

Seagrass Findings — *Zostera marina* Epigenomics

DNA methylation, RdDM machinery, and genetic structure across 20 Zostera marina Nanopore methylomes — the analytical foundation for an RNA-directed methylation strategy.

20

NANOPORE METHYLOMES

~82M

METHYLATION SITES / SAMPLE

20 / 20

RDDM GENES PRESENT

260 Mb

V3.1 REFERENCE GENOME

Overview & Scope

This page reports the computational analyses performed on 20 Zostera marina (eelgrass) samples sequenced by Decibel Bio (Oxford Nanopore, 5-methylcytosine). It is the technical companion to the project overview. Every figure and table here is reproducible from the data bundle linked at the bottom.

Analyses use the chromosome-scale **v3.1** *Zostera marina* assembly (Ma et al. 2021, F1000Research; 6 pseudo-chromosomes, 260.5 Mb, 21,483 genes) — the assembly the methylation was called against. An earlier scaffold-level assembly (v2.1 / GCA_001185155.1, Olsen et al. 2016) is a distinct reference and is not used for coordinate-based analyses here.

What was measured

- **DNA methylation** in all three plant sequence contexts (CG, CHG, CHH), per site, from modkit bedMethyl calls
- **RdDM machinery** — presence/conservation of the 20-gene RNA-directed DNA methylation pathway
- **Genetic structure** — 3.98M biallelic SNPs across the 20 samples, clonality and relatedness
- **Feature-level methylation** — methylation partitioned into promoters, exons, introns, gene bodies, intergenic regions, and repeats/TEs
- **Comparative** — RdDM conservation across six seagrass genomes spanning all four families (Thalassia, Halophila, Cymodocea, Amphibolis, Posidonia vs. *Zostera*)

Why these analyses matter

- The project's non-GMO strategy depends on the **natural RdDM pathway being intact** — now demonstrated (20/20 genes present)
- **Promoters carry the highest non-CG methylation** — the natural handle RdDM uses for targeted gene control
- The pathway is **conserved across all six seagrass genomes** (all four families), so methods developed in compact *Zostera* should transfer across the group
- A methylation + genetic baseline across 20 individuals defines the natural variation any engineered change must be read against

1. The RdDM Machinery Is Intact

*RNA-directed DNA methylation (RdDM) is the plant's natural, non-transgenic route to targeted gene silencing. The whole strategy depends on that pathway being present and complete in *Zostera marina*, so we inventoried all 20 of its core genes.*

All 20 core RdDM genes are present and conserved in *Zostera marina*.

The RNA-directed DNA methylation pathway is complete in *Zostera marina*: 13 strong orthologs, 6 moderate, 1 weak, none absent. Every functional module is represented, from the Pol IV/Pol V transcription apparatus through the siRNA-biogenesis arm to the de novo methyltransferase and the ROS1/DME demethylases that reverse marks. Any RdDM-based approach requires this machinery to be in place, and in *Zostera* it is.

Twenty reviewed *Arabidopsis thaliana* RdDM / DNA-methylation reference proteins were searched (DIAMOND blastp, --very-sensitive, $e < 1e-5$) against the *Zostera marina* proteome (20,648 proteins). Each was scored by amino-acid identity and query coverage into Strong / Moderate / Weak ortholog-confidence calls.



The complete pathway is present: Pol IV (NRPD1) and Pol V (NRPE1) as distinct genes, the shared second subunit NRPD2, the siRNA-biogenesis arm (RDR2, DCL3), the AGO4/AGO6 effectors, the recruitment/ chromatin module (SHH1, CLSY1, DRD1, DMS3, RDM1, SUVH2/9), the de novo methyltransferase DRM2, the maintenance methyltransferases (MET1, CMT3, CMT2), and — relevant for epiallele stability — the demethylases ROS1 and DME.

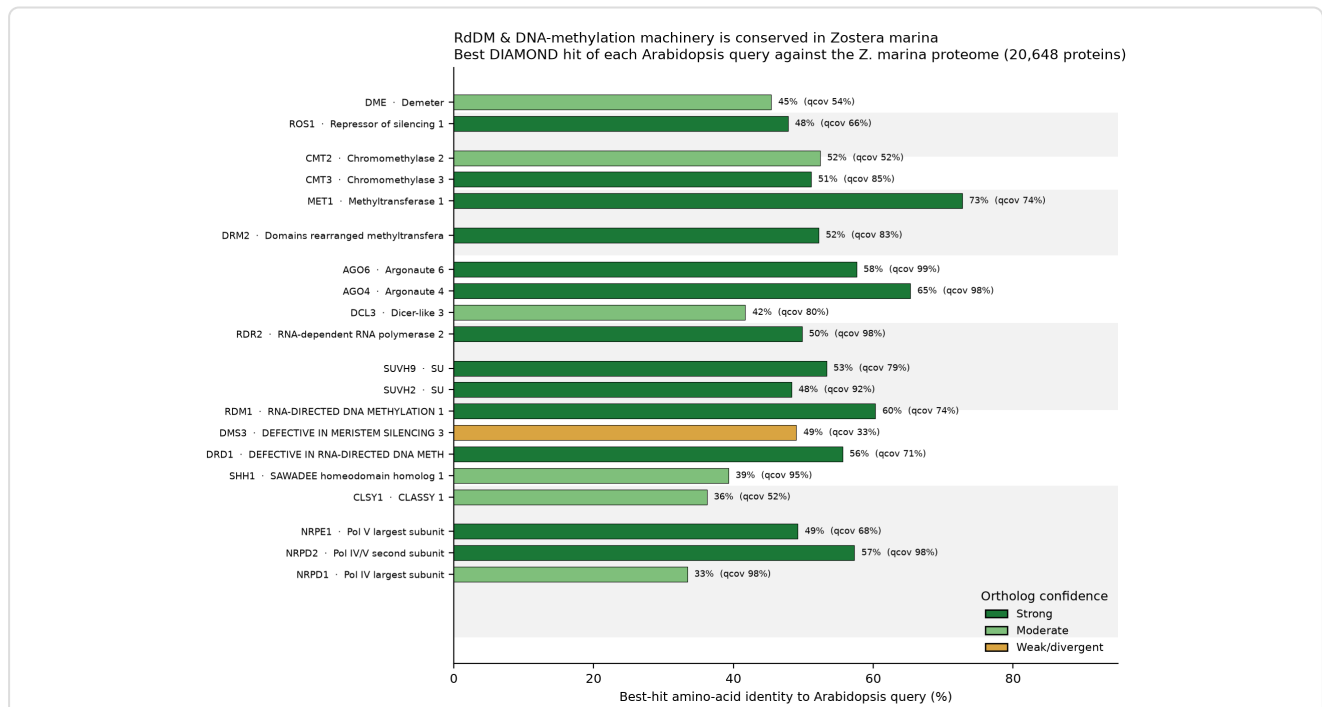


Figure 1. RdDM and DNA-methylation machinery is conserved in *Zostera marina*. Best DIAMOND hit of each Arabidopsis query against the *Z. marina* proteome; bar length = amino-acid identity, colour = ortholog confidence. No component is absent.

The complete inventory

All 20 components, grouped by pathway module, with the best ortholog hit and confidence call in each species. Both seagrasses carry every component; *Thalassia* values come from the comparative scan in §2.

GENE	MODULE	ROLE	ZOSTERA HIT	Z. %ID	Z. CALL	THALASSIA HIT	T. %ID	T. CALL
<i>NRPD1</i>	Pol IV/V	Pol IV largest subunit	KMZ70232.1	33%	Moderate	Thate01g01330	36%	Moderate

GENE	MODULE	ROLE	ZOSTERA HIT	Z. %ID	Z. CALL	THALASSIA HIT	T. %ID	T. CALL
<i>NRPE1</i>	Pol IV/V	Pol V largest subunit	KMZ58649.1	49%	Strong	Thate01g10480	41%	Moderate
<i>NRPD2</i>	Pol IV/V	Pol IV/V 2nd subunit	KMZ68176.1	57%	Strong	Thate05g12060	58%	Strong
<i>CLSY1</i>	Recruitment	Chromatin remodeler	KMZ57147.1	36%	Moderate	Thate01g02240	37%	Moderate
<i>SHH1</i>	Recruitment	Pol IV recruitment	KMZ64381.1	39%	Moderate	Thate09g03450	42%	Moderate
<i>RDR2</i>	siRNA biogenesis	dsRNA synthesis	KMZ58731.1	50%	Strong	Thate08g04430	54%	Strong
<i>DCL3</i>	siRNA biogenesis	24-nt siRNA dicer	KMZ59536.1	42%	Moderate	Thate02g35430	43%	Moderate
<i>AGO4</i>	Effector	siRNA effector	KMZ71001.1	65%	Strong	Thate01g03750	68%	Strong
<i>AGO6</i>	Effector	siRNA effector	KMZ69032.1	58%	Strong	Thate03g26310	61%	Strong
<i>DRD1</i>	DDR / recruitment	DDR complex helicase	KMZ74703.1	56%	Strong	Thate02g12400	58%	Strong
<i>DMS3</i>	DDR / recruitment	DDR complex	KMZ73474.1	49%	Weak/divergent	Thate09g08050	51%	Strong
<i>RDM1</i>	DDR / recruitment	DDR complex	KMZ64644.1	60%	Strong	Thate03g04120	57%	Strong
<i>SUVH2</i>	Pol V recruitment	Pol V recruitment	KMZ64142.1	48%	Strong	Thate03g07440	46%	Strong
<i>SUVH9</i>	Pol V recruitment	Pol V recruitment	KMZ64142.1	53%	Strong	Thate04g01200	58%	Strong
<i>DRM2</i>	Methyltransferase	de novo (CHH)	KMZ59545.1	52%	Strong	Thate02g36290	53%	Strong
<i>MET1</i>	Methyltransferase	CG maintenance	KMZ71047.1	73%	Strong	Thate08g10510	65%	Strong
<i>CMT3</i>	Methyltransferase	CHG maintenance	KMZ58195.1	51%	Strong	Thate05g07140	48%	Strong
<i>CMT2</i>	Methyltransferase	CHH maintenance	KMZ58195.1	52%	Moderate	Thate05g07140	57%	Moderate
<i>ROS1</i>	Demethylase	active demethylation	KMZ69251.1	48%	Strong	Thate08g07090	45%	Moderate
<i>DME</i>	Demethylase	active demethylation	KMZ69251.1	45%	Moderate	Thate07g01050	48%	Moderate

Confidence calls: **Strong** $\geq 45\%$ identity & $\geq 70\%$ query coverage; **Moderate** $\geq 30\%$ & $\geq 50\%$; **Weak/divergent** below that but a clear reciprocal match. Hit IDs are the best-scoring protein in each proteome (*Zostera* KMZ from *GenBank*; *Thalassia* Thate from the annotated proteome).

