REPORT

Reef-associated crustacean fauna: biodiversity estimates using semi-quantitative sampling and DNA barcoding

L. Plaisance · N. Knowlton · G. Paulay ·
C. Meyer

Abstract The cryptofauna associated with coral reefs accounts for a major part of the biodiversity in these ecosystems but has been largely overlooked in biodiversity estimates because the organisms are hard to collect and identify. We combine a semi-quantitative sampling design and a DNA barcoding approach to provide metrics for the diversity of reef-associated crustacean. Twenty-two similar-sized dead heads of *Pocillopora* were sampled at 10 m depth from five central Pacific Ocean localities (four atolls in the Northern Line Islands and in Moorea, French Polynesia). All crustaceans were removed, and partial cytochrome oxidase subunit I was sequenced from 403 individuals, yielding 135 distinct taxa using a species-level criterion of 5% similarity. Most crustacean species were rare; 44% of the OTUs were represented by a single individual, and an additional 33% were represented by several specimens found only in one of the five localities. The Northern Line Islands and Moorea shared only 11 OTUs. Total numbers estimated by species richness statistics (Chao1 and ACE) suggest at least 90 species of crustaceans in Moorea and 150 in the Northern Line Islands for this habitat type. However, rarefaction curves for each region failed to approach an asymptote, and Chao1 and ACE estimators did not stabilize after sampling eight heads in Moorea, so even these diversity figures are underestimates. Nevertheless, even this modest sampling effort from a very limited habitat resulted in surprisingly high species numbers.

Keywords Coral reefs · Biodiversity · Crustaceans · Pacific Ocean · DNA barcoding

Communicated by Biology Editor Dr. Ruth Gates

L. Plaisance (✉) · N. Knowlton
Center for Marine Biodiversity and Conservation, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093-0202, USA
e-mail: plaisancel@si.edu

L. Plaisance · N. Knowlton · C. Meyer
Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, MRC 163, PO Box 37012, Washington, DC 20013-7012, USA

G. Paulay
Florida Museum of Natural History, University of Florida, Gainesville, FL 32611-7800, USA

Introduction

To date, approximately 250,000 marine species have been described (Reaka-Kudla 1997; Groombridge and Jenkins 2000, Bouchet 2006). Our assumptions about total marine biodiversity (described and undescribed) are based on broad estimates from different methods. Examples include extrapolations from quantitative marine samples from specific ecosystems (e.g., the deep sea; Grassle and Maciolek 1992), extrapolations from the described fauna from better known regions or groups (e.g., European seas or Brachyura; Bouchet 2006), or comparisons with estimates of terrestrial biodiversity (e.g., tropical rain forests; Reaka-Kudla 1997). These have led to estimates of total marine species diversity spanning three orders of magnitude, from 5×10^5 (May 1994) to 10^8 (Grassle and Maciolek 1992) [see Bouchet (2006) for more examples of biodiversity estimates]. These extrapolations have created much controversy (Lambshead and Boucher 2003) as there is no easy and straightforward way of estimating global marine

biodiversity reliably, leaving the door open for much conjecture and debate.

Uncertainty about the biodiversity of coral reefs is especially high. Although coral reefs represent less than 0.2% of the area of the ocean, they are the most diverse of all marine ecosystems on a per area basis, and perhaps absolutely as well, the deep sea being the other major repository of marine biodiversity (Sala and Knowlton 2006). Reaka-Kudla (1997) divided coral reef biodiversity into three main components: fishes, reef-building organisms and cryptofauna. For some groups, we have a relatively good understanding of the patterns of diversity and endemism because they are easily collected, and taxonomic expertise is long-standing (especially corals, fishes and some macroinvertebrates) (Veron 1995; Roberts et al. 2002; Karlson et al. 2004). Some biodiversity knowledge can be accessed from databases such as Fishbase (http://www.fishbase.org/search.php), Hexacorallia (http://www.kgs.ku.edu/Hexacoral) and Indo-Pacific Marine Mollusks (http://data.acnatsci.org/obis/find_mollusk.html), or printed resources such as the Systema Brachyurorum (Ng et al. 2008). However, most coral reef diversity is made up of small, cryptic species and species from poorly known groups. Thus, we do not know to even the nearest order of magnitude how many species are associated with coral reef ecosystems, with published estimates ranging from ∼1 to 10 million species (Reaka-Kudla 1997; Small et al. 1998).

Coral reefs are also one of the most endangered marine ecosystems (Knowlton 2001; Bellwood et al. 2004), and dramatic declines in corals and fishes have been well documented (Gardner et al. 2003; Pandolfi et al. 2003; Bruno and Selig 2007; Knowlton and Jackson 2008). However, most reef biodiversity lies outside these two groups, and we have almost no understanding of how reef degradation threatens this biodiversity. The best example we have is the study by Idjadi and Edmunds (2006) who documented positive relationships between various aspects of reef condition (most notably topographic complexity) and the generic diversity of associated invertebrates.

Lately, an increasing awareness of these problems in the scientific community has led to several large-scale initiatives to inventory coral reef biodiversity. These include the investigation of the mollusk fauna in New Caledonia (Bouchet et al. 2002), the marine biodiversity survey of Guam and the Marianas (Paulay 2003), the Santo 2006 expedition in Vanuatu (http://www.santo2006.org), the Moorea Biocode Project (http://bscit.berkeley.edu/biocode), and the Census of Marine Life—Census of Coral Reefs (http://www.creefs.org) survey of French Frigate Shoals (northwestern Hawaiian Islands) in 2006 and in Australia in 2008. These expeditions have put a special emphasis on small and understudied organisms, particularly invertebrate, algal and microbial species.

