Roles of adenosine and cytosine methylation changes and genetic mutations in adaptation to different temperatures

Matthieu Bruneaux[1*†], Ilkka Kronholm[1*], Roghaieh Ashrafi[1], Tarmo Ketola[1]

October 28, 2019

[1]Department of Biological and Environmental Science, University of Jyväskylä, Finland
[*]Joint first authors
[†]Corresponding author: Matthieu Bruneaux, matthieu.bruneaux@ens-lyon.org

Abstract

Epigenetic modifications have been found to be involved in evolution, but the relative contributions of genetic and epigenetic variation in adaptation are unknown. Furthermore, previous studies on the role of epigenetic changes in adaptation have nearly exclusively focused on cytosine methylation in eukaryotes. We collected phenotypic, genetic, and epigenetic data from populations of the bacterium *Serratia marcescens* that had undergone experimental evolution in contrasting temperatures to investigate the relationship between environment, genetics, epigenetic, and phenotypic traits. The genomic distribution of methylated adenosines (m6A) pointed to their role in regulation of gene expression, while cytosine methylation (m4C) likely has a different role in *S. marcescens*. We found both environmentally induced and likely spontaneous methylation changes. There was very little indication that methylation changes were under genetic control. Decomposition of phenotypic variance suggested that both genetic and epigenetic changes contributed to phenotypic variance with slightly higher contribution from genetic changes. Overall, our results suggest that while genetic changes likely are responsible for the majority of adaptation, adenosine methylation changes have potential to contribute to adaptation as well.

Keywords: Adenosine methylation, single molecule real-time sequencing, partitioning of phenotypic variance, experimental evolution.

1 Introduction

The traditional view of evolution is that adaptation proceeds via DNA sequence changes. However, this view has been challenged in recent years as some epigenetic changes, such as DNA methylation changes, have been found to be heritable. Epigenetic changes that are inherited could potentially affect evolution (Jablonka and Raz, 2009; Day and Bonduriansky, 2011; Danchin et al., 2011; Kronholm and Collins, 2016). While convincing cases of epigenetic inheritance do exist, the role of epigenetic variation in evolution has also been met with skepticism (Charlesworth et al., 2017), the main argument for caution being that despite frequent observations we know very little about the relative contributions of genetic and epigenetic variation to adaptation.

Epigenetic variation can be divided into two groups: spontaneous epigenetic variation and induced epigenetic variation (Kronholm, 2017). Spontaneous epigenetic variation is analogous to genetic mutations, such that epigenetic changes occur at a certain rate and are random with respect to fitness. Mutation accumulation experiments in plants have shown that cytosine methylation changes do exhibit these kind of changes and they occur at much higher rates than genetic mutations (Becker et al., 2011; Schmitz et al., 2011; van der Graaf et al., 2015). Modeling studies have shown that spontaneous epigenetic variation has the potential to affect evolutionary dynamics. The different rates of epigenetic and genetic changes can cause a two-phase dynamic where adaptation happens first via epigenetic changes, with genetic changes eventually replacing epigenetic changes (Klironomos et al., 2013; Kronholm and Collins, 2016). The second category of epigenetic changes are induced changes as a result of a specific environmental signal or developmental stage and are guided by an underlying genetic program. These changes can be seen as a mechanism of phenotypic plasticity or transgenerational effects, and there are many examples of such phenomena in plants (Luna and Ton, 2012; Wibowo et al., 2016; Herman and Sultan, 2016; Zheng et al., 2017).

So far empirical results have lacked behind theoretical models as it has been difficult to disentangle the contributions of epigenetic and genetic variation to adaptation.

Several studies have investigated the extent of natural epigenetic variation (Richards et al., 2017), but in many cases effects of genetic changes cannot be excluded. It has also been shown that considerable amount of DNA methylation variation is under genetic control (Dubin et al., 2015; Hagmann et al., 2015). Nevertheless, evolutionary experiments with microbes suggest that epigenetic changes can contribute to adaptation (Wang et al., 2015; Kronholm et al., 2017) and that a two-phase dynamic of epigenetic changes followed by genetic adaptation can happen (Stajic et al., 2019).

While examples of inherited epigenetic changes and their involvement in adaptation exist, we don't understand the relative importance of epigenetic and genetic variation in evolution. Moreover, the majority of empirical work has focused on DNA methylation, in particular studying the role of 5-methylcytosine in eukaryotes. Other modifications such as adenosine methylation, that is common in prokaryotes (Ratel et al., 2006) but which occurs in eukaryotes as well (Iyer et al., 2016), have received much less attention (but see Ma et al. (2019)).

Prokaryotes exhibit several types of methylated DNA bases in their genomes: C5-methylcytosine (m5C), which is historically the best studied methylated base in eukaryotes, N4-methylcytosine (m4C) and N6-methyladenosine (m6A) (Ratel et al., 2006). Such DNA modifications can influence gene expression (Bird, 2002; Casadesús and Low, 2006) and some of them are heritable (Bird, 2002). Some works in prokaryotes have pointed the potential for adenosine methylation to be involved in adaptation (Adam et al., 2008; Atack et al., 2015). The roles of adenosine methylation in bacteria are multiple: protection against foreign DNA by restriction-modification systems, gene expression regulation, DNA replication and repair, cell-cycle regulation and phase variation (Sánchez-Romero et al., 2015). Adenine methyltransferase (MT) genes can either be essential for cell viability (Stephens et al., 1996), result in global transcription changes when mutated (Casselli et al., 2018) or even be mutated without affecting cell survival nor transcription levels (Seshasayee, 2007).

We studied the role of epigenetics in evolution by addressing the following points:

(1) quantifying in detail the epigenetic variation related to m4C and m6A in the bacterial model *Serratia marcescens* evolving in different temperatures, (2) determining if epigenetic variation contributed to adaptation, and (3) comparing the relative contributions of epigenetic and genetic variation to adaptation. We address these questions by phenotyping and single molecule real-time (SMRT) sequencing of bacterial clones from an experimental evolution experiment conducted at three contrasting thermal regimes: 24–38°C fluctuating and 31 °C and 38 °C constant environments (Ketola et al., 2013).

2 Methods

2.1 Origin of sequenced clones (experimental evolution)

We used bacteria clones that were obtained from a previous evolution experiment (Ketola et al., 2013). Briefly, we let populations of *Serratia marcescens* initiated from a single common ancestor clone evolve under either constant or fluctuating temperatures during three weeks (treatments: constant 31 °C, constant 38 °C or daily variation between 24 °C and 38 °C) (Ketola et al., 2013) (Supplementary Figure S1). Note that lines evolved in 38 °C were not reported in Ketola et al. (2013). The experiment lasted approximately 70 generations. After experimental evolution, individual clones were stored at −80 °C in 50 % glycerol. We randomly selected one clone from each population for sequencing (10 from the 31 °C treatment, 8 from the 38 °C treatment and 10 from the fluctuating treatment, i.e. 28 evolved clones sequenced in total). As a reference, and since unfortunately the frozen stocks of the common ancestor used to initiate the evolved populations could not be successfully revived prior to sequencing, we also sequenced the stock clone received from ATCC and from which the single common ancestor itself was derived (Supplementary Figure S1).

2.2 Phenotypic measurements

We used the phenotypic data from Ketola et al. (2013), which included measurements of growth rate and yield at constant 24 °C, 31 °C and 38 °C for the evolved clones. These phenotypes will be called *temperature-related traits* from now on. In addition, growth rate and yield were measured in a series of novel environments: under redox balance stress (1 mg/ml dithiotreitol), in the presence of the ciliate predator *Tetrahymena thermophila* and in the presence of the lytic bacterophage PPV. These traits will be called *coselected traits* from now on. For further details on these phenotypic measurements, see Ketola et al. (2013).

This phenotypic dataset was enriched with two additional traits: *strain virulence* (*S. marcescens* is an opportunistic pathogen of insects (Grimont and Grimont, 1978; Flyg et al., 1980)) and *prophage activation* (the genome of the reference strain contains several prophages, and one of them could be activated in our experimental conditions). The full technical detail of how these traits were measured is presented elsewhere (Bruneaux et al., 2019), but a brief description is given here. Strain virulence was estimated by measuring the survival time of wax moth larvae (*Galleria mellonella*) after injection with 5 µl of an overnight clone culture in SPL 1 % at 31 °C. Larvae were kept at 24 °C or 31 °C after injection, providing a measure of clone virulence at two temperatures. Prophage activation was measured by growing the clones under five different temperature treatments in SPL 1 %. Each treatment lasted two days, and the temperatures for the first and second day for each treatment were (first/second day temperatures): 31/31°C, 24/24°C, 38/38°C, 24/38°C and 38/24°C. After the second day and in order to distinguish between phage DNA inside bacteria and inside free-floating phage particles, each clone culture was harvested into two paired samples: one native sample and one supernatant sample obtained after mild centrifugation to pellet most of the cells while leaving phage particles in suspension. Both samples were then DNase-treated and incubated at 95 °C to release DNA from bacteria cells and potential phage particles. The amount of chromosomal DNA and of prophage DNA was

quantified with qPCR in all samples using two pairs of primers targetting the prophage sequence and a chromosomal sequence outside the prophage region. A Bayesian model using the four DNA quantities generated from each culture well (prophage DNA and non-prophage DNA in both native sample and supernatant) allowed to estimate the proportion of prophage DNA copies which were not contained in bacterial cells in the cultures.

2.3 Sequencing and genome annotation

We used single molecule real-time sequencing using the PacBio platform to sequence the evolved clones and the reference. Since no template amplification takes place prior to sequencing on a PacBio platform in order to detect base modifications, relatively large amounts of DNA per clone are necessary. Selected clones were thawed and grown overnight in 150 ml of liquid medium and DNA was extracted using the Wizard Genomic DNA Purification Kit from Promega (WI, USA). One DNA sample (20 to 60 µg) per clone was sequenced by the DNA Sequencing and Genomics Laboratory of the University of Helsinki on a PacBio RS II sequencing platform using P6-C4 chemistry. Two single-molecule real-time sequencing (SMRT) cells were run per DNA sample.

PacBio software and recommended protocols were used with default parameters for the assembly and modification calling pipeline. For each strain, reads from the RS II instrument were assembled with PacBio `RS_HGAP_Assembly.3`, as implemented in SMRTportal 2.3.0. The resulting assembly for each strain was processed with the Gap4 program to generate *de novo* a first draft sequence for this strain and to circularize it. PacBio `RS_Resequencing.1` protocol was then run 2 to 3 times for each sample to map the reads to the draft sequence and generate a consensus sequence for each strain. The average coverage of the draft chromosome per strain was high (from 102 to 413, average 268). Given the high coverage of the chromosome sequences across all strains, the base calling was considered accurate and genetic variants were directly called from an alignment of the 29 strains chromosomes built using Mugsy (Angiuoli and Salzberg,

2011).

Inter-pulse duration (IPD) ratios were used with the PacBio `RS_Modification and Motif_Analysis` protocol to detect modified bases and methylation sequence motifs. Positions detected as modified by this protocol were labelled as either m6A, m4C or "modified base" if the modification type could not be identified. The estimated fraction of modified copies (methylation fraction) was provided for m6A and m4C bases. The protocol only reports bases for which IPD ratio is significantly different from 1, which makes it highly coverage-dependent for modifications with weak signal such as m4C amd m5C. In order to extract the estimated IPD ratios for all bases and not only the ones detected as modified by this protocol, we also ran `ipdSummary` separately using all the aligned subreads for each sample.

Genome annotation

We used a previously published and annotated genome for *Serratia marcescens* strain ATCC 13880 (RefSeq entry `GCF_000735445.1`) to annotate the chromosome sequence of the reference strain used in our experiment. CDS from the RefSeq entry were aligned to the reference genome using blast (Camacho et al., 2009) and for each CDS the best high-scoring segment pair was used to propagate annotation to the reference genome if its length was at least 99% of the CDS length. After annotating CDS, they were assembled into operons using the Operon Mapper server (`http://biocomputo.ibt.unam.mx/operon_mapper/`, Taboada et al. (2018)). The location of the origin of replication was determined using the DoriC server (Gao et al., 2012).

2.4 Determination of tetramer composition bias in *Serratia marcescens* genome

Oligonucleotide usage bias in prokaryotic genomes exists as a result of evolutionary constraints, such as codon usage and palindrome avoidance (Rocha et al., 1998). In order to test if the target oligonucleotide sequence of the adenosine methylase (5′-

GATC-3′) was under differential selection depending on the genomic context, which could indicate a potential role of m6A methylation in cell function, we calculated the tetramer composition bias in two subsets of genomic segments: genes (CDS) and promoter regions (defined as 200-bp-long regions immediately upstream of operon-leading CDS). We used the sequences corresponding to the (+) strand of each gene or of the CDS downstream of each promoter region. For each of the 256 possible tetramers, we counted the number of observed occurrences in each of those sequence sets, N_{tet}. We compared those observed values with expected number of occurrences, determined either by permutation of the bases within each genomic segment or based on a Markov chain. The Markov chain takes into account the underlying biases that might exist in the frequencies of dimers and trimers comprising each tetramer, while the permutation approach only takes into account biases in frequencies of the A, T, G, C monomers. In the permutation approach, the expected number of occurrences of a given tetramer for a given genome subset is the average of the number of occurrences obtained across n_{perm} permutations of the bases within each segment of this subset: $\mathrm{E}_{\text{tet}}^{\text{perm}} = \frac{1}{n_{\text{perm}}} \sum_{i=1}^{n_{\text{perm}}} N_{\text{tet},i}$. In the Markov chain approach, the expected number of occurrences of a given tetramer of composition $b_1b_2b_3b_4$ is: $\mathrm{E}_{\text{tet}}^{\text{MC}} = \frac{\mathrm{N}_{b_1b_2b_3} \times \mathrm{N}_{b_2b_3b_4}}{\mathrm{N}_{b_2b_3}}$ (Rocha et al., 1998; Pride et al., 2003). Deviations in the usage of each tetramer from expectation were calculated as $D_{\text{perm}} = \frac{\mathrm{N}_{\text{tet}} - \mathrm{E}_{\text{tet}}^{\text{perm}}}{\mathrm{E}_{\text{tet}}^{\text{perm}}}$ and $D_{\text{MC}} = \frac{\mathrm{N}_{\text{tet}} - \mathrm{E}_{\text{tet}}^{\text{MC}}}{\mathrm{E}_{\text{tet}}^{\text{MC}}}$ for the permutation and Markov chain approaches, respectively.

2.5 Detection of methylated positions (MP) and regions (MR) of interest

While some methylated positions (MP) in prokaryotes occur at specific loci (such as GATC motifs) targeted by restriction-modification systems or by orphan methylases, other MP might not be associated with an identified motif. For those, identifying larger methylated regions (MR) spanning several bases and using their average methylation status for strain comparison instead of individual MP can be a more biologically

meaningful approach, as is done in eukaryotes for m5C (e.g. Hagmann et al. (2015)).

We aimed at identifying both methylated *positions* (m6A) and methylated *regions* (m4C) of interest, focusing on partially methylated m6A in GATC motifs for the former and using a kernel density (KD) approach for the latter to detect regions where cytosine modification into m4C was more frequent than expected by chance.

Detection of partially methylated m6A loci in GATC motifs

Serratia marcescens, like other gamma-proteobacteria, harbours a Dam enzyme methylating adenosines of GATC motifs (Blow et al., 2016). Most GATC motifs are usually fully methylated, but hemimethylated locations can regulate DNA replication and cell division while some other locations can be left unmethylated when another regulatory protein already binds DNA and prevents Dam from methylating it (Casadesús and Low, 2006). This competition for access to GATC motifs between Dam and regulatory proteins allows for heritable regulation of gene expression in bacteria lineages (Braaten et al., 1994). To identify GATC loci which were not fully methylated in our dataset, we considered for each GATC locus the estimated fractions of modified adenosines on the plus and minus strand for each strain, as reported by the PacBio pipeline. Loci for which no fraction was reported for a given strain were assigned a value of 0 (i.e. completely unmethylated) for that strain. Assuming that the distribution of modified fractions on the plus and minus strand for fully methylated loci would follow a truncated bivariate distribution centered around (1, 1) (i.e. both strands fully methylated), we defined the set of partially methylated GATC loci of interest for downstream analyses as the loci which presented modified fractions on the plus and minus strands which deviated from the point of full methylation at coordinates (1, 1) more than four times the average quadratic distance to (1, 1), in at least one strain (Supplementary Figure S2). We checked that low estimated values of methylated fractions were not due to low coverage of the corresponding GATC loci by examining the relationship between coverage bins and estimated methylated fraction (Supplementary Figure S3).

Detection of clustered m4C based on kernel density estimates

While studies of m5C in eukaryotes have profitably used Hidden Markov Models (HMM) to identify clusters of methylated cytosines, our attempts at using HMM were unsuccessful as most of the predicted segments after HMM fitting only contained one or two m4C, which might be due to the dispersed nature of modified cytosines in *Serratia marcescens* genome compared to modified cytosines in eukaryotes which tend to be aggregated in CpG islands. We thus chose to use a simpler approach to detect regions which were enriched in m4C based on kernel density estimates.

Each cytosine base on either strand of the reference genome was flagged as m4C if it was detected as such in at least one of the sequenced strains, yielding 77 478 m4C locations with an average spacing of 66 bases between them. Using a grid of 2^{17} evenly spaced points along the reference genome (i.e. about 40 bases away fom each other, slightly less than the average spacing of observed m4C positions), we calculated a kernel density estimate with a Gaussian kernel with a standard deviation of 65 bases. This observed density estimate was then compared with density estimates generated by randomly permuting the 77 478 m4C locations among the 3 059 758 cytosine locations on the reference genome: a kernel density estimate was produced for each permuted dataset at the same grid points and with the same kernel parameters as for the observed density estimate. We performed 100 000 permutations and identified candidate grid points for which the proportion of permuted estimates above the observed estimate was < 0.001, suggesting that the frequency of m4C bases was higher than expected by chance in the genome grid cells centered at those grid points. The advantage of this permutation approach is that it takes into account the local distribution of cytosine bases across the genome, since it is maintained in the permuted datasets. Consecutive candidate grid cells were assembled into segments while single candidate grid cells were discarded, yielding a final number of 167 segments that were used as candidate "clusters" of m4C. For each sequenced strain and each "cluster", we calculated the value of the corresponding m4C epiallele by averaging the methylated fraction of all

m4C observed in the cluster region. Positions which were not detected as m4C in a given strain (but were in at least another one) were assigned a methylation fraction of 0 for this strain before calculating the average epiallele. We note that this approach could result in underestimating the m4C fraction for cytosines with lower coverage as we observed that average coverage for m4C bases with low estimated methylation fraction tended to be slightly higher than for bases with high methylation fraction (Supplementary Figure S4), which is expected given the relatively weak kinetic signal produced by m4C modification.

2.6 Association between epigenetic changes and genetic mutations

The association between epigenetic changes and genetic mutations was investigated using the methylated positions (for m6A) or regions (for m4C) of interest identified as described above and the genetic mutations for which the minor allele was present in at least two of the sequenced strains (i.e. genetic mutations present in a single strain were not used). For each genetic mutation, we built a quantile-quantile curve comparing a uniform distribution of p-values on $[0, 1]$ with the distribution of observed p-values from t-tests between the genetic mutation and the epigenetic changes. In order to determine if the genetic mutation under consideration was more strongly associated with epigenetic changes than expected by chance, we then compared this quantile-quantile curve with a set of similar quantile-quantile curves obtained from permutated datasets where the clones labels were randomly shuffled. This permutation approach allowed us to account for the effect of our dataset structure (such as allele frequencies and distribution of methylated levels) on the expected distribution of p-values.

2.7 Decomposition of phenotypic variance

To quantify the relative contributions of genetics and epigenetics to the phenotypic variance, we used random effect models in which the variance component was split into genetic, epigenetic and residual variance (Thomson et al., 2018). The model we used for each trait was:

$$\mathbf{y} = \text{MVN}\left(\mathbf{0}, \sigma^2_{gen} \times G\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{epi} \times E\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{res} \times I\right) \quad (1)$$

where $\mathbf{y}$ is a column vector containing the trait values for each strain centered to a mean of 0 and scaled to a standard deviation of 1, MVN is the multivariate normal distribution parameterized by a vector of means and a variance-covariance matrix, $\mathbf{0}$ is a column vector of zeros, σ^2_{gen}, σ^2_{epi} and σ^2_{res} are the genetic, epigenetic and residual variances for the trait, G and E are the matrices of genetic and epigenetic similarity between strains, respectively, and I is the identity matrix. The genetic similarity matrix G was built from the 54 variable genetic loci observed in our dataset and the similarity between any two strains was calculated as the proportion of those loci for which the strains shared identical alleles. The epigenetic similarity matrix E was built from the methylation fractions for the partially methylated m6A epiloci in GATC motifs identified previously. First, an Euclidean distance matrix with elements d_{ij} was built from the methylation fraction data. The similarity measure between any two clones was then calculated as $1 - (d_{ij}/(2D))$, where D was the average distance between the reference strain and all the evolved strains. This is conceptually equivalent to measuring a phylogenetic similarity between two species as the shared evolutionary distance from the root of the phylogeny to their last common ancestor divided by the average distance from the root of the phylogeny to all observed species.

