Population genomics, polyploidy, climate resilience and the microbiota of two habitat-building coralline algae

Tom L. Jenkins

t.l.jenkins@exeter.ac.uk

University of Exeter https://orcid.org/0000-0001-5774-0278

Magnus Axelsson

Natural England

Angela Gall

Natural England

Frances Ratcliffe

Natural Resources Wales

Charlie D. Ellis

University of Exeter

Jamie R. Stevens

University of Exeter

Research Article

Keywords:

Posted Date: July 15th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7035383/v1

Additional Declarations: The authors declare no competing interests.

Abstract

Maerl bed habitats support an array of marine biodiversity and are recognised as important blue carbon ecosystems. However, little is currently known about the population structure, spatial distribution of genetic diversity, or microbiota of the coralline algae that build these habitats, or how resilient they are to climate change. Here, we investigate these knowledge gaps in two key maerl-forming species, *Phymatolithon calcareum* and *Lithothamnion corallioides*. In southern England, we identified genetically distinct populations of maerl of the two species in Dorset, while there was evidence of admixture between maerl beds in Cornwall, the extent of which differed among sites and species. A genetically distinct clonal lineage, however, was revealed in the Helford River for *L. corallioides*. Additionally, we distinguished the genetically distinct coarse growth form of *P. calcareum* in Falmouth as a unique, apparently triploid, phenotype of *P. calcareum*. In south-west Wales, only *L. corallioides* was found in Milford Haven, a population that is genetically diverse and isolated from populations in England. Analysis of climate change resilience indicated that *P. calcareum* maerl in Cornwall may have slightly higher risks of being maladapted to temperatures and salinities predicted in 2050 compared to other *P. calcareum* maerl sampled. Lastly, metagenomics analysis revealed differences in microbiota between dense healthy maerl beds compared to scattered dead maerl. As a habitat of conservation priority, our findings will be central to ensuring maerl beds receive appropriate levels of protection to prevent further loss or degradation by human activity and climate change.

INTRODUCTION

Conservation genomics has revolutionised our understanding of the levels, distribution and functional significance of genetic variation in natural populations (Allendorf et al. 2010). One area where genomics has progressed our understanding is in the detection of local adaptation and how species or populations are likely to respond to rapid climate change (Aguirre-Liguori et al. 2021; Bernatchez et al. 2024). It has also provided new insights into genetic diversity and population structure through the increased resolution and precision offered by the analysis of many thousands of genetic markers from across the genome (Davey et al. 2011). This is particularly important for conservation biology as it aids the accurate identification of biologically relevant management units and connectivity scales (Funk et al. 2012).

Progress in conservation genomics has laid the groundwork for the development of Essential Biodiversity Variables (EBVs) for genetic composition (Hoban et al. 2022). The EBVs concept categorises biodiversity into several classes, one of which is genetic composition, and proposes standardised and scalable metrics that are designed to identify and monitor the status and trends of biodiversity at all levels (genetic, species, and ecosystems) over space and time (Pereira et al. 2013; Kissling et al. 2018). Hoban et al. (2022) proposed four genetic EBVs: (i) genetic diversity (e.g. heterozygosity), (ii) genetic differentiation (e.g. F_{ST}), (iii) inbreeding (e.g. F_{IS}), and (iv) effective population size (N_e), which the authors advocate for integration into conservation management strategies. Although genetic diversity has been generally overlooked in global biodiversity policy (Laikre 2010; Laikre et al. 2020), there are encouraging signs given the recent commitments and improved targets set by the Convention on Biological Diversity to monitor and conserve genetic diversity (Hoban et al. 2023). This would be a positive shift since genetic diversity is crucial for providing resilience to rapid environmental change (e.g. climate, land use change, pollution, etc.) and for resisting virulent or emerging diseases. However, despite the information gained from genetic / genomic studies, management

recommendations by researchers and scientists do not always impact regional conservation practices due to the 'conservation genetics gap' (Sandström et al. 2019), whereby the challenges of translating genetic data to non-specialists inhibit their effective integration into policy (Shafer et al. 2015; Jenkins & Stevens 2018). To overcome this gap, it is therefore important to engage and design research projects in conjunction with policy and conservation advisors. Here, we have adopted such a partnership with two UK non-departmental public bodies, Natural England and Natural Resources Wales, both of whom advise on the natural environment and have a central role in shaping policies to protect species and habitats of conservation priority.

Coralline algae are calcareous red seaweeds that are found in almost every coastal ecosystem globally and are highly diverse in appearance and life history, with many species recognised as important ecosystem engineers (Schubert et al. 2020). In terms of genomics research, red macroalgae as a group have been underrepresented compared to other eukaryotic groups (Borg et al. 2023), although recent initiatives have been established to improve genomic resources, such as Rhodoexplorer (Lipinska et al. 2023) and macro-ALG-ALL-CODE (Nelson et al. 2024). For coralline algae, there have also been developments in transcriptomic resources (Page et al. 2019; Xue et al. 2022). Given the overall decreases in sequencing costs and development of more sophisticated bioinformatics tools, the generation and availability of reference genomes and transcriptomes will likely increase over the next decade, enabling more conservation genomics research in red macroalgae.

Free-living, unattached sporophytes of coralline algae on the seafloor are called maerl (or rhodoliths). Maerl-forming species can build complex and extensive three-dimensional 'carpets' of maerl on the seabed, known as maerl beds or rhodolith beds, which provide a vital biogenic reef-like habitat in shallow coastal waters for animals and other seaweeds (Kamenos et al. 2004; Peña et al. 2014; Tuya et al. 2023), and are important blue carbon ecosystems (Mao et al. 2020; James et al. 2024). However, maerl is extremely slow-growing, with a growth rate of 0.5–1.5 mm per thallus tip per year (Blake & Maggs 2003), resulting in very slow recovery potential. Moreover, given the global distribution and ecosystem services provided by maerl beds, they are greatly understudied compared to other coastal habitats (Tuya et al. 2023), despite being threatened by climate change (Qui-Minet et al. 2019) and by disturbance or pollution from human activities (Hall-Spencer 2000; Bernard et al. 2019; Legrand et al. 2024). Therefore, more research is needed to understand the adaptability and resilience of maerl bed habitats and how they may be impacted by climate change.

Until very recently, genetic research on maerl has focused primarily on species delimitation using DNA barcoding (via mitochondrial and chloroplast gene fragments), and on exploring the geographical distribution of maerl-forming species. This has likely been driven by the difficulty in identifying maerl to species level based on external morphology (Pardo et al. 2014). Subsequently, this research has generated new information about the species diversity and the distribution of maerl-forming species, notably in the Atlantic and the Mediterranean (Pardo et al. 2014, 2017; Carro et al. 2014; Melbourne et al. 2017). For example, these data have revealed that *Phymatolithon calcareum* and *Lithothamnion corallioides* are the main two species that build maerl beds in the north-east Atlantic (Pardo et al. 2014). More recently, microsatellite markers were used to investigate genetic diversity, clonality and population structure in *P. calcareum* from sites across the European Atlantic (Pardo et al. 2019), which was followed up with a whole genome sequencing (WGS) study using single nucleotide polymorphisms (SNPs) on a subset of sites from the original study (Jenkins et al. 2021). These studies provided the first insights of clonal diversity and genetic differentiation in *P. calcareum*. However,

sampling of maerl within Britain was particularly limited. Furthermore, an analysis of genetic diversity and population structure in *L. corallioides* has yet to be undertaken.

Interpreting genetic patterns in maerl-forming species is challenging due to their complex life cycle, which is perhaps best understood in *P. calcareum* (Fig. 1 in Pardo et al. 2019). This species is partially clonal, meaning it has the capacity to alternate between sexual and clonal reproduction, the latter of which is defined here as an individual which produces new individuals that are genetically identical to the ancestor at all positions in the genome except for sites which have experienced somatic mutations (Stoeckel et al. 2021). In addition, the life cycle is haplodiplontic, and is characterised by a haploid phase (attached gametophyte) and a diploid phase (attached or unattached sporophyte) (Pardo et al. 2019). One of the advantages of isolating genome-wide SNP markers using WGS is the enhanced ability to detect unattached sporophyte (maerl) clones and to more accurately assess genetic diversity within populations.

In England and Wales, maerl bed habitats and two maerl-forming species, *P. calcareum* and *L. corallioides*, are protected by the NERC Act 2006 and the Conservation of Habitats and Species Regulations 2017. As a result, there are several Marine Protected Areas (MPAs), such as Special Areas of Conservation (SAC), that include maerl beds as a feature in their designation order. In Britain, recent maerl bed surveys by Natural England have provided new information about the locations of live (and dead) maerl beds in southern and south-west England (Envision Marine LTD 2024), for which genomic profiling would circumvent the sampling limitations apparent in previous genetic studies of maerl across this region. These new surveys also revealed variation in the number and cover of dense live maerl beds versus mixed or dead maerl beds, as well as sites where only sparsely scattered live or dead maerl is found. This provides an opportunity to explore how genetic diversity and the microbiota of maerl compare across sites with different levels of live and dead maerl, which, currently, is not well understood. The microbiota of crustose coralline algae are known to play important roles in the recruitment and settlement of marine invertebrates (Wilson et al. 2020), and there is conjecture that maerl microbiota may play a similar role in the settlement of commercially important bivalve species (Scales 2022). Therefore, the study of maerl microbiota could provide valuable insights for maerl beds characterised by varying levels of condition.

The main aims of this study were to assemble draft reference genomes for *P. calcareum* and *L. corallioides* and to perform a population genomics analysis with existing WGS data from a previous study (Jenkins et al. 2021) and new data from novel sites sampled in England and Wales. Our research questions were formulated to investigate genetic EBVs and climate adaptation in these maerl-forming species and to explore the microbiota associated with live or degraded/dead maerl beds: (i) What are the levels of genetic diversity and population structure across sites and species? (ii) Are any loci putatively associated with climate adaptation and if so what is the risk of maladaptation to future environmental conditions? (iii) Does the microbiota community of maerl differ across sites and/or in relation to maerl bed condition?

MATERIALS AND METHODS

Sampling locations and DNA extraction

Samples of maerl were collected via SCUBA diving at 4-23m depth from eleven locations along the south-west coast of England, and from two locations in south-west Wales and north Wales; of these, only two had been

previously assessed for maerl species identification or genetic analysis (Table 1, Fig. 1). Collection of samples in both non-protected and protected marine sites were coordinated in partnership with Natural England and Natural Resources Wales to ensure all sampling complied with priority species and habitat regulations. Where possible, transects were deployed and samples were collected at least one metre apart to avoid potential over-sampling of locally dominant clonal lineages, with the aim of maximising the species and genetic variation captured from each maerl bed. The number of samples collected per site varied depending on maerl bed density. All maerl samples were fully immersed in 95–100% ethanol as soon as possible and ethanol was replaced after 24 hours. Samples were stored at 4^{o}C until DNA extraction.

Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen), with some modifications to the protocol designed to increase DNA yield from maerl samples (see Figure S1 for detailed protocol). This was required, particularly for Nanopore sequencing, because the outer layer of live algae is very thin and typically yields extremely low amounts of DNA using the standard protocol. The quantity and concentration of DNA was assessed by fluorometry using a Qubit 1X dsDNA High Sensitivity Assay Kit (Invitrogen). The quality of each aliquot was checked using gel electrophoresis on a 1% agarose gel and only DNA samples showing a ~ 20Kb band or higher with low or no evidence of degradation were selected for whole genome sequencing.

Whole genome sequencing

Illumina short-read libraries were prepared for all samples using NEBNext Ultra II FS following the protocol for inputs below 50ng (half reactions were performed). Adaptors were diluted 1:10 for all samples as per the protocol. No size selection was implemented and 0.8X bead purification was conducted twice prior to seven cycles of PCR enrichment. Libraries were checked by TapeStation D1000 and pooled equimolar. All samples were sequenced on a NovaSeq 6000 using a paired-end strategy with a read length of 150 bases. Additionally, whole genome sequencing data for maerl samples from a previous study (Jenkins et al. 2021) were included in our study (Table 1, Fig. 1); these DNA samples were prepared and sequenced using the same methods described above. Illumina raw reads were filtered and trimmed using fastp v0.23.4 (Chen 2023) with the following parameters: (i) polyG tail trimming enabled, (ii) a Phred quality score > = 30 in at least 60% of bases in a read, and (iii) a minimum length of 100bp.

Nanopore long-read libraries were prepared for two samples, one representing *P. calcareum* and one representing *L. corallioides*, both of which were confirmed to be the correct species by DNA barcoding (see below) and originated from St Mawes, Falmouth (Table 1). For both samples, the same DNA aliquot used to prepare the Illumina libraries was used to prepare Nanopore libraries so that the Illumina data could be used to polish the genome assembled with long reads. Standard Oxford Nanopore Technology libraries were prepared and sequenced on a PromethION 24 flow cell with R9.4.1 chemistry. Guppy v6.3.9 with super-accurate basecalling was used for basecalling and only reads flagged as passed (minimum quality score of 10) were retained in FASTQ format. Porechop v0.2.4 was used to detect and remove adapters from reads passing the basecalling filters.

Species identification via DNA barcoding

A DNA barcoding approach was used to confirm species identity of each sample; this is necessary because of the difficulty in distinguishing maerl-forming species based on external morphology (Pardo et al. 2014). First,

published mitochondrial and chloroplast genomes of *P. calcareum* (MW357900.2, OQ417768.1) and *L. corallioides* (MW357901.2, OQ417769.1) were downloaded from GenBank in FASTA format to function as seed sequences for each organelle. The GenBank reference files for each of these accessions were also downloaded for use in the organelle gene annotation step. Second, the Illumina trimmed reads for each sample were aligned to the mitogenome and plastome seed sequences separately using Bowtie2 v2.5.3 (Langmead & Salzberg 2012). Mapped reads were extracted from each bam file to a FASTQ file using SAMtools v1.19 (Danecek et al. 2021). These target reads were then assembled using Unicycler v0.5.0 (Wick et al. 2017), with the aim of obtaining a complete mitochondrial and chloroplast genome for each individual sample. Third, for organelle genomes successfully assembled, both mitochondrial and chloroplast genomes were annotated using a custom python script (see data archiving statement) using the GenBank reference files. Lastly, for each sample, nucleotide coding sequences (CDS) of the *COI* mitochondrial gene and the *psbA* chloroplast gene were extracted to generate two FASTA files containing CDS for *COI* and *psbA*, respectively. These genes were chosen because they are the most common genes used in coralline algae DNA barcoding studies (Pardo et al. 2014, 2017; Peña et al. 2020), and have the highest intra- and inter-species representation on GenBank. These two FASTA files were submitted to the NCBI *blastn* online server to compare our CDS (query) to the nucleotide database (subject). For a given maerl sample, a species identification was assigned if the percentage identity was > = 99% (Carro et al. 2014).

