From marine park to future genomic observatory? Enhancing marine biodiversity assessments using a biocode approach

Yin Cheong Aden Ip[‡], Ywee Chieh Tay[§,|], Su Xuan Gan[‡], Hui Ping Ang[¶], Karenne Tun[¶], Loke Ming Chou[‡,#], Danwei Huang[‡,#], Rudolf Meier[‡,#]

‡ Department of Biological Sciences, National University of Singapore, Singapore, Singapore
§ National University of Singapore, Singapore, Singapore
| Temasek Life Sciences Laboratory, Singapore, Singapore
¶ National Parks Board, Singapore, Singapore
# Tropical Marine Science Institute, National University of Singapore, Singapore, Singapore

Corresponding author: Ywee Chieh Tay (yweechieh@gmail.com)

Abstract

Few tropical marine sites have been thoroughly characterised for their animal species, even though they constitute the largest proportion of multicellular diversity. A number of focused biodiversity sampling programmes have amassed immense collections to address this shortfall, but obstacles remain due to the lack of identification tools and large proportion of undescribed species globally. These problems can be partially addressed with DNA barcodes ("biocodes"), which have the potential to facilitate the estimation of species diversity and identify animals to named species via barcode databases. Here, we present the first results of what is intended to be a sustained, systematic study of the marine fauna of Singapore's first marine park, reporting more than 365 animal species, determined based on DNA barcodes and/or morphology represented by 931 specimens (367 zooplankton, 564 macrofauna including 36 fish). Due to the lack of morphological and molecular identification tools, only a small proportion could be identified to species solely based on either morphology (24.5%) or barcodes (24.6%). Estimation of species numbers

for some taxa was difficult because of the lack of sufficiently clear barcoding gaps. The specimens were imaged and added to "Biodiversity of Singapore" (http://singapore.biodiversity.online), which now contains images for > 13,000 species occurring in the country.

Keywords

DNA barcoding, marine park, genomic observatory, COI, biocodes
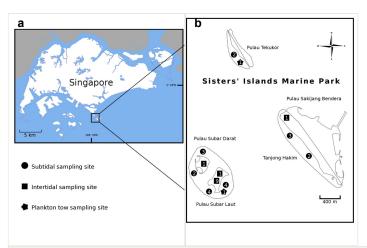
Introduction

In recent decades, it has become clear that biodiversity loss is an increasingly serious problem and many species are expected to become extinct before discovery and description (Costello et al. 2013, Laurance 2013). It is thought that only 226,000 of the estimated 0.7–1 million marine species have been described (Appeltans et al. 2012). Poor sampling of marine fauna in biodiverse regions and a large backlog of species that have yet to be described have rendered most marine species unidentifiable and often unknown to science (Mora et al. 2011, Mora et al. 2013). This incomplete knowledge of species diversity prevents accurate biodiversity assessments and monitoring and limits our understanding of ecosystem functioning (Isaac et al. 2004). Determining species diversity, using traditional taxonomic techniques, requires skilled taxonomists to accurately identify or describe species based on detailed keys and careful study of morphology. However, this approach is manpower-intensive, slow (Miller 2007) and costly (approximately $39,000–122,000 per species; Carbayo and Marques 2011). Consequently, alternative strategies are being developed to expedite the processes of species discovery and delimitation (Wang et al. 2018). While convincing solutions for large-scale species description are lacking, the problem is starting to attract the attention of many animal taxonomists (Riedel et al. 2013).

Molecular techniques have dramatically increased the rate of species discovery and facilitated species identification for those species that have been barcoded. DNA barcoding was initially proposed as a means to identify animal species, although it is now increasingly used for species discovery (Hebert et al. 2003, Ratnasingham and Hebert 2007, Hajibabaei et al. 2007, Goldstein and DeSalle 2011). This technique uses a short DNA fragment as a standard marker for species description and discovery. For metazoans, the mitochondrial cytochrome c oxidase subunit I (COI) is the barcoding locus of choice, having been popularised by Hebert et al. 2003. However, as a prerequisite to successful species identification via barcodes, a comprehensively curated reference database is required (Ekrem et al. 2007). Public databases, such as GenBank (Benson et al. 2018) and the Barcode of Life Data (BOLD) System (Ratnasingham and Hebert 2007), collect such reference sequences and, while they contain many sequences for vertebrates (Hebert et al. 2004, Ward et al. 2005, Kerr et al. 2007), the coverage for invertebrate species is more limited (Barrett and Hebert 2005, Hajibabaei et al. 2006, Elias et al. 2007, Grant and Linse 2009, Hausmann et al. 2011, Park et al. 2011). These databases are continually updated

as new discoveries are made and presently the BOLD System contains > 5 million barcode sequences belonging to 262,679 species (as of December 2017). However, these identifications should be interpreted with caution, because many are for predicted species (Barcode Index Numbers; "BINS") and few specimens have been formally identified or verified as species (Kwong et al. 2012). Despite these potential issues, research is uncovering unexpected diversity for many taxa in many habitats (Hebert et al. 2004, Brower 2006, Witt et al. 2006, Havermans et al. 2011). These discoveries further reinforce the utility of DNA barcoding in surveys of biodiversity (Smith et al. 2005).

For highly biodiverse regions such as Southeast Asia, these global reference databases remain particularly incomplete and poorly curated (Giam et al. 2010, Jinbo et al. 2011), partially due to the prohibitive costs associated with molecular sequencing (Meier 2008). The sheer number of species found in biodiversity hotspots also poses a considerable challenge, as many of the barcodes recovered differ from those in the databases by more than 3%, meaning accurate species identification is not possible for these animals (Kwong et al. 2012) and several of which may also be new species (Wong et al. 2011, Wang et al. 2018). These problems will only be resolved by greater sampling effort and expansion of curated databases. Unfortunately, a shortage of taxonomic expertise for biodiverse regions is compounded by a lack of the necessary skills required to perform the field collections, a consequence of the decline in biodiversity appreciation (Ríos-Saldaña et al. 2018). It is here that digital reference collections, such as "Biodiversity of Singapore" (https://singapore.biodiversity.online/), can make a difference by helping to stimulate an interest in biodiversity and conservation, with verifying putative species identifications (Will et al. 2005, Ang et al. 2013, Chan et al. 2014).

Singapore is situated just outside the southwest corner of the biodiverse Coral Triangle biodiversity hotspot and, like other countries in the region, its marine biodiversity remains relatively poorly understood. To address this shortfall, we describe the first results of a programme that aims to build a comprehensive animal species identification database for Singapore's first marine park—the Sisters' Islands Marine Park (SIMP; Fig. 1). Prior to this study, preliminary estimates suggested that the SIMP may be home to > 100 fish and > 1,000 macroinvertebrate species (K. Tun, pers. obs.). Barcoding the fauna is part of a larger initiative to make Singapore's biodiversity identifiable with molecular tools and constitutes the first steps towards building a national genomic observatory (Davies et al. 2014). Located approximately 6 km south of mainland Singapore, SIMP is recognised as a locally important area of biodiversity in terms of coral species richness, functional and phylogenetic diversity (Wong et al. 2018). Previous work modelling coral larval dispersal has indicated that reefs in the area are potentially strong source reefs that can seed other reefs in Singapore (Tay et al. 2012, Chang 2015). Furthermore, the marine park also aims to serve as an outreach platform to encourage public interest in marine life and participation in biodiversity conservation. It is open to the public and visitors can interact with the natural environment via guided intertidal walks and subtidal dive trails. Thus, a systematic and regular documentation of the biodiversity at the SIMP is important for the management and conservation of marine ecosystems in Singapore. SIMP was established as a marine park on 15 July 2015 and spans an area of 40 hectares that encompasses the

namesake Sisters' Islands (Pulau Subar Laut and Pulau Subar Darat), the western shores of Pulau Tekukor and Pulau Sakijang Bendera (Fig. 1).

Figure 1.

Map depicting the intertidal and subtidal sampling sites in the Sisters' Islands Marine Park (SIMP), Singapore.

Dotted lines define the SIMP's boundaries. Number of sampling events per site are indicated within the sampling event icons in the inset map. Adapted from http://commons.wikimedia.org/wiki/File:Singapore_Outline.svg.

a: Singapore `doi`
b: Sisters' Islands Marine Park `doi`

The work performed here will help consolidate sampling records and molecular data obtained from the SIMP will form an important baseline for monitoring Singapore's marine species. It will also provide better and more complete understanding of marine biodiversity in Singapore, with further utility throughout Southeast Asia where work of this nature is still in its infancy and which is inadequately represented in global databases (Koh and Sodhi 2010, Webb et al. 2010). Finally, it serves as a resource for future work relying on curated databases for species detection and discovery of species interactions, such as environmental DNA (eDNA) (Ficetola et al. 2008, Jerde et al. 2011, Thomsen et al. 2011, Yamamoto et al. 2017, Djurhuus et al. 2018). Given the rapid increase in interest in using molecular data for environmental monitoring in recent years (Harper et al. 2018, Darling 2019, Rey et al. 2019), these will likely have crucial applications for the conservation and management of marine resources.

Materials and Methods

Literature review

We first compiled existing records and published DNA barcodes relevant to SIMP through a literature keyword search for marine fauna found at the SIMP. Records for SIMP species

predominantly came from two survey projects aimed at documenting and/or discovering local biodiversity (without DNA barcodes): (i) a large-scale 'Comprehensive Marine Biodiversity Survey' of Singapore (Tan et al. 2013), which sampled marine fauna across Singapore using many different sampling methods, from hand sampling to dredging and (ii) a BioBlitz initiative by the National Parks Board of Singapore, which is a series of visual surveys involving volunteer scientists and members of the public (Suppl. material 1) that documented, for example, 105 scleractinian coral species at the SIMP. Barcode data for some of the species on the list were obtained via GenBank (109 species; Suppl. material 2) and added to the SIMP database.

Macrofauna

Field collection

Samples were collected from all four islands of the SIMP:

1. Pulau Subar Laut (Big Sister's Island; 1.21417°N, 103.83444°E),
2. Pulau Subar Darat (Small Sister's Island; 1.215788°N, 103.832705°E),
3. Pulau Sakijang Bendera (specifically Tanjong Hakim; 1.213823°N, 103.851107°E) and
4. Pulau Tekukor (1.232139°N, 103.836604°E).

Collections were authorised by the National Parks Board (permit number NP/RP15-088) and were carried out at the accessible intertidal reef, sandy beach, seawall and much of the shallow subtidal reef areas (Fig. 1) over a span of two years from July 2015 to July 2017.