We run the models using the MCMCglmm package in R (Hadfield, 2010). We used an inverse-Gamma prior for each variance component, with shape 0.5 and scale 0.5 (corresponding to $V = 1$ and $nu = 1$ in MCMCglmm parameterization). Four chains

were run for 10 000 iterations, of which the first half was discarded as burn-in, with a thinning of 10. Total phenotypic variance was calculated as $\sigma^2_{tot} = \sigma^2_{gen} + \sigma^2_{epi} + \sigma^2_{res}$, and r^2_{gen}, r^2_{epi} and r^2_{res} were calculated as $\sigma^2_{gen}/\sigma^2_{tot}$, $\sigma^2_{epi}/\sigma^2_{tot}$ and $\sigma^2_{res}/\sigma^2_{tot}$, respectively.

Calculation of partial variance components. In our experiment, epigenetic similarities could be due to genetic control, and both epigenetic and genetic similarities could be due to the evolutionary treatment. To disentangle the joint and disjoint effects of treatment, genetics and epigenetics on phenotypic variance, we fitted seven mixed models similar to the one described above, built from all possible combinations of the treatment, genetic and epigenetic variance components. The fullest model was:

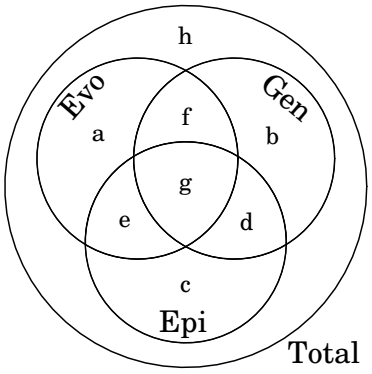
$$\begin{aligned}\mathbf{y} =& \text{MVN}\left(\mathbf{0}, \sigma^2_{evo} \times T\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{gen} \times G\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{epi} \times E\right) + \\ & \text{MVN}\left(\mathbf{0}, \sigma^2_{res} \times I\right)\end{aligned} \tag{2}$$

where σ^2_{evo} is the treatment variance for the trait and T is the design matrix describing the evolutionary treatment corresponding to each clone. The six other models were:

$$\begin{aligned}
\mathbf{y} &= \text{MVN}\left(\mathbf{0}, \sigma^2_{gen} \times G\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{epi} \times E\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{res} \times I\right) \\
\mathbf{y} &= \text{MVN}\left(\mathbf{0}, \sigma^2_{evo} \times T\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{epi} \times E\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{res} \times I\right) \\
\mathbf{y} &= \text{MVN}\left(\mathbf{0}, \sigma^2_{evo} \times T\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{gen} \times G\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{res} \times I\right) \\
\mathbf{y} &= \text{MVN}\left(\mathbf{0}, \sigma^2_{epi} \times E\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{res} \times I\right) \\
\mathbf{y} &= \text{MVN}\left(\mathbf{0}, \sigma^2_{gen} \times G\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{res} \times I\right) \\
\mathbf{y} &= \text{MVN}\left(\mathbf{0}, \sigma^2_{evo} \times T\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{res} \times I\right)
\end{aligned} \tag{3}$$

For each of those models, we calculated the proportion of the phenotypic variance described by non-residual variance components, using the mean of the MCMC chain as a point estimate. We thus obtained seven values of proportions of explained variance for each phenotypic trait: $r^2_{evo,gen,epi}$, $r^2_{gen,epi}$, $r^2_{evo,epi}$, $r^2_{evo,gen}$, r^2_{epi}, r^2_{gen} and r^2_{evo}. We then calculated the joint and disjoint proportions explained by treatment, genetics and

epigenetics using a variation partitioning approach (Legendre and Legendre, 1998) as
316 depicted in the following representation:



using the following formulas:

$$a = r^2_{evo,gen,epi} - r^2_{gen,epi}$$

$$b = r^2_{evo,gen,epi} - r^2_{evo,epi}$$

$$c = r^2_{evo,gen,epi} - r^2_{evo,gen}$$

$$(d+g) = r^2_{gen} + r^2_{epi} - r^2_{gen,epi}$$

$$(f+g) = r^2_{gen} + r^2_{evo} - r^2_{evo,gen}$$

$$(e+g) = r^2_{epi} + r^2_{evo} - r^2_{evo,epi}$$

$$g = (d+g) + (f+g) + (e+g) - r^2_{evo} - r^2_{gen} - r^2_{epi} + r^2_{evo,gen,epi}$$

$$d = (d+g) - g$$

$$e = (e+g) - g$$

$$f = (f+g) - g$$

$$h = 1 - r^2_{evo,gen,epi}$$

Taking into account the effect of genetic loci associated with epigenetic changes. After having identified the genetic loci potentially associated with epigenetic changes, we re-ran the same phenotypic variance decomposition as presented above but adding in Equations 1, 2 and 3 the fixed effect of one genetic locus (or haplotype) at a time and removing the locus (or haplotype) from the calculation of the

genetic similarity matrix. For example, when testing for the effect of haplotype a on phenotypic variance decomposition, Equation 1 would become:

$$\mathbf{y} = \beta_a \times \text{genotype}_a + \text{MVN}\left(\mathbf{0}, \sigma^2_{gen} \times G_a\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{epi} \times E\right) + \text{MVN}\left(\mathbf{0}, \sigma^2_{res} \times I\right)$$

where G_a is the genetic similarity matrix calculated after removing the haplotype a from the table of genetic variants. Equations 2 and 3 would be modified similarly. The obtained variance decomposition is thus providing the contribution of evolutionary treatment, genetic and epigenetic variations to phenotypic variance after removing the effect of haplotype a on the phenotypes.

2.8 Correlations between genetic, epigenetic and phenotypic distances

In this study, we were interested in the overall relationships between evolutionary treatment, genetic, epigenetic and phenotypic datasets to understand how they globally coincide with each other and draw general conclusions about evolution, rather than in pin-pointing specific genetic or epigenetic mutations responsible for phenotypic changes. To this effect, and as an addition to the variance decomposition approach, we used Mantel tests as implemented in the R package vegan (Oksanen et al., 2019) to investigate the association between distance measures based on each of those datasets. If strains which are close based on one distance measure also tend to be close based on another one, then it suggests a dependence between the underlying sets of measurements used to calculate those distances (Mantel and Valand, 1970). It should be noted that the Mantel test does not test for the independence between the sets of variables which were used to calculate the distance matrices, and that the R^2 from a Mantel test is not the same as the R^2 from correlation or regression analysis (Legendre et al., 2015).

3 Results

3.1 Genome sequences and genetic variants

The reference strain chromosome was 5 117 300 bp long, with a GC content of 59.8%. The genome annotation propagated 4628 out of 4697 CDS from the RefSeq entry to the reference genome (98.5%). 54 mutations were identified from the aligned genomes of the reference and evolved clones (Table 1 and Figure 1). Most of the mutations that occurred in coding regions were frameshifts or missense mutations. A striking feature of the variant map is the presence of 11 genetic mutations associated in a single haplotype and for which the minor allele is observed in 5 of the evolved strains from 38 °C and in the reference genome, but in no other evolved strain (haplotype *a* in Supplementary Table S1). This suggests that the ancestor clone used to initiate the replicated populations in the evolution experiment, and which was itself derived from the reference strain sequenced here, actually contained at least two lineages which were preserved in some populations at 38 °C but in no other evolutionary treatment. We are thus careful to consider the haplotype *a* variants separately from the rest of the genetic changes in the rest of this manuscript, since those variants are likely to have arisen prior to the start of the evolution experiment.

Even when not taking into account variants from haplotype *a*, multiple parallel substitutions were observed among the evolved clones (Figure 1), and some genes in particular exhibited several independent mutations occurring in different strains: two independent mutations occurred in a deacetylase, three occurred in a galactokinase and three in a glycosyltransferase (Supplementary Table S1). Remarkably, those three genes are all involved in some steps of the biosynthesis of lipopolysaccharide. In the evolution experiment population sizes were on the order of 8×10^6 cells per 400 µL culture well at plateau, and the fraction transferred to the next generation was 1/10 of the previous culture. This yields a large effective population size $N_e = 2.6 \times 10^6$ during the experiment ($N_e = N_0 g$, where $N_0 = 8 \times 10^5$ cells is the bottleneck size and $g = 3.3$ is

Type	Location	Effect on protein sequence
Indels (31, 30*)	coding regions (13, 12*)	frame shift (11, 10*) no frame shift (2, 2*)
	non-coding regions (18, 18*)	
SNPs (23, 13*)	coding regions (17, 9*)	non-synonymous (14, 8*) synonymous (3, 1*)
	non-coding regions (6, 4*)	

Table 1: Summary of genetic variants across the reference and the 28 evolved strains. Counts are given in parentheses. Numbers with asterisk are counts when the variants comprising haplotype *a* are not taken into account.

the number of generations between transfers (Lenski et al., 1991)). Thus, genetic drift cannot explain the fixation of multiple parallel mutations in independent populations, suggesting instead that selection favoured abolishing the function of those particular genes. As the aim of the present study is to determine the overall relationships between treatment, genetics, epigenetics and phenotypes rather than linking a particular mutation to phenotypic changes, we do not present here any results for the association between specific mutations and phenotypic traits, but such results are presented in details in a separate study (Bruneaux et al., 2019).

3.2 Overview of methylation in *Serratia marcescens* genome

3.2.1 Methylated bases and methylation motifs

The role of methylation in bacteria is still being actively investigated, so we first examined the overall patterns of methylation in the *Serratia* genome. For adenine methylation, out of 2 057 542 adenine bases present in the bacterial chromosome, 90 804 (4.4 %) were detected as m6A in at least one strain by the PacBio protocol. The vast majority of those m6A were occurring in GATC motifs: out of the 90 804 positions detected as m6A, 76 241 (84 %) were in a GATC context. Since a total of 76 300 adenines in GATC context exist in the genome, this corresponds to a very high rate of adenine methylation

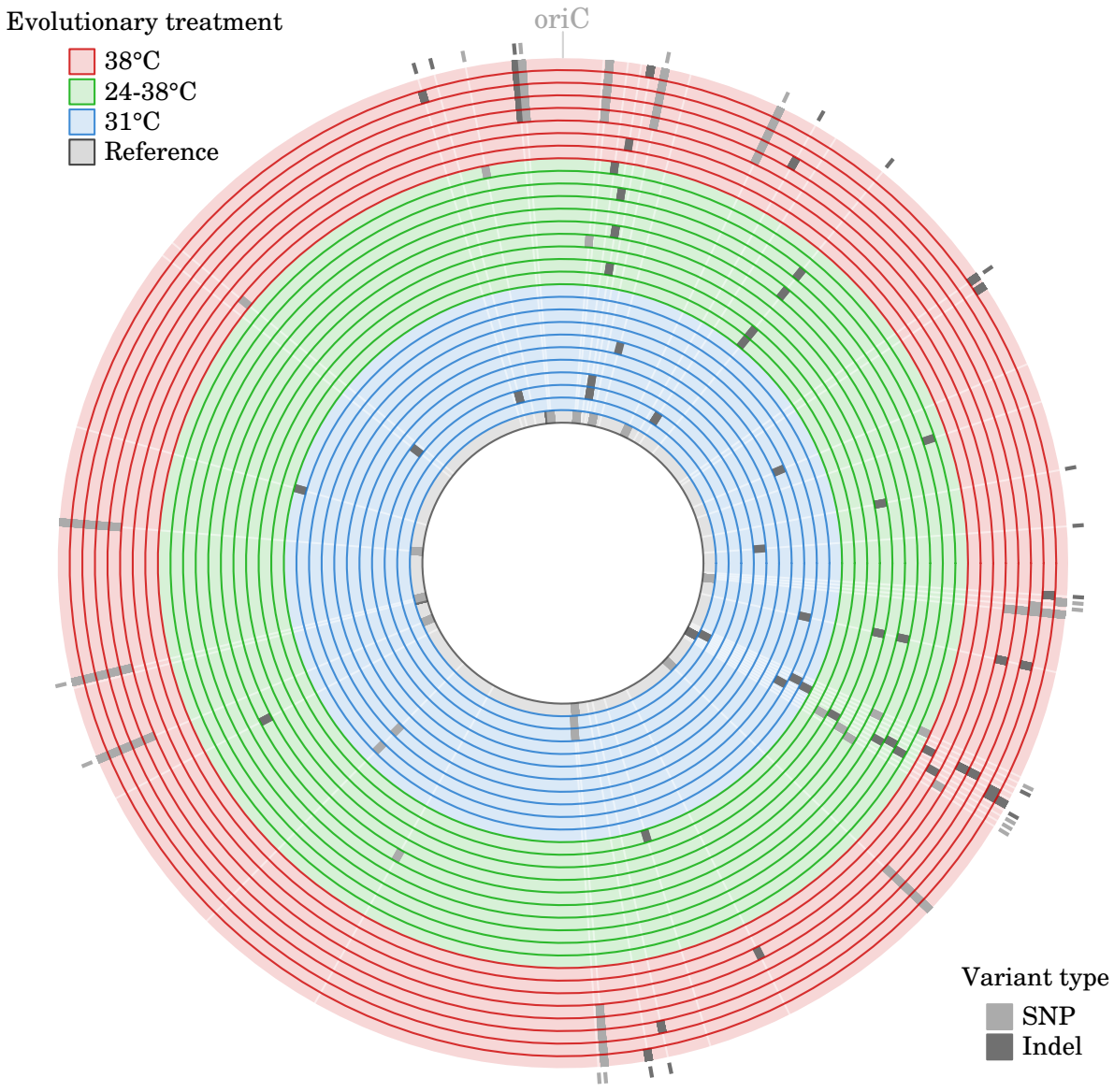


Figure 1: Distribution of genetic variants along *S. marcescens* chromosome. Each circular lane represents the chromosome sequence of one clone. Genetic variants (minor alleles) are depicted in light gray (SNP) and dark gray (indel). Markers on the outer part of the map highlight non-synonymous variants (i.e. indels resulting in a frame shift and non-synonymous SNPs).

in GATC motifs: 99.9 % of GATC adenines are detected as m6A in at least one strain and 99.7 % were detected in all 29 strains. Additionally, 67 711 other adenine bases were detected as modified (but without identifying the specific modification type) in at least one strain, of which only 25 821 were detected in at least two strains.

Out of 3 059 758 cytosine bases present in the genome, 77 481 (2.5 %) were identified as m4C in at least one of the 29 sequenced strains. The average number of m4C detected per strain was 15 965, with a standard deviation (s.d.) of 3200. In addition, 106 378 C bases (3.5 % of the genome C content) were detected as modified in at least two strains, but without a specific modification type being identified by the PacBio protocol. Some of those unidentified modifications could be m5C, for which the kinetic signal is weaker than for m4C and m6A and which requires either very high coverage to identify the weak signal unambiguously or sequencing using TET-modification to generate a stronger kinetic signal for easier identification (Clark et al., 2013).

When detected as modified, m4C bases had an average methylated fraction of 71 % (s.d. 23 %) and m6A bases had an average methylated fraction of 63 % (s.d. 30 %) outside GATC motifs, and of 97 % (s.d. 5 %) inside GATC motifs.

Methylation target sequences. The motif finder algorithm of the PacBio pipeline detected two sets of motifs for adenosine methylation (Table 2) but no reliable motif for cytosine methylation. One motif set for m6A was the GATC palindrome and the other was the much rarer pair AAAGNNNNNNTCG/TTTCNNNNNNAGC. For both sets, almost all genomic locations (> 99 %) were detected as modified. For m4C modifications, although no specific motif was found, the context around modified C bases was distinctively enriched in G, with average G abundances of 55 % for the base immediately before an m4C and of 58 % and 51 % for the two following bases in the $5' \rightarrow 3'$ direction, compared to the average genomic abundance of 30 % of G.

Motif	Occurrences on chromosome	Detected as methylated
5′−G<u>A</u>TC−3′ 3′−CT<u>A</u>G−5′	38150 38150	99.8 %
5′−AA<u>A</u>GNNNNNNTCG−3′ 3′−TTTCNNNNNN<u>A</u>GC−5′	878 878	99.7 % 99.8 %

Table 2: Cognate sequence motifs for m6A modification. Methylated positions are underlined. Percentages of occurrences detected as methylated are reported as mean across the 29 sequenced strains.

3.2.2 Tetramer composition bias in *Serratia marcescens* genome

To investigate the potential link between adenosine methylation and regulation of gene expression, we searched for evidence of differential usage bias of the 5′-GATC-3′ target sequence of the adenosine methyltransferase in promoters and gene bodies. In practice, we compared the usage bias of this target sequence with the usage bias of all other nucleotide tetramers. The usage bias in a given genomic region is positive if a tetramer is more abundant than expected by chance, and negative if it is rarer. Overall, the range of tetramer usage bias in genes and promoters was larger when measured from a permutation approach (deviation ranging from -0.80 to +1.23) than from a Markov chain approach (-0.42 to 0.83). The correlation between values obtained from the two approaches was moderate (Spearman's $\rho = 0.30$), indicating that tetramer biases are strongly related to biases in dimers and trimers usage, such as codon usage bias, which are taken into account by the Markov chain approach but not by the permutation approach. Most of the variation in tetramer usage bias was positively correlated between genes and promoters (e.g. Spearman's $\rho = 0.58$ for Markov chain estimates), but some tetramers exhibited large differences between their in-gene and in-promoter usage biases (Figure 2). Remarkably, 5′-GATC-3′ showed one of the largest distortions in usage bias between gene and promoter regions among all tetramers, in both approaches. The 5′-GATC-3′ tetramer was more abundant in genes and rarer in promoters than expected by chance, even when taking into account biases in dimer and trimer usage (Figure 2).

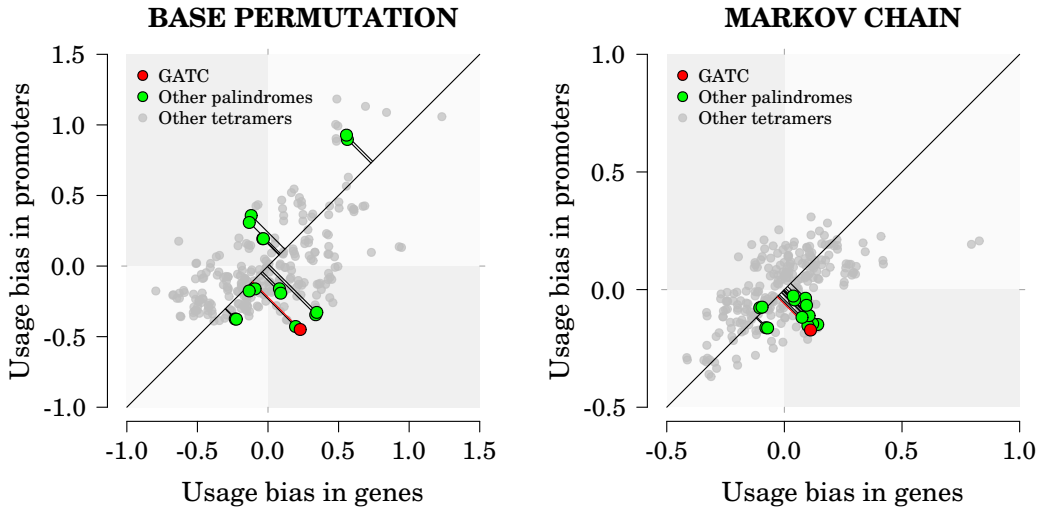


Figure 2: Tetramer usage bias in genes and promoters. Left side panel, tetramer usage bias determined from observed frequencies of single nucleotides (permutation approach). Right side panel, tetramer usage bias taking into account the observed frequencies of dimers and trimers (Markov chain approach). A positive usage bias means that a given tetramer is observed more frequently than expected by chance. The distance between an observation and its projection on the identity line (shown for palindromic tetramers) shows how imbalanced is the usage bias between promoters and genes: observations below the identity line represents tetramers which are rare in promoters compared to genes, and vice-versa.

3.2.3 Genomic methylation profiles

We investigated the profiles of m4C and m6A methylated fractions at the boundaries between the promoter regions (which were defined as the regions immediately upstream of the leading CDS of the predicted operons) and the coding regions. From a base-composition perspective, GC content was lower in promoter regions. In coding regions, we observed a trimodal distribution of GC content in the three codon positions, which is a sign of codon usage bias (Figure 3, top panel). Concerning the methylation profiles, we did not observe any striking spatial pattern for average base m4C methylation in relation with operon and CDS structure: levels of m4C methylation were fairly stable around operon starts and were consistent with genome-wide average m4C levels (Figure 3, middle panel), even though a t-test comparing the average base m4C methylation $\pm$ 500 bp around the operon starts suggested a slightly higher average m4C methylation in the promoter region (0.407 %) than in the coding region (0.370 %)

($t = -4.95$, $df = 975.78$, $p < 0.001$). On the other hand, we observed that levels of adenosine methylation into m6A clearly decreased before the translation start site (Figure 3, bottom panel), implying that adenosine methylation in promoters might interfere with transcription. After translation start adenosine methylation levels quickly returned to background levels in the operon coding region. While we cannot determine if low m6A methylation levels are a cause or a consequence of transcription initiation, these observations suggest that adenosine methylation (m6A) is functionally related to transcription in *S. marcescens* while m4C is not. This does not preclude that other cytosine modifications such as m5C might have a role in transcription regulation: for example, the average IPD ratio profile of cytosine also shows a decrease before the translation start site of the leading CDS, which could be a side effect of the m6A methylation profile in this region but could also indicate that cytosine modifications other that m4C are less frequent in these regions (Supplementary Figure S5).