Genome assembly, QC and contig filtering

A reference genome assembly was built for both *P. calcareum* and *L. corallioides* using Nanopore long reads. Prior to assembly, long reads were taxonomically classified using BugSeq (Fan et al. 2021), which classifies reads against the BugSeq default microbial database. Because of the potential for non-target DNA to be isolated from the sample during extraction, i.e. microbes living on or inside the calcium carbonate skeletal structure of maerl, the rationale here was to identify and remove as many reads of microbial origin as possible prior to the assembly pipeline. From the classification results, only reads classified as 0 (unclassified) or 1 (other) were retained, thus removing all reads classified as Bacteria, Virus or other microbes in the database. Flye v2.9.3 (Kolmogorov et al. 2019) was used to assemble the reads into contigs. The input read length was adjusted to explore the optimum minimum read length for input into Flye; for both samples > = 3000bp was deemed the optimum. For Flye, default parameters were used, alternative contigs were not kept (--no-alt-contigs), and only contigs longer than 3000bp were retained. The assemblies were polished with Illumina short read data using Polypolish v0.6.0 (Wick & Holt 2022).

BlobToolKit v4.2.1 (Challis et al. 2020) was used to assess contamination and quality of the assemblies. Both *blastn* v2.14.1+ (Camacho et al. 2009) and *diamond blastx* v2.1.8 (Buchfink et al. 2021) were run using the assemblies as queries to the NCBI nucleotide and protein databases, respectively. In addition, minimap2 v2.26 (Li 2018) was used to map Nanopore reads to each assembly. The taxonomic hits and sorted bam alignments, along with the assembly FASTA file, were used as inputs to the *blobtools create* command to create a BlobDir. Contigs were removed from the assembly if the following conditions were met: (i) contig was of organelle genome origin, (ii) contig was not supported by mapped reads (at least one read supporting each base across the contig), and (iii) contig matched to Bacteria, Archaea, Virus, or any Eukaryote other than Rhodophyta or Streptophyta. Assembly statistics, including total length, number of contigs, N50 and GC content, were computed using the *stats* command from *seqkit* v2.5.1 (Shen et al. 2016).

BUSCO phylogeny

BUSCO v5.6.1 (Manni et al. 2021) was executed using the eukaryota_odb10 dataset to assess the completeness of near-universal single-copy orthologs in both genome assemblies. Subsequently, a phylogeny of coralline algae (and several other red algal taxa) was built using BUSCO protein sequences, the rationale of which was two-fold: (i) to examine the tree position of *P. calcareum* and *L. corallioides* to further confirm validity of the genome assemblies, and (ii) to compute and visualise a phylogeny of coralline algae based on BUSCO genes. Genome and/or transcriptome assemblies were downloaded for all available coralline algae and several red algal species from the Florideophyceae (Table S1). A BUSCO analysis was run for each assembly and the outputs were supplied as inputs to a modified version of the BUSCO phylogenomics pipeline (see data archiving statement). This pipeline identifies complete single-copy BUSCOs present in at least ten assemblies, then aligns sequences, trims alignments, and constructs gene trees using IQ-TREE (Nguyen et al. 2015). Finally, Astral v5.7.8 (Mirarab & Warnow 2015) was used to build a coalescent-based species tree, with *Porphyra umbilicalis* (Bangiophyceae) used as the outgroup for visualisation.

Variant calling and SNP filtering

After each DNA sample was identified to species level, the trimmed Illumina reads were mapped to the *P. calcareum* or *L. corallioides* genome assemblies using Bowtie2 v2.5.3 (Langmead & Salzberg 2012). The alignments were sorted by coordinate, then read groups were added and duplicates were marked using GATK v4.5 (McKenna et al. 2010). Variant calling was conducted with Freebayes v1.3.8 (Garrison & Marth 2012) in parallel on 100,000bp chunks using *freebayes-parallel* and the *fasta_generate_regions.py* script. The following parameters were used: minimum mapping quality > = 20, minimum base quality > = 30, compute genotype qualities, and the ploidy was set to two. BCFtools v1.21 (Danecek et al. 2021) and VCFtools v0.1.16 (Danecek et al. 2011) were used to filter variants. First, poor quality variants were removed by enforcing a maximum missing threshold of 0.70, a minor allele count of 3 and a minimum quality score of 30. Second, all samples from Bembridge were removed due to a very high proportion (50–95%) of missing genotypes, likely because of insufficient DNA extracted from the target species (Bembridge samples comprised mostly dead maerl). Third, the following parameters were enforced using BCFtools: (i) no missing data, (ii) genotype quality > = 20, (iii) minimum depth > = 3, (iv) allele depth > = 10, minor allele frequency > = 5%, and allele balance > = 0.10. Lastly, the VCF was read into R v4.3.2 and the following filters were applied to create a high quality biallelic SNP dataset: (i) distance thinning such that SNPs within 1000bp on a given contig were removed (to avoid potential linkage), (ii) SNPs with a minimum depth of 15 and a maximum depth of 100 over all samples were removed, and (iii) monomorphic loci were removed.

Ploidy, population structure, clonality and genetic diversity

Ploidy of each sample was assessed using the ‘determining ploidy 1’ method outlined in the vcfR R package documentation (Knaus & Grünwald 2017). This method examines allele balance information, i.e. the number of times an allele was sequenced, at each heterozygous SNP. In diploids, an allele frequency ratio of 50:50 is expected (both alleles are sequenced equally), although there is inevitably variation around this ratio introduced by sequencing bias and variant calling software. When the allele frequencies of all heterozygous SNPs are visualised, we would expect a peak at 1/2. In contrast, in triploids, we would expect two peaks at 1/3 and 2/3. Preliminary analysis of allele balance indicated that read depth and minimum allele depth were

important for interpreting the histograms, so, prior to analysis, SNPs were removed where the depth of at least one allele was < 10, and only samples with a median read depth > = 30 were considered for ploidy determination.

Genetic structure was assessed using non-model based principal component analysis (PCA) and model-based admixture inference using PopCluster v1.4.0 (Wang 2022). For admixture analyses, the admixture model with unequal allele frequencies and medium scaling was run independently ten times for each ancestral population (*K*). No sample grouping or location of origin information (priors) were supplied to the algorithm. The optimal run and *K* statistics (determined by D_{LK2} and F_{STIS} statistics) (Wang 2022) were extracted from the results. Individual admixture proportions were visualised as barplots, and the mean average proportion per cluster per location were visualised as admixture maps using mapmixture v1.1.4 (Jenkins 2024). Locations that had fewer than three individuals were not visualised in the admixture maps. Genetic differentiation among sites was assessed by computing pairwise F_{ST} values (Weir & Cockerham 1984).

Clonality and genetic diversity were investigated for each genetic unit (i.e. each putative population identified based on the genetic structure analyses). Clonal lineages and the number of clones were assessed using the R package poppr v2.9.6 (Kamvar et al. 2014). First, pairwise Prevosti genetic distances were calculated between all individuals. Second, *cutoff_predictor()* was run to define a clone threshold based on the genetic distance matrix. Defining a clone threshold is necessary because somatic mutations can introduce small genetic differences between clones. Lastly, *mlg.filter()* was run using the distance matrix, clone cut-off threshold and the nearest neighbour algorithm to identify multi-locus genotypes (MLGs); each MLG represents a putative clonal lineage. Finally, Pareto β, observed heterozygosity, expected heterozygosity, and F_{IS} (inbreeding coefficient) were computed for each genetic unit. Pareto β characterises clonal diversity and skewness in the distribution of clones within a population; the presence of a few large clonal lineages and many small ones will result in lower β values, while more balanced lineages of similar sizes will result in higher β values (Arnaud-Haond et al. 2007).

Genotype-environment association and genomic offset analysis

Environmental data were downloaded from Bio-Oracle v3.0 (Assis et al. 2024) at a resolution of 0.05 degrees (~ 5.5 km at the equator). NetCDF files were obtained for environmental variables which have been shown or inferred to be important for maerl growth and survival (Martin & Hall-Spencer 2017): ocean temperature (^{o}C), salinity (PSU), pH and oxygen concentration (mmol m^{-3}). For each environmental variable, the following parameters were selected: baseline (2000–2010), benthic (mean average depth per cell) and variable mean (mean average across time period). Future layers (2050–2060) for these variables based on the Shared Socioeconomic Pathway (SSP) SSP245 (‘middle of the road’) scenario from CMIP6 were also obtained. For each individual sample, latitude and longitude coordinates were used to extract values for each variable to use as a set of predictor variables in the analysis. There were only very small differences in pH among locations (8.06–8.08), while multicollinearity checks among the variables revealed a very high correlation between oxygen concentration and ocean temperature (r^2=-0.96). Therefore, only temperature (11.3-15.0^{o}C) and salinity (33.99–35.06 PSU) were retained for genotype-environment association (GEA) analysis. To mitigate spatial autocorrelation, detrending of environmental variables was conducted prior to modelling. First, least-cost geographic distances along coasts were calculated between locations and the output was used to compute

dbMEMs (distance-based Moran's eigenvector maps). Second, a linear model for each environmental variable was modelled as a function of the dbMEMs and the residuals were extracted. The residuals, representing the detrended environmental variables, were used as predictors for the GEA models.

Two GEA methods were used to investigate associations between the environmental data and the SNP genotypes. First, a latent factor mixed model (LFMM) was run using the *lfmm2()* function from LEA v3.12.2 (Gain & François 2021) with $K = 9$ and the default lambda value (1e-05). Using the *lfmm.test()* function, *p*-values were adjusted for latent factors and recalibrated using the genomic inflation factor. Multiple comparisons were controlled for using the Benjamini-Hochberg method and only SNPs with an adjusted *p*-value < 0.05 were considered as candidate loci. Second, a redundancy analysis (RDA) was run using the *rda()* function from vegan v2.6.4. To control for population structure, the first two principal components of the PCA were used as conditions in the model. The significance of each constrained axis was checked by running the *anova.cca()* function and SNP loadings were extracted only for significant RDA axes ($p < 0.01$). SNPs were considered candidate loci if they were present in the tails of the SNP loading histograms using a standard deviation threshold of 3.0. Lastly, only candidate loci found in both methods were considered as GEA loci potentially under selection (climate-associated SNPs).

Putative climate-associated SNP genotypes were used as input for the *genetic.offset()* function from LEA. A matrix of climate data representing both present-day baseline environments (2000–2010) and future environments (2050–5060) were used as covariates. This function calculates a geometric genomic offset value for each sample location using a genetic gap algorithm (Gain et al. 2023). In a population genetic context, the geometric genomic offset can be interpreted as the average value of Nei's D_{ST} (divided by two) for the set of loci assumed to be involved in local adaptation (Gain et al. 2023). Offset estimates predict the relative allele frequency change needed for a population to lower their risk of being potentially maladapted to future environments.

Microbiota assignment using metagenomics

For each maerl sample, Illumina reads that did not align to the reference genome assemblies were extracted to FASTQ files using the *fastq* command from SAMtools. Kraken2 v2.1.3 (Wood et al. 2019) was used to taxonomically classify unmapped reads against a custom marine prokaryotic database, which was built by extracting all sequences from the NCBI nucleotide database that matched taxonomic IDs in the MarDB v1.6 marine metagenomics database (Klemetsen et al. 2018). The following parameters were used in Kraken2: (i) a minimum hit groups of 3, (ii) a minimum base quality of 20, and (iii) a confidence threshold of 0.10. The report output files for all samples were amalgamated to a single tabular file for visualisation using a custom python script (see data archiving statement). After initial inspection of the results, we focused on certain microbial groups that were common across samples or have been reported in previous studies of coralline algae microbiota (Brodie et al. 2016; Valdespino-Castillo et al. 2021), or groups that appeared to differ across sites. To statistically assess differences among these groups among sites, proportions of each group were modelled as a function of sites using an Analysis of Variance (ANOVA), and the Tukey honestly significant difference (HSD) test was used to investigate differences between sites using a 95% confidence interval. To minimise the risk of taxonomic misassignments, we only considered assignments to the level of class.

This analysis was conducted with six sites, all of which were selected to represent different quality/density categories of maerl bed habitats based on a recent Natural England report (Axelsson 2023). This report details a standardised classification system for maerl bed habitats in England. The first three categories cover the range of sites sampled in our study: “Dense Maerl” (Category A), “Maerl Sediment” (Category B), and “Sparse/Scattered Maerl” (Category C), which also have sub-categories (from 1 up to 3) describing physical size, structure, % cover, live/dead proportion and substratum. Thus, we grouped the following sites using this classification system: St Mawes (dense maerl – mostly A1), Helford River (dense maerl – mostly A3), Milford Haven (maerl sediment – mostly B1), The Manacles (maerl sediment – mostly B3), Swanage (maerl sediment – mostly B3) and Bembridge (scattered maerl – mostly C2). This broadly represents a gradient of maerl bed status from a dense, healthy maerl bed (St Mawes) to mostly scattered dead maerl (Bembridge).

RESULTS

Species identification

Of the 207 samples with whole genome sequencing data, DNA barcoding analysis using *COX1* and *psbA* identified 136 as *Phymatolithon calcareum* and 66 as *Lithothamnion corallioides* (Table 1). Five samples from Bembridge could not be identified to species. Elsewhere in Britain, there were some sites where all samples were identified to be a single species, while at other sites samples of both species were observed (Fig. 1A). Sites that only contained samples identified as *P. calcareum* were: The Manacles, Bizzies Reef, Gerrans Bay, Nare Head (all southern Cornwall), and Bembridge (Isle of Wight). In contrast, the following sites only contained samples identified as *L. corallioides*: Helford River (southern Cornwall), Milford Haven (south-west Wales), and St Tudwal’s Island (north-west Wales). Sites with both species were St Mawes, St Austell Bay, Little Gribbin (all southern Cornwall), and Weymouth and Swanage (both Dorset). The remaining samples from Northern Ireland, France and Spain were all identified as *P. calcareum*, as expected from previous analysis (Jenkins et al. 2021).

