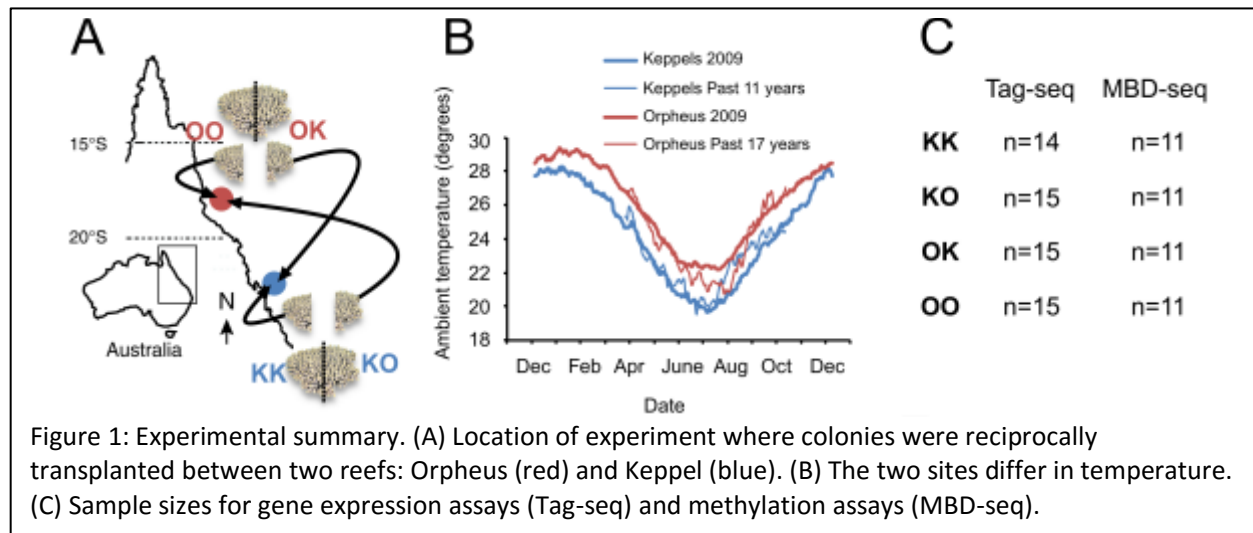


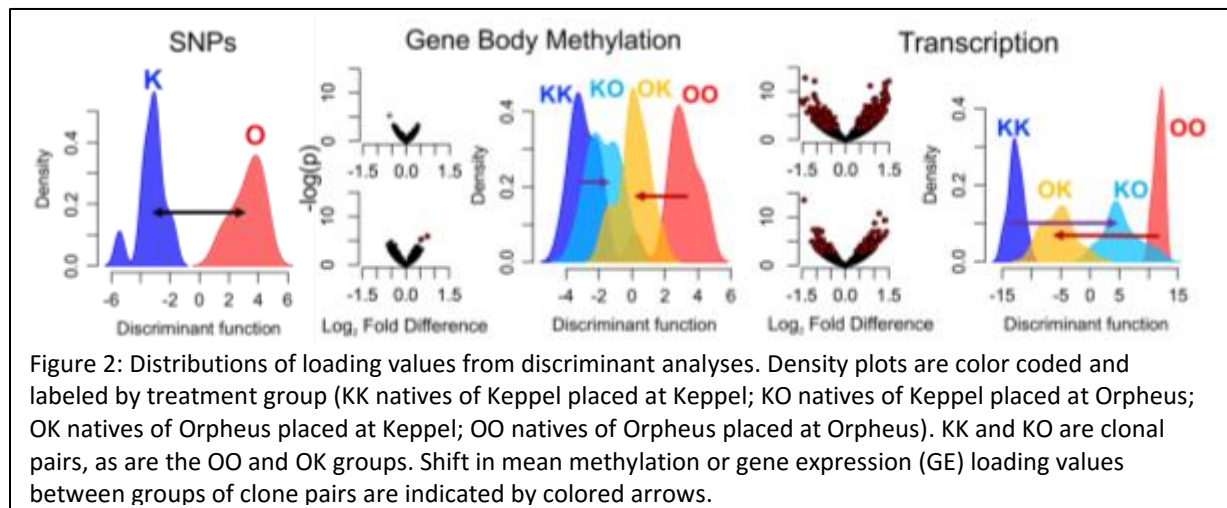
Introduction:

