Received: 1 April 2020 | Revised: 27 January 2021 | Accepted: 10 February 2021

DOI: 10.1111/mec.15851

SPECIAL ISSUE

MOLECULAR ECOLOGY WILEY

Testing small-scale ecological gradients and intraspecific differentiation for hundreds of kelp forest species using haplotypes from metabarcoding

Peter Shum[1,2] | Stephen R. Palumbi[2]

[1]School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK

[2]Hopkins Marine Station, Department of Biology, Stanford University, Pacific Grove, CA, USA

Correspondence
Peter Shum, School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK.
Email: p.shum@ljmu.ac.uk

Abstract

DNA metabarcoding has been increasingly used to detail distributions of hundreds of species. Most analyses focus on creating molecular operational taxonomic units (MOTUs) from complex mixtures of DNA sequences, but much less common is use of the sequence diversity within these MOTUs. Here we use the diversity of COI haplotypes within MOTUs from a California kelp forest to infer patterns of population abundance, dispersal and population history from 527 species of animals and algae from 106 samples of benthic habitats in Monterey Bay. Using haplotypes as a unit we show fine-grained differences of abundance across locations for 15 species, and marked aggregation from sample to sample for most of the common species of plants and animals. Previous analyses could not distinguish these patterns from artefacts of amplification or sequence bias. Our haplotype data also reveal strong population genetic differentiation over small spatial scales for 48 species of red algae, sponges and Bryozoa. Last, phylogenetic analysis of mismatch frequencies among haplotypes show a wide variety of demographic histories from recent expansions to long, stable population sizes. These analyses show that abundant, small-bodied marine species that are often overlooked in ecological surveys can have strikingly different patterns of ecological and genetic structure leading to population, ecological and perhaps adaptive differences between habitats. MOTU diversity data from the same sequencing efforts that generate species-level analyses can greatly increase the scope and value of metabarcoding studies.

KEYWORDS
COI, demographic reconstruction, DNA metabarcoding, haplotype diversity, kelp forest, Tajima's *D*

1 | INTRODUCTION

In marine ecosystems, regime shifts triggered by ocean warming and acidification are putting increasing pressure on habitat-forming species, such as corals and large seaweeds, resulting in less complex and less productive habitats (Arafeh-Dalmau et al., 2020). Monitoring biodiversity has been crucial in clarifying the ways ecological disturbances, such as abrupt heat waves or shifts in grazing or predation pressure, affect the geographical distribution of habitat-forming species (Melis et al., 2019). The ability to monitor changes in community

composition across climate gradients at multiple scales is fundamental to understanding the biogeographical response to climate change (Rogers-Bennett & Catton, 2019).

Monitoring programmes harnessing DNA have revolutionized our understanding of biodiversity and our capacity to manage species and ecosystems (Vernooy et al., 2010). One notable application that quickly gained momentum in an era of next-generation sequencing is DNA metabarcoding. DNA metabarcoding is now a well-established method aimed to generate biodiversity data to detect environmental change efficiently and rapidly (Bush et al., 2019). Many examples characterize diversity to examine the relative contribution of species in a sample and to identify trends in species richness among communities (Sales et al., 2020; Shum et al., 2019; Wangensteen et al., 2018), including recent methodological efforts in estimating abundance and biomass (Di Muri et al., 2020). Despite the growth of metabarcoding as a tool for diversity measurement at the species level or above, there is a source of diversity data in metabarcoding studies that is typically ignored: intraspecific diversity.

There is mounting interest to consider the potential of metabarcoding data sets that could unlock population-specific patterns included in intraspecific biodiversity (Azarian et al., 2020; Sigsgaard et al., 2020). This allows the simultaneous study of intraspecific diversity and phylogeographical patterns for hundreds of species, referred to as the study of "meta-phylogeography" (Turon et al., 2020). Early contributions revealed the potential of haplotype variability in environmental water samples through targeted molecular approaches (Parsons et al., 2018; Sigsgaard et al., 2016), but identifying haplotypes from DNA metabarcoding data of whole communities is a major challenge due to bias in polymerase chain reaction (PCR) amplification and sequencing error. Elbrecht, Vamos, et al., (2018) recognized this challenge and developed a DNA metabarcoding–haplotyping strategy to extract haplotype-level information from metabarcoding data sets. Their approach involved denoising sequences that identify exact amplicon sequence variants (i.e., haplotypes) which are then grouped into MOTUs (molecular operational taxonomic units) to form species-level information, preserving the sequence variation contained in the haplotypes. They validated this approach using a controlled single-species mock community with 15 known haplotypes and used a denoising approach (Edgar, 2016) to then extract intraspecific genetic diversity. In a later study, Turon et al. (2020) realized the paucity in ground truth or mock community data available for many metabarcoding data sets and developed a strategy to adapt existing data sets amenable to phylogeographical studies by separating erroneous sequences from true intra-MOTU variation. They applied their approach to an existing cytochrome oxidase I (COI)-based metabarcoding survey of biodiversity of sublittoral marine benthic communities and found instances of higher haplotype diversity for species than compared to traditional population genetic data collected in similar locations. While these studies show promise in establishing population structure for many species, our ability to reconstruct the past demographic history of species is required in order to strengthen species- and community-level inferences.

Intraspecific data could be a powerful addition to metabarcoding data sets for at least three reasons. First, intraspecific haplotype diversity within a single environmental sample implies there were multiple individuals in that sample from that species. This can provide a more sensitive assay of species abundance than simple presence/absence. Second, intraspecific haplotypes that differ from sample to sample can signal population structure and provide an estimate of dispersal barriers. Third, the distribution of haplotype differences within species has been used to estimate demographic parameters, long-term population sizes and demographic shifts within species. Metabarcoding approaches to community DNA samples can thus provide a wealth of information on the ecology and history of populations if haplotype diversity data can be reliably obtained.

Kelp forests are one of the most diverse ecosystems on the planet and harbour an astounding diversity of invertebrates, fish, algae and other taxa. Large-scale monitoring through traditional methods have produced detailed taxonomic inventories through presence/absence lists and ecological estimates (Port et al., 2016). For example, cobbles with attached crustose coralline algae are routinely used as a standard sampling unit to survey the presence of larval settlement for a wide range of species (Hart et al., 2020; Rogers-Bennett et al., 2016). This type of sample has been recently incorporated into metabarcoding studies as a Wild Reef Monitoring System for biodiversity assessments (Shum et al., 2019) and algae (Akita et al., 2019). The combination of bulk sampling and next-generation sequencing allows a fast-paced sampling of hundreds of species including common charismatic and rare elusive species. Therefore, metabarcoding-based haplotyping may provide vital information to effectively monitor kelp forests affected by environmental changes.

Here, we provide a re-analysis of DNA metabarcoding data (see Shum et al., 2019) to extract haplotype information from cobble communities found within kelp forest ecosystems. Cobbles are a major substrate for a range of benthic animals and algae and offer a unique opportunity to examine intraspecific patterns in a highly diverse ecosystem. Therefore, we aim to (i) validate haplotypes by examining denoising parameter values of alpha (α) and compare patterns of DNA substitution between empirical and DNA metabarcoding haplotype data, (ii) assess the haplotype genetic diversity across different taxonomic groups, (iii) examine population genetic structure compared to traditional data sets and (iv) assess demographic patterns for hundreds of co-inhabiting species that may offer clues about the status of the ecosystem.

2 | MATERIALS AND METHODS

2.1 | Description of kelp forest cobble metabarcoding data

We retrieved the COI raw Illumina reads (SRA database under BioSample accession no. SAMN12220032) generated from a DNA metabarcoding biodiversity study of cobble communities (Shum

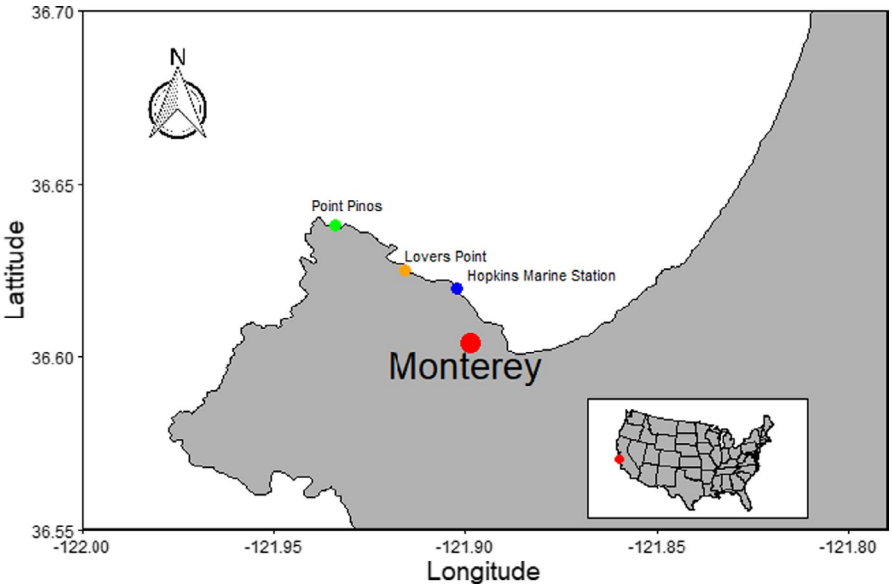

FIGURE 1 Cobble sampling sites within kelp forests along the southern Monterey Peninsula, California. Inset: sampling location along the west coast of California in the USA

et al., 2019). This study sampled a total of 106 medium-sized cobbles (67–3278 cm^3) collected in 2012 as part of a kelp forest monitoring study along a 3-km stretch of coastline in southern Monterey Peninsula at three locations: Point Pinos (Pt. Pinos, 36°38′16.99″N, 121°56′2.04″W, *n* = 24), Lovers Point (LovP, 36°37′30.30″N, 121°54′56.07″W, *n* = 42) and Hopkins Marine Life Refuge (HMS, 36°37′12.44″N, 121°54′8.28″W, *n* = 40) (Figure 1). In brief, each cobble was placed in two polythene ziplock bags (1-gallon) with easy zip closure underwater using SCUBA inside areas of dense kelp forest. Cobbles were gently transported to the surface and samples processed on shore within 3 hr of collection. Processing involved gently scrubbing each cobble using a clean soft brush and filtering contents onto an 80-μm mesh sieve. The sieved contents of each sample were rinsed with 70% ethanol and washed back into glass jars containing 95% ethanol and stored at −5°C until further processing. The equipment used during sampling was washed thoroughly under high-pressure filtered seawater between each sample. This is a routine sampling method targeting sessile and mobile species attached to the cobble (Hart et al., 2020; Rogers-Bennett et al., 2016; Shum et al., 2019), and although seawater was processed, a source of environmental DNA (i.e., water eDNA) such as cell sloughing would present a negligible proportion of the community after sieving (see Curtis et al., 2020 and Kumar et al., 2020 for eDNA best practices).

Each sample was filtered onto sterile 0.45-μm cellulose nitrate membrane filters (47 mm diameter; Whatman, GE Healthcare) using 47-mm magnetic filtration units (Pall Corporation) and samples were rinsed with purified water to remove residual ethanol. Filtration blanks consisting of purified water were processed before the first cobble sample and subsequently every 25th sample. Samples were briefly homogenized by pestle and mortar and up to 10 g of collected material including the filter was DNA extracted using the PowerMax Soil DNA Isolation Kit following the manufacturer's instructions. Cobble samples, filtration blanks, a PCR-negative and one individual positive control (queen conch, *Strombus gigas*) were amplified using a versatile primer set (COI, mICOIintF: GGWACWGGWTGAACWGTWTAYCCYCC; Leray et al., 2013, matched to jgHCO2198: TAIACYTCIGGRTGICCRAARAAYCA; Geller et al., 2013), targeting broadly metazoan phyla and algae. A DNA extract of the positive control individual was obtained from a previous pioneering single nucleotide polymorphism (SNP) genotyping study of queen conch (Natesh et al., 2019). A 1:100 dilution of the PCR product was added to the pooled PCR for library preparation. PCR amplification followed a two-step protocol by first amplifying samples with nonindexed primers, followed by a second PCR with indexed primers containing 8 bp oligo tags differing in at least three bases and a variable number of degenerate bases "N" (two, three or four) to increase the variability of amplicon sequences during sequencing. Illumina libraries were prepared using the Nextflex PCR-free library prep kit and sequenced on an Illumina MiSeq using version 3 chemistry (2 × 300 bp) with a final molarity of 12 pM and 12.5% PhiX (GNomEx, University of Utah). Further details of cobble sampling, PCR conditions, library preparation and data sets are outlined in Shum et al. (2019).