Draft reference genome and BUSCO phylogeny

A draft reference genome was assembled for each species using Nanopore long reads. The assembly statistics were comparable to or improved upon the two coralline red algal species, *Amphiroa fragilissima* and *Porolithon onkodes*, for which draft reference genomes are currently available (Table S2). For *P. calcareum*, the reference genome was 150 Mbp in length, 6,375 contigs, with an N50 of 35 Kbp and an overall GC content of 43%. For *L. corallioides*, the reference genome was 145 Mbp in length, 9,939 contigs, with an N50 of 19 Kbp and an overall GC content of 40%. Using the Eukaryota orthoDB v10 database ($N = 255$), the BUSCO results for *P. calcareum* were C:62.7% [S:59.6%, D:3.1%], F:12.5%, M:24.8%; and for *L. corallioides* were C:40.8% [S:39.2%, D:1.6%], F:11.8%, M:47.4%. The species tree of coralline algae and other Florideophyceae red algal species was built using 157 concatenated BUSCO genes that were complete and single-copy in at least ten species (Fig. 1C). This tree showed that all coralline algae were positioned on a branch separate from all other Florideophyceae species. Within this branch, the evolutionary relationships amongst species were as expected based on previous research on coralline algae phylogenetics (Peña et al. 2020). For the maerl-forming species in our study, the tree (Fig. 1C) showed that both *Lithothamnion* species, *L. corallioides* and *L. proliferum*, were monophyletic, sharing a recent common ancestor, and that both of these species shared their most recent common ancestor with *P. calcareum*.

SNP dataset summary

For *P. calcareum*, filtering the raw variant calls from Freebayes resulted in 124,751 SNPs. Five individuals were removed at this stage because of very high missing data (> 95%); these individuals were all from Bembridge, which is probably because these samples had very little live algal material. Further filtering of these variants by distance thinning (linkage disequilibrium), mean read depth and keeping only biallelic SNPs resulted in a final dataset of 131 individuals genotyped at 15,330 SNPs for *P. calcareum*. Note this is less than the 136 individuals identified as *P. calcareum* because of the removal of five individuals from Bembridge (see above). For *L. corallioides*, filtering the raw variant calls from Freebayes resulted in 192,041 SNPs. Further filtering of these variants resulted in a final dataset of 66 individuals genotyped at 10,215 SNPs for *L. corallioides*.

Triploidy in P. calcareum coarse growth form

At the St Mawes maerl bed, we found two growth forms both identified as *P. calcareum* by DNA barcoding, one of which was typically around 2cm in diameter with characteristic delicate thalli, while the other was typified by a much larger rhodolith (4-5cm in diameter) with far bulkier branching structures (Fig. 1, Figure S2). Hereafter, we refer to the latter growth form in St Mawes as the 'coarse' form (MawC). Allele balance analysis of all coarse maerl samples showed a clear pattern of triploidy (Fig. 2), with two peaks at 1/3 and 2/3 as per expectations of triploids. These samples were compared with all other *P. calcareum* samples which demonstrated that all other samples were diploid, as evidenced by a single peak for allele balance at 1/2 (Fig. 2). All *L. corallioides* samples with sufficient median read depth showed a peak at 1/2 and were deemed to be diploids.

Genetic structure

For *P. calcareum*, PopCluster was first run using sites only from Britain and Northern Ireland, because we were primarily interested in genetic structure of newly sampled maerl across this region. PopCluster statistics (D_{LK2} and F_{STIS}) showed highest support for five *K* ancestral populations (genetic clusters) (Figure S3, S4), so $K = 5$ was used for visualising admixture results in the structure plot and admixture map (Fig. 3). The results indicate that most maerl samples (individuals) from the following sites were primarily derived from four separate ancestral source populations: Zara Shoal (blue cluster), Weymouth (green cluster), Nare Head (grey cluster), and the coarse triploid maerl from St Mawes (orange cluster). All coarse maerl (regardless of whether they were collected in 2011 or 2022) were virtually genetically distinct from all other sites, while two individuals in Weymouth and a single individual in Nare Head had shared ancestry with other sites. The remaining sites across Cornwall: The Manacles, St Mawes (non-coarse maerl), the Bizzies, Gerrans Bay and St Austell, all showed evidence of admixture with one another (mostly blue and pink clusters, Fig. 3). These patterns of genetic structure were supported by PCA when the first three principal components were visualised (Figure S5). To check whether any newly identified site-specific genetic clusters found in Britain (e.g. Weymouth and Nare Head) were also found in other European sites, PopCluster was run a second time including sites from France and Spain; this showed that all four sites from France and Spain were genetically differentiated from Weymouth and Cornwall (Figure S6), but with some site-specific admixture, which was expected based on previous research (Jenkins et al. 2021).

For *L. corallioides*, PopCluster analyses showed a similar range of structuring at different values of *K*, with *K* = 4 and *K* = 5 showing the highest statistical support overall (Figure S7, S8). Because there was support for five genetic clusters in the PCA (Figure S9), *K* = 5 was used for visualising admixture results in the structure plot and admixture map (Fig. 4). The results indicate that most individuals from Weymouth and Swanage are primarily derived from a single ancestral population (green cluster) that is mostly found in this area (although there is evidence of some admixture with other sites), and in St Tudwal's Island. At *K* = 8, there was evidence that St Tudwal's Island might also be genetically differentiated (Figure S8), although only two samples were available for this site. Individuals from St Austell Bay were mostly derived from a single ancestral source population (pink cluster), which appeared to be admixed with St Mawes and with some individuals from Helford River. The remaining seven individuals from Helford River were mostly derived from a different ancestral population (blue cluster) that is primarily found in this area. In Milford Haven, individuals from both sites were mostly derived from a single ancestral population (red cluster), but with some admixture with other sites, and an additional distinct ancestral population was observed (yellow cluster) which was dominant in four individuals.

Pairwise genetic differentiation measures of F_{ST} ranged from 0.02 (Man-Biz) to 0.42 (Zar-MawC) in *P. calcareum* and from 0.001 (Mil1-Mil2) to 0.11 (Hel-Swa) in *L. corallioides* (Figure S10). The most differentiated site in *P. calcareum* was the coarse triploidy maerl from St Mawes (F_{ST} range: 0.33–0.42) followed by Weymouth (F_{ST} range: 0.19–0.33), while in *L. corallioides* the most differentiated sites were Helford River (F_{ST} range: 0.07–0.11) and Swanage (F_{ST} range: 0.02–0.11). Overall, patterns of pairwise F_{ST} emulated the patterns of genetic structure.

Clonal lineages and genetic diversity

Clonality and genetic diversity is only reported for sites in Britain as data for *P. calcareum* from sites in France and Spain have been analysed in a previous study (Jenkins et al. 2021). Based on Prevosti's genetic distances, the clone cut-off threshold (i.e. the threshold at which multi-locus lineages should be defined), was estimated as 0.068 in *P. calcareum* and 0.127 in *L. corallioides* (Figure S11, S12). For *P. calcareum*, this translated to 78 clonal lineages (out of 131 individual samples, which includes the samples from France and Spain), while for *L. corallioides*, this translated to 52 clonal lineages (out of 66 individual samples). The number of clonal linages and the abundance of clones in each lineage was visualised for both species (Fig. 5). Note in *L. corallioides* we present clonality and genetic diversity results for Weymouth and Swanage as a single genetic unit (putative maerl population), as informed by the genetic structure analyses. This is also the case for both St Austell Bay sites in both species, and for both Milford Haven sites in *L. corallioides*. Additionally, these analyses were not carried out for Swanage (*P. calcareum*) or St Tudwal's Island (*L. corallioides*) due to their low sample sizes (less than three individuals).

For *P. calcareum*, one notable result is a single clonal lineage for all coarse triploidy maerl in St Mawes (MawC), which also showed much higher observed heterozygosity and lower negative F_{IS} (Table 2). In contrast, clonal diversity was variable across the other putative maerl populations, evidenced by the range of Pareto β values (Table 2). For instance, a single clonal lineage was dominant at Nare Head, which also has the lowest Pareto β value (excluding the coarse maerl). In comparison, there are lineages of more even distribution in other sites, such as at The Manacles, St Mawes (Maw), the Bizzies, and St Austell Bay. Two lineages were

found across two different sites: MLG19 was found in both St Austell Bay and in St Mawes (Maw) and MLG22 was found in both St Austell Bay and the Bizzies. Observed heterozygosity in *P calcareum* ranged from 0.26–0.65, while expected heterozygosity ranged from 0.19–34. For *L. corallioides*, we identified one large clonal lineage and five smaller ones in Helford River, which is evidenced by its lower Pareto β value (Table 2). In contrast, the other five sites showed either more balanced lineages of similar sizes (Milford Haven and St Mawes), supported by their slightly higher Pareto β values, or no clones were detected (St Austell Bay and Weymouth/Swanage). Observed heterozygosity in *L. corallioides* ranged from 0.32–0.39, while expected heterozygosity ranged from 0.29–0.30 (Table 2). For both species, and in sites where clones were detected, Pareto β values did not exceed 2.10. In addition, expected heterozygosity was lower than observed heterozygosity in all putative maerl populations and all F_{IS} values were negative.

Genotype-environment associations and genomic offset

For *P. calcareum*, using benthic ocean temperature and salinity as climate predictors, 2,476 and 267 GEA outlier loci were identified by LFMM and RDA, respectively. Of these, 138 were common across both methods and were considered as putative climate-associated SNPs. RDA was modelled for only the 138 climate-associated SNPs, which explained 19.2% (adjusted r^2) of the variation in the allele frequencies, and the results were visualised as an RDA biplot (Fig. 6A). This showed that most individuals from Zara Shoal, Weymouth and Trévignon were associated with lower salinities and generally lower temperatures. The Bizzies Reef was generally associated with lower temperatures and mostly higher salinities. In contrast, individuals from Morlaix were associated with higher temperatures. The remaining sites were generally positioned in the centre of the RDA biplot, with some minor exceptions within sites (such as a group of individuals from Nare Head that were linked to higher salinities).

For *L. corallioides*, salinity was highly correlated with temperature (r^2 = 0.94); therefore, genotypes were only modelled as a function of temperature. Very few GEA outlier loci were detected: 11 by LFMM and 34 by RDA, with only a single locus common across both methods. As a result, no further GEA analyses were conducted with *L. corallioides*.

All values for future climate variables are predicted to increase (temperature) or decrease (salinity) by 2050 for the SSP245 scenario. The degree of these changes, however, varied by site. The predicted changes for each variable were as follows: (i) ocean temperature increases ranging from 0.41–0.44^{o}C (Galicia) to 0.68^{o}C (Weymouth), and (ii) salinity decreases ranging from − 0.26 PSU (Trévignon) to -0.42–0.44 PSU (all sites from Cornwall and Zara Shoal). Genomic offsets were averaged for each site and visualised on a map (Fig. 6B). The lowest genomic offsets for 2050 SSP245 were observed in Trévignon, while the highest genomic offsets were observed in all sites from Cornwall, Zara Shoal and Illa de Ons.

Microbiota community of maerl across sites

Microbiota site averages showed bacteria were dominant (78–96%) relative to archaea (4–22%), the latter of which were found in higher proportions in Bembridge and Swanage, likely due to the presence of nitrifying Archaea (see below). Overall, one of the most common group of bacteria detected were the Pseudomonadota, ranging from 26–51%, which were mainly composed of Alphaproteobacteria (7–23%) and Gammaproteobacteria (5–26%) (Fig. 7). Actinobacteria and Cyanobacteria were detected but in much lower

proportions. The mean proportion of Flavobacteria appears to follow a downward trend from St Mawes (dense live maerl – mostly category A1) to Bembridge (scattered mostly dead maerl – mostly category C2). Flavobacteria were on average 18% (13–23%, $p < 0.001$) and 12% (7–18%, $p < 0.001$) higher in St Mawes and Helford River, respectively, compared to Bembridge. Additionally, Flavobacteria were on average 5.4% (0.55-10%, $p = 0.024$) higher in Milford Haven compared to Bembridge. No differences in the proportion of Flavobacteria were observed amongst The Manacles, Swanage and Bembridge sites.

In contrast, the opposite trend was apparent in Nitrospirota (Bacteria) and Nitrososphaerota (Archaea), whereby there was an upward trend in proportion from St Mawes to Bembridge. Nitrospirota were on average 2.9–3.8% higher in Bembridge compared to each of St Mawes, Helford River, Milford Haven and The Manacles (all comparisons $p < 0.001$). Similarly, Nitrososphaerota were on average 8.8–17% higher in Bembridge compared to each of St Mawes, Helford River and Milford Haven (all comparisons $p < 0.001$), and The Manacles ($p = 0.016$). Swanage showed on average a slightly lower proportion of Nitrospirota compared to Bembridge (1.0%, $p = 0.008$), but no difference was observed in Nitrososphaerota ($p = 0.997$). The microbiota proportions for all six sites and for each individual sample are available in the Supporting Information (Figure S13).

DISCUSSION

Combining whole genome sequencing and *de novo* genome assembly, we investigated population genomics, climate adaptation and microbiota communities in two species of coralline algae that form maerl bed habitats in the north-east Atlantic. Our study presents the first draft reference genome assemblies for two maerl-forming species: *Phymatolithon calcareum* (GCA_040759855) and *Lithothamnion corallioides* (GCA_040759835). We built a species tree using BUSCO protein sequences extracted from all currently available coralline algae genomes and transcriptomes; this showed very similar evolutionary relationships to a previous comprehensive phylogenetic study of coralline algae that used seven organelle and nuclear genes (Peña et al. 2020). Whole genome sequence data were mapped to the draft genomes of *P. calcareum* and *L. corallioides* to generate SNP datasets for investigating population genomics and potential genetic adaptations to climate, while unmapped reads were used to explore the microbiota community living on or within maerl thalli. These results, as well as the identification of genetic management units and conservation implications, are discussed below.