Epigenetic modifications influence gene expression and can change throughout the life of an organism, making them likely mechanisms for mediating phenotypic plasticity. Perhaps the best studied epigenetic modification is DNA methylation, referring to the covalent addition of methyl groups to DNA bases, most often the 5' carbon of cytosine. Here we investigate whether a particular type of DNA methylation, gene body methylation (GBM), is involved phenotypic plasticity of a reef-building coral.

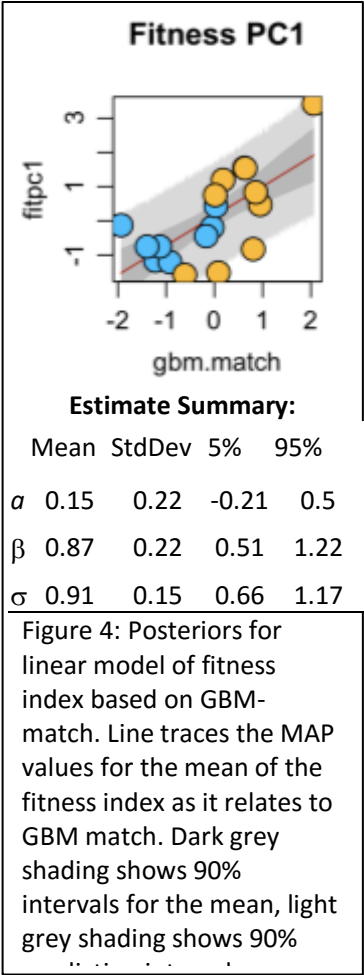
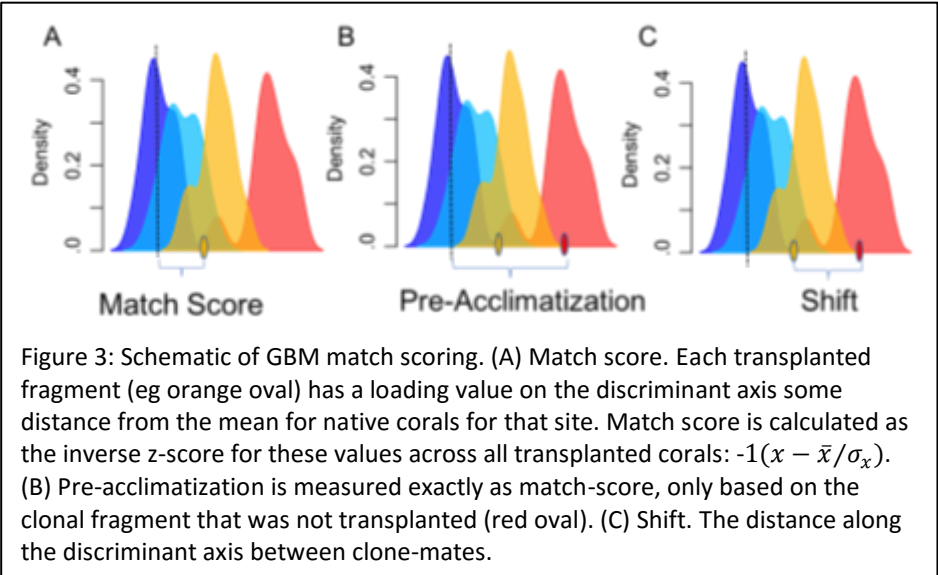
Fifteen colonies of the coral *Acropora millepora* were divided in two and reciprocally transplanted between environmentally distinct sites in the Great Barrier Reef: a warmer site, Orpheus, and a cooler site, Keppel, for three months (fig1 A-B). After three months, tissues were collected and assayed for gene expression (GE) using Tag-seq and GBM using MBD-seq (fig1 C). Reads from the MBD-seq data were also used to call SNPs for each genotype.