3.3 Methylated positions (m6A) and methylated regions (m4C) of interest

As mentioned above, the vast majority of adenosines present in GATC motifs were detected methylated in all sequenced strains (99.7 %). Moreover, these methylated bases had consistently high methylation fractions, with almost all GATC adenosines being close to full methylation. However, we identified 907 GATC adenosines not fully methylated ("low-meth m6A") using our filtering criteria, i.e. 1.2 % of adenosines in GATC, which were located in 458 distinct GATC palindromes. Based on the genomic distribution of all GATC motifs, these low-meth m6A were more frequent than expected in promoter regions ($\chi^2 = 530.49$, $df = 1$, $p < 0.001$) and rarer than expected in operons ($\chi^2 = 1674.5$, $df = 1$, $p < 0.001$).

The detection of m4C methylated regions (i.e. clusters of m4C positions investigated with the kernel density approach) yielded 167 segments distributed across the genome. The majority of those segments (75 %) comprised two or three of the density

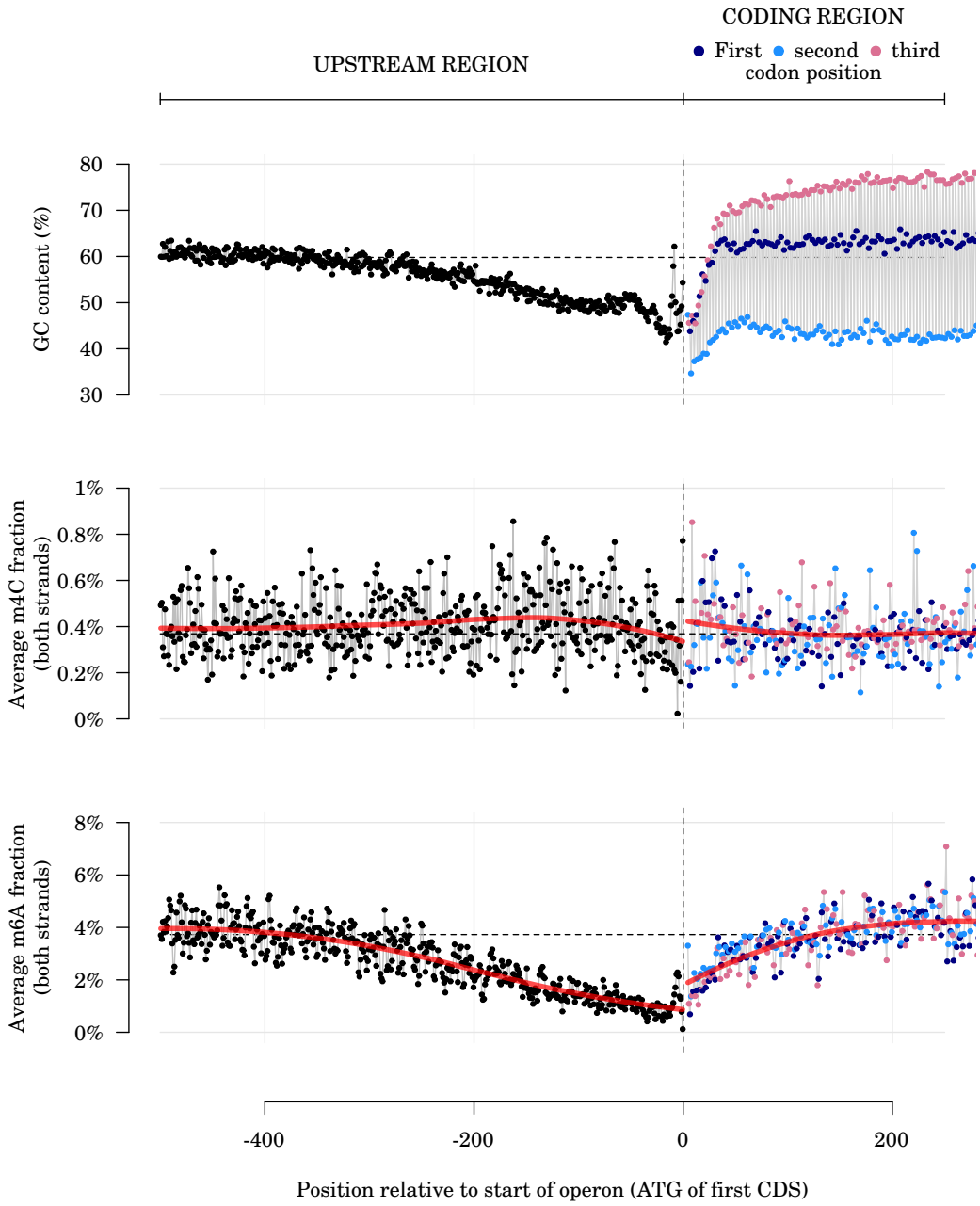


Figure 3: Profiles of nucleotide composition and average methylated fractions for m6A and m4C around the start positions of operon-leading CDS. Thick red line: LOESS regression (span = 0.75). Vertical dashed lines show the limit between upstream non-coding regions and the first codon of the leading CDS of predicted operons. Plotted values are averaged over each position relative to the leading CDS initiation codon based on operons predicted in the reference genome. Horizontal dashed lines show the genome-wide average values. Values for the three first bases on the coding sequences (usually ATG) are dropped from the plot to keep the y-scale reasonably narrow.

estimation cells, which correspond to genome stretches 79 and 118 bp long. The largest segments were 8 cell long (313 bp). About 11 % of the cytosines in those segments were detected as m4C in at least one sequenced strain, compared to a genome-wide average of 2.5 %. Those m4C segments were not evenly distributed between operon and promoter regions, even when taking into account GC-content genomic distribution: like low-meth m6A positions, m4C segments were more frequent than expected in promoter regions ($\chi^2 = 16.5$, $df = 1$, $p < 0.001$) and rarer than expected in operon regions ($\chi^2 = 390.7$, $df = 1$, $p < 0.001$).

Association between genetic mutations and epigenetic changes. We found tentative evidence of association between genetic mutations and m6A methylation changes only for haplotype *a* (which encompasses 11 mutations), and between m4C methylation changes and mutation 39 (Figure 4). In both cases, the number of epiloci associated with the genetic change was limited: using an uncorrected *p*-value threshold of $p < 0.01$ for the associations which were outside the 95 % *p*-value inflation envelope from permuted datasets, 28 out of 907 m6A epiloci were associated with haplotype *a* and 17 out of 167 m4C epiloci were associated with mutation 39. No relationship between distance from the epiloci to the mutation locus and *p*-value of association was observed. All in all, this provides very little support for a genetic control of epigenetic changes by one or a few major loci in our dataset. Additionally, given that haplotype *a* is likely to be the result of a genetic lineage pre-existing the start of the evolution experiment, we adopted a conservative approach when investigating the methylation changes happening during the evolution experiment and removed all m6A epiloci associated with haplotype *a* (uncorrected *p*-value < 0.01) from downstream analyses.

Methylation changes during experimental evolution. We investigated the changes in methylation for the m6A and m4C epiloci of interest across strains by first examining the per-epiloci variability of methylated fraction along the genome, calculated as the standard deviation of methylated fractions observed for all sequenced strains for each

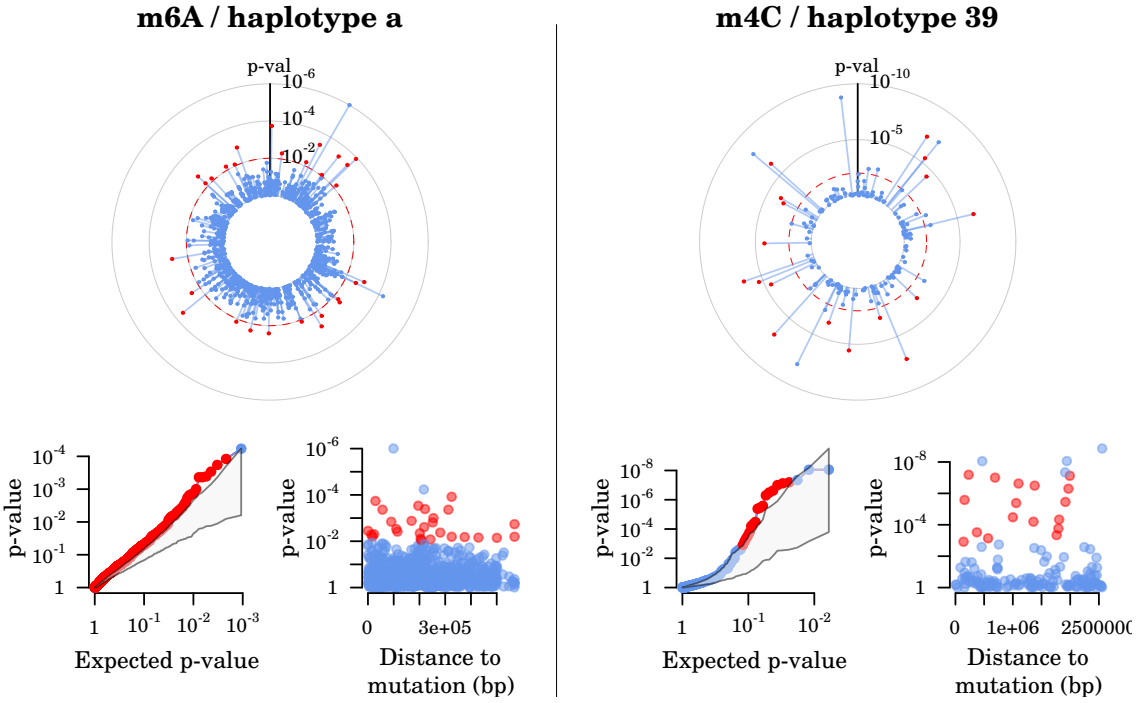


Figure 4: Association between genetic variants and epigenetic changes. Left, association between m6A and haplotype *a*; right, association between m4C and locus 39. Circular diagram: genomic map of the epiloci and their associated p-values, with *oriC* located at the top of the map. Red points indicate an epiloci for which the observed p-value is < 0.01 and is above the 95 % inflation permutation envelope. Bottom left: comparison between inflation for observed p-values and the 95 % (one-tailed) envelope of inflation for p-values from permuted datasets. Bottom right: distribution of epiloci p-values in relation to their genomic distance to the genetic mutation. For the haplotype *a*, the distance is the distance to the closest mutation. The plots for the association between m6A or m4C and the other genetic loci are presented in Supplementary Figure S6.

epiloci. For m4C, methylated fractions were more variable inside operons compared to outside (Wilcoxon test p-value $= 0.043$) but there was no difference between inside and outside promoter regions (p-value $= 0.52$). For m6A, methylated fractions were more variable inside promoter regions compared to outside (Wilcoxon test p-value < 0.001) and were less variable inside operons compared to outside (p-value < 0.001).

In order to determine if methylation changes happened during the evolution experiment, we investigated the methylation changes for the m6A and m4C epiloci of interest across strains by classifying those epiloci into low variability and high variability sets. The high variability set was further divided into loci with continuous and discontinous methylation variation. This was done using a simple heuristic based on their observed methylated fraction profiles. For each epilocus, we sorted the methylated fractions observed in the 28 evolved strains and calculated (i) the range of methylated fractions, (ii) the largest increment between successive methylated fractions and (iii) the ratio between this largest increment and the methylated fraction range (Supplementary Figure S7, panels A and B). We considered epiloci for which the methylated fraction range was less than 0.2 to be *low variability* epiloci. High variability epiloci with a methylation fraction range greater than 0.2 were further split into *continuous* epiloci, i.e. epiloci showing a relatively smooth gradient from lowest to highest methylated fractions and *discontinuous* epiloci, i.e. epiloci for which the largest methylated fraction increment was greater than 0.3 times the methylated fraction range (Supplementary Figure S7, panels C, D and E). For *discontinuous* epiloci, strains were assigned to a "low" or "high" epiallele based on their methylated fraction relative to the location of the largest increment. As discontinuous epiloci seemed more likely to have an effect on gene regulation, we used these when examining the distribution of epigenetic changes across strains and evolutionary treatment.

The procedure to identify *discontinuous* epiloci was performed separately for m4C and m6A. The chosen heuristic resulted in 43 (26 %) of the 167 m4C epiloci and 256 (28 %) of the 907 m6A epiloci being assigned to this category (Supplementary Figure

S7). Six epiloci out of those 256 were found associated with haplotype *a* earlier and thus removed from downstream analyses. The vast majority of those epiloci had their minor epialleles present in only one strain, both for m6A and m4C (Figure 5 A,D). When examining only epiloci with minor epialleles shared by at least two strains, there was no clear association between the occurrence of epigenetic changes and the evolutionary treatment (Figure 5 B,E and C,F).

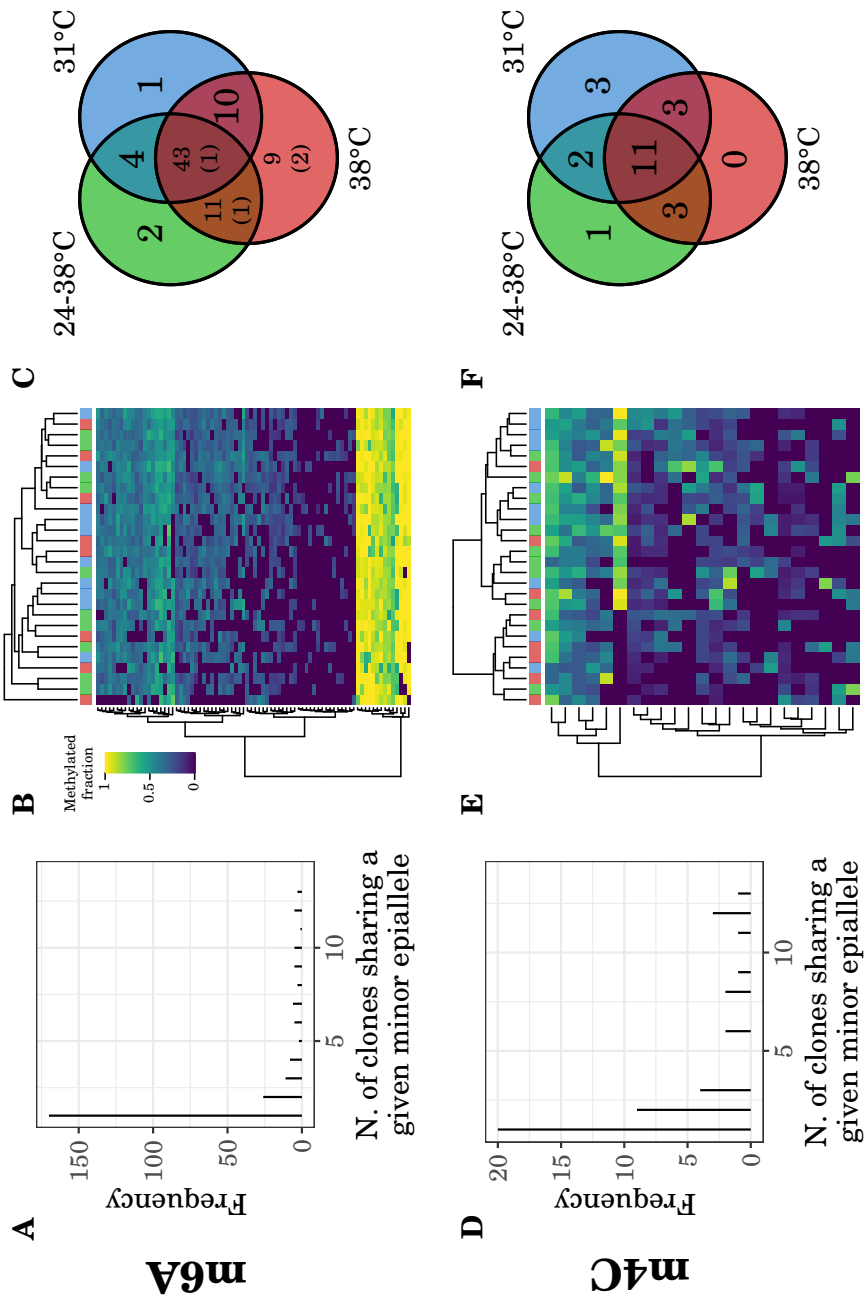


Figure 5: Frequency and distribution of epigenetic changes across strains. Only epiloci assigned to the *discontinuous* category are presented (see Supplementary Figure S7). A, D: distributions of shared epigenetic changes across strains. B, E: heatmap of the methylated fraction for epiloci shown in A, D with minor alleles in at least two strains. Strains are color-coded by evolutionary treatment. C, F: Venn diagrams showing shared minor alleles for epiloci from B, E across evolutionary treatments. For m6A, numbers in parentheses indicate the number of epiloci which were associated with phenotypic traits in Bruneaux et al. (2019).

3.4 Decomposition of phenotypic variance

To evaluate how much genetic and epigenetic changes contributed to phenotypic variation, we used a random effect model incorporating similarity matrices based on genotypes and epigenotypes. Our approach is conceptually equivalent to estimating a genetic heritability and an epigenetic heritability for each phenotypic trait. Comparing the phenotypic variance decomposition between a purely random-effect model (i.e. including haplotype *a* in the genetic similarity matrix, Figure 6) and a model including haplotype *a* as a fixed effect (Supplementary Figure S8) enabled us to estimate the overall effect of genetic variants, regardless of their origin, and the effect of genetic variants appeared during the evolution experiment, respectively (assuming that haplotype *a* was a genetic lineage present at the beginning of the evolution experiment).

When all genetic loci were included into the genetic similarity matrix (including haplotype *a*), estimates of genetic and epigenetic variances varied between traits, but some general patterns can be derived from the posterior distributions of r^2_{GEN}, r^2_{EPI} and r^2_{RES} (Figure 6A): for virulence and phage activation traits (first two columns in Figure 6A), the largest component was the genetic variance, followed by the epigenetic variance which tended to be larger than the residual variance. For temperature-related traits and co-selected traits (last two columns in Figure 6A), the uncertainty was generally larger, with a possibly equal contribution of genetic and epigenetic variances and again a smaller contribution of residual variance. When including a fixed effect of haplotype *a*, the remaining genetic heritability was decreased in favor of an increased epigenetic heritability for virulence and phage activation traits, but the previous patterns remained mostly unchanged for other traits (Supplementary Figure S8, panel A)

To estimate the joint and disjoint proportions of phenotypic variance explained by evolutionary treatment, genetics and epigenetics, we used the total proportions of phenotypic variance explained by all possible combinations of the corresponding similarity matrices in order to calculate point estimates for each of the cells in the Venn

diagrams presented in Figure 6B. Importantly, those estimates did not carry over the uncertainty in proportions of explained variance from the MCMC posteriors, which are usually large in our models. Caution must thus be exercised when interpreting those estimates, and here we only examine the largest numerical differences between those values. When haplotype *a* was included in the genetic similarity matrix, the largest variance component for most traits was the joint contribution of genetics and epigenetics which often accounted for more than half of the phenotypic variance. This joint contribution of genetics and epigenetics was itself split into a large fraction (often more than half) overlapping with the treatment variance. The disjoint contributions of treatment, genetics and epigenetics were much smaller and hard to estimate reliably. This large overlap between the genetic and epigenetic contributions to the phenotypic variances is consistent with the relatively large correlations between genetic, epigenetic and phenotypic distances estimated using Mantel's R (Supplementary Figures S12 and S13). When haplotype *a* was included as a fixed effect, the joint contribution of genetics and epigenetics in variance decomposition decreased for virulence and phage activation traits but remained important, while other phenotypic variance decomposition patterns remained mostly unchanged overall (Supplementary Figure S8, panel B)

Additionally, to characterize the specificity of the evolutionary trajectories among evolutionary treatments, we applied the same variance decomposition approach as shown in Figure 6 using one pair of treatments at a time instead of the full dataset (Supplementary Figure S9). The overall patterns in variance decomposition for growth rates and yields are very similiar whether the full dataset or any pair of treatments is used. However, for virulence and phage activation traits, the genetic heritability seems to play a smaller role when comparing 31 °C and 24–38°C treatments than when comparing any of those treatments with the 38 °C treatment (Supplementary Figure S9, A panels). The decrease of the genetic component is also observed in the Venn diagram decomposition (Supplementary Figure S9, B panels), most markedly through a decrease in the joint contribution of genetics, epigenetics and evolutionary treatment.