Intertidal specimens were obtained using hand tools and nets during low spring tides, 0.0 m to 0.2 m above chart datum. These tools were likewise used for subtidal sampling via SCUBA diving to depths of up to 15 m. The search included around, under and inside potential hideouts. Any metazoans encountered during these visual surveys that were not already in our collection, were collected. Fish were collected using two 'bubu' traps, each measuring 0.072 m^3, deployed twice, for periods of one day each, during the sampling period. Up to three individuals of each species were collected, avoiding gravid females and juveniles to reduce sampling impact on natural populations.

Sample processing and imaging

Samples were provisionally imaged *in situ* using a Canon Powershot G10 (Canon Inc., Japan) or Olympus Stylus Tough TG-4 compact camera (Olympus Corporation, Japan). In the laboratory, invertebrate specimens were relaxed in 7.5% (w/v) $MgCl_2$ buffered in seawater (Messenger et al. 1985), while fish specimens were handled according to NUS Institutional Animal Care and Use Committee (IACUC) guidelines (IACUC Protocol B15-1403). A Canon EOS 750D or a dissecting microscope (Leica S8 APO with Canon EOS 750D mounted; 1–8× magnification) was used for specimen imaging. Tissue subsamples were then taken from each specimen before fixation or preservation. Hard-

bodied specimens were preserved in 70% (v/v) molecular grade ethanol, while soft-bodied organisms were first fixed in 4% (v/v) formaldehyde overnight, then transferred to 70% ethanol for long-term preservation. All macrofaunal vouchers were deposited at the Zoological Reference Collection (ZRC) of the Lee Kong Chian Natural History Museum (LKCNHM) as voucher specimens (Suppl. material 3) and the available image data made available online at the "Biodiversity of Singapore", a digital reference collection for Singapore's biodiversity (http://singapore.biodiversity.online; Fig. 2).

Figure 2.

Example of a screenshot of the SIMP collection in the Biodiversity of Singapore portal.

a: Animal taxa are organised by taxonomic identity doi

b: A thumbnail is available for each taxon that links to an individual web page with more detailed information and/or photos   doi

Tissue subsampling, digestion and DNA extraction

For each large soft-bodied specimen, a small piece of tissue (20–40 mm3) was excised, while for each arthropod, one to two legs from the same side of the body were detached for DNA extraction. The tissues were digested overnight at 55°C in 900 µl CTAB (hexadecyltrimethylammonium bromide) with 0.4 mg proteinase K, after which DNA was purified by phase separation with phenol: chloroform: isoamyl-alcohol (25:24:1).

COI barcode amplification

The COI gene region was amplified using different primer pairs described in Folmer et al. 1994, Leray et al. 2013, Lobo et al. 2013, henceforth referred to as the 'Folmer', 'Leray' and 'Lobo' primers, respectively. Reactions were performed using one of three mixes:

1. BioReady rTaq DNA polymerase, 1× reaction buffer (v/v) (Bulldog Bio Inc., China) with the Folmer primer pair targeting the 658-bp barcode region of the COI gene;
2. GoTaq® DNA polymerase (Promega Corporation, U.S.A.) with the Lobo primer pair amplifying the same 658-bp COI region; or
3. GoTaq® Green Master Mix (Promega Corporation, U.S.A.) with the 'Lobo reverse and Leray forward' primer combination targeting a shorter 313-bp COI region for samples that were particularly challenging to amplify.

Most of the macrofaunal samples were subjected to Sanger barcoding. Each 12.5-µl reaction contained 0.5 µM of each primer (uniquely tagged primers for 46 samples only; untagged for the rest), 0.5 µg BSA (bovine serum albumin), 2 µl template DNA and 1× GoTaq®/BioReady rTaq DNA polymerase and reagents mastermix (v/v), according to the manufacturer's recommendations. The thermal cycling profile for (1) using the Folmer primer pair was 94°C for 60 s; 35 cycles of denaturation at 94°C for 45 s, annealing at 48°C for 45 s, extension at 72°C for 90 s; and a final extension at 72°C for 3 mins. The thermal cycling profile for (2) and (3), using the Lobo primers, included a step-up annealing profile of 94°C for 60 s; 5 cycles of 94°C for 30 s, 48°C for 120 s, 72°C for 60 s; 35 cycles of 94°C for 30 s, 54°C for 120 s, 72°C for 60 s; and 72°C for 5 mins.

DNA barcoding using Sanger barcoding

Successful PCR amplicons were purified using SureClean Plus (Bioline Inc., London, UK) and prepared for Sanger sequencing using the BigDye Terminator Cycle Sequencing Kit v. 1.1 and PureSEQ (Aline Biosciences), on an Applied Biosystems 3730XL DNA Analyzer (Thermo Fisher Scientific, U.S.A.), following the manufacturer's instructions. COI barcodes, obtained via Sanger sequencing, were assembled and edited using Geneious R11 v11.0.2 (Biomatters Limited) (Kearse et al. 2012). Although the cost of generating DNA barcodes with Sanger sequencing is expensive (Meier et al. 2015), it was used for barcoding most macrofaunal samples since collections were conducted in numerous small batches (Suppl. material 3) that were too small for cost-effective barcoding via high-throughput sequencing (HTS).

Zooplankton

Field collection

Sampling was performed at sites 1, 3 and 4 listed in the macrofaunal field collection section. A vertical plankton tow with a 100-μm mesh net was used to collect micro- and mesozooplankton (Sieburth et al. 2003) upwards from a depth of 8 m. Zooplankton were concentrated into a 50-ml bottle of seawater from each tow, put on ice and brought back to the laboratory for processing.

Sample processing and imaging

Samples were concentrated through a 100-μm sieve, preserved in 70% ethanol and stored at -30°C prior to sample sorting and imaging. Sorting and imaging were performed under a dissecting microscope (Leica S8 APO with Canon EOS 750D mounted; 1–8× magnification), using soft fine forceps. Specimen identification followed Johnson and Allen (2005) and samples were preliminarily grouped into seven morphotypes at the phylum level (Arthropoda, Annelida, Chaetognatha, Chordata, Cnidaria, Mollusca and Platyhelminthes). Each zooplankton was processed and stored individually in 70% ethanol in 96-well plates at -30°C. Zooplankton identities were later confirmed using DNA barcodes.

Tissue subsampling, digestion and DNA extraction

For larger arthropods, one or two legs from the same side of the body were detached for DNA extraction. For specimens < 5 mm in size, whole individuals were either used for phenol-chloroform extraction or were incubated in 20 μl of 2×-diluted QuickExtract TM DNA extraction solution (Epicenter, BuccalAmp TM) , following the manufacturer's instructions.

COI barcode amplification

Forty-six macrofaunal samples, along with all zooplankton samples, were sequenced using high-throughput sequencing (HTS; Suppl. material 4). PCR was performed on genomic DNA extracted using QuickExtract (Epicenter, BuccalAmpTM) or directly on selected samples for improved time efficiency (Wong et al. 2014). A short 313-bp fragment of the COI gene was targeted using either the mlCO1intF and rmHCO2198 primer pair (Folmer et al. 1994, Meier et al. 2015) or reaction mix (3) (see section on macrofaunal COI barcode amplification) with forward and reverse primers that were uniquely labelled with 9-bp tags (generated with online freeware "Barcode generator"; http://comailab.genomecenter. ucdavis.edu/index.php/Barcode_generator) that differed from one another by ≥ 3-bp (Meier et al. 2015). Each 20-μl PCR reaction contained 1× GoTaq® Green Master Mix, 0.5 μM of each uniquely labelled primer and 2 μl of DNA extract. PCR thermal cycling conditions were as follows: an initial denaturation step at 94°C for 60 s, followed by 35 cycles of 94°C for 60 s, 47°C for 120 s, 72°C for 60 s, and 72°C for 3 mins.

DNA barcoding using high-throughput sequencing

DNA barcoding via HTS ("HTS barcoding"; Wang et al. 2018) can be used to process a large number of specimens (e.g. zooplankton from bulk samples), using a reverse workflow where all specimens are barcoded and pre-sorted into MOTUs that are considered putative species awaiting verification by taxonomic experts (Wang et al. 2018, Gan et al. 2019). HTS barcoding is faster and more than one order of magnitude cheaper than Sanger barcodes (Meier et al. 2015). Tagged amplicons were pooled into four libraries (NEBNext ® UltraTM II DNA Library Prep) for sequencing over five lanes of the Illumina MiSeq platform (v3; 2 × 300 bp; 25 million single reads). Note that these samples only took up 0.2% to 1.6% of each lane. HTS COI barcodes obtained via Illumina MiSeq sequencing were retrieved following the pipeline described in Meier et al. 2015. Briefly, paired-end read data were assembled using PEAR version 0.9.6 (Zhang et al. 2013), data for individual samples were demultiplexed and dominant read sets per sample were identified. To ensure an accurate barcode database, the data were subsequently filtered for sequencing coverage > 50, then filtered by a total barcode count of > 10 and finally against potential contamination by retaining data where the dominant read set was at least four times as abundant as the second dominant read set for each sample (i.e. ratio of coverages of second: first dominant read ≤ 0.2). Samples, for which this ratio was < 0.35, were further evaluated to assess if their sequences could still be used (i.e. if barcodes were consistent with morphology).

Matching zooplankton barcode identities to morphotype data

Four criteria (C1–4) were used to select barcodes that we considered reliable. C1: Zooplankton morphotypes and barcode identities were congruent and samples had a good match (≥ 97%; giving species level identity) to global databases. C2: Barcode had a poor match (> 85%; giving lowest taxonomic identity), but the match was consistent with the morphological sort. C3: In order to accommodate mistakes that may be made during the initial sort of zooplankton, we kept sequences for specimens, even when the morphotypes and barcode identities were incongruent as long as the BLAST match to an existing species in GenBank was high (≥ 97%) and the specimen images were consistent with the BLAST matches. C4: Specimens that failed to yield a barcode due to the violation of filtering thresholds were re-evaluated and retained when all of the following criteria were fulfilled:

- The ratio of first to second dominant read was 0.2–0.35;
- Sequencing coverage of > 50 reads and total barcode count of > 10;
- The dominant read was ≥ 85% match to a taxon that was congruent to morphotype data (see C1); and
- The second dominant read did not match the preliminary assigned morphotype.