To characterize the variation in SNPs, GBM, and GE, we conducted discriminant analysis of principal components (DAPC)^{1,2}. This method is designed to identify the axis in multivariate space that maximizes differences (discriminates) between designated groups. Additional datasets can then be projected onto this axis. We used this method to develop functions discriminating Orpheus and Keppel corals (KK and OO samples) that were replaced at their native sites during the experiment. We then applied this function to their transplanted clone-mates (KO and OK samples) (fig. 2). Similarly, we generated a discriminant function based on SNP data to quantify genetic differences between Orpheus and Keppel colonies.



It seems reasonable that, either due to selection, plasticity, or a combination of each, native corals will possess methylation patterns better optimized to their local environment. If this is the case, then transplanted corals with methylomes more similar to those of native corals should demonstrate greater fitness. We quantify this similarity of methylomes as a ‘match-score’, calculated as the inverse z-score of the distance along the discriminant axis between a transplanted fragment and the mean for native corals of that site (fig. 3A). This match score is a composite of two different values, *pre-acclimation*, measured as the match-score for the clonal fragment that was not transplanted (fig. 3B), and *shift*, measured as the distance between the transplanted fragment its replanted clonal counterpart (fig. 3C). The following analyses test whether GBM matching is associated with physiological measurements indented to approximate fitness, and attempt to dissect the relative importance of pre-acclimation and shift.



Statistical Analyses and Results:

Bivariate linear models (Section 2 R script)

First we constructed bivariate linear models using match-scores based on SNPs, GBM, and GE as predictors and four different physiological measures (weight gain, carbohydrate, protein, and lipid concentration, as well as a summary fitness index (the first principal component explaining 50% of variation of these measures) as the response variables. These models had three parameters (the intercept (α), the coefficient (β), and the standard deviation of the response variable (σ)). The prior probability distributions for these were: α : normal distribution with mean of 0 and standard deviation of 10; β : normal distribution with mean of 0 and standard deviation of 10; σ : a flat distribution from 0 to 10. Maximum a posteriori (MAP) values for these parameters were estimated using quadratic approximation implemented with the MAP function from the rethinking package³. Posteriors for the model of fitness index are shown in fig. 4. Similar results were obtained for each individual fitness measure (data not shown). In contrast to GBM match score, SNP match-score and GE match-score were poor predictors of fitness (data not shown).

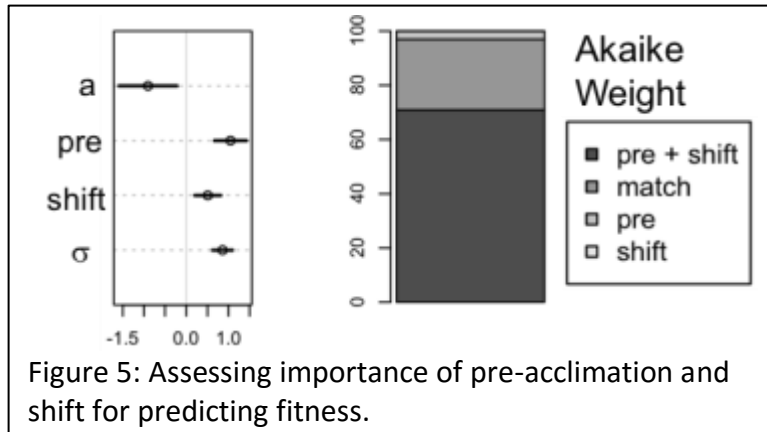
Comparing Origin and GBM match score (Section 3 R script)

Given the distinct conditions of the two sites it seemed possible that site could be a more important predictor than GBM-match. To assess this, we constructed multivariate linear models of fitness index using *Origin* and *GBM-match-score* as predictors, then compared the posterior distributions for the coefficients of the predictors. This analysis demonstrated that when GBM-match-score was known, Origin provided little or no additional predictive power. Similar results were obtained using frequentists methods.