Overall, these results suggest a larger role of genetics in determining the differentiation of virulence and phage activation traits between the 38 °C and the other treatments, compared to between 31 °C and 24–38°C treatments. When comparing those results with results obtained using haplotype *a* as a fixed effect, the relative importance of genetics in comparisons of the 38 °C treatment with the other treatments tended to decrease, supporting the hypothesis that part of the differentiation of the 38 °C treatment compared to the others is due to the selection of the alternate lineage containing haplotype *a* (Supplementary Figure S10).

Finally, we also performed phenotypic variance decomposition for the full dataset after taking into account the fixed effect of the genetic locus 39 which was found to be potentially associated with epigenetic changes for m4C. Removing the fixed effect of locus 39 had overall very little effect on phenotypic variance decomposition (Supplementary Figure S11), which is consistent with the fact that it was associated with m4C changes but not m6A changes.

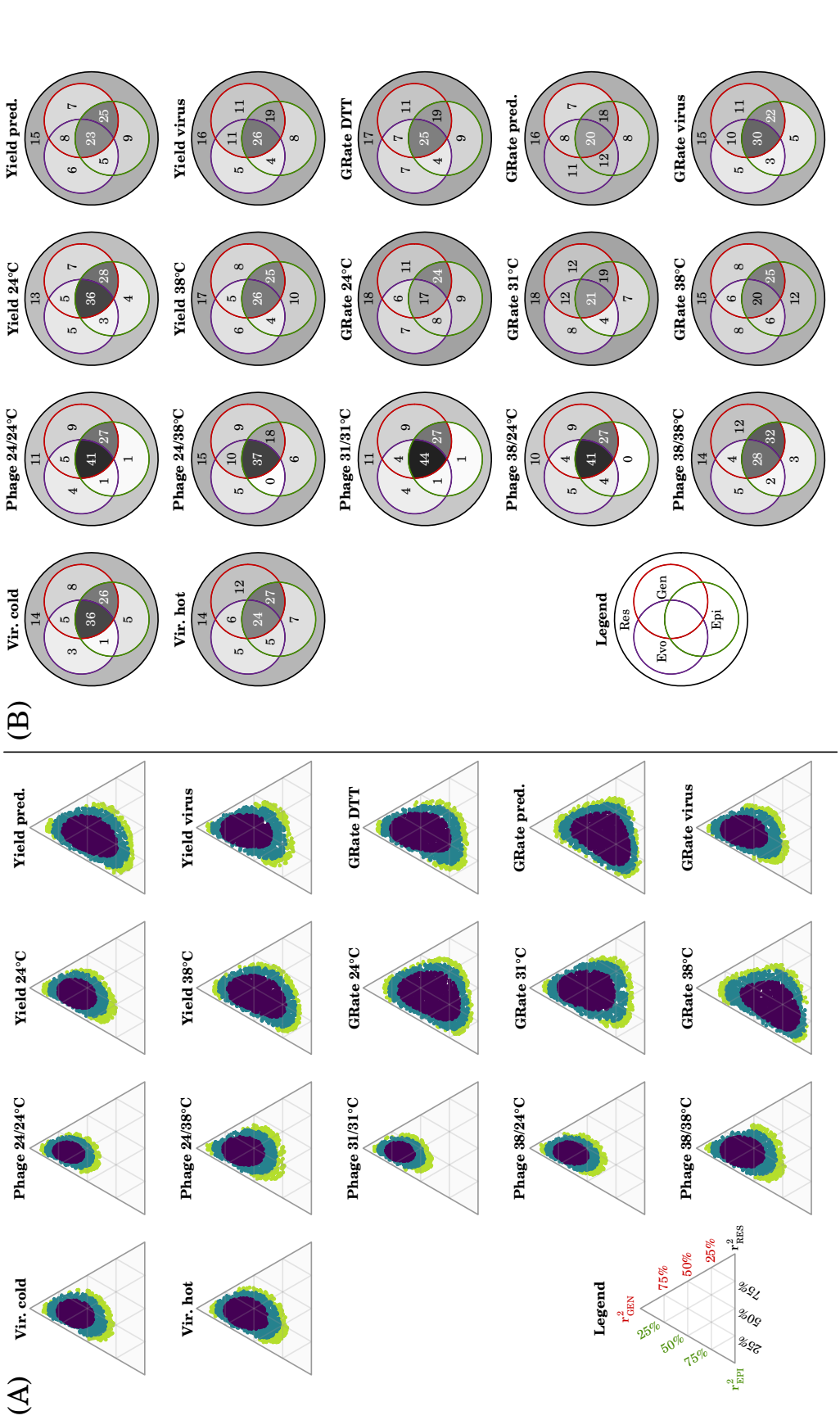


Figure 6: Variance decomposition for each phenotypic trait. (A), 50 %, 80 % and 90 % credible surfaces for the estimates of the genetic, epigenetic and residuals components in total phenotypic variance. (B), point estimates for joint and disjoint variance components described by evolutionary treatment, genetics and epigenetics (in % of total phenotypic variance). Vir. cold and Vir. hot, strain virulence in two temperatures; Phage 24/24 °C and similar, phage activation under five assay temperature treatments; Yield and Growth Rate at different assay temperatures and under novel conditions (in presence of predator, virus and DTT).

4 Discussion

We have shown that substantial variation in methylation arose during experimental evolution in *Serratia marcescens*. The genomic distribution of methylated positions suggested a role of modified adenosines (m6A) in the regulation of gene expression, but not a role of modified cytosines (m4C). Phenotypic variance was in large part (about 40-50% or more) described by a shared contribution of genetics and epigenetics, while evolutionary treatment explained about half of this shared variance. Despite this strong shared component between genetics and epigenetics, only little evidence of genetic control of epigenetic changes was found. Both potentially environmentally-induced variation (changes shared across evolved strains) and spontaneous epigenetic variation (strain-specific changes) were observed in our data. Spontaneous epigenetic variation included variation that was likely neutral along with some potentially adaptive changes.

The function of m6A was suggested by the genomic distribution of the corresponding GATC methylation motif, which pointed to different evolutionary constraints on m6A methylation between promoters and gene bodies: the GATC motif was less frequent in promoter regions and more frequent in gene bodies than expected by chance. Oshima et al. (2002) suggested that GATC in upstream regions could modulate gene expression by interacting with some regulatory proteins. However, Riva et al. (2004) argued that the regulation of expression was due to clusters of GATC situated inside the coding regions, which would affect DNA stability and thus expression based on their methylation status. In the case of *S. marcescens*, the usage bias against GATC in promoters supports a possible selection pressure to preserve transcription regulation in those regions from disturbance due to m6A.

Epigenetic changes did not appear to be under genetic control, as only a small proportion of the methylation variation was associated with genetic mutations. While our statistical power to detect association is limited with our data since there is no segregation among the bacterial clones, extensive genetic control would require assuming that each genetic mutation controls multiple different epigenetic changes in order

for genetic mutations to explain the observed epigenetic variation. In the cases where we observed an association between a mutation and epigenetic changes, a single mutation or haplotype was indeed tentatively associated with multiple epigenetic changes. Moreover, there was no relationship between distance to the mutation and the epigenetic changes, indicating that genetic control over long distance is plausible, mediated perhaps by indirect effects of the mutation. However, considering that we cannot distinguish between a mutation inducing a methylation change and a methylation change hitchhiking with an adaptive genetic mutation, it does not seem plausible that even a majority of the observed methylation changes were induced by genetic mutations. Furthermore, in the case of m6A, association was only observed with haplotype *a* and could be therefore be due to the shared history of those epiloci with haplotype *a* prior to the initiation of the evolution experiment, if we assume that haplotype *a* was part of some unexpected standing genetic variation in the ancestor culture at the time of inoculation of the replicate experimental populations. Low amount of methylation changes that seems to be under genetic control is the same observation made by Kronholm et al. (2017), with *Chlamydomonas*-algae. However, it is in contrast to studies of natural populations of plants, where generally most methylation changes seem to be under genetic control (Dubin et al., 2015; Hagmann et al., 2015). Plants have high rates of spontaneous methylation change (van der Graaf et al., 2015), so it remains to be seen what can explain this discrepancy.

Epigenetic changes can be either spontaneous (van der Graaf et al., 2015) or be induced by the environment (Jiang et al., 2014; Wibowo et al., 2016). Any epigenetic changes that occur in multiple different clones can in principle be changes that are induced by their common environment or spontaneous changes that were fixed by natural selection in multiple populations, thus reflecting paralllel evolution. It is also possible that some epigenetic loci have extremely high forward and back mutations rates, so that some polymorphism is always present. We could classify epiloci of interest in our dataset into three main categories based on their methylated fraction profiles: *low*

variability epiloci, *continuous* epiloci, and *discontinuous* epiloci. Both *continuous* and *discontinuous* epiloci can potentially explain phenotypic diversity, with the *continuous* epiloci possibly having higher rates of change and acting as control knobs of gene regulation at the population level and the *discontinuous* having slower rate of change and acting as gene regulatory switches. When considering the *discontinuous* epiloci, and after discarding any epiloci which might be associated with haplotype *a*, we only observed very few changes that were only shared by clones coming from a particular temperature, but instead the majority of shared methylation changes were shared by some clones from all three treatments. This suggests that either these changes reflect plastic changes in response to the laboratory environment but not to the temperature treatment itself, or that these loci have high rates of change, which could possibly reflect some sort of epigenetic bet-hedging mechanism. Lastly, the majority of observed epigenetic changes for both m6A and m4C occurred in only one or two clones. The most likely explanation is that these were spontaneous methylation changes. Due to the nature of our experiment we cannot investigate whether these rare changes somehow affect the phenotype or are neutral.

Our main objective was to determine the relative contributions of genetics and epigenetics to adaptation in rapidly changing environments, and to what extend those contributions are independent from each other. The decomposition of phenotypic variance showed that residual variance unexplained by either genetics or epigenetics was generally small, and that the shared contribution of genetic and epigenetic variances was generally large for all traits considered. Evolutionary treatment contributed to about half of this shared genetic or epigenetic variance suggesting that, even though treatment had an important effect on evolutionary trajectories during the experimental evolution, contingency was also an important factor. When taking into account the fixed effect of haplotype *a* on phenotypes, and thus controlling for the effect of potential standing genetic variation at the start of the evolution experiment, the joint contribution of genetics and epigenetics to phenotypic variance was decreased but the shared

contribution of genetics, epigenetics and evolutionary treatment to phenotypic variance remained large. This suggests that some of the epigenetic contribution overlaps with the genetic one without this being due to an indirect effect of genetics on phenotype through a genetic control of epigenetic modifications or to a shared history of m6A epiloci associated with haplotype *a* prior to the initiation of the evolution experiment.

Finally, we can interpret our results in the light of the three evolutionary treatments used in the initial experiment: 31 °C, 38 °C and 24–38°C. While we did not find any evidence of more frequent epigenetic changes in any particular treatments, the 38 °C treatment was the only treatment in which the haplotype *a*, which consists of 11 distinct loci, was found (in 5 out of 8 strains). The haplotype *a* was also found in the reference strain from which the ancestor used for the evolution experiment was derived. The probability of 11 mutations arising independently and in succession in several strains in our dataset is quite low given the duration of the evolution experiment. This haplotype thus suggests that the ancestor culture used to initiate all the populations of the experiment might have exhibited some genetic diversity in relation with haplotype *a*, possibly due to cell aggregation occurring during the preparation of the ancestor clone. No sign of this diversity is observed in the sequenced strains from 31 °C and 24–38°C, indicating that it was driven to low frequencies or extinction in 31 °C and 24–38°C conditions while the 38 °C environment allowed for more diverse evolutionary trajectories. The fact that the 38 °C environment is genetically different from the other two is also apparent from the phenotypic variance decompositions performed on pairs of treatments: the genetic heritability is lower when considering the 31 °C/24–38°C pair. Overall, those results suggests that evolving at lower average temperature imposed stronger selective constraints on the genetic variants in our experimental organism while higher temperature allowed for more diverse genetic trajectories, and conversely that epigenetics were more important for differentiation between the constant and fluctuating conditions at the same average temperature.

Epigenetic changes between the evolved strains were moderate at best. However, it

712 is important to note that our experimental design and protocol which included growing the evolved clones in common conditions for a few generations before DNA extraction 714 for sequencing precludes detecting rapidly reset epigenetic changes.

In conclusion, we have shown that substantial epigenetic variation in adenosine 716 methylation exists in *Serratia marcescens* and that some of this variation is likely to have functional consequences and to be adaptive. The role of epigenetic changes may be 718 mediation of initial plastic responses, or just generation of variation as a bet-hedging strategy. Furthermore, at least part of the variation in methylation is likely to be 720 neutral. Genetic variants from unexpected pre-existing standing genetic variation in our experiment seem to be responsible for the majority of divergent adaptation between 722 the 38 °C treatment and the others, but the large shared contribution of genetic and epigenetic variation to phenotype even when taking into account haplotype *a* suggests 724 that genetics and epigenetics both exert a strong control on bacterial traits.

Acknowledgements. We acknowledge the Academy of Finland (Project 278751) 726 and the Centre of Excellence in Biological Interactions for funding and facilities, and CSC-IT center for Science for computational resources.

References

Adam, M., Murali, B., Glenn, N. O., and Potter, S. S. Epigenetic inheritance based evolution of antibiotic resistance in bacteria. *BMC Evolutionary Biology*, 8(1):52, Feb 2008. ISSN 1471-2148. doi: 10.1186/1471-2148-8-52.

Angiuoli, S. V. and Salzberg, S. L. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics*, 27(3):334–342, February 2011. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btq665.

Atack, J. M., Srikhanta, Y. N., Fox, K. L., Jurcisek, J. A., Brockman, K. L., Clark, T. A., Boitano, M., Power, P. M., Jen, F. E.-C., McEwan, A. G., Grimmond, S. M., Smith, A. L., Barenkamp, S. J., Korlach, J., Bakaletz, L. O., and Jennings, M. P. A biphasic epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable *Haemophilus influenzae*. *Nat Commun*, 6:7828, July 2015.

Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., and Weigel, D. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature*, 480:245–249, 2011.

Bird, A. DNA methylation patterns and epigenetic memory. *Genes & Development*, 16(1):6–21, January 2002. ISSN 0890-9369. doi: 10.1101/gad.947102.

Blow, M. J., Clark, T. A., Daum, C. G., Deutschbauer, A. M., Fomenkov, A., Fries, R., Froula, J., Kang, D. D., Malmstrom, R. R., Morgan, R. D., Posfai, J., Singh, K., Visel, A., Wetmore, K., Zhao, Z., Rubin, E. M., Korlach, J., Pennacchio, L. A., and Roberts, R. J. The Epigenomic Landscape of Prokaryotes. *PLOS Genet*, 12(2): e1005854, February 2016. ISSN 1553-7404. doi: 10.1371/journal.pgen.1005854.

Braaten, B. A., Nou, X., Kaltenbach, L. S., and Low, D. A. Methylation patterns in pap regulatory DNA control pyelonephritis-associated pili phase variation in E. coli. *Cell*, 76(3):577–588, February 1994. ISSN 0092-8674, 1097-4172. doi: 10.1016/0092-8674(94)90120-1.

Bruneaux, M., Ashrafi, R., Kronholm, I., Örmälä Odegrip, A.-M., Galarza, J. A., Chen, Z., Mruthyunjay, K. S., and Ketola, T. Environmentally triggered evolutionary cascade across trophic levels in an experimental phage-bacteria-insect system. *submitted to BioRxiv*, 2019.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. BLAST+: architecture and applications. *BMC Bioinformatics*, 10 (1):421, December 2009. ISSN 1471-2105. doi: 10.1186/1471-2105-10-421.

Casadesús, J. and Low, D. Epigenetic Gene Regulation in the Bacterial World. *Microbiology and Molecular Biology Reviews*, 70(3):830–856, September 2006. ISSN 1092-2172, 1098-5557. doi: 10.1128/MMBR.00016-06.

Casselli, T., Tourand, Y., Scheidegger, A., Arnold, W. K., Proulx, A., Stevenson, B., and Brissette, C. A. DNA Methylation by Restriction Modification Systems Affects the Global Transcriptome Profile in Borrelia burgdorferi. *Journal of Bacteriology*,

200(24):e00395–18, December 2018. ISSN 0021-9193, 1098-5530. doi: 10.1128/JB.00395-18.

Charlesworth, D., Barton, N. H., and Charlesworth, B. The sources of adaptive variation. *Proceedings of the Royal Society of London B: Biological Sciences*, 284(1855), 2017. ISSN 0962-8452. doi: 10.1098/rspb.2016.2864.

Clark, T. A., Lu, X., Luong, K., Dai, Q., Boitano, M., Turner, S. W., He, C., and Korlach, J. Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biology*, 11(1):4, January 2013. ISSN 1741-7007. doi: 10.1186/1741-7007-11-4.

Danchin, E., Charmantier, A., Champagne, F. A., Mesoudi, A., Pujol, B., and Blanchet, S. Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. *Nat Rev Genet*, 12(7):475–486, July 2011. ISSN 1471-0056.

Day, T. and Bonduriansky, R. A unified approach to evolutionary consequences of genetic and nongenetic inheritance. *American Naturalist*, 178:E18–E36, 2011.

Dubin, M. J., Zhang, P., Meng, D., Remigereau, M.-S., Osborne, E. J., Paolo Casale, F., Drewe, P., Kahles, A., Jean, G., Vilhjálmsson, B., Jagoda, J., Irez, S., Voronin, V., Song, Q., Long, Q., Rätsch, G., Stegle, O., Clark, R. M., and Nordborg, M. Dna methylation in arabidopsis has a genetic basis and shows evidence of local adaptation. *eLife*, 4:e05255, 2015. doi: 10.7554/eLife.05255.

Flyg, C., Kenne, K., and Boman, H. G. Insect pathogenic properties of Serratia marcescens: phage-resistant mutants with a decreased resistance to Cecropia immunity and a decreased virulence to Drosophila. *Journal of General Microbiology*, 120 (1):173–181, September 1980. ISSN 0022-1287. doi: 10.1099/00221287-120-1-173.

Gao, F., Luo, H., and Zhang, C.-T. DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. *Nucleic Acids Research*, 41(D1):D90–D93, October 2012. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gks990.

Grimont, P. A. D. and Grimont, F. The Genus Serratia. *Annual Review of Microbiology*, 32(1):221–248, 1978. doi: 10.1146/annurev.mi.32.100178.001253.

Hadfield, J. D. Mcmc methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33(2):1–22, 2010.

Hagmann, J., Becker, C., Müller, J., Stegle, O., Meyer, R. C., Wang, G., Schneeberger, K., Fitz, J., Altmann, T., Bergelson, J., Borgwardt, K., and Weigel, D. Century-scale Methylome Stability in a Recently Diverged Arabidopsis thaliana Lineage. *PLoS Genetics*, 11(1):e1004920, January 2015. ISSN 1553-7404. doi: 10.1371/journal.pgen.1004920.

Herman, J. J. and Sultan, S. E. DNA methylation mediates genetic variation for adaptive transgenerational plasticity. *Proceedings of the Royal Society of London B: Biological Sciences*, 283(1838):20160988, 2016. ISSN 0962-8452. doi: 10.1098/rspb.2016.0988.

Iyer, L. M., Zhang, D., and Aravind, L. Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification. *BioEssays*, 38(1):27–40, 2016. doi: 10.1002/bies.201500104.

Jablonka, E. and Raz, G. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Quarterly Review of Biology*, 84:131–176, 2009.

Jiang, C., Mithani, A., Belfield, E. J., Mott, R., Hurst, L. D., and Harberd, N. P. Environmentally responsive genome-wide accumulation of de novo arabidopsis thaliana mutations and epimutations. *Genome Research*, 24(11):1821–1829, 2014. doi: 10.1101/gr.177659.114.

Ketola, T., Mikonranta, L., Zhang, J., Saarinen, K., Örmälä, A.-M., Friman, V.-P., Mappes, J., and Laakso, J. Fluctuating temperature leads to evolution of thermal generalism and preadaptation to novel environments. *Evolution*, 67:2936–2944, 2013.

Klironomos, F., Berg, J., and Collins, S. How epigenetic mutations can affect genetic evolution: Model and mechanism. *BioEssays*, 35:571–578, 2013.

Kronholm, I. Adaptive evolution and epigenetics. In Tollefsbol, T. O., editor, *Handbook of epigenetics: the new molecular and medical genetics*, pages 427–438. Academic Press, London, 2nd edition, 2017. ISBN 978-0-12-805388-1.

Kronholm, I. and Collins, S. Epigenetic mutations can both help and hinder adaptive evolution. *Molecular Ecology*, 25:1856–1868, 2016. ISSN 1365-294X. doi: 10.1111/mec.13296.

Kronholm, I., Bassett, A., Baulcombe, D., and Collins, S. Epigenetic and genetic contributions to adaptation in *Chlamydomonas*. *Molecular Biology and Evolution*, 34(9):2285–2306, 2017.

Legendre, P. and Legendre, L. *Numerical ecology*. Developments in environmental modelling. Elsevier Science, 1998. ISBN 978-0-08-053787-0.