Two structural notes: AGO4 has expanded into a small clade in *Zostera* (distinct AGO4A/AGO4B genes), while several *Arabidopsis* paralog pairs (SUVH2/9, CMT2/3, ROS1/DME) collapse onto single shared *Zostera* genes – the ancestral single-copy state, with the duplications being *Arabidopsis*-lineage specific.

2. RdDM Conservation Across Six Seagrass Genomes

Finding the pathway complete in *Zostera* does not by itself show it is a general seagrass feature. To test that, we searched for the same 20 genes in five more seagrasses covering all four seagrass families.

We searched the same 20 Arabidopsis queries against five additional seagrasses spanning all four seagrass families: *Thalassia testudinum* and *Halophila stipulacea* (Hydrocharitaceae); *Cymodocea nodosa* and *Amphibolis antarctica* (Cymodoceaceae); and *Posidonia oceanica* (Posidoniaceae) — versus *Zostera marina* in Zosteraceae. *Thalassia* was searched against its annotated proteome (25,744 proteins); the other four, whose annotations are not openly available, were searched directly against their genome assemblies by splice-aware protein-to-genome alignment (miniprot). Two of the six — *Amphibolis* and *Halophila* — were added in a second pass to bring within-family replication to two of the four families; both were scanned with the identical query set, aligner, and confidence thresholds.

The RdDM pathway is conserved across all six seagrass genomes and all four families.

All six seagrasses carry the full RdDM pathway: 18 to 20 of the 20 genes are found in each. The central effector and maintenance genes (RDR2, AGO4/6, MET1, CMT3) are a strong match in all six. Where a gene falls short, it is mostly the same few — NRPD1 and SHH1, which are short or fast-changing and hard to detect by genome search — rather than a different gene in each species, the pattern expected when a pathway is genuinely shared rather than matched by chance. This means RdDM is an old, built-in feature of seagrasses, not something unique to *Zostera*, so methods worked out in the small, easy-to-handle *Zostera* genome have a good chance of carrying over to the others.

6

SEAGRASS GENOMES TESTED

18–20 / 20

COMPONENTS DETECTED PER SPECIES

4

SEAGRASS FAMILIES SPANNED

0

COMPONENTS ABSENT IN ALL SIX

The RdDM pathway is conserved across six seagrass genomes

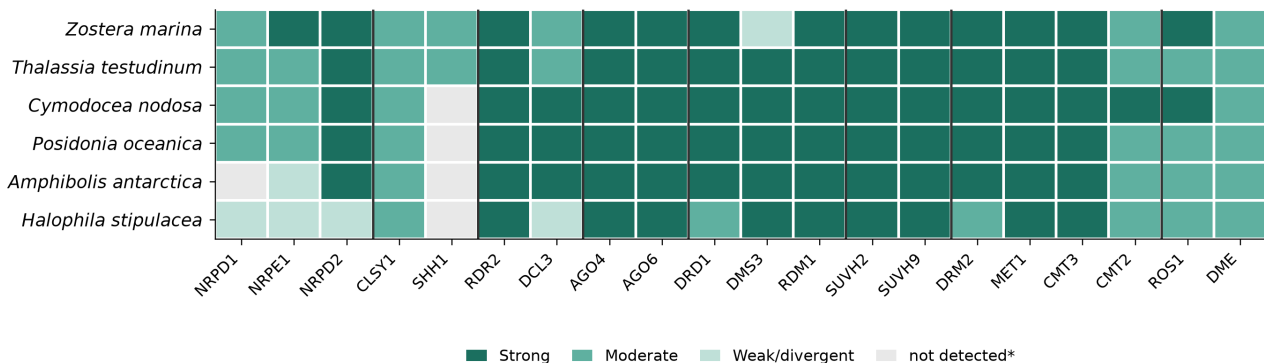


Figure 2. Each row is one seagrass species; each column is one of the 20 genes of the RdDM pathway, grouped by function. Darker cells mean a closer match to the known Arabidopsis version of that gene (Strong to not-detected). The dark band running across every row is the core machinery — it is present in all six species. Methods: *Zostera* and *Thalassia* were searched against their annotated protein sets; the other four against their genome sequences by protein-to-genome alignment. The lighter cells for *Halophila* reflect its fragmented genome assembly (which breaks long genes across pieces and lowers the match score), not a missing gene.

Because the same genes match in the same way across six separate genomes, these are almost certainly the true seagrass versions of the RdDM genes and not lookalikes. One gene, DMS3, looked weak in *Zostera* only because its gene model there is incomplete; it is a strong match in all five other seagrasses, so the pathway is clearly

complete in the group. Two of the six species — *Amphibolis* and *Halophila* — were added as second members of families we already had, which lets us check that the pathway holds up within a family and not just between families. *Amphibolis* matched the same pattern as its relatives *Cymodocea* and *Posidonia*, including the same two hard-to-detect genes, confirming that those gaps are a feature of that plant family rather than a failure of the search. The two genes that remove methylation (*ROS1* and *DME*) are present in every species, which matters because it means any change we add can also be undone by the plant itself. Since the machinery is shared across all four seagrass families, a method that works in *Zostera* is a strong candidate to transfer to the others, most immediately to *Thalassia* (whose genome is about ten times larger) and to the heat-tolerant *Cymodoceaceae*.

Four of the six species were scored by protein-to-genome alignment against their raw assemblies because their annotated proteomes are not openly available. This is slightly less sensitive than a proteome search for small, divergent genes (hence the *NRPD1*/*SHH1* gaps). *Halophila stipulacea* is a special case: its assembly is highly fragmented (3.2 Gb, scaffold N50 ~8 kb), which splits large multi-exon genes across short scaffolds and lowers their match scores — *NRPD2* and *DCL3* drop to Weak/divergent and *DRD1* and *DRM2* to Moderate, where the same genes are Strong in the better assemblies. The main effectors (*RDR2*, *AGO4/6*, *MET1*, *CMT3*) still score Strong, so the pathway is clearly present even in this fragmented genome. Its calls should be read as a lower-confidence presence check; the other five genomes carry the conservation claim, and *Halophila* is consistent with it. When better assemblies or annotated proteomes become available, these columns can be refined to proteome-based calls.

Public genome availability, by species

The table below is the assembly picture from NCBI GenBank at the time of writing. Two things stand out: the best *Zostera marina* assembly is not on GenBank at all (the fragmented 2015 scaffold set is; the chromosome-scale v3.1 this project uses lives at JGI/Phytozome), and *Halodule wrightii* — the most heat-tolerant species — has **no public assembly of any kind**. Where good genomes do exist (*Thalassia*, *Cymodocea*, *Posidonia*, all assembled 2024), there is still no base-resolution methylome for any seagrass beyond the *Zostera* data generated here.

SPECIES	FAMILY	BEST PUBLIC GENOME	LEVEL	CONTIGUITY	NOTE
<i>Zostera marina</i>	Zosteraceae	JGI/Phytozome v3.1 (used here)	Chromosome	6 pseudo-chromosomes, 260.5 Mb	GenBank copy (GCA_001185155.1, 2015) is fragmented (scaffold N50 0.49 Mb)
<i>Halodule wrightii</i>	Cymodoceaceae	none	—	—	No public assembly — highest-value new genome
<i>Thalassia testudinum</i>	Hydrocharitaceae	GCA_037157565.1 (2024)	Scaffold	contig N50 ~112 Mb, 4.26 Gb	Good genome; no methylome
<i>Cymodocea nodosa</i>	Cymodoceaceae	GCA_036874045.1 (2024)	Chromosome	N50 ~21 Mb, 375 Mb	Good genome; no methylome
<i>Posidonia oceanica</i>	Posidoniaceae	GCA_037176725.1 (2024)	Scaffold	contig N50 ~68 Mb, 2.96 Gb	Good genome; no methylome
<i>Amphibolis antarctica</i>	Cymodoceaceae	GCA_024699675.1 (2022)	Contig	N50 ~40 kb, 245 Mb	Fragmented — presence checks only
<i>Halophila stipulacea</i>	Hydrocharitaceae	GCA_047748335.1 (2025)	Scaffold	N50 ~8 kb, 3.2 Gb	Fragmented — presence checks only

This is the evidence base for the collection priorities in §4: generate a genome where none exists (*Halodule*), and generate methylomes broadly, since whole-genome base-resolution methylation maps are effectively absent from the public record for every seagrass.