The current rate of species description using traditional methods is extremely slow, and identifications based on traditional keys typically require specialized expertise. These two bottlenecks have severely limited our understanding of coral reef biodiversity. However, the revolution in molecular genetics has dramatically changed the potential for reef scientists to make progress in this area. DNA barcoding, in particular, has the potential to speed the identification of described species (Hebert et al. 2003), but its use to estimate species numbers, regardless of their formal taxonomic state, is probably even more important for understanding biodiversity patterns and trends. Especially, noteworthy is its ability to detect cryptic species, which are common on coral reefs and are difficult to detect using traditional taxonomic methods (Knowlton 1993, 2000). DNA barcoding has stimulated intense debate, primarily about its reliability at the species level, and it is accepted that a single molecular divergence cutoff for species delimitation is not defensible (Meyer and Paulay 2005). However, studies on crustaceans based on broad datasets have shown the utility of the 5′ end of the mitochondrial cytochrome oxidase subunit I in species delimitation (Lefébure et al. 2006; Costa et al. 2007).

This study focused on the crustacean fauna inhabiting coral reef interstices at five central Pacific Ocean localities, regions of moderate reef diversity well to the east of the coral triangle epicenter of reef diversity (Myers et al. 2000; Hughes et al. 2002; Hoeksema 2007). In order to avoid biases and inaccuracies associated with nonquantitative sampling strategies, the crustacean fauna was sampled from similar-sized and structured reef units—dead heads of *Pocillopora* coral—and extrapolation techniques were used to standardize richness data. Furthermore, sampling was restricted to a single depth and type of reef exposure. This allowed estimation of the crustacean species richness in this habitat, as well as the exploration of diversity patterns and estimation of the number of species still to be documented.

Materials and methods

Sampling

Similar-sized dead heads of *Pocillopora verrucosa* (height + width + depth ≈ 90 cm ± 22%) were collected from four atolls in the Northern Line Islands in August 2005 (Kingman, Palmyra, Tabuaeran and Kiritimati) and in Moorea, Society Islands, French Polynesia, in August 2006. The coral heads were selected to minimize successional and environmental differences. All were collected from a depth of 10 m on the fore-reef. They were colonized by encrusting flora and fauna so that bare

skeleton was obscured, but still remained attached to the reef at the base (so that proximity to the surface of the reef was standardized), and the branching structure of the coral was still present.

The heads were gently broken from the bottom with a hammer and chisel and quickly placed in a 20 l bucket underwater. No mesh or covering was used for this study, but we did not observe any animals escaping (most associates initially cling tightly to dead submerged coral heads if the head itself remains intact). All macroinvertebrates ($>\sim 5$ mm in size) encountered were extracted from the head, shipboard in the Northern Line Islands and in the laboratory at Moorea, as follows: Each branch of the coral head was detached with a hammer and chisel and examined closely for motile invertebrates. The remaining rubble was placed in a bucket of seawater. When all the branches and the base were broken apart and examined, the fragments were then broken up in smaller pieces (~ 5 cm) and examined a second time for remaining creatures. We did not attempt to extract boring organisms. The seawater, in which the coral head and later on the coral fragments were kept, was sieved through a 2 mm sieve.

Decapods and stomatopods were sorted to morphospecies, and abundances of each recorded. Minute, often postlarval, decapods and peracarids were not sorted to morphospecies, but all were set aside for sequence-based identification. Each morphospecies was identified in the field to the lowest taxonomic rank possible [with identification confirmed or modified afterwards by Gustav Paulay and several specialists (see Acknowledgements)]. One to three exemplars were photographed, and up to five individuals of each morphospecies were processed for sequencing from each coral head. For larger organisms, a tissue sample was collected (most commonly a leg) for DNA analysis and frozen at $-80°C$, and the individual was then preserved in 95% ethanol and vouchered at Florida Museum of Natural History (FLMNH). For smaller specimens, the entire organism was frozen at $-80°C$, sequenced, and thus no vouchers were taken ($\approx 15\%$ of all organisms sampled). A single head typically took an entire day to process (collection of the coral, extraction of the associated fauna and vouchering and preservation of the specimens). This same procedure was applied for each new head sampled.

Extraction, PCR amplification and sequencing

Total genomic DNA was extracted from each specimen using DNeasy 96 Blood and Tissue kit (Qiagen) according to the manufacturer's protocol. DNA was eluted in a final volume of 50 μl. A 658 base-pair (bp) fragment of the mitochondrial cytochrome oxidase subunit I gene (COI) was amplified using the primers LCO1490 and HCO2198

(Folmer et al. 1994). Twenty-five μl PCR amplifications were performed with 2 μl of DNA extract, 10 pM of each PCR primer and Ready-To-Go PCR beads (Amersham Pharmacia Biotech), each containing 1.5 U *Taq* polymerase, 10 mM Tris–HCl at pH 9, 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM of each dNTP and stabilizers including bovine serum albumin. The PCR conditions consisted of 1 min at 94°C followed by 5 cycles of 40 s at 94°C, 40 s at 45°C, 60 s at 72°C; followed by 35 cycles of 40 s at 94°C, 40 s at 51°C, 60 s at 72°C; followed by 5 min at 72°C. Successful PCRs, where a single fragment was amplified, were purified using the QIAquick PCR Purification Kit (Qiagen). When several fragments were obtained (low annealing temperatures made it easier to amplify from a broad taxonomic array of organisms, but sometimes resulted in the amplification of pseudogenes), the PCR product was run on an agarose gel (2%) containing EtBr, and the target fragment was excised from the gel and purified using QIAquick Gel Extraction Kit (Qiagen). Automated sequencing was performed directly on purified PCR products using ABI BigDye terminator V3.1. Sequence reactions were purified using Millipore 96-well plates loaded with Sephadex G-50 and run on an ABI 3130xl genetic analyzer (Applied Biosystems). Products were sequenced in both directions using LCO1490 and HCO2198.

