DNA methylation in Mytilus chilensis

Cristian Gallardo[1], Valentina Venezuala[1], Steven Roberts[2]

[1]University of Concepcion,
[2]University of Washington,

Corresponding author: Cristian Gallardo, `cgallardo@udec.cl`

Corresponding author: Valentina Venezuala, `vvenezuala@udec.cl`

Corresponding author: Steven Roberts, `sr320@uw.edu`

Abstract

We investigate genome-wide DNA methylation patterns in Mytilus chilensis from two contrasting seedbeds—Cochamó and Yaldad—with the aim of identifying differentially methylated regions (DMRs) that may underlie local adaptation. By integrating this epigenetic information with known patterns of gene expression and environmental context, we seek to illuminate the molecular basis of ecological divergence and contribute new insights into how environmental and anthropogenic forces shape genomic and epigenomic variation in a key aquaculture species.

Plain Language Summary

This study explores how DNA methylation may help Chilean blue mussels adapt to different environments, providing insights for aquaculture and conservation.

1 Introduction

Understanding how marine species respond to environmental heterogeneity is critical for conserving locally adapted populations and sustaining aquaculture practices in the face of global change. The Chilean blue mussel, Mytilus chilensis, a key endemic species in southern Chile, supports one of the world's largest mussel aquaculture industries. However, the biological foundation of population-level adaptation in this species remains underexplored. Emerging evidence suggests that both genetic and epigenetic mechanisms—particularly DNA methylation—may contribute to phenotypic plasticity and local adaptation in marine invertebrates.

Previous studies have revealed extensive transcriptomic and mitochondrial differences between M. chilensis populations from ecologically distinct seedbeds, such as Cochamó (41°S) and Yaldad (43°S), in the inner sea of Chiloé Island. These differences include differential expression of genes involved in immunity, metabolism, and cellular stress responses, as well as tissue-specific shifts in mitochondrial gene expression linked to local environmental pressures, including aquaculture practices and variable oxygenation levels (Yévenes et al., 2021, 2022). Experimental evidence from reciprocal transplants shows population-specific transcriptional responses and reduced growth performance upon translocation, highlighting functional divergence between seedbeds despite high larval connectivity (Yévenes et al., 2025).

In addition to gene expression differences, recent work has uncovered epigenetic signatures that may contribute to adaptive divergence. Long non-coding RNAs (lncRNAs) were found to be differentially expressed between Cochamó and Yaldad populations, with associated changes in immune and metabolic gene networks (Yévenes et al., 2023). These findings underscore the potential for environmentally responsive regulatory elements—including DNA methylation—to mediate local adaptation.

Yet, the role of DNA methylation in shaping population-specific responses in M. chilensis remains unexplored. DNA methylation, a heritable and reversible epigenetic modification, can influence gene expression and phenotypic traits without altering the underlying DNA sequence. In bivalves, DNA methylation is known to affect developmental and immune processes and may be responsive to environmental stressors such as hypoxia, toxins, and acidification—conditions that vary across the Chilean inner sea and have been shown to modulate gene expression and microbiota composition in M. chilensis (Castillo et al., 2017; Montúfar-Romero et al., 2024, 2025).

In this study, we investigate genome-wide DNA methylation patterns in Mytilus chilensis from two contrasting seedbeds—Cochamó and Yaldad—with the aim of identifying differentially methylated regions (DMRs) that may underlie local adaptation. By integrating this epigenetic information with known patterns of gene

expression and environmental context, we seek to illuminate the molecular basis
of ecological divergence and contribute new insights into how environmental and
anthropogenic forces shape genomic and epigenomic variation in a key aquaculture
species.

2 Methods

To investigate DNA methylation differences between populations of Mytilus chilensis
from Cochamó and Yaldad, whole-genome bisulfite sequencing (WGBS) was per-
formed and analyzed using a reproducible, high-throughput bioinformatics workflow.
DNA was extracted from gill tissue of adult individuals and subjected to bisulfite
treatment and library preparation. Libraries were sequenced on an Illumina NovaSeq
6000 platform, generating 150 bp paired-end reads, with each sample yielding ap-
proximately 120–160 million read pairs, providing $>30\times$ average coverage across the
1.93 Gb M. chilensis genome.

2.1 Read Processing and Alignment

Raw sequencing reads were processed using the nf-core/methylseq pipeline (ver-
sion 2.5), implemented through Nextflow (version 23.04.1). Reads were aligned to
the M. chilensis reference genome (MchilensisGenomeV1.fa), a chromosome-level
assembly comprising 14 scaffolds (Gallardo-Escárate et al. 2023), using Bismark (ver-
sion 0.24.1) with the Bowtie2 aligner (version 2.5.1). Quality trimming was applied
to reverse reads by hard-clipping the final 10 bases (–clip_r2 10) to mitigate low-
quality base calls at the 3 end. Methylation calls were extracted using the Bismark
methylation extractor and aggregated by cytosine context. Pipeline execution was
configured via a custom .config file to specify HPC resources and run with –resume,
–with-trace, –with-report, and –with-timeline flags for reproducibility and track-
ing. Output files were stored in a structured directory and included per-base CpG
methylation calls, quality control metrics, and coverage files.