3. From Pathway to Trait

With the machinery confirmed, this section lays out how it would be put to use: which traits methylation could plausibly change, how the change would be delivered, and how each trait would be measured. Specific target genes await phenotype-linked methylation data.

RdDM is the sequence-targetable arm of the plant methylation system. A 24-nt small-RNA trigger complementary to a locus recruits the pathway to methylate that locus, biased toward the CHH context. Because targeting derives from an RNA sequence rather than a nuclease or an inserted gene, RdDM is compatible with a non-transgenic strategy: the machinery a plant already uses to silence its transposons can be directed to a chosen regulatory region.

The feature-level data locate the natural handle. Promoters carry the highest non-CG methylation of any genic feature (CHH 3.3%; §5), and non-CG methylation is the RdDM-maintained mark, so the pathway already acts at the regulatory regions relevant to gene control. Prior seagrass work supports a methylation–phenotype link: in a clonal *Zostera marina* meadow, gene-body CG methylation covaried with photosynthetic performance independent of genetic background, with the associated genes enriched for light-harvesting and protein-folding functions (Jueterbock et al. 2020). In *Posidonia oceanica* and *Cymodocea nodosa*, warm-origin plants carry higher global methylation than cold-origin plants, and heat-responsive genes are concentrated in the low-gene-body-methylation, high-plasticity fraction of the genome (Entrambasaguas et al. 2021).

Trait candidates and measurement

The traits below are the current candidates for engineering or selection. Each one is tied to a documented cause of seagrass-restoration failure: the 2024 USDA/NOAA report to Congress catalogues why restorations fail (Appendix 4, Orth, Lefcheck and colleagues), and nearly every trait here addresses one of those failure modes. Targeting a known failure mode means a success in the lab translates into better establishment and survival in the field, which is the main bottleneck in restoration. Each trait needs a quantitative assay before it can be used as a selection endpoint; the proposed measurements are listed alongside.

TRAIT	RATIONALE	RESTORATION FAILURE MODE ADDRESSED (FEDERAL REPORT)	PROPOSED QUANTITATIVE READOUT
Pathogen resistance	<i>Labyrinthula zosterae</i> (wasting disease) is a primary threat to <i>Zostera</i> meadows	"Disease (e.g., fungal attack on seeds or seedlings)"	Lesion area / infection score; challenge assay survival
Temperature tolerance	Central climate-resilience goal	Marine heatwaves named the emerging primary threat; assisted gene flow from warm-adapted stock is the stated strategy	Biomass retention and leaf coloration under heat ramp (multi-spectral imaging)
Photosensitivity	Light regime varies across restoration sites	"Too shallow (desiccation) or too deep (insufficient light)"; poor water clarity	Biomass and coloration under controlled irradiance
Salinity tolerance	Estuarine and dredging-affected sites vary widely	Site-selection failures at variable-salinity estuarine and placed-sediment sites	Growth vs. water conductivity gradient
Carbon fixation	Directly tied to blue-carbon value	(value-enhancing rather than failure-linked)	Leaf/biomass accrual plus sediment contribution modeled from root oxygenation
Seed success ratio	Restoration throughput from seed	"Lack of donor material or seed stock (e.g., no flowering)"	Germination and establishment fraction

TRAIT	RATIONALE	RESTORATION FAILURE MODE ADDRESSED (FEDERAL REPORT)	PROPOSED QUANTITATIVE READOUT
Sulfide (H ₂ S) tolerance	Sulfidic sediments limit survival	"Sediment instability ... smothering and burial of seedlings"	Growth/survival across H ₂ S dosing
Hypoxic-sediment durability	Low-oxygen sediments common at degraded sites	Sediment/anchorage failures; night-time hypoxia in dense beds	Survival and rhizome growth under sediment hypoxia
Epiphyte-overgrowth resistance	Epiphyte load reduces photosynthesis; grazers (nudibranchs) mediate cleaning	"Algal blooms and/or excessive epiphyte growth"; "(over)grazing of transplants"	Epiphyte cover; grazer preference assay (harder to standardize)

Non-CG methylation status varies by tissue and ramet age, so assays and sampling must be standardized: roots are methylomically distinct from leaves and rhizomes (thousands of differentially methylated sites, versus ~18 between leaves and rhizomes), and young versus old ramets separate in methylation ordination (Nilsen et al. 2025). Leaves and rhizomes are interchangeable sampling proxies; roots are not.

Identifying which methylation marks to target

Naming a trait is not the same as knowing which cytosines to change to affect it. Going from one to the other is a distinct step. Much of the groundwork for it is what the analyses in this document produce; the final target-ranking is where Decibel's platform comes in.

The target-identification logic runs in five stages:

- 1. Measure the trait in plants that differ in it.** Collect plants that vary in the trait — for heat tolerance, the warm- and cold-adapted eelgrass ecotypes the collection strategy in §4 targets — and score each with the quantitative assay from the table above.
- 2. Read their methylomes.** Sequence the methylome of each plant (Oxford Nanopore 5mC, as was done for the 20 samples here). This is the reference layer this project builds.
- 3. Find the marks that track the trait.** Compare the methylomes of high-trait versus low-trait plants to find differentially methylated regions (DMRs) — stretches of the genome that are consistently more or less methylated in the resilient plants.
- 4. Remove the marks that are just inherited.** This is the step the coupling result in §8 makes unavoidable. Because methylation similarity tracks genetic similarity here (Mantel $r = 0.77$), a naive DMR list will be dominated by marks that are simply linked to genotype and cannot be changed independently. The genetic relatedness has to be regressed out first, leaving the *facultative* marks — the ones that respond to environment and can actually be written or erased. Those are the only useful targets.
- 5. Keep the marks that sit on a control point.** Intersect the remaining DMRs with the genome annotation and keep those over gene promoters or the transposable elements next to genes — the positions where a methylation change actually turns a gene up or down. The feature-methylation analysis in §5 shows this is where the RdDM-type (CHH) signal concentrates, so those are the positions where an applied change has the most leverage.

The output is a short list of candidate sites: a specific promoter or region, in the facultative fraction, on a gene with a plausible tie to the trait. Ranking those candidates, deciding which to write, and designing the small-RNA trigger that writes them is the step a trait-prediction platform such as Decibel's is expected to perform. Decibel's internal methods are proprietary, so this states where their platform fits rather than how it works. In this project's terms: the analyses here produce the reference methylomes, the genome coordinates, the separation of inherited from facultative marks, and the trait-contrasting source plants; the platform uses those to name and prioritize the specific targets to write.

The current gap is stage 1. We have the methylomes (stage 2) and the analytical methods for stages 3–5, but not yet the phenotype-linked collection stage 1 requires. Assembling that trait contrast is the near-term experimental priority, and it is what the collection strategy in §4 is designed to produce.

Delivery routes

There are two fundamentally different ways to get a plant to carry a useful methylation mark, and they lead to three practical routes. The first way is to **add** the mark ourselves — deliver a small-RNA trigger that tells the RdDM pathway to methylate a chosen gene. The second way is to **skip induction entirely** and instead find plants that already carry the mark naturally, then propagate them. The three cards below follow that split: Routes A and B are two ways of adding the mark (they differ only in how the trigger reaches the plant), and Route C is the no-induction alternative. All three are non-transgenic, and each has been shown to work in other plants but not yet in a seagrass.

Route A is the one Decibel Bio's platform performs: it reads the methylome and then writes a new mark using an externally applied spray or seed treatment. Routes B (letting the plant move the trigger internally) and C (selecting plants that already carry the mark) are alternatives that would not rely on Decibel's write step.

Route A — Apply the trigger from outside

- **Externally applied trigger** — a small-RNA trigger is applied to the plant or seed from the outside, and the plant's own RdDM pathway writes the mark. No transgene. This is the mechanism Decibel's platform uses.
- **Spray-induced silencing (SIGS)** — apply RNA that the plant processes into the 24-nt siRNAs that guide RdDM.
- **Water-dispersed or pelletized delivery** — formulations for applying the trigger to established plants underwater (detailed under *Route A in detail* below).
- **Pre-treated seed** — treat seed before sowing, then track establishment.

Route B — Let the plant carry the trigger

- The trigger is introduced once and the plant moves it internally, rather than being applied to every target.
- **Graft-transmissible siRNA** — mobile siRNAs travel from a trigger-donor tissue into the target plant.
- **Co-inoculation with beneficial bacteria** — pair trigger delivery with dispersal of helpful microbial colonies.

Route C — Select, don't induce

- No trigger at all: find plants that already carry the favourable mark and propagate them.
- **Natural-epiallele selection** — identify and clonally propagate favourable methylation states already present in warm-adapted individuals.
- **Reversibility** — the demethylase genes ROS1/DME are present in all six seagrass genomes (§2), so any induced mark can also be actively removed.

Experimental plan

Candidate traits would be screened in aquarium trials using a multi-spectral camera to quantify biomass and coloration. Two starting materials are available on different timelines: eelgrass grown from seed reaches assay size in approximately six months, while live plants collected from multiple geographic locations provide immediate material together with the environmental and trait data of their source sites, supporting comparison of natural variation before any manipulation. Once RdDM material is in hand, water-dispersal and pelletized-dispersal methods can be tested on live plants alongside RdDM pre-treated seed, with parallel trials of co-inoculation with beneficial bacterial colonies dispersed simultaneously.

The near-term priorities are: (1) generate phenotype-linked methylomes to identify specific target loci for the traits above; (2) pilot one delivery route — SIGS has the lowest barrier — against a reporter locus to confirm the

trigger-to-methylation step operates in *Zostera*; and (3) screen warm-adapted ecotypes for pre-existing favourable epialleles that require no induction.