Data analysis

Sequences were assembled and edited using Sequencher v. 4.5 (Gene Codes, Ann Arbor, MI). Each unique sequence served as a blast query to the GenBank database to identify the most similar sequence in GenBank to the queried sequence. This allowed us to remove problematic sequences, which stemmed from contamination problems (non-crustacean sequences) or sequences showing a mismatch with the initial description of the organism sampled (e.g. a crab sequence from an animal initially recorded as a shrimp). A few putative pseudogenes were also removed based on the presence of stop codons or reading frame shifts. The rest of the sequences were subsequently aligned using MacClade 4 (Maddison and Maddison 2000) and submitted to GenBank (accession numbers: GQ260847–GQ260981).

In order to cluster the sequences into Operational Taxonomic Units (OTUs), nucleotide sequence divergences were calculated with DNADIST of the Phylip package (Felsenstein 1989) using the Kimura-two-parameter (K2P) model. The pairwise distances served as an input to DOTUR (Schloss and Handelsman 2005). In order to choose the sequence dissimilarity threshold used for species discrimination, a step function analysis was run, testing the number of OTUs found as a function of the value of the

threshold for crustaceans found in Moorea, the Line Islands and the combined data from Moorea and the Line Islands.

The furthest neighbor clustering algorithm, with a 5% dissimilarity for definition of Operational Taxonomic Units (OTUs), was used for clustering sequences into OTUs, generating rarefaction curves, and calculating the species richness estimators ACE and Chao 1 (Hortal et al. 2006) for both the total crustacean fauna and for the decapods only. Chao1 and ACE are abundance-based nonparametric estimators of species richness that work by examining the number of species in a sample observed more than once (2 times or up to 10 times for Chao1 and ACE, respectively) relative to the number of species that are observed just once. The advantage of using these estimators is that the estimated diversity of samples can be compared, even when the true diversity of the total population is not known. In the absence of complete inventories, nonparametric estimators have been shown to perform better than most other methods, such as observed species richness, species–area curves or asymptotic estimators. Because they depend on total species abundances, they seem to be quite robust despite variations in sample grain size, and when data on abundances are available, ACE and Chao1 show the most precision. However, the precision of abundance-based estimators is dependent on sample coverage (but see Hortal et al. 2006) and both estimators give a lower bound to species richness, thus producing conservative estimates.

Abundance for each OTU was given by the number of similar sequences (using a 5% threshold) plus the numbers recorded in the field for the abundance of unsequenced individuals assigned to each morphospecies. This method potentially underestimates real diversity, as we may have missed some cryptic diversity, but we minimized this problem by assigning individuals to different morphospecies if there were any doubts and by sequencing multiple individuals for abundant and difficult morphospecies. In no case did we recover cryptic genetic lineages within assumed morphospecies, indicating that field identification was effective within each of these locations.

EstimateS (Colwell 2005) was used to test the performance of the diversity estimators according to the number of heads sampled using the Moorea data set. The computation of both Chao 1 and ACE estimates was done for each of one through eight heads sampled, with a randomized order of samples without replacement for 100 runs. A sample-based rarefaction curve was also computed.

Comparison with existing data sets

In addition to blasting against GenBank, we compared our results to the marine invertebrate barcode database of the Moorea Biocode Project (MBP—http://bscit.berkeley.edu/biocode/) in order to evaluate the amount of overlap between the two studies. For the MBP, crustacean species were collected by hand at 1–30 m, from fore-reef and lagoonal habitats, during the summer of 2006 as part of a pilot study to collect genetic vouchers of every species on the island of Moorea, French Polynesia. These data are available in the Barcode of Life Database at www.boldsystems.org in the public project 'MBMIA', Moorea Biocode Marine Inverts A-Crustaceans.

Results

Amplification and sequencing success

From 22 dead *Pocillopora* heads, 403 usable COI sequences were generated from 500 individuals, for an overall success rate of 80.6%; by comparison, the Guelph barcoding center had a success rate of 70.2% (515 of 734) for Moorean crustaceans processed during the MBP pilot project. Of the 97 organisms that were not sequenced, almost half ($N = 46$, 9.2% total) were not successfully amplified and a third ($N = 33$, 6.6% total) produced unreadable sequences (e.g., mixed signal). There was a strong phylogenetic bias, in that 37% of the Caridea failed to sequence. Pseudogenes accounted for one-tenth ($N = 11$, 2.2% total) of the unusable sequences, mainly within the genus Petrolisthes. The remainder ($N = 7$, 1.4%) was removed because of apparent contamination problems (i.e. the sequence did not match the taxonomic group from which the specimen originated).

Diversity and taxonomic distributions

The step function analysis (Fig. 1) shows the number of unique lineages (OTUs) found as a function of increased sequence dissimilarity threshold. For the three data sets (Moorea, the Line Islands and the combined Moorea and the Line Islands), the curves show the same pattern. There is a steep decrease from 0 to 2% representing the coalescent. At around 2% sequence dissimilarity, an inflexion point leads to a plateau that lasts until a threshold value of 14%. The inflexion point represents the switch from intraspecific sequence variability to interspecific sequence variability. Based on this graph and on previous work on other marine invertebrate barcoding projects with better taxonomic control (Meyer and Paulay 2005), a 5% threshold for OTU discrimination was conservatively chosen. Other thresholds ranging from 3 to 14% were applied in order to test for sensitivity with this metric, but the number of OTUs did not vary substantially across this range ($N = 141$ at 3% and $N = 130$ at 14%), as would be expected based on the plateau of the curve.

