding the PrhP protein (phenolic acid decarboxylase regulator) (Zhang et al. 2019). This IS insertion occurred in one of the two investigated clones from population B (Haw35b4) (table 2). All the detected SNPs and IS insertion were confirmed by polymerase chain reaction (PCR) amplification of the mutated region and either Sanger sequencing or gel electrophoresis, respectively.

The Genomic Polymorphisms Detected in the Five Adapted Clones Are Gain of Function Mutations

In order to investigate the fitness advantage of the genomic polymorphisms detected in the adapted clones Haw35a1, Haw35b1, Haw35b4, Haw35e1, and Haw35e3, we created each mutation in the ancestral clone. Five GMI1000 mutants were first generated, mAG68 carrying the soxA1^{C639R} SNP, mAG69 carrying the RSp1574^{V95L} SNP, mAG70 carrying the ISrso9 insertion in the promotor region of the *prhP* gene (prhP^{IS-6}), mAG71 carrying the RSc3094^{R162R} SNP, and mAG72 carrying the SNP in the RSp1136–1137 intergenic region (RSp1136^{C-218A}). The fitness advantage of these mutations were measured by conducting competitive experiments with the GMI1000 WT strain into the stem of Hawaii 7996. These experiments revealed that all the generated mutants had a CI significantly superior to one thus demonstrating that all the mutants had a better fitness in Hawaii 7996 than the WT strain (fig. 2).

When comparing the CI values obtained for each mutant, we found that the CI values of the mAG68 and mAG69 mutants carrying the soxA1^{C639R} and RSp1574^{V95L} SNPs, respectively, were significantly superior to the CI values of the three other mutants (fig. 2). This result suggested that the soxA1^{C639R} and the RSp1574^{V95L} SNPs gave a better fitness advantage than the other mutations. The CI values obtained for the mutants mAG69, mAG71, and mAG72 carrying the mutations RSp1574^{V95L}, RSp1136^{C-218A}, and RSc3094^{R162R}, respectively, fully replicated the levels of fitness gain displayed by the corresponding evolved clones (b1, e3, and e1, respectively; figs. 1 and 2). However, this was not true for the mutants

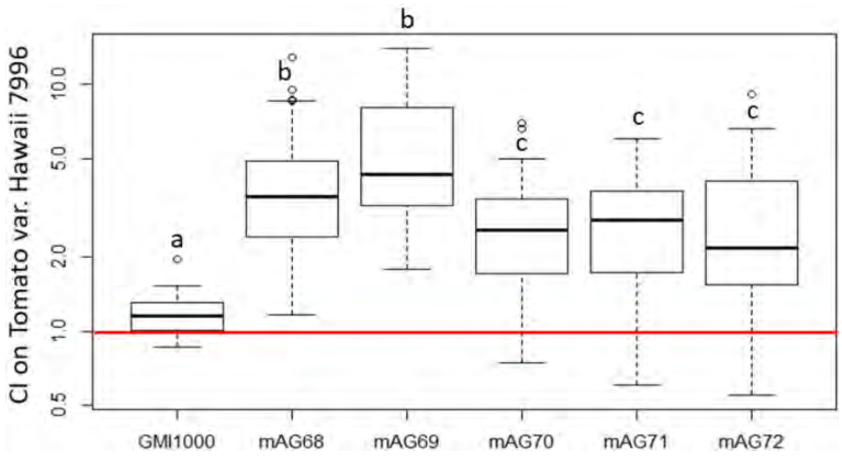


Fig. 2. Box plot of the CI values of GMI1000 and allelic mutants on tomato Hawaii 7996. mAG68, mAG69, mAG70, mAG71, and mAG72 carry the mutation soxA1^{C639R}, RSp1574^{V95L}, prhP$^{IS\text{-}6}$, RSc3094^{R162R}, and RSp1136$^{C\text{-}218A}$ SNP, respectively. A minimum of ten replicates were performed for GMI1000 and each allelic mutant. Statistical analyses were performed using Wilcoxon test. Different letters indicate significantly different CI values ($P < 0.05$) (see fig. 1 for legend).

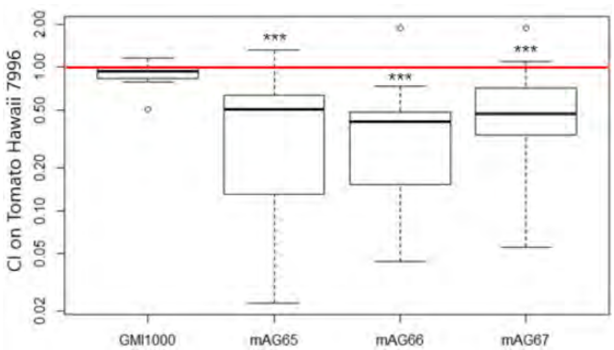


Fig. 3. Box plot of the CI values of GMI1000 and unmarked deletion mutants on tomato Hawaii 7996. mAG65, mAG66, and mAG67 correspond to ΔsoxA1, ΔRSp1574, and ΔRSp3094 respectively. A minimum of ten replicates were performed for GMI1000 and each deletion mutant. Statistical analyses were performed using Wilcoxon test (***$P \leq 0.001$) (see fig. 1 for legend).

mAG68 and mAG70 carrying the mutations soxA1^{C639R} and prhP$^{IS\text{-}6}$, respectively, for which the CI values were less than the CI values of their corresponding evolved clones (a1 and b4, respectively; figs. 1 and 2). The second mutation, RSp1574^{V95L}, detected in b4 could explain this difference. However, for a1, no other mutation was detected, thus suggesting a role of other epigenetic modification in the fitness gain of a1.

In order to test the hypothesis that fitness gain was associated with a loss of function mutation for the three SNPs detected into the coding region of the soxA1, RSp1574, and RSc3094 genes, we deleted each of these genes in the GMI1000 strain. Competitive experiments into the stem of Hawaii 7996 between each of these three deleted mutants and the WT strain revealed that the mutants were less fit than the GMI1000 WT strain into this tomato stem. Indeed, the CI obtained for these three deleted mutants are significantly inferior to one (fig. 3). These analyses demonstrated that the SNPs that appeared during experimental evolution did not cause any loss of function of the genes but rather enhanced the functionalities of the corresponding proteins.

Some Fitness Gains on Resistant Tomato Are Associated with Increased Growth Rates

To evaluate the metabolic efficiency of the evolved clones, we determined their growth rates from in vitro cultures, using minimal medium (MM) supplemented either with glutamine (the most abundant amino acid in tomato xylem) (Zuluaga et al. 2013) or proline, a discriminant marker of metabolic versatility of GMI1000 strain (Perrier et al. 2016; Peyraud et al. 2016). These analyses were conducted for the ten adapted clones sequenced previously, the three deleted mutants and the five reconstructed allelic mutants, all compared with the WT GMI1000 strain. We also used as positive control, the *efpR*-deleted mutant, which is characterized by both enlarged metabolic diversity and enhanced growth rate in the tested media (Perrier et al. 2016). Four of the ten evolved clones (a1, b1, b4, and c1) had a growth rate significantly higher than their WT ancestor in MM + glutamine (fig. 4*A*). In MM + proline, the growth of the WT is greatly reduced but the same four clones also acquired an increase in growth rate (fig. 4*B*). Two groups of clones were distinguished: a1 and c1, which had a growth rate in glutamine and proline similar to the one of the *efpR* mutants, and b1 and b4 with an intermediate gain in the growth rate. These results suggested that these four clones could probably multiply better than the ancestor in planta, but also indicated that fitness gain was not associated with apparent increased growth rate in the remaining six evolved clones.

Concerning the deleted and allelic mutants, only the mAG69 allelic mutant, carrying the RSp1574^{V95L} mutation detected in both b1 and b4, had a growth rate in glutamine and proline significantly better than the WT strain (Welch *t*-test; *P* value = 0.0038 on glutamine; *P* value = 1.03E−6 on proline). Increase in growth rate was in a range similar to that of the adapted clones b1 and b4 (fig. 4). This result suggested that the improved growth rate of these two clones in glutamine- and proline-containing environments depends on the RSp1574^{V95L} mutation, which thus contributes to the observed fitness gain in planta.

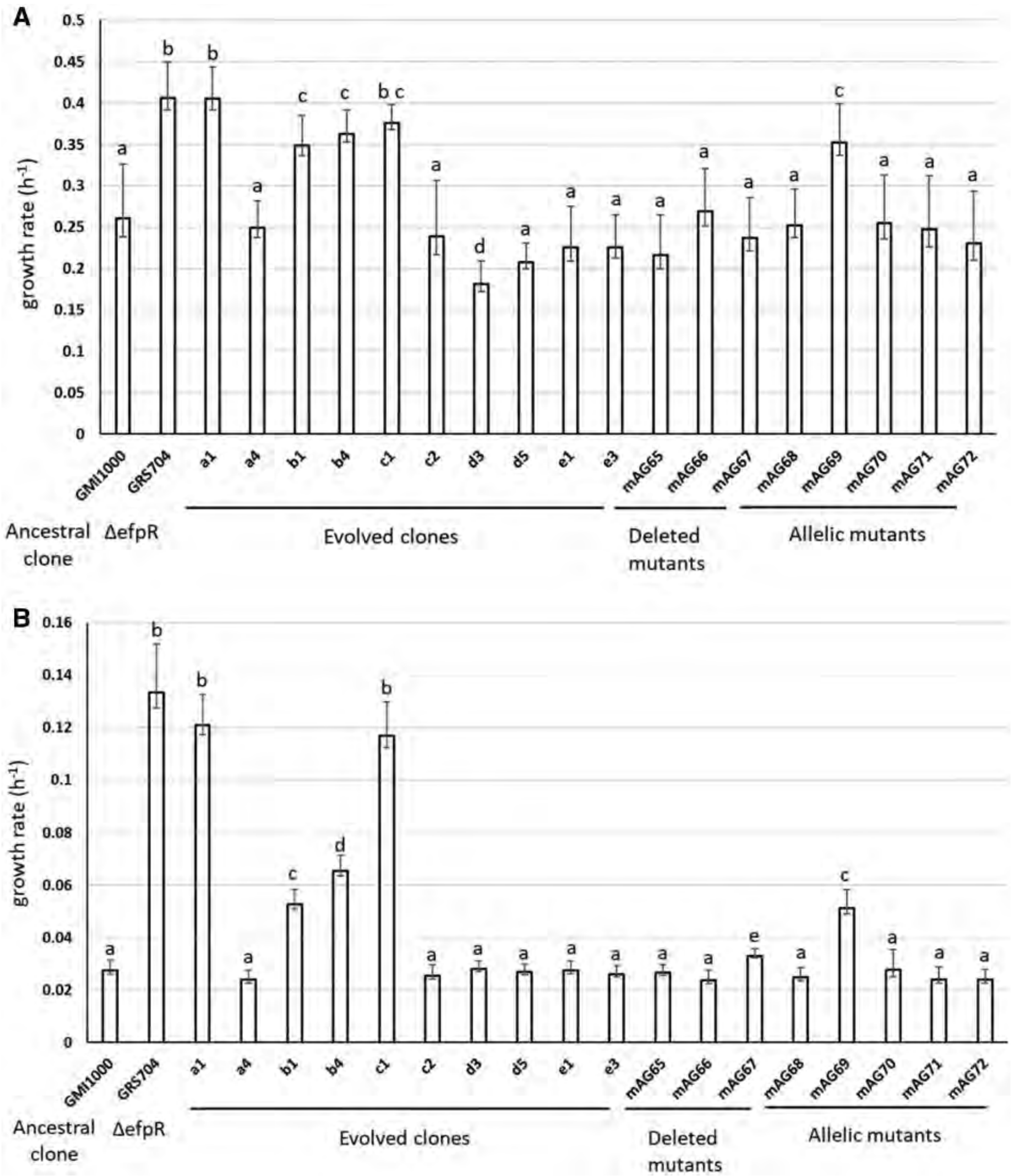


FIG. 4. In vitro growth rate of evolved clones and engineered mutants in MM supplemented with (*A*) 20 mM glutamine and (*B*) 20 mM proline. The clones were grown at 28 °C under shaking and the OD_{600nm} was measured until 50 h. Growth rate was calculated during the exponential growth phase. The deleted mutants mAG65, mAG66, and mAG67 correspond to ΔsoxA1, ΔRSp1574, and ΔRSp3094, respectively. The allelic mutants mAG68, mAG69, mAG70, mAG71, and mAG72 carry the mutation soxA1^{C639R}, RSp1574^{V95L}, prhP$^{IS\text{-}6}$, RSc3094^{R162R}, and RSp1136$^{C\text{-}218A}$ SNP, respectively. Statistical analysis was performed using Welch *t*-test. Different letters indicate significantly different growth rates ($P < 0.01$).

Evidence for Significant Transcriptomic Variations in Evolved Clones Showing No Genomic Polymorphism

To obtain a broader picture of the physiological changes in the better-adapted clones isolated from Hawaii 7996, we studied potential variation of their transcriptomic profile compared with the ancestral founder. We included in this analysis several of the clones for which no genetic alteration could be detected. We thus established the transcriptomes of the ten adapted clones sequenced previously and the GMI1000 ancestral clone using an RNA-sequencing (RNA-seq) approach of strains grown in MM supplemented with 10 mM glutamine.

Analysis of RNA-seq data revealed that all samples rendered between 1.3 and 4.7 million of GMI1000-mapped reads. Differentially expressed genes (DEGs) between an adapted clone and the ancestral clone were considered as those presenting an absolute fold change between strain $|FC| > 2$ and an false discovery rate (FDR)-adjusted *P* value (*P*adj, FDR) $<$ 0.05. Using these cutoff values, we found between 125 and 1,227 DEGs in the ten investigated adapted clones compared with the ancestral clone (see table 3 for summary and complete lists in supplementary table S2, Supplementary Material online). It thus appeared that a significant variation from the WT transcriptomic pattern were recorded in clones in which

Table 3. DEGs between the Clones Experimentally Adapted to Hawaii 7996 and the Ancestral GMI1000 Strain.

	Clones Evolved on Hawaii 7996									
	a1	a4	b1	b4	c1	c2	d3	d5	e1	e3
Mutation	soxA1^{C639R}		RSp1574^{V95L}	RSp1574^{V95L} prhP$^{IS\text{-}6}$					RSp1136$^{C\text{-}218A}$	RSp1136$^{C\text{-}218A}$ RSc3094^{R162R}
Mean CI value	8.6	7.2	6.5	12.9	4.2	4.0	5.4	4.1	3.8	5.4
Median CI value	7	2.1	5.1	8.9	2.1	2.1	2.6	2.4	1.7	2.0
No. of DEGs	1,227	187	478	503	902	272	125	269	245	212
Downregulated genes	617	148	275	290	481	199	91	201	177	154
Upregulated genes	610	39	203	213	421	73	34	68	68	58
Varying expression of known virulence regulators										
efpR	Down (logFC = −3.65)				Down (logFC = −1.91)					
hrpB				Down (logFC = −1.38)	Down (logFC = −1.68)	Down (logFC = −1.56)				
hrpG				Down (logFC = −1.21)						
prhI	Up (logFC = 1.70)				Up (logFC = 1.30)					
prhP	Down (logFC = −1.28)			Down (logFC = −6.36)						
lecM	Down (logFC = −4.46)		Down (logFC = −2.44)	Down (logFC = −1.80)	Down (logFC = −4.86)					
xpsR	Down (logFC = −1.10)									
Main deregulated functions (GO enrichment analysis)[a]										
Protein secretion			Down	Down	Down	Down				
Integral component of membrane	Down		Down		Down	Down				
Macromolecule localization			Down		Down	Down				
Transaminase activity	Down				Down	Down		Down		
Secondary metabolite biosynthetic and metabolic process	Up		Up		Up					
Bacterial-type flagellum assembly and organization	Up				Up					
Chemotaxis	Up				Up					
Fatty acid biosynthetic and metabolic process	Up				Up					
Locomotion	Up				Up					
Monocarboxylic acid biosynthetic process	Up				Up					
Receptor activity	Up				Up					
Signal transducer activity	Up				Up					
Signaling receptor activity	Up				Up					
Structural molecule activity	Up				Up					
Transmembrane receptor activity	Up				Up					

NOTE.—Genes were considered as differentially expressed when the absolute log fold change $|logFC| > 1$ and the FDR-adjusted P value (Padj, FDR) < 0.05.
[a]Full list of functions and statistical threshold used is provided in supplementary table S4, Supplementary Material online.

no genetic alteration could be identified, going up to 900 DEGs in clone c1. Remarkably, this c1 clone is characterized by a significant variation in the expression pattern of major global/virulence regulators, such as *efpR* (Perrier et al. 2016) and *hrpB*, the T3SS regulon transcriptional activator (Genin et al. 1992), or the *lecM* gene that contributes to the quorum-sensing-dependent virulence signaling pathway (Hayashi et al. 2019) (table 3). In the other clones carrying no genetic polymorphism, the number of DEGs is more modest (from 125 to 270 genes) but remains significant. Interestingly, several of these clones with no genetic change showed an obvious overlap in their transcriptomic signatures (see below).

A Convergent Transcriptomic Signature in a Majority of Independently Evolved Clones

In agreement with the previous observations, the number of DEGs in each of the investigated evolved clones was not correlated to the number of mutations detected in the corresponding genomes. For example, the highest number of DEGs (1,227 DEGs) was identified in the a1 clone carrying a single mutation, whereas the b4 and e3 clones, both carrying two mutations, had 503 and 212 DEGs, respectively (table 3).

Because the b1 and b4 clones both carried mutations presumably altering transcriptional regulator functions, we first sought to identify common regulatory targets or overlap with known PrhP targets (Zhang et al. 2019). b1 and b4 share 225 common DEGs, but rather surprisingly, when the comparison includes other clones such as a1 or c1 which do not carry the RSp1574 mutation, the number of genes specifically shared between b1 and b4 drops to 22 (supplementary fig. S2, Supplementary Material online). Seventeen out of these 22 genes are upregulated, which suggests that the RSp1574 mutation conferred gain-of-function for a transcription factor with enhanced activating abilities. Among these, 17 upregulated genes are found the RSp1575 and RSp1576 genes, just neighboring RSp1574, and which both encode transporters for unknown molecules (supplementary fig. S2, Supplementary Material online). Four additional transporters belong to this group of upregulated genes, raising the hypothesis that the fitness gain associated to the RSp1574 mutation is due to an increased uptake of metabolic compounds in planta.

The *prhP* regulatory gene appeared downregulated in two different populations: First, an IS movement was detected in the b4 clone with insertion of the transposable element just 6 bp upstream of the *prhP* start codon, thus probably leading to gene inactivation (table 2). This was confirmed by the transcriptome analysis of clone b4 where the strongest downregulated gene is indeed RSp0309 (*prhP*) (supplementary table S2, Supplementary Material online). Second, we identified that *prhP* was also downregulated in the a1 clone (table 3). When comparing the DEGs from the a1 and b4 clones with the previously identified *prhP*-dependent targets (Zhang et al. 2019), relatively few overlap was found. PrhP was shown to control expression of the T3SS which is also downregulated in the a1 and b4 clones, but several other PrhP targets (e.g., the type IV pili and flagellar genes or the phenolic acid degradation genes) do not follow this pattern in a1 and b4. This observation suggested that *prhP* is probably an important regulatory node to be downregulated or inactivated, but a complex rewiring of the global regulatory network must take place to lead to partial expression of the PrhP regulon.

A wider comparison of DEG repertoires among clones originating from different populations revealed an indisputable overlap in the transcriptomic profiles of adapted clones, which does not depend solely on common genetic polymorphisms. Figure 5 illustrates the DEG repertoire relationship between five clones originating from four independent evolved populations and reveals a strong convergence of the transcriptomic signatures. For example, clone c1 has up to 85% of its DEGs shared with the four other independently evolved clones. The average overlap of DEGs for each of the five clones ranged from 53% to 85%, whether for down- or upregulated genes. This convergence in terms of transcriptomic response can sometimes be observed with a certain degree of variation within populations: For example, clone b1 has more DEGs in common with c1, from another population, than with clone b4 (the same population), 64% versus 47%, respectively. A similar level of convergence in transcriptomic signatures was also observed when comparing evolved clones with smaller pool of genes (a4, c2, d3, d5, e1, and e3) with predominantly overlapping DEG repertoires between clones (e.g., 87% and 74% of overlap for e3 and d5, respectively). We concluded that the experimental evolution carried out in parallel on five populations propagated in Hawaii 7996 led to the selection of clones exhibiting little genetic polymorphism but with a significant reorganization of gene expression, which strongly converged between the different populations.

Remarkably, 29 common DEGs were found in at least eight of the ten evolved clones (supplementary table S3, Supplementary Material online). This list included one-third of proteins of unknown function and one-third of genes belonging to the *hrpB* regulon and associated to the T3SS (further detailed below). Among the last third were two histone-like proteins and an acyl homoserine lactone-synthase, which are associated in the literature with global reorganization of gene expression either through action on DNA supercoiling or through quorum-sensing-dependent gene expression shifts, respectively (Ali et al. 2014; Hawver et al. 2016). These latter genes therefore appear as candidates in the implementation of the gene expression changes observed in the majority of adapted clones.

Expression of the *efpR* and *hrpB* Regulons Is Consistently Downregulated in Fittest Clones Isolated from Resistant Tomato

In order to uncover the functions involved in adaptive mechanisms of the fittest clones on Hawaii 7996, we performed a Gene Ontology (GO) enrichment analysis on DEGs. This revealed that 1) several functions (such as chemotaxis, locomotion, signaling activity, and metabolic processes) were commonly upregulated in the a1 and c1 clones and 2) genes associated to the protein secretion process were significantly enriched in the downregulated genes of the b1, b4, c1, and c2

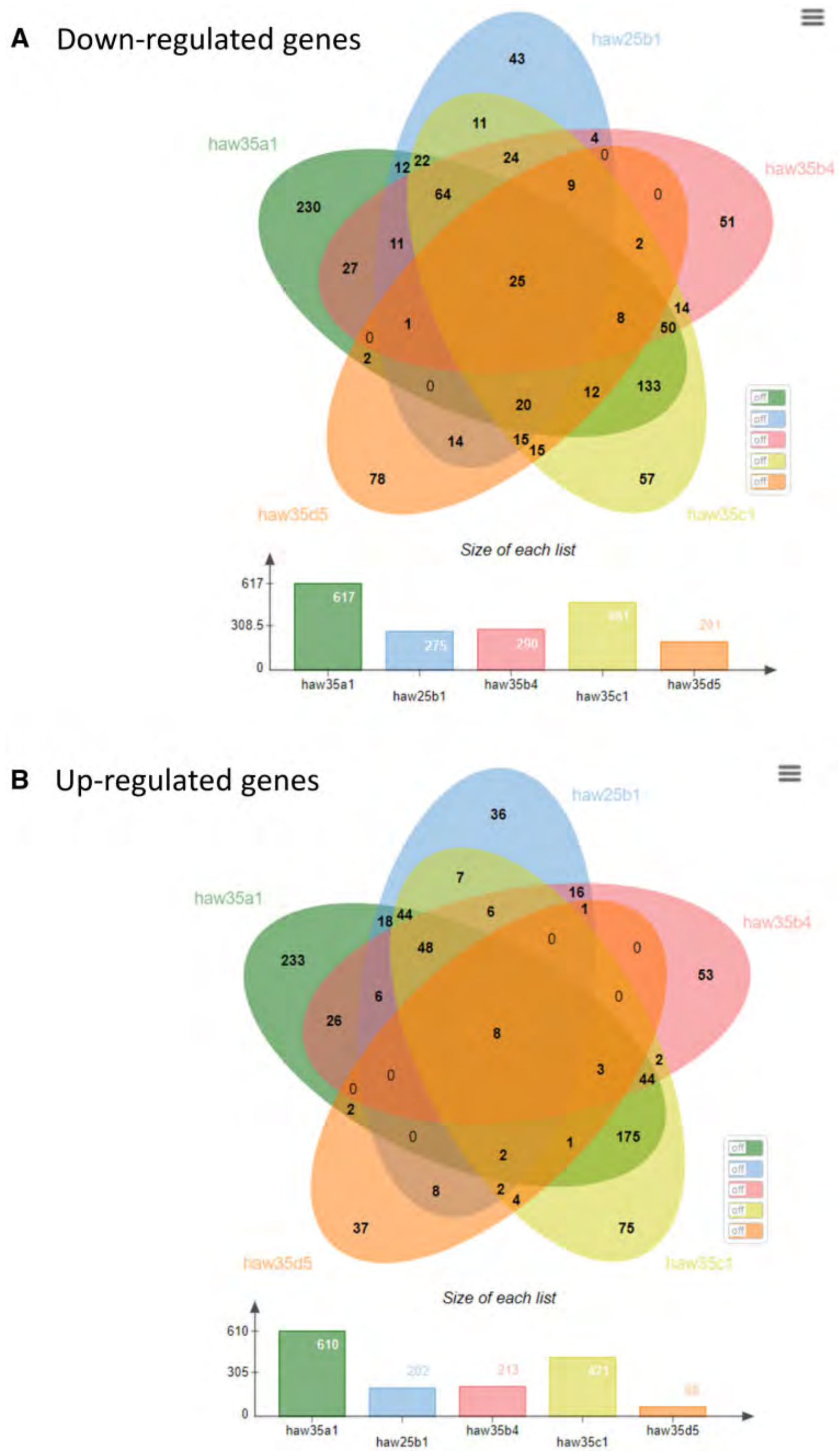


Fig. 5. Grouping of genes differentially expressed (compared with the ancestral clone) of five evolved clones, coming from four independent evolution lines. Clone haw35a1 originates from population A, haw35b1 and haw35b4 from population B, haw35c1 from population C, and haw35d5 from population D. The total number of DEGs in each clone appears at the bottom of Venn's diagram. Genes were considered as differentially expressed with the following thresholds: absolute fold change between strain $|FC| > 2$ and an FDR-adjusted *P*-value (Padj, FDR) < 0.05.

clones (table 3 and supplementary table S4, Supplementary Material online).

Associated to a strong reorganization of gene expression for metabolic functions and increased growth rate compared with the ancestral strain, clones a1 and c1 are characterized by a significant downregulation of the *efpR* gene (table 3). Because *efpR* was previously shown to be a master regulatory gene coordinating expression of multiple virulence and metabolic functions and that loss-of-function mutations were associated to fitness gain in *R. solanacearum* (Perrier et al. 2016; Capela et al. 2017), we suspected that downregulation of *efpR* was a key adaptive event that occurred in a1 and c1. A

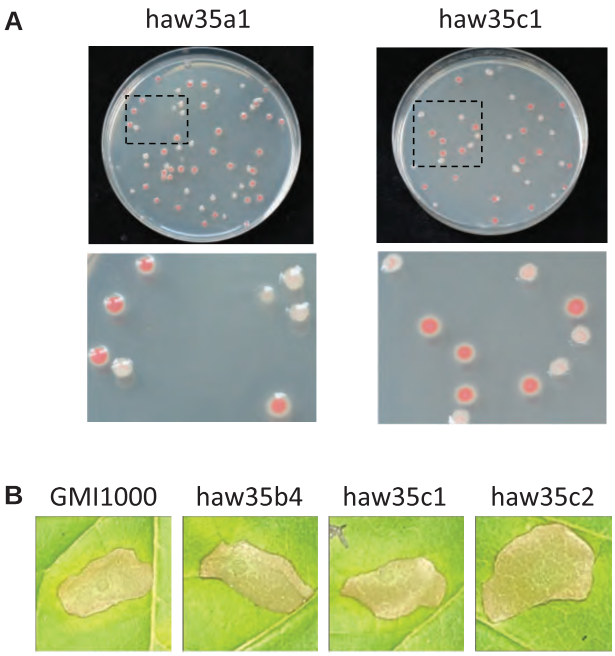


FIG. 6. Phenotypical characterization of evolved clones in which the *efpR* and *hrpB* genes are downregulated. (*A*) The haw35a1 and haw35c1 evolved clones display phenotypic heterogeneity on plates similar to an *efpR* mutant (Perrier et al. 2019). Colonies were grown for 48 h on a complete agar medium supplemented with glucose and triphenyl tetrazolium chloride. (*B*) The haw35b4, haw35c1, and haw35c2 evolved clones characterized by a significant downregulation of the *hrpB* gene show a similar HR response on *N. tabacum* than the ancestral GMI1000 clone. Tobacco leaves were infiltrated with bacterial suspensions at 10^8 CFU/ml. Pictures were taken 48 h after the infiltration. The seven other tested experimentally evolved clones also show a similar HR response on *Nicotiana tabacum* than the ancestral GMI1000 clone (see complete picture in supplementary fig. S6, Supplementary Material online).

closer examination of the DEGs in a1, c1, and an *efpR* mutant (Perrier et al. 2016; Capela et al. 2017) revealed in fact that only one half of the *efpR* regulon is shared with a1 and c1, with a total of 245 DEGs common to the three strains (supplementary fig. S3, Supplementary Material online). This subset of common genes included the chemotaxis and flagellar genes as well as most of the exopolysaccharide (EPS) genes, which are associated to main phenotypic features of the *efpR* mutant (hypermotile, reduced EPS biosynthesis). We investigated if the a1 and c1 clones also showed phenotypic heterogeneity features as described in an *efpR* mutant (Perrier et al. 2019). We thus compared the morphotypes of the colonies on plates. Whereas the WT GMI1000 strain produces colonies with only one morphotype (type S, mucoid), the *efpR* mutant produces two morphotypes, the type S and the nonmucoid type EV (Perrier et al. 2019). The a1 and c1 adapted clones also produced colonies with both the type S and the type EV (fig. 6A), thus confirming the strong downregulation of *efpR* in these two adapted clones and its consequences on major phenotypic traits. The increased growth rate of a1 and c1 in MM supplemented with glutamine or proline similarly as the *efpR* mutant (fig. 4) further supports the idea that downregulation of *efpR* is probably a major contribution to the fitness gain acquired by the a1 and c1 clones in the stem of Hawaii 7996.

The GO enrichment analysis highlighted four clones with significant deregulation of protein secretion processes (b1, b4, c1, and c2), which corresponds to a notable downregulation of the *hrpB* regulon in theses clones (supplementary table S4, Supplementary Material online). In fact, all ten clones except a4 show a more or less pronounced reduction in expression of the *hrpB* regulon. For example, in clones d5, e1, and e3 which have the lower number of DEGs (150–200), there is a subset of 27 downregulated genes belonging to the *hrpB* regulon, mostly comprising genes coding for extracellular products (effectors, pilin) or factors modulating secretion (chaperone, associated helper proteins) (supplementary fig. S4, Supplementary Material online). In clones a1, c1, and b4, downregulation of the T3SS and associated effectors reached up to 40% of the *hprB* regulon (supplementary fig. S5, Supplementary Material online). So we wondered if type 3 secretion was only decreased in its capacities or nonfunctional in the evolved clones. We therefore compared the ability of evolved clones to elicit an HR on resistant plants, a property that depends on the functionality of the T3SS and is controlled by *hrpB* (Genin et al. 1992). The HR test was conducted for the ten evolved clones sequenced previously on both *Nicotiana tabacum* and *N. benthamiana* leaves. The test revealed that the ten evolved clones conserved their ability to elicit an HR on these resistant plants, similarly to the WT GMI1000 clone (fig. 6B and supplementary fig. S6, Supplementary Material online).

Discussion

What Did We Learn from Experimental Evolution of *R. solanacearum* on a Resistant Host?

We used experimental evolution combined with phenotyping, genomics, and transcriptomics to analyze the adaptive potential of the bacterial plant pathogen *R. solanacearum* to overcome the quantitative resistance of the tomato cultivar Hawaii 7996. This breeding line is widely used for management of bacterial wilt as it is the most effective source of resistance against various RSSC strains under different environmental conditions (Grimault et al. 1994; Thoquet et al. 1996; Lebeau et al. 2011; Wang et al. 2013). Hawaii 7996 resistance to the RSSC is expressed in both root and shoot tissues (Planas-Marquès et al. 2020). Previous reports indicating that Hawaii 7996 spatially restricts the bacterial movement in the vasculature (Caldwell et al. 2017) and that the T3SS is still expressed in xylem vessels to counteract plant immunity (Monteiro et al. 2012) support the view that *R. solanacearum* is still exposed to active plant defense responses in our experimental system. Here, we used a stem-inoculation procedure in order to maintain a uniform selective environment within each replicate. After serial passaging over 300 generations, we observed pathogen adaptation to within-plant environment of the resistant cultivar Hawaii 7996 but we did not observe any symptom of the bacterial wilt disease, which suggests that this evolutionary time is not enough to overcome the polygenic resistance of Hawaii 7996 expressed in the shoot tissues. In future work, it might be interesting to compare this result to natural conditions where the pathogen experiences a complete lifecycle including survival in soil and root infection.

Pathogen adaptation was characterized by a better bacterial multiplication rate in planta compared with the ancestor, although this gain was not enough to cause the disease. We showed that three out of the ten tested evolved clones also have a better growth rate than the ancestor in MM supplemented with glutamine, the most abundant amino acid found in tomato xylem (Zuluaga et al. 2013), which most probably explains their better multiplication rate in planta. For two of these clones from population B, we showed that this increased growth rate in presence of glutamine depends on the RSp1574 regulatory gene (fig. 4A). Moreover, the upregulation of transporters in these evolved clones suggests that enhanced growth rates rely on increased metabolic capacities, as we previously observed in other regulatory mutant backgrounds (Perrier et al. 2016; Peyraud et al. 2016). Nevertheless, our study also showed that the fitness gain on tomato Hawaii7996 does not only depend on growth gain (i.e., no significant difference in growth rate between evolved vs. ancestral clone), which suggests the involvement of different mechanism(s) to explain why evolved clones are more competitive than their WT ancestor in planta.

Host Adaptation through Gain of Function Mutations and Regulatory Network Rewiring

Few genomic mutations were detected in the genome of experimentally adapted clones (up to two mutations per clone, half of them having no mutation) after using two distinct sequencing technologies (Illumina/PacBio). This low number of genetic changes is in the range of what was found in a previous evolution experiment of the same *R. solanacearum* strain on other hosts (Guidot et al. 2014). None of the mutations detected in the former experiment were recovered in this one, but two of the mutated genes identified in Hawaii 7996-adapted clones (*soxA1* and RSp1136) were previously found in another experimental evolution aimed to convert *R. solanacearum* into a plant symbiont (Clerissi et al. 2018). Furthermore, we were able to detect an IS movement which leads to inactivation of *prhP*, confirming the role played by mobile genetic elements in bacterial adaptation (Vandecraen et al. 2017).

Using reverse genetic approaches, we demonstrated that five genetic alterations detected in the adapted clones were adaptive mutations, providing a competing advantage over the ancestral strain. With the exception of *prhP*, all these mutations lead to a gain of function. Deletion of the corresponding genes induced instead a decrease in competitiveness (fig. 3), indicating that the WT alleles already contribute to *R. solanacearum*'s fitness in plant. This work uncovered a previously uncharacterized LysR family regulatory protein, RSp1574, for which a nonsynonymous mutation is associated with both better competitiveness, increased growth rate and upregulation of specific metabolic transporters that we hypothesize are RSp1574 regulatory targets. This evolution experiment also shed light on *prhP*, a recently described *R. solanacearum* virulence regulatory gene (Zhang et al. 2019). *prhP* was identified as a positive regulator on detoxification of phenolic acids, a class of secondary metabolites produced by plants and acting as broad antimicrobials (Lowe et al. 2015; Zhang et al. 2019). Interestingly, *prhP* was also found to positively control expression of the T3SS genes, but the mechanism involved remains unclear (Zhang et al. 2019). The transcriptome analysis of the clone b4 carrying the *prhP*$^{IS\text{-}6}$ mutation revealed that *prhP* was indeed strongly downregulated but, surprisingly, the DEGs of clone b4 did not completely reflect the DEGs previously identified in the *prhP* deleted mutant (with the notable exception of T3SS genes). It should be noted, however, that clone b4 also carries an additional mutation in the regulatory gene RSp1574. Altogether, these observations lead us to believe that the adaptive events that occurred in the evolved clones induce a rewiring of the global regulatory network that results in novel gene expression patterns associated to fitness gain. This hypothesis is supported by the finding that this rewiring occurred after modifications affecting important nodes in the regulatory network of *R. solanacearum* (Peyraud et al. 2018) and that these modifications did not appear randomly. It is indeed striking to note that the same regulatory genes are regularly affected by mutations or modifications of expression during *R. solanacearum* experimental evolution. Thus, *efpR* (downregulated in a1 and c1) had already been identified as contributing to adaptation to the plant in two previous evolution experiments (Guidot et al. 2014; Capela et al.. 2017), *prhP* is inactivated by an IS in b4 and downregulated in a1, *hrpB* is also downregulated in three clones, and *lecM*, a lectin involved in a feedback loop on the quorum-sensing signaling (Hayashi et al. 2019), is also downregulated in four of them (table 3). All these genes have been associated, more or less directly, with metabolic or virulence functions (Peyraud et al. 2018).

The rewiring of the regulatory network that occur in evolved clones is certainly complex as illustrated by the divergent expression patterns of the regulatory genes *hrpB* (downregulated in three evolved clones) and *prhI* (upregulated in two evolved clones), previously known to belong to the same signaling pathway (table 3). However, it is striking to observe that although the regulatory network rewiring results from different adaptive events and leads to large scale changes in expression profiling (hundreds of genes), it appears to converge toward largely overlapping transcriptomic or phenotypic responses. For example, clone b4 shares 62% of its DEGs (310 genes) with the independently evolved clone c1 although they do not carry any mutation in common.

Variations exist in the magnitude of the mean CI between the five populations that evolved independently on Hawaii 7996, suggesting either different evolutionary trajectories or a more rapid accumulation of adaptive genomic changes in populations with the highest CI values (Clerissi et al. 2018; Garoff et al. 2020). It is also interesting to note that the majority of clones with the highest CI values are associated with deregulation or mutation of several transcriptional regulators (EfpR, HrpB, PrhP, and RSp1574) which are presumably key nodes of the regulatory network. Their deregulation should efficiently reorientate the RSSC strain GMI1000 physiology to better adapt to the Hawaii 7996 xylem environment.

efpR and *hrpB* Are Major Regulatory Nodes for Bacterial Adaptation to Resistant Tomato

EfpR appeared to be a central node in the control of bacterial physiology and a target for mutation during *R. solanacearum* experimental evolution (Guidot et al. 2014; Perrier et al. 2016; Capela et al. 2017). Here, we did not detect mutations in the *efpR* gene but significant downregulation of this gene in two evolved clones. Transcriptomic analysis revealed that only one half of the known *efpR* regulon (Perrier et al. 2016, 2019; Capela et al. 2017) was shared with these two clones so we suspected that other genetic or epigenetic modifications probably superimpose to the downregulation of *efpR*. In one of the two clones, we detected mutation in the *soxA1* gene. However, recreating the *soxA1*^{C639R} mutation in the WT strain did not enhance the growth rate in glutamine and proline as observed for an *efpR* mutant (fig. 4), suggesting that the *soxA1*^{C639R} mutation is not linked to the downregulation of the *efpR* gene. To support this hypothesis, we also found that recreating the *soxA1*^{C639R} mutation in the WT strain was not enough to replicate the level of fitness gain of the evolved clone, thus suggesting again a probable role of other genetic or epigenetic modifications.

HrpB, the primary activator of T3SS expression and secreted effectors, was significantly downregulated in three clones, but the majority of them showed a reduced expression of several genes belonging to the HrpB regulon. At first glance, this finding seems counterintuitive because T3SS genes are essential for the multiplication of bacteria in the xylem (Vasse et al. 2000) and remains expressed throughout the infection (Jacobs et al. 2012; Monteiro et al. 2012). We showed that all the ten evolved clones conserved their ability to elicit an HR on resistant plants similarly to the WT ancestor, indicating that the T3SS is still functional despite the reduced expression. Two hypotheses can be raised from this result: Either the evolved clones optimized the expression of T3SS in resistant tomatoes by minimizing the cost of its biosynthesis while keeping it functional or/and only a subpart of the HrpB regulon is downregulated to specifically avoid recognition of the plant immune system. This latter hypothesis is raised by the intriguing observation that in several clones, a majority of the HrpB regulon genes found downregulated are not genes directly involved in the biosynthesis of T3SS but mostly effectors or extracellular components of the secretory machinery (Hrp pilin, T3SS chaperones, and helper proteins). For example, effectors, such as RipAA, PopP2, or the harpin RipX, which all elicit an immune response on diverse hosts (Arlat et al. 1994; Deslandes et al. 2003; Poueymiro et al. 2009), had their expression significantly reduced in almost all the evolved clones. One can therefore wonder whether the polygenic resistance of Hawaii 7996 could be attenuated if the ability to recognize effectors by plant receptors is itself reduced due to dampened (or selective) expression of these effectors by bacteria. It is in any case striking to see that the deregulation of part of the *hrpB* regulon (this work) as well as that of *efpR* regulon (Perrier et al. 2016; Capela et al. 2017, this work) is a signature which seems recurrent in a majority of clones having better fitness.

Adaptation Also Probably Relies on Epigenetic Modifications

In this evolution experiment on a resistant tomato, no mutation were detected in several bacterial clones after 300 generations of evolution and extensive resequencing. In contrast, up to 900 DEGs were identified when comparing the transcriptome of clone c1 (with no genetic alteration detected) and the ancestor. We therefore hypothesized that epigenetic modifications could be the cause of such transcriptomic variations and could play a major role in adaptation. Epigenetic modifications, such as DNA methylation, are known to impact gene expression in bacteria (Casadesús and Low 2006; Sánchez-Romero and Casadesús 2019; Vandenbussche et al. 2020). DNA methylation was described in RSSC strains (Erill et al. 2017), thus suggesting that variation of the DNA methylation profile could impact virulence of *R. solanacearum*, as recently reported for the insect pathogen *Photorhabdus luminescens* (Payelleville et al. 2017, 2018). By highlighting the probable role of epigenetic variation in host adaptation, this study encourage future works to consider both genetic and epigenetic mutations in bacterial pathogen adaptive evolution.

Materials and Methods

Bacterial Strains, Plant Material, and Growth Conditions

The evolution experiment was performed with the model *R. pseudosolanacearum* strain GMI1000 (Salanoubat et al. 2002). The list of GMI1000 derivatives used in this work is given in table 4. The bacterial strains were grown at 28 °C (under agitation at 180 rpm for liquid cultures) either in BG complete medium or in MP MM (Plener et al. 2010). The pH of the MP medium was adjusted to 6.5 with KOH. For agar plates, BG medium was supplemented with D-Glucose (5 g/l) and triphenyltetrazolium chloride (0.05 g/l). MP medium was supplemented with L-glutamine (10 mM) and oligoelements (1000 mg/l). Gentamycin (final concentration of 10 mg/l) was added to the media when required.

Table 4. List of GMI1000 Derivatives Used in This Study.

Strain	Genotype	References
GRS540	GMI1000, Gm^R	Guidot et al. (2014)
GRS704	GMI1000, Δ*efpR*, Sp^R	Guidot et al. (2014)
mAG65	GMI1000, ΔRSp0048	This study
mAG66	GMI1000, ΔRSp1574	This study
mAG67	GMI1000, ΔRSc3094	This study
mAG68	GMI1000, *soxA1*^{C639R}, Gm^R	This study
mAG69	GMI1000, RSp1574^{V95L}, Gm^R	This study
mAG70	GMI1000, *prhP*$^{IS\text{-}6}$, Gm^R	This study
mAG71	GMI1000, RSp3094^{R162R}, Gm^R	This study
mAG72	GMI1000, RSp1136$^{C\text{-}218A}$, Gm^R	This study

NOTE.—Sp^R and Gm^R indicate resistance to spectinomycin and gentamycin, respectively.