2.2 | Description of empirical population genetic data sets for comparison

To characterize DNA metabarcoding haplotype diversity to traditional data sets, we used the empirical haplotype abundance data from Kelly and Palumbi (2010) as a comparative analysis to our DNA metabarcoding haplotype data set. Their data examined the COI region in 50 species of marine invertebrates collected along the Pacific coast of North America, and we used data for which COI sequences were generated from species collected from Monterey (16 invertebrate species).

2.3 | Generation of haplotypes

Quality control and filtering of Illumina reads were done using OBITOOLS (Boyer et al., 2016) and generation of haplotypes per MOTU

for each cobble sample was implemented in JAMP (Just Another Metabarcoding Pipeline; Elbrecht, Vamos, et al., 2018). JAMP is a modular R package that implements USEARCH (Edgar, 2013), VSEARCH (Rognes et al., 2016) and CUTADAPT (Martin, 2011) for data processing and applies a denoising method (unoise3; Edgar, 2016) that involves error-correction on amplicon sequences to recover correct sequences by resolving distinct sequences down to a single difference.

FASTQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) was used to check the quality of the raw Illumina reads and *obicut* (OBITOOLS; Boyer et al., 2016) was used to trim reads with a minimum quality threshold of 28. Pair-end reads were aligned using *illuminapairend*, and alignments with a quality score <40 were discarded. The aligned data set was demultiplexed using *ngsfilter*. Reads were strictly filtered for length (313 bp, *obigrep*) and reads containing ambiguous bases, "N", were removed. The data set at this point contains quality controlled and filtered reads assigned to cobble sample names in a single file. To provide JAMP with the correct format for haplotyping, the data were parsed into separate files according to sample name, after which file names were appended. *Denoise()* (JAMP), used to extract haplotypes, performs dereplication (*usearch*), denoising (*unoise3*) and clustering (*usearch*) before generating the final table. *unoise3* extracts haplotype-level sequence information across the entire data set. Haplotypes are then clustered into MOTUs using default parameters (haplotypes within 3% similarity). *Denoise(minsize=1*, *minrelsize=1e-13*, *OTUmin=1e-13*, *minHaploPresence=1)* was used to extract haplotype sequence-level information with effectively no abundance filtering (*1e-13*) that would otherwise remove rare and/or low-abundant haplotypes. This returns the largest number of haplotypes while potentially maintaining false positive artificial sequences in the data set.

Taxonomic assignment of haplotypes was done by generating a reference COI database using an *in silico* PCR (*ecoPCR*, Ficetola et al., 2010) on the R134 release of the EMBL-EBI database. Taxonomic assignment was performed for each haplotype using default parameters in *ecotag* (OBITOOLS), which uses a phylogenetic framework to assign MOTUs to the most conservative lowest common ancestor (Boyer et al., 2016). However, due to the potential taxonomic bias of *in silico* PCR methods (see Clarke et al., 2014), further manual curation of assignments was carried against NCBI's nonredundant database by (i) retrieving the top BLAST hit for each haplotype, and (ii) comparing the percentage identity and taxonomic IDs between the *ecotag* and NCBI taxonomic assignments. The NCBI assignments with 99%–100% species identity was considered if the *ecotag* assignment reported a lower species identity and (iii) the haplotype sequence for each taxonomic discrepancy was assessed by a manual BLAST search (BLASTN) on GenBank and corrected if no overlapping species hits were present.

We used *decontam* version 1.8.0 (Davis et al., 2018) in R (R Core Team, 2020) to statistically evaluate and remove contaminated haplotypes from the data set based on the proportional abundance of reads using the "Prevalence" model (threshold of 0.5).

2.4 | Generation of population genetic data sets and demographic statistics

A classic population genetic data set consists of collecting individuals of a species within their natural geographical range and generating genetic data through Sanger-sequencing that helps to reveal the haplotype variation within a population and demographic genetic structure of a species. The final JAMP haplotype data set contains MOTUs (i.e., species or taxonomic group), each with *n* number of haplotypes as well as read abundance information. However, this approach cannot accurately distinguish between the number of individuals detected in bulk DNA-extracted samples using read abundance. Therefore, to format the haplotype data set to mimic data sets for traditional population genetic analysis, we used one read as a threshold to determine the presence of a haplotype found on a cobble. This approach will identify at least one individual carrying a haplotype, and the resolution to count the number of potentially valid identical haplotypes from multiple individuals present on a cobble is lost. We used the number of cobbles on which that haplotype occurred to represent the number of individuals having that haplotype. For example, if MOTU *q* contains haplotype_1 occurring on 15 of 106 cobbles and haplotype_2 occurring on 55 cobbles, we represent the whole population as 15 identical sequences of haplotype_1 plus 55 sequences of haplotype_2. The sequences for MOTU *q* are then parsed into a separate fasta-formatted file consisting of 70 individuals that consist of two haplotypes. We generated haplotype frequency spectra for hundreds of MOTUs for our entire data, and for each of the three geographically separate sampling localities. This resulted in 2040 population genetic data sets providing a suite of molecular diversity indices, including nucleotide (π) (Nei, 1987), haplotype (*h*) (Nei & Tajima, 1981), F_{ST} (Hudson et al., 1992) and Tajima's *D* (Tajima, 1983) diversities estimated for each data set using DNASP version 5.10 (Librado & Rozas, 2009).

2.5 | Validation of haplotypes

To test the ability of our pipeline to distinguish DNA sequencing errors without removing valid, rare sequences, we applied the *Denoise()* procedure implemented in JAMP to COI data from our single-haplotype individual positive control (*Strombus gigas*) and varied the alpha (α) parameter from 1 to 10. Higher values of the α parameter accept a larger fraction of rare sequences as being real. Lower values reject more of these rare sequences as artefacts, but this higher filtering of potentially incorrect sequences comes at the cost of ignoring some real, rare sequences. We assessed the influence of α values on the pattern of DNA sequence variation by examining the relationship between (i) the absolute number of transitions (Ts) and transversions (Tv) plotted against the sequence divergence under the GTR model (DAMBE version 7.2.136; Xia & Xie, 2001) to analyse nucleotide mutation saturation of COI, and (ii) the number of transitions, transversions, SNPs and amino acid (AA) substitutions against the number of haplotypes for *Strombus*

gigas. These comparisons were also carried out on the 16 empirical population genetic haplotype data sets from Kelly and Palumbi (2010) and our selected haplotype data set (referred to as haploCOBB) to further compare the pattern of denoised *Strombus gigas* data sets. The number of transitions, transversions, SNPs and AA substitutions were calculated in *strataG* (Archer et al., 2016) and *seqinR* (Charif & Lobry, 2007) in R. Nucleotide sequences were translated to amino acid sequences by *transeq* (EMBOSS, http://emboss.sourceforge.net, Rice et al., 2000), using the genetic code indicated from NCBI's taxonomy browser.

2.5.1 | Meta-phylogeographical analysis

We used COI DNA sequences generated by Sanger sequencing for the red abalone (*Haliotis rufescens*), the pink volcano barnacle (*Tetraclita rubescens*), the Pacific purple sea urchin (*Strongylocentrotus purpuratus*) and the bat star (*Patiria miniata*) from GenBank and compared the phylogeographical relationships with the haploCOBB data set. Extensive population genetic studies on *H. rufescens* and *T. rubescens* have evaluated the connectivity among populations over extensive geographical regions (Dawson et al., 2014; Gruenthal et al., 2007; Hart et al., 2020) and resulted in a large collection of representative haplotypes. In contrast, the only adequate COI sequences currently available for *Strongylocentrotus purpuratus* and *P. miniata* are limited to voucher sequences from a DNA barcode reference library for Canadian echinoderms (Layton et al., 2016). These data sets are used to assess the level of shared and new COI haplotypes that can be revealed through DNA metabarcoding. The GenBank and haploCOBB sequences were aligned using MAFFT version 7.407 (Katoh & Standley, 2013), as implemented in GENEIOUS version 7.1.3 (www.geneious.com) and sequences trimmed to the haploCOBB COI fragment length (313 bp). Haplotype genealogies were constructed in the program HAPVIEW (Salzburger et al., 2011), based on a maximum-likelihood tree implemented in PHYLIP version 3.695 (Felsenstein, 1993).

2.6 | Population genetic analysis

To measure the occupancy of haplotypes (i.e., the number of cobbles occupied by a haplotype per MOTU), we used the presence/absence metabarcoding of haplotypes (PAMH) to examine the relationship between the number of haplotypes with its associated occupancy by calculating the number of haplotypes present for each MOTU per cobble and then obtaining the total number of haplotypes present across all cobbles.

2.6.1 | Ecological aggregation

We used PAMH to assay aggregation of haplotypes across sampling sites using the simple dispersion index (I = variance/mean), where the mean and variance are calculated over the total number of unique haplotypes per MOTU across samples. Values above 1 indicate that individuals are "clumped" within samples more than in random samples (although some nonrandom patterns of abundance can have a dispersion index of 1).

2.6.2 | Population structure

The presence/absence data for each haplotype within a MOTU for each cobble generates a data set that allows regional comparison of haplotype frequencies. We recorded the number of cobbles on which each of the haplotypes occurred within the Point Pinos, Lovers Point and Hopkins Marine Station data sets. We used these to estimate haplotype frequencies and mitochondrial SNPs for each MOTU (denoted mt_{SNP}) at these locations, and estimated F_{ST} (as $[mt_{SNP}(between)\text{-}mt_{SNP}(within)]/mt_{SNP}(between)$) for the 176 species in our analysis with more than one haplotype and at least five occurrences in each of the three habitats.

2.7 | Tajima's *D* and mismatch distributions

Phylogenetic trees of neutral genes within populations are constructed by a balance between substitutions and random lineage extinction. An extensive theoretical framework shows that the overall variation in such trees depends on effective population size and substitution rate, but that the shape of these trees also depends on changes in population size over time (Felsenstein, 1992). Bottlenecks trim branches from trees, especially recently diverged ones, so these trees tend to have deeper branches on average. By contrast, population expansion generates many new branches so these trees tend to have shallower branches on average. Tajima's *D* neutrality statistic captures this measure by relating the number of variable nucleotides in the tree to the average nucleotide diversity (Tajima, 1983). To test whether the observed amount of genetic diversity deviates from neutral expectations, we estimated average Tajima's *D* (Tajima, 1983) for each MOTU, considering haplotypes found across sampling sites (not considering MOTUs with significant F_{ST} [n =15]). However, because Tajima's *D* statistics are sensitive to the abundance of each haplotype (Tajima, 1983), the data set was filtered to examine the influence of rare or false positive haplotypes on Tajima's *D*. To achieve this, we compared the haploCOBB data set to two treatments. First, in Filter1, haplotypes that were seen in the data set with only one read were removed. Second, in Filter2, haplotypes that occur on only one cobble, regardless of read abundance, were removed. Diversity statistics including Tajima's *D* were recalculated for the haploCOBB and the two filtered data sets as described above.

2.7.1 | Mismatch distributions

Mismatch distribution analysis is used to evaluate the frequency distribution of pairwise differences between sequences. Large, stable

populations tend to show bimodal mismatch distributions, whereas expanding populations tend to show a prevalence of individuals with identical haplotypes with zero sequence mismatches (Rogers & Harpending, 1992). For example, Marko et al. (2010) generated mismatch distributions for intertidal invertebrates along the northeastern Pacific shore and documented two patterns. One set of species had several dominant haplotypes, and so mismatch distributions showed most comparisons having one or more sequence differences. These species were shown by simulations to probably have been stable or expanded over hundreds of thousands of years. By contrast, about half the species in Marko et al. (2010) showed one strongly dominant haplotype, with the result that the most common comparison among individuals was zero sequence difference. These populations were shown by simulations to have probably expanded since the last glacial maximum 18,000 years ago (Marko et al., 2010). To examine the demographic history of MOTUs, mismatch analysis was performed using arlequin version 3.5.2.2 (Excoffier & Lischer, 2010).