Fig. 1 Step function analysis of the number of species found in dead *Pocillopora* coral in Moorea, the Line Islands and the Line Islands, and Moorea combined as a function of the cytochrome oxidase subunit I sequence dissimilarity threshold
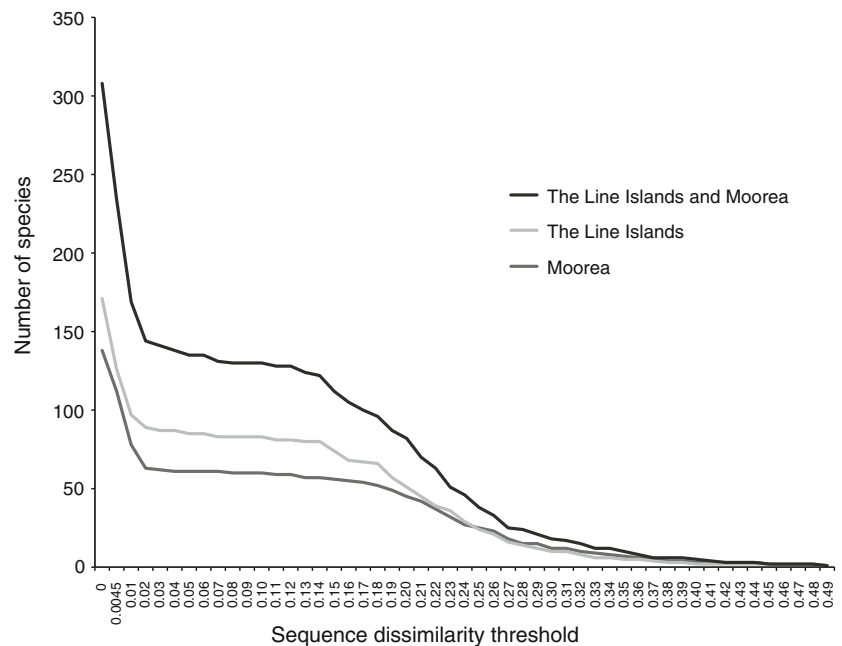
Table 1 Sampling coverage

| Locality	Kingman	Palmyra	Tabuaeran	Kiritimati	N Line Islands	Moorea	Total
No. of heads	4	3	2	5	14	8	22
No. organisms	121	147	83	101	452	337	789
No. tissue samples	85	61	45	68	259	241	500
No. sequences	65	56	41	57	219	184	403
No. assignment	92	125	71	77	365	261	626
No. of OTUs	35	31	26	34	85	61	135

Sampling details for the localities studied: Number of dead *Pocillopora* heads sampled, number of crustaceans counted in the heads, number of organisms for which we have a tissue sample, number of usable sequences, number of organisms assigned to each sequence obtained based on morphological criteria (including sequenced individuals), total number of operational taxonomic units (OTUs)

The 403 individuals sequenced belonged to 135 unique OTUs: 85 in the Northern Line Islands and 61 in Moorea (Table 1). None of the 403 sequences had a sequence identity higher than 91% with any of the GenBank COI sequences. Most sequences (74%) had between 80 and 84% similarity to the most similar GenBank COI sequence (Fig. 2). The lack of a match to GenBank reflects the small number of coral reef crustacean species with published sequences. As of January-28, 2009, the GenBank database had the following numbers of submitted COI sequences in the taxonomic groups recovered from the dead coral heads: Stomatopoda—92; Peracarida—4,741; Caridea—2,782; Anomura—1,167; Brachyura—1,886. However, most collected organisms were morphologically similar to described species, species complexes or genera; of the 108 decapods, 47% ($N = 51$) were identified to species/species complex and an additional 29% ($N = 31$) to genus (details in Electronic Supplementary Material). Two identified species collected from the *Pocillopora* heads had COI sequences in GenBank: *Menaethius monoceros* and

Trapezia rufopunctata. However, none of them matched the GenBank sequences by more than 85%, implying that these taxa are either species complexes (*M. monoceros*) or likely misidentified (*T. rufopunctata* in GenBank).

Most of the crustaceans (86%) collected from the coral heads were decapods (Table 2), which were mostly larger-bodied species. Stomatopods were represented by the fewest species. Peracarids, especially amphipods, were numerous in the Moorea samples but underrepresented in the Line Island samples, probably because they were undersampled in the more difficult, shipboard working conditions. Because decapods were the most reliably collected and also most abundant, we divided subsequent analyses into all Crustacea (135 OTUs) and Decapoda only (108 OTUs).

Most of the OTUs were rare or narrowly distributed. Out of the 135 OTUs, 59 were singletons (i.e., represented by a single individual) (44% of all crustaceans and 33% of the decapods), and 45 were represented by several specimens found from only one island (34% of all crustaceans and 32% of the decapods). Within the Northern Line Islands,
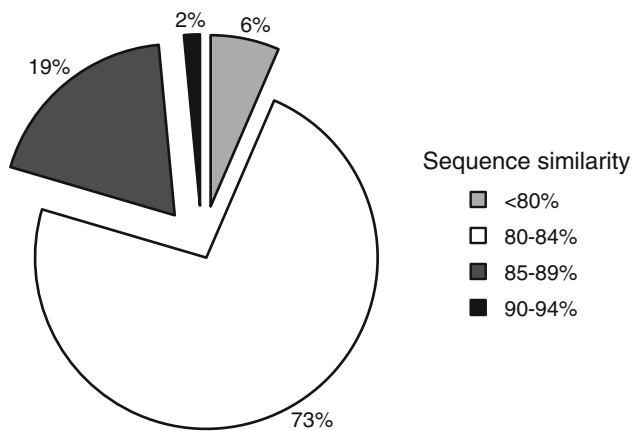
Fig. 2 Distribution of COI sequence identity values between crustaceans living in dead *Pocillopora* heads in the Northern Line Islands and Moorea and the most similar sequence within the GenBank database. The maximum sequence identity observed was 91%

only 17 of 85 OTUs were found on 2 islands, 9 on 3 islands and 2 on all 4 islands (39 were singletons and 18 occurred more than once from just one island). Overlap between the dead head fauna of Moorea and of the Northern Line Islands was even lower; of the 135 crustacean OTUs, only 11 (all decapods) were shared between both localities (Fig. 3). The overlap with the database of the MBP was also relatively low as more than half of the OTUs (35) found in the dead *Pocillopora* heads of Moorea have not been recovered in the MBP sampling. Overall, only 22 OTUs (26%) found in the dead *Pocillopora* heads in the Northern Line Islands have been found in Moorea (*Pocillopora* sampling and the MBP sampling combined).