Legendre, P., Fortin, M.-J., and Borcard, D. Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution*, 6(11):1239–1247, 2015. ISSN 2041-210X. doi: 10.1111/2041-210X.12425.

Lenski, R. E., Rose, M. R., Simpson, S. C., and Tadler, S. C. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. *The American Naturalist*, 138(6):1315–1341, 1991. ISSN 0003-0147.

Luna, E. and Ton, J. The epigenetic machinery controlling transgenerational systemic acquired resistance. *Plant Signaling & Behavior*, 7(6):615–618, June 2012. ISSN 1559-2324.

Ma, C., Niu, R., Huang, T., Shao, L.-W., Peng, Y., Ding, W., Wang, Y., Jia, G., He, C., Li, C.-Y., He, A., and Liu, Y. N6-methyldeoxyadenine is a transgenerational epigenetic signal for mitochondrial stress adaptation. *Nature Cell Biology*, 21(3):319–327, 2019. ISSN 1476-4679.

Mantel, N. and Valand, R. S. A technique of nonparametric multivariate analysis. *Biometrics*, 26(3):547–558, September 1970. ISSN 0006-341X.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H. *vegan: Community Ecology Package*, 2019. R package version 2.5-4.

Oshima, T., Wada, C., Kawagoe, Y., Ara, T., Maeda, M., Masuda, Y., Hiraga, S., and Mori, H. Genome-wide analysis of deoxyadenosine methyltransferase-mediated control of gene expression in Escherichia coli. *Molecular Microbiology*, 45(3):673–695, August 2002. ISSN 0950-382X.

Pride, D. T., Meinersmann, R. J., Wassenaar, T. M., and Blaser, M. J. Evolutionary implications of microbial genome tetranucleotide frequency biases. *Genome Research*, 13(2):145–158, February 2003. ISSN 1088-9051. doi: 10.1101/gr.335003.

Ratel, D., Ravanat, J.-L., Berger, F., and Wion, D. N6-methyladenine: the other methylated base of DNA. *Bioessays*, 28(3):309–315, March 2006. ISSN 0265-9247. doi: 10.1002/bies.20342.

Richards, C. L., Alonso, C., Becker, C., Bossdorf, O., Bucher, E., Colome-Tatche, M., Durka, W., Engelhardt, J., Gaspar, B., Gogol-Doring, A., Grosse, I., van Gurp, T. P., Heer, K., Kronholm, I., Lampei, C., Latzel, V., Mirouze, M., Opgenoorth, L., Paun, O., Prohaska, S., Rensing, S. A., Stadler, P., Trucchi, E., Ullrich, K., and Verhoeven, K. J. Ecological plant epigenetics: Evidence from model and non-model species, and the way forward. *Ecology Letters*, 20:1351–1364, 2017. doi: 10.1101/130708.

Riva, A., Delorme, M.-O., Chevalier, T., Guilhot, N., Hénaut, C., and Hénaut, A. The difficult interpretation of transcriptome data: the case of the GATC regulatory network. *Computational Biology and Chemistry*, 28(2):109–118, April 2004. ISSN 1476-9271. doi: 10.1016/j.compbiolchem.2003.12.004.

Rocha, E. P., Viari, A., and Danchin, A. Oligonucleotide bias in Bacillus subtilis: general trends and taxonomic comparisons. *Nucleic Acids Research*, 26(12):2971–2980, June 1998. ISSN 0305-1048.

Schmitz, R. J., Schultz, M. D., Lewsey, M. G., O'Malley, R. C., Urich, M. A., Libiger, O., Schork, N. J., and Ecker, J. R. Transgenerational epigenetic instability is a source of novel methylation variants. *Science*, 334:369–373, 2011.

Seshasayee, A. S. N. An assessment of the role of DNA adenine methyltransferase on gene expression regulation in E coli. *PloS One*, 2(3):e273, March 2007. ISSN 1932-6203. doi: 10.1371/journal.pone.0000273.

Stajic, D., Perfeito, L., and Jansen, L. E. T. Epigenetic gene silencing alters the mechanisms and rate of evolutionary adaptation. *Nature Ecology & Evolution*, pages –, 2019. ISSN 2397-334X.

Stephens, C., Reisenauer, A., Wright, R., and Shapiro, L. A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. *Proceedings of the National Academy of Sciences*, 93(3):1210–1214, February 1996. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.93.3.1210.

Sánchez-Romero, M. A., Cota, I., and Casadesús, J. DNA methylation in bacteria: from the methyl group to the methylome. *Current Opinion in Microbiology*, 25: 9–16, June 2015. ISSN 1369-5274. doi: 10.1016/j.mib.2015.03.004.

Taboada, B., Estrada, K., Ciria, R., and Merino, E. Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*, 34(23):4118–4120, December 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty496.

Thomson, C. E., Winney, I. S., Salles, O. C., and Pujol, B. A guide to using a multiple-matrix animal model to disentangle genetic and nongenetic causes of phenotypic variance. *PLOS ONE*, 13(10):e0197720, October 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0197720.

van der Graaf, A., Wardenaar, R., Neumann, D. A., Taudt, A., Shaw, R. G., Jansen, R. C., Schmitz, R. J., Colomé-Tatché, M., and Johannes, F. Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proceedings of the National Academy of Sciences*, 112(21):6676–6681, 2015. doi: 10.1073/pnas.1424254112.

Wang, J., Reddy, B. D., and Jia, S. Rapid epigenetic adaptation to uncontrolled heterochromatin spreading. *eLife*, 4:e06179, mar 2015. ISSN 2050-084X. doi: 10.7554/eLife.06179.

Wibowo, A., Becker, C., Marconi, G., Durr, J., Price, J., Hagmann, J., Papareddy, R., Putra, H., Kageyama, J., Becker, J., Weigel, D., and Gutierrez-Marcos, J. Hyperosmotic stress memory in *Arabidopsis* is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. *eLife*, 5:e13546, may 2016. ISSN 2050-084X. doi: 10.7554/eLife.13546.

Zheng, X., Chen, L., Xia, H., Wei, H., Lou, Q., Li, M., Li, T., and Luo, L. Transgenerational epimutations induced by multi-generation drought imposition mediate rice plant's adaptation to drought condition. *Scientific Reports*, 7:39843, January 2017.

5 Appendix

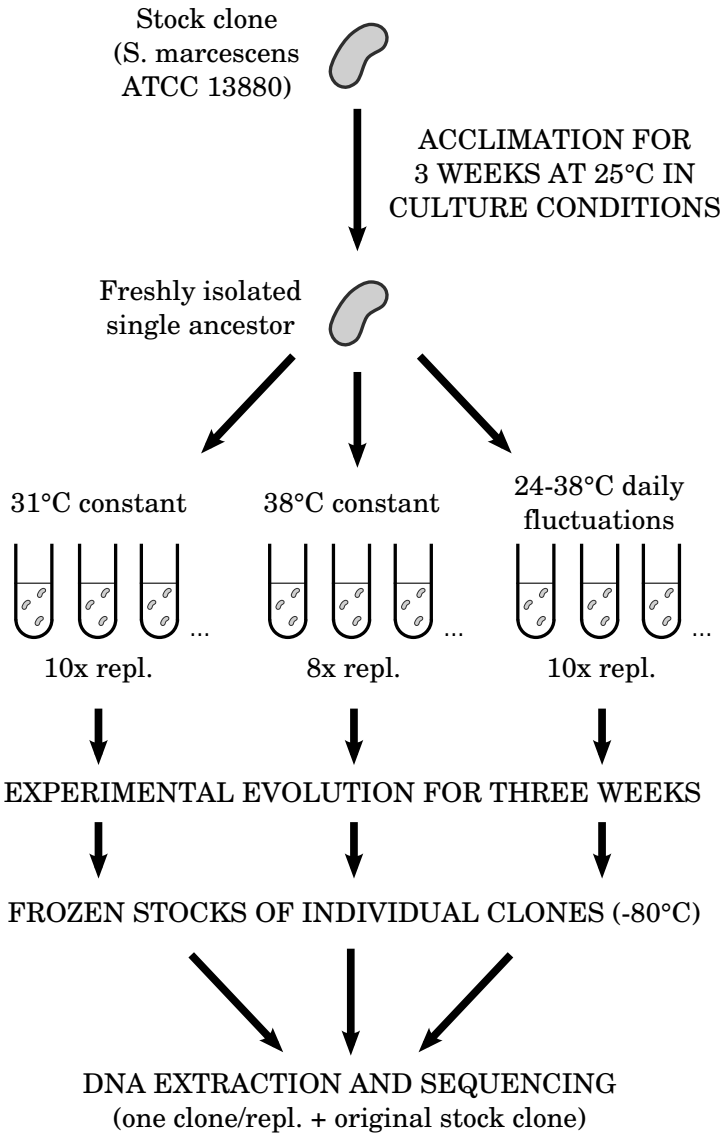
5.1 Supplementary tables

ID	Haplotype	Freq.	Pos. (bp)	Type	Region	Effect	Overlapping or closest (≤500bp) gene	
							Name	Function
1	f	1/28	31 753	indel	non-coding	-	-	-
2	a	*5/28	40 239	indel	non-coding	-	-	-
3	d	1/28	70 546	indel	non-coding	-	-	-
4		6/28	92 159	indel	non-coding	-	hypothetical protein	unknown
5	b	1/28	108 315	indel	non-coding	-	PTS beta-glucoside transporter	carbohydrate import
6	a	*5/28	131 841	SNP	CDS	non-syn.	cellulose biosynthesis protein BcsG	biofilm?
7		1/28	173 551	indel	non-coding	-	protoheme IX biogenesis protein HemY	heme metabolism
8	a	*5/28	328 601	SNP	CDS	non-syn.	condensation protein	non-ribosomal peptide synthesis
9		1/28	391 159	indel	CDS	frameshift	RNA chaperone Hfq	regulation of stress transcription factors
10	g	1/28	429 888	indel	non-coding	-	transcriptional regulator	similar to regulator of E. coli phage Mu
11		4/28	522 878	indel	CDS	frameshift	integrase	prophage DNA Integration/excision
12	b	1/28	751 961	indel	CDS	frameshift	DNA polymerase II	DNA elongation
13	b	1/28	770 532	indel	non-coding	-	2-isopropylmalate synthase	Leucine biosynthesis
14		1/28	914 534	indel	non-coding	-	hydroxyacylglutathione hydrolase	lactate metabolism/response to heat
15		1/28	979 119	indel	non-coding	-	-	-
16	f	1/28	1 093 517	indel	CDS	frameshift	competence protein ComEA	cell surface DNA binding
17		1/28	1 185 019	indel	CDS	frameshift	hypothetical protein	Unknown
18	c	1/28	1 311 662	indel	CDS	no frameshift	galactokinase	Galactose metabolism
19	b	1/28	1 311 735	SNP	CDS	non-syn.	galactokinase	Galactose metabolism
20	h	1/28	1 311 996	SNP	CDS	non-syn.	galactokinase	Galactose metabolism
21	a	*5/28	1 317 345	SNP	CDS	syn.	Mo-dependent transcriptional regulator	regulation of transcription
22		5/28	1 421 879	indel	non-coding	-	acyl carrier protein	Fatty acid/polyketide biosynthesis
23	d	1/28	1 609 697	indel	non-coding	-	transcriptional regulator GalS	regulation of galactose transport/catabolism
24		1/28	1 611 529	SNP	CDS	non-syn.	galactose/galactoside ABC transporter MglA	galactose import
25		1/28	1 612 777	indel	CDS	no frameshift	galactose/galactoside ABC transporter MglC	galactose import
26		5/28	1 648 573	indel	non-coding	-	-	-
27	c	1/28	1 649 017	indel	non-coding	-	-	-
28		*4/28	1 649 038	indel	non-coding	-	-	-
29		4/28	1 665 941	indel	CDS	frameshift	glycosyltransferase	LPS biosynthesis
30		1/28	1 670 147	SNP	CDS	non-syn.	glycosyltransferase	LPS biosynthesis
31		2/28	1 670 356	SNP	CDS	non-syn.	glycosyltransferase	LPS biosynthesis
32	e	1/28	1 670 370	SNP	CDS	non-syn.	glycosyltransferase	LPS biosynthesis

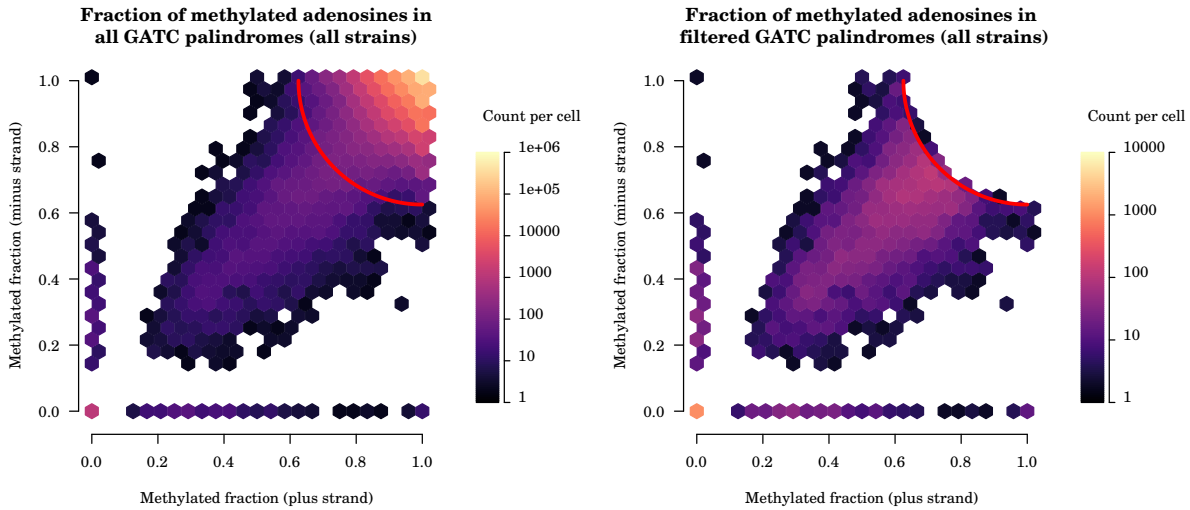
ID	Haplotype	Freq.	Pos. (bp)	Type	Region	Effect	Overlapping or closest (≤500bp) gene	
							Name	Function
33	a	*5/28	1 861 227	indel	non-coding	-	putative transcriptional regulator	regulation of transcription
34		1/28	2 144 682	indel	non-coding	-	hypothetical protein	unknown
35		1/28	2 282 483	indel	non-coding	-	MATE family efflux transporter	Na+/H+ driven multidrug efflux pump
36	h	1/28	2 353 326	indel	CDS	frameshift	fumarase C (iron independent)	TCA cycle
37	b	1/28	2 384 093	indel	CDS	frameshift	HlyD (haemolysin secretion system)	haemolysin/cutinase excretion
38	a	*5/28	2 456 338	SNP	CDS	non-syn.	peptidoglycan synthase	peptidoglycan biosynthesis
39		2/28	2 466 586	SNP	CDS	non-syn.	MmgE/PrpD family protein	propionate metabolism/TCA cycle?
40	e	1/28	2 941 884	indel	non-coding	-	VOC family protein	unknown
41		2/28	3 161 361	SNP	CDS	syn.	serine/threonine protein kinase	regulation of cell processes
42	e	1/28	3 408 594	indel	non-coding	-	nucleoside diphosphate hydrolase	regulation of cell processes
43	a	*5/28	3 477 366	SNP	CDS	non-syn.	transcriptional regulator RcsB	capsule synthesis/cell division/biofilm/motility
44		*0/28	3 600 509	indel	non-coding	-	phospholipid-binding lipoprotein MlaA	Outer membrane maintenance
45	a	*5/28	3 607 617	SNP	CDS	non-syn.	heme exporter protein CcmB	cytochrome c biogenesis
46	a	*5/28	3 869 219	SNP	CDS	syn.	alcohol dehydrogenase	energy metabolism
47		1/28	4 025 724	indel	non-coding	-	acetyl-CoA carboxylase alpha subunit	lipid metabolism
48		1/28	4 337 062	indel	non-coding	-	tRNA-Phe	translation
49		1/28	4 362 753	indel	non-coding	-	glycoporin	carbohydrate import
50	c	1/28	4 845 837	indel	CDS	frameshift	peptidylprolyl isomerase	protein folding chaperone
51	g	1/28	4 872 989	indel	CDS	frameshift	short chain dehydrogenase	oxidoreductase
52	d	1/28	4 924 755	SNP	CDS	non-syn.	threonine dehydratase	amino acid metabolism
53	a	*5/28	5 010 850	indel	CDS	frameshift	deacetylase	LPS biosynthesis
54	a	*5/28	5 010 868	SNP	CDS	non-syn.	deacetylase	LPS biosynthesis

Supplementary Table S1: Summary of the genetic variants observed in the sequenced clones. This table is taken from Bruneaux et al. (2019). Haplotype: letters denote groups of mutations for which alleles are associated together. Freq.: minor allele frequency among the 28 evolved clones. An asterisk denotes loci for which the reference clone carries the minor allele.

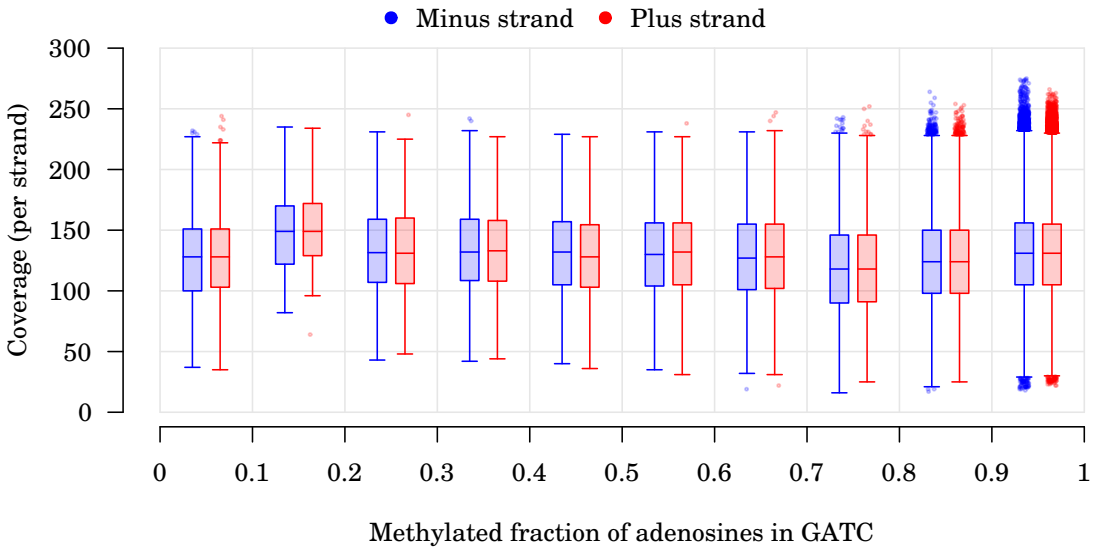
5.2 Supplementary figures



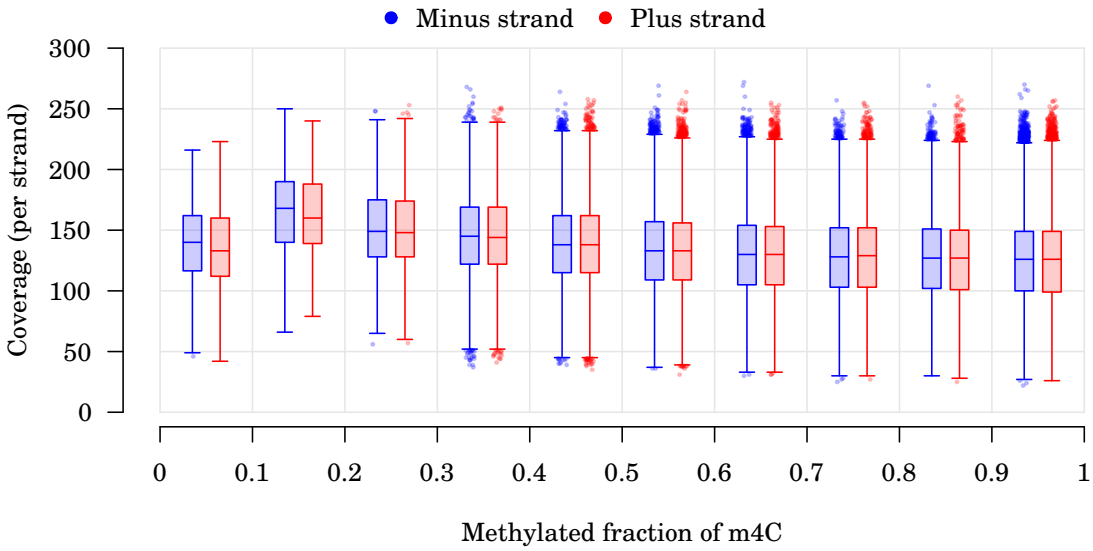
Supplementary Figure S1: Setup of the evolution experiment from which sequenced clones were isolated.