3 | RESULTS

3.1 | Haplotype validation and summary

Haplotype sequences from our community DNA survey of kelp forest cobbles provide estimates of genetic diversity for hundreds of species. A total of 25,524,732 Illumina paired-end reads were analysed; the OBITOOLS filtering procedure for quality and length resulted in 9,842,197 sequences. Following the denoising pipeline in JAMP, we varied the alpha (α) parameter from 1 to 10, where 1 represents deletion of the most low-frequency haplotypes and 10 represents deletion of the least. We applied this variety of parameters to sequences of a positive control (*Strombus gigas*), and found that an alpha value of 5 was the least stringent filtering that still produced the expected single haplotype: values of 6 or more produced multiple haplotypes from our control. The *Strombus gigas* data with higher levels of alpha have large numbers of SNPs, high percentage transversion changes, and much larger numbers of amino acid changes compared with empirical

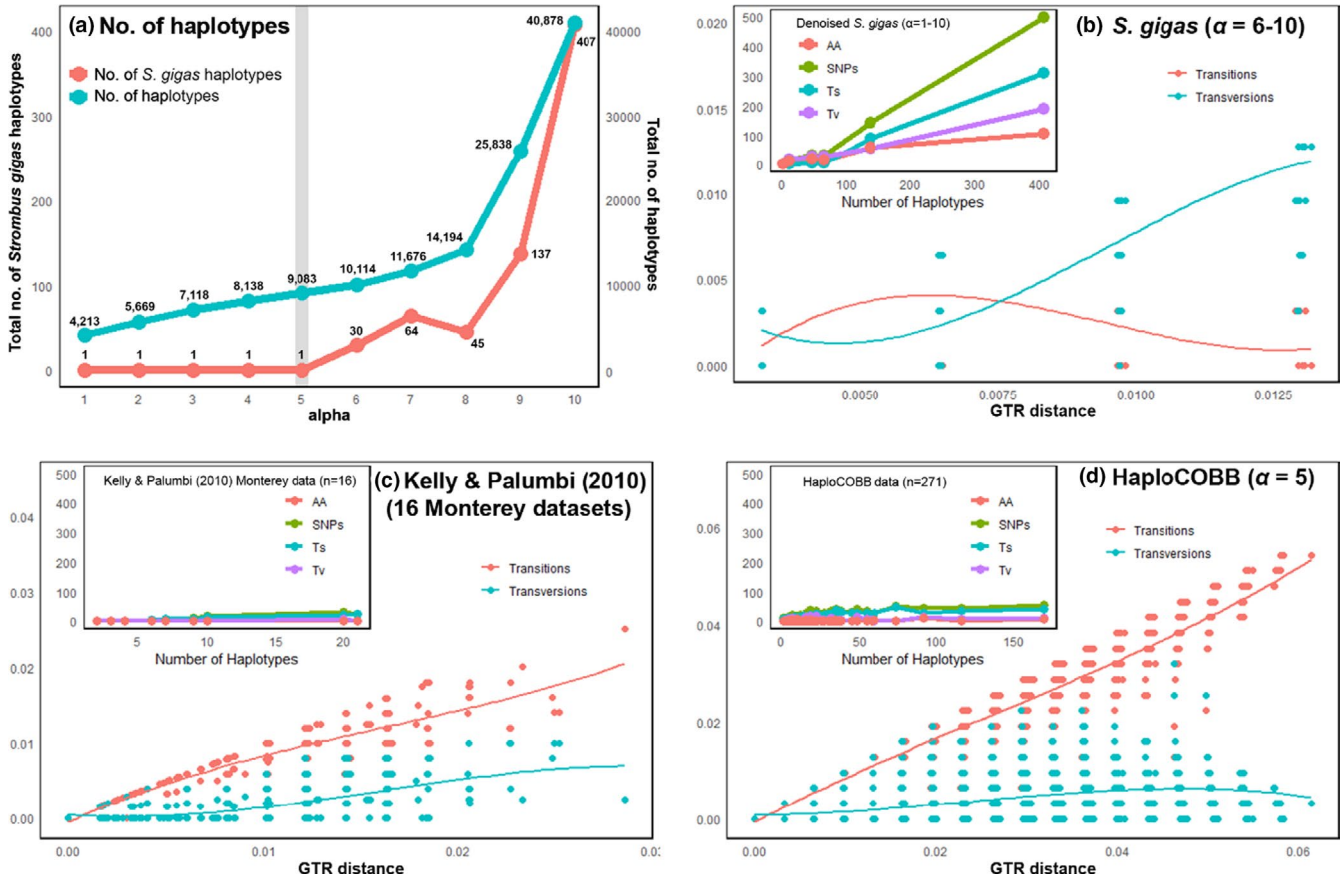

FIGURE 2 Validation of our DNA metabarcoding haplotype data set (haploCOBB). (a) Variation in the number of *Strombus gigas* (positive control) haplotypes and total number of haplotypes recovered with increasing values of alpha (1–10). Grey line highlights the data set selected (haploCOBB) for further analysis, indicating an alpha set to 5 recovered the expected one haplotype for the positive control. (b–d) The relationship between the number of transitions and transversions (*y*-axis) plotted against the pairwise sequence divergence (GTR distance) to demonstrate the level of substitution saturation of the COI sequences. These comparisons show the combined analysis for multiple population genetic data sets. Inset: the number of transitions (Ts), transversions (Tv), single nucleotide polymorphisms (SNPs) and amino acid substitutions (AA) against the number of haplotypes for each data set. The *y*-axis shows an identical scale for comparison. (b) five haplotype data sets of *Strombus gigas* with alpha of 6–10. (c) Empirical population genetic invertebrate haplotype data sets from Monterey (*n* = 16, Kelly & Palumbi, 2010). (d) Haplotype data set (haploCOBB) with alpha set to 5 (*n* = 271 MOTU data sets)

data from Kelly and Palumbi (2010) and the haploCOBB data set (Figure 2). Therefore, the substitution patterns we record from data sets generated by low filtering at high α levels are not consistent with the kinds of substitutions we observe in real population-level data. As a consequence, we used α = 5 for all further analysis (Figure 2a).

Our survey uncovered 9084 COI haplotypes from 2040 unique MOTUs including algae, single-celled eukaryotes and metazoans. Of these MOTUs, 1255 could not be identified to class or phylum, 201 were single-celled eukaryotes or fungi, and 19 were in metazoan or algae groups with low representation (five or fewer MOTUs). After this filtering, there were 527 MOTUs and 2586 COI haplotypes from 19 algal or metazoan classes that we further analysed for patterns of diversity (Table S1). To construct our population sample, we recorded the presence/absence of each haplotype on each cobble. Across all cobbles and MOTUs, we recorded the presence of unique haplotypes, totalling 4559 observations.

MOTUs had a range of haplotype numbers from 1 to 169 (Figure S1a). Haplotype sequence diversity within MOTUs ranged from 0 to 2.8%, with a mean of 0.002% (0.0055% for nonzero MOTUs, Figure S1b). Both of these estimates of genetic diversity increased with the number of cobbles on which a MOTU was found, though there was considerable variation, especially in overall nucleotide diversity (pi; Figure S1c,d).

3.2 | Patterns of diversity within classes

There are wide differences among classes in the distribution of genetic diversity among MOTUs. For example, Anthozoans, Asteroids and Enopla nemerteans show low sequence diversity (Figure 3), and a large number of MOTUs with only one haplotype (see the leftmost histogram in Figure 3). By contrast, a large set of classes show intermediate diversity levels between 0.2% and 0.8% in one third to two thirds of their MOTUs. This group includes MOTUs in the Ascidians, Pycnogonids, Bivalves, Florideophyceae (red algae), Gastropods, Holothurians, Malacostracans, Phaeophyceae (brown algae), Polycheates, Oligichaetes and Polyplacaphorans (chitons; Figure 3). Among this set, there is a broad difference between animals and algae: Metazoan classes average 0.28% nucleotide diversity whereas the two alga groups Florideophyceae and Phaeophyceae show 0.18% and 0.09% respectively. A third group of MOTUs show very high genetic diversity of 1%–2%. This group is dominated by small arthropods (Maxillopods and Malacostracans), but also includes ophiuroids, chitons (Polyplacophorans), polychaetes and bivalves (Figure 3).

3.3 | Low-diversity MOTUs

A total of 259 MOTUs only had one haplotype. In general, these MOTUs were uncommon in our sample, occurring on average on only five cobbles. However, there was a set of exceptions, where common MOTUs had low genetic diversity. For example, a single haplotype of the giant kelp *Macrocystis pyrifera* occurred across all 106 cobbles, with one other haplotype occurring on seven. The two major algal taxa (brown algae, Phaeophyceae; and red algae, Floridiophyceae) show generally low average nucleotide diversity. Similarly, among the five Anthozoa species in our survey, none has more than one haplotype even though these MOTUs were found on an average of 27 cobbles each (range 4–51). One possible reason for

FIGURE 3 Distribution of nucleotide diversity (pi) among MOTUs from 19 classes of invertebrates and algae from kelp forest cobbles

low diversity despite high abundance is a low nucleotide substitution rate across the 313 bases of the conserved COI gene in the region we sequenced. In the Anthozoa, low rates of mitochondrial evolution or high rates of DNA repair led to low sequence evolutionary rates (Shearer et al., 2002). One additional reflection of this low divergence rate is that mitochondrial differentiation among Anthozoan species or even genera is often as low as 1%–2%. In our data set, different genera differed by as little as a single base.

However, other explanations for extremely low diversity are likely in other classes. For example, the sea star *Pycnopodia helianthoides* occurred on 28 cobbles with only one haplotype. Similarly, two other single-haplotype sea stars were common in the data set: *Pisaster giganteus* (on 27 cobbles) and *Pisaster ochraceous* (on 15 cobbles). Across all our seven sea star MOTUs, six showed one haplotype. Only *Patiria miniata*, on 52 cobbles and one of the most abundant invertebrates in kelp forest surveys (Shum et al., 2019), had substantial genetic diversity with nine haplotypes and 0.44% nucleotide diversity. There is no evidence that Echinoderms or Asteroids have low rates of mitochondrial evolution (Britten et al., 1978). Abundant MOTUs with low diversity may reflect populations that have grown from smaller sizes in the recent past (see bottleneck and expansion analysis below).

3.4 | High-diversity MOTUs

On the opposite side of the diversity spectrum are MOTUs with very high molecular diversity. Among high-diversity MOTUs with good class-level identification, the largest (MOTU_55 with 169 haplotypes) is identified as a caprellid amphipod with 82% identity. A nucleotide phylogeny shows five deep clades ~4% divergent (data not shown). Such a large number of haplotypes might be due to amplification of pseudogenes, sequencing errors or other artefacts of the community DNA pipeline. To test this, we translated the sequences from all 169 haplotypes: the amino acid comparison showed nine amino acid substitutions among a small number of sequences distributed across nine small clades. There is also a TTT→TTA mutation shared by six individuals that might be processed by amphipod mitochondria as a stop codon. Overall patterns of substitution follow typical patterns seen in functional COI: a dominance of transition substitutions and low numbers of amino acid substitutions (see Figure 2). These data strongly suggest that these high-diversity haplotypes are in fact COI sequences from active genomes rather than sequence errors or pseudogenes.

The next highest diversity MOTUs are within the ophiuroids, with 1.6% and 1.9% sequence diversity within them, respectively. MOTU_226 with 116 haplotypes shows, when translated, only three amino acid changes, no stop codons and no insertions/deletions, all suggesting that these are valid mitochondrial sequences evolving under natural selection. Likewise, the high-diversity ophiuroid MOTU_57 with 30 haplotypes shows no amino acid changes. In this case, the MOTU shows two large clades that have about 3% sequence divergence (Figure 3).

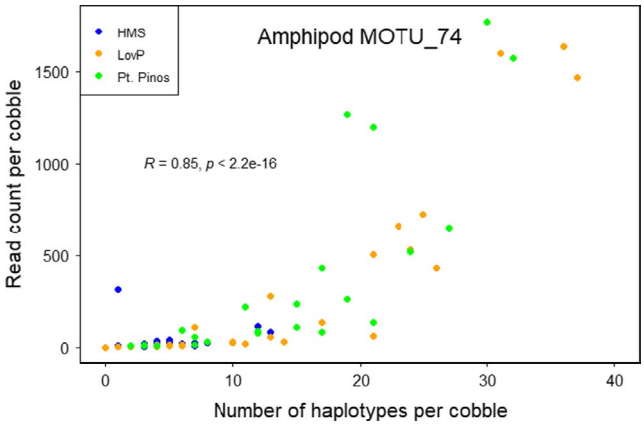

FIGURE 4 An example of the strong positive relationship between read count of a haplotype on a cobble and the number of haplotypes on that cobble. MOTU_74 is identified as a Cyamid amphipod with 60 haplotypes across 88 cobbles. Some cobbles have as many as 36 haplotypes from this single MOTU, and up to 1800 total reads. Pearson correlation is presented and colours represent haplotypes found at each location for Hopkins Marine Station (HMS), Lovers Point (LovP) and Point Poins (Pt. Pinos)