2.2 Methylation Data Processing and Filtering

Downstream analysis was performed in R (version 4.3.1) using the methylKit pack-
age (version 1.10.0). Due to compatibility issues with newer versions of supporting
libraries, data.table was pinned to version 1.14.10. Methylation coverage files were
imported using methRead() with the pipeline = "bismarkCoverage" parameter and
CpG context specified. Six samples—three each from Cochamó (LCo_BSr1–3) and
Yaldad (LYa_BSr1–3)—were analyzed. Samples were assigned to treatment groups
with Cochamó designated as the control (treatment = 0) and Yaldad as the treat-
ment group (treatment = 1).

To ensure data quality, CpG sites were filtered to retain only those with a minimum
read coverage of 10 in all samples, excluding those in the top 2% of coverage val-
ues to reduce PCR bias. Sites were then unified across samples using unite() with
min.per.group = 3 and destrand = FALSE to retain strand-specific methylation
information.

2.2.1 Exploratory and Differential Methylation Analyses

Global methylation patterns were assessed using hierarchical clustering (Ward's
method on correlation distance) and principal component analysis (PCA) to evalu-
ate sample similarity. Population-level differences in DNA methylation were then
tested using methylKit's logistic regression model, incorporating a multinomial
overdispersion correction (overdispersion = "MN") to account for biological variabil-
ity. Statistical testing was performed using the Chi-squared test (test = "Chisq"),
and q-values were adjusted for multiple testing using the SLIM method (adjust =
"SLIM").

Differentially methylated loci (DMLs) were defined as CpG sites with an absolute
methylation difference 25% and FDR-adjusted q-value 0.05. Separate sets of

106 DMLs were extracted to identify hypomethylated (lower in Yaldad) and hyperme-
107 thylated (higher in Yaldad) loci. Results were saved as tab-delimited text files for
108 downstream annotation and visualization.

3 Results

110 Bisulfite sequencing across six libraries yielded approximately 99.8 million reads
111 per sample. Alignment to the reference genome was successful for 60.5–61.4% of
112 reads, with 22–24% unaligned and 15–16% aligning ambiguously. After removal of
113 duplicate reads (2.2–3.5 million per sample), 57.0–58.8 million uniquely aligned reads
114 remained for methylation analysis.

115 Genome-wide cytosine methylation was assessed in CpG, CHG, and CHH contexts.
116 Across samples, ~2.1–2.2 billion cytosines were examined. CpG methylation was
117 the most prominent, with an average of ~36 million methylated CpGs and ~205
118 million unmethylated CpGs per sample, corresponding to a methylation rate of ap-
119 proximately 15%. In contrast, CHG and CHH methylation was much lower, with
120 methylation levels below 1% in both contexts. These patterns are consistent with
121 expectations for invertebrate methylomes, where methylation is primarily restricted
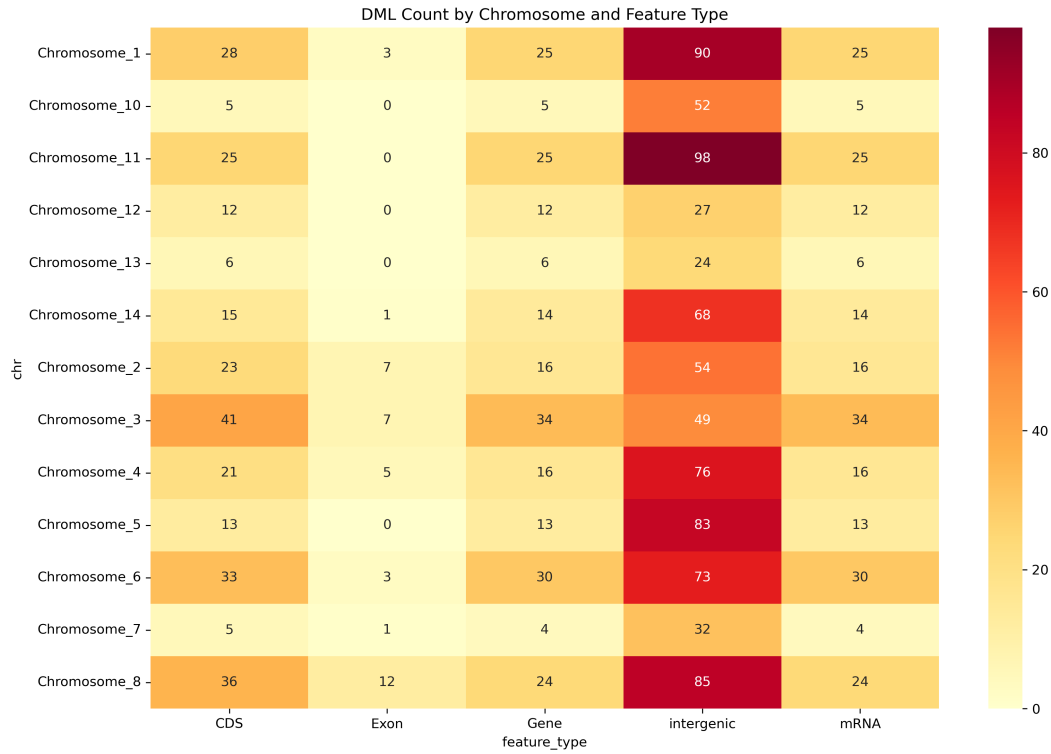122 to CpG sites.