Richness estimation

Rarefaction curves were constructed to estimate the completeness of sampling effort and, therefore, the reliability of diversity estimates (Fig. 4). None of the rarefaction curves reached a plateau, indicating that the number of individuals sequenced and, therefore, the number of dead *Pocillopora* heads examined was insufficient to estimate reliably the total number of crustacean species within this habitat using these curves.

The Chao 1 (Chao 1984) and ACE (Chao and Lee 1992) species diversity estimates are designed to provide estimates of diversity when many species remain to be sampled (Colwell and Coddington 1994). Irrespective of the

Table 2 Taxon abundance

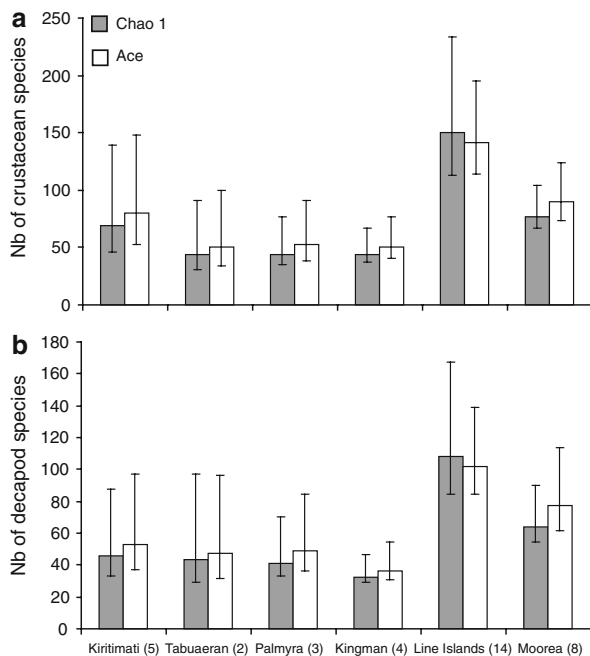|	Kingman	Palmyra	Tabuaeran	Kiritimati	N Line Islands	Moorea	Total
Decapoda							
Anomura	11	47	23	14	95	74	169
Brachyura	42	54	38	56	190	85	275
Caridea	29	23	7	1	60	32	92
Stomatopoda	5	1	0	2	8	2	10
Peracarida	5	0	3	4	12	68	80
Total	92	125	71	77	365	261	626

Number of individuals (assigned to morphospecies) in the different crustacean taxa found in the dead *Pocillopora* heads

Fig. 3 Overlap of species sampled in the dead *Pocillopora* heads in Moorea and the Northern Line Islands and in the Moorea Biocode Project for **a** all crustaceans and **b** decapods only
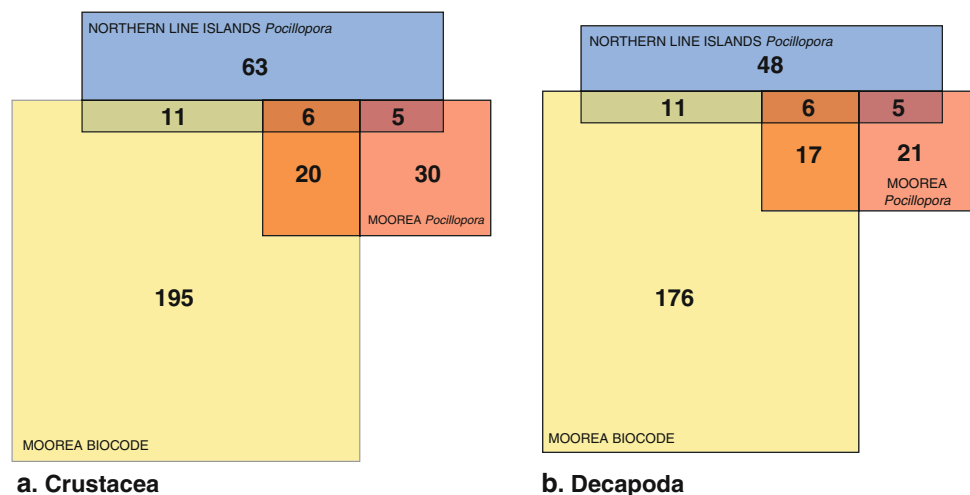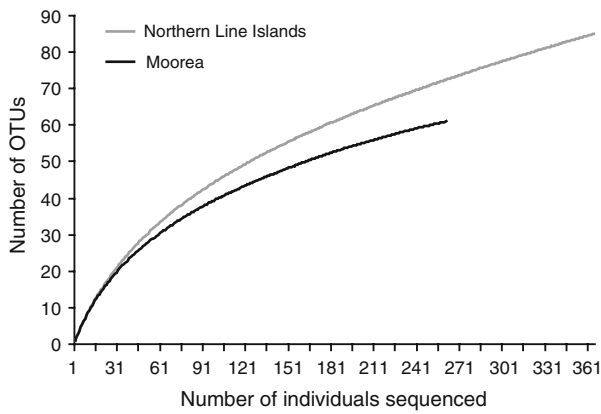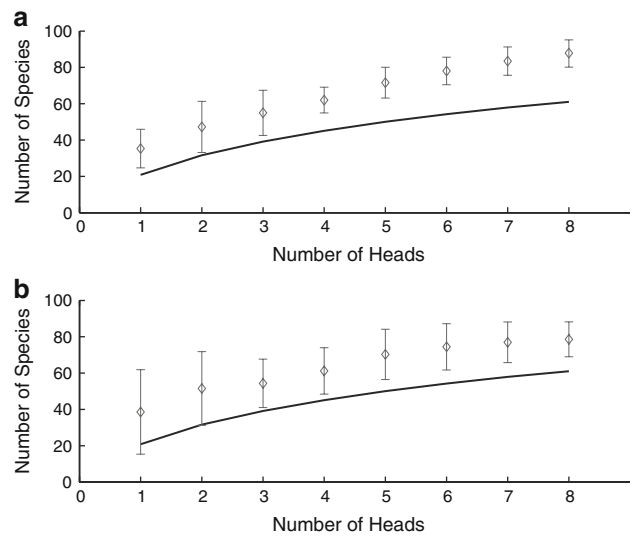
Fig. 4 Rarefaction curves for the crustaceans sampled in dead *Pocillopora* heads in Moorea and in the Northern Line Islands using the furthest neighbor assignment algorithm with partial COI sequences