3.5 | Ecological abundance estimated using haplotype occurrence

Overall, the 20 highest diversity MOTUs were dominated by amphipods (eight of 20), with ophiuroids, polychaeta worms, small bivalves, gastropods and bryozoans making up the remainder. These high-diversity MOTUs tended to occur on many cobbles, and have many haplotypes per cobble, implying high population sizes. Population size has been difficult to measure using standard community and eDNA approaches because sequence read depth of a MOTU from a sample is subject to PCR bias, library variation and variation in body size. A better approach has been to use the number of samples in which a MOTU occurs as an index of its overall abundance across all samples (Presence Absence Metabarcoding, Shum et al., 2019). Identifying multiple haplotypes in a MOTU allows us to extend this latter approach to each haplotype as a separate entity, and use the number of haplotypes of a MOTU in a sample as a proxy for the population size of that MOTU in each sample (PAMH, Figure 4). Using this approach suggests that amphipod MOTU_74 is more abundant at Point Pinos where it has 14.1 haplotypes per cobble (*SD =8*) and Lovers Point (10.4 haplotypes per cobble, *SD =10*) than Hopkins (4.1 haplotypes per cobble, *SD =3*, p <.001 *t*-test). Similarly, the southern barnacle *Tetraclita rufescens* has a higher number of haplotypes at Hopkins and Point Pinos than at Lovers Point (average haplotypes per cobble of 2.6, 1.8 and 0.9 respectively, *t*-test p <.0001), and the sea star *Patiria miniata* was more common at Lovers Point (1.3 haplotypes per cobble compared to 0.7 and 0.9 for Hopkins and Point Pinos, p =.0015, *t*-test). By contrast, the red coralline alga *Calliarthron tuberculosum* and the barnacle *Balanus crenatus* were evenly distributed across sampling locations (1.5, 1.4 and 1.9

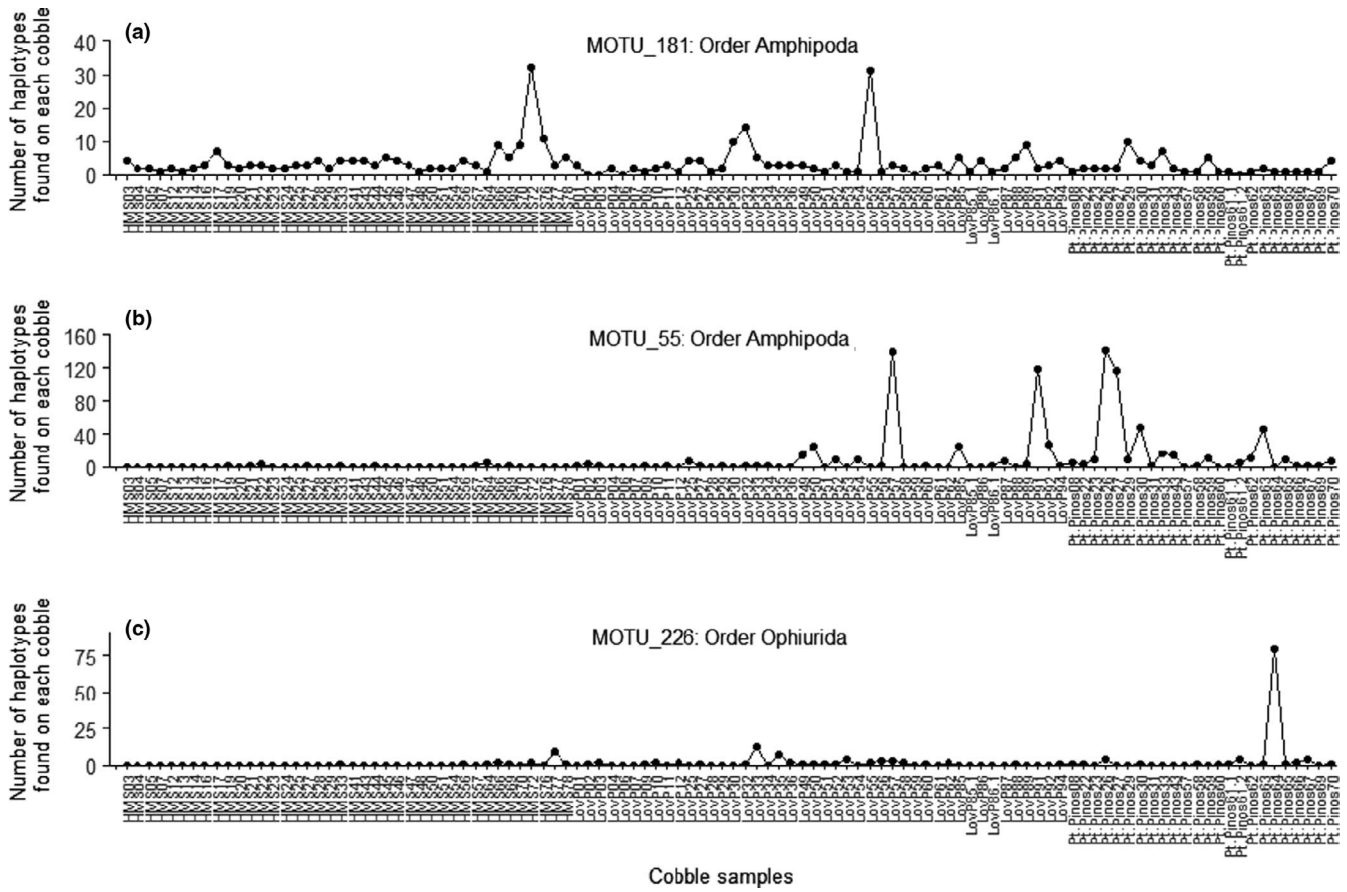

FIGURE 5 The number of haplotypes in three high-diversity, highly abundant MOTUs are clumped among cobbles

haplotypes per cobble, p =.855 t-test for *Calliarthron* and 0.4, 0.2 and 0.4 haplotypes per cobble, p =.26 t-test for *Balanus*).

Overall, there are 15 taxa with significant abundance differences among the three areas we sampled, including eight small crustaceans (amphipods and copepods), bryozoans, ophiuoirds and red algae. In all these cases, emergent distribution differences are more apparent when comparing multiple haplotypes distributed across samples than when comparing only species presence/absence. As a result, when focused on haplotype distribution, PAMH can be a useful tool in establishing patterns of distribution and abundance of species-level MOTUs.

3.6 | Tests of ecological aggregation

A different ecological pattern was seen in many abundant, diverse MOTUs, exemplified by a haplotype analysis of an Aoroides amphipod (MOTU_181, Figure 5a). Here, average haplotype occurrence was similar among locations (p =.5, t-test) but two cobbles in particular showed high numbers of haplotypes whereas others sampled from the same habitat showed none. Cobbles HMS72 and LoversPoint55 contained 32 and 31 haplotypes, respectively (out of 33), and a large number of DNA sequence reads (996 and 1645). These data suggest large populations of this amphipod species on these particular cobbles. Likewise, caprellid amphipod MOTU_55 (Figure 5b) is found on a few samples, but there are over 100 haplotypes recorded on each of four cobbles.

Previously, high read counts for a species in a single sample might have been assumed to be due to artefacts of PCR variation or biomass (Krehenwinkel et al., 2017), and species clustering was therefore hard to prove with community DNA or eDNA data. The PAMH analysis provides higher statistical reliability for tests of clustering because these tests can be done on multiple independent haplotype data sets.

We computed the dispersion index for all of the 527 MOTUs in our sample, based on the sum of the number of unique haplotypes per cobble. Using this approach, clustering is common (Figure S2) and tends to be higher (dispersion index is higher) for more abundant MOTUs: most MOTUs with a sum of haplotypes across cobbles >200 are clumped. The strongest clumping is seen in small animals with many haplotypes, such as amphipods, ophiuroids and an unnamed veneroid bivalve (dispersion index >5). Among the algae, the coralline *Corallina tuberculata* is the most clumped (dispersion index =1.9).

3.7 | Population genetic structure

Most species had low F_{ST} values (median 0.044; Figure S3a), as expected for marine species at locations within 1–3 km of one another (Dawson et al., 2014). However, some abundant MOTUs showed high

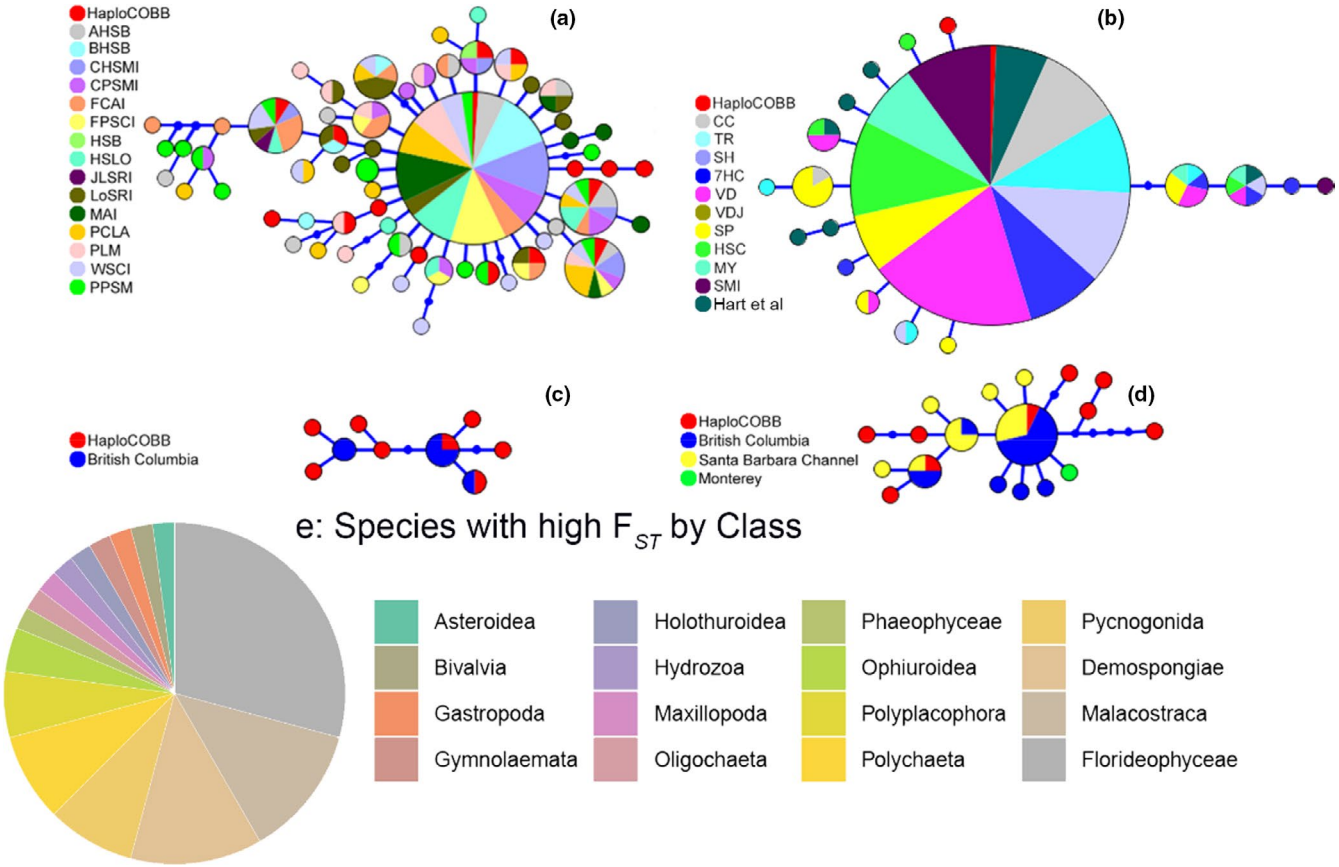

FIGURE 6 Haplotype genealogies visualizing the relationships between DNA sequences. The size of each circle is proportional to the frequency of haplotypes and the lengths of the connecting lines reflect the number of mutations between them. (a) *Tetraclita rubescens*, (b) *Haliotis rufescens*, (c) *Strongylocentrotus purpuratus* and (d) *Patiria miniata*. Labels correspond to different sampling localities (see Dawson et al., 2014 (a), Gruenthal et al. 2007 and Hart et al. 2020 (b) for the corresponding sampling locations). HaploCOBB haplotypes from the present study are shown in red. (e) Distribution of 48 taxa with high levels of F_{ST} (>0.1) among classes

F_{ST} between locations. Overall, 15 MOTUs had F_{ST} >0.20, whereas random location assignments yielded an average of 8.4 MOTUs with F_{ST} values this high. Among the high F_{ST} animals, the sponge *Halichonridia panicea* has brooded larvae (Palumbi, 1984) and showed significant F_{ST} =0.203, driven by 10 separate cobbles with the same COI haplotype at Hopkins (mt_{SNP} =0) compared to mt_{SNP} of around 0.45 for 14–18 samples each from the other locations. Other significant MOTUs included eight different red algae (average F_{ST} =0.301), suggesting that fine-scale population structure is common in these low-dispersal taxa. For example, the highest F_{ST} (0.59) is for a MOTU with 99% identity to a red alga, *Rhodophysema* sp. 2. There are two haplotypes that are four bases different: one is dominant at Lovers Point and Point Pinos, while the other is dominant at Hopkins.