Sequence data analysis

All sequence data were aligned using MUSCLE 3.8.425 (Edgar 2004), translated and screened for stop codons using Geneious R11 v11.0.2 (Biomatters Limited) (Kearse et al. 2012). Only one specimen, *Peronia verrculata* (IP0136), was found to have a deletion of

one codon compared to all other sequences. This deletion was confirmed against 63 other *P. verruculata* specimens previously collected in Singapore (Chang et al. 2018). Data were subject to a final round of contamination check by comparing BLAST (Altschul 1990) matches against GenBank (Benson et al. 2018) and the Barcode of Life Data (BOLD) System (Ratnasingham and Hebert 2007), for consistency with specimen morphology. BLAST matches were used to identify the specimens where query cover was ≥ 80%. The thresholds used were ≥ 80% identity for family, ≥ 90% identity for genus and ≥ 97% identity for species identification. It must be noted that these identities at family and genus levels were conservative, based on general interspecific distances which have been observed amongst various marine taxa (Sun et al. 2016, Trivedi et al. 2015), but should be treated with caution as supraspecific ranks are not consistently defined (Bertrand et al. 2006, Kuntner and Agnarsson 2006). Finally, an objective clustering method, with internal gaps treated as a fifth character, was used to group sequences into MOTUs, based on variable pairwise sequence similarities (Srivathsan, unpublished software; implementation of objective clustering described in Meier et al. 2006). One of the barcodes (IP0303), was excluded from this step due to a large internal gap that interferes with clustering.

Results

Sampling at Sisters' Islands Marine Park (SIMP) yielded 931 specimens, comprising 564 macrofauna (benthic and fish) and 367 zooplankton specimens (Figs 3, 4, Suppl. materials 3, 4). Of all the specimens, 24.5% were identified by morphology to species, based on literature. They were included in a database of ca. 150 species that was assembled prior to barcoding. Most of the sampled fauna belonged to arthropods (38.9% of 931 samples) and molluscs (21.0%), followed by cnidarians (9.5%) and annelids (8.3%) (Fig. 5). Fish were the least represented, with only 36 specimens (3.9%).

Overall, COI amplification success was 68.0% across all phyla (633 out of 931 samples). A total of 297 of the sample barcodes (46.9%) were ≥ 658-bp in length (long; average length 677-bp), while 336 samples (53.1%) had sequence lengths varying between 229- and 657-bp (short; average length 350-bp) (Suppl. materials 3, 4). No deletions, insertions or stop codons were observed in any of the COI sequences, suggesting that the data did not include nuclear DNA sequences, originating from mitochondrial DNA (NUMTs) (Bensasson 2001).

Amplification and sequencing success were variable across different phyla and primer pair combinations. Molluscs were generally easy to amplify, while echinoderms were challenging and required more PCR optimisation. Specifically, primer pairs in reaction mix (1) yielded approximately 50% amplification success, reaction mix (2) gave approximately 80% success and reaction mix (3) yielded the highest amplification success at ≥ 95%.

PCR amplification success for zooplankton samples was 70.6% (of 367 samples) and 259 tagged amplicons were sequenced using high-throughput sequencing (HTS) barcoding. Due to uneven amplicon pooling, data for only 191 of these amplicons were retrieved, for which 411,201 reads were demultiplexed and sequence quality filters resulted in coverage

of 19 to 4,048 reads per barcode. Overall, sequencing success was moderate, with 174 out of 259 samples (67.2%) passing all filtering criteria. Of those failing the criteria, nine specimens were retained following criteria C4.
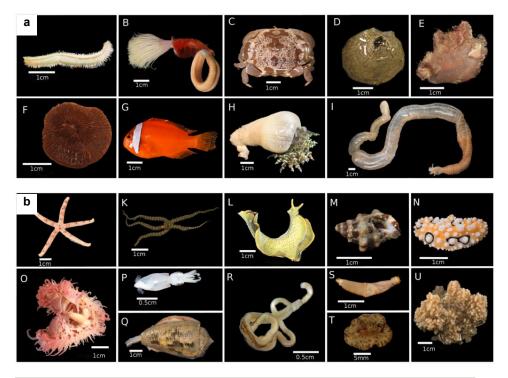

Figure 3.

Representative images of all sampled benthic and fish phyla except Acoelomorpha.

Sample numbers and corresponding museum codes (ZRC) are indicated, where available.

a: (A, B) Phylum Annelida. (A) *Eurythoe* sp., IP0133; (B) *Protula* sp., IP0315. (C, D) Phylum Arthropoda. (C) *Atergatis floridus*, IP0450; (D) *Tetraclita squamosa*, IP0106 [ZRC 2017.1114]. (E, G) Phylum Chordata. (E) F. Pyuridae, IP0070; (G) *Amphiprion frenatus*, IP0479. (F, H) Phylum Cnidaria. (F) *Lithophyllon scabra*, IP0336; (H) *Phymanthus* sp., IP0122 [ZRC.CNI.1249]. doi

b: (I, J, K) Phylum Echinodermata. (I) F. Synaptidae, IP0142; (J) *Nepanthia* sp., IP0321 [ZRC.ECH.1253]; (K) *Ophiactis* sp., IP0207 [ZRC.ECH.1243]. (L, M, N, O, P, Q) Phylum Mollusca. (L) *Elysia ornata*, IP0269 [ZRC.MOL.010720]; (M) *Tenguella* sp., IP0096 [ZRC.MOL.010698]; (N) *Phyllidia ocellata*, IP0192; (O) *Limaria* sp., IP0263 [ZRC.MOL.010718]; (P) *Idiosepius pygmaeus*, IP0115 [ZRC.MOL.010704]; (Q) F. Conidae, IP0144. (R) Phylum Nemertea, IP0108 [ZRC.MIS.0006]. (S) Phylum Sipuncula, IP0085 [ZRC.SIP.0030]. (T) Phylum Platyhelminthes. *Pseudobiceros damawan*, IP0447. (U) Phylum Porifera, IP0396.   doi
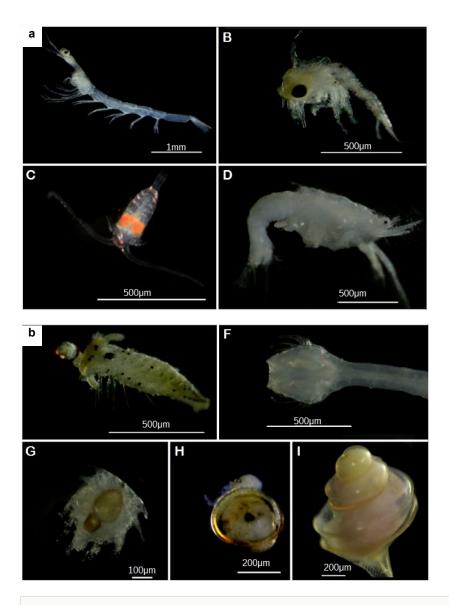
Figure 4.

Representative images of major zooplankton morphotypes sampled.

a: (A, B, C, D, G) Phylum Arthropoda. (A) F. Luciferidae, ZP024 (Naomi et al. 2006); (B) Order Decapoda, ZP328; (C) Order Calanoida, unsequenced live copepod; (D) *Acetes indicus*, ZP332; (G) *Tetraclita singaporensis*, ZP277;   doi

b: (E) Phylum Annelida. *Polydora aura*, ZP288; (F) Phylum Chaetognatha. F. Sagittidae, ZP278; (H, I) Phylum Mollusca. (H) *Dendostrea frons*, ZP016; (I) F. Turridae, ZP312.   doi
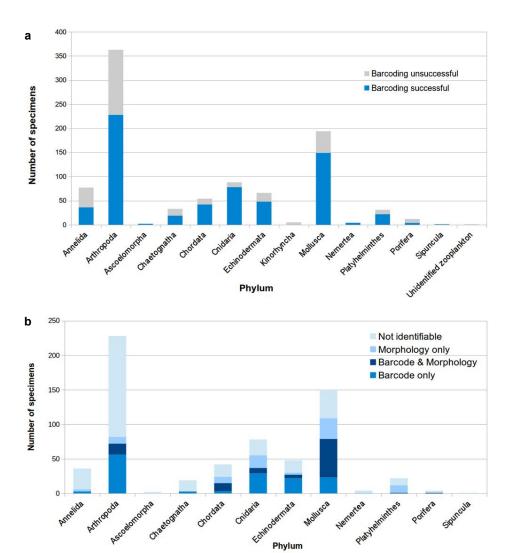
Figure 5.

Summary of specimen collection, DNA barcoding and species identification successes across phyla.

a: Barcoding success across specimens collected. Numbers of specimens for which COI barcoding was successful and not successful, are indicated by blue and grey shades, respectively. doi

b: Identification success to species level for specimens, for which barcodes were successfully obtained. Numbers of specimens identified based only on morphology, only on barcode matches or via both methods which were congruent, are represented by different shades of blue. doi

Well-studied and morphologically distinct groups such as corals, sea anemones, echinoderms, molluscs and crustaceans were easily recognised, but most specimens could not be identified to species (75.3% unidentifiable; only 230 specimens were identifiable to 155 species by morphology) without DNA barcode-assisted identification. DNA barcodes obtained for 633 specimens clustered into 351–395 species dependent on clustering criterion (i.e. MOTUs), of which 83 specimens were identifiable only via DNA barcode (48 species). This adds up to an approximately 36% increase in the number of specimens that could be delimited to at least species level.

The final set of 633 COI specimen barcodes obtained clustered into 351 molecular operational taxonomic units (MOTUs, i.e. putative species). This was based on a species delimitation threshold of 3%, which was defined by assessing the percentage pairwise differences across all sampled taxon groups (Suppl. material 5). Clustering at 3% provides a conservative estimate of species, especially for Anthozoa—there may be up to 42 additional species of anthozoans and another species of *Costasiella* slug and synaptid sea cucumber (Suppl. material 5), bringing the final DNA barcode-based species count up to 395. Although barcodes for 305 of the 633 specimens were matched successfully (i.e. ≥ 97%) to global databases (Suppl. material 6), 49 of these matched to database sequences that were not identified to species (e.g. uncultured zooplankton and "sp."), while another 28 were not congruent with our morphological identification. Hence, only 229 specimens (36.2% of 633) were successfully identified based on COI barcode matching alone, which represented 116 MOTUs/species at the 3% threshold. Morphological identification for 90 specimens without COI-based identification (< 97% identity match, ≥ 97% match to unidentified database sequences or mismatch to morphological identification) allowed for 304 specimen barcodes to be tagged with species names (48.0% of 633), which represented 149 species (Suppl. material 6). With the addition of a further 14 species, based on morphology alone, a total of 163 known species were identified across the 931 specimens collected at the SIMP. Ten of these were new (zooplankton) species records for Singapore. Including MOTUs not identifiable to species, we hereby report a total of more than 365 animal species collected from SIMP, based on morphological and/or genetic identifications. All sequence data have been deposited in GenBank (Accession numbers MN689967–MN690599) and BOLD.