Four to five-week-old resistant tomato (*Solanum lycopersicum*) cultivar Hawaii 7996 plants were used for the experimental evolution and the in planta bacterial competition assays. Six-week-old *N. tabacum* cv. Bottom Special plants and 4-week-old *N. benthamiana* plants were used for the HR tests. All plants were grown in a greenhouse. Plant experiments were conducted in a growth chamber under the following conditions: 75% humidity, 12 h light at 28 °C and 12 h darkness at 27 °C.

Experimental Evolution

The evolution experiment of GMI1000 strain was performed on the resistant tomato cultivar Hawaii 7996 using the protocol described in Guidot et al. (2014). Overnight culture in liquid complete BG medium from a single colony of GMI1000 was diluted to 10^6 CFU/ml and 10 µl was injected into the plant stem 0.5 cm above the cotyledons using a microsyringe. SPEs were then conducted as follows. At each SPE, the bacterial population was recovered from the plant stem 15 days postinoculation. For that purpose, the whole stem and petioles were cut at 1 cm above the inoculation point, truncated into 1-cm segments, and incubated in 3 ml sterile water at room temperature for 30 min to allow the bacteria to diffuse from stem and petiole to water. Ten microliters of the recovered bacterial suspension diluted 100× was directly injected into the stem of a healthy plant and 1 ml was stored in glycerol at −80 °C. In order to estimate the number of bacterial generations at each SPE, the effective number of bacterial cells injected into the healthy plant at SPE_n was compared with the effective number of cells recovered from the infected plant at SPE_{n+1}. The effective number of bacterial cells recovered from the infected plant and injected into the healthy plant was estimated by plating serial dilutions of the recovered bacterial suspension onto BG complete medium, incubation at 28 °C for 48 h and enumeration of the CFU. Five biological replicates were conducted thus generating five parallel lineages (populations) of clones experimentally derived from GMI1000.

Bacterial Competition Assay

The bacterial competition assay was carried out to determine the in planta fitness of the derived clones and mutants as described previously (Macho et al. 2010; Guidot et al. 2014; Perrier et al. 2016). The assay was performed using a mixed inoculum, consisting of equivalent CFU of a derived clone or mutant and the GMI1000 WT strain or gentamycin-resistant GMI1000 derivative (GRS540 strain). The GMI1000 WT strain was coinoculated with the gentamycin-resistant mutants, and the GRS540 strain was coinoculated with the derived clones and other gentamycin sensitive mutants. As a control, GMI1000 was coinoculated with GRS540. The inoculation dose is an important factor conditioning potential interference between coinoculated strains (complementation and/or dominant negative). Interference takes place when using a high dose of inoculum, whereas lower doses completely avoid this interference (Macho et al. 2007). Here, we injected 10 μl of the mixed inoculum at a 10^6 CFU/ml concentration into the stem of Hawaii 7996 plant as previously used (Guidot et al. 2014). Bacteria were recovered from the plant stem after 15 days, serial diluted and plated on BG complete medium with and without gentamycin. A CI was calculated as the ratio of derived clone (or mutant)/GRS540 (or GMI1000) strain obtained from the plant stem (output) divided by the ratio in the inoculum (input). A minimum of ten replicates were performed for each derived clone and each mutant. Differences between mean CI values were tested using a Wilcoxon test performed in the R statistical software.

In Vitro Growth Analysis

The growth rates of the derived clones and mutants were compared with the growth rate of the ancestral GMI1000 strain in cultures growing in MP MM supplemented with L-glutamine or L-proline at a 10 mM final concentration. Overnight cultures grown at 28 °C and 180 rpm shaking in MP MM supplemented with L-glutamine 10 mM were used to inoculate 200 μl of fresh MP MM supplemented with L-glutamine or L-proline 10 mM with an initial OD_{600nm} at 0.05. Bacterial growth was performed in 96-well microplates incubated at 28 °C, 700 rpm shaking and monitored using a microplate spectrophotometer (FLUOstar Omega, BMG Labtech, Offenburg, Germany). OD_{600nm} was measured every 5 min during 50 h. Three technical and three biological repeats were performed. Differences between in vitro growth rates were tested using a Student *t*-test with the R statistical software.

HR Assays

For HR assays, the bacterial strains were grown overnight in liquid BG complete medium at 28 °C. A bacterial suspension at 10^8 CFU/ml was infiltrated both in the leaves of 6-week-old *N. tabacum* plants and in the leaves of 4-week-old *N. benthamiana* plants with a 1-ml needless syringe. HR assays were performed in a controlled environment room (75% humidity, 12 h light/12 h dark, 28/27 °C). The cell death was checked 48 h after infiltrations.

Genome Sequencing

Based on CI experiment results, ten evolved clones (two from each of the five populations) were selected for genomic sequence analysis. A single colony of each of the ten evolved clones and the ancestral GMI1000 clone were grown overnight in 50 ml MP MM supplemented with 10 mM L-glutamine at 28 °C and 180 rpm shaking. The morning, OD_{600} was adjusted to 0.075 in a final volume of 50 ml MP MM supplemented with glutamine and incubated again at 28 °C and 180 rpm shaking during 8 h. OD_{600} was then adjusted to 0.001 in a final volume of 100 ml for overnight culture. The next morning, bacterial growth was monitored by measuring the OD_{600} every 60 min until the bacterial culture reaches mid exponential growth phase, then OD_{600} was monitored every 15 min until the bacterial culture reaches the stationary phase. Twenty milliliters of bacterial culture was collected at the beginning of stationary phase and centrifuged at 5,000 rpm for 10 min. The pellet was washed with ultrapure water, centrifuged again, and stored at −80 °C until DNA extraction.

DNA extraction was performed as previously described for high molecular weight genomic DNA (Mayjonade et al. 2016; Erill et al. 2017). DNA concentration and quality were measured by spectrometry using the nanodrop (Thermo Fisher Scientific) and fluorometry using the Qubit dsDNA HS Assay Kit (Life Technologies).

The genomic DNAs were first sequenced using the Illumina technology in order to detect any potential SNPs or small InDels between the derived and ancestral clones. DNA sequencing (DNAseq) was performed at the GeT-PlaGe core facility, INRAE Toulouse, France. DNAseq libraries were prepared according to Biooscientific's protocol using the Biooscientific PCR free Library Prep Kit (Perkin-Elmer). Briefly, DNA was fragmented by sonication, size selection was performed using CLEAN CleanPCR beads, and adaptators were ligated to be sequenced. Library quality was assessed using an Advanced Analytical Fragment Analyzer and libraries were quantified by qPCR using the Kapa Library Quantification Kit (Kapa). DNA-seq experiments were performed on an Illumina Miseq using a paired-end read length of 2 × 150 bp with the Illumina Miseq Reagents micro V2 kits (Illumina). More than 12 million paired-end reads (2 × 150 bp) were generated leading to an ~100× total coverage of the reference genome.

All genomic DNAs were also sequenced using the single molecule, real time (SMRT) DNA sequencing technology in order to detect any potential large genomic rearrangements between the derived and ancestral clones. Library preparation was performed at GeT-PlaGe core facility, INRAE Toulouse, France and SMRT sequencing at Gentyane core facility, INRAE Clermont-Ferrand, France. Two libraries of five multiplex samples were performed according to the manufacturer's instructions "Procedure-Checklist-Preparing-Multiplexed-Microbial-SMRTbell-Libraries-for-the-PacBio-Sequel-System." At each step, DNA was quantified using the Qubit dsDNA HS Assay Kit (Life Technologies) and DNA purity was tested using the nanodrop (Thermo Fisher Scientific). Size distribution and degradation were assessed using the Fragment analyzer (AATI) and High Sensitivity Large Fragment 50 kb Analysis Kit (Agilent). Purification steps were performed using AMPure PB beads (PacBio). The ten individual samples (2 µg) were purified, then sheared at 10 kb using the Megaruptor1 system (Diagenode). Using SMRTBell template Prep Kit 1.0 (PacBio), samples (1 µg) were Exo VII-treated before independently going through DNA Damage Repair and End-Repair. Then, barcoded adapters were ligated to each sample separately. Following ligation, 2 × 5 samples were pooled, then digested with Exo III and Exo VII. The two libraries of five samples were purified tree times. The two SMRTbell libraries were sequenced on two SMRTcell on Sequel1 instrument at 6pM with 120-min preextension and 10-h movies using Sequencing Primer V4, polymerase V3, diffusion loading. Using these conditions, the mean reference genome coverage obtained was 320×.

Detection of Genomic Modifications

The BWA v0.7.15-r1140 software was used to map Illumina reads on *R. pseudosolanacearum* GMI1000 genome with BWA-MEM algorithm (bwa mem –M) (Li and Durbin 2010). Mapping results was filtered with samtools v1.3.1 to keep only properly paired reads with mapq ≥1 and remove PCR duplicates (Li et al. 2009). SNP variants were called with samtools mpileup and VarScan v2.4.3 mpileup2snp (–min-coverage 33, –min-read2 15, –min-var-freq 0.4, –min-avg-qual 20) and InDel variants were called with mpileup2indel algorithm and the same parameters as SNP calling (Koboldt et al. 2012).

For Structural Variant (SV) calling, we used the PacBio tools (https://github.com/PacificBiosciences/pbbioconda). The first step was to build Circular Consensus Sequencing (CCS) contigs with pbccs v4.1.0 in order to have high-quality contigs for the following steps. Then, the CCS contigs were mapped with pbmm2 v1.1.0v on *R. pseudosolanacearum* GMI1000 genome with these parameters –min-length 1000 and –preset CCS. Finally, the SV calling was performed with pbsv v2.2.2 and output Variant Call Format (VCF) was filtered to remove artifacts caused by circular genome.

All detected genomic modifications were checked by PCR amplification and sequencing with the Sanger technology using the primers reported in supplementary table S5, Supplementary Material online.

RNA Extraction, Depletion of rRNA, and RNA-Seq

The ten evolved clones selected for genomic sequence analysis were also investigated for transcriptomic analysis. These transcriptomic analyses were conducted from the same bacterial cultures prepared for DNA extraction grown overnight in 50 ml MP MM supplemented with 10 mM L-glutamine. This condition was preferred to the in planta environment to avoid biases associated with bacteria extraction and sequencing, and because glutamine is by far the most abundant amino acid in tomato xylem (Zuluaga et al. 2013). Briefly, 20 ml of the bacterial culture was collected at the beginning of stationary phase for RNA extraction. Three biological replicates were conducted for each of the ten clones and the GMI1000 strain. The bacterial cultures were stopped growing by the addition of 1 ml ethanol/phenol (95:5) to the 20 ml culture and mixed well by vortexing 1 min. The mixture was then centrifuged at 10,000 rpm for 20 min at 4 °C and the pellets were stored at −80 °C until RNA extraction. Total RNA was extracted and ribosomal RNAs were depleted as previously described (Perrier et al. 2016). In order to optimize ribosomal RNA depletion specifically targeting RSSC rRNAs, three novel oligonucleotides were added to the oligonucleotide set used for the ribosomal RNA depletion, 10Sa-75 5′-ATTPATTAACPAGPTGACGPGTC-3′, 10Sa-294 5′-TCAGTLATTTPATTTAALCGPCG-3′, and 23S-1668 5′-GTACLAATTTPCCPAGTTLCTTC-3′. The two first primers target 10Sa RNA (tmRNA) and the last the 23S rRNA. The concentration and quality of the RNA samples were measured by spectrometry using the nanodrop (Thermo Fisher Scientific) and fluorometry using the Qubit (Life Technologies).

Oriented paired-end RNA-seq was performed at the GeT-PlaGe core facility, INRAE Toulouse, France. RNAseq 30 libraries were prepared according to Illumina's protocols using the

Illumina TruSeq Stranded mRNA sample prep kit (Illumina) to analyze mRNA. Briefly, mRNA was selected using poly-T beads. Then, RNA was fragmented to generate double-stranded cDNA, and adaptators were ligated to be sequenced. A total of 11 cycles of PCR were applied to amplify libraries. Library quality was assessed using a Fragment Analyser and libraries were quantified by quantitative PCR using the Kapa Library Quantification Kit (Kapa). RNA-seq experiments were performed on two lanes of an Illumina HiSeq3000 using a paired-end read length of 2 × 150 bp with the Illumina HiSeq3000 sequencing kits (Illumina).

Mapping and Analysis of RNA-seq Data

RNA-seq read pairs were mapped on *R. pseudosolanacearum* GMI1000 genome using the Glint v1.0 rc12 software (https://forge-dga.jouy.inra.fr/projects/glint) with parameters set as follows: matches ≥75 nucleotides, ≤4 mismatches, no gap allowed, only best-scoring hits taken into account. Ambiguous matches (the same best score) were removed.

DEGs were detected with EdgeR Bioconductor package version 3.30.3 (Robinson and Smyth 2008). Genes with no counts across all libraries were discarded prior to further analysis. Normalization was performed using trimmed mean of *M* values method (Robinson and Oshlack 2010). Quality control plots of normalized data sets and reproducibility of biological repeats were generated by principal component analysis using Ade4 version 1.7-15 package (Dray and Dufour 2007). Correlation between biological repeats was estimated by calculating the correlation coefficient of Spearman. differences and similarities in gene expression between clones were tested by calculating the Euclidean distance and shown on heatmaps. Heatmaps were obtained with the package pheatmap version 1.0.12 (https://CRAN.R-project.org/package=-pheatmap) on sample-to-sample Euclidean distances.

Fitted generalized linear models (GLM) with a design matrix multiple factor (biological repeat and factor of interest) were designed for further analyses. The Cox-Reid profile-adjusted likelihood method in estimating dispersions was then used. DEGs were called using the GLM likelihood ratio test using an FDR (Benjamini and Yekutieli 2001) adjusted *P* value <0.05. Clustering on filtered DEGs (*P* value <0.05 in at least one biological condition) was generated with heatmap.2 function as available in the gplots Bioconductor package version 3.0.1. (https://CRAN.Rproject.org/package=gplots) using Ward's minimum variance clustering method on Euclidean (Murtagh and Legendre 2014). Enrichment analysis considering GO was then conducted using the topGO package version 2.40.0 (Alexa et al. 2006). The Venn diagrams were constructed using the jvenn tool (Bardou et al. 2014).

Construction of Mutants

All mutants were constructed in GMI1000 using the multiplex genome editing by natural transformation protocol (MuGENT) as described (Dalia et al. 2014). Primers were designed to amplify 3–3.5 kb arms of homology from either side of the SNPs. PCR was performed on the genomic DNA of the evolved clones. The purified PCR products act as the unselected products, whereas an insertional plasmid cAG10 linearized by *Sca*I acts as the selected product. Natural transformation of *R. solanacearum* was achieved using 300 ng of selected product and 1 μg of unselected marker with the addition of 50 μl competent cells. The transformants were selected on selective medium (rich medium supplemented with gentamycin 10 mg/l) and were validated by PCR amplification and Sanger sequencing using the respective primers (supplementary table S5, Supplementary Material online). The mutants were named mAG68, mAG69, mAG70, mAG71, and mAG72 corresponding to *soxA1*^{C639R}, Rsp1574^{V95L}, Rsp0309$^{ins1.4kb}$, Rsp3094^{R162R}, and Rsp1136–37 respectively.

Gene deletion was performed on GMI1000 for each SNPs (soxA1, Rsp1574, and Rsc3094) using pK18 plasmid. The plasmid has an *sacB* gene and antibiotic marker (kanamycin) for selection with restriction sites *Eco*RI and *Hin*dIII. Two border fragments of each gene (700 bp–1 kb) were PCR amplified separately as follows: 1) downstream fragment with flanking regions—*Eco*RI on the left end and *Nhe*I on the right end and 2) upstream fragment with flanking regions—*Nhe*I on the right end and *Hin*dIII on the left end. The up and downstream fragments were restriction digested using appropriate enzymes and the pK18 plasmid was restriction digested using *Eco*RI and *Hin*dIII. The digested plasmid pK18, upstream, and downstream fragments were ligated overnight using T4DNA ligase. The ligated vector was transformed into competent DH5α *Escherichia coli* cells followed by blue/white selection and the ligation was verified by sequencing. The plasmid DNA was isolated from the transformants and transformed into competent GMI1000 cells. Selection was performed on sucrose medium, since pK18 plasmid has *sacB* gene that hinders their growth on sucrose. This was followed by recombination where the crossover occurs resulting in Δ*soxA1*, ΔRsp1574, and ΔRsp3094 mutants.

The mutants were tested for in planta fitness using the CI method as described above. The insertion mutants competed with the control ancestral GMI1000 clone, whereas the deletion mutants competed with the control Gm^{R} GRS540 clone in the stem of tomato Hawaii 7996. Bacterial recovery was performed at 15 dpi and the CI was calculated.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

The authors thank the members of the RAP team and Julien Brillard for support and advices during the course of this work. R.G.N. was funded by a PhD fellowship from the French Laboratory of Excellence project "TULIP" (Grant Nos. ANR-10-LABX-41; ANR-11-IDEX-0002-02). This work was supported by the French National Research Agency (Grant No. ANR-17-CE20-0005-01), the French Laboratory of Excellence project "TULIP" (Grant No. ANR-10-LABX-41; ANR-11-IDEX-0002-02), and the French Research Federation "FR AIB" (Agrobiosciences Interactions and Biodiversity). Part of this work was carried out on the Toulouse Plant-Microbe

Phenotyping “TPMP” facility, LIPM—UMR INRAE441/ CNRS2594. This work was performed in collaboration with the GeT core facility, Toulouse, France (http://get.genotoul.fr) and was supported by France Génomique National Infrastructure, funded as part of “Investissement d’avenir” program managed by the French National Research Agency (Grant No. ANR-10-INBS-09) and by the GET-PACBIO program (Programme Operationnel FEDER-FSE MIDI-PYRENEES ET GARONNE 2014-2020). Raw sequencing data were deposited on the NCBI Sequence Read Archive, SRP276307.

References

Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22(13):1600–1607.

Ali SS, Soo J, Rao C, Leung AS, Ngai DH-M, Ensminger AW, Navarre WW. 2014. Silencing by H-NS potentiated the evolution of Salmonella. *PLoS Pathog*. 10(11):e1004500.

Arlat M, Van Gijsegem F, Huet JC, Pernollet JC, Boucher CA. 1994. PopA1, a protein which induces a hypersensitivity-like response on specific Petunia genotypes, is secreted via the Hrp pathway of Pseudomonas solanacearum. *EMBO J*. 13(3):543–553.

Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. 2014. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* 15(1):293.

Barrick JE, Lenski RE. 2013. Genome dynamics during experimental evolution. *Nat Rev Genet*. 14(12):827–839.

Benjamini Y, Yekutieli D. 2001. The control of the false discovery rate in multiple testing under dependency. *Ann Stat*. 29:1165–1188.

Caldwell D, Kim B-S, Iyer-Pascuzzi AS. 2017. Ralstonia solanacearum differentially colonizes roots of resistant and susceptible tomato plants. *Phytopathology* 107(5):528–536.

Capela D, Marchetti M, Clérissi C, Perrier A, Guetta D, Gris C, Valls M, Jauneau A, Cruveiller S, Rocha EPC, et al. 2017. Recruitment of a lineage-specific virulence regulatory pathway promotes intracellular infection by a plant pathogen experimentally evolved into a legume symbiont. *Mol Biol Evol*. 34:2503–2521.

Carmeille A, Caranta C, Dintinger J, Prior P, Luisetti J, Besse P. 2006. Identification of QTLs for *Ralstonia solanacearum* race 3-phylotype II resistance in tomato. *Theor Appl Genet*. 113(1):110–121.

Casadesús J, Low D. 2006. Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev*. 70(3):830–856.

Clerissi C, Touchon M, Capela D, Tang M, Cruveiller S, Genthon C, Lopez-Roques C, Parker MA, Moulin L, Masson-Boivin C, et al. 2018. Parallels between experimental and natural evolution of legume symbionts. *Nat Commun*. 9(1):2264.

Corwin JA, Kliebenstein DJ. 2017. Quantitative resistance: more than just perception of a pathogen. *Plant Cell* 29(4):655–665.

Dalia AB, McDonough E, Camilli A. 2014. Multiplex genome editing by natural transformation. *Proc Natl Acad Sci U S A*. 111(24):8937–8942.

Deslandes L, Olivier J, Peeters N, Feng DX, Khounlotham M, Boucher C, Somssich I, Genin S, Marco Y. 2003. Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proc Natl Acad Sci U S A*. 100(13):8024–8029.

Dray S, Dufour A-B. 2007. The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw*. 22:1–20.

Ebert D. 1998. Experimental evolution of parasites. *Science* 282(5393):1432–1436.

Erill I, Puigvert M, Legrand L, Guarischi-Sousa R, Vandecasteele C, Setubal JC, Genin S, Guidot A, Valls M. 2017. Comparative analysis of *Ralstonia solanacearum* methylomes. *Front Plant Sci*. 8:504.

Flavier AB, Clough SJ, Schell MA, Denny TP. 1997. Identification of 3-hydroxypalmitic acid methyl ester as a novel autoregulator controlling virulence in *Ralstonia solanacearum*. *Mol Microbiol*. 26(2):251–259.

Garoff L, Pietsch F, Huseby DL, Lilja T, Brandis G, Hughes D. 2020. Population bottlenecks strongly influence the evolutionary trajectory to fluoroquinolone resistance in *Escherichia coli*. *Mol Biol Evol*. 37(6):1637–1646.

Genin S. 2010. Molecular traits controlling host range and adaptation to plants in *Ralstonia solanacearum*. *New Phytol*. 187(4):920–928.

Genin S, Boucher C. 2004. Lessons learned from the genome analysis of *Ralstonia solanacearum*. *Annu Rev Phytopathol*. 42(1):107–134.

Genin S, Gough CL, Zischek C, Boucher CA. 1992. Evidence that the hrpB gene encodes a positive regulator of pathogenicity genes from *Pseudomonas solanacearum*. *Mol Microbiol*. 6(20):3065–3076.

Grimault V, Gélie B, Lemattre M, Prior P, Schmit J. 1994. Comparative histology of resistant and susceptible tomato cultivars infected by *Pseudomonas solanacearum*. *Physiol Mol Plant Pathol*. 44(2):105–123.

Guidot A, Jiang W, Ferdy JB, Thébaud C, Barberis P, Gouzy J, Genin S. 2014. Multihost experimental evolution of the pathogen *Ralstonia solanacearum* unveils genes involved in adaptation to plants. *Mol Biol Evol*. 31(11):2913–2928.

Hawver LA, Jung SA, Ng W-L. 2016. Specificity and complexity in bacterial quorum-sensing systems. *FEMS Microbiol Rev*. 40(5):738–752.

Hayashi K, Kai K, Mori Y, Ishikawa S, Ujita Y, Ohnishi K, Kiba A, Hikichi Y. 2019. Contribution of a lectin, LecM, to the quorum sensing signalling pathway of *Ralstonia solanacearum* strain OE1-1. *Mol Plant Pathol*. 20(3):334–345.

Hayward AC. 1991. Biology and epidemiology of bacterial wilt caused by *Pseudomonas solanacearum*. *Annu Rev Phytopathol*. 29(1):65–87.

Jacobs JM, Babujee L, Meng F, Milling A, Allen C. 2012. The in planta transcriptome of *Ralstonia solanacearum*: conserved physiological and virulence strategies during bacterial wilt of tomato. *mBio* 3(4):e00114.

Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* 444(7117):323–329.

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 22(3):568–576.

Kwak M-J, Kong HG, Choi K, Kwon S-K, Song JY, Lee J, Lee PA, Choi SY, Seo M, Lee HJ, et al. 2018. Rhizosphere microbiome structure alters to enable wilt resistance in tomato. *Nat Biotechnol*. 36(11):1100–1109.

Lebeau A, Daunay M-C, Frary A, Palloix A, Wang J-F, Dintinger J, Chiroleu F, Wicker E, Prior P. 2011. Bacterial wilt resistance in tomato, pepper, and eggplant: genetic resources respond to diverse strains in the *Ralstonia solanacearum* species complex. *Phytopathology* 101(1):154–165.

Lenski RE. 2017. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J*. 11(10):2181–2194.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

Lowe TM, Ailloud F, Allen C. 2015. Hydroxycinnamic acid degradation, a broadly conserved trait, protects *Ralstonia solanacearum* from chemical plant defenses and contributes to root colonization and virulence. *Mol Plant Microbe Interact*. 28(3):286–297.

Macho AP, Guidot A, Barberis P, Beuzón CR, Genin S. 2010. A competitive index assay identifies several *Ralstonia solanacearum* type III effector mutant strains with reduced fitness in host plants. *Mol Plant Microbe Interact*. 23(9):1197–1205.

Macho AP, Zumaquero A, Ortiz-Martín I, Beuzón CR. 2007. Competitive index in mixed infections: a sensitive and accurate assay for the genetic analysis of *Pseudomonas syringae*-plant interactions. *Mol Plant Pathol*. 8(4):437–450.

Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, Dow M, Verdier V, Beer SV, Machado MA, et al. 2012. Top 10 plant

pathogenic bacteria in molecular plant pathology. *Mol Plant Pathol*. 13(6):614–629.
Mayjonade B, Gouzy J, Donnadieu C, Pouilly N, Marande W, Callot C, Langlade N, Muños S. 2016. Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *BioTechniques* 61(4):203–205.
McDonald BA, Linde C. 2002. Pathogen population genetics, evolutionary potential, and durable resistance. *Annu Rev Phytopathol*. 40(1):349–379.
McGarvey JA, Denny TP, Schell MA. 1999. Spatial-temporal and quantitative analysis of growth and EPS I production by *Ralstonia solanacearum* in resistant and susceptible tomato cultivars. *Phytopathology* 89(12):1233–1239.
Meaden S, Koskella B. 2017. Adaptation of the pathogen, *Pseudomonas syringae*, during experimental evolution on a native vs. alternative host plant. *Mol Ecol*. 26(7):1790–1801.
Monteiro F, Genin S, van Dijk I, Valls M. 2012. A luminescent reporter evidences active expression of *Ralstonia solanacearum* type III secretion system genes throughout plant infection. *Microbiology (Reading)* 158(8):2107–2116.
Murtagh F, Legendre P. 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J Classif*. 31(3):274–295.
Nakaho K, Inoue H, Takayama T, Miyagawa H, Nakaho K, Inoue H, Takayama T, Miyagawa H. 2004. Distribution and multiplication of *Ralstonia solanacearum* in tomato plants with resistance derived from different origins. *J Gen Plant Pathol*. 70(2):115–119.
Payelleville A, Lanois A, Gislard M, Dubois E, Roche D, Cruveiller S, Givaudan A, Brillard J. 2017. DNA adenine methyltransferase (Dam) overexpression impairs photorhabdus luminescens motility and virulence. *Front Microbiol*. 8:1671
Payelleville A, Legrand L, Ogier J-C, Roques C, Roulet A, Bouchez O, Mouammine A, Givaudan A, Brillard J. 2018. The complete methylome of an entomopathogenic bacterium reveals the existence of loci with unmethylated Adenines. *Sci Rep*. 8(1):12091.
Peeters N, Guidot A, Vailleau F, Valls M. 2013. *Ralstonia solanacearum*, a widespread bacterial plant pathogen in the post-genomic era. *Mol Plant Pathol*. 14(7):651–662.
Perrier A, Barlet X, Rengel D, Prior P, Poussier S, Genin S, Guidot A. 2019. Spontaneous mutations in a regulatory gene induce phenotypic heterogeneity and adaptation of *Ralstonia solanacearum* to changing environments. *Environ Microbiol*. 21(8):3140–3152.
Perrier A, Peyraud R, Rengel D, Barlet X, Lucasson E, Gouzy J, Peeters N, Genin S, Guidot A. 2016. Enhanced in planta fitness through adaptive mutations in EfpR, a dual regulator of virulence and metabolic functions in the plant pathogen *Ralstonia solanacearum*. *PLoS Pathog*. 12(12):e1006044-23.
Peyraud R, Cottret L, Marmiesse L, Genin S. 2018. Control of primary metabolism by a virulence regulatory network promotes robustness in a plant pathogen. *Nat Commun*. 9(418). Doi:10.1038/s41467-017-02660-4
Peyraud R, Cottret L, Marmiesse L, Gouzy J, Genin S. 2016. A resource allocation trade-off between virulence and proliferation drives metabolic versatility in the plant pathogen *Ralstonia solanacearum*. *PLOS Pathog*. 12(10):e1005939.
Planas-Marquès M, Kressin JP, Kashyap A, Panthee DR, Louws FJ, Coll NS, Valls M. 2020. Four bottlenecks restrict colonization and invasion by the pathogen *Ralstonia solanacearum* in resistant tomato. *J Exp Bot*. 71(6):2157–2171.
Plener L, Manfredi P, Valls M, Genin S. 2010. PrhG, a transcriptional regulator responding to growth conditions, is involved in the control of the type III secretion system regulon in *Ralstonia solanacearum*. *J Bacterioal*. 192(4):1011–1019.
Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ. 2009. Shades of gray: the world of quantitative disease resistance. *Trends Plant Sci*. 14(1):21–29.
Poueymiro M, Cunnac S, Barberis P, Deslandes L, Peeters N, Cazale-Noel A-C, Boucher C, Genin S. 2009. Two type III secretion system effectors from *Ralstonia solanacearum* GMI1000 determine host-range specificity on tobacco. *Mol Plant Microbe Interact*. 22(5):538–550.
Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 11(3):R25.
Robinson MD, Smyth GK. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* 9:321–332.
Salanoubat M, Genin S, Artiguenave F, Gouzy J, Mangenot S, Arlat M, Billault A, Brottier P, Camus JC, Cattolico L, et al. 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* 415(6871):497–502.
Sánchez-Romero MA, Casadesús J. 2019. The bacterial epigenome. *Nat Rev Microbiol*. 18:7–20.
Song S, Fu S, Sun X, Li P, Wu J, Dong T, He F, Deng Y, Song S, Fu S, et al. 2018. Identification of cyclic dipeptides from *Escherichia coli* as new antimicrobial agents against *Ralstonia solanacearum*. *Molecules* 23(1):214.
Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. 2012. The molecular diversity of adaptive convergence. *Science* 335(6067):457–461.
Thoquet P, Olivier J, Sperisen C, Rogowsky P, Laterrot H, Grimsley N. 1996. Quantitative trait loci determining resistance to bacterial wilt in tomato cultivar Hawaii7996. *Mol Plant Microbe Interact*. 9(9):826–836.
Trivedi P, Wang N. 2014. Host immune responses accelerate pathogen evolution. *ISME J*. 8(3):727–731.
Vandecraen J, Chandler M, Aertsen A, Van Houdt R. 2017. The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev. Microbiol*. 43(6):709–730.
Vandenbussche I, Sass A, Pinto-Carbó M, Mannweiler O, Eberl L, Coenye T. 2020. DNA methylation epigenetically regulates gene expression in Burkholderia cenocepacia and controls biofilm formation, cell aggregation, and motility. *mSphere* 5(4):e00455-20.
Vasse J, Genin S, Frey P, Boucher C, Brito B. 2000. The hrpB and hrpG regulatory genes of *Ralstonia solanacearum* are required for different stages of the tomato root infection process. *Mol Plant Microbe Interact*. 13(3):259–267.
Wang J-F, Ho F-I, Truong HTH, Huang S-M, Balatero CH, Dittapongpitch V, Hidayati N. 2013. Identification of major QTLs associated with stable resistance of tomato cultivar 'Hawaii 7996' to *Ralstonia solanacearum*. *Euphytica* 190(2):241–252.
Wicker E, Grassart L, Coranson-Beaudu R, Mian D, Guilbaud C, Fegan M, Prior P. 2007. *Ralstonia solanacearum* strains from Martinique (French West Indies) exhibiting a new pathogenic potential. *Appl Environ Microbiol*. 73(21):6790–6801.
Wicker E, Grassart L, Coranson-Beaudu R, Mian D, Prior P. 2009. Epidemiological evidence for the emergence of a new pathogenic variant of *Ralstonia solanacearum* in Martinique (French West Indies). *Plant Pathol*. 58(5):853–861.
Zhang Y, Zhang W, Han L, Li J, Shi X, Hikichi Y, Ohnishi K. 2019. Involvement of a PadR regulator PrhP on virulence of *Ralstonia solanacearum* by controlling detoxification of phenolic acids and type III secretion system. *Mol Plant Pathol*. 20(11):1477–1490.
Zuluaga AP, Puigvert M, Valls M. 2013. Novel plant inputs influencing *Ralstonia solanacearum* during infection. *Front Microbiol*. 4:349.

Chapter 3. Two-fold profiling of the experimentally evolved clones: genomic and transcriptomic level

3.1. Introduction

The previous chapter specifically studied the clones evolved on Hawaii tomato. The present chapter focuses on the clones evolved on other hosts such as Eggplant var 'Zebrina' and Bean var blanc 'precoce' of which Zebrina is susceptible to GMI1000 (original host) and Bean is tolerant to the bacterium (distant host). The experimental evolution of GMI1000 was previously performed on various host plants including Zebrina and Bean by serial passage experiment (SPE) for over 300 bacterial generations. A total of 26 SPE was required for the bacteria to reach 300 generations in both the host plants Zebrina and Bean. Five biological replicates were conducted thus generating five independent lineages (population) of clones experimentally derived from GMI1000 (named A, B, C, D and E). From each population five clones were randomly chosen for further experiments (such as a1, a2, a3…e4, e5) and their corresponding CI (competitive index) were already measured (Guidot *et al.*, 2014). This earlier study analyzed 125 clones evolved from various hosts including Zebrina (Zeb) and Bean showed that 80% of the clones had a better fitness in their experimental host. The whole genome sequencing of 50 adapted clones revealed an average of 2.3 genomic polymorphisms per evolved clone in comparison to the ancestral clone. Importantly, the study identified mutations in multiple independently evolved populations in the regulatory gene, *efpR* and its association with fitness gain in the experimental host (Guidot *et al.*, 2014; Perrier *et al.*, 2016). Therefore, in the present study we decided to choose other clones (evolved from Zeb and Bean) for which the genomic and transcriptomic information were not available, in order to identify new possible mechanisms and genes involved in adaptation.

	Zebrina evolved clones							
	Population B			Population C		Population D	Population E	
	b1	**b5**	**c2**	**c3**	**c4**	**d1**	**e1**	**e3**
Mutation	$RSp0083^{IS\text{-}9}$		$ISRso1^{F91L}$	$ISRso1^{F91L}$	dld^{R135S}			
mean CI value	2.7	3.7	2.1	1.6	2.1	<u>0.9</u>	3.6	<u>1.4</u>
Wilcoxon test (*p*-value)	2.65E-05	2.51E-04	1.43E-04	1.21E-02	1.21E-02	0.659	4.76E-05	0.179
No. of DEGs	88	151	6	169	6	19	932	1193
Downregulated genes	21	77	1	126	4	12	366	510
Upregulated genes	67	74	5	43	2	7	566	683
Expression of important genes Genes involved in virulence								
RSc0289 (vsrA)	-	-	-	-	-	-	-	-
RSc1097 (efpR)	-	DOWN logFC = -1.305	-	-	-	-	-	-
RSc2736 (phcS)	-	-	-	-	-	-	DOWN logFC = -1.339	UP logFC = 1.090
RSc2748 (phcA)	-	-	-	-	-	-	-	-
RSp0309 (prhP)	-	-	-	-	-	-	-	-
RSp0338 (epsR)	-	-	-	-	-	-	-	UP logFC = 1.157
RSp0852 (hrpG)	-	-	-	-	-	-	-	-
RSp0873 (hrpB)	-	-	-	DOWN logFC = -1.0866	-	-	-	-
Mutated genes in Zebrina clones								
RSc2664 (dld)	-	-	-	-	-	-	-	-
RSp0127 (tISRso1)	-	-	-	-	-	-	-	-
RSp0083 (Hyp. protein)	-	-	-	-	-	-	DOWN logFC = -1.139	-

Table 3 Characteristics of the analyzed clones evolved on Eggplant Zebrina

The table includes the genomic polymorphism, CI (competitive index), DEGs (differentially expressed genes) observed in the Zebrina evolved clones investigated. The differential expression of some of the well-known virulence regulators were also included. Statistical test were performed using Wilcoxon test for CI and their p-value is given in comparison to the ancestral GMI1000 clone. Statistically not significant CI are underlined.

	Bean evolved clones					
	Population A		Population B			Population C
	a4	**a5**	**b1**	**b3**	**b4**	**c1**
Mutation	RSc2508$^{GA\to G,392}$ rpoB^{D428Y}	RSc2508$^{GA\to G,392}$ RSp0778$^{GA\to G,32}$	RSc2508$^{GA\to G,392}$	RSc2508$^{GA\to G,392}$	RSc2508$^{GA\to G,392}$	efpR^{P93Q} purF$^{C\to T,-88}$
mean CI value	6.1	6.5	4.6	5.3	5.5	6.6
Wilcoxon test (*p*-value)	2.00E-07	2.80E-04	2.92E-05	3.05E-06	1.98E-06	1.03E-06
No. of DEGs	879	1145	1146	406	387	379
Downregulated genes	395	555	534	267	244	254
Upregulated genes	484	590	612	139	143	134
Expression of important genes						
Genes involved in virulence						
RSc0289 (vsrA)	-	-	-	-	-	-
RSc1097 (efpR)	-	-	-	-	-	DOWN logFC = -5.970
RSc2748 (phcA)	-	-	-	-	-	-
RSc2736 (phcS)	-	-	-	-	-	-
RSp0309 (prhP)	DOWN logFC = -1.922	DOWN logFC = -1.786	DOWN logFC = -1.867	DOWN logFC = -1.664	DOWN logFC = -1.545	-
RSp0338 (epsR)	-	-	-	-	-	-
RSp0852 (hrpG)	-	-	-	-	-	-
RSp0873 (hrpB)					DOWN logFC = -1.758	
Mutated genes in bean clones						
RSc1976 (purF)	UP logFC = 1.703	UP logFC = 1.778	UP logFC = 1.877	-	-	-
RSc2508 (hyp. protein)	-	-	-	-	-	-
RSc3034 (rpoB)	UP logFC = 2.893	UP logFC = 2.646	UP logFC = 2.731	UP logFC = 1.724	UP logFC = 2.011	-

Table 4 Characteristics of the analyzed clones evolved on Bean

The table includes the genomic polymorphism, CI (competitive index), DEGs (differentially expressed genes) observed in the Bean evolved clones investigated. The differential expression of some of the well-known virulence regulators and the genes' with genomic polymorphisms detected were also included. . Statistical test were performed using Wilcoxon test for CI and their p-value is given in comparison to the ancestral GMI1000 clone

CI		Zebrina	Hawaii
		p-value Student t-test	
	Mean±SE	2.27±0.333	5.86±0.942
Bean	5.77±0.281	7.91E-06	0.9329
Hawaii	5.86±0.942	0.00572	

Table 5 Statistical analyses of the mean CI obtained from the clones of various experimental host

The statistical analysis to compare the mean CI from each experimental host were performed using student t-test. The underlined value denotes that the mean CI was not significantly different.

3.2 Materials and methods

3.2.1 Selection of clones

The following number of evolved clones were chosen from each host plant: eight clones from eggplant Zebrina and six clones from Bean var Bartolo. Clones evolved on Eggplant Zebrina included those that (i) had no or few mutations (ii) had fitness gain with no mutation and (iii) two clones showing no significant fitness gain compared to the ancestral clone that will act as negative control. Majority of the clones evolved on Bean were chosen with no *efpR* mutation, since the importance of this gene has been studied previously in the team (Perrier *et al.*, 2016, 2018, 2019). The list of the clones chosen along with their CI and genomic polymorphisms is given in Table 3 and Table 4.

3.2.2 Bacterial growth conditions

The bacterial strains were grown at 28°C (under agitation at 180 rpm for liquid cultures) either in BG complete medium or in MP minimal medium. The pH of the MP medium was adjusted to 6.5 with KOH. For agar plates, BG medium was supplemented with D-Glucose (5 g/l) and triphenyltetrazolium chloride (0.05 g/l). The MP medium was supplemented with L-Glutamine (10 mM) and oligo elements (1000 mg/l).

3.2.3 Genomic DNA preparation

The DNA samples for genomic analysis were prepared as described previously (Gopalan-Nair et al., submitted). Briefly, each of the evolved clones and the ancestral clone GMI1000 were grown in MP medium with 10mM glutamine. Samples were collected at the beginning of stationary phase for both DNA and RNA preparation. For whole genome sequencing, 20 ml of the bacterial culture was centrifuged at 5000g for 10 minutes followed by washing the pellets with water and centrifuged again. The pellets were stored at -80°C until DNA extraction. The DNA were prepared based on the protocol described for high molecular weight genomic DNA (Mayjonade *et al.*, 2017).

3.2.4 Detection of genomic polymorphisms

The chosen 14 evolved clones and GMI1000 were sequenced using both Illumina and PacBio sequencing technologies. Illumina technology was used to identify SNPs (Single Nucleotide Polymorphism) and small InDels (Insertion-Deletion), whereas PacBio technology was used to identify large InDels.

3.2.5 Ribosome depletion and RNA sequencing

The samples for RNA sequencing were collected from the same bacterial culture as it was for the DNA genomic analysis. There were three biological replicates per clone. The growth of bacterial cultures were stopped by the addition of 1ml ethanol/phenol (95:5) to 20ml culture and mixed well. The mixture was centrifuged at 10000g for 20 min at 4°C and the pellets were stored at -80°C until extraction. Total RNA was extracted and ribosomal RNAs were depleted as previously described (Perrier et al. 2018; Gopalan-Nair et el., submitted). The quality and quantity of the RNA samples were analyzed using Agilent RNA 6000 Nano and Qubit high sensitivity assay.

3.3 Results

3.3.1 Genomic sequencing analysis divulged between 0 and 2 polymorphisms per clone

The whole genome of the 14 clones evolved on Zebrina and Bean plants were sequenced using both Illumina and Pacbio sequencing technologies to identify the genomic polymorphisms that occurred during experimental evolution.

The comparison of the genomic sequences of the ancestral clone GMI1000 and the eight Zebrina clones unveiled between 0 and 1 polymorphism per clone (Table 3). Both sequencing methods did not detect any polymorphism in four clones of which, two clones showed a fitness gain compared to the ancestral clone (clones Zeb b5 and Zeb e1 having a CI value significantly more than 1; Table 3); and two clones showed no fitness gain (Zeb d1 and Zeb e3 with a CI value not significantly different from 1; Table 3). In Zeb b1, an IS insertion of around 1 kb was detected downstream of the RSp0083 gene, which encodes a hypothetical protein. This insertion mutation was validated by PCR and Sanger sequencing. Clones Zeb c2 and Zeb c3 possessed the same non-synonymous mutation in the RSp0127 gene which encodes an ISRSo1-transposase protein. In clone Zeb c4, another non-synonymous mutation was detected on the RSc2264 gene, which encodes the Dld protein (a putative D-lactate dehydrogenase (cytochrome) oxidoreductase protein). These two mutations on RSp0127 (Zeb c1 and c2) and RSc2664 (Zeb c3) were confirmed by PCR and Sanger sequencing.

The comparison of GMI1000 and the Bean evolved clones showed between 1 and 2 mutations per clone (Table 4). The detected genomic polymorphisms were either a frameshift mutation or nonsynonymous mutation. One of the frameshift mutation affected the gene RSc2508, encoding a HipA domain protein and was detected in all the investigated bean clones except Bean c1. In clone Bean a4, an additional non-synonymous mutation was observed in the RSc3034 gene, which encodes the RpoB protein (DNA-directed RNA polymerase subunit beta). Genomic sequencing was previously performed in Bean c1 (Guidot *et al.*, 2014). The same two mutations were detected in the resequencing using PacBio technology in the present work, in the RSc1097 gene encoding EfpR and in the upstream region (-88 bp) of the RSc1976 gene which encodes the PurF protein (amidophosphoribosyltransferase PurF protein). These two mutations were validated in the previous work using PCR and Sanger sequencing.