3.8 | Metabarcoding haplotype genealogy compared to empirical haplotypes

Haplotype networks were constructed to test the relationships among DNA metabarcoding haplotypes to those previously reported from traditional summaries of individually amplified and sequenced COI sequences. MOTU_412 contained 17 haplotypes with 99%–100% identity to *T. rubescens* which we compared to 328 COI DNA sequences of *T. rubescens* collected at 15 locations along the Californian coast (Dawson et al., 2014). Their data contain full-length COI sequences (~1300 bp) with 227 haplotypes and reported no clear relationship between sampled sites. Our merged and trimmed data set (Figure 6a) resulted in 59 total haplotypes with six new haplotypes identified: 11 of our 17 haplotypes are shared.

MOTU_538 comprised three haplotypes with 99%–100% sequence identity to the red abalone, *H. rufescens*. Similarly, we collected COI sequence information from Gruenthal et al. (2007) who generated COI sequences from archived tissue collected between 1998 and 2003 from nine locations along the Californian coast. Their ~580 bp fragment produced 43 haplotypes for which they report little evidence for population structure. We supplemented *H. rufescens* comparisons with 605 bp COI sequences (12 haplotypes) from Hart et al. (2020) who collected red abalone larvae from 2012 to 2016 from Monterey. A comparison of these data sets (Figure 6b) aligned to our *H. rufescens* haplotypes revealed a total of 18 haplotypes, with two of our haplotypes nested within one major haplotype with one newly identified haplotype.

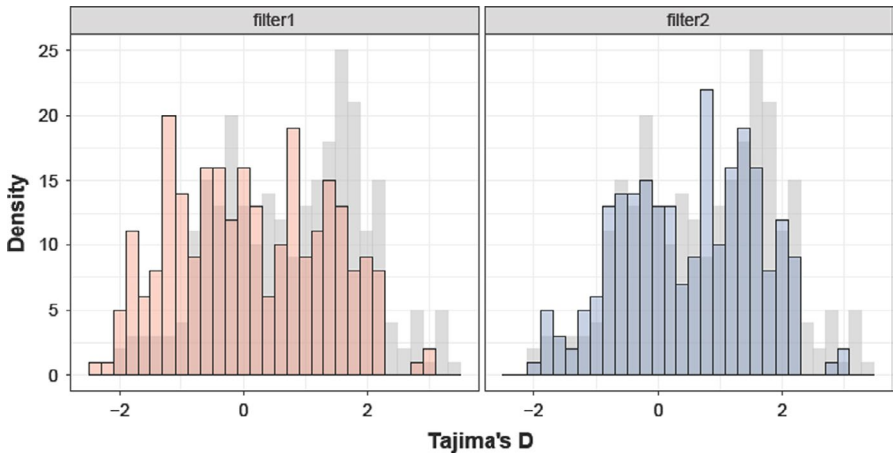

FIGURE 7 Histogram of the Tajima's *D* distribution using COI metabarcoding data comparing the HaploCOBB (grey) against filtered haplotype data sets Filter1 (one-read haplotypes removed) and Filter2 (haplotypes occurring one time removed, regardless of the number of reads). Tajima's *D* values based on individual COI sequences from a subset of these species were generally negative, similar to the blue histograms. MOTUs for which Tajima's *D* could not be computed are not shown

MOTU_398 and MOTU_33 identified with 99%–100% sequence identity to *Strongylocentrotus purpuratus* (eight haplotypes) and *P. miniata* (nine haplotypes) respectively. The COI sequences of these species are not well represented in GenBank and therefore inferences on their phylogeographical patterns are limited. Nonetheless, we collected available voucher COI sequences for these species that were used to construct the DNA barcode reference library for Canadian echinoderms (Layton et al., 2016). For *Strongylocentrotus purpuratus*, five COI sequences from animals collected in British Columbia (Canada) revealed three haplotypes and only two haplotypes were shared between the data sets with five unique haplotypes found (Figure 6c). For *P. miniata*, haplotypes were available from British Columbia (six haplotypes), Santa Barbara Channel (California, seven haplotypes) and Monterey (California, one haplotype). Our haplotype data set shows that two haplotypes are shared with the voucher sequences and we identified seven unique haplotypes (Figure 6d).

Overall, these tests strongly suggest that our pipeline for detecting and verifying haplotypes works well, and produces results largely in accordance with traditional population genetic approaches for COI.

3.9 | Population history

3.9.1 | Demographic patterns and data filters

We used the mismatch distribution of COI sequences within species and Tajima's *D* statistic to investigate population demographic changes. These values are sensitive to the shape of the intraspecific phylogenetic tree, particularly the relative fractions of short and long branches. Because of this sensitivity, we first examined typical data filtering and quality control approaches used in metabarcoding for their effect on results.

We compared results from three different data filters to explore the best assumption sets for detailed phylogenetic analysis of metabarcoding data. The first data set is unfiltered: it includes all the haploCOBB sequences that were produced by *Denoise* ($\alpha = 5$). This set captures all the diversity of all samples, but may contain some sequences produced by Illumina or PCR errors allowed by the $\alpha = 5$ setting. In the second data set (filter1), we removed all haplotypes that occurred as only a single read (seven haplotypes across seven MOTUs), under the assumption that these rarest singletons had the greatest chance of being sequencing errors. In the third data set (filter2), we removed all haplotypes that occurred only on one cobble (with any number of reads, 508 haplotypes across 145 MOTUs), under the assumption that this would leave only highly vetted haplotypes that were seen in two independent cobble samples. This third data set has the least chance of including spurious sequences, but probably eliminates a number of rare haplotypes that are true sequences in the population.

3.9.2 | Tajima's D

The HaploCOBB $\alpha = 5$ data set produced positive Tajima's *D* estimates for 187 MOTUs from 12 phyla (including Unclassified MOTUs) and negative estimates from 84 MOTUs (Figure 7; Table S2). By comparison, COI data from 16 invertebrate species collected by Kelly and Palumbi (2010) and 14 species reported by Marko et al. (2010) had overwhelmingly negative estimates: 15/16 for Kelly and Palumbi (2010) and 11/14 for Marko et al. (2010). This difference suggests that the HaploCOBB data set does not provide a reliable estimate of phylogenetic history. For the second metabarcoding data set (filter1), filtered by removing all one-read haplotypes, 128 Tajima's *D* values (50%) were positive and 130 negative (50%). This difference also suggests that removing singleton sequences does not fully bring this analysis into alignment with true population data. For the data

set (filter2) where we filtered out all haplotypes that occur on only a single cobble, there are 148 (63%) negative and 87 positive (37%) values. These analyses suggest that rare haplotypes produced by uncommon metabarcoding sequence errors have an effect on intraspecific tree structure for which Tajima's *D* is particularly sensitive. As a result, Tajima's *D* may not be a valid reflection of population history using MOTU haplotypes.

3.9.3 | Mismatch distributions

By contrast, values for overall nucleotide diversity and haplotype number are much less sensitive to data filters (the relationships between nucleotide diversity, haplotype number and Tajima's *D* for the haploCOBB data set against filter1 and filter2 are shown in Figure S4). One way that these less sensitive data have been used in understanding the demographic history of populations is in their use to create and analyse mismatch distributions—the pairwise differences among individuals in their DNA sequences. To distinguish between these species in our data set, we examined each MOTU for the fraction of pairwise comparisons between individuals within identical haplotypes vs those showing one or more sequences differences between different haplotypes. As an example of this variety of distributions, three MOTUs with three to four haplotypes are shown in Figure 8. Figure 8a (MOTU_4) shows a red alga (Gigartinales) in which one haplotype dominates, a typical distribution for recently expanded populations. Figure 8b is a chiton (MOTU_286, order Neoloricata) with three haplotypes and a bimodal mismatch distribution. Figure 8c (MOTU_261) is a bryozoan with two major haplotypes and a minor haplotype with one base different from each other, a distribution typical of a stable population or an ancient expansion.

To summarize mismatch distributions, we compared the frequency of identical haplotypes (e.g., mismatch =0) to the frequency of divergent haplotypes. Among 238 multihaplotype MOTUs, 86 are dominated by one haplotype and showed more than 50% of the sequence comparisons with zero divergence. These are the MOTUs that fall below the 1:1 line in Figure 9: based on similar analyses in Marko et al. (2010), these low-diversity taxa probably represent species that have probably expanded within the last 20,000 years. Above the line in Figure 9 are MOTUs with much higher levels of DNA divergence between haplotypes, and many fewer comparisons of identical sequences between individuals. These have an average sequence mismatch greater than zero, indicating population stability or more ancient expansion.

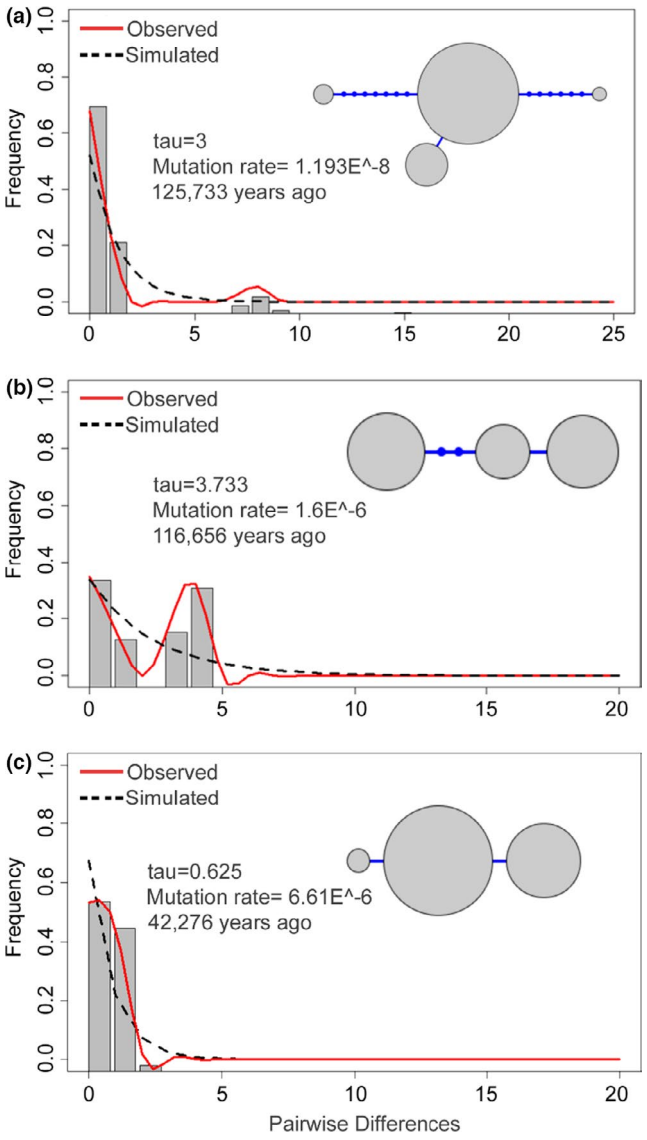

FIGURE 8 Mismatch distributions for three MOTUs with three or four haplotypes showing patterns that suggest different population histories. (a) MOTU_4 (red algae, order Gigartinales) has one major haplotype, and most comparisons of individuals showed zero sequence difference. (b) MOTU_286 (chiton, order Neoloricata) shows moderate diversity, and only about half the comparisons had zero sequence difference. (c) MOTU_261 (bryozoan, order Cheilostomatida) shows no dominant haplotype and most comparisons between individuals showed one or more sequence differences

4 | DISCUSSION

This study uses the potential of DNA metabarcoding-based haplotyping to test small-scale patterns in ecological abundance and genetic structure for 527 kelp forest species simultaneously. Tracking the similar but distinct DNA sequences that are typically binned into MOTUs in metabarcoding studies can greatly expand the power of these data to estimate population size, ecological aggregation, genetic diversity, genetic differentiation and population size changes over time. As a result, splitting MOTUs and using the haplotype data that are contained within them represents a significant increase in the power and utility of environmental and community DNA approaches in ecological monitoring. For our collection of animal and algal species, this approach reveals remarkable aggregation of small crustaceans, fine-scale genetic structure over 1–3 km in red algae, depauperate genetic variability in most sea stars and hidden species barriers in many taxa.