Discussion

In recent years, the process of species discovery has been enhanced with DNA barcoding approaches (Hebert et al. 2003). Relatedly, large-scale marine sampling programmes and expeditions such as the Moorea Biocode Project (Check 2006, Leray et al. 2011) and SANTO 2006 (Bouchet et al. 2009), focusing on South Pacific islands, have expedited species discovery and diversity estimation. Such biodiversity sampling expeditions are important, not just for documenting biodiversity; the massive collections also provide material for improving our understanding of the evolutionary histories of various groups (Holford et al. 2008,Redmond et al. 2013, da Silva Oliveira et al. 2017). In this study, nearly one thousand specimens belonging to more than 365 marine animal species were

collected and processed over 24 months across 13 sites at the SIMP. These species represent ca. 300 macrofauna and 70 zooplankton species, which include ten new (zooplankton) species records for Singapore and 58 species that are hitherto not included or misidentified in two of the largest global COI barcode databases (GenBank and BOLD). While morphological identification of most groups could be performed to the family level, species-level sorting and identification was a key challenge before the use of DNA barcoding. This was demonstrated in morphologically distinct groups that were easily recognised, where DNA barcoding resulted in approximately 36% increase in the number of specimens that could be delimited to at least species level. These advances underscore the importance of COI barcoding, especially for large-scale biodiversity surveys.

The morphological study of small animals and zooplankton is particularly time-consuming because large numbers of specimens are usually collected (e.g. Schmoker et al. 2014, Gan et al. 2019). This includes planktonic larvae, which are some of the most difficult developmental stages to identify and are traditionally reliant on laboratory-reared larvae to match juveniles with adults (Miller et al. 1989, Johnson and Allen 2005). Barnacles (Cirripedia) in the water column have never been identified to species before (Schmoker et al. (2014), likely because their planktonic nauplii tend to be morphologically similar between species (Gaonkar et al. 2014). In the present study, a combination of conventional sorting via microscopy and barcode matching to the GenBank database yielded new zooplankton species records for Singapore within just two plankton-tow samples. In particular, adult-larva matching was achieved for four barnacle species in the plankton samples, as all could be successfully identified to species by their DNA barcodes. Furthermore, planktonic molluscs have high potential for phenotypic plasticity and have been shown to exhibit cryptic speciation, displaying substantial morphological overlap between species and distinct morphotypes in different life stages (Kocot et al. 2016, Sun et al. 2016). Being able to quickly associate larvae with adults using DNA barcodes is thus important to improve the sensitivity of species detection. Larvae typically provide a much larger pool of species' signals that can increase the chances of species detection than with adults, which may often be rare or go undetected when adult females are collected. Expanding beyond the usual morphological work on adult male specimens through the use of DNA barcoding allows for a more comprehensive understanding of species' ecology, such as their life histories and phenologies, since the life history stages of dozens of species can be cost-effectively matched in a single study (Yeo et al. 2018).

Morphological identification can be challenging, even for charismatic animals due to the presence of cryptic species. Our analyses revealed at least two pairs of morphologically indistinguishable species with high COI sequence divergence. These possibly sympatric cryptic species groups include two *Ligia* isopods with a 22% pairwise distance, as well as two *Peronia* slugs (Mollusca: Gastropoda: Onchidiidae) with a 5.4% pairwise distance. In the latter case, Chang et al. (2018) have shown, using an integrative taxonomic approach, that *Peronia* onchidiids in Singapore form a cryptic species complex. Cryptic diversity in the SIMP onchidiids were initially confirmed via BLAST matches in which IP0287 and IP0136 matched 99.8% and 99.9% to *Peronia* sp. 2 and *Peronia* "Singapore Clade", respectively and corresponding morphological differences were found (Chang et al. 2018, Chan et al.

2019). More detailed morphological work is necessary to determine whether the *Ligia* species also belong to a species complex.

Indeed, DNA barcodes can help with species delimitation and cryptic species detection. DNA barcodes also allow for obtaining abundance and distribution information, but they tend to be of limited value for their original purpose, i.e. species identification, as only 36.2% of barcodes obtained here had species-level matches. A substantial number of our sequences that were matched to GenBank sequences at < 90% identity yielded only very tentative genus- or family-level identities. Even well-studied and common taxa such as molluscs, arthropods and fishes (e.g. *Hyselodoris*, *Dendrodoris*, *Ashtoret, Grapsus and Pomacentrus*) lacked barcodes in GenBank. Furthermore, in some taxa, the genus-level identities were of questionable accuracy. For example, amongst *Alpheus* shrimps, up to 20 specimens were recovered in the incorrect lineage with < 90% identity, with the closest match being a Caridea sp. at 82% to 88% sequence similarity (Suppl. material 3). The same situation was observed with synaptid sea cucumbers (Synaptidae) which lacked close matches to GenBank sequences. Some of these problems can be overcome by building a local database for Singapore that is supported by reliable morphological identifications.

The inadequacy of the barcode databases was particularly problematic for understudied groups such as annelids, platyhelminths, poriferans and zooplankton, such as chaetognaths. Amongst the 17 barcoded platyhelminth flatworm samples, for instance, all GenBank matches were < 88% in sequence identity and accurate only to the phylum level for 14 samples, while seven samples were assigned to the incorrect genus (see also Vanhove et al. 2013). Most platyhelminth flatworm studies use the 28S rDNA gene for phylogenetic analysis (Litvaitis and Rohde 1999, Litvaitis and Newman 2001, Bolaños et al. 2007, Bahia et al. 2017) and thus few COI sequences are represented in the global databases. Furthermore, the commonly used barcoding region *sensu* Folmer et al. (1994) does not overlap with the COI region that is informative for flatworms and thus lacks resolving power to delimit flatworm species (Vanhove et al. 2013) with the usual 3% barcoding threshold used across metazoans (Hebert et al. 2003). Amongst the zooplankton samples, we found up to 15 chaetognath MOTUs that could not be identified to species, either based on morphology or barcodes. Chaetognath species, such as *Krohnitta pacifica* and *Aidanosagitta crassa*, have been previously recorded in Singapore (Schmoker et al. 2014) but both species still lack COI barcodes in GenBank.

Overall, our study confirms that a substantial number of the sequences in the global databases are misidentified and that one should carefully distinguish between the use of the barcode sequences for, for example, obtaining distributional data and the use of barcode identification in the database. This is particularly important for understudied taxa (Zhang and Zhang 2014). For example, GenBank sequences of the copepod *Paracalanus aculeatus* had 100% matches to two different MOTUs which differed by an uncorrected p-distance of 9.3% (ZP011/ZP344 and ZP232 clusters). Conversely, a single cluster/ sequence (e.g. ZP344) in our database matched at 100% to different *Paracalanus* species on GenBank, when compared to the BOLD database. Even well-studied taxa including fish, arthropods and molluscs, were not free from misidentification (e.g. Buhay 2009)—poor

identity matches were obtained for our specimens of *Centrogenys vaigiensis*, *Ashtoret lunaris*, *Jorunna funebris*, for example, despite the presence of barcodes in GenBank that were filed under these names. We recommend that this gap be bridged by working with taxonomic experts in each pre-sorted group and, subsequently, supplementing local or global (i.e. GenBank, BOLD) databases with COI barcode sequences that are tagged with accurate species identities. This will facilitate future faunal identification studies.

For more than a decade, the COI locus has been popularised for barcoding a wide range of metazoan species (Hebert et al. 2003). Small intraspecific variation coupled with correspondingly large interspecific variation in the COI locus amongst most metazoan species sometimes yield a 'barcoding gap', which allows for accurate species identification using a generalised threshold of 3% between intra- and interspecific variabilities (Meyer and Paulay 2005, Meier et al. 2006). In reality, the distinctions between intra- and interspecific distances vary amongst taxa and the barcoding gap may not be present (Virgilio et al. 2010), especially in recently-diverged species (Meier et al. 2008, van Velzen et al. 2012). Such overlap can be caused by slow evolution of COI in some taxa (Hellberg 2006, Huang et al. 2008, Shearer et al. 2008), hybridisation events between sympatric species (Steinke et al. 2009, Ward et al. 2009) or by high sequence divergence, coupled with morphological stasis (Gómez et al. 2002). In particular, delimitation at the species and even genus level is difficult for many anthozoans, platyhelminth flatworms and some fish because of the lack of COI divergence (Steinke et al. 2009, Vanhove et al. 2013). For these groups, other markers need to be explored for DNA barcoding. To this end, combinations of different genes, including mitochondrial and nuclear loci, have been proposed and used for different groups of anthozoans (Huang et al. 2011) and the 28S rDNA for flatworms (Litvaitis and Rohde 1999, Litvaitis and Newman 2001). In anthozoans, species-level resolution is still limited due to frequent hybridisation (Quattrini et al. 2019), resulting in low interspecific divergence (Shearer and Coffroth 2008, Dohna and Kochzius 2015), though these markers are possibly useful at the genus level (Hsu et al. 2014). Other mitochondrial markers such as cytochrome *b* (Ward et al. 2005, Sevilla et al. 2007), 12S rRNA (Miya et al. 2015) and D-loop control region (Lee et al. 1995) have been suggested for use on barcoding fish species. Phylogenomic analyses in these groups are emerging (Egger et al. 2015, Lin et al. 2016, Marcionetti et al. 2018) and the data will help in the design of taxon-specific nuclear markers for future DNA barcoding work. With reducing costs of high-throughput DNA sequencing, multiple-gene DNA barcoding should become viable in the near future, which can help improve the accuracy of species identification.

The large number of species from many divergent lineages, examined here, would typically require a wide range of taxonomic expertise to sort the specimens into putative species, based on morphological data. This expertise was not readily available, so we use molecular tools for rapid and cost-effective species delimitation (Baloğlu et al. 2018, Wang et al. 2018) and occasionally for species identification if sequences can be matched to accurately-identified databases. Eventually, this barcoding exercise would follow the reverse workflow described in Wang et al. (2018), in which DNA barcodes act as precursory guides to direct the verification and evaluation by skilled taxonomists. While we have only managed to put names on 163 species, less than half of the > 365 species

delimited here (Fig. 5), the data have already spawned collaborations with various specialists to focus on the more understudied fauna. In particular, taxonomic work elicited by our results is ongoing or recently accomplished for taxa such as corallimorpharians (Oh et al. 2019), corals (Poquita-Du et al. 2017), anemones (Yap et al. 2019) and onchidiid slugs (Chang et al. 2018, Chan et al. 2019), including potentially new species initially spotlighted by DNA barcodes generated here (Chang et al. 2018, Chan et al. 2019, Oh et al. 2019).