Comparing subcomponents of GBM match score (Section 3 R script)

The overall match score depends on two components: how much GBM patterns already resembled the target mean (*pre-acclimation*; fig.3B), and how much GBM patterns changed during the experiment (*shift*; fig.3C). Understanding the relative importance of these components could shed light on the mechanism linking GBM-match-score and fitness. First we generated multivariate models including both *pre-acclimation* and *shift* and compared the

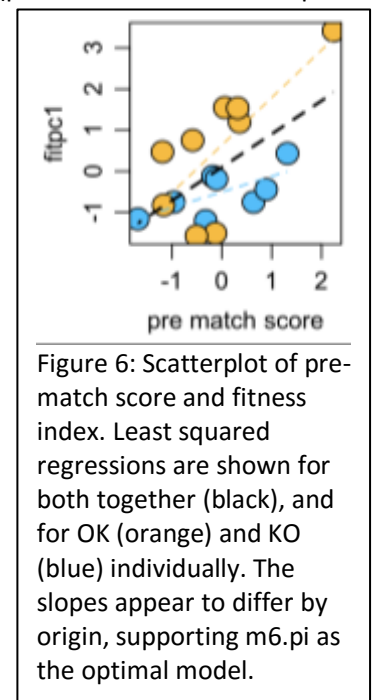
posterior distributions for their coefficients. Both the estimates for *pre-acclimation* and *shift* were most likely greater than zero, indicating that each variable supplied predictive power in addition to the other. We then constructed four linear models of the fitness index using *GBM-match-score*, *pre-acclimation*, and *shift* alone, as well as a model with both *pre-acclimation* and *shift*. We estimated the predictive accuracy of these models using Widely Applicable Information Criterion (WAIC), and found that the model including both *pre-acclimation* and *shift* was likely the optimal model for



distributions for *pre-acclimation* and *shift* were stable across models (data not shown), indicating that these predictors remained important even when genetic and gene expression predictors are added. By comparing the models using WAIC we found that the top three models were either m0 (pre-acclimation + shift only), m6.pi (pre-acclimation + shift + pre-acclimation*origin interaction) and m6.bi (pre-acclimation + shift + pre-acclimation*origin interaction + shift*origin interaction). These models appeared roughly equal in their predictive accuracy, although m6.pi was often weighted higher. The same analysis using AIC on frequentist models found m6.pi to be the optimal model. This indicates that the effect of pre-acclimation varied based on origin (or transplantation site, as these are effectively the same variable within the transplanted corals). The scatterplot of the individual relationships supports this (fig. 6).

Conclusions:

We found that both components of GBM-match score were important for predicting fitness. This suggests that both standing epigenetics variation and the magnitude of plastic response are linked with individual fitness. It is surprising that GBM correlated with fitness, whereas genotype and gene expression were poorly correlated. This certainly does not mean that genotype and gene expression are unimportant for fitness, rather for this particular dataset, GBM happens to have the greatest predictive power. This could be because the MBD-seq data were of higher quality than the Tag-seq and SNP data. However it is also possible that GBM provides a better readout of physiologically relevant genomic mechanisms than either SNPs, or RNA-seq. In either case, these analyses demonstrate a link between both pre-acclimation and plasticity of GBM and fitness under altered environmental conditions.



References:

1. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94 (2010).
2. Kenkel, C. & Matz, M. V. Enhanced gene expression plasticity as a mechanism of adaptation to a variable environment in a reef-building coral. *Nat. Ecol. Evol.* **1**, 14 (2016).
3. McElreath, R. *Statistical Rethinking A Bayesian Course with Examples in R and Stan.* (CRC Press, 2016).