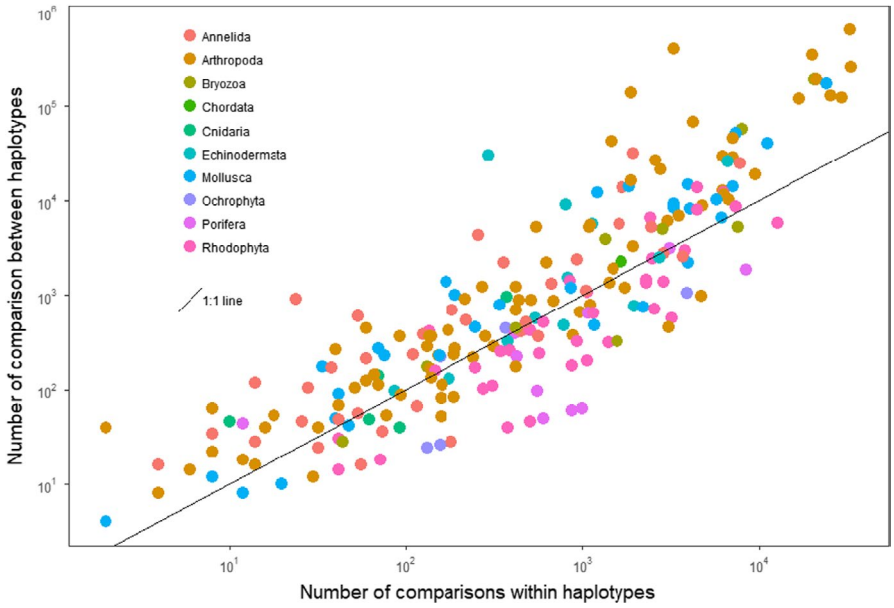

FIGURE 9 The number of comparisons within haplotypes (that have zero sequence difference) vs the number between haplotypes (with one or more sequence differences) among all MOTUs. The points below the line represent MOTUs with a dominant haplotype, probably due to low mutation rates, low population size or historical bottlenecks. Points above the line represent MOTUs with a wider diversity of haplotypes from populations with a long history of large population size. MOTUs with only one haplotype are not represented in this plot

4.1 | Testing data pipelines

To robustly examine haplotypes from metabarcoding data, we followed previous work (Elbrecht , Vamos, et al., 2018) and adjusted the data pipeline to estimate the impact of random sequencing errors on the results. By using a known set of sequences from one individual of a control species, we showed that adjusting the denoising procedure by changing the alpha (α) parameter changes how many spurious sequences are excluded, an observation similar to other examples of barcoded haplotype analyses (Figure 2). Data from using α = 5 also provided patterns of DNA substitution for the haplotype data sets that match empirical population genetic data sets of local invertebrate species (Kelly & Palumbi, 2010): higher numbers of transition substitutions and low numbers of amino acid substitutions (Figure 2). However, Elbrecht, Vamos, et al. (2018) revealed the potential for spurious haplotypes present in mock communities in validating the haplotyping approach, and as a result, testing haplotype patterns for sequence variation remains an important component of haplotype analysis of metabarcoding data.

4.2 | Diversity within MOTUs

Average COI diversity was low across the 527 MOTUs we focused on, averaging 0.28% for animals and 0.14% for algae (Table S1). Our metabarcoding DNA fragment is shorter (313 bp) than the COI fragment generally sequenced in single species studies (550–650 bp), and as a result, we see a higher fraction of MOTUs with only one haplotype. The low diversity observed is perhaps due to limited marker resolution and is exaggerated by the possibility of clonal spread across adjacent cobbles, such as algae (Arnaud-Haond et al., 2005). Nevertheless, summaries of larger COI regions reported similar values for overall nucleotide diversity. Marko et al. (2010) reported average nucleotide diversity of 0.2% for 15 species of Northeast Pacific intertidal animals, and Kelly and Palumbi (2010) found 0.005% among 34 species.

Abundant taxa with low nucleotide and haplotype diversity include species such as the sea anemones that are known for low divergence even between genera, probably due to low sequence evolutionary rates (Emblem et al., 2014; Shearer et al., 2002). Another possible case is within the brown algae (Phaeophyceae) where we find the giant kelp *Macrocystis pyrifera* on nearly every cobble sampled, most with the same COI sequence. Extensive surveys of the genus *Macrocystis* across the Pacific, Atlantic and Indian Oceans report very low sequence diversity in COI (Macaya & Zuccarello, 2010), suggesting widespread occurrence of one species. However, the potential of low rates of sequence evolution in these brown algae to generate low sequence divergence is poorly understood.

On the opposite side of the diversity scale, many MOTUs of small crustaceans, brittle stars and worms (classes Malacostraca, Maxillipoda, Ophiuroidea and Polychaeta) show extremely high nucleotide diversity, and high abundance. The numerous occurrences of different haplotypes of the same MOTU within and between samples provide an expanded opportunity to use eDNA data to measure ecological patterns.

4.3 | Ecological patterns discerned from haplotype abundance

The density of haplotype data in many species allows tests of differential abundance by using the presence/absence of each haplotype as an index of population size. Typical eDNA and community DNA data allow limited ecological inferences because the number of DNA reads of a particular MOTU correlate only poorly with the number of individuals of this species in the sample (Elbrecht et al., 2017; Lamb et al., 2019; Shum et al. 2019). Adding different haplotypes increases the accuracy of population estimates because

the number of different haplotypes in a sample must represent at least the same number of individuals. For example, we detected over 80 haplotypes of an ophiuroid MOTU (OTU_226; Figure 5c) on a single cobble even though the vast majority of cobbles have none. As a result, these data strongly suggest that many individuals of this species co-inhabited this single sample. By contrast, traditional metabarcoding analysis of high read counts of a single MOTU would be unable to distinguish a high population size from high biomass of a few individuals. In the case of OTU_226, brittle stars have high-dispersal larvae but have been known to aggregate in dense beds of more than 1000 individuals per square metre (Dauvin et al., 2013). In particular, species of *Ophiothrix* (similar genetically to OTU_226) have been reported to settle at densities of 50 per cm^2 on favourable substrates such as sponges (Turon et al., 2000). Geraldi et al. (2017) suggested that brittle star clumps may provide ecological benefits similar to sessile reef-forming species, but further investigation on the genetic patterns of species aggregations is required to determine the ecology of these dense aggregations.

Haplotype occurrence from sample to sample is well suited to analyses of overall patterns of clustering. The dispersion index (*I*, Pielou, 1969) has been used as one way to estimate the aggregation of individuals of a species across a landscape (see Johnson & Zimmer, 1985). Across our sample, the dispersion index based on metabarcoding haplotypes shows that the biota of individual cobbles in kelp forest is clumped, with high levels of aggregation especially in small arthropods, bryozoa, echinoderms and red algae (Figure S2). These results are based on clumping of haplotypes, not high numbers of sequence reads, and so are not biased by large body size or PCR artefacts. The clumping of species on cobbles may be due to ecological associations between predators and their prey, larvae settlement cues, ecological facilitation or other factors known to operate in marine communities to produce differences among local assemblages, such as in tide pools (Metaxas & Scheibling, 1993).

4.4 | Traditional population genetics data sets compared to meta-phylogeography

To study population genetics and phylogeography requires a lengthy process of concentrated sampling for each species. Despite these obstacles, tailored sampling strategies have allowed the successful collection of taxa for numerous genetic investigations that involve extracting adequate DNA, PCR amplification and Sanger-sequencing of individuals, often in the hundreds and for multiple species. For example, Kelly and Palumbi (2010) collected over 3000 individual specimens representing 34 intertidal invertebrate species of Annelids, Arthropods, Cnidaria, Echinoderms and Molluscs along the Northeastern Pacific coast and found different levels of genetic structure in support of known biogeographical breaks and oceanographic discontinuities that help define genetic connectivity on an ecosystem scale (Blanchette et al., 2008). Similarly, Marko et al. (2010) compared mitochondrial signatures of 14 species from this area. To contrast this to the moderate number of cobbles collected across a 3-km distance along the southern Monterey Peninsula, metabarcoding generated 527 taxonomic records and 2586 haplotypes for population genetic analysis.

The phylogeographical examples of *T. rubescens* and *H. rufescens* provide a comparative database of haplotypes against which we can test our metabarcoding data (Dawson et al., 2014; Gruenthal et al., 2007; Hart et al., 2020). These comparisons reveal the detection of shared haplotypes in our assessment as well as uncovering unique haplotypes not previously reported. For instance, we report 16 haplotypes of *T. rubescens*, of which four are shared among Dawson et al.'s (2014) Monterey collections. For the sea urchins *Strongylocentrotus purpuratus* and the sea star *P. miniata* we found haplotypes in common for animals collected as far apart as Santa Barbara and British Columbia but also revealed 39%–60% more haplotype diversity than previously reported (Figure 6c,d; Layton et al., 2016).

4.5 | Patterns in haplotype phylogeny and inferences about population history

Historical processes have played a central role in shaping evolutionary trajectories among species (O'Dea et al., 2016). In particular, characterizing genetic diversity within species allows tests of historical population abundance. For instance, Marko et al. (2010) reconstructed the demographic history of 14 invertebrate species in the Northeastern Pacific and suggested that some have expanded recently—probably after the last Glacial Maximum—whereas others appear to have maintained long-term demographic stability over the last 150,000 years. Here, we show that DNA metabarcoding haplotype information facilitates tests of the history of population size for hundreds of species.

A common way to quantify demographic signatures of population growth is to estimate Tajima's *D* (T_D) statistic (Tajima, 1983). In principle, a positive T_D points to a population with a deficiency of rare haplotypes that has experienced a recent bottleneck, while a negative T_D suggests a historical population expansion experiencing an excess of rare haplotypes. The effects of random, demographic and selective processes for any single marker can cause deviations from the neutral-equilibrium expectation, and it remains a challenge to determine which of these forces has had the most influence on DNA variation. Our analysis shows that T_D is particularly sensitive to assumptions in the quality control of sequence data. The number of rare haplotypes fluctuates greatly between data sets with different filters, and subsequently affects estimates of T_D. As a result, inferences of population histories based Tajima's *D* are sensitive to filtering assumptions.

By contrast, mismatch distribution tabulates pairwise differences among all DNA sequences for each MOTU and is less sensitive to rare haplotypes. In our analysis, mismatch analysis was far more stable to filtering assumptions (Figure S4). Our analysis

of mismatch distributions shows that a good fraction of abundant kelp forest species are dominated by a single haplotype, implying a recent expansion from a bottleneck. Even for taxa with more than one haplotype, past bottlenecks are suggested by low mismatch distributions for many understorey kelp forest species of red and brown algae as well as sponges (points below the line in Figure 9). By contrast, a large number of common arthropod, mollusc and echinoderm species show high diversity and high mismatch, suggesting longer periods of high population size in this ecosystem. Marko et al. (2010) uncovered similar variation, based on Tajima's *D*, and suggested that high-diversity species had persisted through climate cycles imposed by changing Ice Ages but that low-diversity species had become common much more recently, perhaps if kelp forest communities had succeeded in a broad expansion since that time.

Low population diversity may have negative consequences in changing environments and may indicate low adaptive capacity. For example, many of the sea star species susceptible to the recent outbreak of sea star wasting disease have little or no COI diversity in our data set or in traditional population surveys. By contrast, the sea star with less susceptibility, *Patiria miniata*, retains large populations sizes and substantial genetic diversity. Though the small piece of COI used in metabarcoding is not an adequate proxy for genome-wide genetic diversity (Wares, 2010), future research may confirm that diversity in this gene is a relative measure of overall adaptive capacity.

4.6 | Kelp forest community patterns from metabarcoding phylogeography

Our survey of 527 species over a limited geography reveals some previously hidden ecological and evolutionary patterns in this wide range of taxa. Species richness is high across cobble substrates, but many species are ecologically aggregated. This is especially true for the most abundant species: for those with mean abundance of haplotypes per cobble of >2, nearly all (22 of 26) show an aggregated pattern with a dispersion index >1 (mean 4.01 of 26 MOTUs). Aggregations of marine invertebrates can derive from gregarious larval settlement (e.g., Hadfield & Paul, 2001) or post-settlement interactions (e.g., Kristensen et al., 2013). In our case, cobbles may be like tide pools in that they draw from a common source pool of species but represent different ecological trajectories over small scales (Dethier, 1984).

Species from cobbles can show surprising population genetic structure over small spatial scales. In our survey, 48 species, about 18% of species with genetic diversity, showed high population differentiation with F_{ST} values above 0.1 (Figure 6e). These were dominated by red algae, sponges, isopods and amphipods, taxa known to generally show low-dispersal larvae (Grantham et al., 2003; Harwell & Orth, 2002). Most surveys of ecological communities for genetic structure find far fewer species with such fine-grained differentiation (e.g., Kelly & Palumbi, 2010; Marko et al., 2010; Toonen, 2001). However, these surveys rarely include large numbers of low-dispersal taxa such as provided by our metabarcoding survey. Such high levels of genetic structure imply very strong ecological isolation at the population over a scale of 1–3 km, and could allow high levels of local adaptation across environmental mosaics. The role of local adaptation has recently been highlighted in several species of intertidal brown algae (e.g., Nicastro et al., 2020) and to a lesser extent for the red algae (e.g., López et al., 2017). Our data suggest that local population differentiation among red algae may be common even over small spatial scales and that patterns of local dispersal have an opportunity to generate local adaptation.