Fig. 5 Species richness estimates (Chao 1 and ACE) based on the number and frequency of COI gene sequences of **a** crustaceans and **b** decapods in the five localities sampled (Moorea, French Polynesia; Kingman, Palmyra, Tabuaeran and Kiritimati in the Northern Line Islands; number of heads sampled for each in parentheses). Maximum and minimum values were calculated with 95% confidence intervals

Fig. 6 Rarefaction curves (*line*) and values of diversity estimators ACE (**a**) and Chao 1 (**b**) with their 95% confidence intervals for Moorea as a function of the number of heads entered into the analysis

taxa studied (all crustaceans or decapods only), the estimated richness was always higher in Moorea than in any of the four Northern Line Islands sampled (Fig. 5) with an estimated richness of 90 species of crustaceans and 80 species of decapods. In the Northern Line Islands, the highest diversity was found in Kiritimati (80 species of crustaceans, 50 species of decapods), and the lowest diversity was observed in Kingman (44 species of crustaceans, 30 species of decapods), but differences in diversity

among the four Northern Line Islands were not significant. Overall, the estimated diversity in the Northern Line Islands was 150 species of crustaceans and 110 species of decapods.

However, these patterns could have been affected by differences in the number of coral heads sampled from the different islands (eight heads in Moorea vs. two to five heads in the Northern Line Islands). To test this, we examined diversity estimates for all the crustaceans in Moorea as a function of the number of heads included in the statistical analysis. The plots show that although confidence intervals narrow with increasing number of heads included, ACE values do not plateau even after sampling eight heads (Fig. 6a), and Chao1 values only start to plateau after six heads (Fig. 6b).

Discussion

Semi-quantitative sampling and diversity estimators

The standardized, semi-quantitative sampling and molecular techniques applied in this work allowed us to estimate the diversity of the crustacean fauna in a comparable and reproducible way and with greater speed and precision compared to that obtained by previous evaluations based on conventional sampling methods. Conventional collection methods used on reefs, such as hand collecting, play a very important role in investigating and cataloguing the biologic diversity and allow for taxonomic inventories and species discovery. However, these methods are poorly adapted to estimate the likely global diversity of coral reefs

because they are difficult to standardize, which makes it difficult to compare and combine independent estimates. They are also usually biased in favor of the larger, more numerous and easily studied plants and animals (but see Bouchet et al. 2002). Quantitative sampling methods permit more rigorous biodiversity estimates and comparisons of biodiversity among habitats and sites. Thus, they also have the potential to more accurately evaluate the extent to which biodiversity is being lost as a function of anthropogenic reef degradation.

It is extremely challenging to apply quantitative sampling on coral reefs because of their heterogeneous, rigid and complex structure; this makes them unsuitable for quantified sieving, as is used for soft sediments (Markmann and Tautz 2005). Because of the extraordinary diversity of coral reefs, exhaustive inventories of reef-associated fauna also still remain impractical. Thus, developing standardized sampling methods that work for reefs is a high priority.

We focused on the crustacean fauna inhabiting dead *Pocillopora* heads in the Northern Line Islands and in Moorea (French Polynesia). In our sampling, crustaceans accounted for a substantial fraction (30–40%) of the macrofaunal diversity encountered in this habitat, but other groups remain to be analyzed. Moreover, our sampling technique does not allow for a thorough sampling of the microcrustaceans, such as amphipods, which are too small to be easily detected on this heterogeneous surface; quantitative sampling of the microfaunal diversity requires alternative approaches such as the use of a mesh bag for the rubble collection, smaller sizes of sieves and dissecting microscopes for the extraction of the organisms. Therefore, even the crustacean diversity reported here is clearly a substantial underestimate for the coral heads sampled, which are themselves a tiny part of the coral reef habitat as a whole.

This study also highlighted the need to sample more heads of *Pocillopora* in order to have a more precise estimate of the number of species, even for the taxa and habitat type analyzed. In any community, species richness estimates are always tied to sampling effort (Hughes et al. 2001). In this study, the rarefaction curves did not approach a constant value, even for this relatively restricted habitat type (fore-reef, 10 m) (Fig. 4). ACE and especially Chao I estimates performed better, but only for the highest numbers of heads sampled.

This study was based on the use of molecular sequences as a way to determine species diversity. We are aware that species delimitation based on a single molecular marker and the use of a molecular threshold may raise questions about the validity of these results. However, no single approach can provide a definitive conclusion on species boundaries, and challenges exist for any taxonomic and

DNA-based methods. Lefébure et al. (2006) have demonstrated that artifacts for species delimitation due to the use of a sole maternally inherited molecule have little effect in crustaceans. Moreover, intraspecific and congeneric divergences in COI for crustaceans, as shown by Fig. 1, overlap only weakly (Lefébure et al. 2006). Costa et al. (2007), studying the ability of DNA barcodes to provide species level identifications in crustaceans, found that sequence divergence among congeneric species averaged 17.16% whereas intraspecific variation averaged 0.46%, and therefore, that species recognition was straightforward in 95% of the cases. DNA barcoding seems to be a useful tool in the discrimination of crustacean species and in diversity studies in this group. DNA sequence-based identification is especially useful and informative for juveniles and microcrustaceans.