Route A in detail: Decibel's platform and the underwater delivery problem

Route A above is the path we expect to use. We plan to make the methylation change using technology from **Decibel Bio**, a Berkeley company that has already sequenced the 20 *Zostera marina* genomes and methylomes described in these findings. Decibel's platform has two parts: it reads a plant's methylation pattern and predicts traits from it, and it writes new methylation using a spray or seed treatment that does not alter the plant's DNA. The change is reversible, keeps the plant's own genetic diversity, and can be applied at planting or later in the season — the same conditions this project requires. Decibel's published example treats a crop so it protects its photosynthesis under extreme heat, which is the trait we lead with in eelgrass. One caution: Decibel's public materials describe the platform only in general terms and do not spell out the molecular chemistry, so the exact design of the trigger should be confirmed with the company rather than assumed from press coverage.

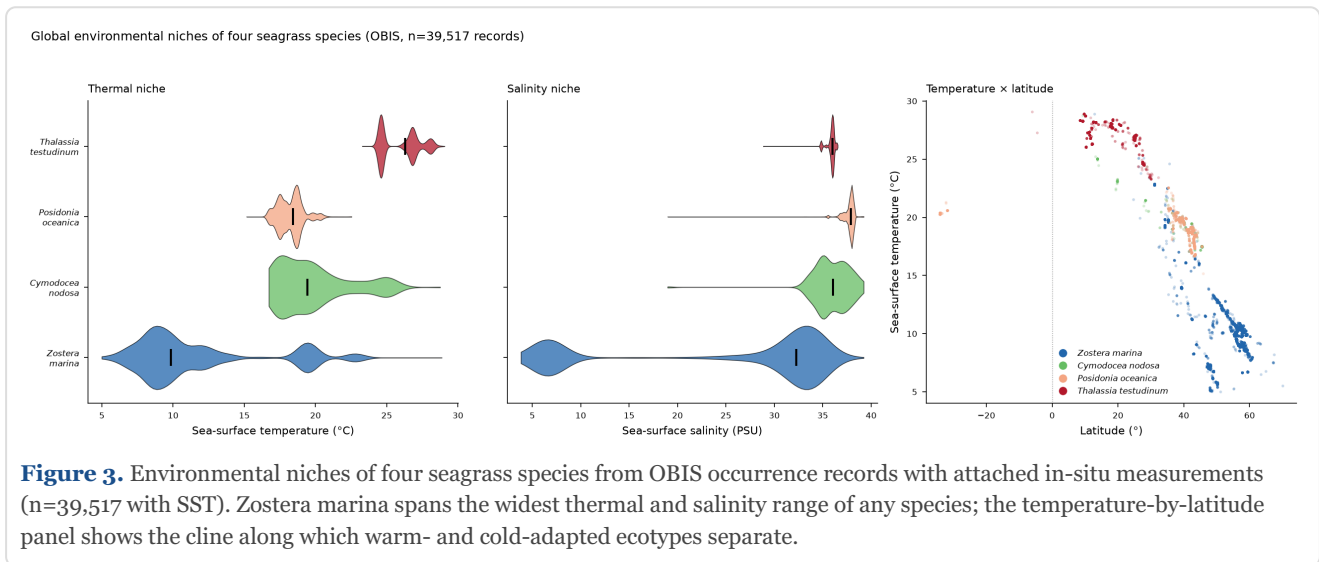
Decibel's platform was built for land crops, where a spray lands on leaves in the air. Eelgrass grows underwater, so the main thing to work out is how to apply the treatment to a submerged plant. There are three ways to do this. The simplest is to **treat the seed before we plant it**: this uses Decibel's existing seed treatment with little change, and since we plant from seed anyway it needs no new underwater step. Treated seed alone is enough to move the project forward, so the underwater methods below would add reach rather than being required. The second way is to **release the treatment into the water or sediment** around established plants. The third is to **deliver it in pellets** — a technology being developed by Chris Oakes (former CEO and founder of ReefGen) that would protect the treatment and release it in a controlled way at the seafloor, which helps with the dilution and wash-away problems of loose release. The second and third methods would let us treat existing meadows, not just newly planted seed; they are more valuable but less proven, and are the ones to test next.

The underlying method is known to work in a land plant: spraying short RNA molecules onto tobacco leaves added methylation to a target gene and switched it off (Walz et al., "Induction of promoter methylation and transcriptional gene silencing upon high-pressure spraying of 24-nt small RNAs in *Nicotiana benthamiana*," bioRxiv 2022, doi:10.1101/2022.03.14.484340). So the biology is sound; what remains is adapting the delivery for use underwater, which is the practical problem to work through with Decibel and the pelletization partner. A fuller write-up of Decibel's public technology is in `decibel_bio_research_brief.md` (Downloads).

4. Source Populations & Collection Strategy

To breed or select for a trait like heat tolerance, we need starting plants that already vary in that trait. This section maps where each project species lives, the environmental conditions each population experiences, and which populations are practically collectable — with emphasis on temperature and on sites shippable to the Bay-Area sequencing location. Data: 65,932 georeferenced records from OBIS (marine, with per-record sea-surface temperature, salinity and depth) and GBIF (freshwater).

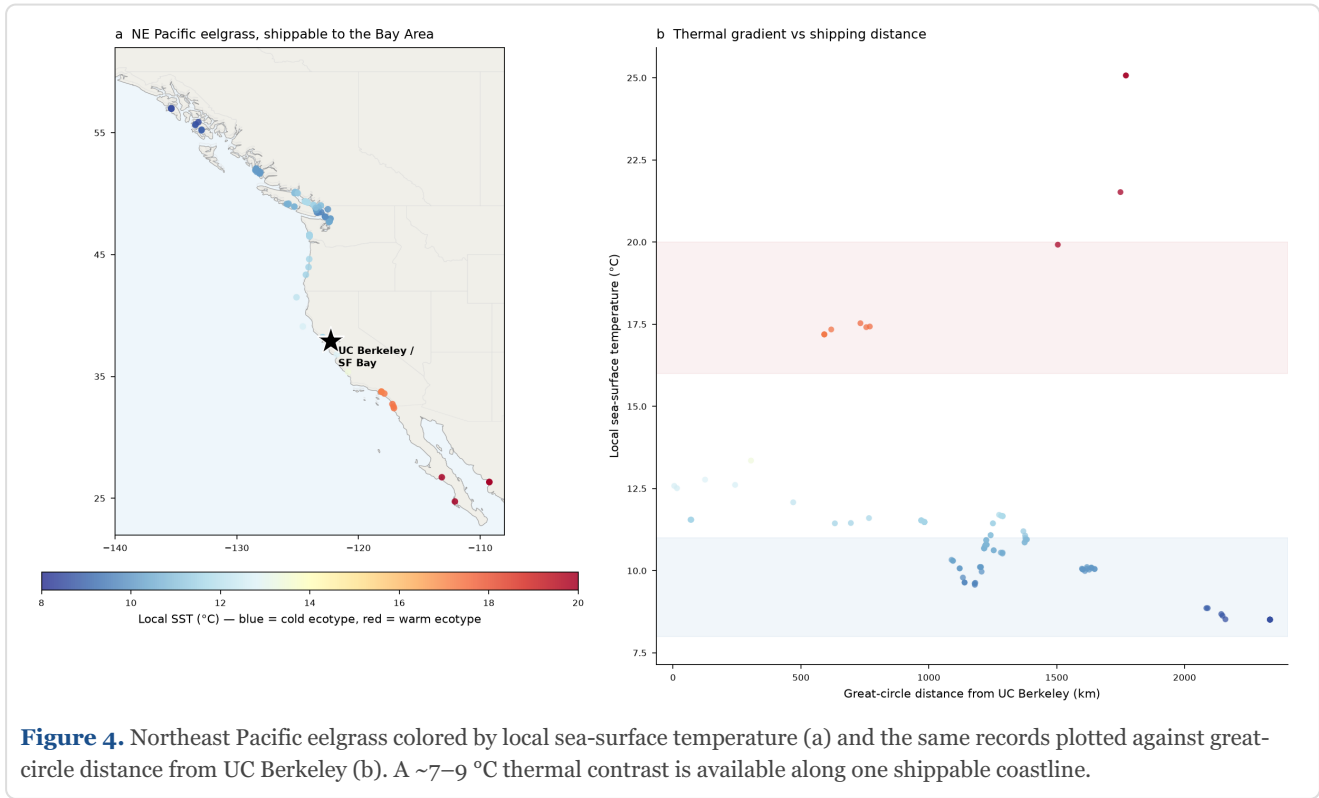
Among the six project species, *Zostera marina* occupies by far the widest thermal niche — a 24 °C span from 5 °C in Arctic Norway, Alaska and the White Sea to 29 °C at its warm range edge — and the widest salinity range, from full marine to the near-fresh Baltic. Warm- and cold-adapted ecotypes of the *same genome* are therefore available to collect and contrast, which is exactly the raw material an epigenetic temperature program requires. The strictly warm species (*Thalassia* median 26 °C, *Halodule* 25 °C) and the thermally narrow Mediterranean endemic *Posidonia* (15–22 °C) offer much less internal contrast.



Each occurrence carries its in-situ SST, salinity and depth, so every collected population arrives with an environmental label already attached — the substrate for associating methylation marks with thermal origin (cf. Jueterbock et al. 2020; Entrambasaguas et al. 2021, §3).

A Bay-Area-centered collection plan

The Northeast Pacific is the most practical first collection axis: a single coastline carrying a strong thermal gradient, reachable from the Bay Area by cold-chain courier. Occurrences of *Zostera marina* were binned by latitude and ranked by distance from Berkeley.