A second implication of local population dispersal for taxa such as the 48 high F_{ST} species in our data set is that they are much more likely to respond positively to local protection in Marine Protected Areas (e.g., Krueck et al., 2017). The species in our high F_{ST} list include none that are heavily exploited, but some may provide juvenile habitat or food for those that are, such as fin fish.

The third implication of our data is regarding the stability of the kelp forest community over time, and the ability of these communities to adapt to changing conditions. Marko et al. (2010) hypothesized that genetically depauperate species might have less adaptive capacity in the face of climate change. Our COI region includes a very small number of bases, and species with zero variation in COI have nevertheless been shown to have population structure at nuclear loci (e.g., *Macrocystis* kelp, Camus et al., 2018; *Pisaster* sea star, Schiebelhut et al., 2018). Nevertheless, species with higher levels of mitochondrial variation may in fact harbour more nuclear genetic diversity accumulated over historical periods with different environmental conditions. This notion remains an important idea for species comparative testing.

4.7 | Present limitations and potential for future practical applications

The field of population genetics has been instrumental in enhancing the use of molecular markers that allow us to determine the genotypes of individuals and are a powerful means of investigating and quantifying genetic diversity, spatial structure and gene flow (Allendorf & Luikart, 2009). Genetic data facilitate in-depth analyses of different demographic scenarios that shape genetic patterns of diversity and can clarify species vulnerability and population status to extinction risk under future climate change (Harrisson et al., 2014). In the past decade, the field of molecular ecology has witnessed the development of effective metabarcoding-based tools for ecological assessment from various sample types in a variety of different ecosystems (van der Heyde et al., 2020; Mariani et al., 2019; Wangensteen et al., 2018). However, the intraspecific diversity of species from the enormous availability of metabarcoding biodiversity data sets remains hidden, and until recently, few studies have attempted to harness this concealed genetic component from environmental and community DNA (Elbrecht, Vamos, et al., 2018; Parsons et al., 2018; Sigsgaard et al., 2016; Tsuji et al., 2020; Turon et al., 2020; Zizka et al., 2020).

Examples from eDNA highlight the potential to capture intraspecific diversity from charismatic marine species using targeted assays for the mitochondrial control region (Parsons et al., 2018; Sigsgaard et al., 2016). Sigsgaard et al. (2016) compared mitochondrial control region haplotypes of tissue samples of whale sharks (*Rhincodon typus*) to water eDNA samples collected in the presence of large whale shark aggregations in the Arabian Gulf. They reported greater mitochondrial diversity in the eDNA samples with the addition of two haplotypes not observed in the tissue samples and were able to calculate meaningful estimates of effective population size from scaled eDNA haplotypes. In a later study, Parsons et al. (2018) screened the mitochondrial control region of the harbour porpoise (*Phocoena phocoena*) in Alaska and used prior knowledge of haplotypic sequence diversity to baseline data for comparison with eDNA haplotypes. Of the eight eDNA haplotypes found, six were observed from tissue samples previously collected and two were newly identified haplotypes with high read count. However, while these studies demonstrated encouraging evidence of population genetic inferences using targeted approaches, drawing conclusions from intraspecific genetic data using DNA metabarcoding of whole communities is fraught with limitations such as PCR amplification, sequencing error and the bioinformatics procedure used.

Haplotype-based DNA metabarcoding is accompanied by PCR amplification bias, and the choice of primer and level of degeneracy comes at a cost of undesired amplification of nontarget taxa (Krehenwinkel et al., 2017) and/or amplicon variation due to slippage (Elbrecht et al., 2018). Amplicon chimera formation and sequencing error can cause spurious haplotypes, so that error-correction methods are required to exclude them (Edgar, 2016). Work led by Elbrecht, Vamos, et al. (2018) introduced a methodological framework (JAMP) to analyse intraspecific diversity from metabarcoding data, and they tested their approach by using a mock community of known haplotypes to baseline a multispecies macroinvertebrate community data set. Their bioinformatic assessment involved denoising reads ($\alpha = 5$) to remove low-abundant haplotypes not distinguishable from sequencing noise. Turon et al. (2020) suggested an α of 5 and a minimal read depth of 20 as an optimal denoising strategy.

In the present study, we aimed to characterize haplotype diversity for hundreds of kelp forest species using DNA metabarcoding data from a popular set of COI primers. However, several caveats should be considered for this approach. First, we used the default setting provided in JAMP to perform the clustering of denoised haplotypes ($\alpha = 5$) into MOTUs. The JAMP procedure relies on constructing clusters using a strict distance cut-off (USEARCH; Edgar, 2013), and we used a cut-off threshold to cluster sequences with at least 97% sequence similarity for a highly diverse COI data set. However, the use of a single similarity threshold may not accommodate all taxa across the COI and implementing a flexible threshold approach (e.g., SWARM, Mahé et al., 2014) may improve resolution of clusters. Second, the COI fragment is 313 bp and considerably shorter than barcoded specimens at this region (~650 bp). This means the resolution of the haplotypes is limited and we found diversity estimates for this DNA segment to be lower than observed from single-sequence data (see Figure 6). As a result, the fraction of MOTUs with zero haplotype diversity is high. Third, while the denoising parameter alpha set to produced haplotypes consistent with common patterns of DNA sequence evolution (see Figure 2), further haplotype filtering by removing low-abundant reads was necessary to remove spurious haplotypes (Elbrecht, Vamos, et al., 2018). However, abundance filtering is a challenging issue and is likely to be optimal at different levels between studies or even analyses within studies (Turon et al., 2020). Our Tajima's *D* analyses illustrate this problem: three data sets show contrasting results for Tajima's *D* but not for other estimates of intraspecific phylogenetic patterns (Figure 7). Lastly, primer bias—which appears strongly in some multispecies amplifications—might also impact which intraspecific variants are amplified. To combat this bias, we adopted a presence/absence approach that scores a haplotype as present in a sample, and assigns a population abundance for that haplotype based on how many samples in which it occurs. Despite these obstacles, further research will reveal robust support toward a more calibrated framework for metabarcoding-based haplotype approaches for effective biomonitoring.

5 | CONCLUSION

DNA metabarcoding studies have focused on measuring biodiversity, identifying abundantly more taxa than morphological assessments. The development of metabarcoding-based haplotyping allows access to the genetic variability of MOTUs to reveal population structure. We took this approach further to unravel population genetic inferences to trace historical demographic connectivity and expansion–contraction dynamics for hundreds of co-inhabiting kelp forest species that offer clues about the community response to environmental changes and status of the ecosystem. Our approach allowed us to track intraspecific diversity and facilitated enhanced biodiversity monitoring, and these patterns yield interesting inferences about the intensity at which taxonomic groups experience past events, and importantly, infer which groups will be potentially affected by the effects of climate change.

ACKNOWLEDGEMENTS

We thank Vasco Elbrecht for assistance in troubleshooting the JAMP pipeline; and Aurélie Bonin and three anonymous reviewers for constructive feedback offered to the final version of the manuscript. This work was funded by the David and Lucile Packard Foundation through the Marine Life Observatory and PISCO, and with funding from the Monterey Bay Aquarium.

AUTHOR CONTRIBUTION

P.S. and S.R.P. conceptualized the study, carried out statistical analyses and wrote the manuscript.

DATA AVAILABILITY STATEMENT

Raw Illumina sequences can be found on NCBI's SRA database BioProject ID: PRJNA552876. The data set including MOTUs, haplotypes and associated sequences, taxonomic assignment, abundances as well as OBITOOLS and JAMP bioinformatics scripts, R scripts, and sample barcodes has been deposited in Dryad (https://doi.org/10.5061/dryad.cvdncjt3q). Bioinformatic scripts are also available at: https://github.com/shump2/haploCOBB.

ORCID

Peter Shum https://orcid.org/0000-0001-8154-9828

REFERENCES

Akita, S., Takano, Y., Nagai, S., Kuwahara, H., Kajihara, R., Tanabe, A. S., & Fujita, D. (2019). Rapid detection of macroalgal seed bank on cobbles: Application of DNA metabarcoding using next-generation sequencing. *Journal of Applied Phycology*, *31*(4), 2743–2753.

Allendorf, F. W., & Luikart, G. (2009). *Conservation and the genetics of populations*. John Wiley & Sons.

Arafeh-Dalmau, N., Schoeman, D. S., Montaño-Moctezuma, G., Micheli, F., Rogers-Bennett, L., Olguin-Jacobson, C., & Possingham, H. P. (2020). Marine heat waves threaten kelp forests. *Science*, *367*(6478), 635.

Archer, F. I., Adams, P. E., & Schneiders, B. B. (2016). strataG: An R package for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources*. *17*(1), 5–11. https://doi.org/10.1111/1755-0998.12559

Arnaud-Haond, S., Alberto, F., Teixeira, S., Procaccini, G., Serrao, E. A., & Duarte, C. M. (2005). Assessing genetic diversity in clonal organisms: Low diversity or low resolution? Combining power and cost efficiency in selecting markers. *Journal of Heredity*, *96*(4), 434–440.

Azarian, C., Foster, S., Devloo-Delva, F., & Feutry, P. (2020). Population differentiation from environmental DNA: Investigating the potential of haplotype presence/absence-based analysis of molecular variance. *Environmental DNA*. https://doi.org/10.1002/edn3.143

Blanchette, C. A., Melissa Miner, C., Raimondi, P. T., Lohse, D., Heady, K. E., & Broitman, B. R. (2008). Biogeographical patterns of rocky intertidal communities along the Pacific coast of North America. *Journal of Biogeography*, *35*(9), 1593–1607.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*(1), 176–182.

Britten, R. J., Cetta, A., & Davidson, E. H. (1978). The single-copy DNA sequence polymorphism of the sea urchin *Strongylocentrotus purpuratus*. *Cell*, *15*(4), 1175–1186.

Bush, A., Compson, Z. G., Monk, W., Porter, T. M., Steeves, R., Emilson, E. J., Gagne, N., Hajibabaei, M., Roy, M., & Baird, D. (2019). Studying ecosystems with DNA metabarcoding: Lessons from biomonitoring of aquatic macroinvertebrates. *Frontiers in Ecology and Evolution*, *7*, 434.

Camus, C., Faugeron, S., & Buschmann, A. H. (2018). Assessment of genetic and phenotypic diversity of the giant kelp, *Macrocystis pyrifera*, to support breeding programs. *Algal Research*, *30*, 101–112.

Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: a contributed package to the R Project for statistical computing devoted to biological sequences retrieval and analysis. In U. Bastolla, M. Porto, H. E. Roman, et al. (Eds.), *Structural approaches to sequence evolution*. Biological and medical physics, biomedical engineering (pp. 207–232). Berlin: Springer. http://dx.doi.org/10.1007/978-3-540-35306-5_10

Clarke, L. J., Soubrier, J., Weyrich, L. S., & Cooper, A. (2014). Environmental metabarcodes for insects: In silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, *14*(6), 1160–1170.

Curtis, A. N., Larson, E. R., & Davis, M. A. (2020). Field storage of water samples affects measured environmental DNA concentration and detection. *Limnology*, *22*(1), 1–4.

Dauvin, J. C., Méar, Y., Murat, A., Poizot, E., Lozach, S., & Beryouni, K. (2013). Interactions between aggregations and environmental factors explain spatio-temporal patterns of the brittle-star *Ophiothrix fragilis* in the eastern Bay of Seine. *Estuarine, Coastal and Shelf Science*, *131*, 171–181.

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, *6*, 226–230.

Dawson, M. N., Hays, C. G., Grosberg, R. K., & Raimondi, P. T. (2014). Dispersal potential and population genetic structure in the marine intertidal of the eastern North Pacific. *Ecological Monographs*, *84*(3), 435–456.

Dethier, M. N. (1984). Disturbance and recovery in intertidal pools: maintenance of mosaic patterns. *Ecological Monographs*, *54*(1), 99–118.

Di Muri, C., Lawson Handley, L., Bean, C. W., Li, J., Peirson, G., Sellers, G. S., Walsh, K., Watson, H. V., Winfield, I. J., & Hänfling, B. (2020). Read counts from environmental DNA (eDNA) metabarcoding reflect fish abundance and biomass in drained ponds. *Metabarcoding and Metagenomics*, *4*, e56959. https://doi.org/10.3897/mbmg.4.56959

Edgar, R. C. (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*(10), 996.

Edgar, R. C. (2016). UNOISE2: Improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv*, 081257.

Elbrecht, V., Hebert, P. D., & Steinke, D. (2018). Slippage of degenerate primers can cause variation in amplicon length. *Scientific Reports*, *8*(1), 1–5.