The next steps of this work involve identifying the impact of the observed mutations in the fitness by constructing each mutation individually in the ancestral clone and conducting competitive index (CI) experiment of the mutants with the ancestral clone.

Cutoff	Zebrina								Bean						Hawaii									
	b1	b5	c2	c3	c4	d1	e1	e3	a4	a5	b1	b3	b4	c1	a1	a4	b1	b4	c1	c2	d3	d5	e1	e3
-1<logFC<+1 ; pvalue FDR<0,05	88	151	6	169	6	19	932	1193	1518	1912	1848	812	735	790	1227	187	478	503	902	272	125	269	245	212
-1<logFC<+1 ; pvalue FDR<0,01	39	78	5	54	5	5	534	952	1300	1809	1655	603	550	601	1136	100	403	422	804	211	89	186	196	165
-1<logFC<+1 ; pvalue FDR<0,001	20	31	4	25	2	4	376	339	991	1376	1354	393	368	397	980	35	293	340	681	128	51	118	129	114
1.5<logFC<+1.5 ; pvalue FDR<0,05	51	53	3	77	4	14	432	599	879	1145	1146	406	387	379	563	49	173	175	385	103	40	71	74	55
-1.5<logFC<+1.5 ; pvalue FDR<0,01	30	34	3	36	4	5	360	546	853	1103	1103	350	345	329	543	30	153	161	366	91	27	54	64	43
-1.5<logFC<+1.5 ; pvalue FDR<0,001	17	17	3	23	2	4	262	423	720	1008	1009	267	257	264	498	14	124	139	329	54	16	39	47	35
-2<logFC<+2 ; pvalue FDR<0,05	17	23	2	20	3	6	346	300	477	654	664	207	190	194	253	13	60	81	184	38	16	27	22	18

Table 6 Gradient logFC (log fold change) and FDR (pvalue adjusted) for all the studied evolved clones

The ideal cutoff chosen for evolved clones from each host are heighted in yellow.

3.3.2 Transcriptomic analyses of the clones obtained from Zebrina and Bean

Comparative transcriptomic analyses performed on Hawaii tomato adapted clones and the ancestral clone revealed significant variations even in clones with no genomic polymorphisms (Gopalan-Nair et al., submitted). A similar approach was used here to identify the differential transcriptomic profiles of the clones obtained from Zebrina and Bean in comparison to the ancestral clone. The RNA-sequencing was performed on samples grown in minimal media supplemented with 10mM glutamine.

3.3.2.1 Identification of the best cutoffs for DEGs analysis

In order to identify the best cutoffs for the log FC and FDR-adjusted *p*-value for DEGs analysis, an analysis with gradient log FC and FDR-adjusted p-value (*p*value, FDR) was performed. Table 6 consists of the different cutoffs tested and the respective DEGs for each evolved clone. For Zebrina clones we used the same cutoffs as for the Hawaii clones (-1 < logFC > 1; *p*value FDR<0.05). However, for Bean clones due to high number of DEGs, the logFC was increased (-1.5 < logFC > 1.5; *p*value FDR<0.05).

3.3.2.2 Transcriptomic variations in the Zebrina evolved clones

The DEGs between the Zebrina evolved clones and the ancestral clone were analyzed using the aforementioned cutoff (-1 < logFC > 1; *p*value FDR<0.05). The number of DEGs varied significantly between the Zebrina clones: between 6 and 1193 genes were differentially expressed. A comparison of downregulated genes of the clones b1, b5, c3, e1 and e3 (clones having a significant number of DEGS) revealed that only 1 gene (RSc0152 – lipoprotein) was common (Figure 12). Comparison of the upregulated genes of Zeb b1, b5, c3, e1 and e3 also revealed only 1 gene (RSp0724 – hypothetical protein) that was common among the clones (Figure 13).

The numbers of DEGs could not be correlated to the CI value or to the number of genomic polymorphisms. For example, the highest number of DEGs (1193 DEGs) was observed in clone Zeb e3 having a CI value of 1.4 (not significantly different from ancestral clone) and no genomic

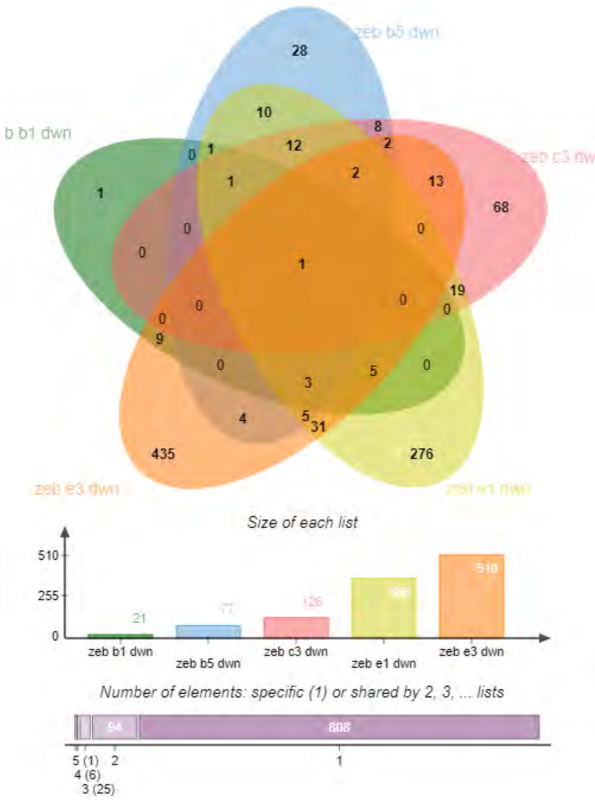


Figure 12 Downregulated genes in Zebrina clones

Venn diagram representing the genes downregulated in Zeb b1, b5, c3, e1 and e3. The graph represents the number of genes in each list. Clones Zeb c2, c4 and d1 were not included in this analysis because the number of down-regulated genes was not significant.

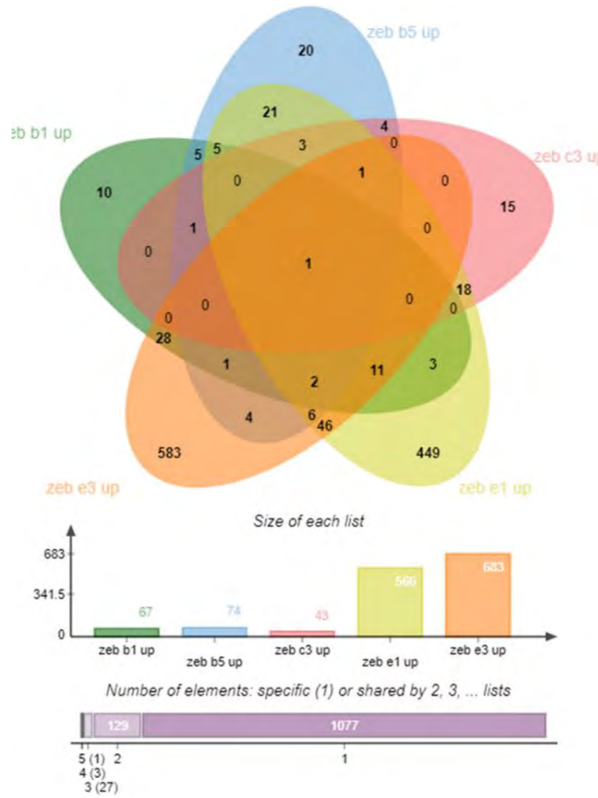


Figure 13 Upregulated genes in Zebrina clones

Venn diagram representing the genes upregulated in Zeb b1, b5, c3, e1 and e3. The graph represents the number of genes in each list. Clones Zeb c2, c4 and d1 were not included in this analysis because the number of up-regulated genes was not significant.

polymorphism. While the lowest number of DEGs (6 DEGs) was detected in two clones from population C (Zeb c2 and c4) having both a CI value of 2.1 (significantly different from the ancestral clone) and one nonsynonymous mutation (Table 3). The clone Zeb c4 carried a nonsynonymous mutation in the *dld* gene, which did not affect the *dld* gene expression (Table 3). A hypothesis is that the mutation enhanced the Dld protein activity without affecting the dld gene expression nor other gene expression. However, experiments still need to be conducted to know if this mutation is an adaptive mutation. The clone Zeb c2 shared the same nonsynonymous mutation in the RSp0127 gene (ISRso1 transposase protein) with the clone Zeb c3. However, Zeb c2 possessed only 6 DEGs while Zeb c3 possessed 169 DEGs (Table 3). Other non-genetic modifications in Zeb c3 could affect gene expression. Interestingly in Zeb c3, the hrpB gene was significantly down-regulated with a logFC = -1.08 (Table 3). However, only 20% of the HrpB regulon could be detected among the Zeb c3 DEGs. The same result has been observed for Hawaii clones and a hypothesis of a dampening of the hrpB activity that control the T3SS, through genetic or epigenetic mechanisms, was suggested.

The two clones from population B, Zeb b1 and b5 showed a significant fitness gain compared to the ancestor with a mean CI of 2.7 and 3.7, respectively, and a significant number of DEGs (88 and 151, respectively) (Table 3). However, the two clones shared only 21 DEGs. In clone b1, the mutation in the RSp0083 gene did not affect its expression. Again, a hypothesis is that the mutation could enhanced the activity of the hypothetical protein RSp0083 without affecting the RSp0083 gene expression. However, similarly to the *dld* gene, experiments still need to be conducted to know if this mutation is an adaptive mutation. In clone b5, no mutations could be detected thus suggested a probable role of non-genetic modifications in transcriptomic variation. Interestingly, in this clone, the *efpR* gene was significantly down-regulated with a logFC = -1.3 (Table 3). Among the 151 DEGs detected in b5, 49 genes (32%) belonged to the EfpR regulon. However, these 49 genes only represented 7% of the EfpR regulon. Again, a hypothesis was that non-genetic modifications could have induced a dampening effect of the EfpR activity.

The clone Zeb d1 with no fitness gain *in planta* had only 19 DEGs compared to the ancestral clone, including the genes *hrpX*, *hrpK* and *ripBJ* that were downregulated. However, before drawing conclusions, the down-regulation of these three important genes needs to be validated using a RT-qPCR approach.

High number of genes were differentially expressed in both clones from population E, Zeb e1 and e3 (932 and 1193 DEGs respectively) in which no mutations were detected. A hypothesis is that non-genetic modifications occurred in these clones and affected gene expression. However, surprisingly, Zeb e3 showed no significantly fitness gain compared to the ancestor, and despite the fact that the clones were from the same population, the list of genes differentially expressed were quite different (only 47 downregulated genes and 67 upregulated genes were shared by the two clones ; Figure 12 and Figure 13). In addition, when analyzing the lists of DEGs, no clear pattern could be extracted. For example, in Zeb e1, three important genes involved in virulence (*phcS*, *phcR*, *phcB)* were downregulated while, *phcS* was upregulated in Zeb e3, and *phcR* and *phcB* were not differentially expressed. Some of the type III secretion proteins (*ripA5* and *ripQ*), prhR were upregulated in Zeb e1 while the same genes were downregulated in Zeb e3. In Zeb e3, the genes associated with the EPS production (EpsB and EpsR) were downregulated and the genes were not found to be differentially expressed in Zeb e1. While few flagellar proteins (FlhC, FlhD and FliK) and pilus protein (PilB) were downregulated in Zeb e1, the flagellar proteins (FlgC, FlhA, FliG, FliL, FliM, FliQ and FliS) and the pilus proteins (PilA, PilB, PilC, PilH, PilJ and PilQ) were upregulated in Zeb e3. These very surprising data needs to be validated by conducting additional biological repeats.

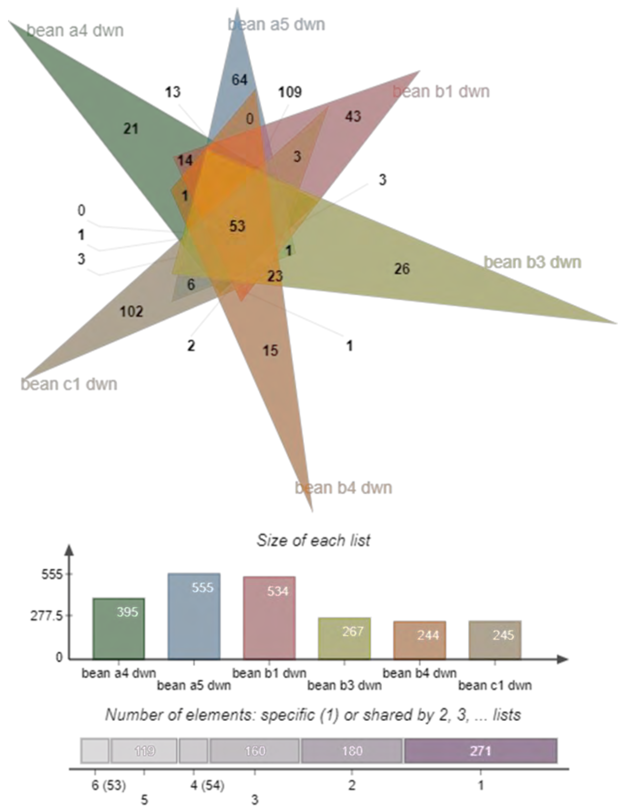


Figure 14 Common downregulated genes in the clones evolved on Bean

The comparison includes the downregulated genes of all six investigated bean clones (a4, a5, b1, b3, b4 and c1). The graph represents the number of genes in each list.

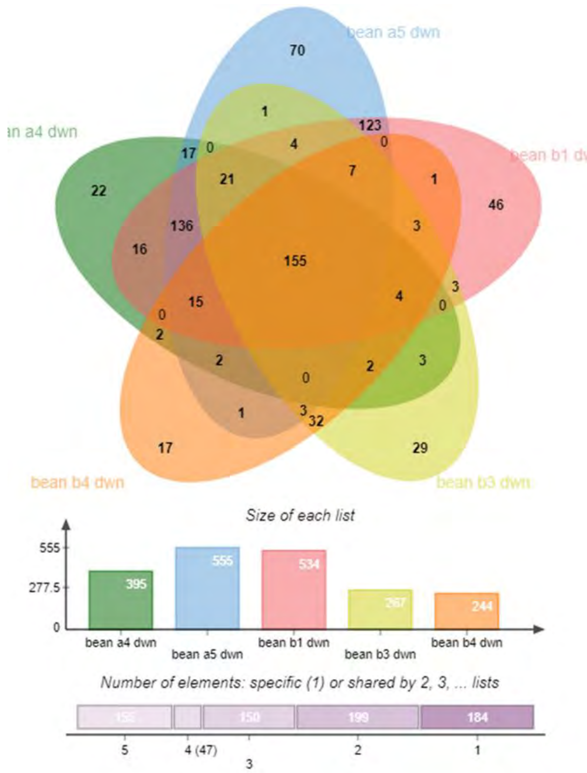


Figure 15 Common downregulated genes in five clones evolved of Bean

The comparison includes the downregulated genes of five of the six investigated bean clones (a4, a5, b1, b3 and b4). Excluding c1 (*efpR* mutated clone) increased the common DEGs shared among the clones. The graph represents the number of genes in each list.

Gene ID	Gene	Description	Classification
RSc1443		putative ligase protein	Macromolecule metabolism
RSc0611	apbE	thiamine biosynthesis lipoprotein apbE precursor transmembrane	Misc.
RSc1454		putative signal peptide protein	Misc.
RSc2279		signal peptide protein	Misc.
RSp1081		signal peptide protein	Misc.
RSc0550		response regulator transcription regulator protein	Misc.
RSp1142		putative transcription regulator protein	Misc.
RSc0055		putative signal peptide protein	Small molecule metabolism
RSc0158		arginase	Small molecule metabolism
RSc0219		putative acylphosphatase protein	Small molecule metabolism
RSc1527	folP	7,8-dihydropteroate synthase protein	Small molecule metabolism
RSc2107	rsl	L-fucose-binding lectin RSL	Small molecule metabolism
RSc2433		putative oxidoreductase protein	Small molecule metabolism
RSc2644	hutI	imidazolonepropionase	Small molecule metabolism
RSc2645	hutG	Formimidoylglutamase	Small molecule metabolism
RSc2646	hutH	histidine ammonia-lyase	Small molecule metabolism
RSc2647	hutU	urocanate hydratase	Small molecule metabolism
RSc2648	hutC	histidine utilization repressor transcription regulator protein	Small molecule metabolism
RSc3286	solI	AUTOINDUCER SYNTHESIS PROTEIN SOLI	Small molecule metabolism
RSc3288	rsl2	Mannose-binding lectin RS-IIL	Small molecule metabolism
RSp0647		putative enoyl-coenzyme A hydratase protein	Small molecule metabolism
RSp0652		acyl-CoA dehydrogenase oxidoreductase protein	Small molecule metabolism
RSc0159		acetylornithine aminotransferase protein	Small molecule metabolism
RSc0161		transmembrane aldehyde dehydrogenase oxidoreductase protein	Small molecule metabolism
RSc2730	pntAa	NAD(P) transhydrogenase alpha subunit	Small molecule metabolism
RSp1111		lipase protein	Small molecule metabolism
RSc1586		lipoprotein	Transport associated
RSc2652		lipoprotein	Transport associated
RSp0569		putative Mannose-binding lectin	Transport associated
RSc0218		hypothetical protein	Unknown function
RSc0221		hypothetical protein	Unknown function
RSc0232		hypothetical protein	Unknown function
RSc0236		hypothetical protein	Unknown function
RSc0293		putative signal peptide protein	Unknown function
RSc1727		hypothetical protein	Unknown function
RSc1874		hypothetical protein	Unknown function
RSc1910		hypothetical protein	Unknown function
RSc2509		hypothetical protein	Unknown function
RSc2886		hypothetical protein	Unknown function
RSc2888		hypothetical protein	Unknown function
RSc3265		hypothetical protein	Unknown function
RSc3292		hypothetical protein	Unknown function
RSc3370		hypothetical protein	Unknown function
RSp0026		hypothetical protein	Unknown function
RSp1109		hypothetical protein	Unknown function
RSp1526		hypothetical protein	Unknown function
RSp1527		hypothetical protein	Unknown function
RSp1663		hypothetical protein	Unknown function
RSc2902		lipoprotein	Unknown function
RSc3315		putative prolin-rich transmembrane protein	Unknown function
RSp0169		putative transmembrane protein	Unknown function
RSc1800	ripG4	T3E	Virulence associated
RSc3272	ripAM	T3E	Virulence associated

Table 7 List of genes downregulated in all the studied Bean clones

The list of genes includes description of the genes with their ID and their functional classification.

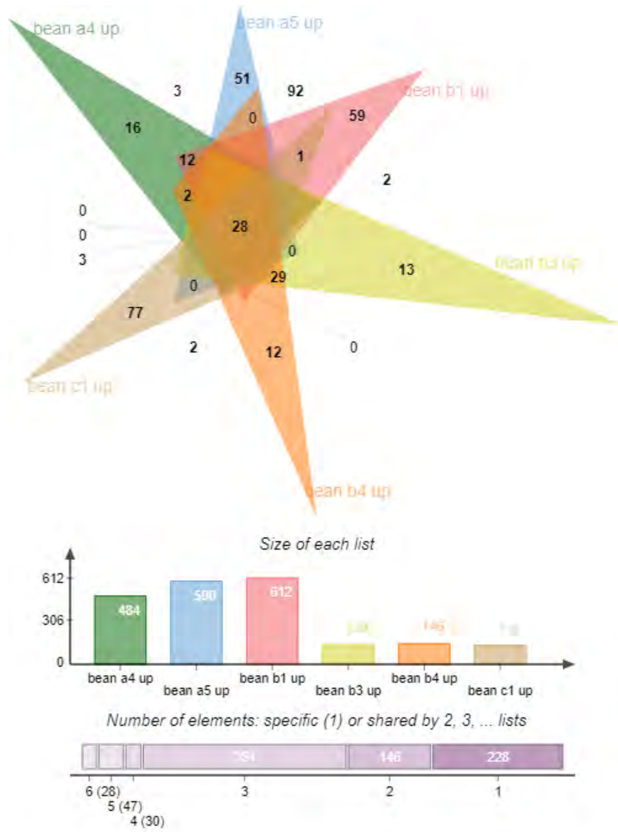


Figure 16 Common upregulated genes in the clones evolved on Bean

The comparison includes the upregulated genes of all six investigated bean clones (a4, a5, b1, b3, b4 and c1). The graph represents the number of genes in each list.

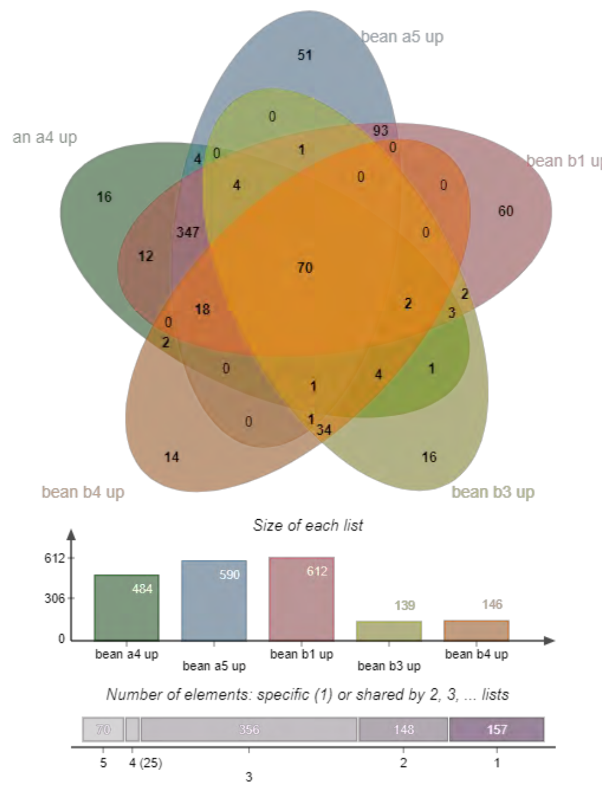


Figure 17 Common upregulated genes in five clones evolved of Bean

The comparison includes the upregulated genes of five of the six investigated bean clones (a4, a5, b1, b3 and b4). Excluding c1 (*efpR* mutated clone) increased the common DEGs shared among the clones. The graph represents the number of genes in each list.

Gene ID	Gene	Description	Classification
RSc1164	ppiB	peptidyl-prolyl cis-trans isomerase B	Chaperoning
RSc3236		putative bacteriophage-related protein	External elements
RSp0490		putative signal peptide protein	Misc.
RSc1276		cytochrome C oxidase subunit I	Small molecule metabolism
RSc1991	gltA	type II citrate synthase	Small molecule metabolism
RSc2244		substrate-binding periplasmic ABC transporter protein	Small molecule metabolism
RSc2749	gapA	glyceraldehyde-3-phosphate dehydrogenase	Small molecule metabolism
RSp0779		autoinducer synthase	Small molecule metabolism
RSp0781		Cys/Met metabolism, pyridoxal phosphate-dependent enzyme	Small molecule metabolism
RSp0782		putative aminoacyl-tRNA synthetase protein	Small molecule metabolism
RSp0783		Acyl carrier protein-like	Small molecule metabolism
RSp0784		putative omega-3 fatty acid desaturase transmembrane protein	Small molecule metabolism
RSp0787		putative transmembrane protein	Small molecule metabolism
RSp1047		transporter transmembrane protein	Small molecule metabolism
RSp1431		putative oxidoreductase signal peptide protein	Small molecule metabolism
RSc2967		outer membrane signal peptide protein	Structural elements
RSc0114		hypothetical protein	Unknown function
RSc2009		hypothetical protein	Unknown function
RSp0724		hypothetical protein	Unknown function
RSp0773		putative signal peptide protein	Unknown function
RSp0785		hypothetical protein	Unknown function
RSp0786		hypothetical protein	Unknown function
RSp1427		hypothetical protein	Unknown function
RSp1428		hypothetical protein	Unknown function
RSp1432		hypothetical protein	Unknown function
RSp1433		hypothetical protein	Unknown function
RSp1434		hypothetical protein	Unknown function
RSp1437		hypothetical protein	Unknown function

Table 8 List of genes upregulated in all the studied Bean clones

The list of genes includes description of the genes with their ID and their functional classification.

Gene ID	Gene	Description	Classification
RSp1499		putative pili assembly chaperone transmembrane protein	Cell processes#
RSp0277	treA	periplasmic alpha,alpha-trehalase signal peptide protein	Cell processes$
RSp0333	osmB2	putative osmotically inducible lipoprotein B2 transmembrane	Cell processes$
RSc0216		transport transmembrane protein	Cell processes* (+9 genes)
RSp0579		putative bacteriophage-related protein	Elements of external origin
RSc2287		putative thiol:disulfide interchange protein	Macromolecule metabolism
RSc3152		putative dTDP-4-dehydrorhamnose reductase	Macromolecule metabolism
RSp0162	egl	endoglucanase precursor (endo-1,4-BETA- glucanase) protein	Macromolecule metabolism
RSp0212		Lignostilbene-alpha,beta-dioxygenase related enzyme	Macromolecule metabolism
RSp0239		1,4-alpha-glucan branching enzyme (Alpha amylase catalytic domain)	Macromolecule metabolism
RSp0275		hydrolase transmembrane protein	Macromolecule metabolism
RSp0924		Glycoside hydrolase, putative chitinase	Macromolecule metabolism
RSp1573		putative metallopeptidase	Macromolecule metabolism
RSc0590		putative transcription regulator protein	Not classified regulator (+7 genes)
RSc0050		hypothetical protein	Small molecule metabolism
RSc0215		short chain dehydrogenase	Small molecule metabolism
RSc0217		oxidoreductase transmembrane protein	Small molecule metabolism
RSc0343	amtB	ammonium transporter	Small molecule metabolism
RSc0536	fabG	3-ketoacyl-(acyl-carrier-protein) reductase	Small molecule metabolism
RSc1013		putative ribokinase protein	Small molecule metabolism
RSc1501	mel	Tyrosinase	Small molecule metabolism
RSc1505		alcohol dehydrogenase-like oxidoreductase protein	Small molecule metabolism
RSc1546		Peroxidase / Pyruvate formate lyase - related protein	Small molecule metabolism
RSc1728		hypothetical protein	Small molecule metabolism
RSc1730		dehydrogenase/reductase oxidoreductase protein	Small molecule metabolism
RSc2025		hypothetical protein	Small molecule metabolism
RSc2921	ggt2	gamma-glutamyltranspeptidase precursor	Small molecule metabolism
RSc3138		oxidoreductase protein	Small molecule metabolism
RSc3151		L-Phenylalanine oxidase	Small molecule metabolism
RSc3287	solR	TRANSCRIPTION REGULATOR TRANSCRIPTIONAL ACTIVATOR PROTEIN SOLR	Small molecule metabolism
RSp0163		putative phosphopantetheinyl transferase protein	Small molecule metabolism
RSp0483		ornithine cyclodeaminase	Small molecule metabolism
RSp0487		hypothetical protein	Small molecule metabolism
RSp0648		enoyl-CoA hydratase	Small molecule metabolism
RSp0649	mmsB	3-hydroxyisobutyrate dehydrogenase oxidoreductase protein	Small molecule metabolism
RSp0650	mmsA	methylmalonate-semialdehyde dehydrogenase oxidoreductase protein	Small molecule metabolism
RSp0702		hypothetical protein	Small molecule metabolism
RSp0722		L-aspartate dehydrogenase	Small molecule metabolism
RSp1058	adc	acetoacetate decarboxylase	Small molecule metabolism
RSp1059		3-hydroxybutyrate dehydrogenase	Small molecule metabolism
RSp1060		putative transmembrane protein	Small molecule metabolism
RSp1124		putative NADP-dependent zinc-type alcohol dehydrogenase oxidoreductase protein	Small molecule metabolism
RSp1187	ndeD	N-acyl-D-glutamate deacylase protein	Small molecule metabolism
RSp1418		hydrolase protein	Small molecule metabolism
RSp1419	ralA	Ralfuranone biosynthesis protein	Small molecule metabolism
RSp1530		putative L-ascorbate oxidase (ASCORBASE) oxidoreductase protein	Small molecule metabolism
RSp1636		putative oxidoreductase protein	Small molecule metabolism
RSp1498	papC	P pilus assembly protein, porin PapC	Structural elements
RSc0334		hypothetical protein	Unknown function (+37 genes)

Table 9 List of downregulated genes specific to the studied Bean clones except c1

The list of genes includes description of the genes with their ID and their functional classification. Cell processes# - Chaperoning; Cell processes$ - Osmotic adaptation; Cell processes* - Transport of small molecules

Gene ID	Gene	Description	Classification
RSc2526	sodB	SUPEROXIDE DISMUTASE [FE]	Cell processes+
RSc1818	metG2	methionyl-tRNA synthetase	Macromolecule metabolism
RSc3033	rpoC	DNA-directed RNA polymerase subunit beta'	Macromolecule metabolism
RSc3034	rpoB	DNA-directed RNA polymerase subunit beta	Macromolecule metabolism
RSp1098		putative HNS-like transcription regulator protein	Not classified regulator
RSc0391		ferredoxin protein	Small molecule metabolism
RSc1275		Cytochrome oxidase maturation protein cbb3-type	Small molecule metabolism
RSc1277		cytochrome C oxidase subunit II	Small molecule metabolism
RSc1279		cytochrome C oxidase subunit III	Small molecule metabolism
RSc1280		iron-sulfur 4Fe-4S ferredoxin transmembrane protein	Small molecule metabolism
RSc1431	guaA	bifunctional GMP synthase/glutamine amidotransferase protein	Small molecule metabolism
RSc2075	ilvC	ketol-acid reductoisomerase	Small molecule metabolism
RSc3284	hemN	coproporphyrinogen III oxidase	Small molecule metabolism
RSc3316	atpC	F0F1 ATP synthase subunit epsilon	Small molecule metabolism
RSp0678	pdxH2	pyridoxamine 5'-phosphate oxidase	Small molecule metabolism
RSp0780		hypothetical protein	Small molecule metabolism
RSp0788		putative transferase protein	Small molecule metabolism
RSp0790		putative hydrolase protein	Small molecule metabolism
RSp0791		hypothetical protein (Acyl carrier protein-like)	Small molecule metabolism
RSp0792		putative fatty acid desaturase transmembrane protein	Small molecule metabolism
RSp1424	ralD	Ralfuranone biosynthesis protein [aminotransferase]	Small molecule metabolism
RSp1425	pvdA	L-ornithine 5-monooxygenase oxidoreductase protein	Small molecule metabolism
RSp1659		acyl-carrier-protein	Small molecule metabolism
RSc0394	rplY	50S ribosomal protein L25/general stress protein Ctc	Structural elements
RSc1048	rpmF	50S ribosomal protein L32	Structural elements
RSc1401		Pseudouridine synthase	Structural elements
RSc3000	rplO	50S ribosomal protein L15	Structural elements
RSc3001	rpmD	50S ribosomal protein L30	Structural elements
RSc3007	rplE	50S ribosomal protein L5	Structural elements
RSc3010	rpsQ	30S ribosomal protein S17	Structural elements
RSc3012	rplP	50S ribosomal protein L16	Structural elements
RSc3020	rpsJ	30S ribosomal protein S10	Structural elements
RSc3035	rplL	50S ribosomal subunit protein L7/L12	Structural elements
RSc3038	rplK	50S ribosomal protein L11	Structural elements
RSp1016	wecC	UDP-N-acetyl-D-mannosamine dehydrogenase	Structural elements
RSc1282		hypothetical protein	Unknown function
RSc2832		hypothetical protein	Unknown function
RSp0261		putative transmembrane protein	Unknown function
RSp0760		hypothetical protein	Unknown function
RSp1426		hypothetical protein	Unknown function
RSp1430		hypothetical protein	Unknown function
RSp1277	ripQ	T3E	Virulence

Table 10 List of upregulated genes specific to the studied Bean clones except c1

The list of genes includes description of the genes with their ID and their functional classification. Cell processes+ - Detoxification

3.3.2.3 Transcriptome variations in the Bean evolved clones

Owing to its high number of DEGs in all the clones evolved on Bean, the fold change cutoff of the genes to be assessed as differential expression in comparison to the ancestral clone GMI1000 was increased (-1.5 < logFC > 1.5; *p*value FDR<0.05). The analysis with these cutoffs perceived between 379 and 1146 DEGs in the examined six Bean clones in comparison to the ancestral clone GMI1000 (Table 4).

Comparison of the downregulated genes from the six Bean clones found 53 genes that were shared among all the clones (Figure 14). A major proportion of these genes were associated with metabolic functions and unknown function (38% and 42% respectively) while a few genes were associated with virulence (2 genes) and transport (3 genes) (Table 7). It was interesting to see that Bean c1 did not share many genes among the clones and had the most clone specific downregulated genes (42%). Therefore, another analysis of the bean clones excluding bean c1 was performed and indeed, the genes shared among the five bean clones increased to 155 genes (Figure 15) and the genes that were specific to only these five bean clones were 102 downregulated genes (Table 9). Majority of the genes encoded for unknown function (38 genes), small molecule metabolism (33 genes) while eight genes were involved in macromolecule metabolism (including *egl* known for its involvement in virulence) and regulators (not classified) each. A significant proportion of the genes were involved in cell processes (13%) such as transport of small molecules, osmotic adaptation and chaperoning (Table 9).

A parallel comparison of the upregulated genes of the six bean clones showed that 28 genes were shared among the clones (Figure 16). Again, majority of the upregulated genes were also associated with metabolic functions and unknown functions (49% each) (Table 8). The same analysis excluding Bean c1 clone increased the common genes shared among Bean a4, a5, b1, b3 and b4 clones to 70 genes (Figure 17). The number of genes specific to this group were 42 upregulated genes and the classification majorly composed of genes encoding small molecule metabolism and structural elements (Table 10).

The frameshift mutation detected in the HipA domain protein (RSc2508) of five of the Bean clones did not affect the gene expression in any of the bean clones (Table 4). However, experiments still need to be conducted to analyze its importance in adaptation.

A parallelism between the bean clones was observed in expression modification of the *rpoB*, and *purF* genes that were mutated in at least one of the clones. The gene *rpoB* (RNA polymerase subunit beta) was mutated in clone Bean a4 and was found to be upregulated in all the investigated Bean clones except c1 (Table 4; Table 10). The gene *purF* (amidophosphoribosyl transferase protein) was mutated in clone Bean c1, but surprisingly, was not affected in gene expression in this clone but was upregulated in Bean a4, a5 and b1 in which the gene was not mutated (Table 4). Gene expression of the *purF* gene has to be validated in the Bean c1 clone using a RT-qPCR approach. Interestingly, the *prhP* gene that was mutated in a Hawaii clone was down regulated in all Bean clones but c1 (Table 4). These results thus suggested a parallelism in the genes that are targeted by either genetic or non-genetic modifications, affecting gene expression and probably involved in bacterial adaptation.

3.3.2.4 DEGs of the evolved clones and the expression of important (global) regulators of *R. solanacearum* GMI1000

As a rule of thumb, the list of genes differentially expressed were looked for the virulence genes and global regulators that are known to mediate adaptation of the pathogen: the genes that were particularly considered were *vsrA*, *efpR*, *phcA*, *phcS*, *prhP*, *epsR*, *hrpG* and *hrpB* (Genin and Denny, 2012; Perrier *et al.*, 2016). The expression of the global regulatory genes phcA, vsrA and hrpG was never found to be affected in neither of the investigated evolved clones. The *efpR* gene was downregulated in one of the clones from Zebrina (Zeb b5). In the Bean clones, *efpR* was downregulated only in clone Bean c1 in which the gene was mutated. However, in the other investigated Bean clones (except Bean b3) the log FC for the *efpR* gene was close to the -1.5 cutoff. The *phcS* gene was downregulated in Zeb e1 (logFC = -1.339) and the same gene was upregulated in Zeb e3 (logFC = 1.090) but this has to be confirmed using a RT-qPCR approach. The transcription regulator PrhP was found to be downregulated in all the Bean clones except

Bean c1 and in none of the clones evolved on Zebrina. The negative regulator of exopolysaccharide production protein (EpsR) was upregulated only in Zeb e3. The regulatory transcription protein (HrpB) was found to be downregulated in Zeb c3 and in Bean b4. These results highlight the importance of de-regulation of the *prhP*, *hrpB* and *efpR* genes in pathogen adaptation to various host plants.

3.3.2.5 Study of all the differentially expressed genes to understand the globally affected genes from each host

An analysis of the globally affected genes from each host was conducted by combining all the differentially expressed genes from each host to obtain an extensive list of DEGs (same cutoff as before for each host). This strategy allowed comparing and studying the complete list of genes that were differentially expressed in a particular host. The comparison included the DEGs of clones evolved on Hawaii (Gopalan-Nair et al., submitted), in addition to Zebrina and Bean to observe the pattern in the genes that were differentially expressed on varying host.

A comparison of all the DEGs observed on the three evolved clones revealed that a significant number of genes (668 genes) were common among the clones in spite of the varying experimental host (Figure 18). This number represented 34%, 32% and 40% of all the DEGs from Hawaii, Zebrina and Bean clones, respectively.

The distribution of the host specific DEGs (excluding the genes regulated by HrpB and EfpR) based on their classification for both Zebrina (420 genes) and Bean (158 genes) clones are given in donut plots (Figure 19). In Zebrina evolved clones, a major proportion (42%) of genes differentially expressed belonged to metabolic functions while 23% of the genes were hypothetical proteins with unknown functions. Approximately 1/10th of the DEGs belonged to unclassified regulator genes and transport genes. A small portion of DEGs were involved in cellular

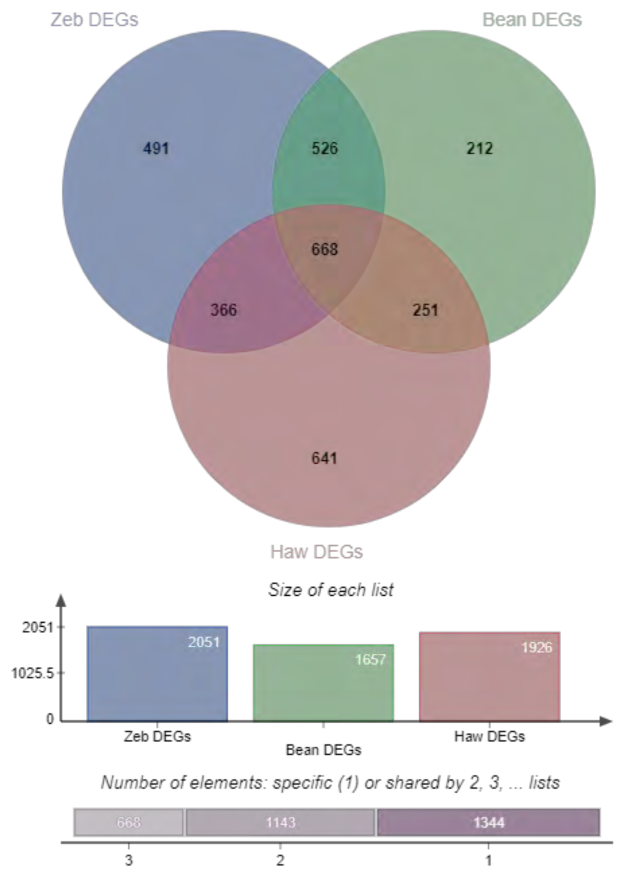
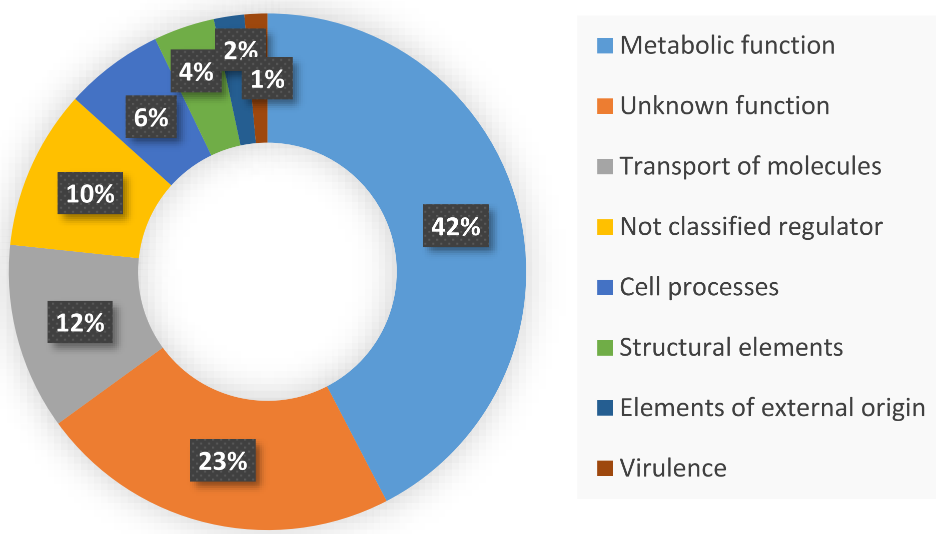


Figure 18 Comparison of all DEGs from Hawaii, Zebrina and Bean clones

The graph represents the number of genes in each list.

a.



b.

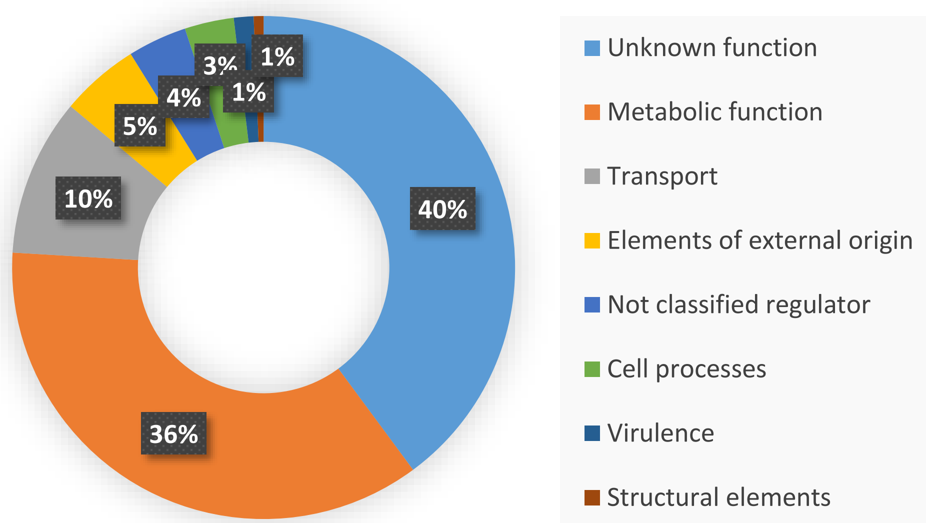


Figure 19 Functional distribution of host specific DEGs in Zebrina and Bean

(a) Host specific differentially expressed genes in Zebrina evolved clones
(b) Host specific differentially expressed genes in Bean evolved clones

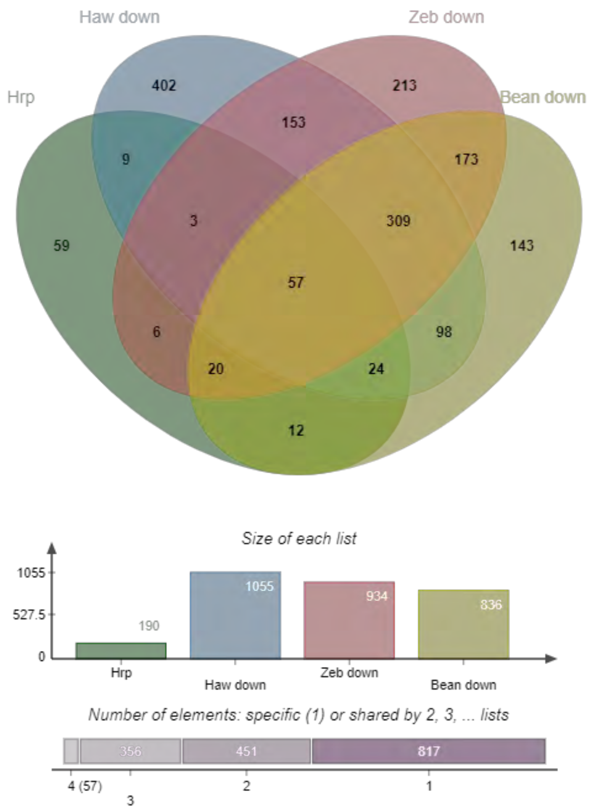


Figure 20 Comparison of downregulated DEGs from the evolved clones (Hawaii, Zebrina, and Bean) to the downregulated gens of hrpB regulon

The graph represents the number of genes in each list.