Our work here is only the beginning of further molecular ecological work in this biodiverse region. It follows recent, successful, large-scale biodiversity sampling exercises, such as the Moorea Biocode Project (Check 2006), which has helped pave the way for numerous other biodiversity studies and related applications, such as environmental DNA metabarcoding (Leray et al. 2011, Leray et al. 2013, Leray and Knowlton 2015), uncovering microbial diversity (McCliment et al. 2011), estimating biodiversity (Plaisance et al. 2009, Hubert et al. 2012), standardised reef biodiversity sampling (Leray and Knowlton 2015, Ransome et al. 2017), larval species identification (Hubert et al. 2010), designing metazoan-specific primers (Geller et al. 2013, Leray et al. 2013) and mapping entire island ecosystems (Achterberg et al. 2018) to designate genomic observatories (Field and Davies 2015). We can leverage on the DNA barcode database built for the SIMP to motivate the development of applications for better documenting species diversity in the region, thus strengthening the case for developing the marine park into a genomic observatory.

To understand why this is advantageous, we note that survey windows at the SIMP are limited in the intertidal areas by the tidal regime and in the subtidal by strong currents, so rapid and non-intrusive sampling methods such as environmental DNA (eDNA) would enable more regular surveys (Rees et al. 2014, Comtet et al. 2015). To this end, our DNA barcode database could serve as *the* reference library for matching and discovering species found in eDNA samples. Exploratory eDNA experiments based on 26 two-litre water samples collected from eight localities have led to the detection of > 500 metazoan MOTUs (Y.C.A. Ip, Y.C. Tay & J.J.M. Chang, unpublished data). Notably, 20 of these MOTUs were matched only to our local database and not the global databases GenBank and BOLD. This is remarkable, given the large repository of > 2.68 million COI sequences on GenBank. Indeed, the enhanced database resolution, resulting from thorough sampling at the SIMP, is crucial as eDNA has emerged as one of the main technology-driven tools for environmental monitoring and management.

Conclusion

The collection and barcoding of marine animals at Sisters' Islands and the surrounding islands began more than two years ago, at a time when these locations were designated Singapore's first marine park. This is part of a larger initiative to make Singapore's biodiversity identifiable, as well as to provide molecular identification tools for future work. Despite only a collection frequency of 34 times over a span of two years on foot and via SCUBA across a large 40-ha area and using only simple hand tools, nets and traps, our study managed to sample more than 365 species across a wide range of marine animals.

A more systematic sampling approach, covering a larger area and using grabs, trawls, dredges and various nets will uncover greater diversity and more taxa, including infaunal and meiofaunal groups.

Being able to quantify and identify species diversity is important for many reasons, including the provision of a community baseline against which future surveys can be compared (Resh and Unzicker 1975). It is particularly critical for making better-informed decisions with respect to coastal reclamation and urban redevelopment (Chee et al. 2017), as well as monitoring the influx of introduced species. Records with geographic data are important for conservation when prioritising sites for protection (Anderson and Martínez-Meyer 2004) and for allowing biogeographic patterns at the regional and even global scales to be uncovered more precisely (Abell et al. 2008, Huang et al. 2018, Yip et al. 2019). Furthermore, our data comprise species records supported by linked photographic images and COI barcode sequences, potentially paving the way for more efficient biomonitoring applications such as eDNA testing. Fundamentally, these methods are revolutionising biodiversity studies, thus not only allowing scientists to discover species on Earth, but also allowing for more ready access to DNA-based identification via new small-sized sequencers whose use requires minimal amounts of laboratory equipment (Srivathsan et al. 2018).

Acknowledgements

Funding program

Hosting institution

National University of Singapore

Ethics and security

No ethical violations or security breaches were made during the course of this study.

Author contributions

R.M., Y.C.T., K.T., D.H. and L.M.C. conceived the idea. R.M., Y.C.T. and D.H. designed the experiments. Y.C.A.I., Y.C.T., S.X.G., D.H. and H.P.A. conducted the fieldwork, while Y.C.A.I., Y.C.T., S.X.G. and D.H. conducted the experiments and analysed the data. Y.C.A.I., D.H. and Y.C.T. wrote the main manuscript and all authors actively revised and approved the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

- Abell R, Thieme M, Revenga C, Bryer M, Kottelat M, Bogutskaya N, Coad B, Mandrak N, Balderas SC, Bussing W, Stiassny MJ, Skelton P, Allen G, Unmack P, Naseka A, Ng R, Sindorf N, Robertson J, Armijo E, Higgins J, Heibel T, Wikramanayake E, Olson D, López H, Reis R, Lundberg J, Sabaj Pérez M, Petry P (2008) Freshwater ecoregions of the world: a new map of biogeographic units for freshwater biodiversity conservation. BioScience 58 (5): 403-414. https://doi.org/10.1641/b580507
- Achterberg E, Buttigieg P, Janssen F, Peplies J, Compere C (2018) Metrology best practice manuals. AtlantOS D6.4: 1-62. https://doi.org/10.3289/ATLANTOS_D6.4
- Altschul S (1990) Basic local alignment search tool. Journal of Molecular Biology 215 (3): 403-410. https://doi.org/10.1006/jmbi.1990.9999
- Anderson R, Martínez-Meyer E (2004) Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. Biological Conservation 116 (2): 167-179. https://doi.org/10.1016/s0006-3207(03)00187-3
- Ang Y, Puniamoorthy J, Pont A, Bartak M, Blanckenhorn W, Eberhard W, Puniamoorthy N, Silva V, Munari L, Meier R (2013) A plea for digital reference collections and other science-based digitization initiatives in taxonomy: Sepsidnet as exemplar. Systematic Entomology 38 (3): 637-644. https://doi.org/10.1111/syen.12015
- Appeltans W, Ahyong ST, Anderson G, Angel MV, Artois T, Bailly N, Bamber R, Barber A, Bartsch I, Berta A, Błażewicz-Paszkowycz M, Bock P, Boxshall G, Boyko CB, Brandão SN, Bray RA, Bruce NL, Cairns SD, Chan T, Cheng L, Collins AG, Cribb T, Curini-Galletti M, Dahdouh-Guebas F, Davie PJF, Dawson MN, De Clerck O, Decock W, De Grave S, de Voogd NJ, Domning DP, Emig CC, Erséus C, Eschmeyer W, Fauchald K, Fautin DG, Feist SW, Fransen CHJM, Furuya H, Garcia-Alvarez O, Gerken S, Gibson D, Gittenberger A, Gofas S, Gómez-Daglio L, Gordon DP, Guiry MD, Hernandez F, Hoeksema BW, Hopcroft RR, Jaume D, Kirk P, Koedam N, Koenemann S, Kolb JB, Kristensen RM, Kroh A, Lambert

G, Lazarus DB, Lemaitre R, Longshaw M, Lowry J, Macpherson E, Madin LP, Mah C, Mapstone G, McLaughlin PA, Mees J, Meland K, Messing CG, Mills CE, Molodtsova TN, Mooi R, Neuhaus B, Ng PKL, Nielsen C, Norenburg J, Opresko DM, Osawa M, Paulay G, Perrin W, Pilger JF, Poore GCB, Pugh P, Read GB, Reimer JD, Rius M, Rocha RM, Saiz-Salinas JI, Scarabino V, Schierwater B, Schmidt-Rhaesa A, Schnabel KE, Schotte M, Schuchert P, Schwabe E, Segers H, Self-Sullivan C, Shenkar N, Siegel V, Sterrer W, Stöhr S, Swalla B, Tasker ML, Thuesen EV, Timm T, Todaro MA, Turon X, Tyler S, Uetz P, van der Land J, Vanhoorne B, van Ofwegen LP, van Soest RWM, Vanaverbeke J, Walker-Smith G, Walter TC, Warren A, Williams GC, Wilson SP, Costello MJ (2012) The magnitude of global marine species diversity. Current Biology 22 (23): 2189-202. https://doi.org/10.1016/j.cub.2012.09.036

- Bahia J, Padula V, Schrödl M (2017) Polycladida phylogeny and evolution: integrating evidence from 28S rDNA and morphology. Organisms Diversity & Evolution 17 (3): 653-678. https://doi.org/10.1007/s13127-017-0327-5
- Baloğlu B, Clews E, Meier R (2018) NGS barcoding reveals high resistance of a hyperdiverse chironomid (Diptera) swamp fauna against invasion from adjacent freshwater reservoirs. Frontiers in Zoology 15 (1). https://doi.org/10.1186/s12983-018-0276-7
- Barrett RH, Hebert PN (2005) Identifying spiders through DNA barcodes. Canadian Journal of Zoology 83 (3): 481-491. https://doi.org/10.1139/z05-024
- Bensasson D (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology & Evolution 16 (6): 314-321. https://doi.org/10.1016/s0169-5347(01)02151-6
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW (2018) GenBank. Nucleic Acids Research 46 (D1): D41-D47. https://doi.org/10.1093/nar/gkx1094
- Bertrand Y, Pleijel F, Rouse GW (2006) Taxonomic surrogacy in biodiversity assessments, and the meaning of Linnaean ranks. Systematics and Biodiversity 4 (2): 149-159. https://doi.org/10.1017/s1477200005001908
- Bolaños DM, Quiroga S, Litvaitis M (2007) Five new species of cotylean flatworms (Platyhelminthes: Polycladida) from the wider Caribbean. Zootaxa 1650: 1-23. https://doi.org/10.11646/zootaxa.1650.1.1
- Bouchet P, Le Guyader H, Pascal O (2009) The SANTO 2006 Global Biodiversity Survey: an attempt to reconcile the pace of taxonomy and conservation. Zoosystema 31 (3): 401-406. https://doi.org/10.5252/z2009n3a0
- Brower AZ (2006) Problems with DNA barcodes for species delimitation: 'Ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). Systematics and Biodiversity 4 (2): 127-132. https://doi.org/10.1017/s147720000500191x
- Buhay J (2009) "COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. Journal of Crustacean Biology 29 (1): 96-110. https://doi.org/10.1651/08-3020.1
- Carbayo F, Marques A (2011) The costs of describing the entire animal kingdom. Trends in Ecology & Evolution 26 (4): 154-155. https://doi.org/10.1016/j.tree.2011.01.004
- Chan A, Chiang L, Hapuarachchi HC, Tan C, Pang S, Lee R, Lee K, Ng L, Lam-Phua S (2014) DNA barcoding: complementing morphological identification of mosquito species in Singapore. Parasites & Vectors 7 (1). https://doi.org/10.1186/preaccept-1085463044116874
- Chang A (2015) Singapore's mother of all coral reefs. https://www.straitstimes.com/singapore/environment/singapores-mother-of-all-coral-reefs. Accessed on: 2019-11-06.