Species identification of maerl sampled in England and Wales

We sampled novel sites in England and Wales containing maerl that have yet to be assigned species identifications in the scientific literature and in UK maerl bed conservation. Our results confirm the coralline algal species forming maerl beds (or scattered live maerl) at these locations: Bembridge (*P. calcareum*), the Bizzies Reef (*P. calcareum*), Gerrans Bay (*P. calcareum*), Nare Head (*P. calcareum*), St Austell Bay (*P. calcareum* and *L. corallioides*), Weymouth (*P. calcareum* and *L. corallioides*), Swanage (*P. calcareum* and *L. corallioides*), Helford River (*L. corallioides*), Milford Haven (*L. corallioides*), and St Tudwal's Island (*L. corallioides*). It should be noted, however, that sampling may not have covered the full spatial distribution of maerl at each location and, therefore, other maerl-forming species may be present but were undetected in the present study.

Marine Conservation Zones (MCZs), a type of national MPA in England, have been designated to protect a range of nationally important, rare or threatened habitats and species. However, while several MCZs list maerl beds as a feature in their designation (e.g. The Manacles, Purbeck Coast, Bembridge), the particular maerl-forming species is not identified as a feature in the listing. This is primarily because Natural England, and other national public bodies in the UK, emphasise the importance of the maerl habitat rather than the specific species (Axelsson 2023). In addition, because maerl is extremely challenging to identify to species level based on external morphology, and especially in the field, the maerl-forming species at these locations are typically unknown. Consequently, this has also made it challenging to design and carry out effective sampling strategies for specific maerl-forming species, which in part explains why some sites in our study have low sample sizes for *P. calcareum* or *L. corallioides* (the other reason for this is that live maerl is very sparse at some sites, e.g. Bembridge). Nevertheless, given that both *P. calcareum* and *L. corallioides* are identified as priority species under the UK Biodiversity Action Plan (now replaced by the UK Biodiversity Framework 2024), our results enable adding species information to the designation orders and/or to the statutory conservation advice for the MCZs outlined above. Knowledge of the species forming live maerl at each MCZ and each location sampled will be informative for future research on these maerl-forming species and for any restoration projects, where it is important to know the diversity of maerl-forming species, particularly the dominant species, present at a given maerl habitat (Axelsson 2023).

Coarse triploidy maerl in St Mawes

In a previous study, *P. calcareum* from St Mawes was found to be genetically unique compared to all other maerl beds sampled in western Europe (Jenkins et al. 2021). In the current study, additional sampling of maerl beds in and around Falmouth, and higher sequencing depth, has enabled further insight into this finding. First, we found that only the newly recognised coarse growth form of *P. calcareum* in St Mawes is genetically unique. Second, there is strong evidence that this is due to triploidy, whereby one or more chromosomes have three copies rather than the two expected in maerl (Pardo et al. 2019). Triploidy in this *P. calcareum* form from the Fal Estuary was also postulated by Pardo and colleagues using microsatellite genotypes (Pardo et al. 2019), which further supports our hypothesis that triploidy is driving the genetic differentiation observed in the coarse maerl. These findings raise two follow-up questions. First, is triploidy the result of hybridisation (allopolyploidy) or whole genome duplication (autopolyploidy)? And, second, has triploidy contributed to the development of the markedly larger and bulkier growth morphology of this form of *P. calcareum*, and if so, is it a neutral or adaptive change? Given that no genomic data exist for other species in the *Phymatolithon* genus, and that no information is available on the karyology or physiology of the coarse growth form, further research will be needed to resolve these questions.

Population structure and clonal lineages

Analysis of genetic structure and clonality revealed both similarities and differences in locations where both maerl-forming species co-exist in sympatry. For instance, maerl from both species in Weymouth (and Swanage in *L. corallioides*) are genetically differentiated from their respective maerl in Cornwall. This is likely a result of geographical isolation, possibly driven by limited dispersal capacity (Pardo et al. 2019), and indicates that maerl in Weymouth/Swanage should be considered as separate populations for both *P. calcareum* and *L. corallioides*. When comparing clonality in Weymouth/Swanage, no clones were detected in *L. corallioides*, whereas one clonal lineage had three (out of five) individual members in *P. calcareum*, all three of which were

also assigned to a single genetically distinct ancestral cluster with no evidence of admixture with Cornwall. These differences could represent evolutionary histories between *P. calcareum* and *L. corallioides*, possibly driven by different rates of clonal versus sexual reproduction. However, given sample sizes of five (*P. calcareum*) and seven (*L. corallioides*) in these areas, we may not have sampled all potential clonal lineages. Therefore, additional sampling of maerl would enhance inferences of clonality and population structure at these locations.

In Cornwall, discounting the genetically distinct coarse triploidy maerl (discussed above), there are also interesting similarities between species, but some subtle and potentially important differences. For example, in both species the St Mawes maerl bed is admixed with other maerl beds in Cornwall, with similar levels of clonality, characterised by one or two relatively dominant lineages and several singleton lineages. For *P. calcareum*, this contrasts with previous studies where only the coarse triploid maerl was sampled (Pardo et al. 2019; Jenkins et al. 2021), because our results indicate that there is shared ancestry between St Mawes (the non-coarse maerl form) and maerl sampled elsewhere in Cornwall. Indeed, there is evidence of admixture among all sites from Cornwall in *P. calcareum* (with the potential exception of Nare Head). Given that a single maerl thallus is potentially over 100 years old (Foster 2001), and that the dispersal capacity of maerl-forming species is thought to be limited (Pardo et al. 2019), this suggests that the genetic connectivity observed among these locations has likely acted over a very long temporal scale. This is supported by evidence of some mixed ancestry with Zara Shoal (Northern Ireland) and other European maerl beds (Jenkins et al. 2021), which suggests that *P. calcareum* has a complex evolutionary history that has likely been influenced by many factors, such as periodic changes in habitat suitability, connectivity and glacial refugia. Nonetheless, we infer that *P. calcareum* maerl from sites sampled in Cornwall are mostly admixed populations which may represent a contemporary/historic metapopulation.

In *L. corallioides*, we found two ancestral clusters in the Helford River, one of which was only found at this location, and based on our clonal analysis this is a distinct and large clonal lineage. Similarly, we observed a congruent pattern at Nare Head for *P. calcareum*, whereby a dominant clonal lineage was present. This may suggest that the same set of environmental and evolutionary processes have shaped the population structure of maerl at these locations. In contrast, there are some subtle genetic differences between the two species across Cornwall. For instance, while *P. calcareum* maerl in St Austell Bay exhibits shared ancestry from two ancestral sources, *L. corallioides* maerl in St Austell Bay is primarily derived from a single ancestral source. Moreover, no clones were detected in St Austell Bay for *L. corallioides*, whereas two clonal lineages composed of two or three clones were detected for *P. calcareum*. This could reflect the differences in evolutionary history between the two species in this area, whether it be geographical isolation, historic (re)colonisation, connectivity, or a mixture of these.

In Milford Haven, only *L. corallioides* maerl was found at this location in our study, so no comparison could be made with *P. calcareum*. Nevertheless, we found that Milford Haven maerl is genetically differentiated from all other sites sampled in England. In addition, Milford Haven maerl is clonally diverse with high heterozygosity relative to other sites, and the two ancestral sources dominant in this region are mostly found here, with very little or no admixture with maerl from Cornwall or Weymouth/Swanage. This suggests that Milford Haven harbours at least one distinct maerl population that is genetically diverse and geographically isolated from other populations in England.

Climate adaptation in maerl-forming species

Climate adaptation was only assessed for *P. calcareum* because only a single GEA outlier locus was detected for *L. corallioides*, which may be because there is not enough environmental variability across the sites where *L. corallioides* was sampled. For *P. calcareum*, after controlling for spatial autocorrelation (the potential influence of geographic distances correlating with environmental gradients among locations), the GEA analysis found 138 putative climate-associated SNPs. An RDA using just these SNPs showed some site-specific clustering with temperature and salinity. For example, maerl from Morlaix were associated with higher temperatures, more so than maerl from Galicia (Bornalle and Illa de Ons), despite present-day temperatures being on average higher in Galicia, though the next highest benthic temperatures after Galicia (15.0^{o}C) are in Morlaix (13.7^{o}C). This suggests that maerl in Morlaix may have putative genetic adaptations to benthic ocean temperatures. However, despite potential adaptations to temperature, Morlaix appears to have a lower (genomic offset) risk of being maladapted to temperature and salinity conditions in 2050 compared to many other sites. According to our analysis, the maerl bed with the lowest risk of maladaptation is Trévignon; this is probably because this area has the smallest predicted decreases in salinity (0.26 PSU compared to the next decrease of 0.38 PSU in Weymouth) and the second smallest increase in temperature (0.51^{o}C after 0.41^{o}C–0.44^{o}C in Galicia). Interestingly, Bornalle, a maerl bed located within an inner estuary, had a lower maladaptation risk compared with Illa de Ons, a maerl bed located within an outer estuary. This was an unexpected result given that temperature and salinity is broadly similar across both Galician sites, both in terms of present-day and future change. However, coastal configuration and the location of maerl beds within estuaries has been found to shape genetic structure and clonality in *P. calcareum* (Pardo et al. 2019); this may suggest that subtle changes in temperature or salinity could have higher ramifications for maerl populations situated in the outer zones of estuaries in Galicia.

In the UK, maerl from Cornwall and Zara Shoal had higher risks of maladaptation to conditions in 2050 compared to Weymouth and two maerl beds from France. Weymouth has the highest projected temperature increase (0.68^{o}C) of any site, while all sites from Cornwall and Zara Shoal have the highest projected salinity decrease (0.42–0.44 PSU). However, even though Weymouth also has projected salinity decreases (0.35 PSU), it appears that predicted salinity decreases, as well as temperature increases in Zara Shoal and Cornwall (0.53-0.60^{o}C), may result in a greater risk of maladaptation to conditions in 2050 for maerl across these areas. This indicates that *P. calcareum* maerl from Cornwall, Zara Shoal and Illa de Ons will have to have greater shifts in their allele frequencies to mitigate the risk of being maladapted to future environments. It should be noted, however, that the RDA explained 19.2% of the variation in the allele frequencies of the climate-associated SNPs, which suggests that a large portion of variation in these outlier loci is explained by other factors, e.g. other localised environmental parameters or biological processes. Moreover, predictions of maladaptation risk do not consider changes in other conditions that may have additional impacts on the habitat suitability and survival of maerl, such as increased exposure to physical disturbances, pollution or invasive species.

Microbiota differences between maerl beds of varying condition

Metagenomics analysis revealed that Alphaproteobacteria and Gammoproteobacteria, both classes of the Proteobacteria phylum, were the most common microbe found on or within maerl samples across sites

(except at Bembridge and Swanage). These results accord with findings in other coralline algae, such as *Corallina officinalis* (Brodie et al. 2016) and *Neogoniolithon trichotomum* (Valdespino-Castillo et al. 2021), whereby the core microbiota was made up of Proteobacteria, along with several other groups that were also detected in our study, including Bacteroidetes (which includes the Flavobacteria class), Actinobacteria and Cyanobacteria.

Flavobacteria comprised a larger proportion of the microbiota community in St Mawes and Helford River (mostly dense maerl beds with live maerl samples) compared to Bembridge (mostly dead maerl bed with decaying or dead maerl samples). This group of aerobic bacteria can be found free-living, attached to detritus in the water column, or colonising the surface of macroalgae (Mann et al. 2013; Gavriilidou et al. 2020), the latter of which explains why these bacteria were detected on maerl. In contrast, a lower proportion of nitrifying microbes on maerl were found in St Mawes, Helford River and Milford Haven, while a much higher proportion of nitrifying microbes were found on maerl in Bembridge and Swanage. Proportions of microbiota from maerl at The Manacles were roughly in between these two profiles. The apparent replacement of Flavobacteria with nitrifying microbes may suggest a functional role for Flavobacteria on live maerl thalli. Alternatively, the proportional increase in nitrifying microbes could indicate the presence of more nitrogenous compounds in the environment, potentially originating from nearby sewage discharges or fish farms (Legrand et al. 2024). Eutrophication is problematic for maerl because excessive growth of other epiphytic algae can increase sedimentation and create hypoxic conditions, both of which inhibit photosynthesis and respiration (Grall & Hall-Spencer 2003). If this is true, our findings may suggest that the Bembridge maerl bed, and to some extent the Swanage maerl bed, may be (or may have been) in a state of degradation caused by eutrophication. However, this is speculative without further spatial and temporal sampling, as it is not possible to ascertain whether these differences in microbiota are consistent across space and time, and whether the patterns observed are the cause or a symptom of degradation in Bembridge or Swanage maerl.

Genetic units and conservation implications

We assigned locations where maerl was sampled to genetic management units based on all the data collected and analysed in this study (Table 2). Because genetic analysis is conducted on single species, the units are specified for each species (*P. calcareum* and *L. corallioides*). However, as some locations harbour maerl from both species, it seems probable that new or improved conservation actions in one area could benefit more than one maerl-forming species. For each genetic unit, we assessed genetic diversity (alpha diversity) and genetic differentiation (beta diversity) EBVs and used this information to guide our conservation insights (see below). We do not report inbreeding or effective population size (N_e) EBVs because the complex life history and reproduction of maerl-forming species make these metrics difficult to determine with confidence and, particularly for N_e, we do not yet have large enough sample sizes for accurate and meaningful assessments.

Four *P. calcareum* genetic units are currently located within a designated MPA: St Mawes coarse form (Fal and Helford SAC), St Mawes non-coarse form (Fal and Helford SAC), The Manacles (The Manacles MCZ), and Weymouth (Purbeck Coast MCZ). The Fal and Helford SAC also encompasses *L. corallioides* maerl from the Helford River (a distinct genetic unit in this species), which overlaps with the Helford River MCZ, but this MCZ is designated specifically for native oysters. The coarse, triploid maerl from St Mawes are all clones that descend from a single genetically distinct lineage. Thus, this genetic unit of *P. calcareum* has no clonal

diversity but interestingly has high observed heterozygosity (Table 2), which is likely a result of triploidy introducing additional alleles at many SNP loci. Indeed, it could be considered as an Evolutionary Significant Unit (ESU) because it has likely been segregated for many generations and is defined by a unique set of genetic (triploidy) and phenotypic (coarse growth form) traits (Funk et al. 2012). The loss of such an ESU would be equivalent to losing a distinct product of evolutionary history that may have evolved unique and/or adaptive traits (Hoban et al. 2022). In addition, we also found a relatively large clonal lineage in the Helford River that was genetically distinct from all other *L. corallioides* maerl analysed. Accordingly, considering the high maerl bed density and the distinct genetic diversity found in St Mawes and the Helford River, we advocate that the Fal Estuary and Helford River area remains a high priority for marine conservation.