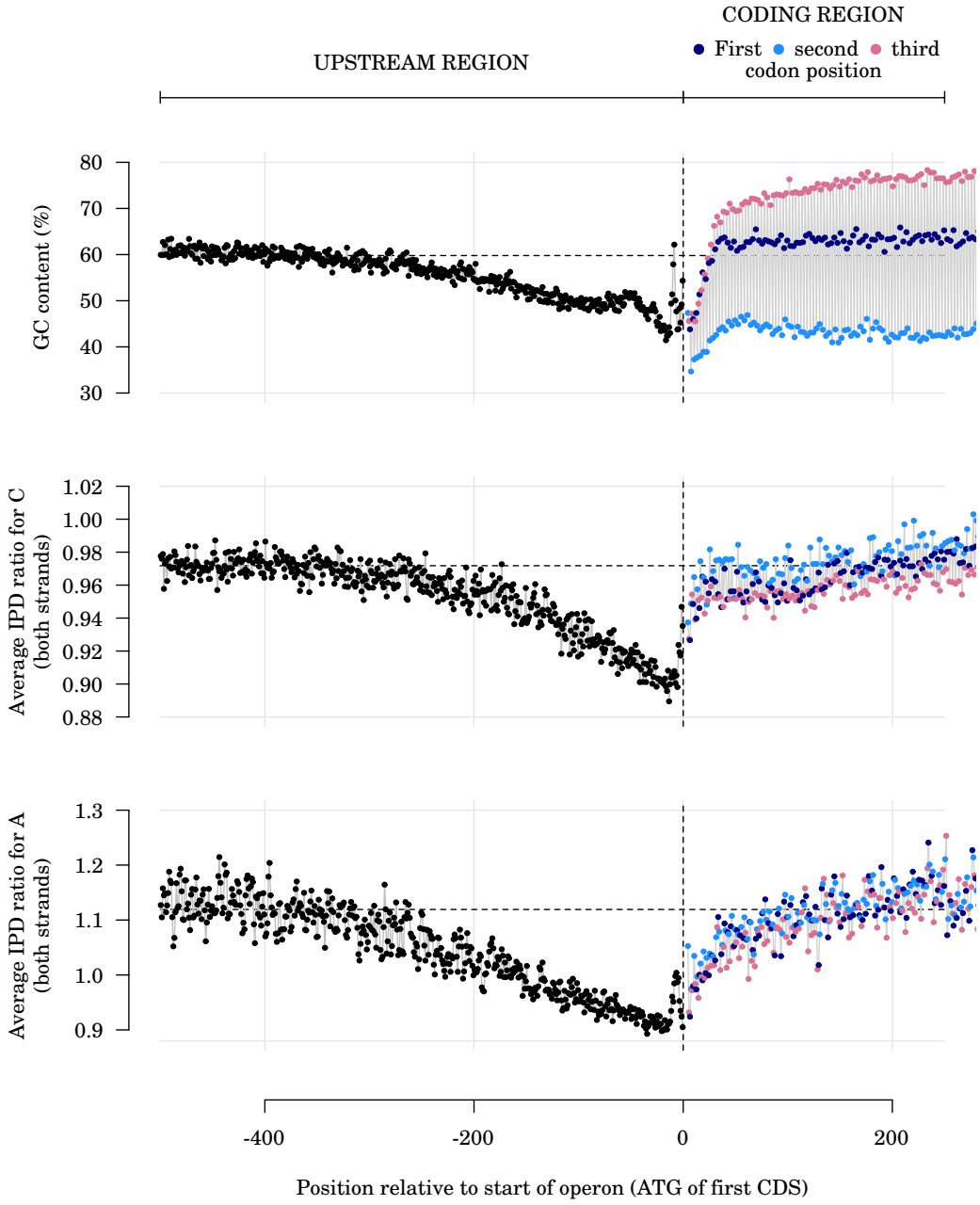
Supplementary Figure S2: Detection of partially methylated GATC loci. Distribution of methylated fractions of adenosines on boths DNA strands for GATC palindromes (showing data for all strains together). Left panel, all GATC palindromes shown; right panel, only GATC palindromes qualified as low methylation sites shown. The red arc in the left panel delimits the observations which are less four times the average quadratic distance to full methylation (point at (1,1)) away from full methylation. GATC palindromes are considered as low methylation sites if they lay outside this area (right panel).



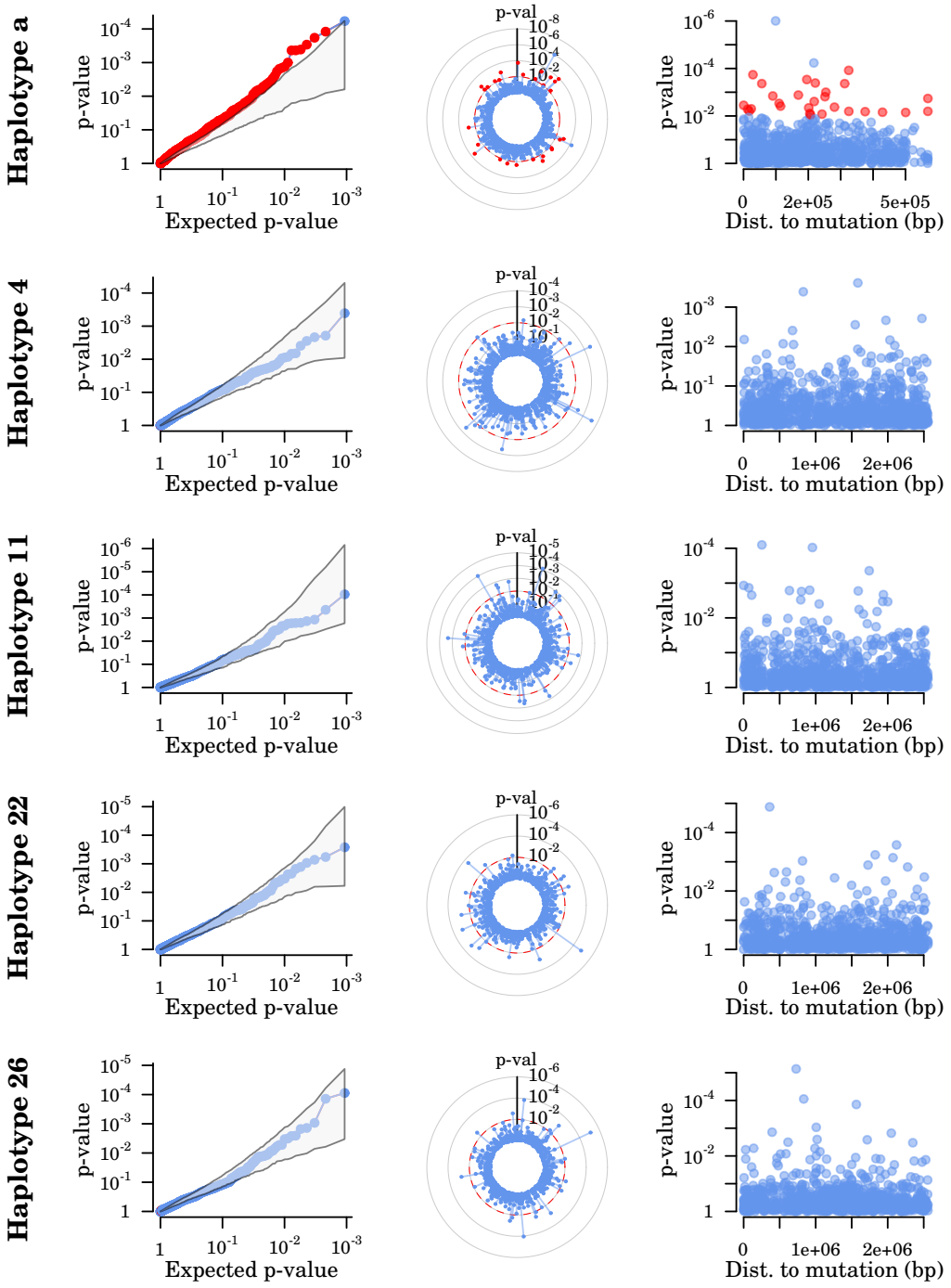
Supplementary Figure S3: Relationship between coverage and estimated fraction of methylated adenosines in GATC motifs. All adenosines present in GATC motifs, from all sequenced strains, are included. Data is binned by intervals of methylated fraction of 0.1 width, as indicated on the x axis. There is no correlation between coverage and estimated methylated fraction (Spearman's ρ = -0.01). However, the slight increase in average coverage for the (0.1, 0.2) methylated fractions compared to the (0, 0.1) fractions suggests that at very low methylation levels (below 0.2), higher coverage is needed to estimate a methylated fraction other than 0.



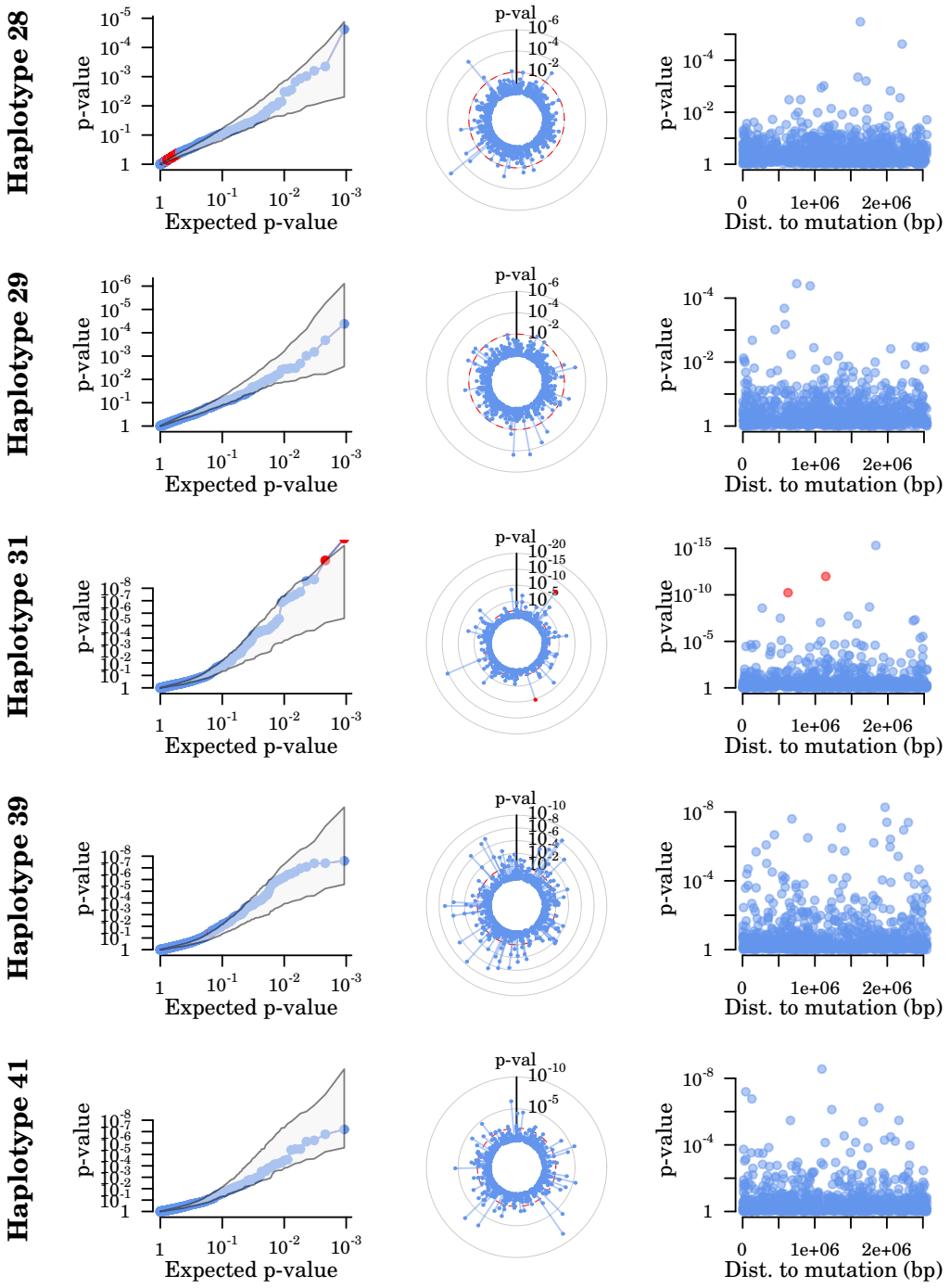
Supplementary Figure S4: Relationship between coverage and estimated fraction of methylated cytosines (m4C). All cytosines detected as m4C, from all sequenced strains, are included. Data is binned by intervals of methylated fraction of 0.1 width, as indicated on the x axis. There is a correlation between coverage and estimated methylated fraction (Spearman's ρ = -0.17, $p < 0.001$), indicating that higher coverage is needed to estimate low methylation fractions. However, 90 % of detected m4C have a methylation fraction > 0.4, suggesting that the bias due to this coverage effect is likely to be small.



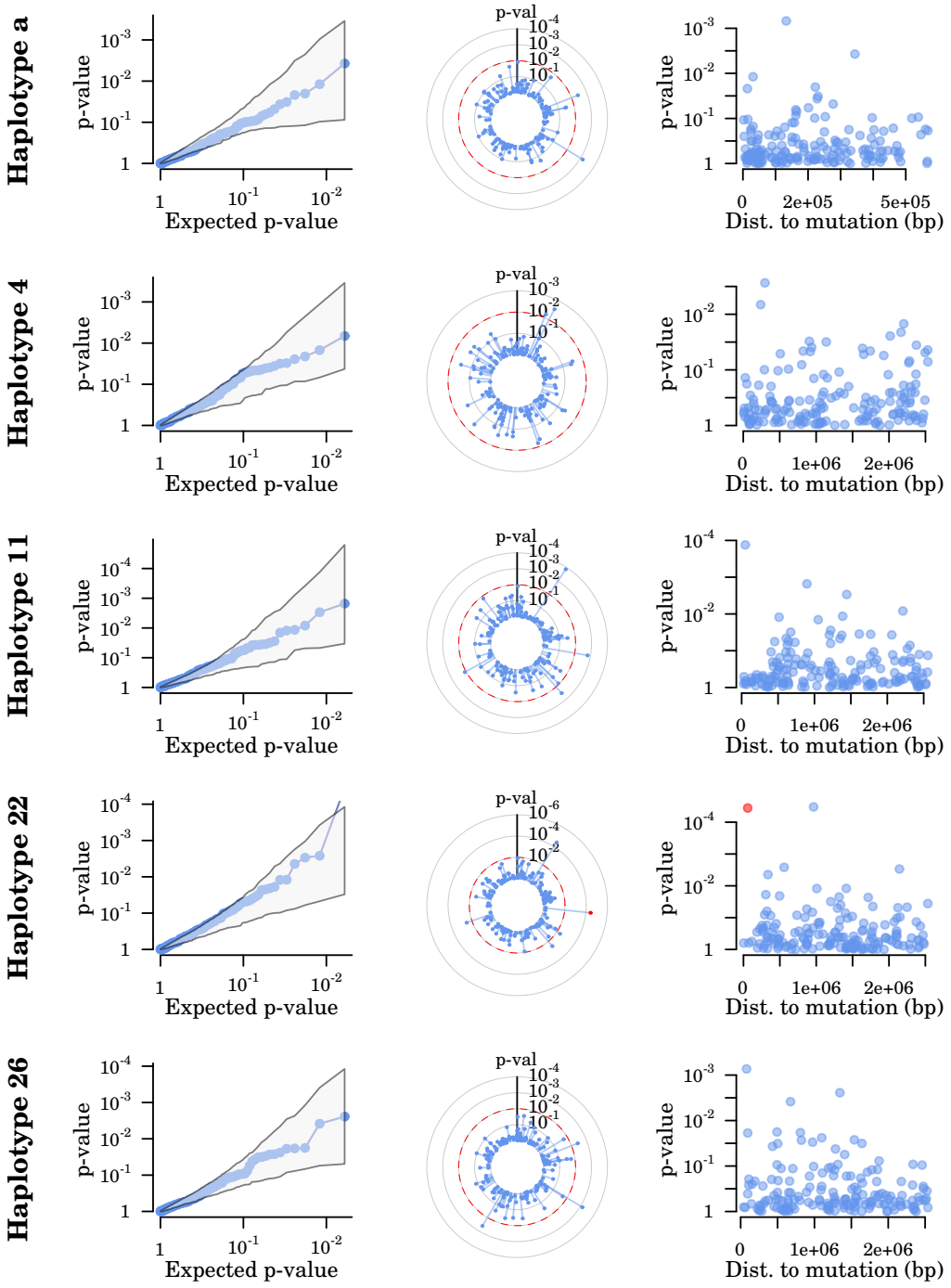
Supplementary Figure S5: Profiles of nucleotide composition and average IPD ratios for adenosines and cytosines around the start position of operon-leading CDS. IPD ratios (inter-pulse duration ratios between the sequenced based and an in-silico control) are related to base modifications, with high IPD ratios suggesting base modification. Analysis of the IPD ratios is complex, since a modified base can influence the IPD ratios of neighbouring bases. Thus, the trend observed for IPD ratios for cytosines could (at least partially) reflect the trend observed for neighbouring m6A, even if the cytosines are not themselves modified. Vertical dashed lines show the limit between upstream regions and the first codon of the leading CDS of predicted operons. Plotted values are averaged over each position relative to the leading CDS initiation codon based on operons predicted in the reference genome. Horizontal dashed lines show the genome-wide average values. Values for the three first bases on the coding sequences (usually ATG) are dropped from the plot to keep the y-scale reasonably narrow.