Gene ID	Gene	Description
RSc0111		sulfur carrier protein ThiS
RSc0235		hypothetical protein
RSc0245	ripB	Type III effector protein
RSc0257		Type III effector protein with ankyrin repeats
RSc0535		putative signal peptide protein
RSc0616		hypothetical protein
RSc0617		putative secreted aspartic protease
RSc0868	popP2	Type III effector protein (acetyltransferase) PopP2
RSc1357	gala7	Type III effector protein GALA7
RSc1474		probable Type III effecor chaperone
RSc1475		Type III effector protein
RSc1735		two-component transmembrane sensor kinase transcription regulator protein
RSc1799		branched-chain alpha-keto acid dehydrogenase subunit E2
RSc1800	gala4	Type III effector protein GALA4
RSc2359		Type III effector protein
RSc2775	popW	Type III effector protein, harpin with pectate lyase domain
RSp0119		hypothetical protein
RSp0120	acnA2	aconitate hydratase
RSp0121	prpC	methylcitrate synthase
RSp0165		hypothetical protein
RSp0213		Type III effector protein
RSp0304		Type III effector protein
RSp0647		putative enoyl-coenzyme A hydratase protein
RSp0648		enoyl-CoA hydratase
RSp0649	mmsB	3-hydroxyisobutyrate dehydrogenase oxidoreductase protein
RSp0822		Type III effector protein
RSp0825		hypothetical protein
RSp0839		hypothetical protein
RSp0840		putative methyl-accepting chemotaxis transducer transmembrane protein
RSp0841		putative lipoprotein
RSp0842		putative Type III effector protein with leucine-rich-repeats
RSp0845		Type III effector protein
RSp0847	awr4	Type III effector protein, AWR family, AWR protein 4
RSp0848		hypothetical protein
RSp0855	hrpY	Hrp pilus subunit HRPY protein
RSp0856	hrpX	hypothetical protein
RSp0857	hrpW	HRPW transmembrane protein
RSp0858	hrpV	hypothetical protein
RSp0859	hrcS	Hrp conserved HRCS transmembrane protein
RSp0860	hrcR	type III secretion system protein
RSp0861	hrcQ	Hrp conserved protein HRCQ
RSp0863	hrcV	Hrp conserved HRCV transmembrane protein
RSp0864	hrcU	type III secretion system protein HrcU
RSp0865	hrpK	HRPK protein
RSp0866	hrpJ	HRPJ protein
RSp0867	hrcJ	Hrp conserved lipoprotein HRCJ transmembrane
RSp0868	hrpH	HRPH protein
RSp0871	hrpD	hypothetical protein
RSp0872	hrcT	Hrp conserved HRCT transmembrane protein
RSp0876	popB	Type III effector protein PopB
RSp0877	popA	Type III effector protein PopA
RSp0879		Type III effector protein
RSp0882		hypothetical protein
RSp0885		Type III effector protein
RSp1024	awr5	Type III effector protein, AWR family, AWR protein 5
RSp1239		Type III effector protein
RSp1566	coxO	cytochrome-C oxidase oxidoreductase protein

Table 11 Common downregulated genes of HrpB cluster shared among the downregulated genes of all Zebrina, Bean and Hawaii clones

The list of genes includes description of the genes with their gene ID.

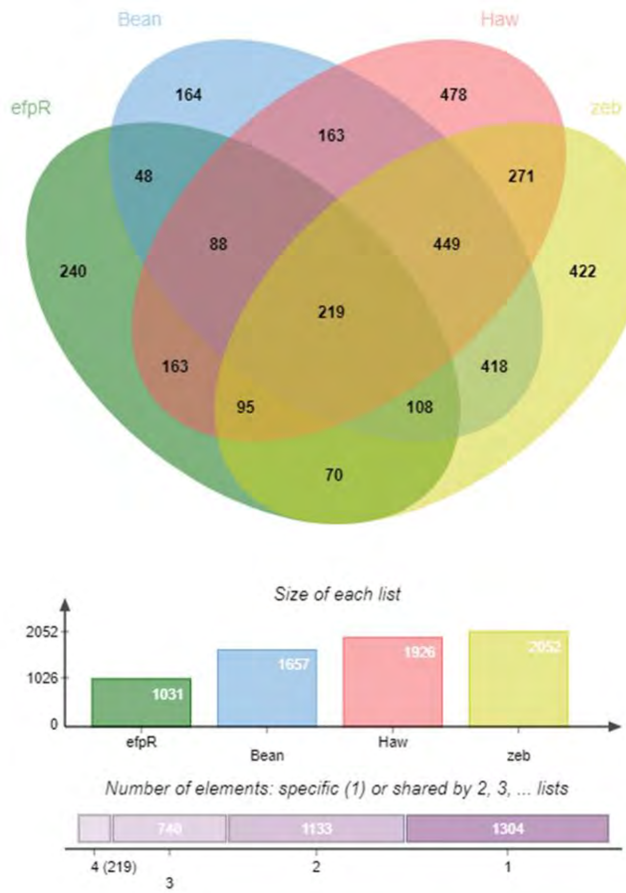


Figure 21 Comparison of DEGs from the evolved clones (Hawaii, Zebrina, and Bean) to the efpR regulon

The graph represents the number of genes in each list.

processes (such as cell division, detoxification, mobility chemotaxis, protection responses) (6%) and virulence (1%). In Bean evolved clones, majority of the DEGs (40%) belonged to hypothetical proteins with unknown functions, the second major DEGs encoded metabolic functions (38%) and 10% involved in transport of molecules (Figure 19). The distribution of host specific DEGs in Hawaii evolved clones were quite similar with a major portion of genes with unknown function and involved in metabolic functions (Annexure 2). Among these host specific genes, very few were virulence determinants.

The HrpB regulatory cascade represents one of the key regulators that are essential for the pathogenicity of *R. solanacearum* (Genin and Boucher, 2002a; Castillo and Greenberg, 2007; Genin and Denny, 2012; Castillo and Agathos, 2019). A comparison of the downregulated genes from HrpB regulon (190 genes) with the downregulated DEGs of evolved clones from all three hosts revealed a significant proportion (57 genes) of the overlap (Figure 20) (Table 11). The results were similar to that of Hawaii clones, which shared a significant proportion of the downregulated DEGs with the *hrpB* regulon (Gopalan-Nair et al., submitted). This significant overlap of the genes from the HrpB regulon despite the varying hosts indicate a prominent association with the adaptation process. Around 9% of the downregulated genes from the clones evolved on Hawaii and Zebrina belonged to the Hrp regulon while the proportion was slightly higher in Bean clones (14%).

A comparison of the extensive list of DEGs of the evolved clones with the 1031 genes of the EfpR regulon (Perrier *et al.*, 2016; Capela *et al.*, 2017) revealed that around 50% of the EfpR regulon was found in the DEGs of the evolved clones 492 (48%), 463 (45%) and 565 (55%) genes in the Zebrina, Bean and Hawaii DEGs, respectively (Figure 21). This result highlights again the importance of the downregulation of the *efpR* gene in *R. solanacearum* adaptation to the host plants. A total of 219 genes were common to all evolved clones (Figure 21). These 219 genes represent more than 20% of the EfpR regulon. The list of common genes between the EfpR regulon and the evolved clones is given in Annexure 3.

3.4 Discussion

The current study aimed to evaluate the molecular determinants of adaptation of the clones experimentally evolved on Eggplant ‘Zebrina’ (original host) and Bean ‘blanc précoce’ (distant host). The former work established that a majority of the clones evolved from GMI1000 had better fitness *in planta* (Guidot *et al.*, 2014). Here, we analyzed both the genomic and transcriptomic variations in the experimentally evolved clones. Two clones with no fitness gain were also included in our study as controls. Whole genome sequencing of the 14 analyzed clones disclosed between 0 and 1 genomic polymorphism in Zebrina clones and between 1 and 2 mutations in Bean clones. This numbers were in coherence with the numbers obtained for Hawaii clones (Hawaii tomato 7996) (Gopalan-Nair et al., submitted).

Overall, the level of fitness gain was significantly higher in the Bean and Hawaii clones compared to Zebrina clones (Table 5). In addition, mutation detected in the Zeb clones affected three genes, which did not affect their expression and the analysis of the transcriptomes revealed high variation in the number of DEGs between Zeb clones and no clear pattern when the number of DEGs were high. A hypothesis was that such results were obtained because Eggplant Zebrina is a susceptible host and the genomic and transcriptomic variations that occurred in the experimentally evolved clones could only be the result of genetic drive. The adaptive advantage of the mutation detected in the Zebrina clones still have to be evaluated to test this hypothesis. Interestingly however, two Zeb clones (b5 and e1) with the highest fitness gain had no mutation but a significant number of DEGs (151 and 932, respectively). The hypothesis was that epigenetic modifications such as DNA methylation occurred in these two clones and affected gene expression and improved the fitness of these clones.

In Bean clones, the mutations were detected in 5 different genes, including 3 transcription regulators. Mutation of genes encoding transcription regulatory hubs have been previously observed in the multiple laboratory evolution experiments of bacteria (Herring *et al.*, 2006; Conrad *et al.*, 2009; Marchetti *et al.*, 2010; Guidot *et al.*, 2014). Consequently, the mutation observed in the gene *rpoB* (RNA polymerase subunit B) on Bean b1 resulted in the change of amino acid from aspartic acid to tyrosine. While the effect of the mutation in the fitness gain is not yet known in this study, the gene is significantly upregulated in all the Bean clones except Bean c1. Various studies suggest that the gene *rpoB* is ubiquitous in the bacterial kingdom (Ogier *et al.*, 2019) and more interestingly, the gene might be involved in the adaptation of the bacteria to extreme environments. Studies conducted on *Mycobacterium tuberculosis* found that mutations in the *rpoB* gene elicited a stress response and thereby increased the bacterial adaptation conferring antibiotic resistance (Bergval *et al.*, 2007). Mutations in *rpoB* gene was also identified for its adaptive potential in the experimental evolution of *E. coli* in minimal medium (Herring *et al.*, 2006).

Interestingly, in Bean clones, a parallelism was observed in the de-regulation of five genes (*rpoB*, *purF*, *prhP* and *efpR*). These genes were either directly targeted by genetic mutation or indirectly through genetic or non-genetic modification affecting their expression. The *efpR* gene has been previously identified as an important gene for RSSC adaptation to the host plant (Guidot *et al.*, 2014; Perrier *et al.*, 2016; Capela *et al.*, 2017). Here, four other genes (*rpoB*, *prhP* and *purF*) were identified as potential new candidates important for RSSC adaptation to new host plants.

Despite the experimental evolution of GMI1000 on different hosts (original and distant), the evolved clones always seem to incline to the similar set of genes that are differentially expressed in the adapted clones. This was similar to the Hawaii evolved clones were convergence of transcriptomic profiles with overlapping DEGs were observed (Gopalan-Nair et al., submitted). The pattern was more consistent with the Bean clones than for the Zebrina. Comparative analyses of all the DEGs from the clones of each host with the regulon of the HrpB and EfpR regulators showed that a significant proportion of these two regulons (up to 50% for the EfpR regulon) was among the DEGs of the evolved clones. These results highlight the importance of de-regulation of the *hrpB* and *efpR* genes in RSSC adaptation to various host plants.

Chapter 4. Methylome analyses of the evolved clones

4.1 Brief introduction

The undeniable role of DNA methylation in the bacterial defense mechanisms and host adaptation is now well-studied in various pathogens (Oliveira and Fang, 2020; Oliveira *et al.*, 2020; Sánchez-Romero and Casadesús, 2020). Methylome analysis by SMRT sequencing is the first step toward understanding the biology and functions of DNA methylation in bacteria. Previous analysis using REBASE identified seven putative DNA MTases in the GMI1000 reference strain of RSSC, of which six were present on the chromosome and one on the megaplasmid (Table 1 Table 1in the general introduction). With the advent of SMRT sequencing technology, it was identified that RSc1982 methylates the adenine base in the G**T**WWA̲C motif (Erill *et al.*, 2017). The analysis revealed a total of 783 G**T**WWA̲C in the GMI1000 genome, of which 518 sites were present on the chromosome and the remaining 265 sites were present on the megaplasmid. The study highlights that 99% of the GTWWAC motifs are methylated in the GMI1000 genome.

In chapters 2 and 3, it was highlighted that most of the clones had better fitness in their respective host plants after the experimental evolution. Further analyses of the adapted clones revealed various genomic polymorphisms and several differentially expressed genes (DEGs) in comparison to the ancestral clone that could explain the fitness gain. Interestingly, we identified adapted clones with no genomic polymorphism but, a significant number of DEGs. The hypothesis was that the fitness gain could be explained by epigenetic modifications such as DNA methylation.

Therefore, the objective of this chapter was to test the hypothesis that the DNA methylation profile could vary during experimental evolution and could influence gene expression that aids the bacterial adaptation. Firstly, the G**T**WWA̲C methylome profile of the clones evolved from Hawaii tomato (10 clones), Zebrina (8 clones) and Bean (6 clones) were

	Hawaii tomato 7996 evolved clones										Mean
	a1	a4	b1	b4	c1	c2	d3	d5	e1	e3	
Mutation	$soxA1^{C639R}$		$RSp1574^{V95L}$	$RSp1574^{V95L}$ $prhP^{IS-6}$					$RSp1136^{C-218A}$	$RSp1136^{C-218A}$ $RSc3094^{R162R}$	0.7
Mean CI value	8.60	7.2	6.5	12.9	4.2	4.00	5.4	4.1	3.8	1.9	5.86
no. of DEGs	1227	187	478	503	902	272	125	269	245	212	442
No. of DMRs	21	18	21	19	18	12	18	17	15	18	18

Table 12 List of Hawaii tomato evolved clones for methylome analysis

The table includes the mean CI values, genetic mutation(s), their position, the DEGs of each evolved clone (Gopalan-Nair et al., submitted) and the number of differentially methylated regions (DMRs)

	Zebrina evolved clones								
	b1	**b5**	**c2**	**c3**	**c4**	**d1**	**e1**	**e3**	**Mean**
Mutation	RSp0083 $^{IS-9}$		ISRso1^{F91L}	ISRso1^{F91L}	dld^{R135S}				0.5
Mean CI value	2.7	3.7	2.1	1.6	2.1	**<u>0.9</u>**	3.6	**<u>1.4</u>**	2.3
No. of DEGs	88	151	6	169	6	19	932	1193	321
No. of DMRs	17	21	17	20	17	13	23	18	18

Table 13 List of Zebrina evolved clones for methylome analysis

The table includes the mean CI values, genetic mutation(s), their position, the DEGs of each evolved clone (unpublished data) and the number of differentially methylated regions (DMRs). The underlined CI represents CI significantly less than the ancestral clone (Wilcoxon test, *p*-value > 0.05)

	Bean evolved clones						
	a4	**a5**	**b1**	**b3**	**b4**	**c1**	**Mean**
Mutation	$RSc2508^{GA\text{->}G,392}$ $rpoB^{D428Y}$	$RSc2508^{GA\text{->}G,392}$	$RSc2508^{GA\text{->}G,392}$	$RSc2508^{GA\text{->}G,392}$	$RSc2508^{GA\text{->}G,392}$	$efpR^{P93Q}$ $purF^{C\text{->}T,-88}$	1.50
Mean CI value	6.1	6.5	4.6	5.3	5.5	6.6	5.77
No. of DEGs	879	1145	1146	406	387	379	724
No. of DMRs	26	35	25	25	26	20	26

Table 14 List of Bean evolved clones for methylome analysis

The table includes the mean CI values, genetic mutation(s), their position, the DEGs of each evolved clone (unpublished data) and the number of differentially methylated regions (DMRs) compared to the G**T**WWAC methylome profile of the ancestral clone. In the second part, the potential impact of the differentially methylated motifs on gene expression was investigated by a comparative analysis of the methylome and transcriptome variations of the evolved clones.

4.1 Material and methods

4.1.1 Bacterial strains and growth conditions

The list of the clones for which the methylome was investigated is given in Table 12, Table 13 and Table 14. The evolved clones were obtained from the experimental evolution of GMI1000 on tomato Hawaii 7996 (Gopalan-Nair et al., submitted), eggplant Zebrina and Bean 'blanc precoce' during 300 bacterial generations (Guidot et al., 2014). The bacterial growth conditions were the same as for the transcriptome analysis. Each of the evolved clones were grown in minimal media supplemented with glutamine (10 mM). Bacteria were grown until the beginning of stationary phase for DNA extraction. No biological replicate was performed.

4.1.2 Single-molecule, real-time (SMRT) sequencing

Genomic DNA was prepared from the bacterial cells grown in minimal media with glutamine collected at the beginning of stationary phase in order to limit the number of cells in division and avoid a bias towards hemimethylated marks. The bacterial samples were collected simultaneously for DNA as done for RNA (Gopalan-Nair et al., submitted). 20 ml of the bacterial culture was centrifuged at 5000g for 10 minutes followed by washing the pellets with water and centrifuged again. The pellets were stored at -80°C until DNA extraction.

The DNA were prepared based on the protocol described for high molecular weight genomic DNA (Mayjonade *et al.*, 2017). Briefly, the pellet was resuspended in 1800 µl of preheated lysis buffer (72 °C) followed by the addition of 6 µl of RNase A (100 mg/ml). The lysis buffer prepared on the day of experiment contains the following components given as final concentration and the respective quantity/volume for a final volume of 10 ml in bracket. 1 %

sodium metabisulfite (0.1 g), 0.5 M NaCl (0.29 g), 100 mM Tris HCl (1 ml of 1M Trsi HCl, pH 8.0), 50 mM EDTA (1 ml of 0.5 M EDTA, pH 8.0) and 1.25% SDS (625 µl of 20 % SDS) adjusted to a final volume of 10 ml using molecular biology grade water. The lysate was incubated at 55 °C for 30 minutes and mixed by inverting the tube every 10 minutes. These steps were carried out in the specific area allocated for quarantine bacteria (L2 in LIPM).

A volume of 600 µl (1/3 volume of lysis buffer) of cold (4 °C) 5M potassium acetate was added to the lysis buffer and mixed to obtain a homogenous solution by vigorous pipetting using P1000 pipette tip cut at 0.5 cm from the extremity (to avoid DNA shearing). The potassium acetate precipitates the protein and polysaccharides and forms a complex with SDS (sodium dodecyl sulfate) which will be removed by centrifugation. The mixture was centrifuged at 5000 g for 10 min at 4 °C.

The supernatant (550 – 600 µl) was then carefully transferred to a 2 ml Eppendorf tube followed by centrifuging again at 5000 g for 10 min at 4 °C. 500 µl of the clear supernatant was transferred to a new 2ml tube followed by equal volume (500 µl) of binding buffer. Binding buffer contains 2 g PEG8000 and 1.75 g of NaCl adjusted to a final volume of 10 ml molecular biology grade water, which was prepared on the day of experiment. PEG8000 was well dissolved until the solution becomes clear as it can affect the yield. 30 µl of Serapure beads (1:18 (V:V) was added and mixed well by inverting the tubes 20 times. Serapure beads solutions contains 2 % Sera-Mag SeedBead magnetic carboxylate modified particles (washed 4 times with H2O to remove sodium azide), 18 % PEG8000, 1M NaCl, 10mM Tris HCl (pH 8.0), 1 M EDTA (pH 8.0). The suspension was incubated for 30 min at room temperature with gentle agitation on a rocker platform. Followed by incubation the tube was spin down for 1 sec and placed on magnetic rack for 3 min (or until the solution becomes clear). The supernatant was removed without disturbing the magnetic beads pellet.

Subsequently the beads pellet were washed quadruple times in total with wash solution (70 % EtOH). For this, the tube was removed from the magnetic rack, mixed with 1 ml of wash solution by inverting the tube 20 times. The tube was spin down for 1 sec and then placed on the magnetic rack for 3 min (or until the solution becomes clear) to remove the supernatant without disturbing the beads pellet. This washing step was repeated three additional times to reach a total of quadruple wash. The tube was spin down for 1 sec, placed on the magnetic rack to remove any remaining wash solution using P10 pipette tips, and the beads were air dried for 1 min with the lid open. It is crucial not to let the beads dry longer than a minute, as it will significantly reduce the elution efficiency. The genomic DNA (gDNA) was eluted using 80 µl of elution buffer (Qiagen EB elution buffer) preheated to 55 °C. The beads were resuspended by flicking the tube to ensure that the beads are not aggregated. The tube was spin down for 1 sec and placed on the magnetic rack for 10 min or until the solution becomes clear. Higher DNA concentrations can increase the elution duration and it is therefore recommended to let the tube on magnetic rack for 2 hours (or overnight) and/or addition of more elution buffer. Finally, the clear gDNA solution was transferred to a new tube followed by quantity and quality check.

The concentration of the gDNA was measured using spectrophotometer (Nanodrop 2000, Thermofischer, USA). A260/280 ratio must be between 1.8 and 2 whereas the A260/230 ratio must be between 2 and 2.2. A higher ratio in either could be an indication of RNA contamination. Additional quantification was performed using Qubit 4 Fluorometer (Thermofischer, USA). The samples were then sent for the high throughput sequencing at GeT-Plage, Toulouse.

4.1.3 Analysis of methylation profile of GTWWAC

All methylation analyses were performed with public GMI1000 genome and annotation. Motif and methylation detection were performed using the pipeline "pbsmrtpipe.pipelines.ds_modification_motif_analysis" from PacBio SMRTLink 6.0. The default settings were used except: compute methyl fraction set as true, minimum required alignment concordance >= 80 and minimum required alignment length >= 1000 (all these analyses were performed by bioinformatician, *Ludovic Legrand*).

Followed by the bioinformatics analyses of the data obtained from SMRT sequencing, methylome profiles of the ten evolved clones were compared to the ancestral clone individually. The analysis showed the methylation profile for G**T**WWAC motif with a score, coverage, IPD ratio, and fraction for each sample (Annexure 10 - Annexure 33). A score above 30 is considered significant and coverage represents the sequencing depth (higher the better). IPD ratio or interpulse duration ratio is the time required for the consequent nucleotide to bind, where the presence of methylated base increases the time required for the nucleotide addition (higher IPD ratio means a higher probability of methylation). The fraction represents the percentage of methylated bases in the genome pool at that particular position. In this experiment, the methylation or hemimethylation of a particular position is considered significant when the fraction is greater than or equal to 0.50 (represents at least 50% of the sequences are methylated at that particular position in the whole genome pool) in addition to the score above 30.

4.2 Results

4.2.1. SMRT sequencing technology for the analysis of GTWWAC methylation

Here, we used SMRT sequencing technology to compare the methylation profiles of the ancestral GMI1000 clone and the experimentally evolved clones generated by SPE on Hawaii 7996, Zebrina and Bean during 300 generations. Genomic DNA for methylome analysis was extracted from *in vitro* bacterial cultures. The bacterial samples were collected at the beginning of stationary phase, assuming the fact that the methylome pattern remains stable during the growth curve as seen in *E. coli* and *P. luminescens* strains (Payelleville *et al.*, 2017). Genomic DNA extraction of the ancestral GMI1000 clone and the evolved clones using the protocol described for high molecular weight genomic DNA (Mayjonade *et al.*, 2017) yielded sufficient concentrations of DNA suitable for SMRT sequencing. The concentration of DNA ranged between 75 ng/µl and 500 ng/µl. The concentration of DNA and their respective purity ratio for each of the analyzed sample is given in Annexure 4 and Annexure 5.

The bioinformatic pipeline used in the present work found that 778 out of the 783 (99%) G**T**WWAC motifs were methylated (m6A modification) in the ancestral clone GMI1000. The genes targeted by a G**T**WWAC motif non-methylated in the GMI1000 ancestral clone are described in Table 15. Three of these motifs were present on the gene body, which includes RSc2612, RSc3393 and RSp1329. The remaining two motifs were detected on the promotor regions of RSc0102-03 and RSp1674-75. Three of these genes corresponds to transposase proteins (RSc0103, RSc3393 and RSp1675) and the two other genes encodes hypothetical transmembrane protein based on STRING database (RSc2612 and RSp1329). The motif on RSp1329 could also be probable promotor of RSp1330. A schematic representation of these five positions is given in figures 26, 27, 28, 29 and 30.

4.2.2. Detection of differential methylation marks between the ancestral clone and the evolved clones

The putative differential methylation marks of the G**T**WWA̲C motif between the ancestral clone and the various evolved clones were divided into two major categories as follows:

(i) Full methylation (methylation on both DNA strands in the evolved clones whereas both strands are not methylated in the ancestral clone) (Table 15)

(ii) Hemimethylation

Type 1: methylation on one DNA strand in the ancestral clone, but nonmethylated in the evolved clone (Table 16)

Type 2: methylation on one DNA strand in the evolved clones, but nonmethylated in the ancestral clone (Table 17)

The analysis revealed between 12 and 35 differential methylation regions (DMRs) in the evolved clones in comparison to the ancestral clone (Tables 12, 13 and 14). The raw data of the differential methylation for each of the analyzed evolved clones is given in Annexure 10 - Annexure 33. The total number of differentially methylated motifs in each evolved clone including both fully methylated and hemimethylated sites are depicted as barplot in Figure 22, Figure 23 and Figure 24. There was a significant difference in the mean number of DMRs observed in the clones evolved on Bean and Zebrina (student t-test, *p*-value = 0.008); and between Bean and Hawaii evolved clones (student t-test, *p*-value = 0.006). However, there was no statistical

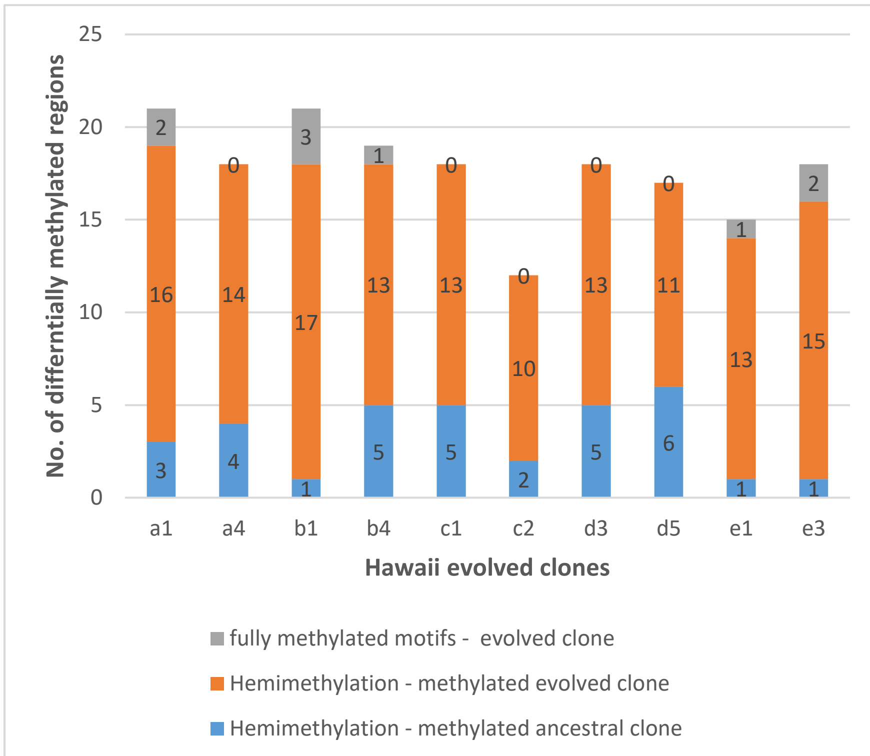


Figure 22 Barplot of the number of differentially methylated motifs between the ancestral and the Hawaii evolved clones

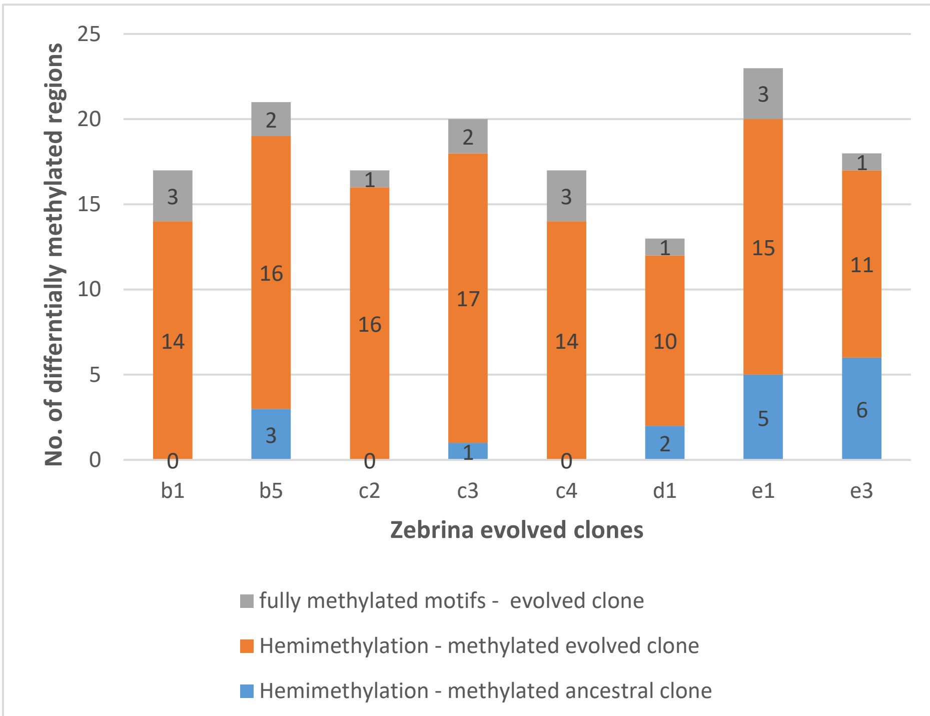


Figure 23 Barplot of the number of differentially methylated motifs between the ancestral and the Zebrina evolved clones

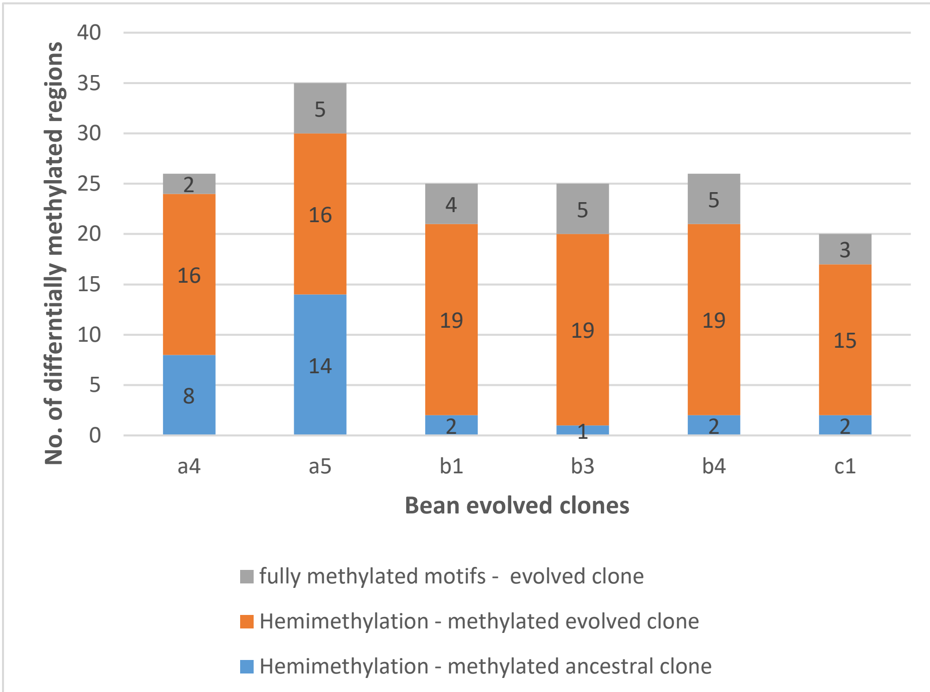


Figure 24 Barplot of the number of differentially methylated motifs between the ancestral and the Bean evolved clones

Gene	Position	Feature	GMI1000	Hawaii evolved clones										Zebrina evolved clones								Bean evolved clones					
				a1	a4	b1	b4	c1	c2	d3	d5	e1	e3	b1	b5	c2	c3	c4	d1	e1	e3	a4	a5	b1	b3	b4	c1
RSc0102	117936	upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)	6A	**6mA**		**6mA**	**6mA**							**6mA**	**6mA**								**6mA**	**6mA**	**6mA**	**6mA**	**6mA**
	117939		6A	**6mA**	6mA	**6mA**	**6mA**			6mA		6mA	6mA	**6mA**	**6mA**	6mA	6mA	6mA	6mA	6mA	6mA	6mA	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**
RSc2612	2813720	RSc2612 (hypothetical protein)	6A		6mA		6mA	6mA	6mA		6mA	6mA		6mA	6mA	6mA	**6mA**	**6mA**	6mA		6mA	6mA	**6mA**	6mA	**6mA**	**6mA**	
	2813723		6A							6mA							**6mA**	**6mA**					**6mA**		**6mA**	**6mA**	6mA
RSc3393	3660528	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	6A		6mA	**6mA**	6mA			6mA	6mA	6mA	**6mA**	6mA		6mA		**6mA**	6mA	**6mA**	6mA	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**
	3660531		6A			**6mA**		6mA					**6mA**				6mA	**6mA**		**6mA**		**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**
RSp1329	1680220	RSp1329 (hypothetical protein)	6A	6mA		**6mA**						**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**
	1680223		6A	6mA		**6mA**						**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**
RSp1675	2087332	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	6A											**6mA**						**6mA**			**6mA**	**6mA**	**6mA**	**6mA**	**6mA**
	2087335		6A	6mA	6mA	6mA	6mA		6mA		6mA	6mA	6mA	**6mA**	6mA	6mA	6mA	6mA	6mA	**6mA**	6mA	6mA	**6mA**	**6mA**	**6mA**	**6mA**	**6mA**
Total number of putative differential methylation positions per evolved clone				2	0	3	1	0	0	0	0	1	2	3	2	1	2	3	1	3	1	2	5	4	5	5	3

Table 15 List of differentially (fully) methylated positions in all the evolved clones

The table includes the gene name, description and the position on the genome. Transposase protein is given in blue font. 6A – nonmethylated base; 6mA – methylated base. Underlined base – fully methylated motif. Empty box refers to the same type of base modification as GMI1000. Orange fill – Up regulated gene; Green fill – Down regulated gene.

Gene	Position	Feature	GMI1000	Hawaii evolved clones										Zebrina evolved clones								Bean evolved clones					
				a1	a4	b1	b4	c1	c2	d3	d5	e1	e3	b1	b5	c2	c3	c4	d1	e1	e3	a4	a5	b1	b3	b4	c1
RS01108	2683489	ACUR	6mA																				**6A**				
RSc0102	117950	upstream putative calcium binding hemolysin protein	6mA																			**6A**					
RSc0109	127850	upstream thiazole synthase thiG	6mA				6A		6A												6A						
RSc0279	313558	upstream acyl-CoA dehydrogenase oxidoreductase protein	6mA																		**6A**						
RSc0608	655714	upstream Type III effector protein RipAA	6mA					6A			6A	6A			6A						6A	6A		6A		6A	
RSc0636	683379	upstream ISRSO5-transposase protein	6mA		**6A**			**6A**	**6A**	**6A**			**6A**														
RSc0831	873770	translocase subunit SecA-related protein SecA2	6mA																	**6A**							
RSc1079	1134729	upstream glucarate dehydratase protein gudD1	6mA																				**6A**				
RSc1116	1173706	upstream putative signal peptide protein	6mA																			**6A**					
RSc1539	1645829	upstream hydrolase protein	6mA																				**6A**				
RSc1705	1821399	Hypothetical protein	6mA																			**6A**					
RSc1705	1821402	Hypothetical protein	6mA	6A	6A																		6A				
RSc2132	2314079	Type III effector protein RipJ	6mA								6A											6A					
RSc2176	2360132	upstream ISRSO5-transposase protein	6mA																6A	6A			6A				
RSc2326	2527133	upstream putative lipoprotein	6mA																				**6A**				
RSc2491	2697793	upstream putative signal peptide protein	6mA																		6A		6A				6A
RSc2654	2856885	upstream serine protease protein	6mA	6A	6A		6A	6A		6A	6A						6A		6A	6A	6A	6A	6A	6A	6A	6A	
RSc2795	3004675	upstream hypothetical protein	6mA				**6A**																				
RSc2825	3039075	type IV pilus assembly protein (pilB)	6mA																				**6A**				
RSc2918	3145217	upstream putative extracytoplasmic function sigma factor (IRON-regulated) transcription regulator protein	6mA																	**6A**							
RSc3251	3506641	Hypothetical protein	6mA																				**6A**				
RSc3393	3660517	upstream ISRSO5-transposase protein	6mA							**6A**																	
RSp0078	87279	upstream hypothetical protein	6mA		**6A**		**6A**	**6A**		**6A**	**6A**																
RSp0216	269728	serine/threonine-protein kinase ; upstream ISRSO5-transposase protein (RSp0127)	6mA																		**6A**						
RSp0339	505523	Hypothetical protein	6mA	**6A**																							
RSp0449	561449	putative RSH-related protein	6mA																				**6A**				
RSp0547	687780	upstream ISRSO8-transposase orfA protein	6mA																			**6A**					
RSp0572	712820	Type III effector protein RipAX2	6mA												6A								6A				
RSp0611	743015	Hypothetical protein	6mA								6A				6A												
RSp0641	792207	peptide synthetase protein	6mA					6A												6A							
RSp0725	912864	Predicted Fe-S oxidoreductase ; upstream RSp0726 (MFS family transporter)	6mA																			**6A**					
RSp0726	913370	MFS family transporter	6mA																				**6A**				
RSp0755	953740	Hypothetical protein	6mA																				**6A**				
RSp1132	1427525	Hypothetical protein	6mA							**6A**																	
RSp1643	2062924	upstream hypothetical protein	6mA			6A	6A				6A																6A
		Total number of putative hemimethylated motifs		3	4	1	5	5	2	5	6	1	1	0	3	0	1	0	2	5	6	8	14	2	1	2	2

Table 16 List of differentially type 1-hemimethylated positions (methylated ancestral clone)

The table includes the gene and the position on the genome. 6A – nonmethylated base; 6mA – methylated base. Empty box refers to the same type of base modification as GMI1000. Orange fill – Up regulated gene; Green fill – Down regulated gene. The red font represents motifs that are present in both the list of differential hemimethylated positions. Underlined 6A represent host specific DMRs.

				Hawaii evolved clones										Zebrina evolved clones								Bean evolved clones					
Gene	**position**	**Feature**	**GMI1000**	**a1**	**a4**	**b1**	**b4**	**c1**	**c2**	**d3**	**d5**	**e1**	**e3**	**b1**	**b5**	**c2**	**c3**	**c4**	**d1**	**e1**	**e3**	**a4**	**a5**	**b1**	**b3**	**b4**	**c1**
RSc0081	94120	upstream hypothetical protein	6A					**6mA**																			
RSc0109	127847	upstream thiazole synthase thiG	6A	6mA			6mA			6mA		6mA			6mA	6mA						6mA	6mA	6mA	6mA	6mA	6mA
RSc0608	655717	upstream Type III effector protein RipAA	6A	6mA	6mA	6mA	6mA	6mA	6mA	6mA		6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA
RSc0636	683376	upstream ISRSO5-transposase protein	6A	6mA		6mA	6mA							6mA						6mA		6mA	6mA	6mA		6mA	6mA
RSc0636	683390	upstream ISRSO5-transposase protein	6A		6mA	6mA	6mA	6mA		6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA
RSc0957	1004579	Hypothetical protein	6A																**6mA**								
RSc2094	2267250	upstream transcription regulator protein; upstream putative xanthine dehydrogenase oxidoreductase protein XdhA (RSc2095)	6A	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA		6mA	6mA	6mA	6mA	6mA	6mA	6mA
RSc2176	2360129	upstream ISRSO5-transposase protein	6A			6mA					6mA		6mA		6mA	6mA	6mA			6mA		6mA		6mA	6mA	6mA	6mA
RSc2176	2360143	upstream ISRSO5-transposase protein	6A	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA		6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA
RSc3178	3435906	putative ATP binding protein	6A		6mA						6mA					6mA	6mA										
RSc3249	3505345	putative signal peptide protein	6A	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA		6mA	6mA	6mA	6mA
RSp0116	133340	upstream putative lipoprotein	6A	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA		6mA	6mA	6mA	6mA
RSp0216	269725	serine/threonine-protein kinase ; upstream ISRSO5-transposase protein (RSp0127)	6A	6mA	6mA						6mA			6mA	6mA	6mA	6mA	6mA		6mA		6mA	6mA		6mA	6mA	6mA
RSp0294	382743	putative hemolysin-type calcium-binding protein	6A	6mA	6mA	6mA	6mA	6mA		6mA				6mA	6mA	6mA	6mA	6mA	6mA		6mA		6mA	6mA	6mA		
RSp0338	445738	upstream negative regulator of exopolysaccharide production transcription regulator protein	6A																								**6mA**
RSp0454	565511	rhs-related protein	6A			6mA	6mA	6mA		6mA	6mA						6mA					6mA		6mA	6mA	6mA	6mA
RSp1026	1298049	upstream signal peptide protein	6A	6mA		6mA							6mA		6mA		6mA	6mA	6mA			6mA	6mA	6mA		6mA	
RSp1152	1452559	upstream ISRSO5-transposase protein	6A			6mA			6mA	6mA		6mA	6mA	6mA	6mA			6mA		6mA		6mA	6mA	6mA	6mA	6mA	6mA
RSp1343	1695111	upstream hypothetical protein	6A	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA		6mA	6mA	6mA			6mA	6mA	6mA	6mA	6mA	6mA	6mA	
RSp1404	1766900	chemotaxis protein	6A	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA	6mA
RSp1529	1916009	upstream 1-aminocyclopropane-1-carboxylate oxidase Efe	6A			6mA		6mA					6mA							6mA		6mA	6mA	6mA	6mA	6mA	
RSp1544	1939052	Hypothetical protein	6A	6mA									6mA										6mA		6mA		
RSp1545	1939296	putative hemagglutinin-related protein	6A	6mA	6mA	6mA				6mA		6mA	6mA	6mA	6mA	6mA	6mA	6mA		6mA			6mA	6mA	6mA	6mA	
RSp1643	2062927	upstream hypothetical protein	6A	6mA	6mA				6mA			6mA		6mA		6mA	6mA	6mA	6mA	6mA	6mA			6mA	6mA	6mA	
RSp1675	2087346	upstream ISRSO5-transposase protein	6A	6mA	6mA	6mA	6mA	6mA	6mA			6mA	6mA	6mA	6mA	6mA	6mA	6mA		6mA	6mA		6mA	6mA	6mA	6mA	6mA
		Total number of putative hemimethylated motifs		16	14	17	13	13	10	13	11	13	15	14	16	16	17	14	10	15	11	16	16	19	19	19	15

Table 17 List of differentially type 2-hemimethylated positions (methylated evolved clone)

The table includes the gene and the position on the genome. 6A – nonmethylated base; 6mA – methylated base. Empty box refers to the same type of base modification as GMI1000. Orange fill – Up regulated gene; Green fill – Down regulated gene. The red font represents motifs that are present in both the list of differential hemimethylated positions. Underlined 6mA represent host specific DMRs.

difference in the number of DMRs between the clones evolved on Hawaii and Zebrina (student t-test, *p*-value = 0.695) (Table 18).