ROLE	CANDIDATE ZONE	MEDIAN SST	NEAREST DISTANCE FROM BERKELEY
Local reference	SF Bay / Monterey	11.5 °C	~5 km
Warm-edge	S California (Channel Is. / San Diego)	17–17.5 °C	~590–730 km
Warm-edge (extreme)	N Baja California	25 °C	~1,770 km

ROLE	CANDIDATE ZONE	MEDIAN SST	NEAREST DISTANCE FROM BERKELEY
Cold-edge	BC / N Washington	9.8–10.9 °C	~1,100–1,600 km
Cold-edge (extreme)	SE Alaska	8.5 °C	~2,100 km

A minimal design takes three points on one coast — San Diego/Channel Islands (warm), SF Bay (local reference), and BC/Puget Sound or SE Alaska (cold) — to bracket a ~7–9 °C thermal contrast within a single genetic lineage and a single shipping corridor. That is enough to build the first phenotype-linked methylome contrast the trait program needs; the Kuroshio Current (S Japan, median 19 °C) and Mediterranean populations extend the warm end further if a wider gradient is wanted later.

The eelgrass population-genetics literature sets the resolution at which provenance must be tracked. *Zostera marina* produces dormant, negatively buoyant seeds that disperse only a few metres, so meadows differentiate genetically on a scale below 100 km, and local adaptation between meadows separated by as little as 10 km can determine transplant survival (DuBois et al. 2022). Genome-wide association work in adjacent bays has resolved specific loci underlying eelgrass thermal tolerance (Schiebelhut et al. 2023), so a heritable, mappable warm/cold signal already exists in this species. Collecting the warm and cold ecotypes above with both genotype and methylome in hand lets the plastic (methylation) and heritable (sequence) components of thermal response be separated in the same material — a stronger design than either layer alone, and one that stays within locally-appropriate genetic backgrounds rather than moving fixed alleles across the range.

Vallisneria americana, the project's one freshwater species, lives in interior lakes and rivers of North America and is included as a heat-tolerant aquatic-plant comparator (Chesapeake ecotypes tolerate 33–36 °C); it is treated separately from the marine collection plan. Full distribution maps, per-species ecosystem rankings, and data caveats (survey bias, undersampling of the immediate California coast) are in the biogeography report under Downloads.

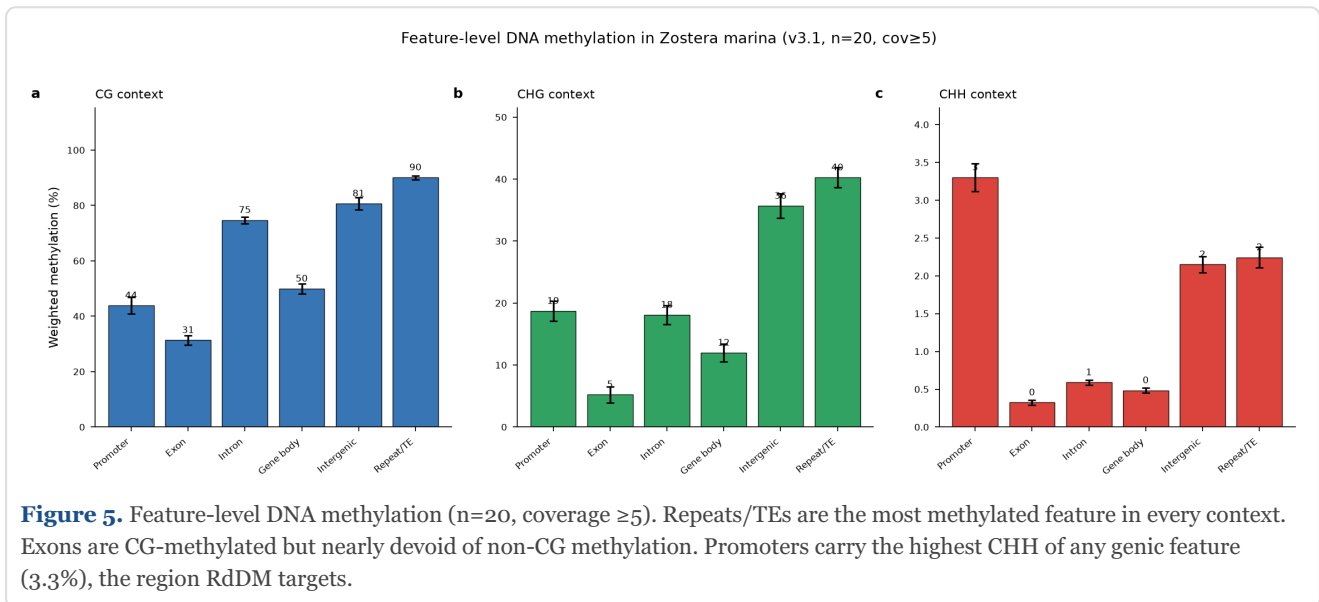
The sequenced sample panel

The 20 *Zostera marina* methylomes analyzed throughout this document come from a structured collection: **five collection groups of four samples each** (sample codes 1 . 1–5 . 4, where the code is *group.replicate*). The group axis is collection site/population; the four replicates capture within-site variation. The full sample-to-group key is provided as a downloadable resource — [sample_metadata.xlsx](#) — so any per-sample result in the sections below (methylation landscape, genetic structure, clustering) can be traced back to its collection group.

5. Feature-level Methylation

Splitting the methylation by genomic feature — promoters, exons, introns, gene bodies, intergenic regions, and repeats — shows which parts of the genome are methylated and, in particular, where RdDM is most active.

Using the v3.1 gene models and a repeat annotation, coverage- ≥ 5 sites in all 20 samples were assigned to promoters (2 kb upstream of the TSS), exons, introns, gene bodies, intergenic regions, and repeats/TEs, and weighted methylation was computed per feature \times context.

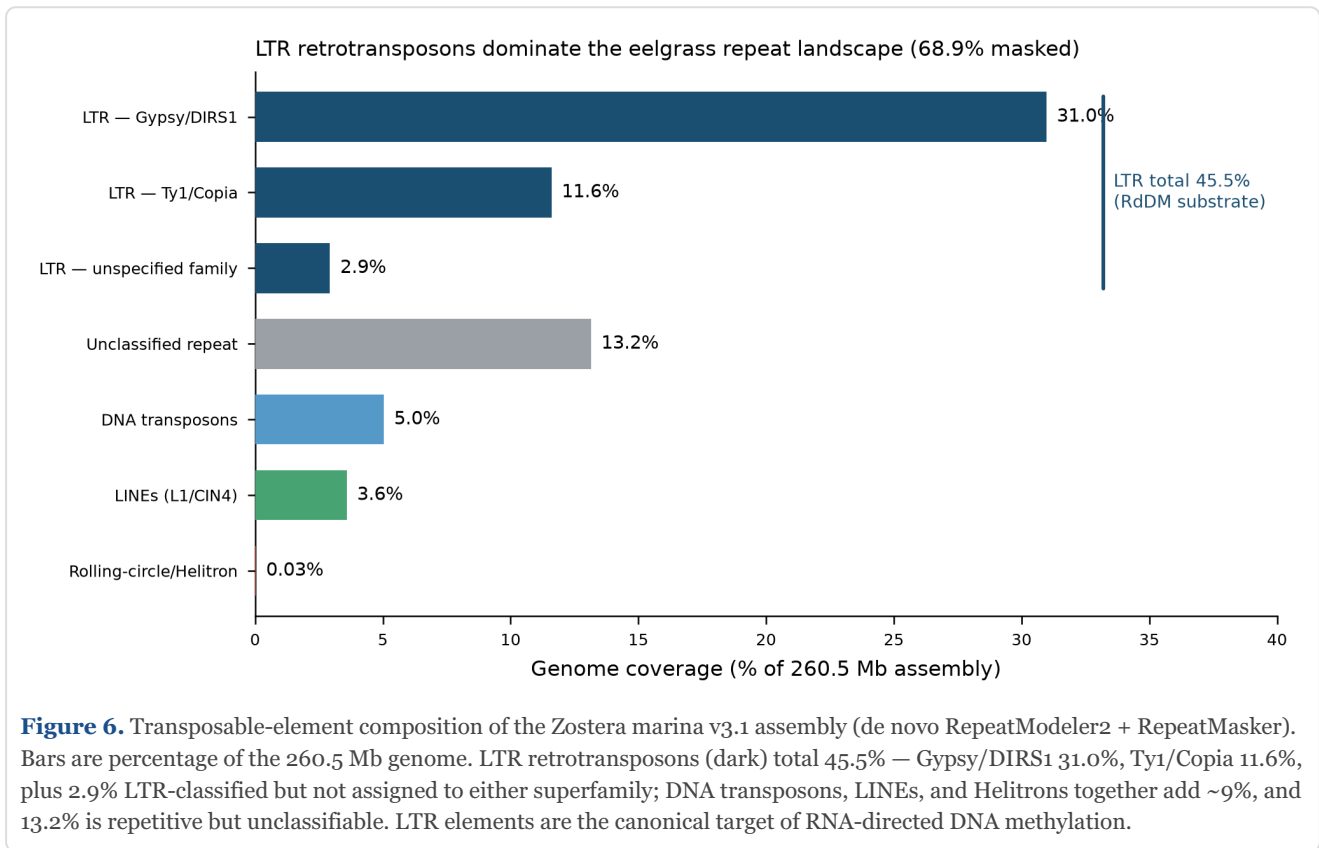


FEATURE	CG	CHG	CHH
Promoter (2 kb up)	43.9%	18.7%	3.3%
Exon	31.3%	5.2%	0.3%
Intron	74.7%	18.1%	0.6%
Gene body	49.9%	12.0%	0.5%
Intergenic	80.6%	35.7%	2.2%
Repeat / TE	90.1%	40.2%	2.2%

Two features of the distribution are relevant here. Repeats/TEs are hypermethylated in all three contexts (CG 90%, CHG 40%, CHH 2.2%), the expected transposon-silencing signature and a check that the methylation calls behave as they should. Promoters carry the highest CHH methylation of any genic feature (3.3%), above intergenic and repeat regions. Because CHH methylation is maintained by ongoing RdDM, promoters are where the pathway concentrates its activity, and are the region an RNA-directed strategy would engage to modulate a target gene without altering its coding sequence.