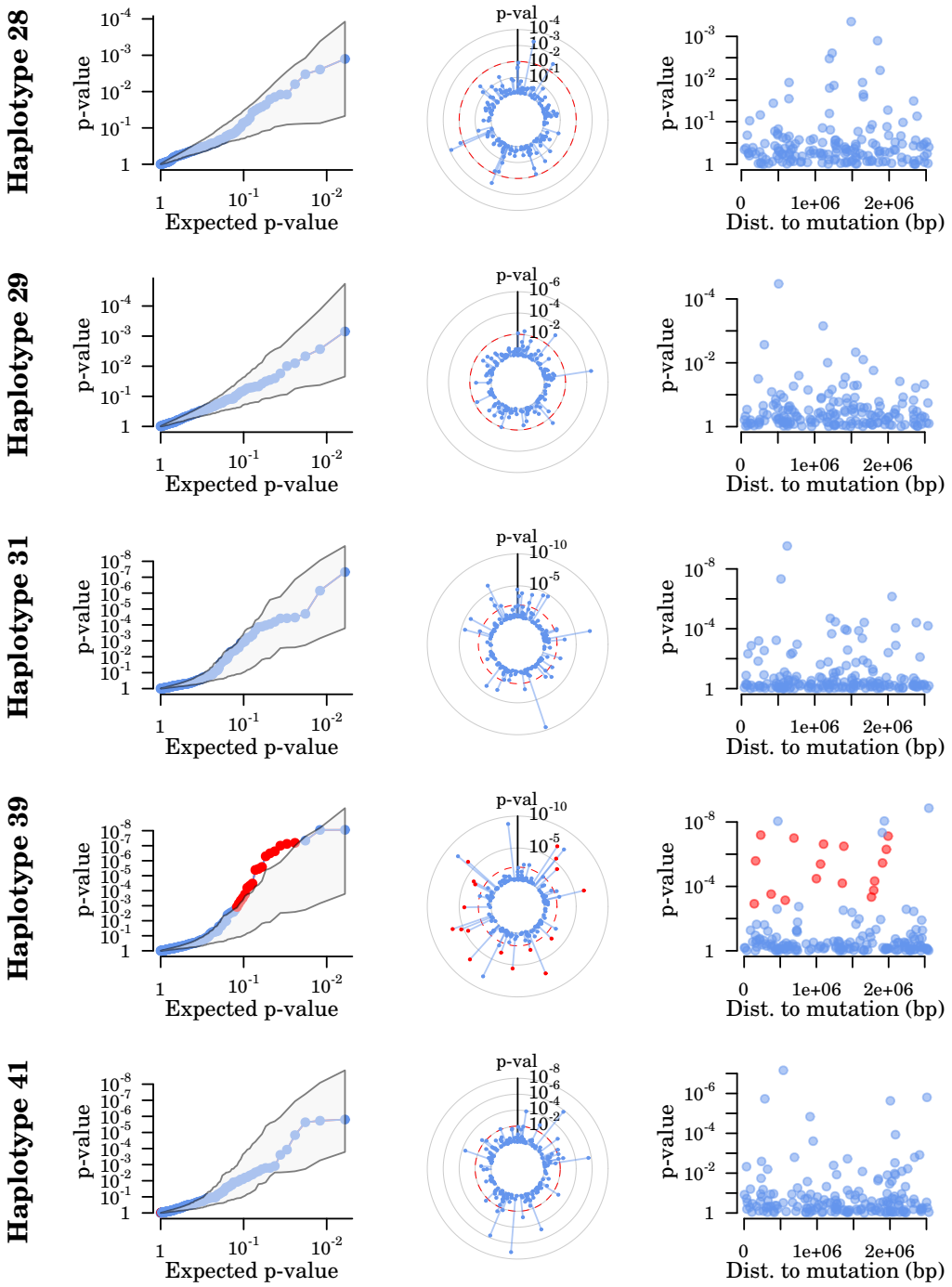
(a) m6A



(b) m6A

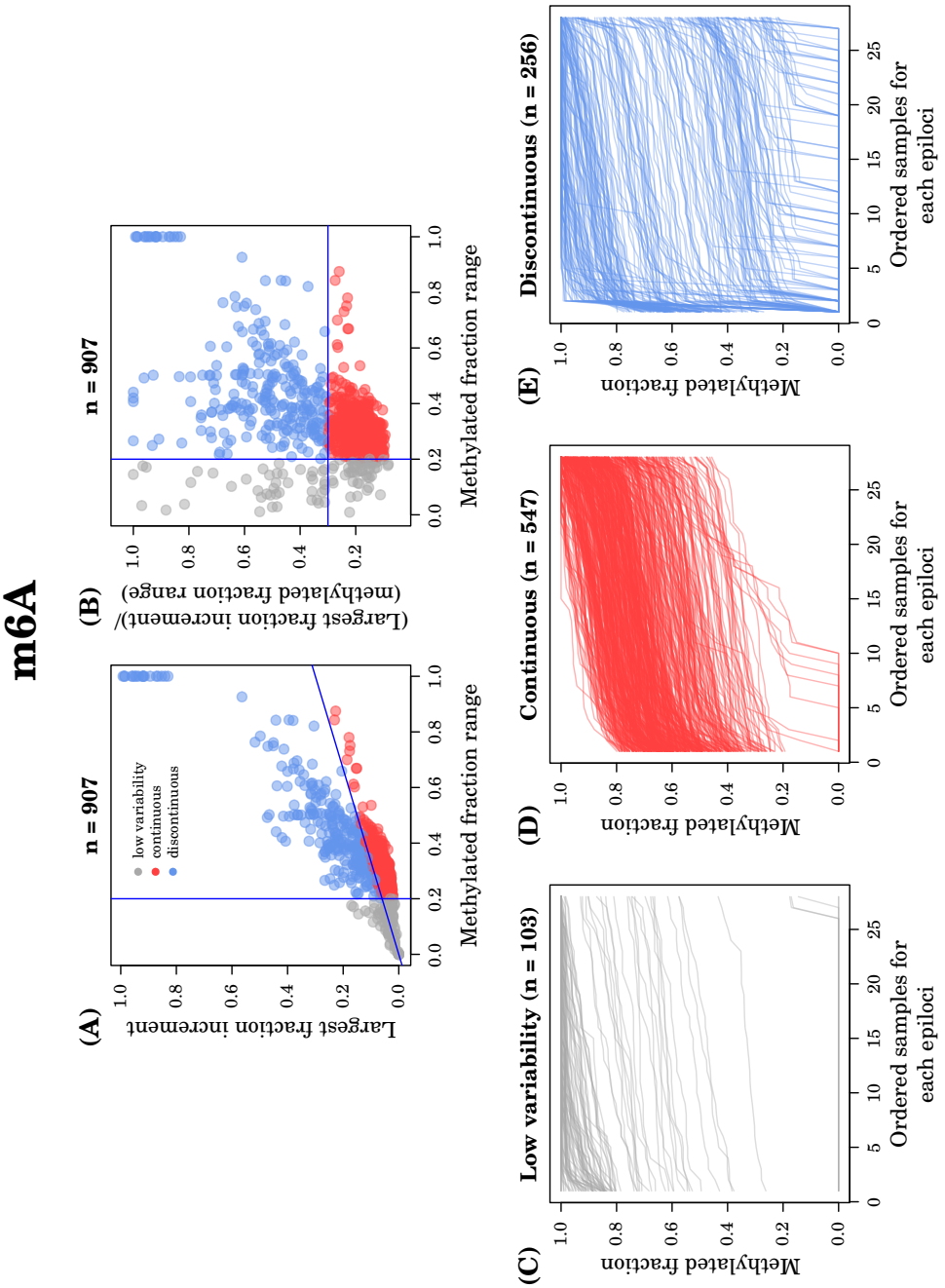


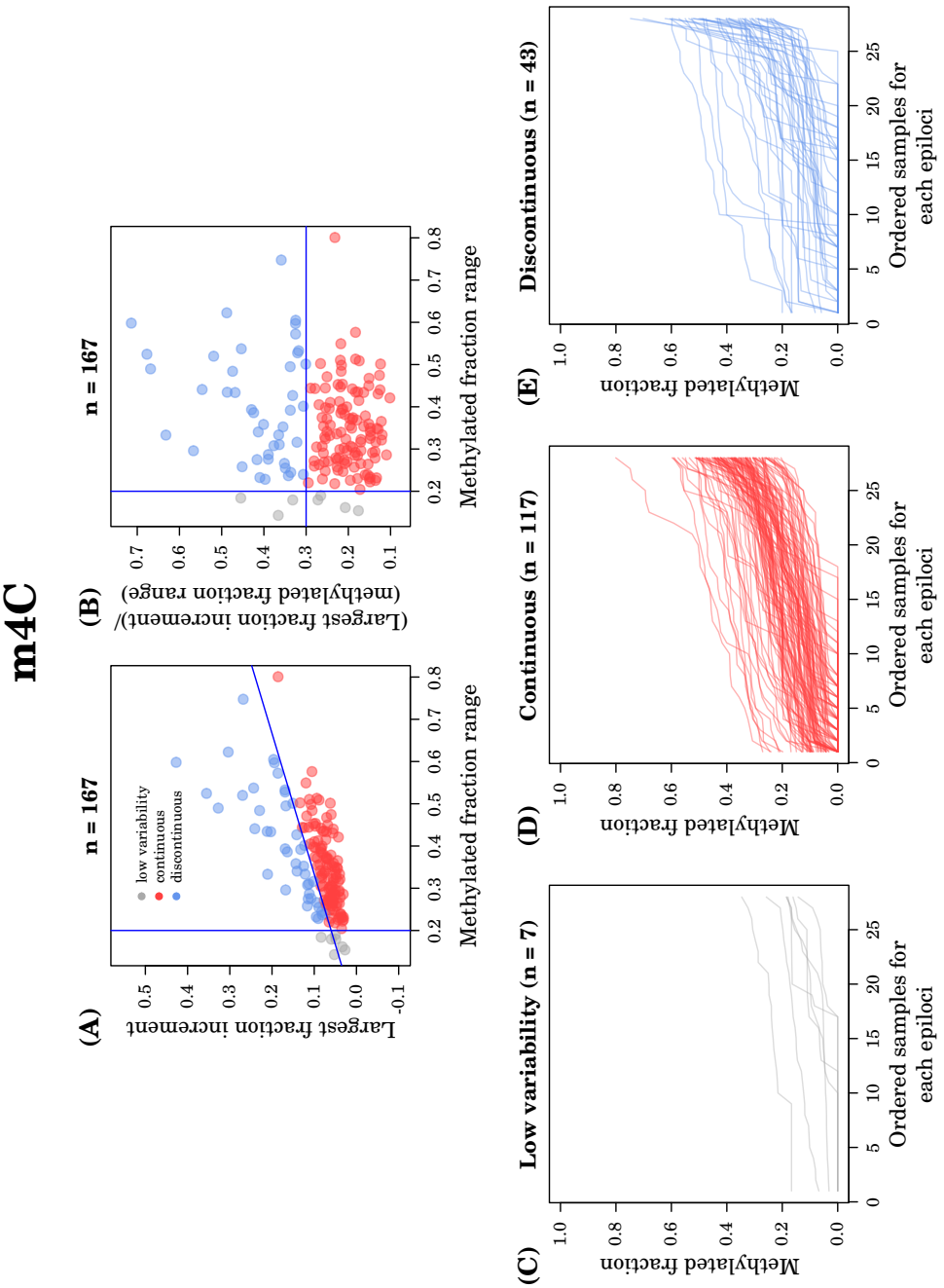
(c) m4C



(d) m4C

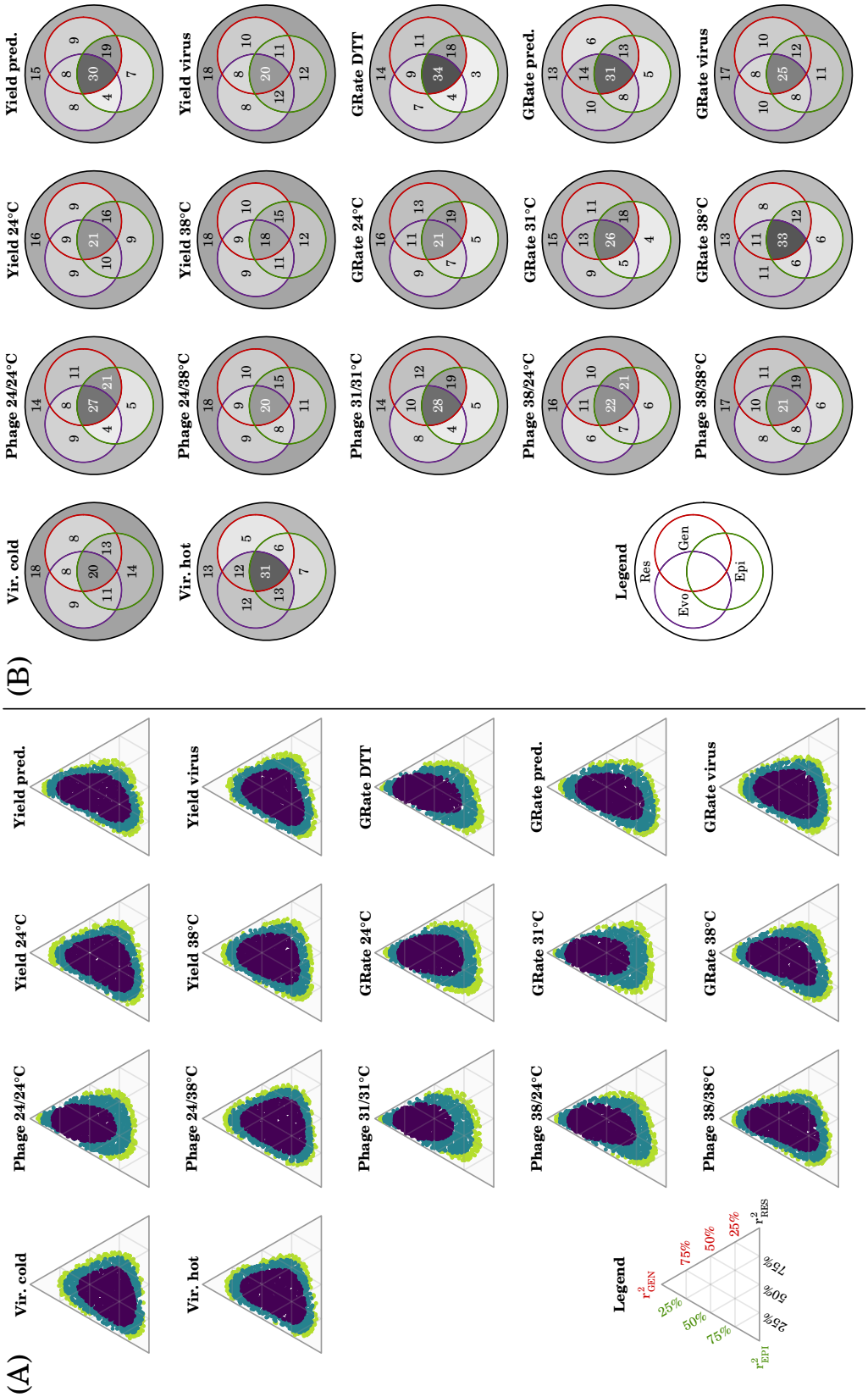
Supplementary Figure S6: Association between genetic and epigenetic changes. Each row corresponds to one genetic locus or haplotype. Left column, comparison between inflation for observed p-values and the 95 % (one-tailed) envelope of inflation for p-values from permuted datasets. Middle column, genomic map of the epiloci and their associated p-values, with *oriC* at the top of the chromosome. Red points indicate an epiloci for which the observed p-value is < 0.01 and is above the 95 % inflation permutation envelope. Right column, distribution of epiloci p-values in relation to their genomic distance to the genetic mutation. For the haplotype *a*, the distance is the distance to the closest mutation.



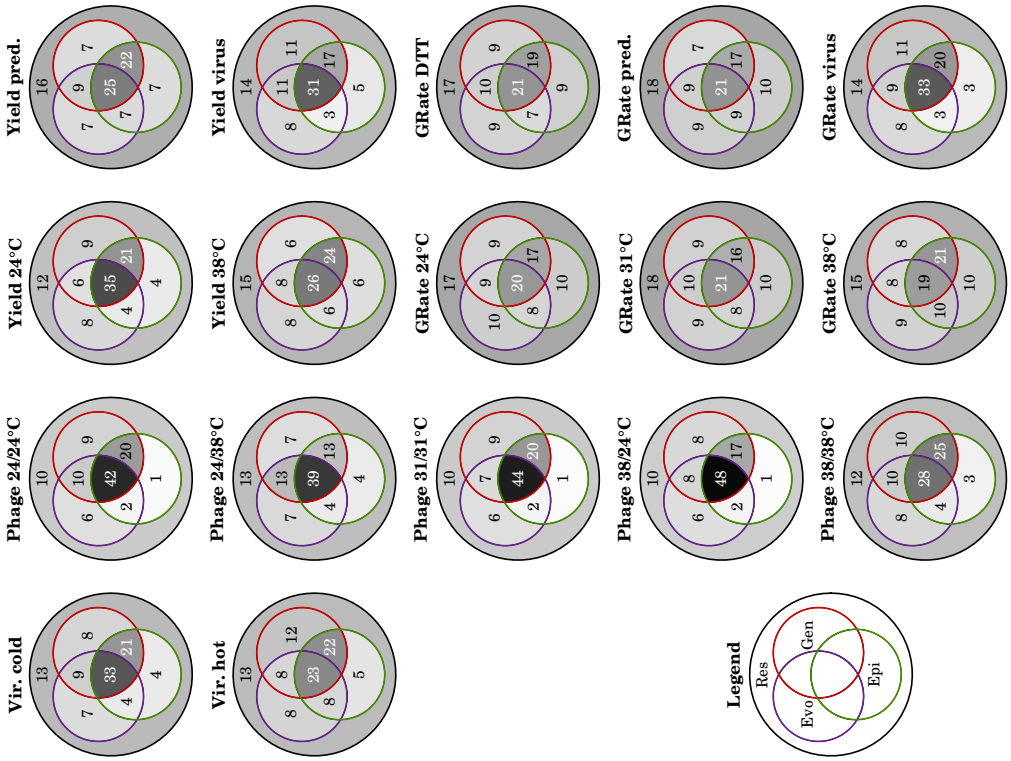


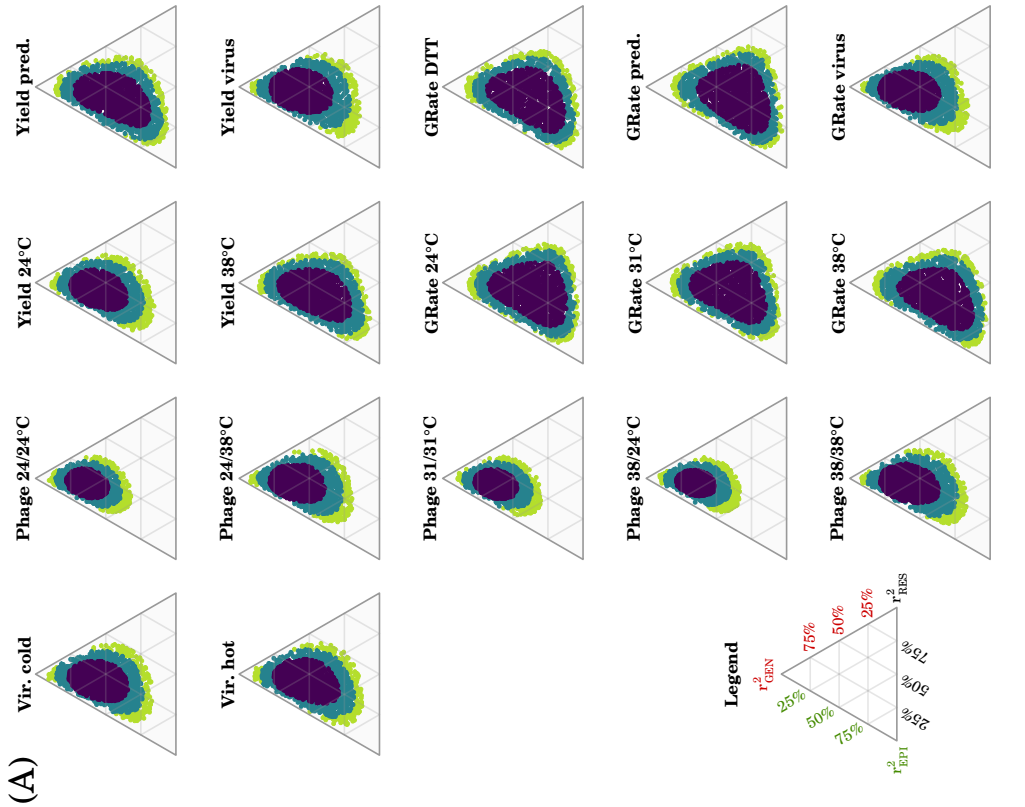
Supplementary Figure S7: Classification of epiloci of interest based on their methylation profiles. Epiloci of interest were m6A exhibiting incomplete methylation in at least one strain ("low-meth m6A") and clustered m4C detected with the kernel density approach. A,B: classification of epiloci based on their profiles of ordered methylated fractions in the 28 evolved strains. Epiloci with fraction range less than 0.2 are classified as *low variability* (light grey). Among the remaining epiloci, epiloci with a largest fraction increment less than 0.3 times their fraction range are classified as *continuous* (red) while others are classified as *discontinuous* (blue). C,D,E: ordered methylation profiles for epiloci classified in each category. The numbers above each panel are the number of epiloci displayed in this panel.

Supplementary Figure S8: Variance decomposition for each phenotypic trait, taking into account the fixed effect of haplotype group *a*. (A), 50 %, 80 % and 90 % credible surfaces for the estimates of the genetic, epigenetic and residuals components in total phenotypic variance. (B), point estimates for joint and disjoint variance components described by evolutionary treatment, genetics and epigenetics (in % of total phenotypic variance). See Figure 6 for legend details.