Elbrecht, V., Peinert, B., & Leese, F. (2017). Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecology and Evolution*, *7*(17), 6918–6926.

Elbrecht, V., Vamos, E. E., Steinke, D., & Leese, F. (2018). Estimating intraspecific genetic diversity from community DNA metabarcoding data. *PeerJ*, *6*, e4644.

Emblem, Å., Okkenhaug, S., Weiss, E. S., Denver, D. R., Karlsen, B. O., Moum, T., & Johansen, S. D. (2014). Sea anemones possess dynamic mitogenome structures. *Molecular Phylogenetics and Evolution*, *75*, 184–193.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567.

Felsenstein, J. (1992). Estimating effective population size from samples of sequences: Inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genetics Research*, *59*(2), 139–147.

Felsenstein, J. (1993). PHYLIP (phylogeny inference package), version 3.5 c. Joseph Felsenstein.

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P., & Pompanon, F. (2010). An in silico approach for the evaluation of DNA barcodes. *BMC Genomics*, *11*(1), 434.

Geller, J., Meyer, C., Parker, M., & Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, *13*(5), 851–861.

Geraldi, N. R., Bertolini, C., Emmerson, M. C., Roberts, D., Sigwart, J. D., & Connor, N. E. (2017). Aggregations of brittle stars can perform similar ecological roles as mussel reefs. *Marine Ecology Progress Series*, *563*, 157–167.

Grantham, B. A., Eckert, G. L., & Shanks, A. L. (2003). Dispersal potential of marine invertebrates in diverse habitats: Ecological archives A013–001-A1. *Ecological Applications*, *13*(s1), 108–116.

Gruenthal, K. M., Acheson, L. K., & Burton, R. S. (2007). Genetic structure of natural populations of California red abalone (*Haliotis rufescens*) using multiple genetic markers. *Marine Biology*, *152*(6), 1237–1248.

Hadfield, M. G., & Paul, V. J. (2001). Natural chemical cues for settlement and metamorphosis of marine invertebrate larvae. *Marine Chemical Ecology*, *13*, 431–461.

Harrisson, K. A., Pavlova, A., Telonis-Scott, M., & Sunnucks, P. (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications*, *7*(9), 1008–1025.

Hart, L. C., Goodman, M. C., Walter, R. K., Rogers-Bennett, L., Shum, P., Garrett, A. D., Watanabe, J. M., & O'Leary, J. K. (2020). Abalone recruitment in low-density and aggregated populations facing climatic stress. *Journal of Shellfish Research*, *39*(2), 359–373.

Harwell, M. C., & Orth, R. J. (2002). Long-distance dispersal potential in a marine macrophyte. *Ecology*, *83*(12), 3319–3330.

Hudson, R. R., Boos, D. D., & Kaplan, N. L. (1992). A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution*, *9*(1), 138–151.

Johnson, R. B., & Zimmer, W. J. (1985). A more powerful test for dispersion using distance measurements. *Ecology*, *66*(5), 1669–1675.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Kelly, R. P., & Palumbi, S. R. (2010). Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS One*, *5*(1).

Krehenwinkel, H., Wolf, M., Lim, J. Y., Rominger, A. J., Simison, W. B., & Gillespie, R. G. (2017). Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports*, *7*(1), 1–12.

Kristensen, E., Delefosse, M., Quintana, C. O., Banta, G. T., Petersen, H. C., & Jørgensen, B. (2013). Distribution pattern of benthic invertebrates in Danish estuaries: The use of Taylor's power law as a species-specific indicator of dispersion and behavior. *Journal of Sea Research*, *77*, 70–78.

Krueck, N. C., Ahmadia, G. N., Green, A., Jones, G. P., Possingham, H. P., Riginos, C., Treml, E. A., & Mumby, P. J. (2017). Incorporating larval dispersal into MPA design for both conservation and fisheries. *Ecological Applications*, *27*(3), 925–941.

Kumar, G., Eble, J. E., & Gaither, M. R. (2020). A practical guide to sample preservation and pre-PCR processing of aquatic environmental DNA. *Molecular Ecology Resources*, *20*(1), 29–39.

Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G., & Taylor, M. I. (2019). How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology*, *28*(2), 420–430.

Layton, K. K., Corstorphine, E. A., & Hebert, P. D. (2016). Exploring Canadian echinoderm diversity through DNA barcodes. *PLoS One*, *11*(11), e0166118.

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, *10*(1), 34.

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*(11), 1451–1452.

López, B. A., Tellier, F., Retamal-Alarcón, J. C., Pérez-Araneda, K., Fierro, A. O., Macaya, E. C., Tala, F., & Thiel, M. (2017). Phylogeography of two intertidal seaweeds, *Gelidium lingulatum* and *G. rex* (Rhodophyta: Gelidiales), along the South East Pacific: Patterns explained by rafting dispersal? *Marine Biology*, *164*(9), 188.

Macaya, E. C., & Zuccarello, G. C. (2010). DNA Barcoding and genetic divergence in the Giant Kelp Macrocystis (Laminariales) 1. *Journal of Phycology*, *46*(4), 736–742.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2014). Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ*, *2*, e593.

Mariani, S., Baillie, C., Colosimo, G., & Riesgo, A. (2019). Sponges as natural environmental DNA samplers. *Current Biology*, *29*(11), R401–R402.

Marko, P. B., Hoffman, J. M., Emme, S. A., McGovern, T. M., Keever, C. C., & Nicole Cox, L. (2010). The 'Expansion-Contraction' model of Pleistocene biogeography: Rocky shores suffer a sea change? *Molecular Ecology*, *19*(1), 146–169.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet Journal*, *17*(1), 10–12.

Melis, R., Ceccherelli, G., Piazzi, L., & Rustici, M. (2019). Macroalgal forests and sea urchin barrens: Structural complexity loss, fisheries exploitation and catastrophic regime shifts. *Ecological Complexity*, *37*, 32–37.

Metaxas, A., & Scheibling, R. E. (1993). Community structure and organization of tidepools. *Marine Ecology Progress Series*, *98*(1-2), 187–198.

Natesh, M., Taylor, R. W., Truelove, N. K., Hadly, E. A., Palumbi, S. R., Petrov, D. A., & Ramakrishnan, U. (2019). Empowering conservation practice with efficient and economical genotyping from poor quality samples. *Methods in Ecology and Evolution*, *10*(6), 853–859.

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.

Nei, M., & Tajima, F. (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics*, *97*, 145–163.

Nicastro, K. R., Assis, J., Serrão, E. A., Pearson, G. A., Neiva, J., Valero, M., Jacinto, R., & Zardi, G. I. (2020). Congruence between fine-scale genetic breaks and dispersal potential in an estuarine seaweed across multiple transition zones. *ICES Journal of Marine Science*, *77*(1), 371–378.

O'Dea, A., Lessios, H. A., Coates, A. G., Eytan, R. I., Restrepo-Moreno, S. A., Cione, A. L., Collins, L. S., De Queiroz, A., Farris, D. W., Norris, R. D., & Stallard, R. F. (2016). Formation of the Isthmus of Panama. *Science Advances*, *2*(8), e1600883.

Palumbi, S. R. (1984). Tactics of acclimation: Morphological changes of sponges in an unpredictable environment. *Science*, *225*(4669), 1478–1480.

Parsons, K. M., Everett, M., Dahlheim, M., & Park, L. (2018). Water, water everywhere: Environmental DNA can unlock population structure in elusive marine species. *Royal Society Open Science*, *5*, 180537.

Pielou, E. (1969). Association tests versus homogeneity tests: Their use in subdividing quadrats into groups. *Vegetatio*, *18*(1), 4–18.

Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., Yamahara, K. M., & Kelly, R. P. (2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*, *25*(2), 527–541.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European molecular biology open software suite. *Trends in Genetics*, *16*, 276–277.

Rogers, A. R., & Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, *9*(3), 552–569.

Rogers-Bennett, L., & Catton, C. A. (2019). Marine heat wave and multiple stressors tip bull kelp forest to sea urchin barrens. *Scientific Reports*, *9*(1), 1–9.

Rogers-Bennett, L., Dondanville, R. F., Catton, C. A., Juhasz, C. I., Horii, T., & Hamaguchi, M. (2016). Tracking larval, newly settled, and juvenile red abalone (*Haliotis rufescens*) recruitment in Northern California. *Journal of Shellfish Research*, *35*(3), 601–609.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2584.

Sales, N. G., McKenzie, M. B., Drake, J., Harper, L. R., Browett, S. S., Coscia, I., Wangensteen, O. S., Baillie, C., Bryce, E., Dawson, D. A., & Ochu, E. (2020). Fishing for mammals: Landscape-level monitoring of terrestrial and semi-aquatic communities using eDNA from riverine systems. *Journal of Applied Ecology*, *57*(4), 707–716.

Salzburger, W., Ewing, G. B., & Von Haeseler, A. (2011). The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. *Molecular Ecology*, *20*(9), 1952–1963.

Schiebelhut, L. M., Puritz, J. B., & Dawson, M. N. (2018). Decimation by sea star wasting disease and rapid genetic change in a keystone species, *Pisaster ochraceus*. *Proceedings of the National Academy of Sciences*, *115*(27), 7069–7074.

Shearer, T. L., Van Oppen, M. J. H., Romano, S. L., & Wörheide, G. (2002). Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). *Molecular Ecology*, *11*(12), 2475–2487.

Shum, P., Barney, B. T., O'Leary, J. K., & Palumbi, S. R. (2019). Cobble community DNA as a tool to monitor patterns of biodiversity within kelp forest ecosystems. *Molecular Ecology Resources*, *19*(6), 1470–1485.

Sigsgaard, E. E., Jensen, M. R., Winkelmann, I. E., Møller, P. R., Hansen, M. M., & Thomsen, P. F. (2020). Population-level inferences from environmental DNA—Current status and future perspectives. *Evolutionary Applications*, *13*(2), 245–262.

Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E. D., Robinson, D. P., Knudsen, S. W., Pedersen, M. W., Al Jaidah, M., Orlando, L., Willerslev, E., & Møller, P. R. (2016). Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution*, *1*(1), 1–5.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*(2), 437–460.

Toonen, R. J. (2001). *Genetic analysis of recruitment and dispersal patterns in the porcelain shore crab*, Petrolisthes cinctipes (Doctoral dissertation. Davis). University of California.

Tsuji, S., Miya, M., Ushio, M., Sato, H., Minamoto, T., & Yamanaka, H. (2020). Evaluating intraspecific genetic diversity using environmental DNA and denoising approach: A case study using tank water. *Environmental DNA*, *2*(1), 42–52.

Turon, X., Antich, A., Palacín, C., Præbel, K., & Wangensteen, O. S. (2020). From metabarcoding to metaphylogeography: Separating the wheat from the chaff. *Ecological Applications*, *30*(2), e02036.

Turon, X., Codina, M., Tarjuelo, I., Uriz, M. J., & Becerro, M. A. (2000). Mass recruitment of *Ophiothrix fragilis* (Ophiuroidea) on sponges: Settlement patterns and post-settlement dynamics. *Marine Ecology Progress Series*, *200*, 201–212.

van der Heyde, M., Bunce, M., Wardell-Johnson, G., Fernandes, K., White, N. E., & Nevill, P. (2020). Testing multiple substrates for terrestrial biodiversity monitoring using environmental DNA metabarcoding. *Molecular Ecology Resources*, *20*(3), 732–745.

Vernooy, R., Haribabu, E., Muller, M. R., Vogel, J. H., Hebert, P. D., Schindel, D. E., Shimura, J., & Singer, G. A. (2010). Barcoding life to conserve biological diversity: Beyond the taxonomic imperative. *PLoS Biology*, *8*(7), e1000417.

Wangensteen, O. S., Palacín, C., Guardiola, M., & Turon, X. (2018). DNA metabarcoding of littoral hard-bottom communities: High diversity and database gaps revealed by two molecular markers. *PeerJ*, *6*, e4705.

Wares, J. P. (2010). Natural distributions of mitochondrial sequence diversity support new null hypotheses. *Evolution: International Journal of Organic Evolution*, *64*, 1136–1142.

Xia, X., & Xie, Z. (2001). DAMBE: Software package for data analysis in molecular biology and evolution. *Journal of Heredity*, *92*(4), 371–373.

Zizka, V. M. A., Weiss, M., & Leese, F. (2020). Can metabarcoding resolve intraspecific genetic diversity changes to environmental stressors? A test case using river macrozoobenthos. *Metabarcoding and Metagenomics*, *4*, e51925. https://doi.org/10.3897/mbmg.4.51925

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Shum P, Palumbi SR. Testing small-scale ecological gradients and intraspecific differentiation for hundreds of kelp forest species using haplotypes from metabarcoding. *Mol Ecol*. 2021;30:3355–3373. https://doi.org/10.1111/mec.15851