- Chang JJM, Tay YC, Ang HP, Tun KPP, Chou LM, Meier R, Huang D (2018) Molecular and anatomical analyses reveal that *Peronia verruculata* (Gastropoda: Onchidiidae) is a cryptic species complex. Contributions to Zoology 87 (3): 149-165. https://doi.org/10.1163/18759866-08703002
- Chan IW, Chang JJM, Huang D, Todd P (2019) Colour pattern measurements successfully differentiate two cryptic Onchidiidae Rafinesque, 1815 species. Marine Biodiversity 49 (4): 1743-1750. https://doi.org/10.1007/s12526-019-00940-4
- Check E (2006) Treasure island: pinning down a model ecosystem. Nature 439 (7075): 378-379. https://doi.org/10.1038/439378a
- Chee SY, Othman AG, Sim YK, Mat Adam AN, Firth L (2017) Land reclamation and artificial islands: Walking the tightrope between development and conservation. Global Ecology and Conservation 12: 80-95. https://doi.org/10.1016/j.gecco.2017.08.005
- Comtet T, Sandionigi A, Viard F, Casiraghi M (2015) DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. Biological Invasions 17 (3): 905-922. https://doi.org/10.1007/s10530-015-0854-y
- Costello MJ, May RM, Stork NE (2013) Can we name Earth's species before they go extinct? Science 339 (6118): 413-416. https://doi.org/10.1126/science.1230318
- Darling J (2019) How to learn to stop worrying and love environmental DNA monitoring. Aquatic Ecosystem Health & Management 1-13. https://doi.org/10.1080/14634988.2019.1682912
- da Silva Oliveira FA, Michonneau F, da Cruz Lotufo TM (2017) Molecular phylogeny of Didemnidae (Ascidiacea: Tunicata). Zoological Journal of the Linnean Society 180 (3): 603-612. https://doi.org/10.1093/zoolinnean/zlw002
- Davies N, Field D, Amaral-Zettler L, Clark MS, Deck J, Drummond A, Faith DP, Geller J, Gilbert J, Glöckner FO, Hirsch PR, Leong J, Meyer C, Obst M, Planes S, Scholin C, Vogler AP, Gates RD, Toonen R, Berteaux-Lecellier V, Barbier M, Barker K, Bertilsson S, Bicak M, Bietz MJ, Bobe J, Bodrossy L, Borja A, Coddington J, Fuhrman J, Gerdts G, Gillespie R, Goodwin K, Hanson PC, Hero J, Hoekman D, Jansson J, Jeanthon C, Kao R, Klindworth A, Knight R, Kottmann R, Koo MS, Kotoulas G, Lowe AJ, Marteinsson VT, Meyer F, Morrison N, Myrold DD, Pafilis E, Parker S, Parnell JJ, Polymenakou PN, Ratnasingham S, Roderick GK, Rodriguez-Ezpeleta N, Schonrogge K, Simon N, Valette-Silver NJ, Springer YP, Stone GN, Stones-Havas S, Sansone S, Thibault KM, Wecker P, Wichels A, Wooley JC, Yahara T, Zingone A (2014) The founding charter of the Genomic Observatories Network. GigaScience 3 (1). https://doi.org/10.1186/2047-217x-3-2
- Djurhuus A, Pitz K, Sawaya NA, Rojas-Márquez J, Michaud B, Montes E, Muller-Karger F, Breitbart M (2018) Evaluation of marine zooplankton community structure through environmental DNA metabarcoding. Limnology and Oceanography Methods 16 (4): 209-221. https://doi.org/10.1002/lom3.10237
- Dohna T, Kochzius M (2015) Obstacles to molecular species identification in sea anemones (Hexacorallia: Actiniaria) with COI, a COI intron, and ITS II. Marine Biodiversity 46 (1): 291-297. https://doi.org/10.1007/s12526-015-0329-5
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32 (5): 1792-1797. https://doi.org/10.1093/nar/gkh340
- Egger B, Lapraz F, Tomiczek B, Müller S, Dessimoz C, Girstmair J, Škunca N, Rawlinson KA, Cameron CB, Beli E, Todaro MA, Gammoudi M, Noreña C, Telford MJ (2015) A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms. Current Biology 25 (10): 1347-1353. https://doi.org/10.1016/j.cub.2015.03.034

- Ekrem T, Willassen E, Stur E (2007) A comprehensive DNA sequence library is essential for identification with DNA barcodes. Molecular Phylogenetics and Evolution 43 (2): 530-542. https://doi.org/10.1016/j.ympev.2006.11.021
- Elias M, Hill RI, Willmott KR, Dasmahapatra KK, Brower AZ, Mallet J, Jiggins CD (2007) Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proceedings of the Royal Society B: Biological Sciences 274 (1627): 2881-2889. https://doi.org/10.1098/rspb.2007.1035
- Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. Biology Letters 4 (4): 423-425. https://doi.org/10.1098/rsbl.2008.0118
- Field D, Davies N (2015) Biocode: The New Age of Genomics. Oxford University Press, Oxford, 288 pp.
- 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. Molecular Marine Biology and Biotechnology 3 (5): 294-299.
- Gan SX, Tay YC, Huang D (2019) Effects of macroalgal morphology on marine epifaunal diversity. Journal of the Marine Biological Association of the United Kingdom https://doi.org/10.1017/s0025315419000900
- Gaonkar C, Khandeparker L, Desai D, Anil AC (2014) Identification of *Balanus amphitrite* larvae from field zooplankton using species-specific primers. Journal of the Marine Biological Association of the United Kingdom 95 (3): 497-502. https://doi.org/10.1017/s0025315414001581
- 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. https://doi.org/10.1111/1755-0998.12138
- Giam X, Ng TH, Yap VB, Tan HW (2010) The extent of undiscovered species in Southeast Asia. Biodiversity and Conservation 19 (4): 943-954. https://doi.org/10.1007/s10531-010-9792-2
- Goldstein PZ, DeSalle R (2011) Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. BioEssays 33 (2): 135-47. https://doi.org/10.1002/bies.201000036
- Gómez A, Serra M, Carvalho GR, Lunt DH (2002) Speciation in ancient cryptic species complexes: evidence from the molecular phylogeny of *Brachionus plicatilis* (Rotifera). Evolution 56 (7): 1431-44. https://doi.org/10.1554/0014-3820(2002)056[1431:SIACSC]2.0.CO;2
- Grant RA, Linse K (2009) Barcoding Antarctic Biodiversity: current status and the CAML initiative, a case study of marine invertebrates. Polar Biology 32 (11): 1629-1637. https://doi.org/10.1007/s00300-009-0662-x
- Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. Proceedings of the National Academy of Sciences of the United States of America 103 (4): 968-971. https://doi.org/10.1073/pnas.0510466103
- Hajibabaei M, Singer GC, Hebert PN, Hickey D (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends in Genetics 23 (4): 167-172. https://doi.org/10.1016/j.tig.2007.02.001

- Harper L, Buxton A, Rees H, Bruce K, Brys R, Halfmaerten D, Read D, Watson H, Sayer C, Jones E, Priestley V, Mächler E, Múrria C, Garcés-Pastor S, Medupin C, Burgess K, Benson G, Boonham N, Griffiths R, Lawson Handley L, Hänfling B (2018) Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. Hydrobiologia 826 (1): 25-41. https://doi.org/10.1007/s10750-018-3750-5
- Hausmann A, Haszprunar G, Hebert PDN (2011) DNA barcoding the geometrid fauna of Bavaria (Lepidoptera): successes, surprises, and questions. PLOS ONE 6 (2): e17134. https://doi.org/10.1371/journal.pone.0017134
- Havermans C, Nagy ZT, Sonet G, De Broyer C, Martin P (2011) DNA barcoding reveals new insights into the diversity of Antarctic species of *Orchomene sensu lato* (Crustacea: Amphipoda: Lysianassoidea). Deep Sea Research Part II: Topical Studies in Oceanography 58: 230-241. https://doi.org/10.1016/j.dsr2.2010.09.028
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences of the United States of America 101 (41): 14812-14817. https://doi.org/10.1073/pnas.0406166101
- Hebert PN, Cywinska A, Ball S, deWaard J (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences 270 (1512): 313-321. https://doi.org/10.1098/rspb.2002.2218
- Hebert PN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. PLoS Biology 2 (10): e312. https://doi.org/10.1371/journal.pbio.0020312
- Hellberg ME (2006) No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. BMC Evolutionary Biology 6: 24. https://doi.org/10.1186/1471-2148-6-24
- Holford M, Puillandre N, Terryn Y, Cruaud C, Olivera B, Bouchet P (2008) Evolution of the Toxoglossa venom apparatus as inferred by molecular phylogeny of the Terebridae. Molecular Biology and Evolution 26 (1): 15-25. https://doi.org/10.1093/molbev/msn211
- Hsu C, de Palmas S, Kuo C, Denis V, Chen CA (2014) Identification of scleractinian coral recruits using fluorescent censusing and DNA barcoding techniques. PLoS ONE 9 (9). https://doi.org/10.1371/journal.pone.0107366
- Huang D, Meier R, Todd P, Chou LM (2008) Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. Journal of Molecular Evolution 66 (2): 167-174. https://doi.org/10.1007/s00239-008-9069-5
- Huang D, Licuanan WY, Baird AH, Fukami H (2011) Cleaning up the 'Bigmessidae': Molecular phylogeny of scleractinian corals from Faviidae, Merulinidae, Pectiniidae and Trachyphylliidae . BMC Evolutionary Biology 11: 37. https://doi.org/10.1186/1471-2148-11-37
- Huang D, Goldberg E, Chou LM, Roy K (2018) The origin and evolution of coral species richness in a marine biodiversity hotspot. Evolution 72 (2): 288-302. https://doi.org/10.1111/evo.13402
- Hubert N, Delrieu-Trottin E, Irisson J, Meyer C, Planes S (2010) Identifying coral reef fish larvae through DNA barcoding: A test case with the families Acanthuridae and Holocentridae . Molecular Phylogenetics and Evolution 55 (3): 1195-1203. https://doi.org/10.1016/j.ympev.2010.02.023
- Hubert N, Meyer CP, Bruggemann HJ, Guérin F, Komeno RJL, Espiau B, Causse R, Williams JT, Planes S (2012) Cryptic diversity in Indo-Pacific coral-reef fishes revealed by

DNA-barcoding provides new support to the centre-of-overlap hypothesis. PLOS ONE 7 (3): e28987. https://doi.org/10.1371/journal.pone.0028987