The highest number of DMRs was observed in clone Bean a5 with 35 G**T**WWA̲C sites differentially methylated compared to the ancestral clone; of which 5 sites were fully methylated (methylated on both +/- strands) and 30 sites were hemimethylated of which 16 were methylated on the evolved clone and the remaining 14 sites were methylated on the ancestral clone. The lowest number of DMRs was observed in clone Haw c2 with 13 G**T**WWA̲C sites differentially methylated compared to the ancestral clone; including zero sites fully methylated on the evolved clone and 13 sites hemimethylated of which 11 sites in haw c2 and 2 remaining sites were methylated in GMI1000.

Majority of the differentially methylated motifs (60% in Hawaii and Bean; 63% in Zeb) were detected in the upstream region of the genes. This observation is in agreement with the previous study on GMI1000 that found that 38% of the total G**T**WWA̲C motif were annotated to the upstream region of genes (Erill *et al.*, 2017). Overall, one-fifth of the DMRs observed in Hawaii and Bean clones, and one-fourth of the DMRs from Zeb clones belonged to transposase protein. This could be interesting because methylation in the transposases are known to affect the gene expression of itself or the neighboring genes (Roberts *et al.*, 1985; Nagy and Chandler, 2004).

4.2.3 Fully methylated differential motifs were observed only on the evolved clones

Fully methylated differential motifs were observed in 19 out of the 24 evolved clones analyzed (at least 1 DMR and max 5 DMRs (Table 15). A schematic representation of these five differential methylation marks is given in figures 26, 27, 28, 29 and 30. In Hawaii clones, the observed number of DMRs (fully methylated) were between 0 and 3; between 1 and 3 in Zebrina clones and in Bean clones there were 2 to 5 DMRs (Table 15).

(a)

All DMRs		Zebrina	Hawaii
		p-value Student t-test	
	Mean±SE	18±1.102	18±0.801
Bean	26±1.817	0.008306	0.00605
Hawaii	18±0.801	<u>0.6945</u>	

(b)

Full methylation		Zebrina	Hawaii
		p-value Student t-test	
	Mean±SE	4±0.306	4±0.330
Bean	4±0.471	0.009946	0.00065
Hawaii	1±0.330	0.03514	

Table 18 Statistical analysis of differentially fully methylated positions between clones evolved on various host plants

Student t-test was used to compare the clones and the table includes mean and SE. p-value less than 0.05 is considered significant and non-significant *p*-values are underlined. (a) All DMRs and (b) Differential full methylation

Of the five motifs that were found to be differentially methylated in comparison to ancestral clone, the motif on the gene RSp1329, which encodes hypothetical protein was observed in 18 out of the 24 evolved clones analyzed (Table 15). The motif on the gene RSc2612 was differentially fully methylated in 5 clones and differentially hemimethylated in 15 clones. The G**T**WWA̲C motif upstream RSp1675 possessed only one clone that was similar to GMI1000 (nonmethylated), while the motif on other clones were either fully methylated (7 clones) or hemimethylated (15 clones). The motif upstream RSc0102 and RSc3393 had 10 clones with differential fully methylation and 11 clones with hemimethylated differential motifs.

Overall, the data reveals that there was a statistically significant difference between the numbers of fully methylated differential regions observed on the clones evolved on Hawaii, Zebrina and Bean (student t-test, $p < 0.05$) (Table 18).

4.2.4 Hemimethylation

Differential hemimethylation concerns only one strand of the DNA sequence methylated either in the ancestral clone and not in the evolved clone or vice versa. Hemimethylated state of the DNA can mostly be transient owing to its regulatory role during cell division and transcription (Marinus and Casadesus, no date; Casadesus and Low, 2006; Low and Casadesús, 2008; Casadesús, 2016; Oliveira and Fang, 2020; Sánchez-Romero and Casadesús, 2020). However, hemimethylation can also be inherited (stable) as seen in some bacteria such as *E. coli* (refer General introduction, section 1.2.3) (Wion and Casadesús, 2006) and thereby making it an interesting objective to explore the probable importance of hemimethylation in host adaptation of the RSSC strain GMI1000.

The type 1 hemimethylation where only one of the strand methylated in the ancestral clone is nonmethylated in the evolved clones was observed in 21 of the 24 evolved clones (Table 16). The highest number of hemimethylated motif was observed on Bean a5 with 14 DMRs (hemimethylation). The type 2 hemimethylation where one of the strand is methylated in the evolved clone in comparison to the ancestral clone was observed for all the 24 evolved clones.

The highest number of DMRs was observed in three clones from the same population evolved on Bean (b1, b3 and b4) with 19 hemimethylated motifs. The least number of DMRs was observed in Haw c2 and Zeb d1 with 10 hemimethylated motifs.

(i) Hawaii evolved clones

In Hawaii clones, a total of 13 positions were found methylated in GMI1000 (on one strand) and nonmethylated in the evolved clones (on one strand) (Table 16). The highest number of DMRs (hemimethylation) was observed in Haw d5 and the least number was in Haw b1, e1 and e3 with 1 DMR. Of the 13 motifs, the motif on RSc2654 (serine protease protein) was differentially methylated in six out of the ten clones (a1, a4, b4, c1, d3 and d5). In parallel, a total of 25 positions were found nonmethylated in GMI1000 and methylated in the evolved clones (on one strand) (Table 17). The highest number of DMRs were detected in Haw b1 with 17 DMRs, while the lowest number was on c2 with 10 DMRs. Six of the motifs were differentially hemimethylated in all of the ten Hawaii clones. The motifs were present upstream RSc2094-95 (transcription regulator protein; *xdhA*, putative xanthine dehydrogenase (subunit A) oxidoreductase protein), upstream RSc2176 (transposase protein), on RSc3249 (putative signal peptide protein), upstream RSp0116 (putative lipoprotein), upstream RSp1343 (hypothetical protein) and RSp1404 (chemotaxis protein). The motif upstream RSc0608 (RipAA) was differentially hemimethylated in nine out of the ten clones.

(ii) Zebrina evolved clones

A total of 12 positions were found methylated in GMI1000 (on one strand) and nonmethylated in the evolved clones (on one strand) (Table 16). The highest number of this category of differential methylation was in Zeb e3 with six DMRs and the least number was observed in Zeb b1, c2 and c4 with zero DMRs. In parallel, a total of 22 positions were found nonmethylated in GMI1000 and methylated in the Zebrina evolved clones (on one strand) (Table 16). Five differentially hemimethylated motifs were present in all the eight Zebrina clones analyzed and they are as follows: RSc0608 (*ripAA*), upstream RSc0636 (transposase protein), RSc3249 (putative signal peptide protein), upstream RSp0116 (putative lipoprotein) and RSp1404 (chemotaxis protein).

(iii) Bean evolved clones

A total of 21 positions were found methylated in GMI1000 (on either strand) and nonmethylated in the evolved clones (on either strand) (Table 16). The highest number of differentially hemimethylated motifs (methylated ancestral clone) out of all the clones analyzed from various hosts was observed in Bean a5 clone with 14 DMRs (Table 16). The motif on RSc2654 (serine protease protein) was observed in all the Bean clones except c1. In parallel, a total of 22 positions were found nonmethylated in GMI1000 and methylated in the Bean evolved clones (on one strand) (Table 16). Seven differentially hemimethylated motifs were present in all the six analyzed bean clones. The motifs were present upstream RSc0109-10 (transposase protein; thiazole synthase), upstream RSc0608 (*ripAA*), upstream RSc0637 (transposase protein), upstream RSc2094-95 (transcription regulator protein; *xdhA*, putative xanthine dehydrogenase (subunit A) oxidoreductase protein), upstream RSc2176 (transposase protein), upstream RSp1152 (transposase protein) and RSp1404 (chemotaxis protein).

(iv) Host specific DMRs

Overall, clone evolved on Bean had the highest number of host specific type 1 hemimethylated motifs at 14 positions, while Zebrina and Hawaii clones had 4 and 6 hemimethylated positions each (Table 16 and Table 17). Of the six positions in Hawaii clones, 2 differential motifs were shared among 5 Hawaii clones. On the other hand, there was only one host specific differential type 2 hemimethylation on clones evolved from each host and most of the differential positions were shared amongst the evolved clones regardless of their experimental host. The motif on RSp1404 (chemotaxis protein) related to virulence was hemimethylated on all the 24 analyzed clones; while 3 positions were differentially hemimethylated in 23 out of the 24 clones. The positions were upstream RSp0608 (RipAA), upstream RSc2094-95 (transcription regulator protein; XdhA, putative xanthine dehydrogenase (subunit A) oxidoreductase protein) and upstream RSp0116 (putative lipoprotein). RipAA secreted by the T3SS has a known functional role in virulence and makes it one of the interesting candidate DMR for further studies.

Figure 25 Circos plot of the distribution of differential methylation among the experimentally evolved clones in comparison to the ancestral clone

The circus plot contains both the chromosome and the megaplasmid in the same figure. The colored circles represent the clones evolved from each host. It includes the differentially hypomethylated regions in blue dot and the differentially hypermethylated regions in red dot. The genomic polymorphisms are highlighted in black line. Hypermethylation – presence of methylation on the evolved clone. Hypomethylation – absence of methylation on the evolved clone

4.2.5 DEGs vs DMRs

The aim of the second part of the methylome analysis was to compare the differentially expressed genes (DEGs) of the evolved clones with the differentially methylation regions (DMRs) of the evolved clones. The DEGs were obtained from the transcriptome analysis of the 24 evolved clones and ancestral clone GMI1000 when grown under the same conditions as for the methylome analyses (Gopalan-Nair et al., submitted). In fact, the samples for methylome and transcriptome analyses were collected at the same time from same culture for a fair comparison. The objective of the comparison was to find a probable correlation between the differential methylation marks and the differential expression of the corresponding genes. The presence of methylation on the mobile elements such as in the transposase proteins and in the promotor/upstream regions of the gene can affect the expression of the genes (Nagy and Chandler, 2004; Casadesus and Low, 2006; Casadesús, 2016). The gene expression is integrated in tables 15, 16 and 17, where the up-regulated genes are highlighted in orange whereas the down-regulated genes are highlighted in green. Presence of the same motif on both the list of hemimethylation (sense and anti-sense strand) is highlighted in red font. The motif on or upstream of transposase proteins are highlighted in blue font in the fully methylated differential motifs. For this analysis, the DEGs were those having a FDR *p*-value < 0.05 regardless of the logFC value to include maximum genes differentially expressed (Figure 25,**Error! Reference source not found.**,Figure 26).

Of the five (fully) differentially methylated regions observed in the evolved clones from all three hosts, two of the corresponding genes had a differential expression in at least one of the evolved clone. In total, there were four DMRs that correlated with the DEGs. The putative differential methylation mark on the promotor region of both RSc0102 (putative calcium binding hemolysin protein) and RSc0103 (ISRSo5 – transposase protein) might have influenced the gene expression of RSc0102, which was upregulated in Haw a1, Haw b4 and Bean c1. On contrary, the clone Haw e3 was hemimethylated at the same motif and the gene was downregulated. The transcriptome analysis revealed that the gene RSp1675 was down-regulated Bean b1, which also

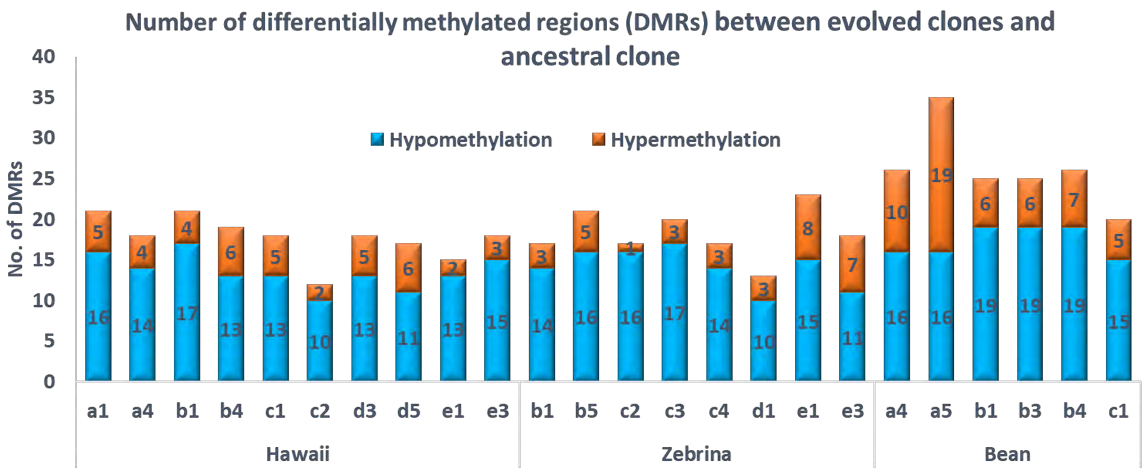


Figure 26 Number of differentially methylated regions (DMRs) between evolved clones and the ancestral clones.

The number of hypomethylated regions are given in blue bar the number of hypermethylated regions in orange.

coincidently has differential methylation mark (fully methylated) in comparison to the ancestral clone. The putative calcium binding hemolysin protein is a structural element and an interesting candidate to explore further. However, these analyses are all putative, until the differential methylation positions are experimentally validated. There were no differential expression in any of the Zebrina clones in correspondence to the DMRs. Association of differential hemimethylation and differential expression of genes from all the evolved clones yielded a 15% presumable correlation in general. Specifically, the association was higher in Hawaii evolved clones (48%) in comparison to the Zebrina (23%) and Bean (28%) evolved clones. At two differential hemimethylated positions, in seven and nine out of the ten Hawaii clones differential expression of the genes RipAA and RSc2176 (transposase protein) were detected, respectively (Table 17). The same observation was not made on Zebrina and Bean clones. However, as a complementary approach to the RNAseq analyses, RT-qPCR (Quantitative reverse transcription PCR) will be performed in order to confirm the presence or absence of the differential expression of the genes. For instance, a potential candidate such as RSc2094 will be subjected to RT-qPCR because the gene is upregulated in Hawaii clones and interestingly in Bean clones, the gene is downregulated.

4.3 Discussion

Changes in DNA methylation can impact gene expression levels by changing the binding affinities of regulatory proteins. On the contrary, DNA binding proteins can block the MTase binding to the specific DNA sequences resulting in non-methylation and alteration of the gene expression levels (Casadesus and Low, 2006). Our study perceived that more than half of the differentially methylated motifs were on the promotor/upstream region of the gene (approximately 60%). The infamous *pap* operon and *ag43* of *E. coli* are some of the examples that highlights the impact of methylation patterns on the respective gene expression (Blyn *et al.*, 1990; Haagmans and Van Der Woude, 2000). Another example is the methylation of the *traJ* gene of the Salmonella virulence plasmid resulting in transcriptional activation. The expression of *traJ* gene, a transcriptional factor, contains one GATC site upstream in the promotor region and is

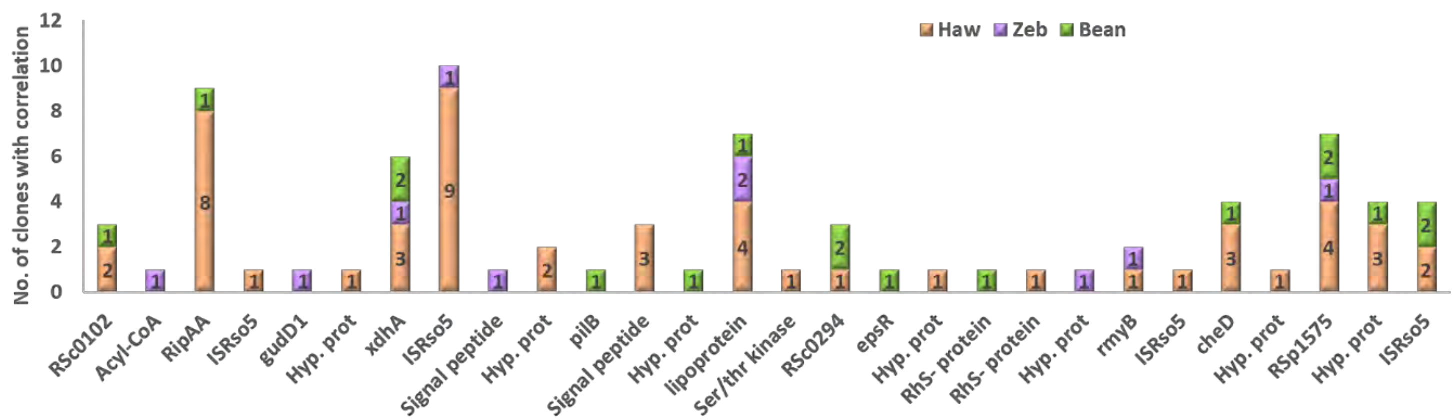


Figure 27 Correlation between the DEGs and DMRs

The y-axis represents the number of clones with the correlation between the DEGs and DMRs. The x-axis represents the various genes and upstream of the genes for which there was a correlation. The number of clones from Hawaii is given in orange, the number of clones from Zebrina in purple and the number of clones from Bean in green.

controlled by multiple cell factors including Lrp. Methylation of the GATC blocks the Lrp binding and prevents *traJ* activation. With the replication initiation, as the GATC site is hemimethylated the Lrp binding activates the *traJ* transcription in one of the daughter plasmid (Camacho and Casadesús, 2005; Casadesús, 2016). It is important to note that these methylation marks were all on the promotor region of the respective genes and that the regulation happens by differential recognition of fully methylated or hemimethylated DNA (Low and Casadesús, 2008).

Conversely, other studies did not find a similar correspondence with differential gene expression and the methylation pattern. The study performed on *Shewanella* sp. concluded that there were no large and direct role of DNA methylation pattern in regulating the gene expression (Bendall *et al.*, 2013). Another interesting study on *P. luminescens* with overexpressing *dam* observed no change in the virulence gene(s) expression, despite a significantly weakened virulence (Payelleville *et al.*, 2017). Our study found no differential methylation motifs on any of the genes with genomic polymorphisms detected on the evolved clones. However, we found some correlation between the differentially expressed genes and the putative differentially methylated regions. High proportion of genes (50%) with differentially methylated regions were also differentially expressed in clones evolved on Hawaii tomato compared to the clones evolved on other hosts (Zeb and Bean). Fully methylated motifs could be experimentally verified by methylation sensitive restriction enzyme (MSRE) and PCR (Krygier *et al.*, 2016). However, because the presence of most hemimethylated DNA is transient the presence of hemimethylation is difficult to experimentally validate (Casadesús, 2016). Nevertheless, the differential expression of the gene with differential hemimethylated motifs could be verified by RT-qPCR.

DNA methylation has been implicated in the regulation of transposase expression and the transposition activity of several TE (transposable elements). Fascinatingly, usually the insertion sequence elements (such as IS*10* and IS*50*) possesses DAM methylation site close to or overlaps the transposase promotors and it has been validated that when fully methylated, the promotor is inefficient (Nagy and Chandler, 2004). As the motif will be transiently hemimethylated subsequently after passage of replication fork, this setup favors the transposon to pass the replication fork before its transposition. Transcriptional activity of IS*10* and IS*50* transposase gene has been shown to operate as mentioned above. Furthermore, it was shown that hemimethylation of IS10 increased transposition (Roberts *et al.*, 1985; Yin *et al.*, 1988). The present study also observed that a considerable proportion (approximately 20%) of the DMRs belonged to the promotor region of the transposase family of proteins. In the fully methylated differential motifs, three motifs belonged to transposase protein, of which only one transposase protein was downregulated. The ISRso5 gene RSp1675 had two differential methylated motifs and the gene was downregulated in clones Haw c1, d5, e3 and Bean a4, b1.

In perspective of this work, the differential methylation marks detected by SMRT sequencing in the evolved clones must be verified using another approach, the MSRE approach (methylation sensitive restriction enzyme) followed by PCR or qPCR (Krygier *et al.*, 2016; Payelleville *et al.*, 2019). The method uses methylation sensitive restriction enzymes to cut the motif followed by PCR and gel electrophoresis to verify the band length of control and the evolved clone. The last part of this project would be to validate the contribution of an individual methylation marks to a phenotypic trait which would be achieved by site-directed strategy. Non-methylated base will be generated by site-directed substitution while keeping the codon identity. This substitution will be performed using PCR and introduced into GMI1000 by natural transformation using the gene replacement strategy as described (Perrier *et al.*, 2016). A similar approach was performed in *Salmonella* to understand the phase variation controlling the expression of lipopolysaccharide modification genes (Broadbent *et al.*, 2010). In addition, complete deletion of the candidate genes will also be performed as a complementary approach that might induce loss of function. This will be accomplished using sacB protocol as described (Gopalan-Nair et al., submitted). Finally, the impact of these mutations *in planta* will be observed using competition experiments (CI) with the ancestral GMI1000 clone (Guidot *et al.*, 2014).

The list of DMRs presented in this work only concerned the G**T**WWAC motif targeted by the RSc1982 MTase. However, 6 other MTase were detected in GMI1000 genome. Unfortunately, a new motif could be detected for only 1 of these 6 MTases, the YGCCGGCR motif targeted by the RSc3438 MTase. A total of 3533 YGCCGGCR motifs were detected on the GMI1000 chromosome and 1923 motifs on the GMI1000 megaplasmid (unpublished data). Methylation of this motif is a 5mC modification. However, SMRT sequencing technology is not well suited to detect 5mC modification. Therefore, the 5mC modifications must be detected using the bisulfite treatment followed by sequencing (Li and Tollefsbol, 2011). However, the *Ralstonia* genome is GC rich and this hinder the efficiency of conversion from unmethylated cytosines to uracil and the consequent amplification. Therefore, the protocol needs to be optimized specifically for *Ralstonia*. Then we could suspect that there will be a higher number of differential methylation marks between the ancestral clone and the evolved clones.

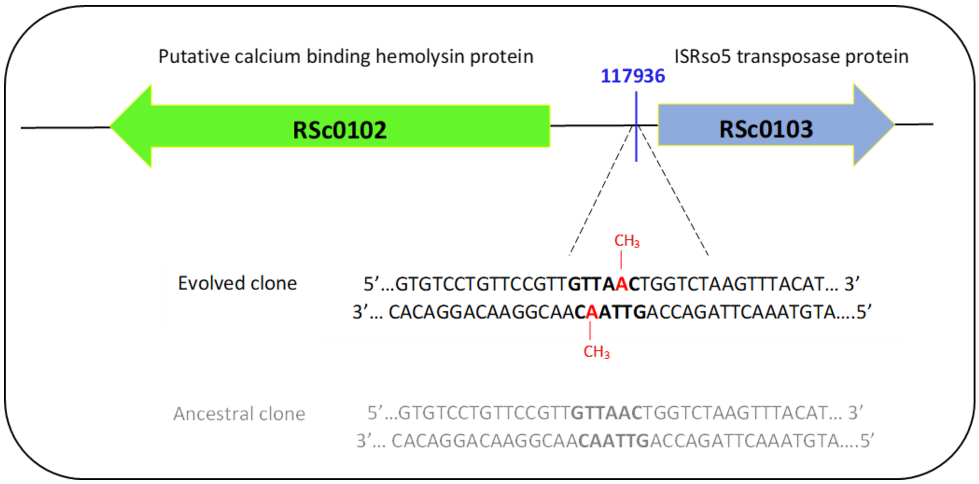


Figure 28 Differentially methylated motif at RSc2612 on R. solanacearum

RSc2612 encodes hypothetical protein (STRING database – putative transmembrane protein). The position of the motif on the chromosome is highlighted in blue font. The presence of methylated base is highlighted in red font as a comparison to the ancestral clone (GMI1000).

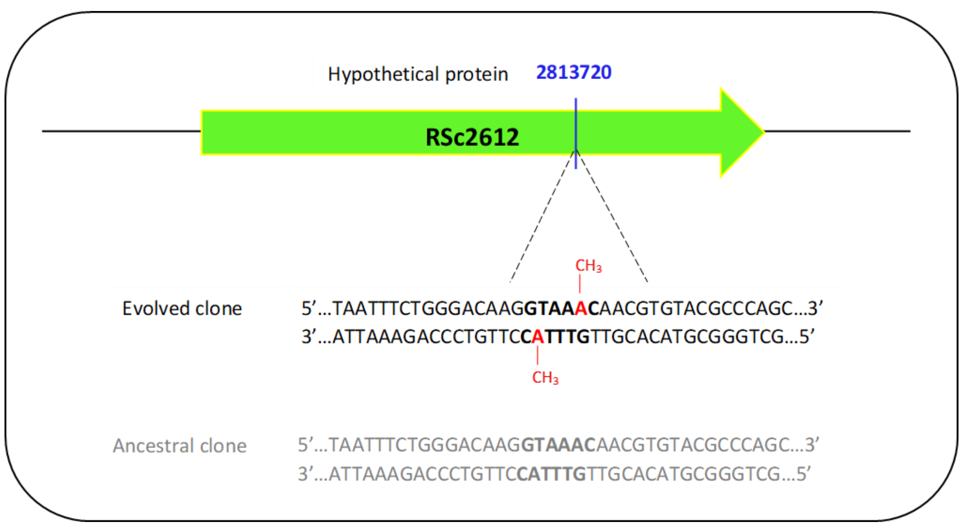


Figure 29 Differentially methylated motif upstream RSc0102-03 on R. solanacearum

RSc0102 encodes putative calcium binding protein and RSc0103 encodes transposase protein. The position of the motif on the chromosome is highlighted in blue font. The presence of methylated base is highlighted in red font as a comparison to the ancestral clone (GMI1000).

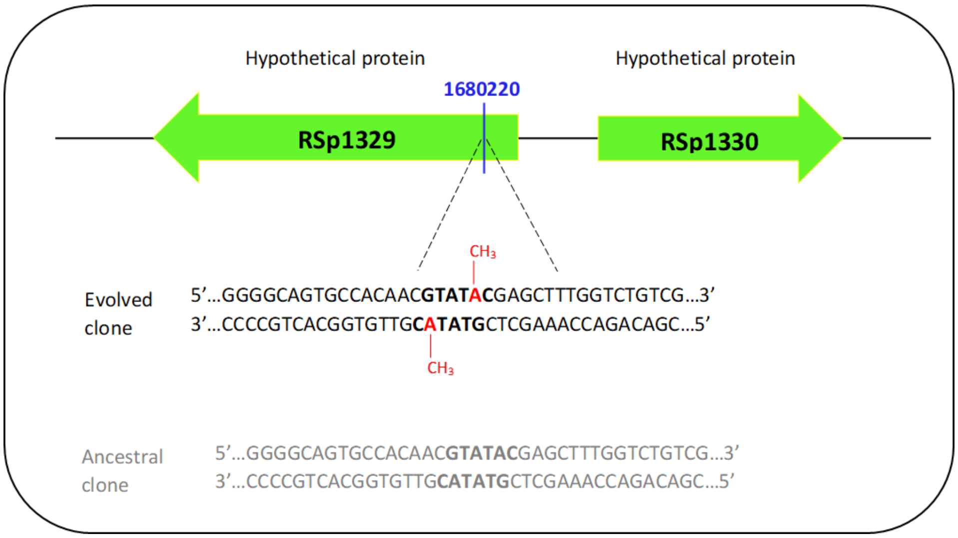


Figure 30 Differentially methylated motif upstream RSp1329-30 on R. solanacearum

RSp1329 encodes hypothetical protein (STRING database – putative transmembrane protein) and RSp1330 encodes hypothetical protein. The position of the motif on the chromosome is highlighted in blue font. The presence of methylated base is highlighted in red font as a comparison to the ancestral clone (GMI1000).

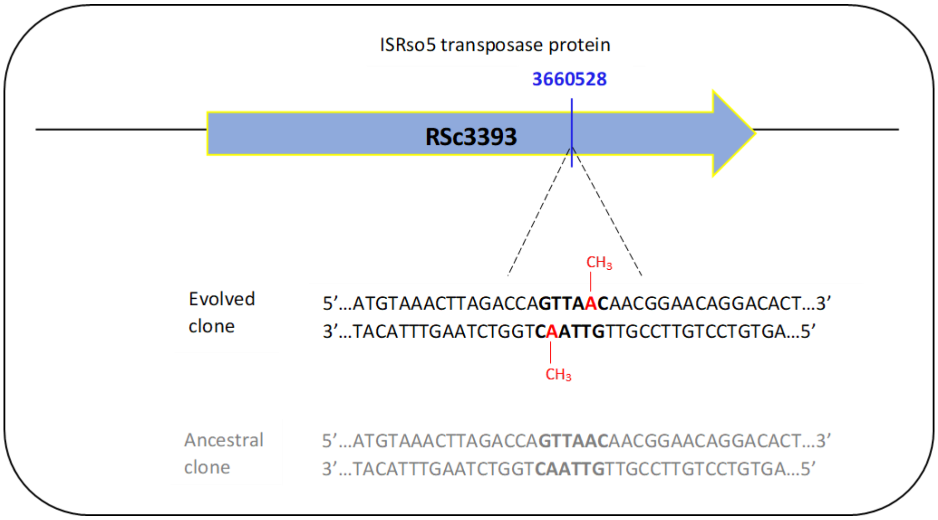


Figure 31 Differentially methylated motif at RSc3393 on R. solanacearum

RSc3393 encodes transposase protein. The position of the motif on the chromosome is highlighted in blue font. The presence of methylated base is highlighted in red font as a comparison to the ancestral clone (GMI1000).

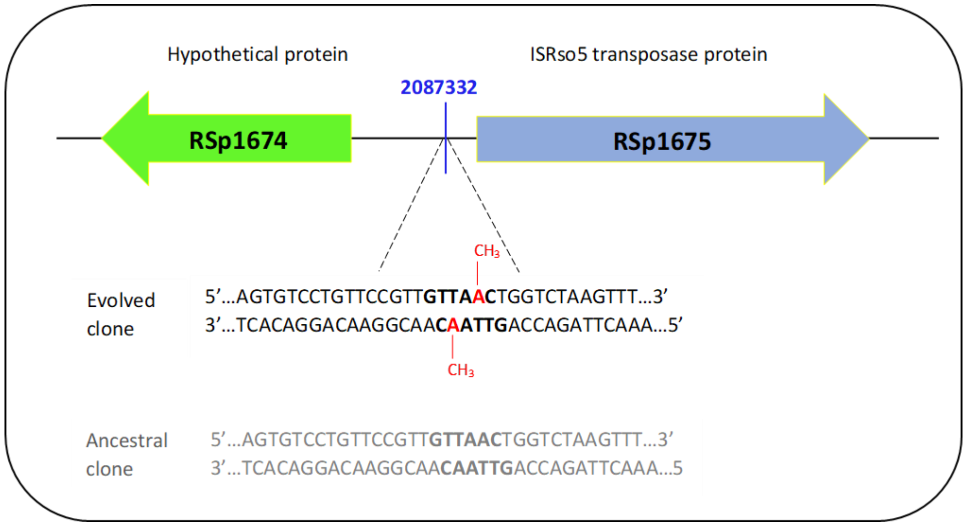


Figure 32 Differentially methylated motif upstream RSp1674-75 on R. solanacearum

RSp1674 encodes hypothetical protein and RSp1675 encodes transposase protein. The position of the motif on the megaplasmid is highlighted in blue font. The presence of methylated base is highlighted in red font as a comparison to the ancestral clone (GMI1000).

Chapter 5. Conclusion and Perspectives

Plant pathogens represent a huge problem to vegetation worldwide that has led to significant impact on both economy and environment (Strange and Scott, 2005). It is vital to improve the disease management approaches of the agricultural crops in particular, for global food security (Martins *et al.*, 2018). An important parameter to consider in order to guide strategies for improvement of disease management approaches is that plant-pathogen interaction is in constant evolution through an arms race of pathogen attack and plant defense (Jones and Dangl, 2006). In order to overcome plant pathogens, it is thus crucial to understand the molecular mechanisms that govern host adaptation. This project focuses on one of the most destructive plant pathogen, the *Ralstonia solanacearum* species complex (RSSC) that affects more than 250 plant species belonging to both monocots and dicots (Mansfield *et al.*, 2012). This multihost pathogen is known for rapid adaptation to new plant species and new environments (Wicker *et al.*, 2007, 2009; van der Gaag *et al.*, 2019). My thesis project was part of a global project aiming at deciphering the genetic and epigenetic bases of RSSC adaptation to new host plants.

The specific objectives of my thesis were (1) to decipher the genetic bases of adaptation of a RSSC strain to a resistant tomato cultivar, (2) to investigate the potential role of epigenetic modifications in host adaptation and (3) to analyze to impact of the plant species on genetic, transcriptomic and epigenetic modifications in RSSC adapted clones. This study was conducted on clones generated by experimental evolution of GMI1000 RSSC strain after 300 generations of serial passages on the resistant tomato ‘Hawaii 7996’ plant, the susceptible eggplant ‘Zebrina’ and the tolerant plant Bean ‘Blanc precoce’.

What did we learn from experimental evolution of *R. solanacearum* on a resistant tomato cultivar?

Before my thesis, Guidot and coworkers conducted several analyses on clones experimentally evolved on susceptible hosts (tomato, eggplant and Pelargonium) and distant hosts (bean and cabbage) (Guidot *et al.*, 2014; Perrier *et al.*, 2016, 2019). However, the clones evolved on the resistant tomato cultivar Hawaii 7996 were never investigated.

The Hawaii 7996 tomato is a reference resistant cultivar, as it has a stable resistance against RSSC. Several studies demonstrated the association of resistance against several strains of RSSC and identified multiple QTLs identified in Hawaii 7996 (Carmeille *et al.*, 2006; Lebeau *et al.*, 2010; Wang *et al.*, 2013). The experimental evolution of *R. solanacearum* on Hawaii 7996 revealed that over 300 generations were not sufficient for the pathogen to overcome the polygenic resistance of Hawaii 7996. Indeed, no disease symptoms were observed during the course of the experimental evolution. However, competition assays with the ancestral clone demonstrated adaptation of the evolved clones for growth into the Hawaii 7996 plant stem (Gopalan-Nair et al., submitted). The CI of the clones evolved on Hawaii was statistically higher than the CI of clones evolved on susceptible host (Zebrina), but not significantly different from CI of clones evolved in distant host (Bean). This is consistent with the previous work on the experimentally evolved clones where higher fitness were observed in clones evolved in distant hosts such as Bean and Cabbage in comparison to the clones evolved in original hosts like Marmande, Zebrina and Pelargonium (Guidot *et al.*, 2014).

Genome sequencing of the Hawaii evolved clones showed no more than 2 mutations per evolved clone. This was in coherence with the low number of genomic polymorphisms observed in the clones evolved in other host plants (Guidot *et al.*, 2014). However, in half of the investigated Hawaii clones, no genomic polymorphism could be detected by both Illumina and Pacbio sequencing technologies. In these evolved clones, the *efpR* gene was not mutated whereas it was mutated in clones evolved in Bean, tomato Marmande and Eggplant (Guidot *et al.*, 2014; Perrier *et al.*, 2016, 2019). Here, mutations were detected in two other regulators, *prhP* and RSp1574. PrhP is known for its importance in RSSC infection and adaptation *in planta* (Zhang *et al.*, 2019) and RSp1574 is a regulator of unknown function. These two regulators represent new candidate genes important for RSSC adaptation to the host plant.

Reverse genetic approaches validated that all five genomic modifications observed in the Hawaii evolved clones were adaptive mutations. In addition, we demonstrated that all five mutations resulted in adaptation through gain of function mutation.

What did we learn from transcriptome analysis of the experimentally evolved clones?

The transcriptomes of the clones evolved on various host plants (Hawaii, Zebrina and Bean) were analyzed in comparison to the ancestral clone GMI1000 for the first time and two important observations were made. Firstly, key regulators such as EfpR, PrhP and HrpB were downregulated in clones evolved on various hosts. The global catabolic repressor and virulence regulator EfpR was downregulated in parallel populations of Hawaii clones while the gene was not mutated. Similarly, the gene was downregulated with no mutation in one of the Zebrina clones. The efpR gene was also significantly downregulated in one of the Bean clone carrying a nonsynonymous mutation in *efpR*. The recently reported virulence regulator PrhP was downregulated in two clones evolved on Hawaii and in five clones evolved on Bean. Interestingly, the down-regulation of the *prhP* gene could be associated to a genetic mutation of the gene in only one of these seven clones. Prevalence of *prhP* downregulation in clones evolved on distant experimental hosts (Hawaii and Bean) exhibits PrhP as a novel candidate critical for adaptation of RSSC to new host plants. The HrpB regulator of the T3SS was also found to be down-regulated in three clones evolved on Hawaii, one clone evolved on Zebrina and one clone evolved on Bean. These results highlight the EfpR, PrhP and HrpB regulators as important key nodes of the virulence regulatory network that were frequently downregulated by either genomic or epigenomic modification during RSSC adaptation.

Secondly, regardless of the experimental host there was a significant overlap in the differentially expressed genes among the clones (Hawaii, Zebrina and Bean). Comparative analyses of the differentially regulated genes from each evolved clones showed that more than half of the HrpB regulon and half of the EfpR regulon were present in the lists of DEGs. These findings show that the adaptive events that occurred during experimental evolution induce a global rewiring of the virulence regulatory network (Gopalan-Nair et al, submitted).

What did we learn from methylome analysis of the experimentally evolved clones?

This work was the first to demonstrate modification of the methylation profile during experimental evolution and foremost an attempt to understand the importance of DNA methylation in adaptation of the plant pathogen, RSSC. The potential impact of differential methylation is supported by the transcriptomic variation in clones with no mutation and the correlation of some of the differentially methylated regions (DMRs) with some DEGs.

Of the seven MTases identified using REBASE platform in the RSSC strain GMI1000, the MTase RSc1982 methylates adenine base with the motif GTWW**A**C. The methylation profile comparison of the motif G**T**WWAC between the ancestral clone and the evolved clones (Hawaii, Zebrina and Bean) revealed between 10 and 35 DMRs. No impact of the CI nor the genomic polymorphisms on the number of DMRs was detected. The number of DMRs detected in clones evolved on Bean was significantly higher than the number of DMRs detected in clones evolved on Hawaii or on Zebrina.

Interestingly, more than half of the DMRs were present in the promotor/upstream region of genes. These potential differentially methylated upstream regions could impact gene expression (Casadesus and Low, 2006). The potential impact of differential methylation mark on gene expression was detected for transposase genes, and genes involved in virulence, chemotaxis and structural stability.

In the perspective of this work, the potential DMRs must be validated. For that purpose, an enzymatic approach could be used, the MSRE (Methylation sensitive restriction enzyme) PCR or qPCR (Krygier *et al.*, 2016; Payelleville *et al.*, 2019). In a second step, the functional characterization of the validated DMRs could be achieved using a site-directed approach. Non-methylated base generated by site-directed substitution while keeping the codon identity using PCR will be introduced into GMI1000 strain by natural transformation using the gene replacement strategy as previously described (Broadbent *et al.*, 2010; Perrier *et al.*, 2016). Complementary to this approach, gene deletion using sacB method will be used, which could induce loss of function (Gopalan-Nair *et al.,* submitted). Ultimately, the mutants obtained by both methods will be subjected to competitive assays in their experimental host to identify their fitness potential.

The methylation profile presented in this work depicts only the G**T**WWA̲C motif methylated by the RSc1982 gene. However, there are other MTases in the RSSC genome, which suggests more methylation and probably more differential methylation marks that might have occurred during experimental evolution.

References

Ackermann, M. *et al.* (2008) 'Self-destructive cooperation mediated by phenotypic noise', *Nature*, 454(7207), pp. 987–990. doi: 10.1038/nature07067.

Al Akeel, R. (2013) 'Role of epigenetic reprogramming of host genes in bacterial pathogenesis', *Saudi Journal of Biological Sciences*. King Saud University, 20(4), pp. 305–309. doi: 10.1016/j.sjbs.2013.05.003.

Alfano, J. R. *et al.* (2000) *The Pseudomonas syringae Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants*. Available at: www.pnas.org (Accessed: 2 May 2020).

Arnoldini, M. *et al.* (2014) 'Bistable expression of virulence genes in salmonella leads to the formation of an antibiotic-tolerant subpopulation', *PLoS Biology*, 12(8). doi: 10.1371/journal.pbio.1001928.

Balaban, N. Q. (2004) 'Bacterial {Persistence} as a {Phenotypic} {Switch}', *Science*, 305(5690), pp. 1622–1625. doi: 10.1126/science.1099390.

Bartoli, C., Roux, F. and Lamichhane, J. R. (2016) 'Molecular mechanisms underlying the emergence of bacterial pathogens: an ecological perspective', *Molecular Plant Pathology*. Blackwell Publishing Ltd, 17(2), pp. 303–310. doi: 10.1111/mpp.12284.

Beck, S. and Rakyan, V. K. (2008) 'The methylome: approaches for global DNA methylation profiling', *Trends in Genetics*. Elsevier Current Trends, pp. 231–237. doi: 10.1016/j.tig.2008.01.006.

Bendall, M. L. *et al.* (2013) 'Exploring the Roles of DNA Methylation in the Metal-Reducing Bacterium Shewanella oneidensis MR-1 Downloaded from', 195, pp. 4966–4974. doi: 10.1128/JB.00935-13.

Bender, C. L. *et al.* (1999) *Pseudomonas syringae Phytotoxins: Mode of Action, Regulation, and*

Biosynthesis by Peptide and Polyketide Synthetases, *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS*.

Bergval, I. L. *et al.* (2007) 'Specific mutations in the *Mycobacterium tuberculosis rpoB* gene are associated with increased *dnaE2* expression', *FEMS Microbiology Letters*. Oxford Academic, 275(2), pp. 338–343. doi: 10.1111/j.1574-6968.2007.00905.x.

Bertolla, F. *et al.* (1999) *During Infection of Its Host, the Plant Pathogen Ralstonia solanacearum Naturally Develops a State of Competence and Exchanges Genetic Material*.

Blount, Z. D., Borland, C. Z. and Lenski, R. E. (2008) 'Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(23), pp. 7899–7906. doi: 10.1073/pnas.0803151105.

Blow, M. J. *et al.* (2016) 'The Epigenomic Landscape of Prokaryotes', *PLoS Genetics*, 12(2), pp. 1–28. doi: 10.1371/journal.pgen.1005854.

Blyn, L. B., Braaten, B. A. and Low, D. A. (1990) 'Regulation of pap pilin phase variation by a mechanism involving differential Dam methylation states', *EMBO Journal*. European Molecular Biology Organization, 9(12), pp. 4045–4054. doi: 10.1002/j.1460-2075.1990.tb07626.x.

Bódi, Z. *et al.* (2017) 'Phenotypic heterogeneity promotes adaptive evolution', *PLOS Biology*. Edited by M. Siegal. Public Library of Science, 15(5), p. e2000644. doi: 10.1371/journal.pbio.2000644.