124 Source: Article Notebook
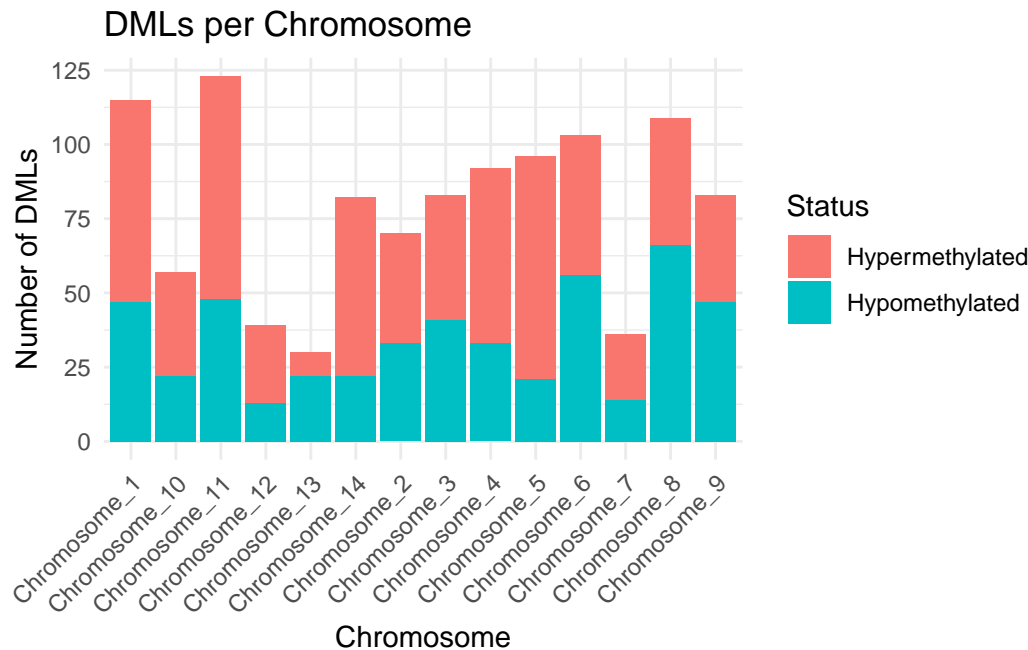
3.1 Location

Figure 1: heatmap

3.2 Genes harboring DMLs and overall methylation pattern

We identified 159 genes containing differentially methylated loci (DMLs). For each gene, methylation differences were calculated by aggregating the absolute methylation change at all DMLs mapped to that gene, using the output from methylKit's logistic regression analysis. Summary statistics (range: -77.04% to $+66.86\%$; SD 40.48%, median: -8.96%, mean: -0.68%) were computed across these gene-level aggregated values, indicating a shift toward hypomethylation among DML-associated genes.

3.3 Gene Ontology (GO): processes, functions, and locations

Across Biological Process terms, DML genes were most frequently annotated to metabolic process (71 genes; 44.7% of all DML genes) and to high-level developmental and cellular organization/biogenesis categories (each 57 genes; 35.8%). Additional frequent terms included nitrogen compound metabolic process (63 genes; 39.6%) and positive regulation of biological process (57 genes; 35.8%). For Molecular Function, broad binding categories predominated (binding: 81 genes, 50.9%; protein binding: 58 genes, 36.5%), consistent with an overrepresentation of regulatory and interaction-prone proteins. Cellular Component assignments placed many targets in cell/intracellular compartments (cell: 104 genes, 65.4%; intracellular: 99 genes, 62.3%), with frequent localization to the cytoplasm (80 genes, 50.3%) and nucleus (53 genes, 33.3%)

These patterns indicate that DMLs are concentrated in genes mediating metabolism, developmental programs, and cellular organization, with many targets residing in cytoplasmic and nuclear contexts where gene regulation and signal processing occur.