What the repeats are

A de novo transposable-element annotation of the v3.1 assembly (RepeatModeler2 built a library of 1,463 families; RepeatMasker applied it) resolves the "repeat/TE" category above into its component classes. **68.9% of the eelgrass genome is repetitive** — and the composition is lopsided in a way that matters directly for RdDM.



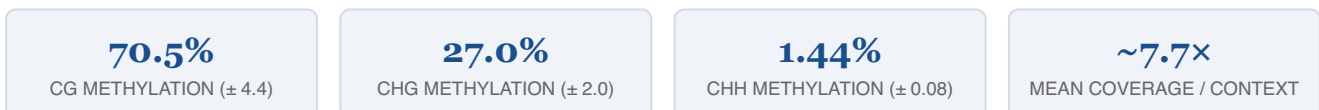
LTR retrotransposons alone occupy 45.5% of the genome — Gypsy/DIRS1 at 31.0%, Ty1/Copia at 11.6%, and a further 2.9% LTR-classified but not assigned to either superfamily — making them by far the largest sequence class in the eelgrass genome. DNA transposons contribute 5.0%, LINEs 3.6%, and Helitrons a negligible 0.03%; a further 13.2% is repetitive but unclassifiable against known families. This is precisely the substrate RdDM evolved to silence: the CHH/CHG hypermethylation of the "repeat/TE" row in the table above is, in the main, methylation of these LTR elements. An assembly this LTR-heavy is a genome under sustained transposon pressure, and a large, intact RdDM inventory (§2) is what keeps that pressure contained. It also delimits where an RNA-directed intervention has natural traction — the pathway is already engaged across nearly half the genome.

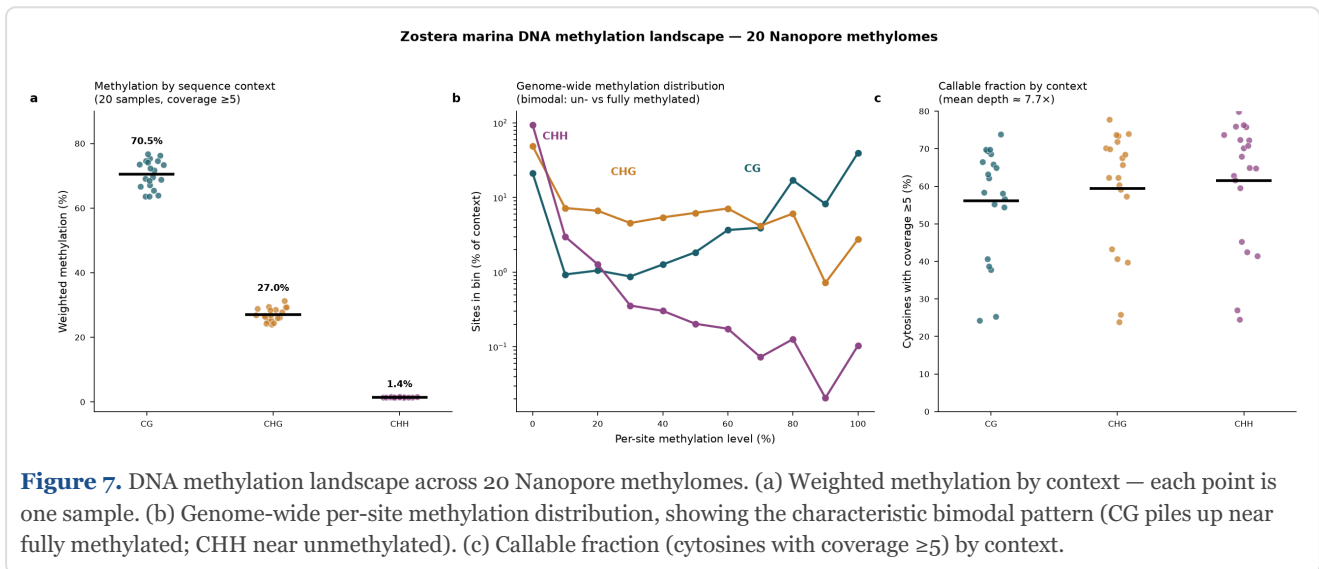
The masked fraction (68.9%) is slightly higher than the 66.5% softmasked repeat annotation used for the feature partition above, because the de novo library recovers eelgrass-specific families a homology-only mask misses; the feature-level methylation percentages are unaffected (they were computed against the softmasked BED, and the two annotations agree on which regions are repeat-dense). A parallel EDTA annotation is running as an independent cross-check and will be noted here if it materially revises these class fractions.

6. Methylation Landscape

Genome-wide DNA methylation across the 20 samples, by sequence context. Weighted methylation = (modified reads) / (modified + canonical reads), pooled over all sites with coverage ≥ 5.

Zostera marina shows the canonical plant methylation hierarchy: high symmetric CG methylation, moderate CHG, and low but non-zero asymmetric CHH. The three contexts are strikingly reproducible across all 20 samples — CHH in particular varies by less than a tenth of a percent between individuals.

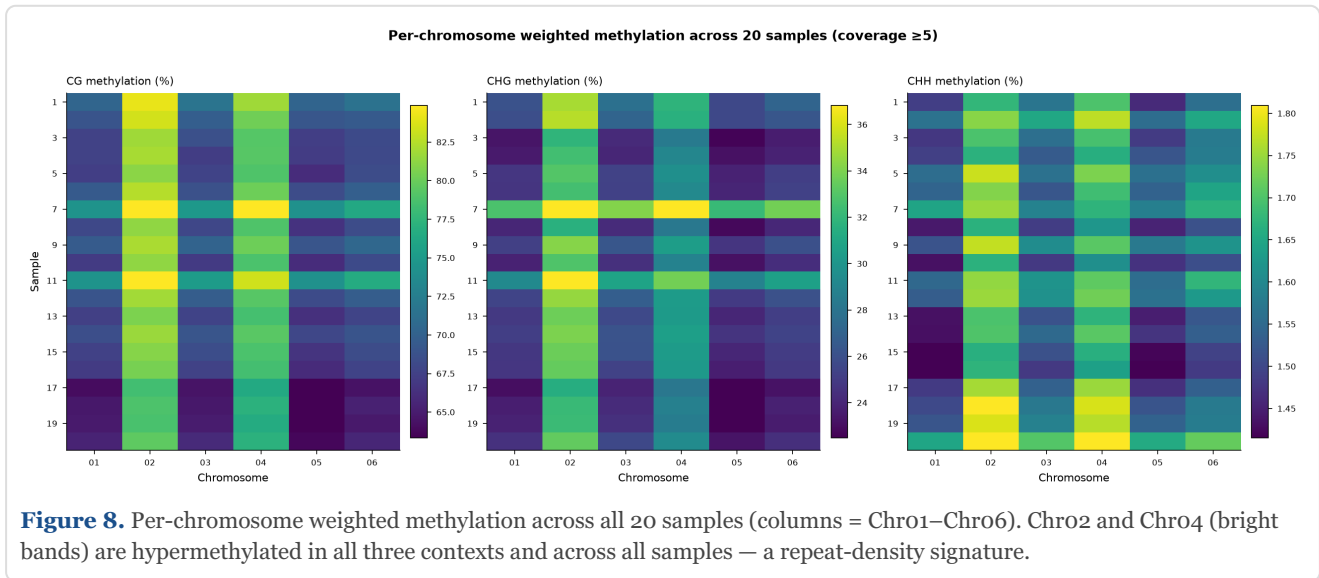




The per-site distribution is bimodal: most CG sites are either near-fully methylated or near-fully unmethylated, while CHH sites are overwhelmingly low-methylated with a small hypermethylated tail — the fraction of the genome under active RdDM/CMT2 control.

Per-chromosome pattern

Methylation is not uniform across the six chromosomes. Chr02 and Chr04 are consistently hypermethylated in every context and in every one of the 20 samples, consistent with higher transposable-element / repeat density on those chromosomes.