Bonneaud, C., Weinert, L. A. and Kuijper, B. (2019) 'Understanding the emergence of bacterial pathogens in novel hosts', *Philosophical Transactions of the Royal Society B: Biological Sciences*. The Royal Society, 374(1782), p. 20180328. doi: 10.1098/rstb.2018.0328.

Broadbent, S. E., Davies, M. R. and Van Der Woude, M. W. (2010) 'Phase variation controls expression of Salmonella lipopolysaccharide modification genes by a DNA methylation-dependent mechanism', *Molecular Microbiology*, 77(2), pp. 337–353. doi: 10.1111/j.1365-2958.2010.07203.x.

Brussow, H., Canchaya, C. and Hardt, W.-D. (2004) 'Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion', *Microbiology and Molecular Biology Reviews*. American Society for Microbiology, 68(3), pp. 560–602. doi: 10.1128/mmbr.68.3.560-602.2004.

Cai, R. *et al.* (2011) 'The Plant Pathogen Pseudomonas syringae pv. tomato Is Genetically Monomorphic and under Strong Selection to Evade Tomato Immunity', *PLoS Pathogens*. Edited by D. S. Guttman. Public Library of Science, 7(8), p. e1002130. doi: 10.1371/journal.ppat.1002130.

Camacho, E. M. and Casadesús, J. (2005) 'Regulation of *traJ* transcription in the Salmonella virulence plasmid by strand-specific DNA adenine hemimethylation', *Molecular Microbiology*, 57(6), pp. 1700–1718. doi: 10.1111/j.1365-2958.2005.04788.x.

Capela, D. *et al.* (2017) 'Recruitment of a Lineage-Specific Virulence Regulatory Pathway Promotes Intracellular Infection by a Plant Pathogen Experimentally Evolved into a Legume Symbiont', *Molecular Biology and Evolution*, 34(10), pp. 2503–2521. doi: 10.1093/molbev/msx165.

Carmeille, A. *et al.* (2006) 'Identification of QTLs for Ralstonia solanacearum race 3-phylotype II resistance in tomato', *Theoretical and Applied Genetics*. Springer, 113(1), pp. 110–121. doi: 10.1007/s00122-006-0277-3.

Casadesús, J. (2016) 'Bacterial DNA methylation and methylomes', *Advances in Experimental Medicine and Biology*. Springer New York LLC, 945, pp. 35–61. doi: 10.1007/978-3-319-43624-1_3.

Casadesus, J. and Low, D. (2006) 'Epigenetic Gene Regulation in the Bacterial World', *Microbiology and Molecular Biology Reviews*, 70(3), pp. 830–856. doi: 10.1128/MMBR.00016-06.

Castillo, J. A. and Agathos, S. N. (2019) 'A genome-wide scan for genes under balancing selection in the plant pathogen Ralstonia solanacearum', *BMC Evolutionary Biology*. BMC Evolutionary Biology, 19(1), pp. 1–16. doi: 10.1186/s12862-019-1456-6.

Castillo, J. A. and Greenberg, J. T. (2007) 'Evolutionary dynamics of Ralstonia solanacearum', *Applied and Environmental Microbiology*, 73(4), pp. 1225–1238. doi: 10.1128/AEM.01253-06.

Cellier, G. *et al.* (2012) 'Phylogeny and population structure of brown rot- and Moko disease-causing strains of Ralstonia solanacearum phylotype II', *Applied and Environmental Microbiology*, 78(7), pp. 2367–2375. doi: 10.1128/AEM.06123-11.

Cianciotto, N. P. and White, R. C. (2017) 'Expanding role of type II secretion in bacterial pathogenesis and beyond', *Infection and Immunity*. American Society for Microbiology, 85(5). doi: 10.1128/IAI.00014-17.

Clough, S. J. *et al.* (1997) *Differential Expression of Virulence Genes and Motility in Ralstonia (Pseudomonas) solanacearum during Exponential Growth*, *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*. Available at: http://aem.asm.org/ (Accessed: 2 September 2020).

Conrad, T. M. *et al.* (2009) 'Whole-genome resequencing of Escherichia coli K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations', *Genome Biology*, 10(10), pp. 1–12. doi: 10.1186/gb-2009-10-10-r118.

Coupat-Goutaland, B. *et al.* (2011) 'Ralstonia solanacearum virulence increased following large interstrain gene transfers by natural transformation', *Molecular Plant-Microbe Interactions*. The American Phytopathological Society , 24(4), pp. 497–505. doi: 10.1094/MPMI-09-10-0197.

Coupat, B. *et al.* (2008) 'Natural transformation in the Ralstonia solanacearum species complex: Number and size of DNA that can be transferred', *FEMS Microbiology Ecology*, 66(1), pp. 14–24. doi: 10.1111/j.1574-6941.2008.00552.x.

Davies, J. (1994) 'Inactivation of antibiotics and the dissemination of resistance genes', *Science*, 264(5157), pp. 375–382. doi: 10.1126/science.8153624.

Davis, B. M., Chao, M. C. and Waldor, M. K. (2013) 'Entering the era of bacterial epigenomics with single molecule real time DNA sequencing', *Current Opinion in Microbiology*. Elsevier Ltd, 16(2), pp. 192–198. doi: 10.1016/j.mib.2013.01.011.

Drenkard, E. and Ausubel, F. M. (2002) 'Pseudomonas biofilm formation and antibiotic resistance are linked to phenotypic variation', *Nature*. Nature Publishing Group, 416(6882), pp. 740–743. doi: 10.1038/416740a.

Elena, S. F. and Lenski, R. E. (2003) 'Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation', *Nature Reviews Genetics*, pp. 457–469. doi: 10.1038/nrg1088.

Elphinstone, J. (2005) 'The current bacterial wilt situation: a global overview'. APS Press, pp. 9–28.

Erill, I. *et al.* (2017) 'Comparative Analysis of Ralstonia solanacearum Methylomes', *Frontiers in Plant Science*, 8. doi: 10.3389/fpls.2017.00504.

Fang, G. *et al.* (2012) 'Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing', *Nature Biotechnology*, 30(12), pp. 1232–1239. doi: 10.1038/nbt.2432.

Flusberg, B. A. *et al.* (2010) 'Direct detection of DNA methylation during single-molecule, real-time sequencing', *Nature Methods*. Nature Publishing Group, 7(6), pp. 461–465. doi: 10.1038/nmeth.1459.

van der Gaag, D. J. *et al.* (2019) 'Pest survey card on potato brown rot, Ralstonia solanacearum', *EFSA Supporting Publications*, 16(2). doi: 10.2903/sp.efsa.2019.en-1567.

Genin, S. (2010) 'Molecular traits controlling host range and adaptation to plants in Ralstonia solanacearum', *New Phytologist*, 187(4), pp. 920–928. doi: 10.1111/j.1469-8137.2010.03397.x.

Genin, S. and Boucher, C. (2002a) 'Ralstonia solanacearum: Secrets of a major pathogen unveiled by analysis of its genome', *Molecular Plant Pathology*. doi: 10.1046/j.1364-3703.2002.00102.x.

Genin, S. and Boucher, C. (2002b) 'Ralstonia solanacearum: Secrets of a major pathogen unveiled by analysis of its genome', *Molecular Plant Pathology*, 3(3), pp. 111–118. doi: 10.1046/j.1364-3703.2002.00102.x.

Genin, S. and Boucher, C. (2004) 'LESSONS LEARNED FROM THE GENOME ANALYSIS OF *RALSTONIA SOLANACEARUM*', *Annual Review of Phytopathology*, 42(1), pp. 107–134. doi: 10.1146/annurev.phyto.42.011204.104301.

Genin, S. and Denny, T. P. (2012) ‘Pathogenomics of the *Ralstonia solanacearum* Species Complex’, *Annual Review of Phytopathology*. Annual Reviews , 50(1), pp. 67–89. doi: 10.1146/annurev-phyto-081211-173000.

van Gestel, J. and Nowak, M. A. (2016) ‘Phenotypic Heterogeneity and the Evolution of Bacterial Life Cycles’, *PLOS Computational Biology*. Edited by C. O. Wilke. Public Library of Science, 12(2), p. e1004764. doi: 10.1371/journal.pcbi.1004764.

Grohmann, E., Muth, G. and Espinosa, M. (2003) ‘Conjugative Plasmid Transfer in Gram-Positive Bacteria’, *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS*, 67(2), pp. 277–301. doi: 10.1128/MMBR.67.2.277-301.2003.

Groisman, E. A. and Casadesús, J. (2005) ‘The origin and evolution of human pathogens’, *Molecular Microbiology*, 56(1), pp. 1–7. doi: 10.1111/j.1365-2958.2005.04564.x.

Guidot A, Prior P, Schoenfeld J, Carrère S, Genin S, B. C. (2007) ‘Genomic Structure and Phylogeny of the Plant Pathogen Ralstonia solanacearum Inferred from Gene Distribution Analysis’, *Journal of Bacteriology*, 189(2), pp. 377–387. doi: 10.1128/JB.00999-06.

Guidot, A. *et al.* (2009) ‘Horizontal gene transfer between ralstonia solanacearum strains detected by comparative genomic hybridization on microarrays’, *ISME Journal*. Nature Publishing Group, 3(5), pp. 549–562. doi: 10.1038/ismej.2009.14.

Guidot, A. *et al.* (2014) ‘Multihost experimental evolution of the pathogen ralstonia solanacearum unveils genes involved in adaptation to plants’, *Molecular Biology and Evolution*, 31(11), pp. 2913–2928. doi: 10.1093/molbev/msu229.

Haagmans, W. and Van Der Woude, M. (2000) ‘Phase variation of Ag43 in Escherichia coli: Dam-dependent methylation abrogates OxyR binding and OxyR-mediated repression of transcription’, *Molecular Microbiology*. Mol Microbiol, 35(4), pp. 877–887. doi: 10.1046/j.1365-2958.2000.01762.x.

Hale, W. B., Van Der Woude, M. W. and Low, D. A. (1994) *Analysis of Nonmethylated GATC Sites in the Escherichia coli Chromosome and Identification of Sites That Are Differentially Methylated* *in Response to Environmental Stimuli*, *JOURNAL OF BACTERIOLOGY*.

Herring, C. D. *et al.* (2006) 'Comparative genome sequencing of Escherichia coli allows observation of bacterial evolution on a laboratory timescale', 38. doi: 10.1038/ng1906.

Hertwig, S. *et al.* (1999) *Generalized Transduction of Small Yersinia enterocolitica Plasmids*, *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*.

Huet, G. (2014) 'Breeding for resistances to Ralstonia solanacearum', *Frontiers in Plant Science*, 5(DEC), pp. 1–5. doi: 10.3389/fpls.2014.00715.

Ichinose, Y., Taguchi, F. and Mukaihara, T. (2013) 'Pathogenicity and virulence factors of Pseudomonas syringae', *Journal of General Plant Pathology*. Springer, pp. 285–296. doi: 10.1007/s10327-013-0452-8.

Jeltsch, A. (2002) 'Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases', *ChemBioChem*. John Wiley & Sons, Ltd, 3(4), pp. 274–293. doi: 10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S.

Johnson, R. (1983) 'Genetic Background of Durable Resistance', in *Durable Resistance in Crops*. Springer New York, pp. 5–26. doi: 10.1007/978-1-4615-9305-8_2.

Jones, J. D. G. and Dangl, J. L. (2006) 'The plant immune system', *Nature*. Nature, pp. 323–329. doi: 10.1038/nature05286.

Jones, P. A. (2012) 'Functions of DNA methylation: islands, start sites, gene bodies and beyond', *Nature Reviews Genetics*. doi: 10.1038/nrg3230.

Julio, S. M. *et al.* (2001) 'DNA Adenine Methylase Is Essential for Viability and Plays a Role in the Pathogenesis of Yersinia pseudotuberculosis and Vibrio cholerae', *INFECTION AND IMMUNITY*, 69(12), pp. 7610–7615. doi: 10.1128/IAI.69.12.7610-7615.2001.

Julio, S. M. *et al.* (2002) 'DNA adenine methylase overproduction in Yersinia pseudotuberculosis alters YopE expression and secretion and host immune responses to infection', *Infection and Immunity*, 70(2), pp. 1006–1009. doi: 10.1128/IAI.70.2.1006-1009.2002.

Kawecki, T. J. *et al.* (2012) 'Experimental evolution', *Trends in Ecology and Evolution*. Elsevier Ltd, 27(10), pp. 547–560. doi: 10.1016/j.tree.2012.06.001.

Khokhani, D. *et al.* (2017) 'A Single Regulator Mediates Strategic Switching between Attachment/Spread and Growth/Virulence in the Plant Pathogen Ralstonia solanacearum Downloaded from'. doi: 10.1128/mBio.00895-17.

King, K. C. *et al.* (2016) 'Rapid evolution of microbe-mediated protection against pathogens in a worm host', *ISME Journal*. Nature Publishing Group, 10(8), pp. 1915–1924. doi: 10.1038/ismej.2015.259.

Kondrashov, F. A. (2012) 'Gene duplication as a mechanism of genomic adaptation to a changing environment', *Proceedings of the Royal Society B: Biological Sciences*, 279(1749), pp. 5048–5057. doi: 10.1098/rspb.2012.1108.

Krygier, M. *et al.* (2016) 'A simple modification to improve the accuracy of methylation-sensitive restriction enzyme quantitative polymerase chain reaction', *Analytical Biochemistry*, 500, pp. 88–90. doi: 10.1016/j.ab.2016.01.020.

Kumar, S. *et al.* (2018) 'N4-cytosine DNA methylation regulates transcription and pathogenesis in Helicobacter pylori', *Nucleic acids research*. NLM (Medline), 46(7), pp. 3429–3445. doi: 10.1093/nar/gky126.

Lebeau, A. *et al.* (2010) 'Bacterial Wilt Resistance in Tomato , Pepper , and Eggplant : Genetic Resources Respond to Diverse Strains in the Ralstonia solanacearum Species Complex', (i).

Lenski, R. E. *et al.* (1991) 'Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2000 generations', *American Naturalist*, 138(6), pp. 1315–1341. doi: 10.1086/285289.

Li, Y. and Tollefsbol, T. O. (2011) 'DNA methylation detection: Bisulfite genomic sequencing analysis', *Methods in Molecular Biology*. NIH Public Access, 791, pp. 11–21. doi: 10.1007/978-1-61779-316-5_2.

Lie Angot, A. *et al.* (2006) *Ralstonia solanacearum requires F-box-like domain-containing type III* *effectors to promote disease on several host plants*. Available at: www.pnas.orgcgidoi10.1073pnas.0509393103 (Accessed: 22 April 2020).

Lindsey, H. A. *et al.* (2013) 'Evolutionary rescue from extinction is contingent on a lower rate of environmental change', *Nature*. doi: 10.1038/nature11879.

Lister, R. and Ecker, J. R. (2009) 'Finding the fifth base: Genome-wide sequencing of cytosine methylation', *Genome Research*. Cold Spring Harbor Laboratory Press, pp. 959–966. doi: 10.1101/gr.083451.108.

Lopes, C. A. and Rossato, M. (2018) 'History and status of selected hosts of the Ralstonia solanacearum species complex causing bacterial wilt in Brazil', *Frontiers in Microbiology*, 9(JUN), pp. 1–6. doi: 10.3389/fmicb.2018.01228.

Low, D. A. and Casadesús, J. (2008) 'Clocks and switches: bacterial gene regulation by DNA adenine methylation', *Current Opinion in Microbiology*. Elsevier Current Trends, pp. 106–112. doi: 10.1016/j.mib.2008.02.012.

Low, D. A., Weyand, N. J. and Mahan, M. J. (2001) 'MINIREVIEW Roles of DNA Adenine Methylation in Regulating Bacterial Gene Expression and Virulence', *INFECTION AND IMMUNITY*, 69(12), pp. 7197–7204. doi: 10.1128/IAI.69.12.7197-7204.2001.

Macho, A. P. *et al.* (2010) 'A Competitive Index Assay Identifies Several Ralstonia solanacearum Type III Effector Mutant Strains with Reduced Fitness in Host Plants', *The American Phytopathological Society*, 23(9), pp. 1197–1205. Available at: https://apsjournals.apsnet.org/doi/10.1094/MPMI-23-9-1197.

Mansfield, J. *et al.* (2012) 'Top 10 plant pathogenic bacteria in molecular plant pathology', *Molecular Plant Pathology*, 13(6), pp. 614–629. doi: 10.1111/j.1364-3703.2012.00804.x.

Marchetti, M. *et al.* (2010) 'Experimental evolution of a plant pathogen into a legume symbiont', *PLoS Biology*, 8(1). doi: 10.1371/journal.pbio.1000280.

Marchetti, M. *et al.* (2014) 'Shaping bacterial symbiosis with legumes by experimental evolution', *Molecular Plant-Microbe Interactions*, 27(9), pp. 956–964. doi: 10.1094/MPMI-03-14-0083-R.

Marcus, S. L. *et al.* (2000) 'Salmonella pathogenicity islands: Big virulence in small packages', *Microbes and Infection*. Elsevier Masson SAS, pp. 145–156. doi: 10.1016/S1286-4579(00)00273-2.

Marinus, M. G. and Casadesus, J. (no date) 'Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more'. doi: 10.1111/j.1574-6976.2008.00159.x.

Martins, P. M. M. *et al.* (2018) 'Persistence in phytopathogenic bacteria: Do we know enough?', *Frontiers in Microbiology*. Frontiers Media S.A. doi: 10.3389/fmicb.2018.01099.

Maurelli, A. T. *et al.* (1998) '"Black holes" and bacterial pathogenicity: A large genomic deletion that enhances the virulence of Shigella spp. and enteroinvasive Escherichia coli', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 95(7), pp. 3943–3948. doi: 10.1073/pnas.95.7.3943.

Mayjonade, B. *et al.* (2017) 'Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules', *BioTechniques*, 62(1), p. xv. doi: 10.2144/000114503.

Mcdonald, B. A. and Linde, C. (2002) 'PATHOGEN POPULATION GENETICS, EVOLUTIONARY POTENTIAL, AND DURABLE RESISTANCE', *Annu. Rev. Phytopathol*, 40, pp. 349–79. doi: 10.1146/annurev.phyto.40.120501.101443.

Mehling, J. S., Lavender, H. and Clegg, S. (2007) 'A Dam methylation mutant of Klebsiella pneumoniae is partially attenuated', *FEMS Microbiology Letters*, 268(2), pp. 187–193. doi: 10.1111/j.1574-6968.2006.00581.x.

Nagy, Z. and Chandler, M. (2004) 'Regulation of transposition in bacteria', *Research in Microbiology*. Elsevier Masson SAS, 155(5), pp. 387–398. doi: 10.1016/j.resmic.2004.01.008.

Nunney, L. *et al.* (2014) 'Large-scale intersubspecific recombination in the plant-pathogenic bacterium xylella fastidiosa is associated with the host shift to mulberry', *Applied and Environmental Microbiology*. American Society for Microbiology, 80(10), pp. 3025–3033. doi: 10.1128/AEM.04112-13.

Nuss, A. M. *et al.* (2016) 'A Precise Temperature-Responsive Bistable Switch Controlling Yersinia Virulence', *PLOS Pathogens*. Edited by D. M. Monack. Public Library of Science, 12(12), p. e1006091. doi: 10.1371/journal.ppat.1006091.

Nye, T. M. *et al.* (2020) 'Methyltransferase DnmA is responsible for genome-wide N6-methyladenosine modifications at non-palindromic recognition sites in Bacillus subtilis', *Nucleic Acids Research*, 48(10), pp. 5332–5348. doi: 10.1093/nar/gkaa266.

Ogier, J. C. *et al.* (2019) 'RpoB, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing', *BMC Microbiology*. BioMed Central Ltd., 19(1), p. 171. doi: 10.1186/s12866-019-1546-z.

Oliveira, P. H. *et al.* (2020) 'Epigenomic characterization of Clostridioides difficile finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis', *Nature Microbiology*. Springer US, 5(1), pp. 166–180. doi: 10.1038/s41564-019-0613-4.

Oliveira, P. H. and Fang, G. (2020) 'Conserved DNA Methyltransferases: A Window into Fundamental Mechanisms of Epigenetic Regulation in Bacteria', *Trends in Microbiology*. Elsevier Ltd, pp. 0–12. doi: 10.1016/j.tim.2020.04.007.

Payelleville, A. *et al.* (2017) 'DNA Adenine Methyltransferase (Dam) Overexpression Impairs Photorhabdus luminescens Motility and Virulence', *Frontiers in Microbiology*. Frontiers Media S.A., 8(SEP), p. 1671. doi: 10.3389/fmicb.2017.01671.

Payelleville, A. *et al.* (2018) 'The complete methylome of an entomopathogenic bacterium reveals the existence of loci with unmethylated Adenines', *Scientific Reports*, 8(1), pp. 1–14. doi: 10.1038/s41598-018-30620-5.

Payelleville, A. *et al.* (2019) 'Role of the Photorhabdus Dam methyltransferase during interactions with its invertebrate hosts', *PLoS ONE*, 14(10), pp. 1–14. doi: 10.1371/journal.pone.0212655.

Peeters, N., Guidot, A., *et al.* (2013) 'Ralstonia solanacearum, a widespread bacterial plant pathogen in the post-genomic era', *Molecular Plant Pathology*, 14(7), pp. 651–662. doi: 10.1111/mpp.12038.

Peeters, N., Carrère, S., *et al.* (2013) 'Repertoire, unified nomenclature and evolution of the Type III effector gene set in the Ralstonia solanacearum species complex', *BMC Genomics*, 14(1). doi: 10.1186/1471-2164-14-859.

Perrier, A. *et al.* (2016) 'Enhanced in planta Fitness through Adaptive Mutations in EfpR, a Dual Regulator of Virulence and Metabolic Functions in the Plant Pathogen Ralstonia solanacearum', *PLoS Pathogens*, 12(12), pp. 1–23. doi: 10.1371/journal.ppat.1006044.

Perrier, A. *et al.* (2018) 'Comparative transcriptomic studies identify specific expression patterns of virulence factors under the control of the master regulator PhcA in the Ralstonia solanacearum species complex', *Microbial Pathogenesis*. Academic Press, 116, pp. 273–278. doi: 10.1016/J.MICPATH.2018.01.028.

Perrier, A. *et al.* (2019) 'Spontaneous mutations in a regulatory gene induce phenotypic heterogeneity and adaptation of *Ralstonia solanacearum* to changing environments', *Environmental Microbiology*, 21(8), pp. 3140–3152. doi: 10.1111/1462-2920.14717.

Peyraud, R. *et al.* (2016) 'A Resource Allocation Trade-Off between Virulence and Proliferation Drives Metabolic Versatility in the Plant Pathogen Ralstonia solanacearum', *PLoS Pathogens*, 12(10), pp. 1–25. doi: 10.1371/journal.ppat.1005939.

Poueymiro, M. and Genin, S. (2009) 'Secreted proteins from Ralstonia solanacearum: a hundred tricks to kill a plant', *Current Opinion in Microbiology*, pp. 44–52. doi: 10.1016/j.mib.2008.11.008.

Poussier, S. *et al.* (2003) 'Host plant-dependent phenotypic reversion of Ralstonia solanacearum from non-pathogenic to pathogenic forms via alterations in the phcA gene', *Molecular Microbiology*, 49(4), pp. 991–1003. doi: 10.1046/j.1365-2958.2003.03605.x.

Preston, G. M., Haubold, B. and Rainey, P. B. (1998) 'Bacterial genomics and adaptation to life on plants: Implications for the evolution of pathogenicity and symbiosis', *Current Opinion in Microbiology*. Elsevier Ltd, 1(5), pp. 589–597. doi: 10.1016/S1369-5274(98)80094-5.

Prior, P. *et al.* (2016) 'Genomic and proteomic evidence supporting the division of the plant pathogen Ralstonia solanacearum into three species', *BMC Genomics*. BioMed Central Ltd., 17(1),

p. 90. doi: 10.1186/s12864-016-2413-z.

Prior, P. (Philippe) and Fegan, M. (2005) ‘Bacterial wilt disease and the Ralstonia solanacearum species complex’, p. 510. Available at: https://espace.library.uq.edu.au/view/UQ:71427.

Prior, P. and Fegan, M. (2005) ‘Recent developments in the phylogeny and classification of ralstonia solanacearum’, *Acta Horticulturae*, 695(May), pp. 127–136. doi: 10.17660/ActaHortic.2005.695.14.

Proctor, R. A. *et al.* (2006) ‘Small colony variants: A pathogenic form of bacteria that facilitates persistent and recurrent infections’, *Nature Reviews Microbiology*, pp. 295–305. doi: 10.1038/nrmicro1384.

Quesada, J. M. *et al.* (2016) ‘Rhizosphere selection of *Pseudomonas putida* KT2440 variants with increased fitness associated to changes in gene expression’, *Environmental Microbiology Reports*. Wiley-Blackwell, 8(5), pp. 842–850. doi: 10.1111/1758-2229.12447.

Roberts, D. *et al.* (1985) ‘IS10 transposition is regulated by DNA adenine methylation’, *Cell*. Cell, 43(1), pp. 117–130. doi: 10.1016/0092-8674(85)90017-0.

Robinson, V. L., Oyston, P. C. F. and Titball, R. W. (2005) ‘A dam mutant of Yersinia pestis is attenuated and induces protection against plague’, *FEMS Microbiology Letters*, 252(2), pp. 251–256. doi: 10.1016/j.femsle.2005.09.001.

Sacristán, S. and García-Arenal, F. (2008) ‘The evolution of virulence and pathogenicity in plant pathogen populations’, *Molecular Plant Pathology*, 9(3), pp. 369–384. doi: 10.1111/j.1364-3703.2007.00460.x.

Safni, I. *et al.* (2014) ‘Polyphasic taxonomic revision of the Ralstonia solanacearum species complex’, *International journal of systemic and evolutionary microbiology*, 64(Pt_9). doi: 10.1099/ijs.0.066712-0.

Salanoubat, M. *et al.* (2002) ‘Genome sequence of the plant pathogen ralstonia solanacearum’, *Nature*, 415(6871), pp. 497–502. doi: 10.1038/415497a.

Sánchez-Romero, M. A. and Casadesús, J. (2020) ‘The bacterial epigenome’, *Nature Reviews*

Microbiology. Nature Research, pp. 7–20. doi: 10.1038/s41579-019-0286-2.

Sánchez-Romero, M. A., Cota, I. and Casadesús, J. (2015) 'DNA methylation in bacteria: From the methyl group to the methylome', *Current Opinion in Microbiology*, 25, pp. 9–16. doi: 10.1016/j.mib.2015.03.004.

Schell, M. A. (2000) *CONTROL OF VIRULENCE AND PATHOGENICITY GENES OF RALSTONIA SOLANACEARUM BY AN ELABORATE SENSORY NETWORK*. Available at: www.annualreviews.org (Accessed: 30 January 2020).

Schicklmaier, P. *et al.* (1995) *Frequency of Generalized Transducing Phages in Natural Isolates of the Salmonella typhimurium Complex From 85 natural isolates of the Salmonella typhimurium complex, including the Salmonella reference collec-tion A (P*, *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*.

Sharma, D. and Singh, Y. (2019) 'Characterization of Ralstonia solanacearum isolates using biochemical , cultural , molecular methods and pathogenicity tests', 8(4), pp. 2884–2889.

Shell, S. S. *et al.* (2013) 'DNA Methylation Impacts Gene Expression and Ensures Hypoxic Survival of Mycobacterium tuberculosis', *PLoS Pathogens*, 9(7), pp. 24–28. doi: 10.1371/journal.ppat.1003419.

Song, S. *et al.* (2018) 'Identification of Cyclic Dipeptides from Escherichia coli as New Antimicrobial Agents against Ralstonia Solanacearum', *Molecules*. Multidisciplinary Digital Publishing Institute, 23(1), p. 214. doi: 10.3390/molecules23010214.

Strange, R. N. and Scott, P. R. (2005) 'Plant Disease: A Threat to Global Food Security', *Annual Review of Phytopathology*, 43(1), pp. 83–116. doi: 10.1146/annurev.phyto.43.113004.133839.

Thomas, C. M. and Nielsen, K. M. (2005) 'Mechanisms of, and barriers to, horizontal gene transfer between bacteria', *Nature Reviews Microbiology*, 3(9), pp. 711–721. doi: 10.1038/nrmicro1234.

Tjou-Tam-sin, N. N. A. *et al.* (2017) 'First report of bacterial wilt caused by Ralstonia solanacearum in ornamental Rosa sp', *Plant Disease*. American Phytopathological Society, 101(2), p. 378. doi: 10.1094/PDIS-02-16-0250-PDN.

Traverse, C. C. *et al.* (2013) 'Tangled bank of experimentally evolved Burkholderia biofilms reflects selection during chronic infections', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 110(3), pp. E250–E259. doi: 10.1073/pnas.1207025110.

Trivedi, P. and Wang, N. (2014) 'Host immune responses accelerate pathogen evolution', *ISME Journal*. Nature Publishing Group, 8(3), pp. 727–731. doi: 10.1038/ismej.2013.215.

Tso, G. H. W. *et al.* (2018) 'Experimental evolution of a fungal pathogen into a gut symbiont', *Science*. American Association for the Advancement of Science, 362(6414), pp. 589–595. doi: 10.1126/science.aat0537.

Wang, Jaw-Fen *et al.* (2013) 'Identification of major QTLs associated with stable resistance of tomato cultivar "Hawaii 7996" to Ralstonia solanacearum', *Euphytica*, 190, pp. 241–252. doi: 10.1007/s10681-012-0830-x.

Weigel, W. A. and Dersch, P. (2018) 'Phenotypic heterogeneity: a bacterial virulence strategy', *Microbes and Infection*. Elsevier Ltd, 20(9–10), pp. 570–577. doi: 10.1016/j.micinf.2018.01.008.

Weissman, S. J. *et al.* (2003) 'Enterobacterial adhesins and the case for studying SNPs in bacteria', *Trends in Microbiology*. Elsevier Ltd, pp. 115–117. doi: 10.1016/S0966-842X(03)00010-6.

Wicker, E. *et al.* (2007) 'Ralstonia solanacearum strains from Martinique (French West Indies) exhibiting a new pathogenic potential', *Applied and Environmental Microbiology*, 73(21), pp. 6790–6801. doi: 10.1128/AEM.00841-07.

Wicker, E. *et al.* (2009) 'Epidemiological evidence for the emergence of a new pathogenic variant of ralstonia solanacearum in martinique (French West Indies)', *Plant Pathology*, 58(5), pp. 853–861. doi: 10.1111/j.1365-3059.2009.02098.x.

Wicker, E. *et al.* (2012) 'Contrasting recombination patterns and demographic histories of the plant pathogen Ralstonia solanacearum inferred from MLSA', *ISME Journal*, 6(5), pp. 961–974. doi: 10.1038/ismej.2011.160.

Wion, D. and Casadesús, J. (2006) 'N6-methyl-adenine: an epigenetic signal for DNA-protein interactions.’, *Nature reviews. Microbiology*, 4(3), pp. 183–192. doi: 10.1038/nrmicro1350.

Van Der Woude, M., Braaten, B. and Low, D. (1996) ‘Epigenetic phase variation of the pap operon in Escherichia coli’, *Trends in Microbiology*. doi: 10.1016/0966-842X(96)81498-3.

Van Der Woude, M. W. and Bäumler, A. J. (2004) ‘Phase and antigenic variation in bacteria’, *Clinical Microbiology Reviews*, 17(3), pp. 581–611. doi: 10.1128/CMR.17.3.581-611.2004.

Wu, D. *et al.* (2015) ‘Oleanolic acid induces the Type III secretion system of ralstonia solanacearum’, *Frontiers in Microbiology*. Frontiers Media S.A., 6(DEC). doi: 10.3389/fmicb.2015.01466.

Yin, J. C. P., Krebs, M. P. and Reznikoff, W. S. (1988) ‘Effect of dam methylation on Tn5 transposition’, *Journal of Molecular Biology*. J Mol Biol, 199(1), pp. 35–45. doi: 10.1016/0022-2836(88)90377-4.

Yoshimochi, T. *et al.* (2009) ‘The global virulence regulator PhcA negatively controls the Ralstonia solanacearum hrp regulatory cascade by repressing expression of the PrhIR signaling proteins’, *Journal of Bacteriology*. American Society for Microbiology Journals, 191(10), pp. 3424–3428. doi: 10.1128/JB.01113-08.

Yuliar, Asi Nion, Y. and Toyota, K. (2015) ‘Recent trends in control methods for bacterial wilt diseases caused by Ralstonia solanacearum’, *Microbes and Environments*, 30(1), pp. 1–11. doi: 10.1264/jsme2.ME14144.

Zhang, Y. *et al.* (2019) ‘Involvement of a PadR regulator PrhP on virulence of Ralstonia solanacearum by controlling detoxification of phenolic acids and type III secretion system’, *Molecular Plant Pathology*, 20(11), pp. 1477–1490. doi: 10.1111/mpp.12854.

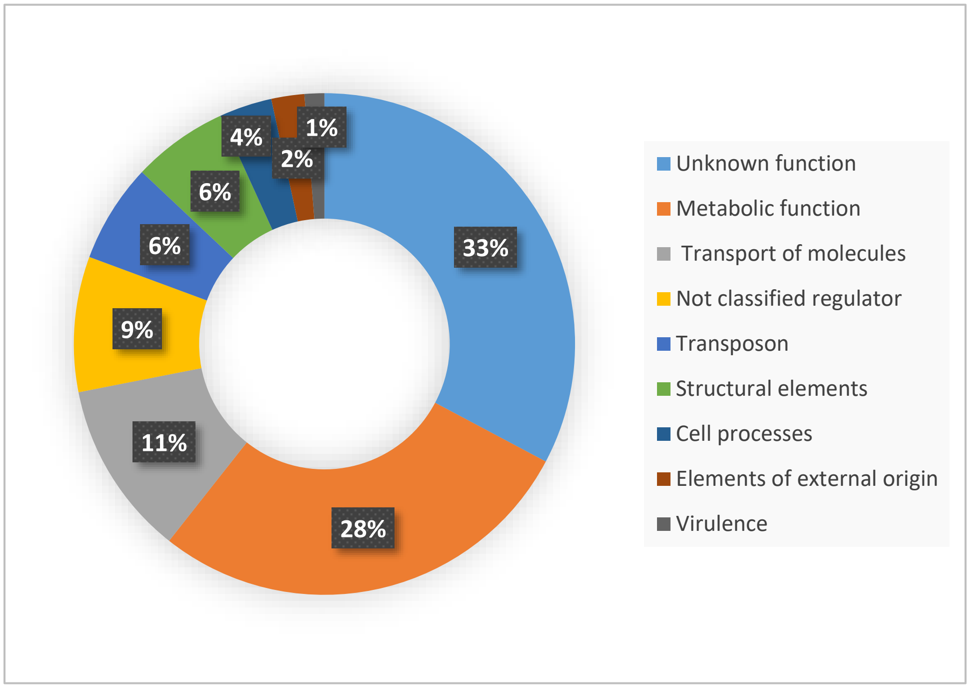
Annexure

RSc0001	RSc0178	RSc0523	RSc0801	RSc1104	RSc1341	RSc1667	RSc1934	RSc2190	RSc2502
RSc0003	RSc0204	RSc0525	RSc0811	RSc1112	RSc1353	RSc1684	RSc1936	RSc2200	RSc2504
RSc0010	RSc0214	RSc0527	RSc0815	RSc1116	RSc1357	RSc1685	RSc1938	RSc2223	RSc2510
RSc0011	RSc0215	RSc0529	RSc0823	RSc1130	RSc1363	RSc1688	RSc1942	RSc2226	RSc2519
RSc0013	RSc0216	RSc0533	RSc0825	RSc1143	RSc1365	RSc1702	RSc1943	RSc2232	RSc2521
RSc0018	RSc0217	RSc0540	RSc0828	RSc1144	RSc1391	RSc1704	RSc1944	RSc2234	RSc2534
RSc0021	RSc0223	RSc0541	RSc0839	RSc1150	RSc1397	RSc1705	RSc1948	RSc2239	RSc2540
RSc0022	RSc0231	RSc0555	RSc0840	RSc1156	RSc1407	RSc1718	RSc1950	RSc2244	RSc2553
RSc0023	RSc0235	RSc0560	RSc0841	RSc1157	RSc1415	RSc1720	RSc1952	RSc2257	RSc2559
RSc0024	RSc0256	RSc0570	RSc0849	RSc1158	RSc1418	RSc1723	RSc1962	RSc2268	RSc2563
RSc0025	RSc0271	RSc0582	RSc0883	RSc1162	RSc1419	RSc1727	RSc1964	RSc2274	RSc2569
RSc0029	RSc0290	RSc0589	RSc0884	RSc1164	RSc1427	RSc1728	RSc1965	RSc2279	RSc2577
RSc0030	RSc0291	RSc0592	RSc0893	RSc1168	RSc1428	RSc1736	RSc1966	RSc2280	RSc2584
RSc0032	RSc0295	RSc0594	RSc0894	RSc1173	RSc1435	RSc1746	RSc1969	RSc2285	RSc2598
RSc0048	RSc0311	RSc0595	RSc0900	RSc1175	RSc1446	RSc1748	RSc1973	RSc2299	RSc2600
RSc0049	RSc0334	RSc0596	RSc0908	RSc1182	RSc1470	RSc1749	RSc1977	RSc2300	RSc2602
RSc0050	RSc0335	RSc0600	RSc0910	RSc1202	RSc1475	RSc1756	RSc1978	RSc2320	RSc2607
RSc0069	RSc0336	RSc0613	RSc0920	RSc1217	RSc1479	RSc1759	RSc1982	RSc2322	RSc2612
RSc0080	RSc0337	RSc0616	RSc0921	RSc1220	RSc1487	RSc1760	RSc1998	RSc2332	RSc2646
RSc0085	RSc0341	RSc0617	RSc0922	RSc1221	RSc1490	RSc1768	RSc2013	RSc2339	RSc2652
RSc0087	RSc0343	RSc0618	RSc0926	RSc1235	RSc1492	RSc1771	RSc2024	RSc2341	RSc2662
RSc0092	RSc0349	RSc0632	RSc0930	RSc1236	RSc1493	RSc1779	RSc2025	RSc2349	RSc2689
RSc0093	RSc0353	RSc0643	RSc0931	RSc1254	RSc1498	RSc1781	RSc2026	RSc2355	RSc2697
RSc0095	RSc0355	RSc0646	RSc0939	RSc1255	RSc1529	RSc1802	RSc2027	RSc2358	RSc2699
RSc0105	RSc0377	RSc0653	RSc0952	RSc1258	RSc1535	RSc1804	RSc2032	RSc2360	RSc2703
RSc0106	RSc0380	RSc0660	RSc0953	RSc1262	RSc1537	RSc1805	RSc2033	RSc2366	RSc2706
RSc0107	RSc0383	RSc0665	RSc0963	RSc1263	RSc1540	RSc1806	RSc2034	RSc2368	RSc2714
RSc0108	RSc0385	RSc0667	RSc0968	RSc1266	RSc1542	RSc1807	RSc2044	RSc2370	RSc2717
RSc0109	RSc0392	RSc0676	RSc0971	RSc1267	RSc1543	RSc1808	RSc2055	RSc2385	RSc2729
RSc0111	RSc0396	RSc0686	RSc0974	RSc1269	RSc1549	RSc1809	RSc2080	RSc2388	RSc2741
RSc0113	RSc0404	RSc0689	RSc0980	RSc1270	RSc1560	RSc1810	RSc2082	RSc2401	RSc2743
RSc0114	RSc0407	RSc0690	RSc0984	RSc1272	RSc1570	RSc1811	RSc2084	RSc2404	RSc2747
RSc0117	RSc0412	RSc0695	RSc0986	RSc1276	RSc1571	RSc1812	RSc2088	RSc2411	RSc2752
RSc0118	RSc0415	RSc0697	RSc0987	RSc1277	RSc1572	RSc1813	RSc2107	RSc2416	RSc2762
RSc0119	RSc0416	RSc0729	RSc0988	RSc1287	RSc1578	RSc1816	RSc2108	RSc2417	RSc2764
RSc0122	RSc0421	RSc0740	RSc0992	RSc1299	RSc1585	RSc1817	RSc2115	RSc2418	RSc2765
RSc0126	RSc0422	RSc0744	RSc0993	RSc1300	RSc1597	RSc1818	RSc2117	RSc2422	RSc2767
RSc0132	RSc0438	RSc0747	RSc1003	RSc1303	RSc1598	RSc1819	RSc2118	RSc2424	RSc2768
RSc0133	RSc0455	RSc0749	RSc1027	RSc1304	RSc1600	RSc1832	RSc2125	RSc2425	RSc2775
RSc0146	RSc0457	RSc0751	RSc1029	RSc1305	RSc1603	RSc1833	RSc2127	RSc2435	RSc2784
RSc0152	RSc0458	RSc0752	RSc1035	RSc1313	RSc1618	RSc1836	RSc2132	RSc2436	RSc2785
RSc0158	RSc0461	RSc0756	RSc1048	RSc1314	RSc1620	RSc1837	RSc2138	RSc2444	RSc2792
RSc0159	RSc0474	RSc0757	RSc1052	RSc1315	RSc1622	RSc1855	RSc2146	RSc2451	RSc2793
RSc0160	RSc0475	RSc0760	RSc1053	RSc1316	RSc1625	RSc1876	RSc2148	RSc2455	RSc2795
RSc0162	RSc0476	RSc0779	RSc1059	RSc1331	RSc1627	RSc1879	RSc2149	RSc2464	RSc2799
RSc0163	RSc0477	RSc0789	RSc1062	RSc1336	RSc1634	RSc1893	RSc2153	RSc2465	RSc2805
RSc0164	RSc0494	RSc0798	RSc1072	RSc1337	RSc1636	RSc1894	RSc2170	RSc2469	RSc2813
RSc0165	RSc0502	RSc0799	RSc1081	RSc1338	RSc1648	RSc1896	RSc2177	RSc2481	RSc2817
RSc0166	RSc0510	RSc0800	RSc1087	RSc1339	RSc1653	RSc1908	RSc2178	RSc2494	RSc2822
	RSc0511		RSc1097	RSc1340	RSc1658	RSc1909	RSc2185	RSc2500	RSc2841

Annexure 1 Genes regulated by the efpR regulon obtained from Perrier et al., 2018 and Capela et al., 2017.