For the remaining sites in Cornwall, only The Manacles has a MCZ designated to protect maerl beds, which has a byelaw to prohibit scallop dredging and bottom towed gear under *The Manacles Marine Conservation Zone (Fishing Restrictions) Byelaw 2017* (Cornwall IFCA). In Nare Head, we found one dominant clonal lineage of *P. calcareum*, further analysis of which suggested that this may be genetically differentiated from other Cornish maerl. In addition, *P. calcareum* maerl from The Manacles, the Bizzies, Gerrans Bay and St Austell Bay are possibly individual populations that have been or are connected by historic or contemporary gene flow. All these sites also have a higher risk of being maladapted to future temperatures and salinities relative to other sites in England and France. We therefore recommend that an additional MPA be considered for Gerrans Bay which would cover both the maerl from Gerrans Bay and Nare Head and safeguard the genetic diversity at these locations. A secondary priority in Cornwall would be to establish another MPA in St Austell Bay to protect both maerl-forming species, the benefit of which could enhance genetic connectivity across the region. In addition, for *L. corallioides*, this would conserve the high clonal diversity observed at this location, which may be indicative of higher rates of sexual reproduction generating new lineages. The prevalence of sexual reproduction could be explored at this site, or indeed any site, by conducting additional surveys to search for attached gametophytes with uniporate conceptacles (Pardo et al. 2019). To our knowledge, the only reported gametophyte of any maerl-forming species has been found in a *P. calcareum* maerl bed in Brittany, France (Pardo et al. 2017). Nevertheless, the detection of these attached gametophytes would have important implications for maerl bed conservation since they are central to introducing novel genetic variants into the population via sexual reproduction.

In Weymouth and Swanage, both maerl-forming species have isolated populations at one or both of these locations and these should be considered as separate genetic units to maerl beds in Cornwall. Currently, Purbeck Coast MCZ encompasses maerl beds at these locations. As mentioned previously for genetically diverse / distinct areas, we recommend that conservation of these sites is prioritised, particularly as there is likely very limited connectivity between these maerl beds and those in Cornwall. On the east coast of the Isle of Wight, we also collected maerl samples from Bembridge, which comprised mostly dead maerl with very little cover of living (pink) algae on each thallus. Purbeck Coast MCZ and Bembridge MCZ list maerl beds as a protected feature in their designation (in addition to other features), with a current management goal of "*recover to favourable condition*". Although we could not determine any genetic EBVs for Bembridge due to insufficient SNP data (likely a result of low DNA yield due to very little algal material on maerl thalli), we were able to identify five samples as *P. calcareum*. Thus, any future restoration project in Bembridge involving the translocation of live maerl from another maerl bed should take this into consideration.

In south-west Wales, all maerl from Milford Haven was classified as a single genetic unit, which has high genetic diversity and is genetically differentiated from other *L. corallioides* maerl beds in England. Milford Haven had the highest and second highest measures of clonal diversity (Pareto β) and observed heterozygosity, respectively. This suggests that, over many past generations, both sexual and clonal reproduction has occurred to produce new lineages and to proliferate existing lineages in the population. It is therefore important that the Pembrokeshire Marine SAC, which encompasses the Milford Haven area, provides adequate protection for maerl from existential threats, since any loss or regression of this maerl bed could erode the unique genetic diversity found in *L. corallioides* at this location. In addition, although we were able to obtain some preliminary data on the genetic structure of maerl from St Tudwal's Island (north-west Wales), more samples will be needed to accurately assess population structure and compute genetic EBVs.

Lastly, although population structure and clonality in other European maerl beds (Northern Ireland, France and Spain) were reported in previous studies for *P. calcareum* (Pardo et al. 2019; Jenkins et al. 2021), by including these WGS data in this study we were able to assess their risk of maladaptation to future climates in 2050 under the SSP245 scenario. Zara Shoal (Strangford Lough, Northern Ireland), Illa de Ons (Galicia, Spain) and all Cornish maerl beds were at elevated risk compared to all other sites. The Strangford Lough SAC lists maerl beds as a protected feature in its conservation objectives, while the Espacio marino de las Rias Baixas de Galicia MPA appears to encompass Illa de Ons (but not Bornalle). At present, therefore, it appears as though the European maerl beds listed above, that have higher risks of maladaptation, are currently positioned within MPAs. However, although analysis of genetic data can tell us about genetic diversity and potential resilience to change, deciding where to prioritise resources should also consider the density and health of maerl beds, as well as potential mechanisms for effectively protecting these habitats from threats in the short-term (e.g. physical disturbances and pollution) and the long-term (e.g. climate change).

Conclusion

In this study, genomic data was used to address important knowledge gaps in the population structure, genetic diversity, climate resilience and microbiota of two maerl-forming species, *P. calcareum* and *L. corallioides*, found across south-west Britain, Northern Ireland, France and Spain. In both species, the discovery of distinct genetic structure suggests that some maerl populations harbour unique genetic diversity and are genetically and/or geographically isolated. For example, maerl of both species in Weymouth/Swanage (Dorset) were found to be genetically differentiated from all maerl in Cornwall. Additionally, in *L. corallioides*, the Milford Haven (south-west Wales) maerl population was genetically diverse and isolated from other maerl populations in England. In Cornwall, we identified the genetically and phenotypically distinct coarse growth form of *P. calcareum* and found that this coarse form co-exists with a non-coarse form in St Mawes, the latter of which, unlike the coarse form, shares ancestry with neighbouring maerl populations (e.g. The Manacles and St Austell Bay). Moreover, in the Helford River, we found a relatively large clonal lineage of *L. corallioides* that is genetically distinct and only found at this location. These findings further emphasise the high conservation priority of these maerl beds around Falmouth, both in terms of maerl bed density and species/genetic diversity, and underline the important role of designated MPAs, such as the Fal and Helford SAC, to protect the Fal Estuary and Helford River area from dredging and bottom towed gear. In addition, we found that populations of *P. calcareum* from Cornwall, Zara Shoal (Northern Ireland) and Illa de Ons (Spain) are predicted to have an elevated risk of maladaptation to projected changes in temperature and salinity in 2050 under the SSP245

$m \leq oftheroad$ climate change scenario. Lastly, we report differences in the microbiota community (i.e. bacteria and archaea) between maerl from dense, healthy maerl beds (e.g. St Mawes) compared to scattered, mostly dead maerl (e.g. Bembridge). Altogether, we use these data to propose genetic management units for maerl populations in south-west Britain, which will inform efforts to prevent further loss or degradation to maerl beds and aid conservation interventions to recover and restore these irreplaceable, high natural capital value habitats.

Declarations

Data archiving statement

Raw DNA sequence data and draft genome assemblies are available from the NCBI (BioProject: PRJNA682082; *Phymatolithon calcareum* assembly: SAMN41910763; *Lithothamnion corallioides* assembly: SAMN41910813). The modified BUSCO phylogenomics pipeline is available from GitHub: Tom-Jenkins/BUSCO_phylogenomics. The custom python scripts used for organelle annotation (extract_CDS.py) and for processing Kraken2 reports (process_kraken2_reports.py) are part of the nextflow pipelines repository available from Zenodo: https://doi.org/10.5281/zenodo.14056754. R code and supporting data used in the analyses are also available from Zenodo: https://doi.org/10.5281/zenodo.14592386.

References

1. Aguirre-Liguori JA, Ramírez-Barahona S, Gaut BS (2021) The evolutionary genomics of species' responses to climate change. *Nature Ecology & Evolution*, **5**, 1350–1360.
2. Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics*, **11**, 697–709.
3. Arnaud-Haond S, Duarte CM, Alberto F, Serrão EA (2007) Standardizing methods to address clonality in population studies. *Molecular Ecology*, **16**, 5115–5139.
4. Assis J, Fernández Bejarano SJ, Salazar VW *et al.* (2024) Bio-ORACLE v3.0. pushing marine data layers to the CMIP6 Earth System Models of climate change research. *Global Ecology and Biogeography*, **33**, e13813.
5. Axelsson M (2023) A categorisation system for maerl bed habitats in England. *Natural England*, **NERR123**.
6. Bernard G, Romero-Ramirez A, Tauran A *et al.* (2019) Declining maerl vitality and habitat complexity across a dredging gradient: insights from in situ sediment profile imagery (SPI). *Scientific Reports*, **9**, 1–12.
7. Bernatchez L, Ferchaud A-L, Berger CS, Venney CJ, Xuereb A (2024) Genomics for monitoring and understanding species responses to global climate change. *Nature Reviews Genetics*, **25**, 165–183.
8. Blake C, Maggs CA (2003) Comparative growth rates and internal banding periodicity of maerl species (Corallinales, Rhodophyta) from northern Europe. *Phycologia*, **42**, 606–612.
9. Borg M, Krueger-Hadfield SA, Destombe C *et al.* (2023) Red macroalgae in the genomic era. *New Phytologist*, **240**, 471–488.
10. Brodie J, Williamson C, Barker GL *et al.* (2016) Characterising the microbiome of *Corallina officinalis*, a dominant calcified intertidal red alga. *FEMS Microbiology Ecology*, **92**, fiw110.

11. Buchfink B, Reuter K, Drost H-G (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, **18**, 366–368.
12. Camacho C, Coulouris G, Avagyan V *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
13. Carro B, Lopez L, Peña V, Bárbara I, Barreiro R (2014) DNA barcoding allows the accurate assessment of European maerl diversity: a proof-of-concept study. *Phytotaxa*, **190**, 176.
14. Challis R, Richards E, Rajan J, Cochrane G, Blaxter M (2020) BlobToolKit – interactive quality assessment of genome assemblies. *G3*, **10**, 1361–1374.
15. Chen S (2023) Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*, **2**, e107.
16. Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
17. Danecek P, Bonfield JK, Liddle J *et al.* (2021) Twelve years of SAMtools and BCFtools. *GigaScience*, **10**, giab008.
18. Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.
19. Envision Marine LTD (2024) Falmouth Bay to St Austell Bay SPA drop down video and acoustic survey maerl mapping: St Austell Bay 2023. *A report for Natural England by Envision Marine LTD*, NECR589.
20. Fan J, Huang S, Chorlton SD (2021) BugSeq: a highly accurate cloud platform for long-read metagenomic analyses. *BMC Bioinformatics*, **22**, 160.
21. Foster MS (2001) Rhodoliths: between rocks and soft places. *Journal of Phycology*, **37**, 659–667.
22. Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, **27**, 489–496.
23. Gain C, François O (2021) LEA 3: factor models in population genetics and ecological genomics with R. *Molecular Ecology Resources*, **21**, 2738–2748.
24. Gain C, Rhoné B, Cubry P *et al.* (2023) A quantitative theory for genomic offset statistics. *Molecular Biology and Evolution*, **40**, msad140.
25. Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv*, **1207**, 3907.
26. Gavriilidou A, Gutleben J, Versluis D *et al.* (2020) Comparative genomic analysis of Flavobacteriaceae: insights into carbohydrate metabolism, gliding motility and secondary metabolite biosynthesis. *BMC Genomics*, **21**, 569.
27. Grall J, Hall-Spencer JM (2003) Problems facing maerl conservation in Brittany. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **13**, S55–S64.
28. Hall-Spencer J (2000) Scallop dredging has profound, long-term impacts on maerl habitats. *ICES Journal of Marine Science*, **57**, 1407–1415.
29. Hoban S, Archer FI, Bertola LD *et al.* (2022) Global genetic diversity status and trends: towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. *Biological Reviews*, **97**, 1511–1538.

30. Hoban S, Bruford MW, da Silva JM *et al.* (2023) Genetic diversity goals and targets have improved, but remain insufficient for clear implementation of the post-2020 global biodiversity framework. *Conservation Genetics*, **24**, 181–191.
31. James K, Macreadie PI, Burdett HL, Davies I, Kamenos NA (2024) It's time to broaden what we consider a 'blue carbon ecosystem'. *Global Change Biology*, **30**, e17261.
32. Jenkins TL (2024) mapmixture: an R package and web app for spatial visualisation of admixture and population structure. *Molecular Ecology Resources*, **24**, e13943.
33. Jenkins TL, Guillemin M-L, Simon-Nutbrown C *et al.* (2021) Whole genome genotyping reveals discrete genetic diversity in north-east Atlantic maerl beds. *Evolutionary Applications*, **14**, 1558–1571.
34. Jenkins TL, Stevens JR (2018) Assessing connectivity between MPAs: selecting taxa and translating genetic data to inform policy. *Marine Policy*, **94**, 165–173.
35. Kamenos NA, Moore PG, Hall-Spencer JM (2004) Nursery-area function of maerl grounds for juvenile queen scallops *Aequipecten opercularis* and other invertebrates. *Marine Ecology Progress Series*, **274**, 183–189.
36. Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281.
37. Kissling WD, Walls R, Bowser A *et al.* (2018) Towards global data products of Essential Biodiversity Variables on species traits. *Nature Ecology & Evolution*, **2**, 1531–1540.
38. Klemetsen T, Raknes IA, Fu J *et al.* (2018) The MAR databases: development and implementation of databases specific for marine metagenomics. *Nucleic Acids Research*, **46**, D692–D699.
39. Knaus BJ, Grünwald NJ (2017) vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, **17**, 44–53.
40. Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, **37**, 540–546.
41. Laikre L (2010) Genetic diversity is overlooked in international conservation policy implementation. *Conservation Genetics*, **11**, 349–354.
42. Laikre L, Hoban S, Bruford MW *et al.* (2020) Post-2020 goals overlook genetic diversity. *Science*, **367**, 1083–1085.
43. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.
44. Legrand E, Taormina B, Grall J *et al.* (2024) Salmon farming alters the structure and functioning of Norwegian maerl bed communities. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **34**, e4142.
45. Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100.
46. Lipinska AP, Krueger-Hadfield SA, Godfroy O *et al.* (2023) The rhodoexplorer platform for red algal genomics and whole-genome assemblies for several *Gracilaria* species. *Genome Biology & Evolution*, **15**, evad124.
47. Mann AJ, Hahnke RL, Huang S *et al.* (2013) The genome of the alga-associated marine Flavobacterium *Formosa agariphila* KMM 3901^T reveals a broad potential for degradation of algal polysaccharides. *Applied and Environmental Microbiology*, **79**, 6813–6822.

48. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM (2021) BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, **38**, 4647–4654.
49. Mao J, Burdett HL, McGill RAR *et al.* (2020) Carbon burial over the last four millennia is regulated by both climatic and land use change. *Global Change Biology*, **26**, 2496–2504.
50. Martin S, Hall-Spencer JM (2017) Effects of ocean warming and acidification on rhodolith/maerl beds. In: *Rhodolith/Maërl Beds: A Global Perspective*, pp. 55–85. Springer International Publishing, Cham.
51. McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.
52. Melbourne LA, Hernández-Kantún JJ, Russell S, Brodie J (2017) There is more to maerl than meets the eye: DNA barcoding reveals a new species in Britain, *Lithothamnion erinaceum* sp. nov. (Hapalidiales, Rhodophyta). *European Journal of Phycology*, **52**, 166–178.
53. Mirarab S, Warnow T (2015) ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, **31**, i44–i52.
54. Nelson DR, Mystikou A, Jaiswal A *et al.* (2024) Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity. *Molecular Plant*, **17**, 747–771.
55. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268–274.
56. Page TM, McDougall C, Diaz-Pulido G (2019) *De novo* transcriptome assembly for four species of crustose coralline algae and analysis of unique orthologous genes. *Scientific Reports*, **9**, 12611.
57. Pardo C, Bárbara I, Barreiro R, Peña V (2017) Insights into species diversity of associated crustose coralline algae (Corallinophycidae, Rhodophyta) with Atlantic European maerl beds using DNA barcoding. *Anales del Jardin Botanico de Madrid*, **74**, 059.
58. Pardo C, Guillemin M-L, Peña V *et al.* (2019) Local coastal configuration rather than latitudinal gradient shape clonal diversity and genetic structure of *Phymatolithon calcareum* maerl beds in north European Atlantic. *Frontiers in Marine Science*, **6**, 149.
59. Pardo C, Lopez L, Peña V *et al.* (2014) A multilocus species delimitation reveals a striking number of species of coralline algae forming maerl in the OSPAR maritime area. *PloS One*, **9**, e104073.
60. Peña V, Bárbara I, Grall J, Maggs CA, Hall-Spencer JM (2014) The diversity of seaweeds on maerl in the NE Atlantic. *Marine Biodiversity*, **44**, 533–551.
61. Peña V, Vieira C, Braga JC *et al.* (2020) Radiation of the coralline red algae (Corallinophycidae, Rhodophyta) crown group as inferred from a multilocus time-calibrated phylogeny. *Molecular Phylogenetics and Evolution*, **150**, 106845.
62. Pereira HM, Ferrier S, Walters M *et al.* (2013) Essential biodiversity variables. *Science*, **339**, 277–278.
63. Qui-Minet ZN, Coudret J, Davoult D *et al.* (2019) Combined effects of global climate change and nutrient enrichment on the physiology of three temperate maerl species. *Ecology and Evolution*, **9**, 13787–13807.
64. Sandström A, Lundmark C, Andersson K, Johannesson K, Laikre L (2019) Understanding and bridging the conservation-genetics gap in marine conservation. *Conservation Biology*, **33**, 725–728.

65. Scales A (2022) Investigating the microbiota associated with maerl beds located in the Fal Estuary, Cornwall. Masters by Research Thesis. University of Plymouth, Plymouth.
66. Schubert N, Schoenrock KM, Aguirre J *et al.* (2020) Coralline algae: globally distributed ecosystem engineers. *Frontiers in Marine Science*, **7**, 352.
67. Shafer ABA, Wolf JBW, Alves PC *et al.* (2015) Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, **30**, 78–87.
68. Shen W, Le S, Li Y, Hu F (2016) SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PloS One*, **11**, e0163962.
69. Stoeckel S, Porro B, Arnaud-Haond S (2021) The discernible and hidden effects of clonality on the genotypic and genetic states of populations: improving our estimation of clonal rates. *Molecular Ecology Resources*, **21**, 1068–1084.
70. Tuya F, Schubert N, Aguirre J *et al.* (2023) Levelling-up rhodolith-bed science to address global-scale conservation challenges. *The Science of the Total Environment*, **892**, 164818.
71. Valdespino-Castillo PM, Bautista-García A, Favoretto F *et al.* (2021) Interplay of microbial communities with mineral environments in coralline algae. *Science of the Total Environment*, **757**, 143877.
72. Wang J (2022) Fast and accurate population admixture inference from genotype data from a few microsatellites to millions of SNPs. *Heredity*, **129**, 79–92.
73. Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358.
74. Wick RR, Holt KE (2022) Polypolish: short-read polishing of long-read bacterial genome assemblies. *PLoS Computational Biology*, **18**, e1009802.
75. Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology*, **13**, e1005595.
76. Wilson B, Fan C-M, Edmunds PJ (2020) An unusual microbiome characterises a spatially-aggressive crustose alga rapidly overgrowing shallow Caribbean reefs. *Scientific Reports*, **10**, 20949.
77. Wood DE, Lu J, Langmead B (2019) Improved metagenomic analysis with Kraken 2. *Genome Biology*, **20**, 257.
78. Xue JY, Hind KR, Lemay MA *et al.* (2022) Transcriptome of the coralline alga *Calliarthron tuberculosum* (Corallinales, Rhodophyta) reveals convergent evolution of a partial lignin biosynthesis pathway. *PLoS One*, **17**, e0266892.

Tables

Table 1. Summary of maerl sampling locations, the number of samples identified as *Phymatolithon calcareum* or *Lithothamnion corallioides*, and the number of samples genotyped at single nucleotide polymorphisms.

Location	Year	Latitude	Longitude	Code	N_{TOTAL}	N_{GENO}	Depth (m)	Reference
Phymatolithon calcareum								
England, Cornwall, St Austell Bay	2023	50.327	-4.728	Aus	2	2	18	This study
England, Cornwall, St Austell Bay, Little Gribbin	2022	50.319	-4.689	Gri	8	8	5-10	This study
England, Cornwall, Falmouth, Bizzies Reef	2022	50.136	-4.987	Biz	10	10	23	This study
England, Cornwall, Falmouth, St Mawes	2015, 2022 REF	50.156	-5.031	Maw	6, 3	9	4-6	This study & Jenkins *et al.* 2021
England, Cornwall, Falmouth, St Mawes COARSE	2011, 2022 REF	50.166	-5.031	MawC	7, 8	15	4-6	This study & Jenkins *et al.* 2021
England, Cornwall, Gerrans Bay	2023	50.199	-4.938	Ger	5	5	18	This study
England, Cornwall, The Manacles	2015	50.039	-5.060	Man	12	12	5-10	This study & Jenkins *et al.* 2021
England, Cornwall, Nare Head, nr Hera Wreck	2022	50.200	-4.907	Nar	10	10	13-15	This study
England, Dorset, Swanage	2022	50.650	-1.910	Swa	1	1	15-19	This study
England, Dorset, Weymouth	2022	50.616	-2.320	Wey	5	5	15-20	This study
England, Isle of Wight, Bembridge	2022	50.637	-1.141	Bem*	10	–	13-21	This study
Northern Ireland, Strangford Lough, Zara Shoal	2011	54.379	-5.564	Zar	11	11	10	Jenkins *et al.* 2021
France, Brittany, Morlaix	2011	48.711	-3.951	Mor	11	11	10-11	Jenkins *et al.* 2021

France, Brittany, Trévignon	2011	47.795	-3.887	Tre	9	9	15	Jenkins *et al.* 2021
Spain, Galicia, Bornalle	2011	42.789	-9.020	Bor	12	12	11	Jenkins *et al.* 2021
Spain, Galicia, Illa de Ons	2011	42.395	-8.915	Ons	11	11	13	Jenkins *et al.* 2021
Lithothamnion corallioides								
England, Cornwall, St Austell Bay	2023	50.327	-4.728	Aus	10	10	18	This study
England, Cornwall, St Austell Bay, Little Gribbin	2022	50.319	-4.689	Gri	4	4	5-10	This study
England, Cornwall, Falmouth, St Mawes	2015, 2022	50.156	-5.031	Maw	2, 3	5	4-6	This study & Jenkins *et al.* 2021
England, Cornwall, Helford River	2015	50.098	-5.123	Hel	13	13	5-10	This study
England, Dorset, Swanage	2022	50.650	-1.910	Swa	3	3	15-19	This study
England, Dorset, Weymouth	2022	50.616	-2.320	Wey	4	4	15-20	This study
Wales, Pembrokeshire, Milford Haven west	2023	51.705	-5.082	Mil1	13	13	5	This study
Wales, Pembrokeshire, Milford Haven east	2023	51.703	-5.076	Mil2	12	12	5	This study
Wales, Llyn Peninsula, St Tudwal's Island	2023	52.809	-4.456	Tud	2	2	12	This study

N_{TOTAL} Total number of maerl samples collected; N_{GENO} Number of maerl samples successfully genotyped.

^{REF} A reference genome was assembled for one individual sample using Nanopore long read sequencing.

* Only five out of ten samples were successfully identified to species level using DNA barcoding.

Table 2 Genetic management units of maerl sampled from England and Wales. Metrics of clonal and genetic diversity for each unit are reported, alongside current information about Marine Protected Area (MPA)

designations and regulations.

Genetic unit	Genetic diversity				Marine Protected Area (MPA)	Designation year	Regulations
	Pareto β	H_o	H_e	F_{IS}			
Phymatolithon calcareum							
St Mawes	1.63	0.31	0.22	-0.35	Fal and Helford SAC	2005	Prohibition of scallop dredging and bottom towed gear under *The Fal & Helford Designated Area (Fishing Restrictions) Order 2008* (Defra).
St Mawes coarse triploid	0.00	0.65	0.34	-0.85	Fal and Helford SAC	–	–
St Austell Bay	1.78	0.28	0.23	-0.22	Falmouth Bay to St Austell Bay SPA	2017	No known regulations to protect maerl beds.
The Bizzies	2.00	0.26	0.21	-0.22	Falmouth Bay to St Austell Bay SPA	–	–
Gerrans Bay	1.00	0.32	0.21	-0.46	Falmouth Bay to St Austell Bay SPA	–	–
Nare Head	0.53	0.26	0.16	-0.38	Falmouth Bay to St Austell Bay SPA	–	–
Weymouth	1.00	0.28	0.19	-0.39	Purbeck Coast MCZ	2019	Partial prohibition of scallop dredging and bottom tower gear under the *Bottom Towed Fishing Dear Byelaw 2016* (Southern IFCA)
The Manacles	2.10	0.26	0.21	-0.20	The Manacles MCZ	2013	Prohibition of scallop dredging and bottom towed gear under *The Manacles Marine*

							Conservation Zone (Fishing Restrictions) Byelaw 2017 (Cornwall IFCA).
Lithothamnion corallioides							
Helford River	0.88	0.39	0.29	-0.25	Fal and Helford SAC Helford Estuary MCZ	– 2019	– No known regulations to protect maerl beds.
St Mawes	2.00	0.37	0.30	-0.21	Fal and Helford SAC	–	–
St Austell Bay	*Inf*	0.34	0.29	-0.11	Falmouth Bay to St Austell Bay SPA	–	–
Milford Haven	2.07	0.38	0.30	-0.20	Pembrokeshire Marine SAC	2004	Partial prohibition of scallop fishing using towed gear under *The Scallop Fishing (Wales) (No.2) Order 2010*. A 'no anchoring' buffer zone drawn around sensitive habitats including maerl and seagrass.
Weymouth/Swanage	*Inf*	0.32	0.29	-0.07	Purbeck Coast MCZ	–	–

MCZ, Marine Conservation Zone; SAC, Special Area of Conservation; SPA, Special Protection Area; IFCA, Inshore Fisheries and Conservation Authority.

Inf, Infinity denotes that no clones were detected (the number of clonal lineages equals the number the samples collected).