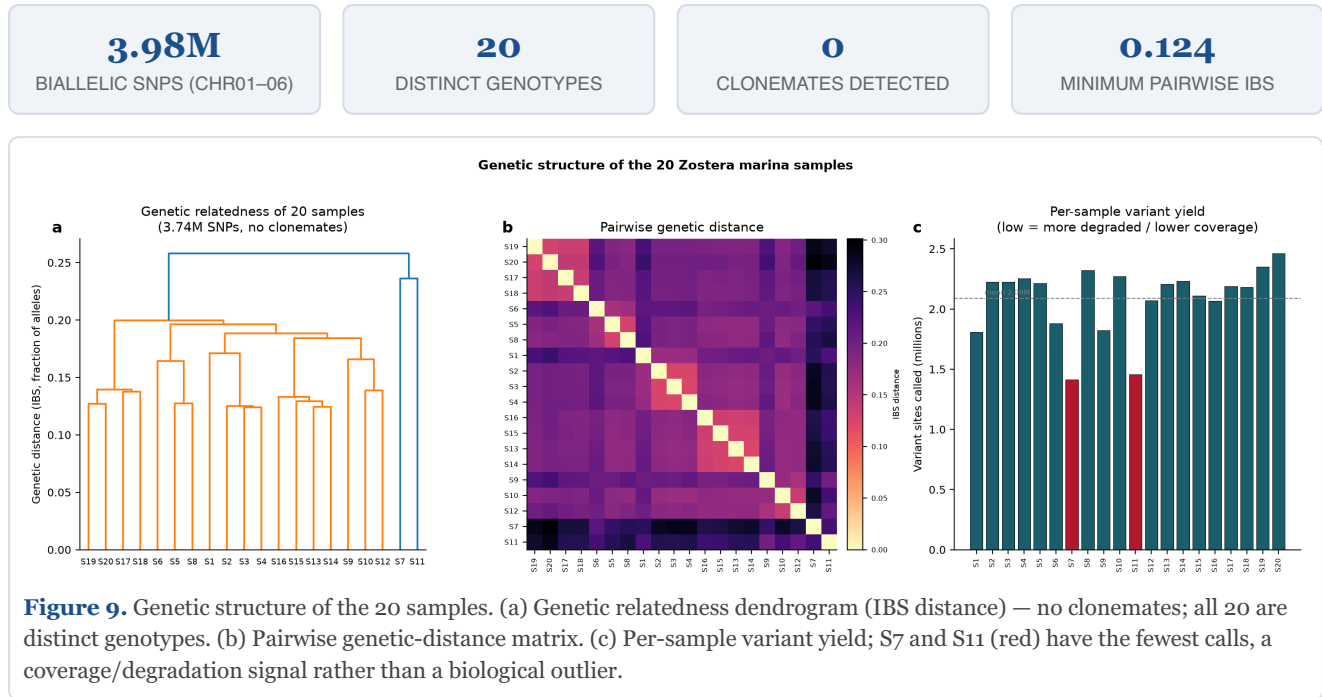


CONTEXT	GENOME-WIDE MEAN	CHR02 (HIGH)	CHR04 (HIGH)	OTHER CHROMOSOMES
CG	70.5%	81.2%	78.8%	~66–68%
CHG	27.0%	33.3%	29.7%	~24–26%
CHH	1.44%	1.74%	1.71%	~1.3–1.4%

7. Genetic Structure & Clonality

Seagrasses reproduce both sexually and clonally, so before interpreting any epigenetic differences we established the genetic relationships among the 20 samples from their SNPs.

Per-sample SNP calls were merged into a union of biallelic sites on the six chromosomes (3,983,427 SNPs; 3,741,213 segregating). Pairwise identity-by-state (IBS) distances were computed across all 20 individuals.

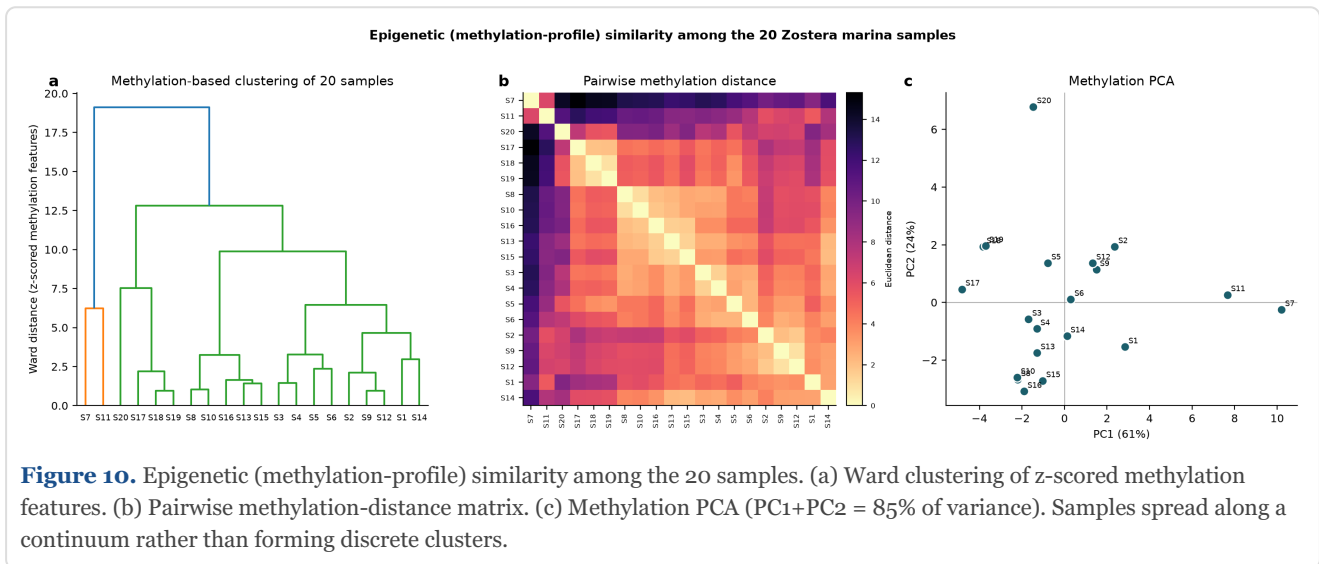


All 20 samples are genetically distinct — the minimum pairwise distance (0.124) is far from zero, so there are no clones in this collection. Samples **S7** and **S11** have the lowest variant yields (1.42M and 1.46M vs. a ~2.1M median), consistent with the note that some samples are more degraded; they appear as apparent outliers in distance space for technical, not biological, reasons.

8. Methylation ↔ Genotype coupling

Some of the methylation differences between samples are inherited with the genome, and some are environmental responses. Separating the two matters for target choice, because only the environmental part is a useful handle — so we tested how closely methylation similarity tracks genetic similarity.

A per-sample methylation feature matrix (per-chromosome × context weighted methylation plus genome-wide context means) was clustered and compared against the genetic distance matrix.



A Mantel test shows methylation distance and genetic distance are significantly correlated ($r = 0.77$, $p = 0.0001$; and still $r = 0.41$ after excluding the low-yield S7/S11). Much of the between-individual methylation variation is therefore genotype-associated. For an RdDM engineering program this is the central practical result: to find *facultative*, environmentally responsive marks (the useful targets), one must first control for the substantial genotype-linked component measured here.

9. Deployment Pathway & Market Context

For the improved plants to have an effect, they have to be planted at scale into a use that is wanted and paid for. Seagrass planting has now been mechanized, and the treated seed this project would produce fits the machines that already exist. The work also sits inside an active, funded U.S. federal restoration agenda. This section covers the planting hardware, the restoration market, the carbon value, and the federal funding.

How the treated seed gets planted

The delivery routes in §3 end with seed reaching the seafloor, and there is now hardware that plants it. **ReefGen** — the restoration-robotics company founded by our delivery partner Chris Oakes — builds seafloor robots that inject seagrass seed and shoots directly into sediment, a mechanized version of the manual seed-injection methods used in the field. Its eelgrass planter (nicknamed *Grasshopper*) weighs about 23 kg and can hold up to 20,000 seeds in a 20-litre bag, planting up to roughly 60 seeds per minute — on the order of ten times a diver's rate, with little disturbance to the surrounding sediment. In its first pilot in 2022 the machine-planted plots grew eelgrass of the same quality and density as hand-planted controls, which established that mechanized planting is feasible at speed. The scale gap this addresses is large: the largest manual seagrass restoration (Chesapeake Bay) plants under 55 acres a year, and the 2024 federal report projects that mechanized planting can increase the area covered roughly a hundredfold. (*Robot specifications from ReefGen and CNN reporting, 2024; scale figures from the 2024 USDA/NOAA report.*)

The robot plants seed into sediment, and the seed can be delivered mixed into a matrix — the same kind of carrier that can hold an RdDM seed pre-treatment (Route A, §3). The environmentally-matched, treated seed this project would produce therefore fits an existing, already-deployed planting system, and reaching established meadows or new sites does not require building new hardware.

Protecting existing meadows, not only replanting lost ones

Replanting is one use of an improved plant; protecting a meadow that is still there is the other, and for climate it is the more valuable of the two. An intact meadow already holds decades to centuries of carbon in its sediment, and if it dies that carbon is released and the die-off stimulates methane production (§ Blue carbon, below). Preventing a loss therefore avoids an emission, on top of preserving the fisheries, water-quality and storm-buffering services the meadow provides — whereas replanting a meadow that has already been lost starts the carbon clock over from zero.

Existing meadows can be given adaptive traits in place. A vulnerable meadow — one sitting at the warm edge of its thermal range, or under recurring heat stress — can be interplanted with RdDM-treated, heat-resilient seed or shoots of the same locally-sourced genotype, raising the resilience of the stand without removing or replacing the plants already growing there. Because RdDM changes methylation state rather than inserting genes and uses local genetic backgrounds (§3), this stays within the same non-transgenic, native-range constraints as new-site planting. In practice the program has two deployment modes on the same delivery hardware: **replant** cleared or degraded sites with treated seed, and **reinforce** at-risk standing meadows in place so they survive the heatwaves that would otherwise convert them from a carbon sink into a carbon and methane source.

The restoration market and its targets

Seagrass restoration is a growing activity, and several programs have set specific acreage goals that create demand for planting stock:

State restoration mandates

Washington State has adopted a kelp and eelgrass restoration target for Puget Sound (reported at **10,000 acres**), and the seafloor planting robots described above have been deployed there in support of it (press reporting; exact statutory figures should be confirmed against the state program before citation).