RSc2845	RSc3048	RSc3300	RSp0134	RSp0339	RSp0423	RSp0646	RSp0858	RSp1058	RSp1236	RSp1488
RSc2847	RSc3058	RSc3309	RSp0137	RSp0340	RSp0424	RSp0651	RSp0861	RSp1059	RSp1240	RSp1503
RSc2855	RSc3070	RSc3311	RSp0139	RSp0341	RSp0425	RSp0652	RSp0865	RSp1060	RSp1241	RSp1511
RSc2858	RSc3071	RSc3315	RSp0152	RSp0342	RSp0427	RSp0658	RSp0866	RSp1066	RSp1250	RSp1513
RSc2861	RSc3072	RSc3324	RSp0154	RSp0343	RSp0430	RSp0676	RSp0873	RSp1069	RSp1252	RSp1515
RSc2864	RSc3086	RSc3340	RSp0156	RSp0344	RSp0433	RSp0679	RSp0874	RSp1070	RSp1255	RSp1526
RSc2881	RSc3090	RSc3341	RSp0157	RSp0345	RSp0437	RSp0686	RSp0876	RSp1072	RSp1257	RSp1527
RSc2883	RSc3091	RSc3343	RSp0161	RSp0346	RSp0442	RSp0691	RSp0877	RSp1074	RSp1262	RSp1529
RSc2885	RSc3098	RSc3344	RSp0164	RSp0347	RSp0443	RSp0693	RSp0879	RSp1077	RSp1268	RSp1530
RSc2886	RSc3104	RSc3349	RSp0172	RSp0348	RSp0449	RSp0694	RSp0897	RSp1080	RSp1269	RSp1534
RSc2894	RSc3124	RSc3356	RSp0173	RSp0349	RSp0480	RSp0695	RSp0898	RSp1081	RSp1270	RSp1543
RSc2902	RSc3138	RSc3357	RSp0174	RSp0350	RSp0481	RSp0696	RSp0900	RSp1083	RSp1271	RSp1544
RSc2903	RSc3146	RSc3358	RSp0178	RSp0351	RSp0482	RSp0697	RSp0901	RSp1094	RSp1274	RSp1546
RSc2909	RSc3147	RSc3359	RSp0179	RSp0352	RSp0483	RSp0698	RSp0902	RSp1098	RSp1280	RSp1555
RSc2910	RSc3148	RSc3373	RSp0181	RSp0355	RSp0484	RSp0699	RSp0903	RSp1103	RSp1287	RSp1580
RSc2911	RSc3149	RSc3374	RSp0184	RSp0369	RSp0487	RSp0700	RSp0904	RSp1105	RSp1295	RSp1581
RSc2924	RSc3151	RSc3386	RSp0188	RSp0371	RSp0491	RSp0701	RSp0906	RSp1107	RSp1362	RSp1589
RSc2931	RSc3161	RSc3398	RSp0197	RSp0372	RSp0492	RSp0708	RSp0923	RSp1108	RSp1366	RSp1592
RSc2933	RSc3175	RSc3399	RSp0198	RSp0374	RSp0497	RSp0712	RSp0941	RSp1111	RSp1367	RSp1604
RSc2942	RSc3186	RSc3411	RSp0206	RSp0375	RSp0501	RSp0724	RSp0942	RSp1112	RSp1375	RSp1606
RSc2943	RSc3187	RSc3412	RSp0207	RSp0376	RSp0502	RSp0726	RSp0957	RSp1113	RSp1376	RSp1607
RSc2954	RSc3189	RSc3416	RSp0208	RSp0377	RSp0506	RSp0728	RSp0964	RSp1114	RSp1383	RSp1608
RSc2963	RSc3190	RSc3420	RSp0211	RSp0378	RSp0507	RSp0729	RSp0983	RSp1120	RSp1390	RSp1615
RSc2966	RSc3196	RSc3422	RSp0212	RSp0379	RSp0516	RSp0730	RSp1003	RSp1121	RSp1391	RSp1628
RSc2967	RSc3197	RSc3434	RSp0213	RSp0382	RSp0517	RSp0732	RSp1004	RSp1122	RSp1392	RSp1630
RSc2970	RSc3198	RSp0002	RSp0215	RSp0383	RSp0518	RSp0733	RSp1005	RSp1123	RSp1393	RSp1644
RSc2971	RSc3205	RSp0005	RSp0221	RSp0384	RSp0526	RSp0734	RSp1006	RSp1124	RSp1401	RSp1652
RSc2972	RSc3215	RSp0021	RSp0240	RSp0385	RSp0531	RSp0736	RSp1007	RSp1125	RSp1402	RSp1656
RSc2973	RSc3224	RSp0027	RSp0242	RSp0388	RSp0535	RSp0737	RSp1008	RSp1128	RSp1403	RSp1659
RSc2974	RSc3234	RSp0028	RSp0243	RSp0389	RSp0540	RSp0745	RSp1009	RSp1129	RSp1405	RSp1663
RSc2975	RSc3236	RSp0031	RSp0244	RSp0390	RSp0541	RSp0753	RSp1010	RSp1139	RSp1406	RSp1665
RSc2977	RSc3237	RSp0032	RSp0245	RSp0391	RSp0542	RSp0756	RSp1011	RSp1147	RSp1407	RSp1680
RSc2983	RSc3238	RSp0033	RSp0247	RSp0392	RSp0543	RSp0771	RSp1012	RSp1153	RSp1408	RSp1682
RSc3001	RSc3240	RSp0036	RSp0248	RSp0393	RSp0544	RSp0773	RSp1013	RSp1157	RSp1409	
RSc3002	RSc3248	RSp0053	RSp0252	RSp0394	RSp0545	RSp0775	RSp1014	RSp1159	RSp1410	
RSc3005	RSc3251	RSp0054	RSp0261	RSp0395	RSp0546	RSp0783	RSp1015	RSp1167	RSp1411	
RSc3006	RSc3261	RSp0058	RSp0264	RSp0403	RSp0549	RSp0786	RSp1016	RSp1171	RSp1412	
RSc3010	RSc3264	RSp0065	RSp0275	RSp0404	RSp0552	RSp0798	RSp1017	RSp1172	RSp1413	
RSc3015	RSc3271	RSp0078	RSp0277	RSp0405	RSp0572	RSp0799	RSp1018	RSp1176	RSp1415	
RSc3016	RSc3274	RSp0080	RSp0295	RSp0406	RSp0578	RSp0800	RSp1019	RSp1178	RSp1417	
RSc3018	RSc3278	RSp0081	RSp0297	RSp0410	RSp0587	RSp0819	RSp1020	RSp1197	RSp1418	
RSc3020	RSc3281	RSp0087	RSp0298	RSp0412	RSp0590	RSp0825	RSp1021	RSp1201	RSp1419	
RSc3021	RSc3282	RSp0091	RSp0304	RSp0413	RSp0591	RSp0828	RSp1022	RSp1202	RSp1423	
RSc3028	RSc3291	RSp0107	RSp0306	RSp0414	RSp0598	RSp0833	RSp1033	RSp1209	RSp1424	
RSc3035	RSc3292	RSp0110	RSp0307	RSp0416	RSp0602	RSp0841	RSp1035	RSp1210	RSp1425	
RSc3036	RSc3293	RSp0117	RSp0313	RSp0417	RSp0603	RSp0844	RSp1041	RSp1211	RSp1439	
RSc3039	RSc3294	RSp0122	RSp0318	RSp0418	RSp0605	RSp0849	RSp1043	RSp1214	RSp1449	
RSc3040	RSc3295	RSp0126	RSp0329	RSp0419	RSp0612	RSp0854	RSp1047	RSp1229	RSp1453	
RSc3045	RSc3296	RSp0127	RSp0332	RSp0421	RSp0635	RSp0855	RSp1052	RSp1233	RSp1469	
RSc3047	RSc3299	RSp0129	RSp0335	RSp0422	RSp0639	RSp0856	RSp1056	RSp1234	RSp1470	

Annexure 1 contd.,



Annexure 2 Functional distribution of host specific DEGs in Hawaii evolved clones

Gene ID	Gene	Description
RSc0013		hypothetical protein
RSc0025		hypothetical protein
RSc0048		putative transcription regulator protein
RSc0050		hypothetical protein
RSc0111		sulfur carrier protein ThiS
RSc0114		hypothetical protein
RSc0132		transcription regulator protein
RSc0162		transcription regulator protein
RSc0166		outer membrane channel lipoprotein
RSc0178	glmS	D-fructose-6-phosphate amidotransferase
RSc0215		short chain dehydrogenase
RSc0216		transport transmembrane protein
RSc0217		oxidoreductase transmembrane protein
RSc0223		hypothetical protein
RSc0231		putative (homoserine/homoserine lactone) efflux transmembrane protein
RSc0235		hypothetical protein
RSc0334		hypothetical protein
RSc0336		signal peptide protein
RSc0349	bfd	putative bacterioferritin-associated ferredoxin protein
RSc0396	ipk	4-diphosphocytidyl-2-C-methyl-D-erythritolkinase
RSc0415		hypothetical protein
RSc0455		hypothetical protein
RSc0523		D,D-heptose 1,7-bisphosphate phosphatase
RSc0529		putative metalloprotease
RSc0582		AvrD-related protein
RSc0596		hypothetical protein
RSc0600		hypothetical protein
RSc0613		hypothetical protein
RSc0616		hypothetical protein
RSc0617		putative secreted aspartic protease
RSc0646		hypothetical protein
RSc0740		hypothetical protein
RSc0751		hypothetical protein
RSc0799		putative purine nucleoside permease protein
RSc0931		signal peptide protein
RSc0953		hypothetical protein
RSc0974		hypothetical protein
RSc0980		Chromate transporter protein
RSc0993		putative transcriptional regulatory DNA-binding transcription regulator protein
RSc1003		carboxypeptidase
RSc1048	rpmF	50S ribosomal protein L32
RSc1097	efpR	transcription regulator protein
RSc1164	ppiB	peptidyl-prolyl cis-trans isomerase B
RSc1202		flavoprotein NADH-dependent oxidoreductase
RSc1235		hypothetical protein
RSc1276		cytochrome C oxidase subunit I
RSc1277		cytochrome C oxidase subunit II
RSc1287	pilE1	type 4 fimbrial biogenesis signal peptide protein
RSc1314		lipoprotein transmembrane
RSc1315		putative signal peptide protein
RSc1339	ssuD	alkanesulfonate monooxygenase
RSc1340	ssuC	aliphatic sulfonate ABC transporter transmembrane protein
RSc1357	gala7	Type III effector protein GALA7

Annexure 3 List of common DEGs in efpR and evolved clones (Hawaii, Zebrina and Bean)

The list of genes includes description of the genes with their gene ID.

Gene ID	Gene	Description
RSc1363	aceB	malate synthase
RSc1407	frr	ribosome recycling factor
RSc1419		putative tRNA/rRNA methyltransferase protein
RSc1475	ripM	Type III effector protein RipM
RSc1487		hypothetical protein
RSc1492	tISRso14b	ISRSO14-transposase orfB protein
RSc1498		putative bacteriophage protein
RSc1603	lpdA	dihydrolipoamide dehydrogenase
RSc1720		hypothetical protein
RSc1727		hypothetical protein
RSc1728		hypothetical protein
RSc1736		two-component response regulator transcription regulator protein
RSc1802		putative transcription regulator protein
RSc1950		methyl-accepting chemotaxis transducer transmembrane protein
RSc2025		hypothetical protein
RSc2080		hypothetical protein
RSc2088		hypothetical protein
RSc2107	rsl	L-fucose-binding lectin RSL
RSc2117		transcription regulator protein
RSc2118		putative xanthine permease transmembrane protein
RSc2149		hypothetical protein
RSc2234		hypothetical protein
RSc2244		substrate-binding periplasmic ABC transporter protein
RSc2279		signal peptide protein
RSc2285		putative signal peptide protein
RSc2299		signal peptide protein
RSc2339		lipoprotein transmembrane
RSc2349	ppa	inorganic pyrophosphatase
RSc2366	dcd	deoxycytidine triphosphate deaminase
RSc2370		hypothetical protein
RSc2401		hypothetical protein
RSc2404		putative signal peptide protein
RSc2502		hypothetical protein
RSc2646	hutH	histidine ammonia-lyase
RSc2652		lipoprotein
RSc2662		Mechanosensitive ion channel, MscS family
RSc2743	rlpB	putative lipoprotein B precursor transmembrane
RSc2775	ripW	Type III effector protein RipW (formerly PopW), harpin with pectate lyase domain
RSc2861	fruB	multiphosphoryl transfer protein
RSc2886		hypothetical protein
RSc2902		lipoprotein
RSc2911		hypothetical protein
RSc2954		hypothetical protein
RSc2967		outer membrane signal peptide protein
RSc3001	rpmC	50S ribosomal protein L30
RSc3002	rpsE	30S ribosomal protein S5
RSc3006	rpsN	30S ribosomal protein S14
RSc3010	rspsQ	30S ribosomal protein S17
RSc3015	rpsS	30S ribosomal protein S19
RSc3016	rplB	50S ribosomal protein L2
RSp0502		hypothetical protein
RSp0518		putative VGR-related protein
RSp0531		P-II-like protein transcription regulator
RSp0658	copC	copper resistance C signal peptide protein
RSp0676	metE	5-methyltetrahydropteroyltriglutamate-- homocysteine methyltransferase

Contd., of Annexure 3

Gene ID	Gene	Description
RSc3020	rpsJ	30S ribosomal protein S10
RSc3021	tuf	elongation factor Tu
RSc3035	rplL	50S ribosomal subunit protein L7/L12
RSc3036	rplJ	50S ribosomal protein L10
RSc3039	nusG	transcription antitermination protein NusG
RSc3040	secE	preprotein translocase subunit SecE
RSc3070		glutathione S-transferase protein
RSc3071		hypothetical protein
RSc3148		hypothetical protein
RSc3161		two component response regulator transcription regulator protein
RSc3189		hypothetical protein
RSc3198		hypothetical protein
RSc3215		hypothetical protein
RSc3236		putative bacteriophage-related protein
RSc3261		hypothetical protein
RSc3292		putative O-acyltransferase
RSc3294	gcvH	glycine cleavage system H protein
RSc3315		putative prolin-rich transmembrane protein
RSc3349		hypothetical protein
RSp0028	gala3	Type III effector protein GALA3
RSp0032		salicylyl-CoA 5-hydroxylase
RSp0036		acyl-CoA dehydrogenase oxidoreductase protein
RSp0065	purU2	formyltetrahydrofolate deformylase
RSp0174		methyltransferase protein
RSp0181		hypothetical protein
RSp0206	tIS0121	TIS1021 transposase
RSp0207		putative signal peptide protein
RSp0208		hypothetical protein
RSp0211		hypothetical protein
RSp0212		dioxygenase signal peptide protein
RSp0213	ripBJ	Type III effector protein RipBJ
RSp0240	treS	Trehalose synthase/alpha-amylase
RSp0245	ahpC1	alkyl hydroperoxide reductase subunit C
RSp0275		hydrolase transmembrane protein
RSp0277	treA	periplasmic alpha,alpha-trehalase signal peptide protein
RSp0304		Type III effector protein
RSp0343	flgC	flagellar basal-body rod protein FlgC
RSp0374	fliQ	flagellar biosynthesis protein FliQ
RSp0384	fliS	flagellar protein FliS
RSp0404		hypothetical protein
RSp0406		hypothetical protein
RSp0410		hypothetical protein
RSp0425		hypothetical protein
RSp0480		amino acid ABC transporter transmembrane protein
RSp0481		substrate-binding periplasmic (PBP) ABC transporter protein
RSp0483		ornithine cyclodeaminase
RSp0484		hypothetical protein
RSp0491	czcC	cobalt-zinc-cadmium outer membrane resistance protein
RSp0492	czcB	cobalt-zinc-cadmium resistance protein
RSp0501		hypothetical protein
RSp0502		hypothetical protein
RSp0518		putative VGR-related protein
RSp0531		P-II-like protein transcription regulator
RSp0658	copC	copper resistance C signal peptide protein
RSp0676	metE	5-methyltetrahydropteroyltriglutamate-- homocysteine methyltransferase
RSp0679		putative diaminopimelate decarboxylase protein

Contd., of Annexure 3

Gene ID	Gene	Description
RSp0693	hdfA	Dioxygenase
RSp0694	hdfB	Tryptophan-2,3-dioxygenase, oxidoreductase protein
RSp0695	hdfC	FAD dependent oxidoreductase protein
RSp0697	hdfE	Short-chain dehydrogenase/reductase protein
RSp0698	hdfF	Aminotransferase protein
RSp0699		peptide chain release factor-like protein
RSp0712		hypothetical protein
RSp0724		hypothetical protein
RSp0771		hypothetical protein
RSp0773		putative signal peptide protein
RSp0783		hypothetical protein
RSp0786		hypothetical protein
RSp0825		Spermidine synthase
RSp0841		putative lipoprotein
RSp0855	hrpY	Hrp pilus subunit HRPY protein
RSp0856	hrpX	hypothetical protein
RSp0858	hrpV	hypothetical protein
RSp0861	hrcQ	Hrp conserved protein HRCQ
RSp0865	hrpK	HRPK protein
RSp0866	hrpJ	HRPJ protein
RSp0873	hrpB	regulatory HRPB transcription regulator protein
RSp0876	ripAB	Type III effector protein RipAB (formerly PopB)
RSp0877	popA	Type III effector protein PopA
RSp0879		Type III effector protein
RSp0903		hypothetical protein
RSp0904		putative dehydrogenase oxidoreductase protein
RSp0906		S1/P1 nuclease
RSp0957		hypothetical protein
RSp1014	epsF	EPS I polysaccharide export inner membrane transmembrane protein
RSp1018	epsB	EPS I polysaccharide export transmembrane protein
RSp1047		transporter transmembrane protein
RSp1056		beta alanine--pyruvate transaminase
RSp1058	adc	acetoacetate decarboxylase
RSp1059		3-hydroxybutyrate dehydrogenase
RSp1074		putative hemolysin activating-like protein
RSp1081		signal peptide protein
RSp1094		hypothetical protein
RSp1111		lipase protein
RSp1124		putative NADP-dependent zinc-type alcohol dehydrogenase oxidoreductase protein
RSp1178		composite two-component sensor histidine kinase and response regulator hybrid transcription regulator protein
RSp1402	cheY2	chemotaxis response regulator transcription regulator protein
RSp1409	cheY	putative chemotaxis response regulator transcription regulator protein
RSp1410	motB	flagellar motor protein MotB
RSp1423		putative transmembrane protein
RSp1424	ralD	Ralfuranone biosynthesis protein [aminotransferase]
RSp1425	pvdA	L-ornithine 5-monooxygenase oxidoreductase protein
RSp1439	cysE1	serine acetyltransferase protein
RSp1449	ttuD1	hydroxypyruvate reductase oxidoreductase protein
RSp1511		hypothetical protein
RSp1526		hypothetical protein
RSp1527		hypothetical protein
RSp1530		putative L-ascorbate oxidase (ASCORBASE) oxidoreductase protein
RSp1580		hypothetical protein
RSp1604		hypothetical protein
RSp1606		hypothetical protein
RSp1607		hypothetical protein
RSp1608		hypothetical protein
RSp1659		acyl-carrier-protein
RSp1663		hypothetical protein
RSp1665		hypothetical protein

Contd., of Annexure 3

Collection number	Sample name	Nanodrop readings Conc (ng/µl)	A260/280	A260/230	Qubit readings (ng/µl)
AG32	Zeb26b1	265.6	1.88	2.01	268
AG36	Zeb26b5	320.9	1.89	2.06	312
AG38	Zeb26c2	352.6	1.91	2.14	470
AG39	Zeb26c3	195.5	1.90	2.06	66.6
AG40	Zeb26c4	55.7	1.81	2.04	58.4
AG42	Zeb26d1	115.6	1.94	2.21	98.2
AG47	Zeb26e1	136.6	1.87	2.09	384
AG49	Zeb26e3	148.3	1.95	2.05	150

Annexure 4 Concentration of DNA of Zebrina evolved clones

Collection number	Sample name	Nanodrop readings Conc (ng/µl)	A260/280	A260/230	Qubit readings (ng/µl)
AG127	Haw35a1	75.3	1.90	2.12	88.6
AG130	Haw35a4	168.9	1.96	2.08	197
AG132	Haw35b1	246.6	1.98	2.12	174
AG135	Haw35b4	317	1.90	2.05	332
AG137	Haw35c1	260.8	1.93	2.10	346
AG138	Haw35c2	301.7	1.92	2.12	398
AG144	Haw35d3	370.2	1.89	2.02	372
AG146	Haw35d5	424.1	1.89	2.02	788
AG147	Haw35e1	458.8	1.87	2.01	528
AG149	Haw35e3	355.5	1.91	2.04	606

Annexure 5 Concentration of DNA of Hawaii evolved clones

Collection number	Sample name	Nanodrop readings Conc (ng/µl)	A260/280	A260/230	Qubit readings (ng/µl)
AG80	Bean26a4	326.4	1.92	2.04	274
AG81	Bean26a5	225.4	1.93	2.13	332
AG82	Bean26b1	312.6	1.97	2.11	262
AG84	Bean26b3	458.1	1.87	2.08	55.8
AG85	Bean26b4	167.4	1.92	2.14	182
AG87	Bean26c1	150.5	1.88	2.01	172
AG1	GMI1000	384.8	1.88	2.01	578

Annexure 6 Concentration of DNA of Bean evolved clones and ancestral clone GMI1000

Collection number	Sample name	Replicate	Agilent readings (ng/µl)	RNA area	Qubit readings (ng/µl)
AG127	Haw35a1	A	168	148.9	89
		B	97	85.9	80.4
		C	159	140.9	85
AG130	Haw35a4	A	139	123.9	84.2
		B	160	142.6	91.6
		C	126	111.9	82.4
AG132	Haw35b1	A	132	117.6	77
		B	169	150.5	86.8
		C	228	203	104
AG135	Haw35b4	A	168	149.5	93
		B	162	79.6	99
		C	195	173.2	95.6
AG137	Haw35c1	A	175	55.3	96.2
		B	180	56.6	84
		C	168	53	87.8
AG138	Haw35c2	A	151	47.6	78.4
		B	126	39.7	72
		C	162	51.1	83.8
AG144	Haw35d3	A	206	64.8	104
		B	189	59.5	83.8
		C	151	47.7	85.2
AG146	Haw35d5	A	218	68.8	100
		B	209	65.7	95.6
		C	172	54.1	89.2
AG147	Haw35e1	A	121	59.9	79.2
		B	137	67.5	80.4
		C	255	262.8	106
AG149	Haw35e3	A	171	84.1	86.8
		B	175	86.2	70.4
		C	139	68.7	66.6

Annexure 7 Concentration of RNA of Hawaii evolved clones for RNAseq analyses

Collection number	Sample name	Replicate	Agilent readings (ng/µl)	RNA area	Qubit readings (ng/µl)
AG32	Zeb26b1	A	190	93.6	79.4
		B	113	116.7	88.2
		C	208	102.7	84.8
AG36	Zeb26b5	A	211	103.9	93.6
		B	195	96.1	90.8
		C	179	114.1	80.2
AG38	Zeb26c2	A	260	165.8	95
		B	110	70.3	46.4
		C	117	120.7	46.2
AG39	Zeb26c3	A	217	138.5	87.6
		B	394	251.3	76.2
		C	243	155.1	93.2
AG40	Zeb26c4	A	254	162.1	97.6
		B	73	37.8	42.6
		C	226	144	87
AG42	Zeb26d1	A	170	108	75.4
		B	194	123.6	75.8
		C	125	68.7	67
AG47	Zeb26e1	A	217	119.9	93.2
		B	172	95.1	85.8
		C	114	62.6	68
AG49	Zeb26e3	A	145	80.2	76.2
		B	117	64.6	69
		C	201	211.6	92.6

Annexure 8 Concentration of RNA of Zebrina evolved clones for RNAseq analyses

Collection number	Sample name	Replicate	Agilent readings (ng/µl)	Agilent readings RNA area	Qubit readings (ng/µl)
AG80	Bean26a4	A	182	191.5	87.6
		B	180	189.2	81
		C	212	223	99
AG81	Bean26a5	A	215	227	97
		B	182	191.5	88
		C	189	199.5	90
AG82	Bean26b1	A	245	257.9	99
		B	124	130.3	67.6
		C	177	186.2	86
AG84	Bean26b3	A	132	138.9	68.8
		B	127	130.5	71
		C	178	183.8	84.8
AG85	Bean26b4	A	112	115.5	62.8
		B	190	97.7	88.4
		C	124	63.7	74.2
AG87	Bean26c1	A	212	218.6	91.8
		B	106	109	61.6
		C	229	105.7	73.6
AG1	GMI1000	A	221	196.5	96.8
		B	179	159	79.4
		C	192	170.8	83.6

Annexure 9 Concentration of RNA of Bean evolved clones and ancestral clone GMI1000 for RNAseq analyses

									GMI1000.sequelG2				AG127 - Haw a1			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					55	146	2.04	0.43
AL646052	*117936*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*57*	*2.09*		*74*	*99*	*3.57*	*0.64*
AL646052	*117939*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*52*	*2.26*		*93*	*101*	*3.79*	*0.8*
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		79	126	2.55	0.57
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					130	143	3.08	0.78
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		73	143	2.26	0.5
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		74	145	2.32	0.48
AL646052	1821402	CTGCGGATGGGTCTATGTTAACTATTTCGCGATAGCGTTTT	m6A	+	GTWWAC	A	RSc1705 (hypothetical protein)	differential	48	68	2.18	0.65	57	102	2.13	
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					124	150	2.85	0.76
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tIsrso5,RSc2176 (ISRSO5-transposase protein)	differential					84	125	2.99	0.58
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52	46	205	1.74	
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		80	126	2.29	0.56
AL646052	3660528	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	39	82	2.21		82	208	2.43	0.48
AL646052	3660531	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	40	74	2.14		81	196	2.11	0.46
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		104	170	2.27	0.7
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					80	149	2.39	0.52
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		87	171	2.25	0.59
AL646053	505523	TGACAGAGCGATGTCAGTAAACCAACAAAATTCCACTTAAA	m6A	-	GTWWAC	A	upstream RSp0399 (hypothetical protein)	differential	40	59	2.04	0.75	39	118	1.76	
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		67	133	2.23	0.48
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					101	144	2.57	0.63
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		69	146	2.25	0.47
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*49*	*22*	*6.35*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*38*	*18*	*4.28*	*1*
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		120	90	4.53	1
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		126	129	3.56	0.77
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					29	104	1.56	0.28
AL646053	1939052	ATTGAAATGACGGAGAGTAAACTTCAAAAGAGGGCGTTGGT	m6A	-	GTWWAC	A	RSp1544 (hypothetical protein)	differential					111	135	3.38	0.68
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		107	166	2.55	0.61
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	96	2.06		101	205	2.3	0.54
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		89	172	2.5	0.59
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		88	151	2.26	0.54

*Annexure 10 Differential methylation positions of G**T**WWA̲C motif raw data between Hawaii evolved clone (Haw a1) and GMI1000*

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

seqid	position	context	type	strand	motif	modified base	feature	status	GMI1000.sequelG2 score	GMI1000.sequelG2 coverage	GMI1000.sequelG2 IPDRatio	GMI1000.sequelG2 fraction	AG130 - Haw a4 score	AG130 - Haw a4 coverage	AG130 - Haw a4 IPDRatio	AG130 - Haw a4 fraction
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					51	173	1.85	0.4
AL646052	117939	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	52	2.26		76	122	2.58	0.64
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					95	182	2.28	0.59
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		59	175	2.01	0.38
AL646052	683379	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	64	70	3	0.76	37	172	1.7	
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		133	161	2.86	0.71
AL646052	1821402	CTGCGGATGGGTCTATGTTAACTATTTCGCGATAGCGTTTT	m6A	+	GTWWAC	A	RSc1705 (hypothetical protein)	differential	48	68	2.18	0.65	41	155	1.62	
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					151	155	3.11	0.86
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					100	146	2.47	0.62
AL646052	2741386	AACATATGACGTTGACGTAAACGTCAATTACCTCGAAGACC	m6A	+	GTWWAC	A	upstream RSc2534 (3-hydroxyacyl-CoA dehydrogenase type II oxidoreductase protein)	differential	64	117	2.27	0.49	33	173	1.6	
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					162	284	2.47	0.63
AL646052	2813723	TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA	m6A	+	GTWWAC	A	RSc2612 (hypothetical protein)	differential					65	268	1.72	0.37
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52	37	145	1.75	
AL646052	3435906	CCCTTCTCCCGCCACAGTTAACCCTGCATTATCGGCGCCAA	m6A	+	GTWWAC	A	RSc3178 (putative ATP binding protein)	differential	34	100	1.76		78	131	2.65	0.6
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		174	199	3.01	0.8
AL646052	3660528	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	39	82	2.21		100	230	2.28	0.57
AL646053	87279	TGGCACGTGACAAATTGTTAACCATGGTTTACAGCGCAGAC	m6A	-	GTWWAC	A	upstream RSp0078 (hypothetical protein)	differential	58	93	2.42	0.57	35	262	1.57	
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		135	182	2.52	0.82
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					116	227	2.27	0.55
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		86	154	2.19	0.62
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		75	164	2.18	0.47
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					60	135	2.1	0.41
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		121	87	3.89	1
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		145	171	2.87	0.74
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					35	209	1.53	0.27
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		78	126	2.21	0.55
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		163	259	2.78	0.65
AL646053	2087332	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential					84	184	2.2	0.48
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		116	177	2.56	0.73
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		112	177	2.51	0.6

*Annexure 11 Differential methylation positions of G**T**WWAC motif raw data between Hawaii evolved clone (Haw a4) and GMI1000*

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG132 - Haw b1			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94117	GAAACAAACGCGTCAAGTTAACGCCGGCGCCGCAGCACACG	m6A	-	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential	41	89	2.08	0.48	53	235	1.83	
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					52	221	1.79	0.37
AL646052	*117936*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*57*	*2.09*		*103*	*158*	*2.72*	*0.6*
AL646052	*117939*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*52*	*2.26*		*108*	*155*	*3.01*	*0.74*
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		89	241	2.12	0.45
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					96	173	2.36	0.6
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		105	227	2.38	0.51
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		146	212	2.83	0.66
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					128	170	2.73	0.71
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					157	187	3.2	0.74
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					127	185	2.75	0.65
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					69	193	1.99	0.42
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		183	217	2.96	0.79
AL646052	*3660528*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*39*	*82*	*2.21*		*147*	*218*	*3.11*	*0.78*
AL646052	*3660531*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*40*	*74*	*2.14*		*101*	*207*	*2.32*	*0.52*
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		138	208	2.47	0.77
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					70	217	1.93	0.4
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		131	215	2.47	0.7
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		98	161	2.5	0.57
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					106	165	2.55	0.61
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		107	176	2.6	0.58
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*71*	*44*	*4.54*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*91*	*43*	*4.37*	*1*
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		149	135	3.49	0.91
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		123	176	2.76	0.62
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					102	195	2.23	0.59
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		122	193	2.35	0.62
AL646053	2062924	CCCAAGATTAACGAACGTTAACAAACGCCAGCGGGAGTGCT	m6A	-	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	90	97	2.84	0.76	66	256	1.96	
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		68	160	2.3	0.51
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		89	139	2.42	0.58

*Annexure 12 Differential methylation positions of G**T**WWA̲C motif raw data between Hawaii evolved clone (Haw b1) and GMI1000*

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG135 - Haw b4			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94117	GAAACAAACGCGTCAAGTTAACGCCGGCGCCGCAGCACACG	m6A	-	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential	41	89	2.08	0.48	47	184	1.82	
AL646052	*117936*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT*	*m6A*	-	*GTWWAC*	*A*	*upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*57*	*2.09*		*103*	*152*	*2.58*	*0.61*
AL646052	*117939*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	+	*GTWWAC*	*A*	*upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*52*	*2.26*		*157*	*156*	*3.93*	*0.98*
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		83	140	2.45	0.55
AL646052	127850	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream thiG,RSc0109 (thiazole synthase)	differential	52	88	2.08	0.58	42	137	1.78	
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					129	162	2.82	0.76
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		112	153	2.72	0.66
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		150	152	3.03	0.84
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					96	145	2.53	0.63
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					98	118	2.51	0.69
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					68	117	2.24	0.52
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52	55	155	1.93	
AL646052	3004675	GTGCATCTGTGTGGCTGTAAACCGGCATGAACATTGGGTTT	m6A	-	GTWWAC	A	upstream RSc2795 (hypothetical protein)	differential	51	92	2.08	0.54	45	141	1.76	
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		61	119	2.09	0.5
AL646052	3660528	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	39	82	2.21		98	178	2.47	0.64
AL646053	87279	TGGCACGTGACAAATTGTTAACCATGGTTTACAGCGCAGAC	m6A	-	GTWWAC	A	upstream RSp0078 (hypothetical protein)	differential	58	93	2.42	0.57				
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		116	174	2.37	0.74
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					74	175	2.07	0.45
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		129	156	2.66	0.85
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		63	111	2.27	0.52
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					62	151	2.07	0.4
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		63	121	2.38	0.49
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		139	99	3.66	0.97
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		125	140	2.92	0.73
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					56	132	1.88	0.46
AL646053	1939052	ATTGAAATGACGGAGAGTAAACTTCAAAAGAGGGCGTTGGT	m6A	-	GTWWAC	A	RSp1544 (hypothetical protein)	differential					49	134	1.93	0.37
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		69	131	2.09	0.49
AL646053	2062924	CCCAAGATTAACGAACGTTAACAAACGCCAGCGGGAGTGCT	m6A	-	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	90	97	2.84	0.76				
AL646053	2087332	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential					79	165	2.16	0.49
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		85	162	2.41	0.6
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		88	147	2.32	0.56

Annexure 13 Differential methylation positions of GTWWAC motif raw data between Hawaii evolved clone (Haw b4) and GMI1000

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG137 – Haw c1			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94117	GAAACAAACGCGTCAAGTTAACGCCGGCGCCGCAGCACACG	m6A	-	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential	41	89	2.08	0.48				
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					47	93	2.16	0.51
AL646052	655714	AAAGCGCTTCTCATTTGTTAACGATTTGCAGGGTTCAACCA	m6A	-	GTWWAC	A	upstream RSc0608 (ripAA)	differential	56	68	2.63	0.6	32	106	1.72	
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					94	105	3.41	0.74
AL646052	683379	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	64	70	3	0.76	38	93	2.02	
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		68	85	2.66	0.64
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					127	116	3.13	0.91
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					59	75	2.43	0.6
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					102	148	2.66	0.62
AL646052	2813723	TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA	m6A	+	GTWWAC	A	RSc2612 (hypothetical protein)	differential					56	134	1.94	0.46
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52				
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		95	103	2.79	0.74
AL646052	3660528	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	39	82	2.21		48	128	1.88	0.46
AL646052	3660531	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	40	74	2.14		69	121	2.24	0.53
AL646053	87279	TGGCACGTGACAAATTGTTAACCATGGTTTACAGCGCAGAC	m6A	-	GTWWAC	A	upstream RSp0078 (hypothetical protein)	differential	58	93	2.42	0.57				
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		53	78	2.07	0.62
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		50	60	2.24	0.68
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		70	100	2.85	0.57
AL646053	792207	CCACGGCTGGGTGGACGTAAACAGCCGGGCGACCTGGCGGT	m6A	-	GTWWAC	A	RSp0641 (peptide synthetase protein)	differential	137	89	4.43	1	36	50	2.07	
AL646053	1452556	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	37	76	1.87	0.48				
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		125	120	3.39	0.89
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		44	56	2.55	0.56
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					76	134	2.1	0.56
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		52	131	2.05	0.42
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		137	118	3.25	0.9

Annexure 14 Differential methylation positions of GTWWAC motif raw data between Hawaii evolved clone (Haw c1) and GMI1000

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG138- haw c2			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	117939	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	52	2.26		47	161	1.88	0.41
AL646052	127850	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream thiG,RSc0109 (thiazole synthase)	differential	52	88	2.08	0.58	40	235	1.63	
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					162	221	3.15	0.67
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		98	244	2.25	0.46
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		125	228	2.51	0.57
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					164	161	3.24	0.86
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					85	173	2.43	0.49
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					123	160	2.95	0.69
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					117	236	2.39	0.53
AL646052	2813723	TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA	m6A	+	GTWWAC	A	RSc2612 (hypothetical protein)	differential					87	225	1.96	0.46
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		125	153	2.7	0.71
AL646052	3660528	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	39	82	2.21		57	240	1.86	0.41
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		70	173	1.96	0.5
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		90	251	2.17	0.42
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					59	138	2.02	0.41
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		91	152	2.74	0.56
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		94	86	3.47	0.86
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		140	187	3.19	0.66
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					36	172	1.58	0.31
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		121	159	2.84	0.71
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		114	168	2.7	0.72
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		81	160	2.36	0.5

Annexure 15 Differential methylation positions of GTWWAC motif raw data between Hawaii evolved clone (Haw c2) and GMI1000

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG144 - Haw d3			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94117	GAAACAAACGCGTCAAGTTAACGCCGGCGCCGCAGCACACG	m6A	-	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential	41	89	2.08	0.48	38	102	1.96	
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					40	103	1.97	0.45
AL646052	117939	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	52	2.26		59	130	2.34	0.5
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		91	154	2.46	0.55
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					56	86	2.38	0.58
AL646052	683379	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	64	70	3	0.76				
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		42	67	2.13	0.5
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					61	73	2.6	0.67
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					50	73	2.28	0.54
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					45	110	1.98	0.4
AL646052	2813723	TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA	m6A	+	GTWWAC	A	RSc2612 (hypothetical protein)	differential					67	113	2.17	0.51
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52	38	83	1.98	
AL646052	3435906	CCCTTCTCCCGCCACAGTTAACCCTGCATTATCGGCGCCAA	m6A	+	GTWWAC	A	RSc3178 (putative ATP binding protein)	differential	34	100	1.76		63	157	2.23	0.47
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		62	104	2.09	0.5
AL646052	3660517	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	+	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	79	75	2.91	0.77	36	102	1.79	
AL646052	3660528	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	39	82	2.21		71	105	2.64	0.66
AL646052	3660531	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	40	74	2.14		55	101	2.47	0.48
AL646053	87279	TGGCACGTGACAAATTGTTAACCATGGTTTACAGCGCAGAC	m6A	-	GTWWAC	A	upstream RSp0078 (hypothetical protein)	differential	58	93	2.42	0.57	38	109	1.99	
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		71	106	2.23	0.66
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		52	67	2.41	0.7
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		97	103	3.02	0.77
AL646053	1427525	CCACAAAAATGGAGAGGTAAACAGGAAATGCTTCATGGTGA	m6A	+	GTWWAC	A	RSp1132 (hypothetical protein)	differential	64	70	2.91	0.67	50	100	2.03	
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		44	71	2.37	0.51
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		83	50	3.84	1
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		73	81	3.5	0.69
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					53	111	1.98	0.47
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		95	107	2.85	0.72
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		50	98	2.13	0.45

Annexure 16 Differential methylation positions of GTWWAC motif raw data between Hawaii evolved clone (Haw d3) and GMI1000

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

seqid	position	context	type	strand	motif	modified base	feature	status	GMI1000.sequelG2 score	GMI1000.sequelG2 coverage	GMI1000.sequelG2 IPDRatio	GMI1000.sequelG2 fraction	AG146 - Haw d5 score	AG146 - Haw d5 coverage	AG146 - Haw d5 IPDRatio	AG146 - Haw d5 fraction
AL646052	94117	GAAACAAACGCGTCAAGTTAACGCCGGCGCCGCAGCACACG	m6A	-	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential	41	89	2.08	0.48				
AL646052	655714	AAAGCGCTTCTCATTTGTTAACGATTTGCAGGGTTCAACCA	m6A	-	GTWWAC	A	upstream RSc0608 (ripAA)	differential	56	68	2.63	0.6				
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		73	80	3	0.7
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					82	105	2.68	0.67
AL646052	2314079	TCCTCTCATCCGCACGGTAAACCCCCTCTTTGTACTGCCCC	m6A	+	GTWWAC	A	RSc2132 (Type III effector protein)	differential	47	83	2.15	0.53	33	66	1.85	
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					55	88	2.22	0.52
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					64	76	2.72	0.65
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					91	111	2.85	0.69
AL646052	2813723	TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA	m6A	+	GTWWAC	A	RSc2612 (hypothetical protein)	differential					58	103	2.01	0.49
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52				
AL646052	3435906	CCCTTCTCCCGCCACAGTTAACCCTGCATTATCGGCGCCAA	m6A	+	GTWWAC	A	RSc3178 (putative ATP binding protein)	differential	34	100	1.76		75	140	2.15	0.59
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		65	68	2.62	0.71
AL646052	3660528	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	39	82	2.21		82	141	2.36	0.62
AL646053	87279	TGGCACGTGACAAATTGTTAACCATGGTTTACAGCGCAGAC	m6A	-	GTWWAC	A	upstream RSp0078 (hypothetical protein)	differential	58	93	2.42	0.57	37	95	1.94	
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		55	65	2.39	0.72
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					71	110	2.52	0.56
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		57	87	2.64	0.54
AL646053	743015	TTCCGTTGTAGTACGAGTAAACGGAATCCGATGCATGCAGC	m6A	+	GTWWAC	A	RSp0611 (hypothetical protein)	differential	65	69	2.61	0.67	48	84	2.1	
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		51	30	3.9	1
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		62	85	2.3	0.56
AL646053	2062924	CCCAAGATTAACGAACGTTAACAAACGCCAGCGGGAGTGCT	m6A	-	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	90	97	2.84	0.76	50	67	2.63	
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		68	105	2.54	0.62
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		52	96	2.16	0.48

*Annexure 17 Differential methylation positions of G**T**WWAC motif raw data between Hawaii evolved clone (Haw d5) and GMI1000*

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

seqid	position	context	type	strand	motif	modified base	feature	status	GMI1000.sequelG2 score	GMI1000.sequelG2 coverage	GMI1000.sequelG2 IPDRatio	GMI1000.sequelG2 fraction	AG147 - Haw e1 score	AG147 - Haw e1 coverage	AG147 - Haw e1 IPDRatio	AG147 - Haw e1 fraction
AL646052	94117	GAAACAAACGCGTCAAGTTAACGCCGGCGCCGCAGCACACG	m6A	-	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential	41	89	2.08	0.48	50	201	1.85	
AL646052	117936	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT	m6A	-	GTWWAC	A	upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	57	2.09		65	121	2.17	0.49
AL646052	117939	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	52	2.26		71	125	2.28	0.62
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		79	156	2.46	0.5
AL646052	655714	AAAGCGCTTCTCATTTGTTAACGATTTGCAGGGTTCAACCA	m6A	-	GTWWAC	A	upstream RSc0608 (ripAA)	differential	56	68	2.63	0.6				
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					95	119	2.76	0.71
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		92	134	2.57	0.62
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					58	88	2.38	0.56
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					80	98	2.57	0.67
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					72	123	2.32	0.54
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		119	112	3.14	0.83
AL646052	3660528	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	39	82	2.21		96	167	2.55	0.64
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		88	144	2.32	0.65
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					68	145	2.11	0.44
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		75	116	2.37	0.57
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*198*	*164*	*4.5*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*219*	*169*	*4.12*	*0.94*
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		146	143	3.02	0.91
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		137	168	2.9	0.71
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					32	119	1.61	0.31
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		82	111	2.4	0.63
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		69	136	2.33	0.5
AL646053	2087332	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential					73	152	2.3	0.47
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		133	167	3.1	0.83
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		82	160	2.17	0.5

*Annexure 18 Differential methylation positions of G**T**WWAC motif raw data between Hawaii evolved clone (Haw e1) and GMI1000*

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG149 - Haw e3			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					49	140	1.89	0.41
AL646052	117939	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	52	2.26		90	105	2.84	0.78
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		63	135	2.21	0.45
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					126	141	3.08	0.79
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		56	116	2.19	0.45
AL646052	683379	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	64	70	3	0.76	37	114	1.8	
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		83	110	3.51	0.62
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					93	106	3.02	0.73
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					76	119	2.33	0.56
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					108	117	2.79	0.75
AL646052	2741386	AACATATGACGTTGACGTAAACGTCAATTACCTCGAAGACC	m6A	+	GTWWAC	A	upstream RSc2534 (3-hydroxyacyl-CoA dehydrogenase type II oxidoreductase protein)	differential	64	117	2.27	0.49	42	121	1.96	
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					78	176	2.16	0.47
AL646052	2813723	TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA	m6A	+	GTWWAC	A	RSc2612 (hypothetical protein)	differential					85	177	2.12	0.49
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		97	147	2.37	0.6
AL646052	*3660528*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*39*	*82*	*2.21*		*62*	*143*	*2.16*	*0.51*
AL646052	*3660531*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*40*	*74*	*2.14*		*98*	*141*	*2.61*	*0.62*
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		87	162	2.14	0.62
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					64	117	2.02	0.49
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					75	108	2.68	0.57
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		75	108	2.77	0.6
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*199*	*157*	*4.69*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*220*	*162*	*4.9*	*0.96*
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		137	107	4.2	0.97
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		121	105	3.37	0.9
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					77	112	2.42	0.63
AL646053	1939052	ATTGAAATGACGGAGAGTAAACTTCAAAAGAGGGCGTTGGT	m6A	-	GTWWAC	A	RSp1544 (hypothetical protein)	differential					114	115	3.13	0.74
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		56	87	2.4	0.54
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		60	149	2.04	0.43
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		73	129	2.24	0.61
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		70	128	2.21	0.51