- Isaac NJB, Mallet J, Mace GM (2004) Taxonomic inflation: its influence on macroecology and conservation. Trends in Ecology & Evolution 19 (9): 464-469. https://doi.org/10.1016/j.tree.2004.06.004
- Jerde C, Mahon A, Chadderton WL, Lodge D (2011) "Sight-unseen" detection of rare aquatic species using environmental DNA. Conservation Letters 4 (2): 150-157. https://doi.org/10.1111/j.1755-263x.2010.00158.x
- Jinbo U, Kato T, Ito M (2011) Current progress in DNA barcoding and future implications for entomology. Entomological Science 14 (2): 107-124. https://doi.org/10.1111/j.1479-8298.2011.00449.x
- Johnson WS, Allen DM (2005) Zooplankton of the Atlantic and Gulf Coasts: A Guide to Their Identification and Ecology. JHU Press
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28 (12): 1647-1649. https://doi.org/10.1093/bioinformatics/bts199
- Kerr KCR, Stoeckle MY, Dove CJ, Weigt LA, Francis CM, Hebert PDN (2007) Comprehensive DNA barcode coverage of North American birds. Molecular Ecology Notes 7 (4): 535-543. https://doi.org/10.1111/j.1471-8286.2007.01670.x
- Kocot KM, Aguilera F, McDougall C, Jackson DJ, Degnan BM (2016) Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. Frontiers in Zoology 13: 23. https://doi.org/10.1186/s12983-016-0155-z
- Koh LP, Sodhi N (2010) Conserving Southeast Asia's imperiled biodiversity: scientific, management, and policy challenges. Biodiversity and Conservation 19 (4): 913-917. https://doi.org/10.1007/s10531-010-9818-9
- Kuntner M, Agnarsson I (2006) Are the linnean and phylogenetic nomenclatural systems combinable? Recommendations for biological nomenclature. Systematic Biology 55 (5): 774-84. https://doi.org/10.1080/10635150600981596
- Kwong S, Srivathsan A, Meier R (2012) An update on DNA barcoding: low species coverage and numerous unidentified sequences. Cladistics 28 (6): 639-644. https://doi.org/10.1111/j.1096-0031.2012.00408.x
- Laurance WF (2013) The race to name Earth's species. Science 339 (6125): 1275. https://doi.org/10.1126/science.339.6125.1275-a
- Lee W, Conroy J, Howell WH, Kocher T (1995) Structure and evolution of teleost mitochondrial control regions. Journal of Molecular Evolution 41 (1): 54-66. https://doi.org/10.1007/bf00174041
- Leray M, Boehm JT, Mills SC, Meyer CP (2011) Moorea BIOCODE barcode library as a tool for understanding predator–prey interactions: insights into the diet of common predatory coral reef fishes. Coral Reefs 31 (2): 383-388. https://doi.org/10.1007/s00338-011-0845-0
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (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: 34. https://doi.org/10.1186/1742-9994-10-34

- Leray M, Knowlton N (2015) DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. Proceedings of the National Academy of Sciences 112: 2076-2081. https://doi.org/10.1073/pnas.1424997112
- Lin MF, Chou WH, Kitahara MV, Chen CLA, Miller DJ, Foret S (2016) Corallimorpharians are not "naked corals": insights into relationships between Scleractinia and Corallimorpharia from phylogenomic analyses. PeerJ 4: e2463. https://doi.org/10.7717/peerj.2463
- Litvaitis M, Rohde K (1999) A molecular test of platyhelminth phylogeny: inferences from partial 28S rDNA sequences. Invertebrate Biology 118 (1): 42-56. https://doi.org/10.2307/3226911
- Litvaitis MK, Newman LJ (2001) A molecular framework for the phylogeny of the Pseudocerotidae (Platyhelminthes, Polycladida). Hydrobiologia 444: 177-182. https://doi.org/10.1023/a:1017503124908
- Lobo J, Costa PM, Teixeira MA, Ferreira MS, Costa MH, Costa FO (2013) Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. BMC Ecology 13: 34. https://doi.org/10.1186/1472-6785-13-34
- Marcionetti A, Rossier V, Bertrand JAM, Litsios G, Salamin N (2018) First draft genome of an iconic clownfish species (*Amphiprion frenatus*). Molecular Ecology Resources 18 (5): 1092-1101. https://doi.org/10.1111/1755-0998.12772
- McCliment EA, Nelson CE, Carlson CA, Alldredge AL, Witting J, Amaral-Zettler LA (2011) An all-taxon microbial inventory of the Moorea coral reef ecosystem. ISME Journal 6 (2): 309-319. https://doi.org/10.1038/ismej.2011.108
- Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Systematic Biology 55 (5): 715-728. https://doi.org/10.1080/10635150600969864
- Meier R (2008) DNA sequences in taxonomy. The New Taxonomy 7: 95-127. https://doi.org/10.1201/9781420008562
- Meier R, Zhang G, Ali F (2008) The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. Systematic Biology 57 (5): 809-813. https://doi.org/10.1080/10635150802406343
- Meier R, Wong W, Srivathsan A, Foo M (2015) $1 DNA barcodes for reconstructing complex phenomes and finding rare species in specimen-rich samples. Cladistics 32 (1): 100-110. https://doi.org/10.1111/cla.12115
- Messenger JB, Nixon M, Ryan KP (1985) Magnesium chloride as an anaesthetic for cephalopods. Comparative Biochemistry and Physiology Part C: Comparative Pharmacology 82 (1): 203-205. https://doi.org/10.1016/0742-8413(85)90230-0
- Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biology 3 (12): e422. https://doi.org/10.1371/journal.pbio.0030422
- Miller K, Blower S, Hedgecock D, Roughgarden J (1989) Comparison of larval and adult stages of *Chthamalus dalli* and *Chthamalus fissus* (Cirripedia: Thoracica). Journal of Crustacean Biology 9 (2): 242-256. https://doi.org/10.1163/193724089X00061
- Miller SE (2007) DNA barcoding and the renaissance of taxonomy. Proceedings of the National Academy of Sciences of the United States of America 104 (12): 4775-4776. https://doi.org/10.1073/pnas.0700466104
- Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T, Yamamoto S, Yamanaka H, Araki H, Kondoh M, Iwasaki W (2015) MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230

subtropical marine species. Royal Society Open Science 2: 150088. https://doi.org/10.1098/rsos.150088

- Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on Earth and in the ocean? PLoS Biology 9 (8): e1001127. https://doi.org/10.1371/journal.pbio.1001127
- Mora C, Wei C, Rollo A, Amaro T, Baco A, Billett D, Bopp L, Chen Q, Collier M, Danovaro R, Gooday A, Grupe B, Halloran P, Ingels J, Jones DB, Levin L, Nakano H, Norling K, Ramirez-Llodra E, Rex M, Ruhl H, Smith C, Sweetman A, Thurber A, Tjiputra J, Usseglio P, Watling L, Wu T, Yasuhara M (2013) Biotic and human vulnerability to projected changes in ocean biogeochemistry over the 21st century. PLoS Biology 11 (10): e1001682. https://doi.org/10.1371/journal.pbio.1001682
- Naomi TS, Antony G, George RM, Jasmine S (2006) Monograph on the Planktonic Shrimps of the Genus *Lucifer* (Family Luciferidae) from the Indian EEZ. Central Marine Fisheries Research Institute, Kochi.
- Oh RM, Neo ML, Yap NWL, Jain SS, Tan R, Chen CA, Huang D (2019) Citizen science meets integrated taxonomy to uncover the diversity and distribution of Corallimorpharia in Singapore. Raffles Bulletin of Zoology 67: 306-321. https://doi.org/10.26107/RBZ-2019-0022
- Park D, Foottit R, Maw E, Hebert PDN (2011) Barcoding bugs: DNA-based identification of the true bugs (Insecta: Hemiptera: Heteroptera). PLOS ONE 6 (4): e18749. https://doi.org/10.1371/journal.pone.0018749
- Plaisance L, Knowlton N, Paulay G, Meyer C (2009) Reef-associated crustacean fauna: biodiversity estimates using semi-quantitative sampling and DNA barcoding. Coral Reefs 28 (4): 977-986. https://doi.org/10.1007/s00338-009-0543-3
- Poquita-Du R, Ng CSL, Loo JB, Afiq-Rosli L, Tay YC, Todd P, Chou LM, Huang D (2017) New evidence shows that *Pocillopora* '*damicornis*-like' corals in Singapore are actually *Pocillopora acuta* (Scleractinia: Pocilloporidae). Biodiversity Data Journal 5: e11407. https://doi.org/10.3897/bdj.5.e11407
- Quattrini A, Wu T, Soong K, Jeng M, Benayahu Y, McFadden C (2019) A next generation approach to species delimitation reveals the role of hybridization in a cryptic species complex of corals. BMC Evolutionary Biology 19: 116. https://doi.org/10.1186/s12862-019-1427-y
- Ransome E, Geller JB, Timmers M, Leray M, Mahardini A, Sembiring A, Collins AG, Meyer CP (2017) The importance of standardization for biodiversity comparisons: A case study using autonomous reef monitoring structures (ARMS) and metabarcoding to measure cryptic diversity on Mo'orea coral reefs, French Polynesia. PLoS ONE 12 (4): e0175066. https://doi.org/10.1371/journal.pone.0175066
- Ratnasingham S, Hebert PN (2007) BARCODING: bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7 (3): 355-364. https://doi.org/10.1111/j.1471-8286.2007.01678.x
- Redmond NE, Morrow CC, Thacker RW, Diaz MC, Boury-Esnault N, Cárdenas P, Hajdu E, Lôbo-Hajdu G, Picton BE, Pomponi SA, Kayal E, Collins AG (2013) Phylogeny and systematics of demospongiae in light of new small-subunit ribosomal DNA (18S) sequences. Integrative and Comparative Biology 53 (3): 388-415. https://doi.org/10.1093/icb/ict078
- Rees H, Maddison B, Middleditch D, Patmore JM, Gough K (2014) REVIEW: The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in

ecology. Journal of Applied Ecology 51 (5): 1450-1459. https://doi.org/10.1111/1365-2664.12306