Federal restoration programs

The National Park Service is funding eelgrass restoration across **five National Seashores** (North Carolina to Massachusetts), with the stated aim of moving heat-tolerant seed north by assisted gene flow — the same approach the collection design in §4 follows.

Beneficial dredge reuse

The U.S. Army Corps of Engineers targets **70% beneficial reuse of dredged sediment by 2030** (up from 30–40% historically — the "70/30 goal"), and seagrass establishment is a qualifying use — creating standing demand for stock suited to placed sediment.

The demand exists because seagrass is being lost. Global seagrass has declined about **29%** since records began, losing roughly **110 km² per year since 1980**, and marine heatwaves are now the leading emerging threat. Each of the common reasons a restoration fails and has to be replanted is one of the trait targets in §3 (see the failure-mode table there).

Blue carbon, quantified

Seagrass meadows store carbon more efficiently per unit area than many temperate forests. Two figures, both calculated with region- and species-specific (IPCC Tier-2) methods, give a sense of the scale:

- The EPA estimates roughly **932,400 metric tons of carbon** stored in the top 30 cm of seagrass meadows from New York to Maine — equivalent to the annual energy use of about **446,653 homes**.

- North Carolina's greenhouse-gas inventory credited its seagrasses with removing about **56.5 kilotons of CO₂-equivalent** in 2020 — comparable to taking **13,455 cars** off the road for a year.

Two caveats go with these numbers. A valuation of restored meadows on Virginia's Eastern Shore found that carbon removal was **less than half** of the total ecosystem-service value; the rest came from fisheries, water quality, storm buffering and biodiversity. The case for seagrass is therefore stronger when it rests on the full set of services rather than carbon alone. The second caveat is that when seagrass dies it releases its stored carbon and stimulates methane production (methane traps roughly 30× more heat than CO₂). A meadow that is restored but then fails can reverse its own carbon gain, so a durable, heat-resilient meadow — the goal of this program — is worth considerably more for climate than one that does not persist.

The work is a funded federal priority

Federal money is already committed to this area. R&D funding into farmed seagrass and seaweed since 2014 exceeds **US\$325 million** across seven agencies (in millions of dollars: Dept. of Commerce ≈121.5, USDA ≈90.4, NSF ≈41, Dept. of the Interior ≈35, DOE ≈23, DHHS ≈14, EPA ≈5). Three points connect that funding to the science in this document:

- The **National Strategic Plan for Aquaculture Research (Action 2.1.2)** names **thermal tolerance** and **disease resistance** as priority breeding goals — the two lead traits in this program.
- The 2024 USDA/NOAA report to Congress recommends developing "**selective breeding, genetically modified organism standards and guidelines for ... restored seagrasses**" — the regulatory lane a *non-transgenic* RdDM approach is designed to occupy.
- The report sets out an escape-risk framework for improved stock (work within native genetic ranges; harvest before reproductive maturity). An RdDM approach meets it without special effort, because it changes methylation state rather than inserting genes and works within locally-sourced genetic backgrounds.

Full extraction of the federal report is in `federal_aquaculture_report_digest.md` (Downloads).

Data & Downloads

Every figure, summary table, and report from these analyses. All files are in the accompanying data bundle; links below are relative to the downloads/ folder.

Figures (PNG)

- `fig1_methylation_landscape.png`
- `fig2_chromosome_methylation.png`
- `fig3_rddm_machinery.png`
- `fig4_methylation_clustering.png`
- `fig5_genetic_structure.png`
- `fig6_feature_methylation.png`
- `fig7_rddm_comparative.png`
- `fig_env_niches.png`
- `fig_berkeley_collection.png`
- `fig_global_map_marine.png`
- `fig_zostera_thermal.png`
- `fig_ecological_map.png`
- `fig_vallisneria_freshwater.png`

Sample metadata (XLSX)

- `sample_metadata.xlsx` — 20-sample panel, 5 collection groups × 4

Methylation tables (CSV)

- methylation_summary_stats.csv
- methylation_per_sample.csv
- methylation_per_chromosome.csv
- methylation_per_chromosome_means.csv
- methylation_context_summary.csv
- feature_methylation_summary.csv
- feature_methylation_per_sample.csv
- methylation_feature_matrix.csv

RdDM & genetics (CSV/TSV)

- rddm_ortholog_inventory.csv
- rddm_hits.tsv
- rddm_thalassia_inventory.csv
- rddm_inventory_6species.csv
- genetic_distance_matrix.csv
- per_sample_varcounts.csv
- seagrass_env_envelopes_6sp.csv
- seagrass_collection_regions_by_species.csv
- berkeley_collection_zones.csv

Reports & references (Markdown)

- consolidated_project_report.md
- methylation_characterization_report.md
- rddm_machinery_report.md
- rddm_comparative_report.md
- genetics_epigenetics_report.md
- methods_reproducibility.md
- seagrass_biogeography_report.md
- federal_aquaculture_report_digest.md
- decibel_bio_research_brief.md
- seagrass_rddm_literature_review.md
- repeat_stopgap_report.md
- zmar_repeatmasker.tbl
- fig_te_composition.png

Methods in brief

Full detail in [methods_reproducibility.md](#) (download above). Summary of pipeline and provenance.

Data & assembly

20 samples, Oxford Nanopore 5mC calls (modkit bedMethyl, Decibel Bio). Reference: *Zostera marina* v3.1 (Ma et al. 2021, Phytozome) — 6 pseudo-chromosomes, 260.5 Mb, 21,483 genes. Sequence context (CG/CHG/CHH) is pre-annotated in the bedMethyl calls.

Methylation

Weighted methylation per context and per feature at coverage ≥ 5 . Feature intervals from v3.1 gene_exons GFF3 (promoters = 2 kb upstream of TSS by strand) and softmasked-repeat BED (66.5% of genome). Aggregation with awk + bedtools.

Orthology

DIAMOND blastp (--very-sensitive, $e < 1e-5$) of 20 reviewed Arabidopsis RdDM proteins vs. each seagrass proteome. Calls: Strong $\geq 45\%$ id & $\geq 70\%$ cov; Moderate $\geq 30\%$ & $\geq 50\%$; Weak $\geq 25\%$ & $\geq 30\%$.

Genetics

Per-sample VCFs merged (bcftools), restricted to biallelic SNPs on Chr01–06. IBS distances and clustering via scikit-allel; methylation–genetics coupling by Mantel test. Note: merge fills absent genotypes as reference, so IBS is a relatedness proxy, not a formal population-genetic estimate.

Contributors & Attribution

Field collection and computational analysis credits for the work reported here.

Species collection

Field collection of the seagrass samples was performed by **Cameron Colby Thomson**, **Morgan Peterson**, **Chris Oakes**, and **Parker Bonnell**.

Computational analysis

Genome, methylome, and comparative analyses reported here were performed by **Cameron Colby Thomson** (Allied Strategy LLC).

Sequencing & methylation platform

Sequencing of the 20 *Zostera marina* genomes and methylomes (Oxford Nanopore, 5mC) and the RdDM methylation-writing platform are provided by **Decibel Bio** – **Brandon Pfannenstiel**, **Travis Bayer**, and **Jack Colicchio**.

Key References

Seagrass epigenetics literature underlying the interpretation above. Full annotated list in [seagrass_rddm_literature_review.md](#) (download above).

- **Ma X, Olsen JL, Reusch TBH, et al. (2021)**. Improved chromosome-level genome assembly of the seagrass *Zostera marina*. *F1000Research* 10:289. — v3.1 reference assembly used throughout.
- **Jueterbock A, Boström C, Coyer JA, et al. (2020)**. The seagrass methylome is associated with variation in photosynthetic performance among clonal shoots. *Frontiers in Plant Science* 11:571646. — methylation–phenotype coupling in *Z. marina*.
- **Entrambasaguas L, Ruocco M, Verhoeven KJF, et al. (2021)**. Gene body DNA methylation in seagrasses: inter- and intraspecific differences and interaction with transcriptome plasticity under heat stress. *Scientific Reports* 11:14343. — gene-body methylation, thermal origin, and heat response in *P. oceanica* and *C. nodosa*.
- **Nilsen AML, et al. (2025)**. The methylome of clonal seagrass shoots shows age-associated variation and differentiation of roots from other tissues. *BBA General Subjects* 1869:130748. — tissue- and age-dependent methylation; sampling design.

- **Olsen JL, Rouzé P, Verhelst B, et al. (2016).** The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature* 530:331–335. — original genome; v2.1 assembly.
 - **DuBois K, Williams SL, Stachowicz JJ (2022).** Local adaptation in a marine foundation species: implications for resilience to future global change. *Global Change Biology* 28:2586–2610. — eelgrass local adaptation over <10 km; provenance sets transplant survival.
 - **Schiebelhut LM, et al. (2023).** Genomic responses to parallel temperature gradients in eelgrass *Zostera marina* in adjacent bays. *Molecular Ecology* 32:2825–2849. — GWAS-resolved thermal-tolerance loci in *Z. marina*.
 - **USDA/NOAA Interagency Working Group (2024).** *Farming Seagrasses and Seaweeds: Responsible Restoration & Revenue Generation*. Report to Congress. — federal policy context; assisted gene flow, restoration methods, blue-carbon economics (see `federal_aquaculture_report_digest.md`, download above).
-

Good Ancestor Foundation

Seagrass Climate Resilience Project — Technical Findings — goodancestor.com

Open science, open technology. All genomes, methods, and data are published for public use. Generated from a canonical Markdown source; regenerate with `build_site.py`.

© 2026 Good Ancestor Foundation and Allied Strategy LLC. All rights reserved.