Annexure 19 Differential methylation positions of GTWWAC motif raw data between Hawaii evolved clone (Haw e3) and GMI1000

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG32 - Zeb b1			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					56	255	1.8	0.37
AL646052	*117936*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*57*	*2.09*		*130*	*253*	*2.4*	*0.57*
AL646052	*117939*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*52*	*2.26*		*122*	*265*	*2.51*	*0.61*
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					188	255	2.95	0.76
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		125	297	2.41	0.5
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		196	278	2.98	0.72
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					174	210	3.03	0.77
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					92	244	2.1	0.45
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					156	221	2.7	0.69
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					140	276	2.43	0.56
AL646052	2813723	TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA	m6A	+	GTWWAC	A	RSc2612 (hypothetical protein)	differential					74	272	1.8	0.41
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		201	318	2.67	0.7
AL646052	3660528	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	39	82	2.21		102	191	2.67	0.62
AL646052	3660531	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	40	74	2.14		77	193	2.1	0.44
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		131	252	2.25	0.68
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					150	302	2.57	0.55
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		169	258	2.59	0.77
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		67	238	1.9	0.36
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					85	179	2.23	0.47
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		140	235	2.65	0.6
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*353*	*297*	*4.7*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*356*	*300*	*4.03*	*0.94*
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		219	229	3.13	0.86
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					33	194	1.5	0.27
AL646053	1939052	ATTGAAATGACGGAGAGTAAACTTCAAAAGAGGGCGTTGGT	m6A	-	GTWWAC	A	RSp1544 (hypothetical protein)	differential					85	256	2.05	0.39
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		140	257	2.55	0.58
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		158	232	2.8	0.7
AL646053	*2087332*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*					*116*	*247*	*2.31*	*0.53*
AL646053	*2087335*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*	*31*	*97*	*1.71*		*126*	*260*	*2.42*	*0.66*
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		122	222	2.47	0.57

*Annexure 20 Differential methylation positions of G**T**WWAC motif raw data between Zebrina evolved clone (Zeb b1) and GMI1000*

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG36.sequelG1. Zeb b5			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					62	183	2.11	0.46
AL646052	*117936*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT*	*m6A*	-	*GTWWAC*	*A*	*upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*57*	*2.09*		*90*	*149*	*2.65*	*0.57*
AL646052	*117939*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	+	*GTWWAC*	*A*	*upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*52*	*2.26*		*76*	*153*	*2.32*	*0.56*
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		80	128	2.45	0.57
AL646052	655714	AAAGCGCTTCTCATTTGTTAACGATTTGCAGGGTTCAACCA	m6A	-	GTWWAC	A	upstream RSc0608 (ripAA)	differential	56	68	2.63	0.6				
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					105	113	3.02	0.76
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		52	106	2.27	0.46
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		88	110	3.07	0.66
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					79	90	2.94	0.75
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					103	160	2.65	0.6
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					100	151	2.46	0.61
AL646052	2741386	AACATATGACGTTGACGTAAACGTCAATTACCTCGAAGACC	m6A	+	GTWWAC	A	upstream RSc2534 (3-hydroxyacyl-CoA dehydrogenase type II oxidoreductase protein)	differential	64	117	2.27	0.49				
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					113	163	2.41	0.65
AL646052	2813723	TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA	m6A	+	GTWWAC	A	RSc2612 (hypothetical protein)	differential					63	155	1.93	0.43
AL646052	3435906	CCCTTCTCCCGCCACAGTTAACCCTGCATTATCGGCGCCAA	m6A	+	GTWWAC	A	RSc3178 (putative ATP binding protein)	differential	34	100	1.76		53	133	1.94	0.47
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		127	158	2.73	0.71
AL646052	3660528	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	39	82	2.21		51	113	2.22	0.47
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		76	127	2.15	0.64
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					104	149	2.91	0.62
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		132	162	2.97	0.8
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		57	149	2.09	0.41
AL646053	712820	CCGTTATGCGGGCAAAGTAAACGAGGCGTTAGGGAAAATTG	m6A	-	GTWWAC	A	RSp0572 (hypothetical protein)	differential	101	84	3.67	0.84	61	88	2.36	
AL646053	743015	TTCCGTTGTAGTACGAGTAAACGGAATCCGATGCATGCAGC	m6A	+	GTWWAC	A	RSp0611 (hypothetical protein)	differential	65	69	2.61	0.67	38	70	1.91	
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					67	112	2.21	0.51
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		72	111	2.36	0.56
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	-	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*239*	*166*	*4.78*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	+	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*190*	*164*	*3.87*	*0.91*
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		156	128	4.06	1
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		127	122	3.35	0.84
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		92	109	2.53	0.69
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		70	150	2.2	0.48
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		113	104	3.31	0.91
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		80	96	2.38	0.66

*Annexure 21 Differential methylation positions of G**T**WWAC motif raw data between Zebrina evolved clone (Zeb b5) and GMI1000*

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG38. Zeb c1			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					66	311	1.83	0.38
AL646052	117936	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT	m6A	-	GTWWAC	A	upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	57	2.09		92	255	2.16	0.44
AL646052	117939	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	52	2.26		138	262	2.71	0.65
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		110	259	2.26	0.5
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					256	295	3.37	0.83
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		96	327	2.04	0.41
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		188	304	2.69	0.67
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					147	214	2.8	0.69
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					99	211	2.21	0.52
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					153	183	2.78	0.76
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					136	244	2.39	0.6
AL646052	2813723	TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA	m6A	+	GTWWAC	A	RSc2612 (hypothetical protein)	differential					99	237	2.06	0.48
AL646052	3435906	CCCTTCTCCCGCCACAGTTAACCCTGCATTATCGGCGCCAA	m6A	+	GTWWAC	A	RSc3178 (putative ATP binding protein)	differential	34	100	1.76		112	360	2.12	0.5
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		174	323	2.48	0.64
AL646052	3660528	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	39	82	2.21		126	351	2.45	0.54
AL646052	3660531	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	40	74	2.14		121	347	2.15	0.48
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		109	231	2.19	0.61
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					167	320	2.8	0.58
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		164	293	2.43	0.71
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		103	328	2.11	0.43
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					71	229	1.88	0.36
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		85	234	2.12	0.44
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*106*	*67*	*4.72*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*97*	*70*	*3.77*	*0.96*
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		181	126	3.73	0.99
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		187	246	2.82	0.73
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					53	229	1.68	0.35
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		124	247	2.33	0.55
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		154	286	2.52	0.63
AL646053	2087332	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential					108	264	2.32	0.48
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		142	282	2.56	0.67
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		140	249	2.42	0.6

Annexure 22 Differential methylation positions of GTWWAC motif raw data between Zebrina evolved clone (Zeb c1) and GMI1000

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG39. Zeb c3			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					46	251	1.69	0.34
AL646052	117936	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT	m6A	-	GTWWAC	A	upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	57	2.09		83	290	1.98	0.4
AL646052	117939	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	52	2.26		151	287	2.71	0.69
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					197	266	2.85	0.76
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		77	253	1.99	0.4
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		153	244	2.86	0.63
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					167	174	3.62	0.89
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					92	187	2.25	0.52
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					108	180	2.42	0.59
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					183	284	2.8	0.67
AL646052	2813723	TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA	m6A	+	GTWWAC	A	RSc2612 (hypothetical protein)	differential					135	277	2.42	0.56
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52	61	239	1.86	
AL646052	3435906	CCCTTCTCCCGCCACAGTTAACCCTGCATTATCGGCGCCAA	m6A	+	GTWWAC	A	RSc3178 (putative ATP binding protein)	differential	34	100	1.76		107	243	2.53	0.54
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		142	221	2.4	0.64
AL646052	3660531	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	40	74	2.14		118	259	2.44	0.51
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		103	198	2.21	0.65
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					147	270	2.48	0.6
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		152	215	2.55	0.79
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		100	185	2.46	0.54
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					100	186	2.33	0.52
AL646053	1680220	CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG	m6A	-	GTWWAC	A	RSp1329 (hypothetical protein)	differential					75	40	5.04	1
AL646053	1680223	GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC	m6A	+	GTWWAC	A	RSp1329 (hypothetical protein)	differential					61	43	4.08	0.88
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		169	155	3.38	0.93
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		186	210	3.41	0.75
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					47	260	1.61	0.32
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		105	223	2.28	0.52
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		137	275	2.44	0.58
AL646053	2087332	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential					91	244	2.11	0.45
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		144	251	2.82	0.69
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		170	231	3.03	0.7

*Annexure 23 Differential methylation positions of G**T**WW<u>A</u>C motif raw data between Zebrina evolved clone (Zeb c3) and GMI1000*

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG40 - Zeb c4			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					66	271	1.87	0.4
AL646052	117939	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	52	2.26		110	195	2.48	0.67
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		87	269	2.12	0.4
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					224	314	3.08	0.75
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		82	309	1.98	0.38
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		206	289	3.04	0.72
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					166	233	2.78	0.72
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					95	231	2.28	0.47
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					139	214	2.89	0.63
AL646052	*2813720*	*GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC*	*m6A*	*-*	*GTWWAC*	*A*	*RSc2612 (hypothetical protein)*	*differential*					*150*	*271*	*2.51*	*0.6*
AL646052	*2813723*	*TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA*	*m6A*	*+*	*GTWWAC*	*A*	*RSc2612 (hypothetical protein)*	*differential*					*141*	*254*	*2.34*	*0.6*
AL646052	3435906	CCCTTCTCCCGCCACAGTTAACCCTGCATTATCGGCGCCAA	m6A	+	GTWWAC	A	RSc3178 (putative ATP binding protein)	differential	34	100	1.76		49	297	1.64	0.35
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		223	252	2.93	0.83
AL646052	*3660528*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*39*	*82*	*2.21*		*117*	*216*	*2.72*	*0.64*
AL646052	*3660531*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*40*	*74*	*2.14*		*111*	*210*	*2.41*	*0.55*
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		148	240	2.37	0.77
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					123	286	2.32	0.52
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		152	242	2.54	0.73
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		101	224	2.31	0.49
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					114	166	2.57	0.62
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		131	227	2.61	0.59
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*44*	*31*	*4.24*	*0.89*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*51*	*30*	*3.76*	*0.95*
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		270	338	3.32	0.78
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					64	307	1.72	0.38
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		192	250	2.9	0.76
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		141	261	2.51	0.61
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		161	247	2.87	0.76
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		174	233	2.83	0.72

*Annexure 24 Differential methylation positions of G**T**WWAC motif raw data between Zebrina evolved clone (Zeb c4) and GMI1000*

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG42 - Zeb d1			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	117936	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT	m6A	-	GTWWAC	A	upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	57	2.09		59	214	1.85	0.34
AL646052	117939	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	52	2.26		136	209	3.02	0.7
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					188	200	3.72	0.82
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		102	260	2.48	0.44
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		159	246	2.85	0.65
AL646052	1004579	GACTTTGTTATTGTTTGTTAACAAAAAAAGCACTGATTTTG	m6A	+	GTWWAC	A	RSc0957 (hypothetical protein)	differential					72	166	2.29	0.51
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					119	162	2.72	0.69
AL646052	2360132	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential	38	54	2.87	0.56	39	93	1.93	
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					51	93	2.2	0.47
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					112	195	2.36	0.58
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52	37	179	1.74	
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		150	189	2.84	0.71
AL646052	3660528	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	39	82	2.21		113	241	2.43	0.62
AL646052	3660531	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	40	74	2.14		81	228	2.1	0.43
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		88	207	2.06	0.55
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					80	209	2.1	0.44
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		89	169	2.1	0.62
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		54	188	1.87	0.35
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					97	175	2.35	0.53
AL646053	1452556	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	37	76	1.87	0.48				
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*148*	*98*	*4.79*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*119*	*95*	*3.45*	*0.87*
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		132	181	2.79	0.66
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		76	199	1.93	0.44
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		126	219	2.53	0.6
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		111	215	2.4	0.65
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		90	218	2.12	0.47

Annexure 25 Differential methylation positions of GTWWAC motif raw data between Zebrina evolved clone (Zeb d1) and GMI1000

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

seqid	position	context	type	strand	motif	modified base	feature	status	GMI1000.sequelG2				AG47 - Zeb e1			
									score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94117	GAAACAAACGCGTCAAGTTAACGCCGGCGCCGCAGCACACG	m6A	-	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential	41	89	2.08	0.48				
AL646052	117939	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	52	2.26		39	52	2.54	0.61
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					72	65	3.23	0.81
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		46	64	2.41	0.54
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		74	63	3.76	0.89
AL646052	873770	GGCAGGCGAGGTATAAGTAAACGCGCCCGCCTTGACCTCAA	m6A	-	GTWWAC	A	RSc0831 (hypothetical protein)	differential	71	91	2.53	0.61	40	36	2.88	
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					80	71	3.95	0.89
AL646052	2360132	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential	38	54	2.87	0.56	43	69	2.49	
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					69	60	3.3	0.9
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					56	108	2.42	0.47
AL646052	2813723	TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA	m6A	+	GTWWAC	A	RSc2612 (hypothetical protein)	differential					62	110	2.05	0.49
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52				
AL646052	3145217	TTACCCGCAAAGCCGCGTAAACCCTGGGCGAGCGATCCAGT	m6A	-	GTWWAC	A	upstream RSc2918 (putative extracytoplasmic function sigma factor (IRON-regulated) transcription regulator protein)	differential	84	90	3.49	0.82				
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		57	68	2.57	0.66
AL646052	*3660528*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	-	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*39*	*82*	*2.21*		*62*	*88*	*2.81*	*0.65*
AL646052	*3660531*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG*	*m6A*	+	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*40*	*74*	*2.14*		*69*	*89*	*2.57*	*0.62*
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		38	50	2.25	0.62
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					94	90	2.94	0.82
AL646053	792207	CCACGGCTGGGTGGACGTAAACAGCCGGGCGACCTGGCGGT	m6A	-	GTWWAC	A	RSp0641 (peptide synthetase protein)	differential	137	89	4.43	1				
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		54	66	3.57	0.69
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	-	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*102*	*62*	*5.01*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	+	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*86*	*58*	*4.47*	*0.98*
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		117	91	3.98	1
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		110	82	3.96	0.97
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					93	110	2.66	0.75
AL646053	1939052	ATTGAAATGACGGAGAGTAAACTTCAAAAGAGGGCGTTGGT	m6A	-	GTWWAC	A	RSp1544 (hypothetical protein)	differential					44	119	1.94	0.39
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		105	111	3.35	0.8
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		80	80	3.13	0.79
AL646053	*2087332*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC*	*m6A*	-	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*					*64*	*60*	*2.64*	*0.75*
AL646053	*2087335*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	+	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*	*31*	*97*	*1.71*		*48*	*62*	*2.36*	*0.66*
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		70	62	2.75	0.8

*Annexure 26 Differential methylation positions of G**T**WWAC motif raw data between Zebrina evolved clone (Zeb e1) and GMI1000*

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG49 - Zeb e3			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	117939	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	52	2.26		94	136	2.9	0.69
AL646052	127850	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream thiG,RSc0109 (thiazole synthase)	differential	52	88	2.08	0.58	41	142	1.7	
AL646052	313558	CACAGTTGACGTTTACGTAAACGTCAATACAATAGCGCGGC	m6A	-	GTWWAC	A	upstream RSc0279 acyl-CoA dehydrogenase oxidoreductase protein	differential	56	78	2.37	0.57	39	102	1.93	
AL646052	655714	AAAGCGCTTCTCATTTGTTAACGATTTGCAGGGTTCAACCA	m6A	-	GTWWAC	A	upstream RSc0608 (ripAA)	differential	56	68	2.63	0.6	42	123	1.86	
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					100	121	3.22	0.72
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		85	127	2.73	0.6
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					84	109	3	0.68
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					71	110	2.34	0.56
AL646052	2697793	TTTTATAGATCAGTGAGTTAACCAACGACCCGCTAAATGAG	m6A	-	GTWWAC	A	upstream RSc2491 (putative signal peptide protein)	differential	72	88	2.43	0.72	47	90	2	
AL646052	2741386	AACATATGACGTTGACGTAAACGTCAATTACCTCGAAGACC	m6A	+	GTWWAC	A	upstream RSc2534 (3-hydroxyacyl-CoA dehydrogenase type II oxidoreductase protein)	differential	64	117	2.27	0.49	44	91	2.23	
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					91	125	2.98	0.64
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52	41	88	2.22	
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		72	118	2.11	0.53
AL646052	3660528	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc3393 (ISRSO5-transposase protein)	differential	39	82	2.21		67	159	2.02	0.55
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		54	98	2.06	0.54
AL646053	269728	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential	44	80	2.09	0.55	49	173	1.74	
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		60	103	2.07	0.6
AL646053	1432320	TCAAATGATCACTTGAGTTTACCGCTGAGCATACTGTTACC	m6A	+	GTWWAC	A	upstream RSp1136 (ISRSO18-transposase protein)	differential	42	86	2.01	0.45				
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*42*	*25*	*3.9*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*32*	*25*	*4.46*	*0.83*
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		68	45	3.55	1
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		139	166	3.1	0.71
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		70	104	2.53	0.59
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		123	164	2.88	0.76
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		87	148	2.28	0.55

*Annexure 27 Differential methylation positions of G**T**WW<u>A</u>C motif raw data between Zebrina evolved clone (Zeb e3) and GMI1000*

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG80 - Bean a4			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	117939	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream RSc0102 (putative calcium binding hemolysin protein)	differential	33	52	2.26		61	58	3.14	0.83
AL646052	117950	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream RSc0102 (putative calcium binding hemolysin protein)	differential	37	50	2.33	0.54				
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		86	141	2.42	0.56
AL646052	655714	AAAGCGCTTCTCATTTGTTAACGATTTGCAGGGTTCAACCA	m6A	-	GTWWAC	A	upstream RSc0608 (ripAA)	differential	56	68	2.63	0.6	41	70	2.12	
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					70	63	4.62	0.89
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		91	103	2.92	0.71
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		105	94	4.01	0.88
AL646052	1173706	TCATGGCCCCGAGATTGTAAACTTCCGGCTGGCTGGCGAGG	m6A	+	GTWWAC	A	upstream RSc1116 (putative signal peptide protein)	differential	79	54	3.43	1	34	44	1.95	
AL646052	1821399	ACGCTATCGCGAAATAGTTAACATAGACCCATCCGCAGCCT	m6A	-	GTWWAC	A	RSc1705 (hypothetical protein)	differential	79	71	2.49	0.8	49	102	1.89	
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					74	92	3.08	0.73
AL646052	2314079	TCCTCTCATCCGCACGGTAAACCCCCTCTTTGTACTGCCCC	m6A	+	GTWWAC	A	RSc2132 (Type III effector protein)	differential	47	83	2.15	0.53	33	59	1.98	
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					51	57	3.53	0.71
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					45	60	2.28	0.55
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					70	84	2.36	0.65
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52				
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		134	150	2.92	0.76
AL646052	*3660528*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*39*	*82*	*2.21*		*124*	*158*	*2.93*	*0.8*
AL646052	*3660531*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*40*	*74*	*2.14*		*131*	*154*	*2.92*	*0.75*
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		52	58	2.44	0.73
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					72	74	2.87	0.74
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		69	110	2.39	0.56
AL646053	687780	TTCCAGCAGGTCGCACGTAAACTCGACCCCAAGGAAGGCAA	m6A	+	GTWWAC	A	upstream tISRso8a,RSp0547 (ISRSO8-transposase orfA protein)	differential	122	84	4.26	0.99	44	62	2.72	
AL646053	912864	CCAGCGACATCAATTGGTAAACGTTATCACGAATCTCATCG	m6A	-	GTWWAC	A	RSp0725 (Predicted Fe-S oxidoreductase) ; upstream RSp0726 (MFS family transporter)	differential	46	68	2.5	0.53	36	57	2.2	
AL646053	1202810	CTTGATTGAAGTCCCGGTAAACGGCGGCCGTAACTATAACG	m6A	-	GTWWAC	A	RRNA,RS05423 (23S)	differential	50	83	2.05	0.49				
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					68	93	2.37	0.58
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		72	78	2.82	0.72
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*148*	*95*	*4.26*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*129*	*93*	*4.3*	*1*
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		111	102	3.52	0.94
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		96	104	3.03	0.7
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					68	96	2.19	0.62
AL646053	2087335	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	31	97	1.71		62	88	2.23	0.67

*Annexure 28 Differential methylation positions of G**T**WWAC motif raw data between Bean evolved clone (Bean a4) and GMI1000*

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000.sequelG2				AG81 - Bean a5			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94117	GAAACAAACGCGTCAAGTTAACGCCGGCGCCGCAGCACACG	m6A	-	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential	41	89	2.08	0.48				
AL646052	*117936*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*57*	*2.09*		*45*	*63*	*2.85*	*0.56*
AL646052	*117939*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*52*	*2.26*		*52*	*59*	*2.91*	*0.72*
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		62	58	3.07	0.75
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					126	100	4.89	0.97
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		68	84	2.53	0.64
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		74	82	2.82	0.7
AL646052	1134729	TGCAGGTACACCCCTGGTAAACCCCGAACACCCCACCGGAA	m6A	-	GTWWAC	A	upstream gudD1,RSc1079 (glucarate dehydratase protein)	differential	82	97	2.66	0.76				
AL646052	1645829	TGCCCGAAGGCGCACGGTATACCTGCCCGGCATTAGACGCG	m6A	-	GTWWAC	A	upstream RSc1539 (hydrolase protein)	differential	107	86	3.85	0.98				
AL646052	1821402	CTGCGGATGGGTCTATGTTAACTATTTCGCGATAGCGTTTT	m6A	+	GTWWAC	A	RSc1705 (hypothetical protein)	differential	48	68	2.18	0.65	36	56	2.05	
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					48	44	2.76	0.77
AL646052	2360132	AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG	m6A	+	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential	38	54	2.87	0.56	32	54	2.05	
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					50	45	2.83	0.81
AL646052	2527133	CGTGGTTCTTTCGTCCGTAAACGCTAACCATGCGGGTCATG	m6A	+	GTWWAC	A	upstream RSc2326 (putative lipoprotein)	differential	68	76	2.75	0.65	44	48	2.5	
AL646052	2683489	CTCCCCGCCATTGACCGTAAACCCGCGTCACGCATGAATCG	m6A	+	GTWWAC	A	ACUR (RS01108)	differential	50	80	2.23	0.6	31	34	2.36	
AL646052	2697793	TTTTATAGATCAGTGAGTTAACCAACGACCCGCTAAATGAG	m6A	-	GTWWAC	A	upstream RSc2491 (putative signal peptide protein)	differential	72	88	2.43	0.72				
AL646052	*2813720*	*GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC*	*m6A*	*-*	*GTWWAC*	*A*	*RSc2612 (hypothetical protein)*	*differential*					*53*	*75*	*2.02*	*0.53*
AL646052	*2813723*	*TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA*	*m6A*	*+*	*GTWWAC*	*A*	*RSc2612 (hypothetical protein)*	*differential*					*61*	*75*	*2.26*	*0.59*
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52				
AL646052	3039075	GCGCAGCAGCGCCTCGGTAAACAGCTTCTGCAGGAAGCGCA	m6A	-	GTWWAC	A	pilB,RSc2825 (type IV pilus assembly protein)	differential	59	66	3.7	0.73	32	36	2.88	
AL646052	3506641	GAACTTTTCCAATGCTGTAAACCATTCCTTTTTCGAAGCCT	m6A	+	GTWWAC	A	RSc3251 (hypothetical protein)	differential	49	101	2.08	0.51	35	89	1.74	
AL646052	*3660528*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*39*	*82*	*2.21*		*83*	*96*	*2.9*	*0.77*
AL646052	*3660531*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*40*	*74*	*2.14*		*72*	*101*	*2.45*	*0.6*
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					122	133	3.07	0.77
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		74	94	2.48	0.73
AL646053	561449	TGCTGACGAAAAGAAGGTAAACGAGTACGTGCAGAGCATCG	m6A	-	GTWWAC	A	RSp0449 (putative RSH-related protein)	differential	126	117	3.37	0.87				
AL646053	712820	CCGTTATGCGGGCAAAGTAAACGAGGCGTTAGGGAAAATTG	m6A	-	GTWWAC	A	RSp0572 (hypothetical protein)	differential	101	84	3.67	0.84	53	70	2.55	
AL646053	913370	CGAGGCTGCGGCGGGGGTAAACCGCGCCCTGACATATGCCA	m6A	+	GTWWAC	A	RSp0726 (MFS family transporter)	differential	76	69	3.12	0.89				
AL646053	953740	GCTCTCTGATTTTTCCGTAAACGTCGGGGCTATACCGCTTA	m6A	-	GTWWAC	A	RSp0755 (hypothetical protein)	differential	78	67	3.2	0.92	38	40	2.34	
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					41	53	2.32	0.53
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		53	60	3.14	0.67
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*182*	*137*	*4.76*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*165*	*144*	*4.19*	*0.87*
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		131	124	3.46	0.9
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		75	74	3.13	0.76
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					75	59	3.1	0.89
AL646053	1939052	ATTGAAATGACGGAGAGTAAACTTCAAAAGAGGGCGTTGGT	m6A	-	GTWWAC	A	RSp1544 (hypothetical protein)	differential					67	85	2.69	0.58
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		70	88	2.67	0.64
AL646053	*2087332*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*					*60*	*64*	*2.75*	*0.68*
AL646053	*2087335*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*	*31*	*97*	*1.71*		*70*	*65*	*3.13*	*0.85*
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		82	63	3.09	0.9

Annexure 29 Differential methylation positions of GTWWAC motif raw data between Bean evolved clone (Bean a5) and GMI1000

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000				AG82 - Bean b1			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					42	136	1.78	0.42
AL646052	*117936*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT*	*m6A*	-	*GTWWAC*	*A*	*upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*57*	*2.09*		*99*	*178*	*2.45*	*0.55*
AL646052	*117939*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	+	*GTWWAC*	*A*	*upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*52*	*2.26*		*133*	*185*	*3*	*0.76*
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		98	195	2.3	0.52
AL646052	655714	AAAGCGCTTCTCATTTGTTAACGATTTGCAGGGTTCAACCA	m6A	-	GTWWAC	A	upstream RSc0608 (ripAA)	differential	56	68	2.63	0.6	44	173	1.76	
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					143	167	3.28	0.76
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		143	213	2.86	0.65
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		201	200	3.68	0.88
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					91	112	2.8	0.68
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					76	85	2.76	0.69
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					98	79	3.13	0.91
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					80	155	2.35	0.51
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52	48	148	2.03	
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		86	137	2.22	0.56
AL646052	*3660528*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	-	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*39*	*82*	*2.21*		*158*	*175*	*3.68*	*0.88*
AL646052	*3660531*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG*	*m6A*	+	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*40*	*74*	*2.14*		*123*	*184*	*2.79*	*0.64*
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		117	169	2.37	0.77
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		128	205	2.37	0.72
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		145	178	3.22	0.73
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					72	123	2.35	0.52
AL646053	1432320	TCAAATGATCACTTGAGTTTACCGCTGAGCATACTGTTACC	m6A	+	GTWWAC	A	upstream RSp1136 (ISRSO18-transposase protein)	differential	42	86	2.01	0.45	36	122	1.74	
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		128	152	3.34	0.71
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	-	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*502*	*400*	*5.1*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	+	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*482*	*414*	*4.56*	*1*
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		245	212	3.73	0.97
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		123	115	3.84	0.83
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					169	173	2.8	0.91
AL646053	1939052	ATTGAAATGACGGAGAGTAAACTTCAAAAGAGGGCGTTGGT	m6A	-	GTWWAC	A	RSp1544 (hypothetical protein)	differential					69	236	1.9	0.36
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		135	235	2.57	0.59
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		78	145	2.34	0.53
AL646053	*2087332*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC*	*m6A*	-	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*					*145*	*183*	*2.88*	*0.72*
AL646053	*2087335*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	+	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*	*31*	*97*	*1.71*		*163*	*183*	*3.71*	*0.87*
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		160	175	3.34	0.78

Annexure 30 Differential methylation positions of GTWWAC motif raw data between Bean evolved clone (Bean b1) and GMI1000

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

seqid	position	context	type	strand	motif	modified base	feature	status	GMI1000 score	GMI1000 coverage	GMI1000 IPDRatio	GMI1000 fraction	AG84 - Bean b3 score	AG84 - Bean b3 coverage	AG84 - Bean b3 IPDRatio	AG84 - Bean b3 fraction
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					49	236	1.78	0.34
AL646052	*117936*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*57*	*2.09*		*182*	*236*	*3.09*	*0.72*
AL646052	*117939*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*52*	*2.26*		*164*	*244*	*3.07*	*0.74*
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		157	233	2.71	0.67
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					153	216	3.01	0.69
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		96	274	2.14	0.44
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		198	255	2.99	0.75
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					148	139	3.91	0.9
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					164	227	2.97	0.68
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					160	213	2.81	0.72
AL646052	*2813720*	*GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC*	*m6A*	*-*	*GTWWAC*	*A*	*RSc2612 (hypothetical protein)*	*differential*					*173*	*262*	*3.02*	*0.66*
AL646052	*2813723*	*TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA*	*m6A*	*+*	*GTWWAC*	*A*	*RSc2612 (hypothetical protein)*	*differential*					*129*	*255*	*2.29*	*0.56*
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52	51	171	1.87	
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		215	275	2.88	0.77
AL646052	*3660528*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*39*	*82*	*2.21*		*151*	*212*	*3.11*	*0.77*
AL646052	*3660531*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*40*	*74*	*2.14*		*164*	*209*	*3.47*	*0.73*
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		108	147	2.36	0.78
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					201	229	3.43	0.82
AL646053	382743	CGGTGCCGAGTTTGTTGTTAACCCGGATGGGTCGCGAGAGA	m6A	+	GTWWAC	A	RSp0294 (putative hemolysin-type calcium-binding protein)	differential	51	73	2.34		155	215	2.59	0.79
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		165	207	3.13	0.74
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					86	257	2.04	0.4
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		146	168	3.3	0.75
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*630*	*611*	*4.52*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*695*	*616*	*4.72*	*1*
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		482	586	3.59	0.97
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		167	159	3.92	0.84
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					222	334	2.65	0.82
AL646053	1939052	ATTGAAATGACGGAGAGTAAACTTCAAAAGAGGGCGTTGGT	m6A	-	GTWWAC	A	RSp1544 (hypothetical protein)	differential					138	179	3.02	0.65
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		127	176	2.67	0.67
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		100	201	2.37	0.53
AL646053	*2087332*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*					*108*	*214*	*2.24*	*0.55*
AL646053	*2087335*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*	*31*	*97*	*1.71*		*145*	*205*	*2.81*	*0.8*
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		166	207	2.84	0.76

Annexure 31 Differential methylation positions of GTWWAC motif raw data between Bean evolved clone (Bean b1) and GMI1000

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000				AG85 - Bean b4			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	*117936*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*57*	*2.09*		*137*	*158*	*2.94*	*0.75*
AL646052	*117939*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*52*	*2.26*		*119*	*164*	*2.89*	*0.78*
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		83	93	2.95	0.68
AL646052	655714	AAAGCGCTTCTCATTTGTTAACGATTTGCAGGGTTCAACCA	m6A	-	GTWWAC	A	upstream RSc0608 (ripAA)	differential	56	68	2.63	0.6	38	99	1.93	
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					86	101	2.88	0.73
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		112	149	2.83	0.66
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		105	143	2.79	0.66
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					51	73	2.62	0.57
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					115	133	2.93	0.72
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					118	127	2.94	0.77
AL646052	*2813720*	*GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC*	*m6A*	*-*	*GTWWAC*	*A*	*RSc2612 (hypothetical protein)*	*differential*					*168*	*181*	*2.91*	*0.81*
AL646052	*2813723*	*TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA*	*m6A*	*+*	*GTWWAC*	*A*	*RSc2612 (hypothetical protein)*	*differential*					*122*	*180*	*2.87*	*0.63*
AL646052	2856885	CGAATACGGACTAGGTGTAAACCCTGGTGTCGGCCGTTTGC	m6A	-	GTWWAC	A	upstream RSc2654 (serine protease protein)	differential	47	76	2.65	0.52	34	117	1.84	
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		67	123	2.19	0.51
AL646052	*3660528*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*39*	*82*	*2.21*		*140*	*157*	*3.04*	*0.85*
AL646052	*3660531*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*40*	*74*	*2.14*		*110*	*158*	*2.58*	*0.63*
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		77	109	2.2	0.7
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					118	159	2.8	0.68
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		91	122	2.8	0.64
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					60	86	2.58	0.55
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		99	142	2.7	0.64
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*155*	*117*	*4.37*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*177*	*119*	*3.75*	*1*
AL646053	1695111	GTGGCGCGGCCCGCTTGTTAACTCGAGGACAGCGGGCTGTT	m6A	+	GTWWAC	A	upstream RSp1343 (hypothetical protein); ACUR (RS04768)	differential	37	34	2.74		150	158	3.22	0.86
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		111	119	3.47	0.74
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					65	121	2.08	0.52
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		100	177	2.28	0.56
AL646053	2062927	ACTCCCGCTGGCGTTTGTTAACGTTCGTTAATCTTGGGGGT	m6A	+	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	43	96	2.06		89	173	2.37	0.53
AL646053	*2087332*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*					*102*	*98*	*3.01*	*0.81*
AL646053	*2087335*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*	*31*	*97*	*1.71*		*112*	*91*	*3.34*	*1*
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		77	92	2.45	0.65

*Annexure 32 Differential methylation positions of G**T**WWAC motif raw data between Bean evolved clone (Bean b4) and GMI1000*

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

									GMI1000				AG87 - Bean c1			
seqid	position	context	type	strand	motif	modified base	feature	status	score	coverage	IPDRatio	fraction	score	coverage	IPDRatio	fraction
AL646052	94120	GTGCTGCGGCGCCGGCGTTAACTTGACGCGTTTGTTTCAAT	m6A	+	GTWWAC	A	upstream RSc0081 (hypothetical protein)	differential					43	137	1.82	0.39
AL646052	*117936*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGT*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc0103 (ISRSO5-transposase protein) ; upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*57*	*2.09*		*111*	*146*	*3.03*	*0.67*
AL646052	*117939*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream RSc0102 (putative calcium binding hemolysin protein)*	*differential*	*33*	*52*	*2.26*		*132*	*149*	*3.23*	*0.85*
AL646052	127847	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0110 (ISRSO5-transposase protein) ; upstream thiG,RSc0109 (thiazole synthase)	differential	31	89	1.91		96	157	2.53	0.58
AL646052	655717	TTGAACCCTGCAAATCGTTAACAAATGAGAAGCGCTTTCTA	m6A	+	GTWWAC	A	upstream RSc0608 (ripAA)	differential					120	196	3.07	0.58
AL646052	683376	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	RSc0636 (hypothetical protein) ; upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	41	72	2.29		154	197	2.99	0.72
AL646052	683390	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSc0637 (ISRSO5-transposase protein)	differential	32	69	2.1		165	193	3.28	0.76
AL646052	2267250	AGATCGCGGCCGGTGGGTAAACCCCCGATTCTGGCGTTCCG	m6A	+	GTWWAC	A	upstream xdhA,RSc2095 (putative xanthine dehydrogenase (subunit A) oxidoreductase protein)	differential					122	131	3.18	0.8
AL646052	2360129	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					119	137	2.86	0.73
AL646052	2360143	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISrso5,RSc2176 (ISRSO5-transposase protein)	differential					106	134	2.74	0.68
AL646052	2697793	TTTTATAGATCAGTGAGTTAACCAACGACCCGCTAAATGAG	m6A	-	GTWWAC	A	upstream RSc2491 (putative signal peptide protein)	differential	72	88	2.43	0.72	52	184	1.77	
AL646052	2813720	GCTGGGCGTACACGTTGTTTACCTTGTCCCAGAAATTAATC	m6A	-	GTWWAC	A	RSc2612 (hypothetical protein)	differential					43	214	1.61	0.3
AL646052	2813723	TAATTTCTGGGACAAGGTAAACAACGTGTACGCCCAGCAAA	m6A	+	GTWWAC	A	RSc2612 (hypothetical protein)	differential					121	205	2.36	0.58
AL646052	3505345	CGGCACCACCACCACGGTAAACGGCTACAACGCACAAACTG	m6A	+	GTWWAC	A	RSc3249 (putative signal peptide protein)	differential	38	93	1.8		119	161	2.38	0.67
AL646052	*3660528*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*39*	*82*	*2.21*		*164*	*149*	*3.57*	*0.97*
AL646052	*3660531*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream tISRso5,RSc3393 (ISRSO5-transposase protein)*	*differential*	*40*	*74*	*2.14*		*147*	*148*	*3.51*	*0.83*
AL646053	133340	GGTCGGCGATTTCCTGGTTAACCCGGGGGCGGGGAAATATT	m6A	+	GTWWAC	A	upstream RSp0116 (putative lipoprotein)	differential	45	81	2.02		108	194	2.19	0.67
AL646053	269725	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	-	GTWWAC	A	RSp0216 (serine/threonine-protein kinase) ; upstream tISRso5,RSp0217 (ISRSO5-transposase protein)	differential					204	240	3.16	0.78
AL646053	445723	TGGGGTAAACAAAAAGGTTTACATCCCATTGATGGGTGCAA	m6A	-	GTWWAC	A	upstream RSp0338 (negative regulator of exopolysaccharide production transcription regulator protein)	differential					54	112	1.85	0.48
AL646053	445738	GGATGTAAACCTTTTTGTTTACCCCAATCGGGCGCTGTGCA	m6A	+	GTWWAC	A	upstream RSp0338 (negative regulator of exopolysaccharide production transcription regulator protein)	differential					76	110	2.55	0.58
AL646053	565511	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC	m6A	+	GTWWAC	A	RSp0454 (rhs-related protein)	differential	34	88	2.01		82	161	2.42	0.5
AL646053	1298049	ATCGAGCGCGTTTTAAGTAAACGCTGGATAACGTAAAACCT	m6A	+	GTWWAC	A	upstream RSp1026 (signal peptide protein)	differential					68	143	2.06	0.44
AL646053	1452559	ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGG	m6A	+	GTWWAC	A	upstream tISRso5,RSp1152 (ISRSO5-transposase protein)	differential	31	75	1.87		105	186	2.48	0.56
AL646053	*1680220*	*CGACAGACCAAAGCTCGTATACGTTGTGGCACTGCCCCTGG*	*m6A*	*-*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*196*	*138*	*5.08*	*1*
AL646053	*1680223*	*GGGGCAGTGCCACAACGTATACGAGCTTTGGTCTGTCGGAC*	*m6A*	*+*	*GTWWAC*	*A*	*RSp1329 (hypothetical protein)*	*differential*					*185*	*145*	*4.84*	*0.96*
AL646053	1766900	CAAGTACAGGGATTACGTAAACAGTTGCAGGCGCGAAGCCG	m6A	+	GTWWAC	A	RSp1404 (chemotaxis protein)	differential	35	64	1.99		160	150	3.84	0.86
AL646053	1916009	CATTGCACCGCACAACGTTTACCCATAAGGCAAAGGCACAT	m6A	-	GTWWAC	A	upstream efe,RSp1529 (1-aminocyclopropane-1-carboxylate oxidase (Ethylene-forming enzyme))	differential					71	168	1.97	0.49
AL646053	1939296	ATCGCTCTCACTAAGAGTAAACTGGCTCGCATTCACCGTCG	m6A	+	GTWWAC	A	RSp1545 (putative hemagglutinin-related protein)	differential	43	77	1.99		88	223	2.06	0.47
AL646053	2062924	CCCAAGATTAACGAACGTTAACAAACGCCAGCGGGAGTGCT	m6A	-	GTWWAC	A	upstream RSp1643 (hypothetical protein); ACUR (RS02201)	differential	90	97	2.84	0.76				
AL646053	*2087332*	*ATGTAAACTTAGACCAGTTAACAACGGAACAGGACACTAGC*	*m6A*	*-*	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*					*102*	*230*	*2.22*	*0.5*
AL646053	*2087335*	*AGTGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGG*	*m6A*	*+*	*GTWWAC*	*A*	*upstream tISRso5,RSp1675 (ISRSO5-transposase protein)*	*differential*	*31*	*97*	*1.71*		*193*	*225*	*3.55*	*0.89*
AL646053	2087346	GATGCCAACATCCTATGTAAACTTAGACCAGTTAACAACGG	m6A	-	GTWWAC	A	upstream tISRso5,RSp1675 (ISRSO5-transposase protein)	differential	46	90	2.04		186	213	3.11	0.79

Annexure 33 Differential methylation positions of GTWWAC motif raw data between Bean evolved clone (Bean c1) and GMI1000

Threshold of fraction is greater than or equal to 0.50 (to be considered methylated). Fully methylated motifs are highlighted in italics. Empty boxes in fraction means no methylation on the given position.

Summary

Author: Rekha Gopalan-Nair
Email: rekha.gopalan-nair@inrae.fr
Address: LIPM, INRAE

The *Ralstonia solanacearum* species complex (RSSC) is a destructive plant pathogen that infects more than 250-plant species including tomato, potato, pelargonium, ginger and banana. In addition, this multihost pathogen is known for rapid adaptation to new plant species and new environments. In order to overcome this pathogen, it is important to understand the molecular mechanisms that govern host adaptation. The objectives of this thesis were (1) to decipher the genetic bases of adaptation of a RSSC strain to a resistant cultivar, (2) to investigate the potential role of epigenetic modifications in host adaptation and (3) to analyze to impact of the plant species on genetic, transcriptomic and epigenetic modifications in RSSC adapted clones. This study was conducted on clones generated by experimental evolution of GMI1000 RSSC strain after 300 generation of serial passages on the resistant tomato 'Hawaii 7996' plant, the susceptible eggplant 'Zebrina' and the tolerant plant Bean 'Blanc precoce'. Competitive experiments with the GMI1000 ancestral clone demonstrated that 95% of the clones evolved on Hawaii 7996 were better adapted to the growth into this tomato plant than the ancestral clone. Genomic sequence analysis of these adapted clones found between 0 and 2 mutations per clone and we demonstrated that they were adaptive mutations. Transcriptome analysis of the Hawaii, Zebrina and Bean evolved clones revealed a convergence towards a global rewiring of the virulence regulatory network as evidenced by largely overlapping gene expression profiles. Three transcription regulators, HrpB, the activator of the type 3 secretion system regulon, EfpR, a global regulator of virulence and metabolic functions, and PrhP, involved in virulence and adaptation, emerged as key nodes of this regulatory network that were frequently targeted by either genetic or potential epigenetic modification affecting their expression. Significant transcriptomic variations were also detected in evolved clones showing no mutation, suggesting a potential role of epigenetic modifications in adaptation. Comparison of the DNA methylation profiles between the evolved clones and the ancestral clone revealed between 13 and 35 differentially methylated regions (DMRs). No clear impact of the host plant on the list of DMRs appeared. Some of these DMRs targeted genes that were identified to be differentially expressed between the evolved clones and the ancestral clone. This result supported the hypothesis that epigenetic modifications regulate gene expression and could play a major role in RSSC adaptation to new host plants.

Keywords: Phytopathogenic bacteria; RSSC; experimental evolution; host adaptation; adaptive mutations; DNA methylation