The one-to-one correspondence between adult decapod morphospecies identified on the basis of field appearance and genetic OTUs is encouraging. This indicates that field taxonomic methods at a single locality can be very effective in delineating species, even in those taxa where sibling species are common. However, DNA-based identification is likely to reveal more complex patterns in among-site comparisons, where allopatric sibling species complexes are often encountered (e.g., Meyer and Paulay 2005). These were also evident in our data set; for example, individuals identified as *Perinia tumida* from the Society and Line Islands are deeply divergent. Moreover, standard sequence data (DNA barcodes) permit research teams to cross-correlate morphospecies identified by different workers, which is particularly useful for species, currently without names.

It is now possible to bring state of the art sequencing technology to reefs located in the heart of marine biodiversity. Therefore, DNA barcodes could be used as a real-time guide to traditional surveys.

Comparisons with other studies

Past studies of the cryptofauna inhabiting *Pocillopora* heads have mainly focused on organisms associated with live corals, and crustaceans are also the main component of this fauna. In living *Pocillopora damicornis*, examining 30 to 66 heads, more than 50 species of decapods have been recorded for the Gulf of Panama (Abele 1976; Abele and Patton 1976), 101 species of crustaceans for the Great Barrier Reef (Austin and Austin 1980) and 54 species of crustaceans for Western Australia (Black and Prince 1983). Live corals harbor fewer species than dead corals as they mostly host obligate symbionts, whereas species known from a variety of reef habitats can be found on dead *Pocillopora*. For example, during our sampling campaigns, seven heads of live *Pocillopora* were also examined from

three localities of the Northern Line Islands (Palmyra, Tabuaeran, and Kiritimati). Using the same molecular techniques as described above, we found that the live *Pocillopora* heads harbored at least 28 species of crustaceans (all decapods), compared to 69 spp. of crustaceans and 60 spp. of decapods from dead heads from the same three islands. Of the crustaceans from living *Pocillopora*, 16 also were found in dead *Pocillopora* heads.

One of the striking results of this study is the high proportion of rare species: 44% of all species were singletons, and an additional 33% of species were sampled several times but only from one island. Even the overlap between the crustacean fauna inhabiting the dead *Pocillopora* heads and the crustacean fauna recorded by the Moorea Biocode Project was low (43%), especially considering the intensity of the sampling effort for the latter (~6 weeks of collecting macro-invertebrates at 48 collecting stations). This pattern has been found before with other reef-associated cryptofaunal groups such as isopods (Kensley 1998) and mollusks (Bouchet et al. 2002); in the latter study, 32% of the species collected in a survey of mollusk diversity in New Caledonia were collected at a single station, and 20% of the species were represented by single specimens. Therefore, the bulk of the reef cryptofauna diversity is made up of low-abundance species. Similarly, singletons in tropical arthropods surveys averaged 32% (Coddington et al. 2009), highlighting the need to increase sampling size to obtain robust and reliable species richness estimates.

None of the species sampled in this study matched COI sequences in GenBank [the highest sequence identity being 91%, indicating that they are probably different species within species complexes (e.g. Knowlton et al. 1993)], and underscoring the still limited availability of DNA barcoding data for identifying coral reef invertebrates. The building of a comprehensive COI barcode database for marine invertebrates is a challenge that several large-scale projects are undertaking (e.g., the Marine Barcode of Life and the Moorea Biocode Project). The assemblage of DNA barcodes will allow for an effective identification system that will be very useful in understanding the structure of reef diversity.

Biodiversity in coral reefs is exceptionally high. As an illustrative example, after centuries of taxonomic inventories in one of the most intensively and comprehensively inventoried regions of the world, only 212 species of true crabs (Brachyura) have been listed for European Seas (Bouchet 2006). When we examined 22 dead *Pocillopora* heads, sampled from a single depth in five localities from the central Pacific, we found 65 OTUs that could be identified as brachyurans among the 403 sequences distinguished at the 5% level. This represents ~30% of the recorded brachyuran diversity in Europe, and ~1% of the global brachyuran fauna [6,793 recognized species (Ng et al. 2008)]. The implied challenges for documenting the diversity of species on coral reefs are enormous. Bouchet (2006) estimated that, at the current rate of species description, it would take another 250–1,000 years to complete the inventory of marine biodiversity. Clearly new approaches are needed that are repeatable and cross-comparable, not only because the diversity of reefs is so staggering, but also because we have no idea of how biodiversity on reefs is threatened due to rapid rates of reef degradation. In this effort, standardized sampling coupled with molecular analyses will play a key role.

References

Abele LG (1976) Comparative species composition and relative abundance of decapod crustaceans in marine habitats of Panama. Mar Biol 38:263–278

Abele LG, Patton WK (1976) The size of coral heads and the community biology of associated decapod crustaceans. J Biogeogr 3:35–47

Austin AD, Austin SA (1980) Community structure of the fauna associated with the coral *Pocillopora damicornis* (L.) on the Great Barrier Reef. Aust J Mar Freshw Res 31:163–174

Bellwood DR, Hughes TP, Folke C, Nystrom M (2004) Confronting the coral reef crisis. Nature 429:827–833

Black R, Prince J (1983) Fauna associated with the coral *Pocillopora damicornis* at the southern limit of its distribution in Western Australia. J Biogeogr 10:135–152

Bouchet P (2006) The magnitude of marine biodiversity. In: Duarte C (ed) The exploration of marine biodiversity: scientific and technological challenges. Fundación BBVA, Bilbao, Spain, pp 31–64

Bouchet P, Lozouet P, Maestrati P, Heros V (2002) Assessing the magnitude of species richness in tropical marine environments: exceptionally high numbers of molluscs at a New Caledonia site. Biol J Linn Soc 75:421–436