Figures

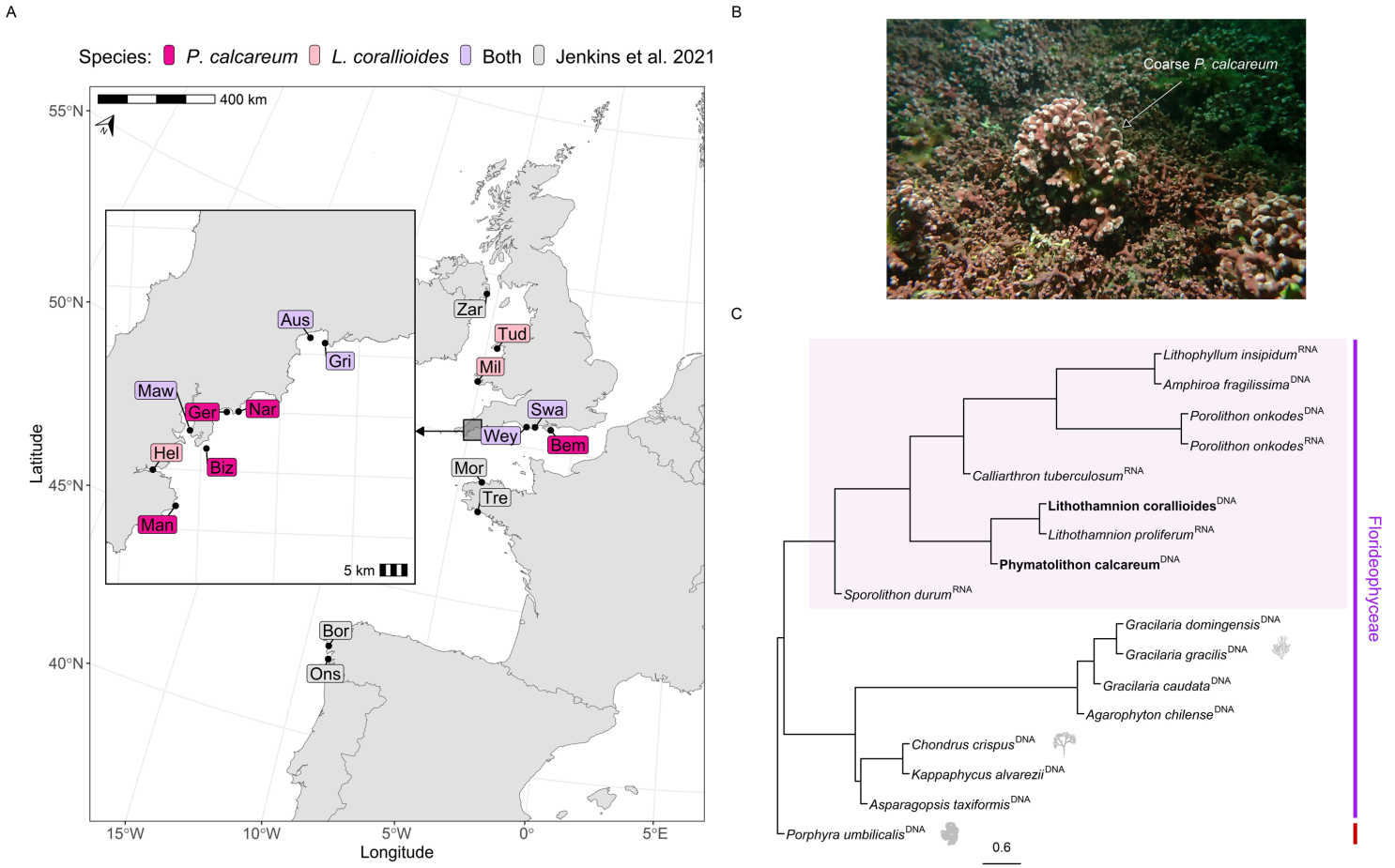

Figure 1

Map of sampling locations and BUSCO species tree. (A) Map of sites where maerl was sampled. Coloured labels denote the maerl-forming species that was identified at each site. (B) Photo of St Mawes maerl bed (image credit: Matt Slater, Cornwall Wildlife Trust). The coarse growth form of *Phymatolithon calcareum* (MawC) is highlighted by the white arrow; the surrounding maerl is a mix of non-coarse *P. calcareum* and *Lithothamnion corallioides*. (C) Red algae species tree of the Florideophyceae (purple vertical line) and one outgroup (Bangiophyceae) (red vertical line), built using 157 BUSCO complete and single-copy genes that were present in at least ten species. The two species analysed in our study are highlighted in bold. At the end of each species name, the superscript label denotes whether the assembly used for BUSCO analysis was a genome (DNA) or a transcriptome (RNA). The light purple box highlights all calcifying red algal lineages in the Corallinophycidae, which includes the coralline algae orders Corallinales and Hapalidiales.

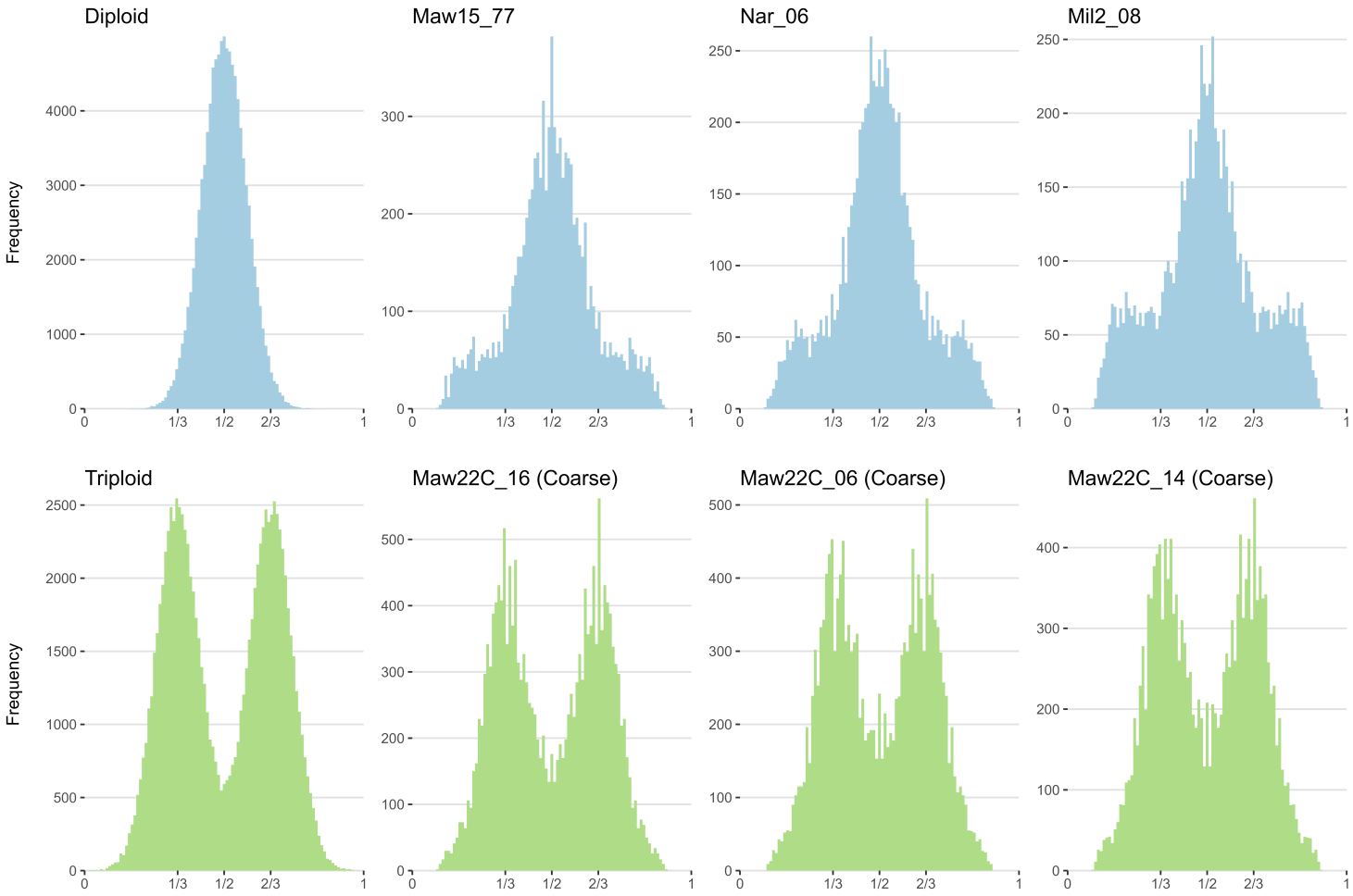

Figure 2

Allele balance and ploidy analysis. The first column shows allele frequency histograms of the expected distribution for a diploid (single peak at 1/2) and a triploid (two peaks at 1/3 and 2/3). The remaining columns show histograms of allele balance proportions from heterozygous SNPs in samples from St Mawes (*Phymatolithon calcareum*), Nare Head (*P. calcareum*) and Milford Haven (*Lithothamnion corallioides*); coarse *P. calcareum* maerl samples are presented on the bottom panel. Colours denote whether the pattern is indicative of diploidy (blue) or triploidy (green).

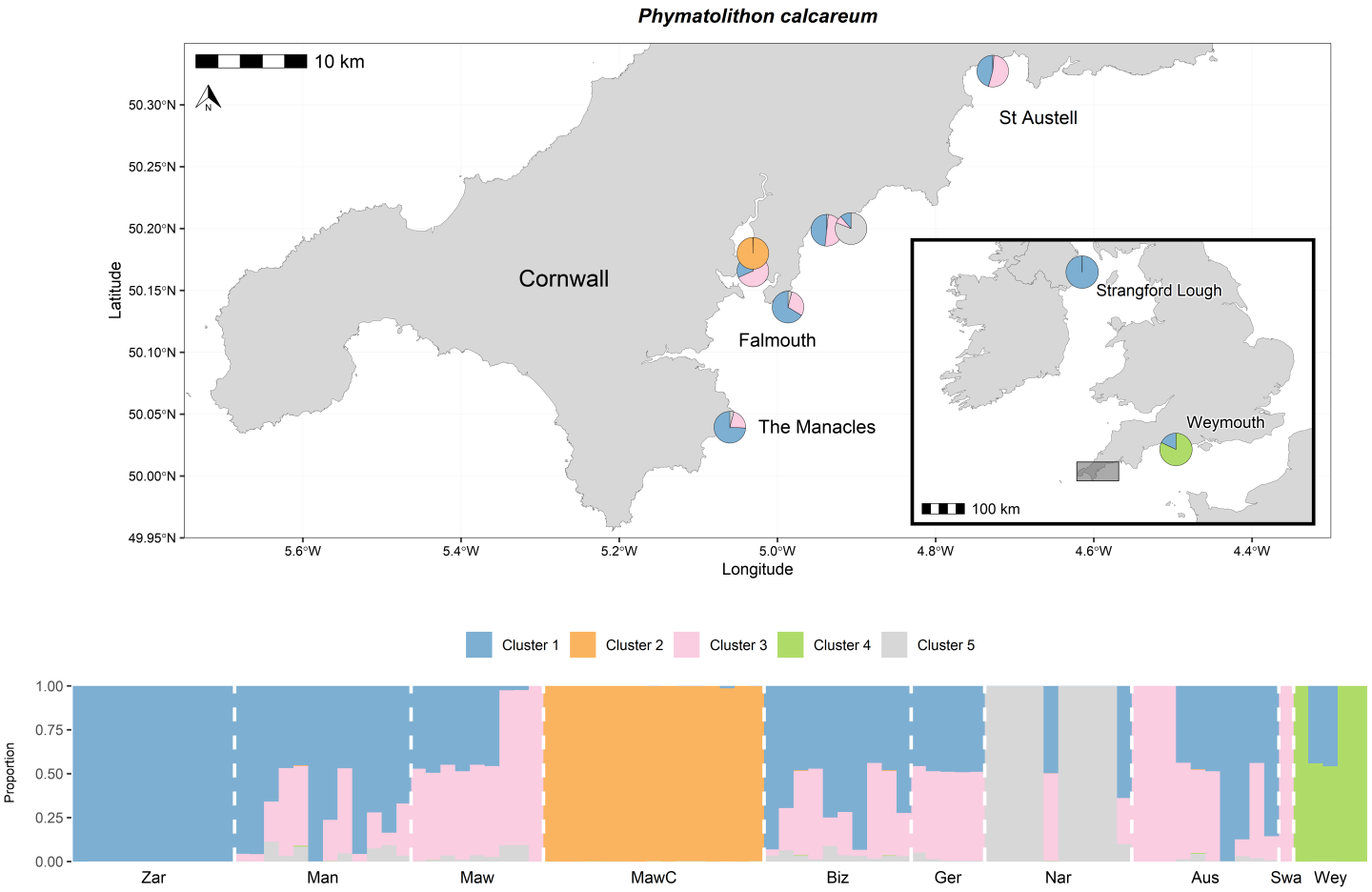

Figure 3

Phymatolithon calcareum admixture map and structure plot. The admixture map (top panel) shows pie charts on a zoomed in map of Cornwall (England), where each pie represents the average admixture proportion across all individuals for a given cluster at a location. An inset admixture map shows the same data for Zara Shoal (Strangford Lough) and Weymouth (Dorset). The structure plot (bottom panel) shows the same admixture data but per individual, where each bar represents the proportion of their genome derived from each *K* source ancestral population.

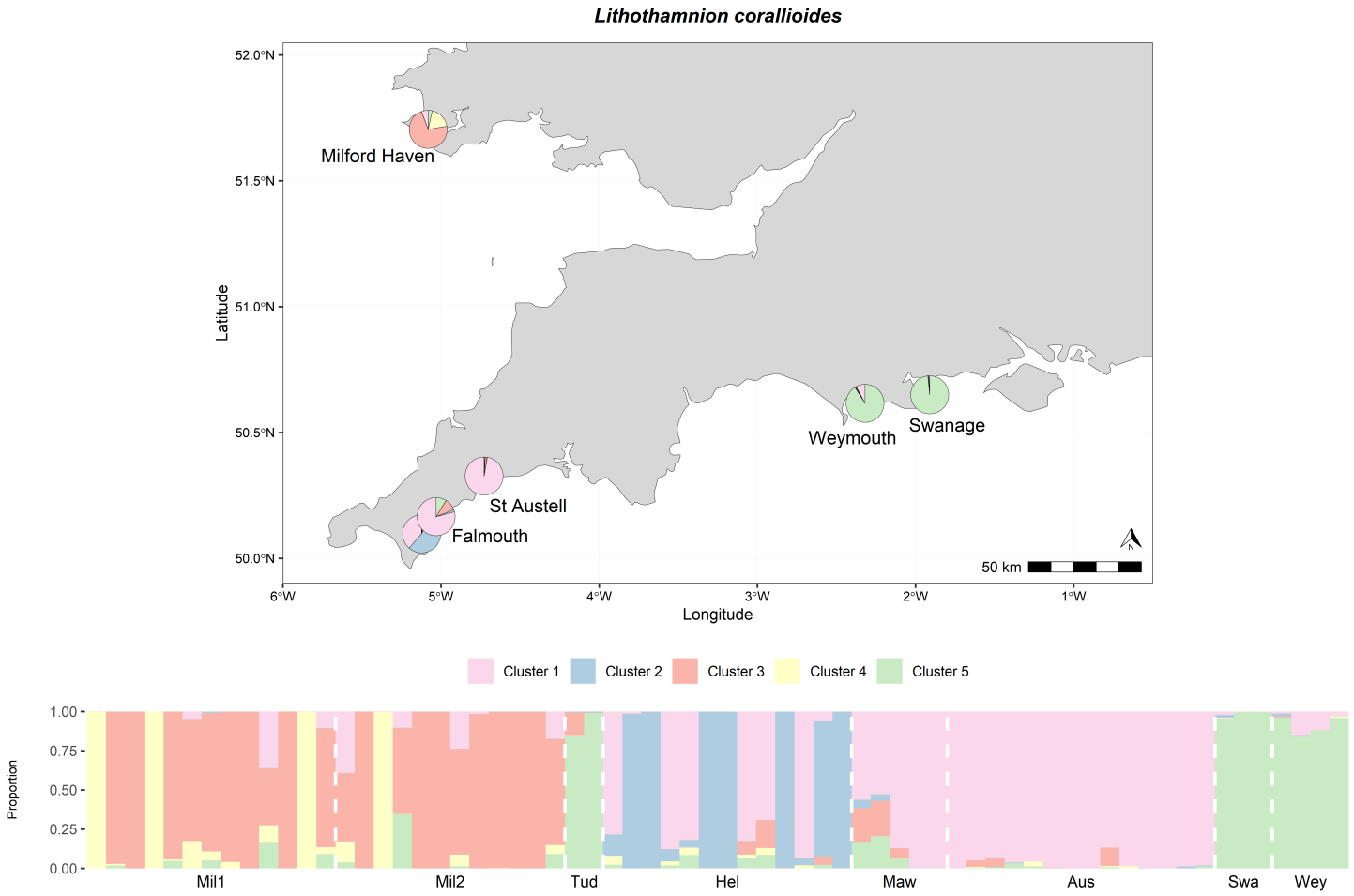

Figure 4

Lithothamnion corallioides admixture map and structure plot. The admixture map (top panel) shows pie charts on a zoomed in map of south-west England and southern Wales, where each pie represents the average admixture proportion across all individuals for a given cluster at a location. The structure plot (bottom panel) shows the same admixture data but per individual, where each bar represents the proportion of their genome derived from each *K* source ancestral population.