3.4 KEGG pathways

DML-associated genes mapped to 122 unique KEGG pathways, with most pathways represented by a single gene, reflecting broad functional dispersion rather than concentration in a small number of routes. Prominent pathway classes included amino acid, carbohydrate, lipid, and nucleotide metabolism, as well as signal transduction and cellular process categories. The prevalence of core metabolic routes aligns with the GO-based enrichment of metabolic processes and suggests potential epigenetic modulation of energy balance and biosynthesis.

3.5 Pfam domains

We detected 118 unique Pfam domains among DML-associated genes, generally with low redundancy (most domains present in one or two genes), underscoring protein family diversity. Recurrently observed domains included the RNA recognition motif (PF00076)—implicating post-transcriptional regulation—alongside regulatory/interaction modules such as BTB/POZ (PF07707) and ankyrin repeats (PF00023). We also observed immunoglobulin domains (PF00047) and cytochrome P450 families (PF00061/ PF00067), the latter consistent with roles in xenobiotic and lipid metabolism. Several domains of unknown function (DUFs) were present, indicating potentially novel or poorly characterized targets of methylation.

3.6 Directionality of methylation changes

Given the negative median methylation difference, hypomethylation was more common among DML genes. Functionally, hypomethylated targets were enriched for regulatory proteins, signaling molecules, and transporters, whereas hypermethylated targets included transcription factors, metabolic enzymes, and structural proteins. Collectively, these patterns are consistent with widespread epigenetic remodeling across regulatory and metabolic layers, rather than changes confined to a single pathway or compartment.

3.7 Summary

Together, these results show that genes harboring DMLs in Mytilus chilensis are functionally diverse. However, they converge on metabolism, development, and cellular organization, with many targets positioned in cytoplasmic and nuclear environments. The distribution of DMLs is skewed toward hypomethylation. DML-associated genes are dispersed across many KEGG pathways and contain regulatory or interaction-rich Pfam domains. These findings suggest a system-wide epigenetic response with potential consequences for metabolic control and developmental programming.

Castillo, N., Saavedra, L. M., Vargas, C. A., Gallardo-Escárate, C., & Détrée, C. (2017). Ocean acidification and pathogen exposure modulate the immune response of the edible mussel *mytilus chilensis*. *Fish & Shellfish Immunology*, *70*, 149–155. https://doi.org/10.1016/j.fsi.2017.08.047

Montúfar-Romero, M., Valenzuela-Muñoz, V., Valenzuela-Miranda, D., & Gallardo-Escárate, C. (2024). Hypoxia in the blue mussel *mytilus chilensis* induces a transcriptome shift associated with endoplasmic reticulum stress, metabolism, and immune response. *Genes*, *15*(6), 658. https://doi.org/10.3390/genes15060658

Montúfar-Romero, M., Valenzuela-Miranda, D., Valenzuela-Muñoz, V., Morales-Rivera, M. F., & Gallardo-Escárate, C. (2025). Microbiota dysbiosis in *mytilus chilensis* is induced by hypoxia, leading to molecular and functional consequences. *Microorganisms*, *13*(4), 825. https://doi.org/10.3390/microorganisms13040825

Yévenes, M., Núñez-Acuña, G., Gallardo-Escárate, C., & Gajardo, G. (2021). Adaptive differences in gene expression in farm-impacted seedbeds of the native blue mussel *mytilus chilensis*. *Frontiers in Genetics*, *12*, 666539. https://doi.org/10.3389/fgene.2021.666539

Yévenes, M., Núñez-Acuña, G., Gallardo-Escárate, C., & Gajardo, G. (2022). Adaptive mitochondrial genome functioning in ecologically different farm-impacted

202 natural seedbeds of the endemic blue mussel *mytilus chilensis*. *Comparative*
203 *Biochemistry and Physiology Part D: Genomics and Proteomics*, *42*, 100955.
204 https://doi.org/10.1016/j.cbd.2021.100955
205 Yévenes, M., Gallardo-Escárate, C., & Gajardo, G. (2023). Epigenetic variation me-
206 diated by lncRNAs accounts for adaptive genomic differentiation of the endemic
207 blue mussel *mytilus chilensis*. *Heliyon*, *10*(1), e23695. https://doi.org/
208 10.1016/j.heliyon.2023.e23695
209 Yévenes, M., Gajardo, G., & Gallardo-Escárate, C. (2025). Decoding local adapta-
210 tion in the exploited native marine mussel *mytilus chilensis*: Genomic evidence
211 from a reciprocal transplant experiment. *International Journal of Molecular*
212 *Sciences*, *26*(3), 931. https://doi.org/10.3390/ijms26030931