Bruno JF, Selig ER (2007) Regional decline of coral cover in the Indo-Pacific: Timing, extent, and subregional comparisons. PLoS ONE 8:e711

Chao A (1984) Non-parametric estimation of the number of classes in a population. Scand J Stat 11:265–270

Chao A, Lee SM (1992) Estimating the number of classes via sample coverage. J Am Stat Assoc 87:210–217

Coddington JA, Agnarsson I, Miller JA, Kuntner M, Hormiga G (2009) Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys. J Anim Ecol 78:573–584

Colwell RK (2005) EstimateS: statistical estimation of species richness and shared species from samples. Version 7.5. User's guide and application published at: http://purl.oclc.org/estimates

Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. Philos Trans R Soc Lond B 345:101–118

Costa FO, deWaard JR, Boutillier J, Ratnasingham S, Dooh RT, Hajibabaei M, Hebert PDN (2007) Biological identifications through DNA barcodes: the case of the Crustacea. Can J Fish Aquat Sci 64:272–295

Felsenstein J (1989) PHYLIP—phylogeny inference package (version 3.2). Cladistics 5:164–166

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3:294–299

Gardner TA, Cote IM, Gill JA, Grant A, Watkinson AR (2003) Long-term region-wide declines in Caribbean corals. Science 301:958–960

Grassle J, Maciolek NJ (1992) Deep-sea species richness: regional and local diversity estimates from quantitative bottom samples. Am Nat 139:313–341

Groombridge B, Jenkins MD (2000) Global biodiversity: earth's living resources in the 21st century. World Conservation Monitoring Centre, Cambridge, United Kingdom

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc, B 270:313–321

Hoeksema BW (2007) Delineation of the Indo-Malayan centre of maximum marine biodiversity: the coral triangle. In: Renema W (ed) Biogeography, time, and place: distributions, barriers and islands. Springer, Netherlands, pp 117–178

Hortal J, Borges PAV, Gaspar C (2006) Evaluating the performance of species richness estimators: sensitivity to sample grain size. J Anim Ecol 75:274–287

Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. Appl Environ Microbiol 67:4399–4406

Hughes TP, Bellwood DR, Connolly SR (2002) Biodiversity hotspots, centres of endemicity, and the conservation of coral reefs. Ecol Lett 5:775–784

Idjadi JA, Edmunds PJ (2006) Scleractinian corals as facilitators for other invertebrates on a Caribbean reef. Mar Ecol Prog Ser 319:117–127

Karlson RH, Cornell HV, Hughes TP (2004) Coral communities are regionally enriched along an oceanic biodiversity gradient. Nature 429:867–870

Kensley B (1998) Estimate of species diversity of free-living marine isopod crustaceans on coral reefs. Coral Reefs 17:83–88

Knowlton N (1993) Sibling species in the sea. Annu Rev Ecol Syst 24:189–216

Knowlton N (2000) Molecular genetic analyses of species boundaries in the sea. Hydrobiologia 420:73–90

Knowlton N (2001) The future of coral reefs. Proc Natl Acad Sci USA 98:5419–5425

Knowlton N, Jackson JBC (2008) Shifting baselines, local impacts, and global change on coral reefs. PLoS Biology 6:e54

Knowlton N, Weigt LA, Solorzano LA, Mills DK, Bermingham E (1993) Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the Isthmus of Panama. Science 260:1629–1632

Lambshead JD, Boucher G (2003) Marine nematode deep-sea biodiversity - hyperdiverse or hype. J Biogeogr 30:475–485

Lefébure T, Douady CJ, Gouy M, Gibert J (2006) Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. Mol Phyl Evol 40:435–447

Maddison WP, Maddison DR (2000) MacClade. Sinauer Associates, Sunderland, MA

Markmann M, Tautz D (2005) Reverse taxonomy: an approach towards determining the diversity of meiobenthic organisms based on ribosomal RNA signature sequences. Philos Trans R Soc Lond B Biol Sci 360:1917–1924

May RM (1994) Biological diversity: differences between land and sea. Philos Trans R Soc Lond B Biol Sci 343:105–111

Meyer CP, Paulay G (2005) DNA Barcoding: error rates based on comprehensive sampling. PLoS Biol 3:2229–2238

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403:853–858

Ng PKL, Guinot D, Davie PJF (2008) Systema Brachyururum: part I. An annotated checklist of extant brachyuran crabs of the world. Raffles B Zool 17:1–286

Pandolfi JM, Bradbury RH, Sala E, Hughes TP, Bjorndal KA, Cooke RG, McArdle D, McClenachan L, Newman MJH, Paredes G, Warner RR, Jackson JBC (2003) Global trajectories of the long-term decline of coral reef ecosystems. Science 301:955–958

Paulay G (2003) Marine biodiversity of Guam and the Marianas: overview. Micronesica 35–36:3–25

Reaka-Kudla M (1997) The global biodiversity of coral reefs: a comparison with rain forests. In: Reaka-Kudla M, Wilson DE, Wilson EO (eds) Biodiversity II: understanding and protecting our biological resources. Joseph Henry Press, Washington, D.C., pp 83–108

Roberts CM, McClean CJ, Veron JEN, Hawkins JP, Allen GR, McAllister DE, Mittermeier CG, Schueler FW, Spalding M, Wells F, Vynne C, Werner TB (2002) Marine biodiversity hotspots and conservation priorities for tropical reefs. Science 295:1280–1284

Sala E, Knowlton N (2006) Global marine biodiversity trends. Annu Rev Environ Resour 31:93–122

Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 71:1501–1506

Small A, Adey A, Spoon D (1998) Are current estimates of coral reef biodiversity too low? The view through the window of a microcosm. Atoll Res Bull 458:1–20

Veron JEN (1995) Corals in space and time. University of New South Wales Press, Townsville, Qld