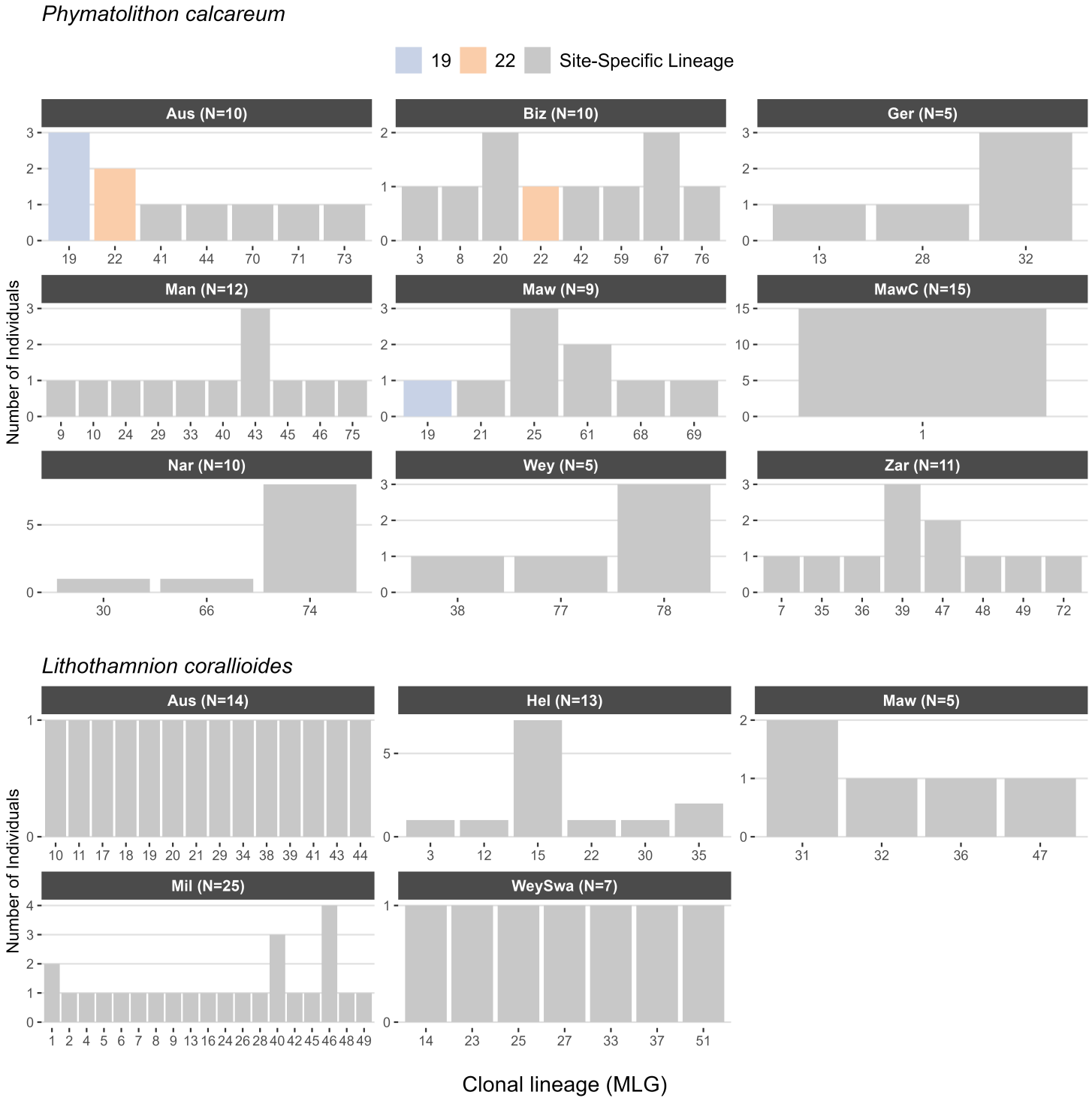

Figure 5

Clonality analysis per genetic unit in *Phymatolithon calcareum* and *Lithothamnion corallioides*. The clonal lineage (multi-locus genotype, MLG) is numbered on the x-axis and the number of individuals in each clonal lineage is represented by bars on the y-axis. Each clonal lineage was identified by defining a clone threshold (see main text), which is necessary because somatic mutations can introduce very small genetic differences between clones. Grey bars indicate that a clonal lineage is only present in that location, whereas coloured bars indicate that a lineage was found in multiple locations.

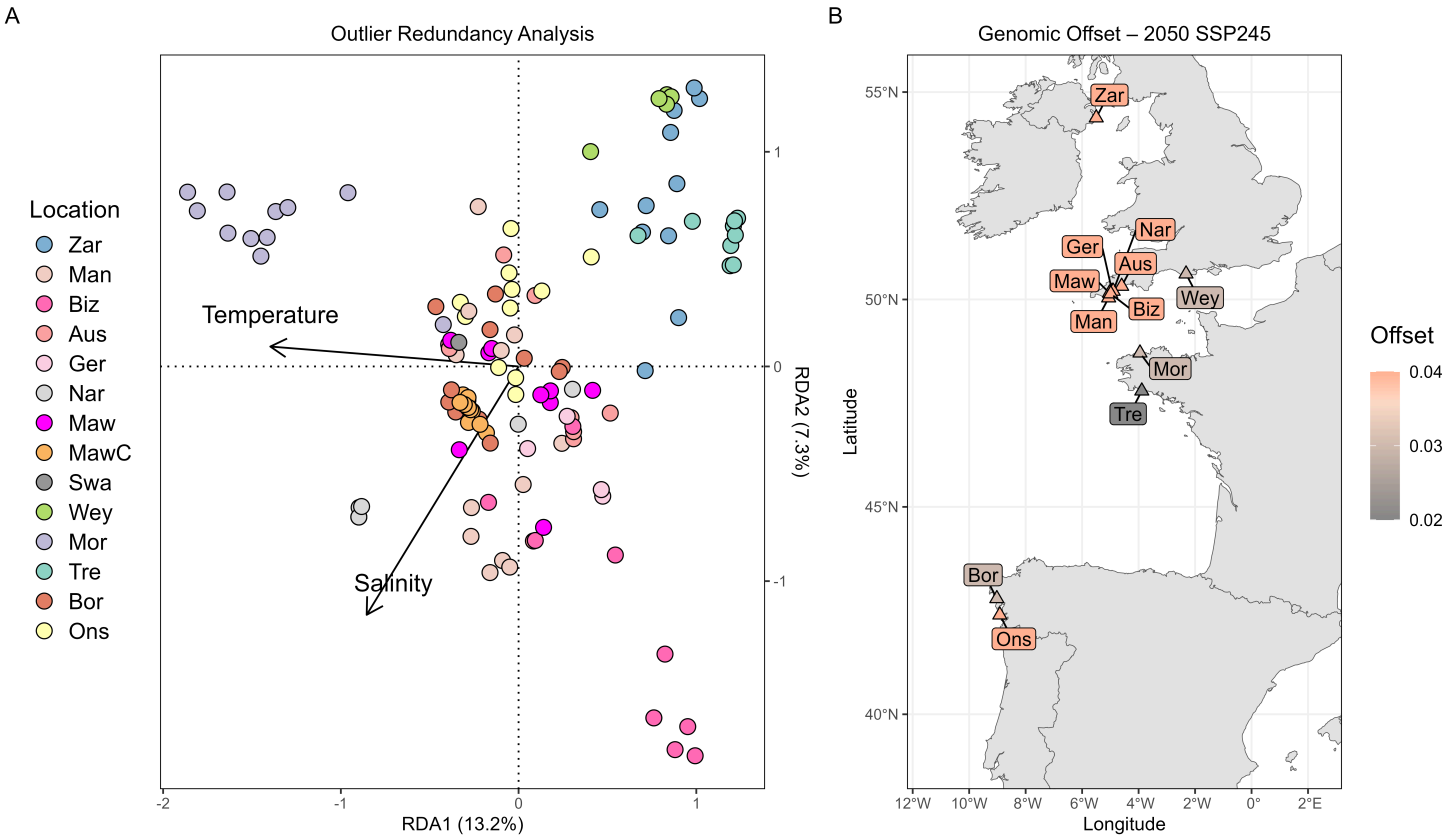

Figure 6

Phymatolithon calcareum outlier redundancy analysis and genomic offset map. (A) Redundancy analysis using allele frequencies of the outlier loci, and temperature and salinity as predictor variables. Each point represents an individual, and colours denote the location of origin. (B) Genomic offset map using allele frequencies of the outlier loci and future climate predictions for temperature and salinity under the shared socioeconomic pathway (SSP) 245 scenario. The offset is a relative measure, meaning that the risk of being maladapted to future climates is only relevant in this analysis (not comparable across studies).

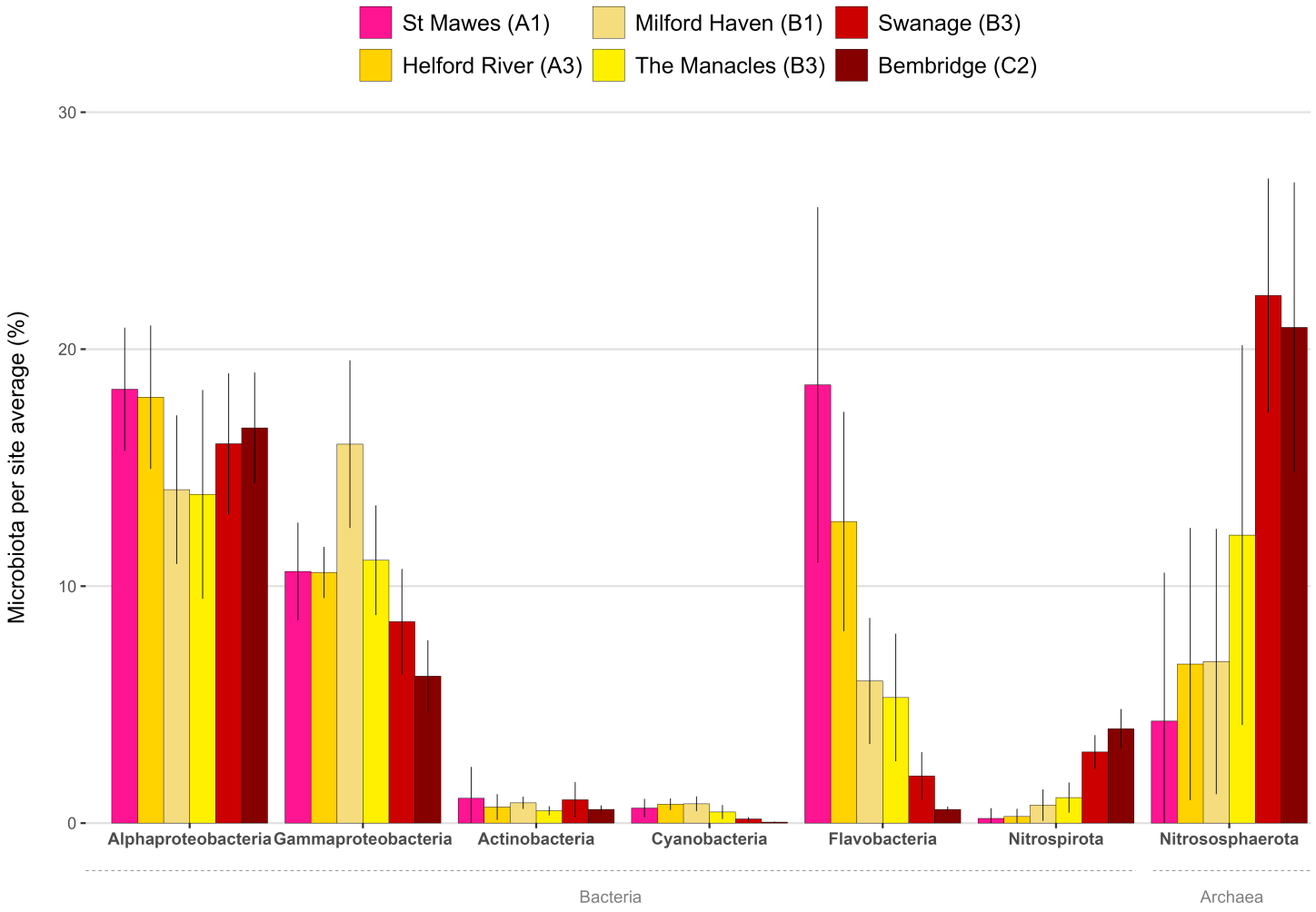

Figure 7

Microbiota community of maerl assessed using metagenomics. The mean average proportion for groups of bacteria or archaea are presented for six locations; these locations broadly represent a gradient of maerl bed status from a dense, healthy maerl bed (St Mawes) to scattered, mostly dead maerl (Bembridge). See main text for classification descriptions in parentheses.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1DNAextractionprotocol.docx
- FigureS2S12.docx
- FigureS13microbiotaproportions.zip
- TableS1BUSCOassemblies.xlsx
- TableS2Assemblystats.docx