(a) Evolutionary treatments used for variance decomposition: 31 °C and 24–38°C

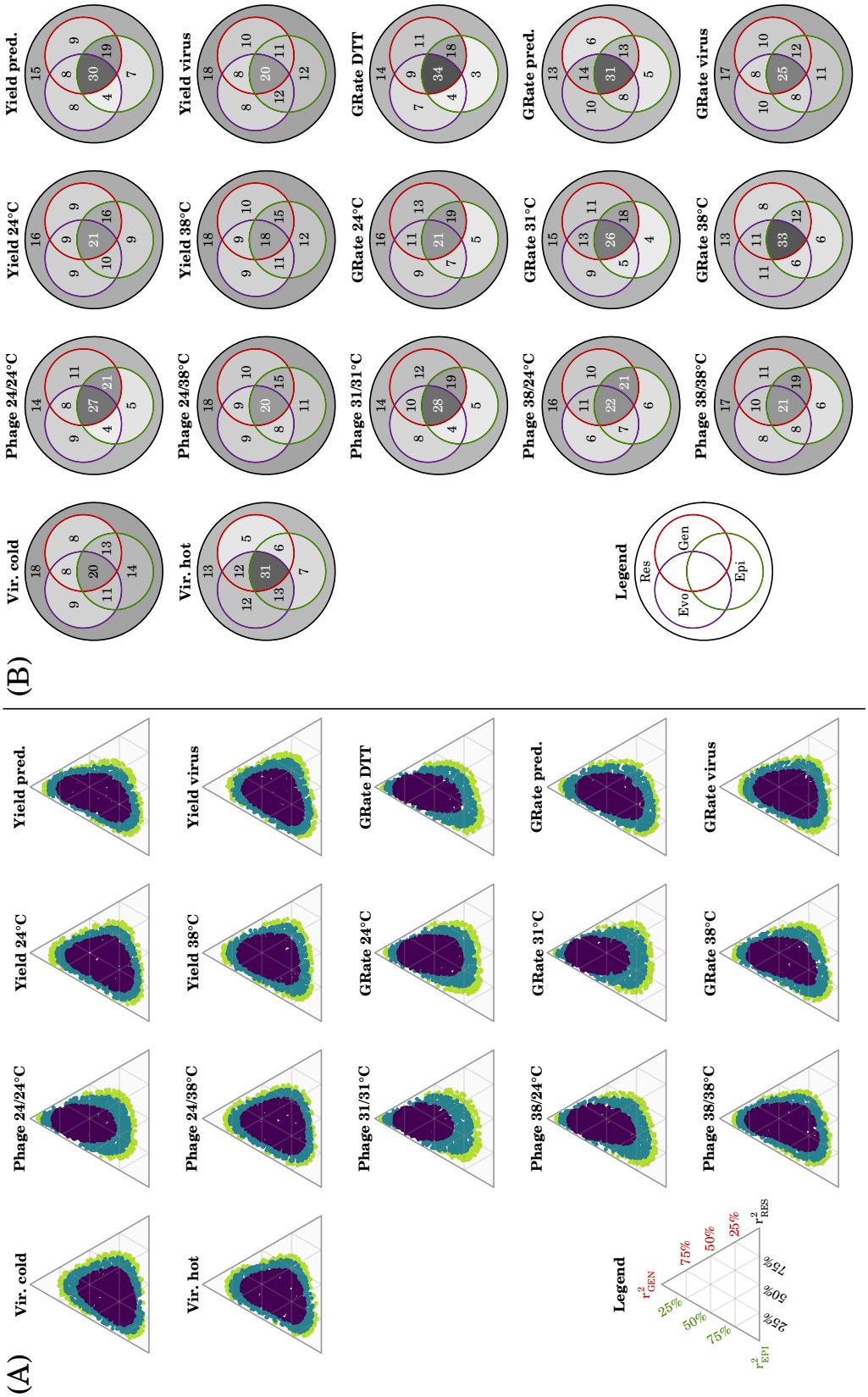




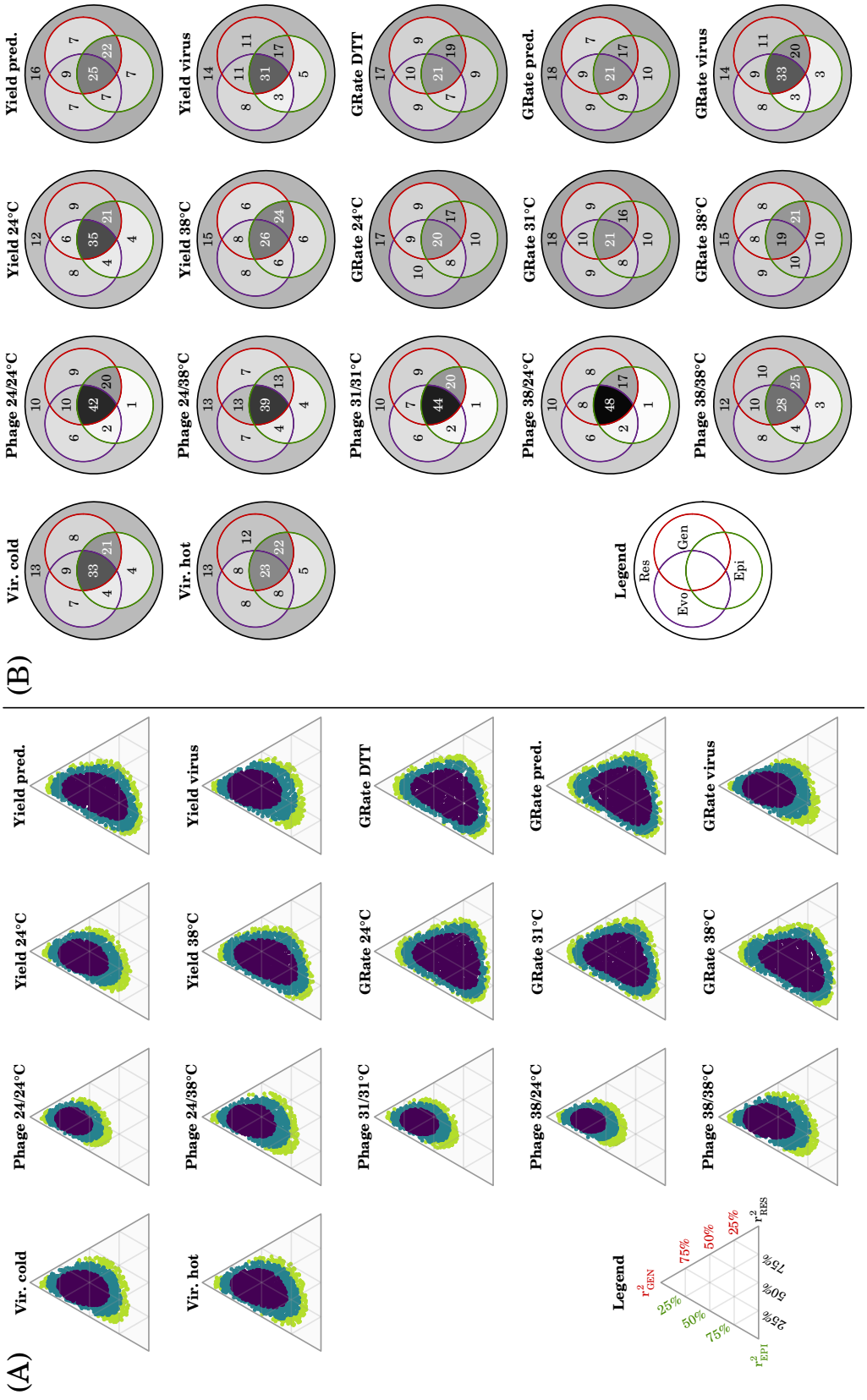
(b) Evolutionary treatments used for variance decomposition: 31 °C and 38 °C

(c) Evolutionary treatments used for variance decomposition: 24–38°C and 38 °C

Supplementary Figure S9: Variance decomposition for each phenotypic trait, using data from pairs of evolutionary treatments. (A), 50 %, 80 % and 90 % credible surfaces for the estimates of the genetic, epigenetic and residuals components in total phenotypic variance. (B), point estimates for joint and disjoint variance components described by evolutionary treatment, genetics and epigenetics (in % of total phenotypic variance). See Figure 6 for legend details.



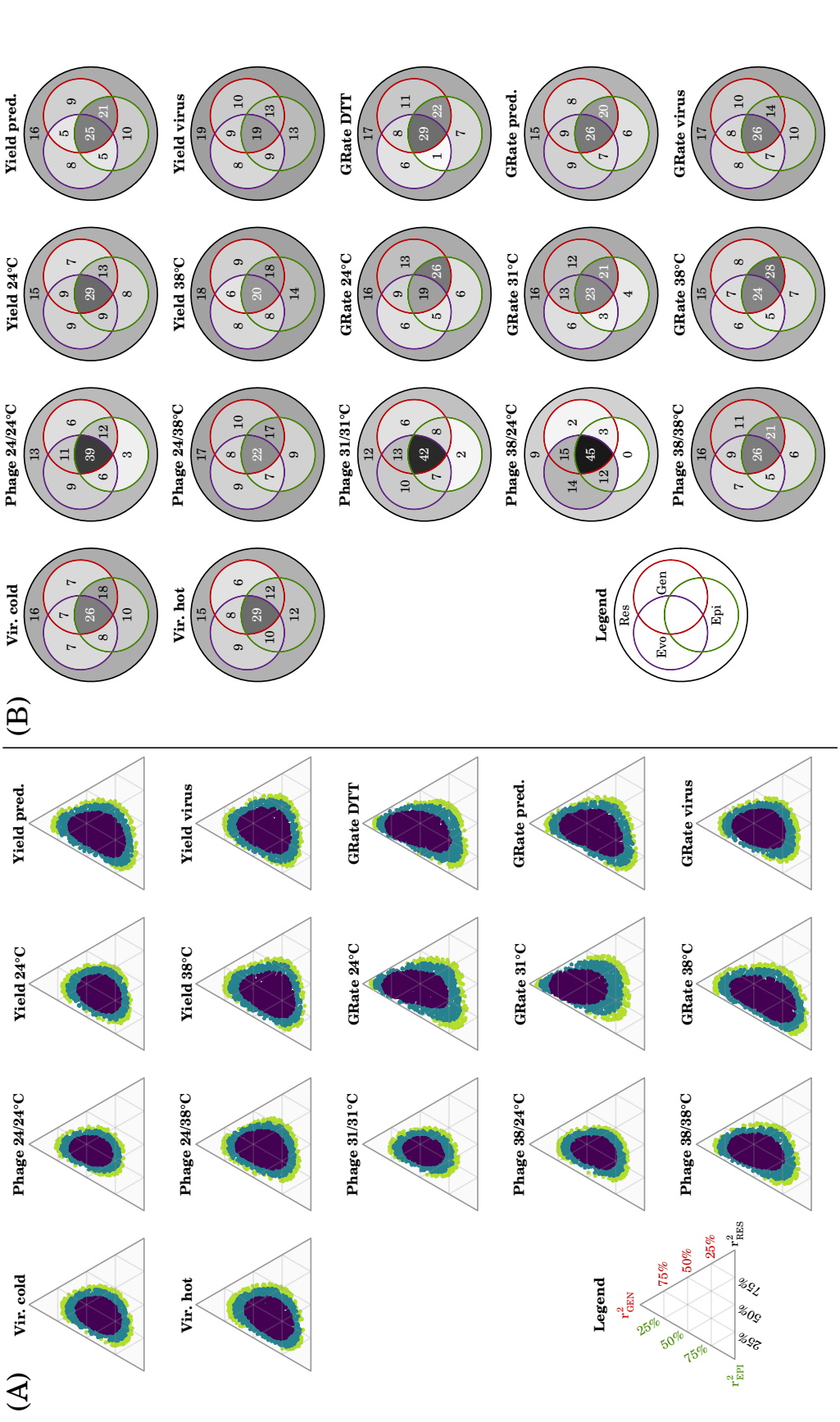
(a) Evolutionary treatments used for variance decomposition: 31 °C and 24–38°C (with haplotype a as fixed effect)



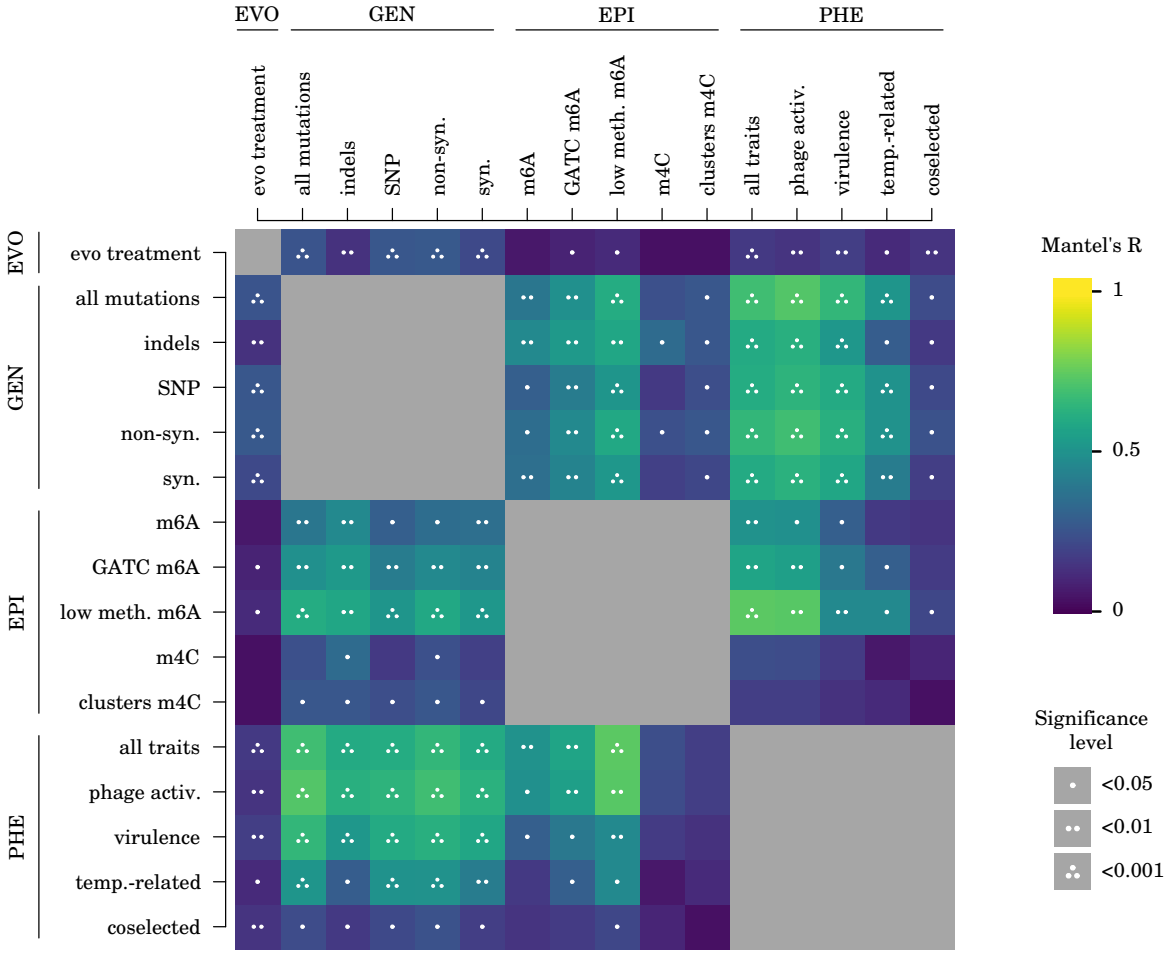
(b) Evolutionary treatments used for variance decomposition: 31 °C and 38 °C (with haplotype a as fixed effect)

(c) Evolutionary treatments used for variance decomposition: 24–38°C and 38 °C (with haplotype *a* as fixed effect)

Supplementary Figure S10: Variance decomposition for each phenotypic trait, using data from pairs of evolutionary treatments and taking into account the fixed effect of haplotype group *a*. (A), 50 %, 80 % and 90 % credible surfaces for the estimates of the genetic, epigenetic and residuals components in total phenotypic variance. (B), point estimates for joint and disjoint variance components described by evolutionary treatment, genetics and epigenetics (in % of total phenotypic variance). See Figure 6 for legend details.



Supplementary Figure S11: Variance decomposition for each phenotypic trait, taking into account the fixed effect of locus 39. (A), 50 %, 80 % and 90 % credible surfaces for the estimates of the genetic, epigenetic and residuals components in total phenotypic variance. (B), point estimates for joint and disjoint variance components described by evolutionary treatment, genetics and epigenetics (in % of total phenotypic variance). See Figure 6 for legend details.



Supplementary Figure S12: Correlogram for distance measures based on treatment, genetic, epigenetic and phenotypic data. EVO, GEN, EPI, PHE: distance measures grouped in evolutionary treatment, genetic, epigenetic and phenotypic-based measures, respectively. In GEN: *non-syn.*, indels resulting in frame shift and non-synonymous SNPs; *syn.*, all other indels and SNPs. In EPI: *m6A*, all m6A fractions; *GATC m6A*, m6A fractions in GATC motifs; *low meth. m6A*, m6A fractions for GATC loci detected as partially methylated (see Methods); *m4C*, all m4C fractions; *clusters m4C*, m4C epialleles calculated for m4C clusters based on the kernel density permutation approach. In PHE: *phage activ.*, rate of activation of prophage KSP20; *temp.-related*, temperature-related traits from Ketola et al. (2013); *coselected*, coselected traits from Ketola et al. (2013).

		EVO	GEN					EPI					PHE				
		evo treatment	all mutations	indels	SNP	non-syn.	syn.	m6A	GATC m6A	low meth. m6A	m4C	clusters m4C	all traits	phage activ.	virulence	temp.-related	coselected
EVO	evo treatment		**0.27** <0.001	**0.15** 0.001	**0.27** <0.001	**0.28** <0.001	**0.22** <0.001	0.07 0.071	**0.1** **0.032**	**0.12** **0.021**	0.05 0.12	0.04 0.168	**0.17** <0.001	**0.15** **0.006**	**0.19** **0.001**	**0.12** **0.013**	**0.16** **0.006**
GEN	all mutations	**0.27** <0.001						**0.4** **0.006**	**0.5** **0.001**	**0.62** <0.001	0.25 0.057	**0.27** **0.019**	**0.7** <0.001	**0.73** <0.001	**0.67** <0.001	**0.52** <0.001	**0.23** **0.016**
GEN	indels	**0.15** 0.001						**0.48** **0.004**	**0.54** **0.002**	**0.6** **0.002**	**0.35** **0.012**	**0.27** **0.03**	**0.61** <0.001	**0.63** <0.001	**0.53** <0.001	**0.3** **0.04**	**0.17** **0.049**
GEN	SNP	**0.27** <0.001						**0.31** **0.024**	**0.42** **0.001**	**0.53** <0.001	0.17 0.114	**0.24** **0.022**	**0.62** <0.001	**0.65** <0.001	**0.61** <0.001	**0.52** <0.001	**0.23** **0.014**
GEN	non-syn.	**0.28** <0.001						**0.36** **0.012**	**0.48** **0.002**	**0.6** <0.001	**0.25** **0.05**	**0.27** **0.014**	**0.66** <0.001	**0.7** <0.001	**0.64** <0.001	**0.51** <0.001	**0.25** **0.01**
GEN	syn.	**0.22** <0.001						**0.36** **0.008**	**0.45** **0.001**	**0.53** <0.001	0.19 0.106	**0.21** **0.049**	**0.61** <0.001	**0.64** <0.001	**0.6** <0.001	**0.42** **0.002**	**0.19** **0.028**
EPI	m6A	0.07 0.071	**0.4** **0.006**	**0.48** **0.004**	**0.31** **0.024**	**0.36** **0.012**	**0.36** **0.008**						**0.51** **0.007**	**0.5** **0.01**	**0.31** **0.039**	0.17 0.151	0.15 0.108
EPI	GATC m6A	**0.1** **0.032**	**0.5** **0.001**	**0.54** **0.002**	**0.42** **0.001**	**0.48** **0.002**	**0.45** **0.001**						**0.59** **0.004**	**0.57** **0.006**	**0.4** **0.018**	**0.3** **0.046**	0.18 0.074
EPI	low meth. m6A	**0.12** **0.021**	**0.62** <0.001	**0.6** **0.002**	**0.53** <0.001	**0.6** <0.001	**0.53** <0.001						**0.75** <0.001	**0.74** **0.002**	**0.48** **0.006**	**0.48** **0.01**	**0.21** **0.038**
EPI	m4C	0.05 0.12	0.25 0.057	**0.35** **0.012**	0.17 0.114	**0.25** **0.05**	0.19 0.106						0.24 0.082	0.23 0.11	0.18 0.137	0.07 0.342	0.11 0.176
EPI	clusters m4C	0.04 0.168	**0.27** **0.019**	**0.27** **0.03**	**0.24** **0.022**	**0.27** **0.014**	**0.21** **0.049**						0.18 0.116	0.19 0.116	0.15 0.148	0.13 0.196	-0.04 0.667
PHE	all traits	**0.17** <0.001	**0.7** <0.001	**0.61** <0.001	**0.62** <0.001	**0.66** <0.001	**0.61** <0.001	**0.51** **0.007**	**0.59** **0.004**	**0.75** <0.001	0.24 0.082	0.18 0.116					
PHE	phage activ.	**0.15** **0.006**	**0.73** <0.001	**0.63** <0.001	**0.65** <0.001	**0.7** <0.001	**0.64** <0.001	**0.5** **0.01**	**0.57** **0.006**	**0.74** **0.002**	0.23 0.11	0.19 0.116					
PHE	virulence	**0.19** **0.001**	**0.67** <0.001	**0.53** <0.001	**0.61** <0.001	**0.64** <0.001	**0.6** <0.001	**0.31** **0.039**	**0.4** **0.018**	**0.48** **0.006**	0.18 0.137	0.15 0.148					
PHE	temp.-related	**0.12** **0.013**	**0.52** <0.001	**0.3** **0.04**	**0.52** <0.001	**0.51** <0.001	**0.42** **0.002**	0.17 0.151	**0.3** **0.046**	**0.48** **0.01**	0.07 0.342	0.13 0.196					
PHE	coselected	**0.16** **0.006**	**0.23** **0.016**	**0.17** **0.049**	**0.23** **0.014**	**0.25** **0.01**	**0.19** **0.028**	0.15 0.108	0.18 0.074	**0.21** **0.038**	0.11 0.176	-0.04 0.667					

Supplementary Figure S13: Correlogram (values) for distance measures based on treatment, genetic, epigenetic and phenotypic data. In each cell, the top number is Mantel's R and the bottom number is the corresponding p-value. Cells with content in bold have $p < 0.05$. EVO, GEN, EPI, PHE: distance measures grouped in evolutionary treatment, genetic, epigenetic and phenotypic-based measures, respectively. In GEN: *non-syn.*, indels resulting in frame shift and non-synonymous SNPs; *syn.*, all other indels and SNPs. In EPI: *m6A*, all m6A fractions; *GATC m6A*, m6A fractions in GATC motifs; *low meth. m6A*, m6A fractions for GATC loci detected as partially methylated (see Methods); *m4C*, all m4C fractions; *clusters m4C*, m4C epialleles calculated for m4C clusters based on the kernel density permutation approach. In PHE: *phage activ.*, rate of activation of prophage KSP20; *temp.-related*, temperature-related traits from Ketola et al. (2013); *coselected*, coselected traits from Ketola et al. (2013).