- Resh VH, Unzicker JD (1975) Water quality monitoring and aquatic organisms: the importance of species identification. Journal - Water Pollution Control Federation 47 (1): 9-19.
- Rey A, Carney K, Quinones L, Pagenkopp Lohan K, Ruiz G, Basurko O, Rodríguez-Ezpeleta N (2019) Environmental DNA metabarcoding: a promising tool for ballast water monitoring. Environmental Science & Technology 53 (20): 11849-11859. https://doi.org/10.1021/acs.est.9b01855
- Riedel A, Sagata K, Surbakti S, Tänzler R, Balke M (2013) One hundred and one new species of *Trigonopterus* weevils from New Guinea. ZooKeys 280: 1-150. https://doi.org/10.3897/zookeys.280.3906
- Ríos-Saldaña CA, Delibes-Mateos M, Ferreira C (2018) Are fieldwork studies being relegated to second place in conservation science? Global Ecology and Conservation 14: e00389. https://doi.org/10.1016/j.gecco.2018.e00389
- Schmoker C, Mahjoub M, Calbet A, Hsiao S, Russo F, Larsen O, Trottet A, Drillet G (2014) A review of the zooplankton in Singapore waters. Raffles Bulletin of Zoology 62: 726-749.
- Sevilla R, Diez A, Norén M, Mouchel O, Jérôme M, Verrez-Bagnis V, Van Pelt H, Favre-Krey L, Krey G, Consortium TF, Bautista J (2007) Primers and polymerase chain reaction conditions for DNA barcoding teleost fish based on the mitochondrial cytochrome b and nuclear rhodopsin genes. Molecular Ecology Notes 7 (5): 730-734. https://doi.org/10.1111/j.1471-8286.2007.01863.x
- Shearer TL, Coffroth MA (2008) DNA BARCODING: Barcoding corals: limited by interspecific divergence, not intraspecific variation. Molecular Ecology Resources 8 (2): 247-255. https://doi.org/10.1111/j.1471-8286.2007.01996.x
- Shearer TL, van Oppen MJH, Romano SL, Wörheide G (2008) Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). Molecular Ecology 11 (12): 2475-2487. https://doi.org/10.1046/j.1365-294x.2002.01652.x
- Sieburth J, Smetacek V, Lenz J (2003) Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions 1. Limnology and Oceanography 23 (6): 1256-1263. https://doi.org/10.4319/lo.1978.23.6.1256
- Smith MA, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philosophical Transactions of the Royal Society of London. Series B, Biological sciences 360 (1462): 1825-1834. https://doi.org/10.1098/rstb.2005.1714
- Srivathsan A, Baloğlu B, Wang W, Tan W, Bertrand D, Ng AQ, Boey EH, Koh JY, Nagarajan N, Meier R (2018) A MinION™-based pipeline for fast and cost-effective DNA barcoding. Molecular Ecology Resources 18 (5): 1035-1049. https://doi.org/10.1111/1755-0998.12890
- Steinke D, Zemlak TS, Hebert PDN (2009) Barcoding nemo: DNA-based identifications for the ornamental fish trade. PLOS ONE 4 (7): e6300. https://doi.org/10.1371/journal.pone.0006300
- Sun S, Li Q, Kong L, Yu H, Zheng X, Yu R, Dai L, Sun Y, Chen J, Liu J, Ni L, Feng Y, Yu Z, Zou S, Lin J (2016) DNA barcoding reveal patterns of species diversity among northwestern Pacific molluscs. Scientific Reports 6: 33367. https://doi.org/10.1038/srep33367

- Tan KS, Koh KS, Ng JY, Goh L (2013) The Comprehensive Marine Biodiversity Survey Singapore Strait International Workshop. Raffles Bulletin of Zoology S34: 1-7.
- Tay Y, Todd P, Rosshaug P, Chou L (2012) Simulating the transport of broadcast coral larvae among the Southern Islands of Singapore. Aquatic Biology 15 (3): 283-297. https://doi.org/10.3354/ab00433
- Thomsen PF, Kielgast J, Iversen LL, Wiuf C, Rasmussen M, Gilbert MTP, Orlando L, Willerslev E (2011) Monitoring endangered freshwater biodiversity using environmental DNA. Molecular Ecology 21 (11): 2565-2673. https://doi.org/10.1111/j.1365-294X.2011.05418.x
- Trivedi S, Aloufi AA, Ansari AA, Ghosh SK (2015) Role of DNA barcoding in marine biodiversity assessment and conservation: an update. Saudi Journal of Biological Sciences 23 (2): 161-171. https://doi.org/10.1016/j.sjbs.2015.01.001
- Vanhove M, Tessens B, Schoelinck C, Jondelius U, Littlewood T, Artois T, Huyse T (2013) Problematic barcoding in flatworms: A case-study on monogeneans and rhabdocoels (Platyhelminthes). ZooKeys 365: 355-379. https://doi.org/10.3897/zookeys.365.5776
- van Velzen R, Weitschek E, Felici G, Bakker FT (2012) DNA barcoding of recently diverged species: relative performance of matching methods. PLOS ONE 7 (1): e30490. https://doi.org/10.1371/journal.pone.0030490
- Virgilio M, Backeljau T, Nevado B, De Meyer M (2010) Comparative performances of DNA barcoding across insect orders. BMC Bioinformatics 11: 206. https://doi.org/10.1186/1471-2105-11-206
- Wang W, Srivathsan A, Foo M, Yamane S, Meier R (2018) Sorting specimen-rich invertebrate samples with cost-effective NGS barcodes: Validating a reverse workflow for specimen processing. Molecular Ecology Resources 18 (3): 490-501. https://doi.org/10.1111/1755-0998.12751
- Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 360 (1462): 1847-1857. https://doi.org/10.1098/rstb.2005.1716
- Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. Journal of Fish Biology 74 (2): 329-356. https://doi.org/10.1111/j.1095-8649.2008.02080.x
- Webb C, Slik JWF, Triono T (2010) Biodiversity inventory and informatics in Southeast Asia. Biodiversity and Conservation 19 (4): 955-972. https://doi.org/10.1007/s10531-010-9817-x
- Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. Systematic Biology 54 (5): 844-851. https://doi.org/10.1080/10635150500354878
- Witt JS, Threloff D, Hebert PN (2006) DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. Molecular Ecology 15 (10): 3073-3082. https://doi.org/10.1111/j.1365-294x.2006.02999.x
- Wong JY, Chan YKS, Ng CSL, Tun KP, Darling E, Huang D (2018) Comparing patterns of taxonomic, functional and phylogenetic diversity in reef coral communities. Coral Reefs 37 (3): 737-750. https://doi.org/10.1007/s00338-018-1698-6
- Wong LL, Peatman E, Lu J, Kucuktas H, He S, Zhou C, Na-nakorn U, Liu Z (2011) DNA barcoding of catfish: species authentication and phylogenetic assessment. PLoS ONE 6 (3): e17812. https://doi.org/10.1371/journal.pone.0017812

- Wong WH, Tay YC, Puniamoorthy J, Balke M, Cranston P, Meier R (2014) 'Direct PCR' optimization yields a rapid, cost-effective, nondestructive and efficient method for obtaining DNA barcodes without DNA extraction. Molecular Ecology Resources 14 (6): 1271-1280. https://doi.org/10.1111/1755-0998.12275
- Yamamoto S, Masuda R, Sato Y, Sado T, Araki H, Kondoh M, Minamoto T, Miya M (2017) Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. Scientific Reports 7: 40368. https://doi.org/10.1038/srep40368
- Yap NWL, Tan R, Yong CLX, Tan KS, Huang D (2019) Sea anemones (Cnidaria, Actiniaria) of Singapore: redescription and taxonomy of *Phymanthus pinnulatus* Martens in Klunzinger, 1877. ZooKeys 840: 1-20. https://doi.org/10.3897/zookeys.840.31390
- Yeo D, Puniamoorthy J, Ngiam RWJ, Meier R (2018) Towards holomorphology in entomology: rapid and cost-effective adult-larva matching using NGS barcodes. Systematic Entomology 43 (4): 678-691. https://doi.org/10.1111/syen.12296
- Yip ZT, Quek RB, Huang D (2019) Historical biogeography of the widespread macroalga *Sargassum* (Fucales, Phaeophyceae). Journal of Phycology https://doi.org/10.1111/jpy.12945
- Zhang J, Kobert K, Flouri T, Stamatakis A (2013) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30 (5): 614-620. https://doi.org/10.1093/bioinformatics/btt593
- Zhang R, Zhang B (2014) Prospects of using DNA barcoding for species identification and evaluation of the accuracy of sequence databases for ticks (Acari: Ixodida). Ticks and Tick-borne Diseases 5 (3): 352-358. https://doi.org/10.1016/j.ttbdis.2014.01.001

Supplementary materials

Suppl. material 1: Supplementary Review List S1a `doi`

Authors: Yin Cheong Aden Ip, Ywee Chieh Tay, Su Xuan Gan, Hui Ping Ang, Karenne Tun, Loke Ming Chou, Danwei Huang, Rudolf Meier
Data type: Compiled literature of SIMP biodiversity
Brief description: Existing published records documenting marine fauna at the Sisters' Islands Marine Park (SIMP), based on a keyword search of the literature. Literature keyword search was performed by entering keywords in the following order: "Sisters' Islands" or "Sisters' Islands Marine Park" or "Pulau Subar Laut" or "Pulau Subar Darat" or "Pulau Sakijang Bendera" or "Tanjong Hakim".
Download file (44.81 kb)

Suppl. material 2: Supplementary Review List S1b `doi`

Authors: Yin Cheong Aden Ip, Ywee Chieh Tay, Su Xuan Gan, Hui Ping Ang, Karenne Tun, Loke Ming Chou, Danwei Huang, Rudolf Meier
Data type: Compiled literature of SIMP COI barcodes
Brief description: Compiled literature of species records for which COI barcodes are available on GenBank. Species names from the compiled literature of SIMP biodiversity were searched on GenBank, returning 109 species with COI sequences that were sequenced elsewhere and these were compiled separately as a reference sequence database.
Download file (561.16 kb)

Suppl. material 3: Supplementary Table S1a: SIMP macrofauna `doi`

Authors: Yin Cheong Aden Ip, Ywee Chieh Tay, Su Xuan Gan, Hui Ping Ang, Karenne Tun, Loke Ming Chou, Danwei Huang, Rudolf Meier
Data type: Metadata on macrofaunal specimens collected
Brief description: Information on macrofaunal specimen image availability, taxonomic information, genetic identity match to both GenBank and BOLD system databases, collection information, barcode availability, barcode length (long ≈ 658bp; short ≈ 313bp), LKCNHM Zoological Reference Collection (ZRC) catalogue numbers and GenBank numbers of all collected specimens.
Download file (33.18 kb)

Suppl. material 4: Supplementary Table S1b: SIMP zooplankton `doi`

Authors: Yin Cheong Aden Ip, Ywee Chieh Tay, Su Xuan Gan, Hui Ping Ang, Karenne Tun, Loke Ming Chou, Danwei Huang, Rudolf Meier
Data type: Metadata on zooplankton specimens collected
Brief description: Information on zooplankton specimen image availability, taxonomic information, genetic identity match to both GenBank and BOLD system databases, collection information, barcode availability, barcode length (long ≈ 658bp; short ≈ 313bp), LKCNHM Zoological Reference Collection (ZRC) catalogue numbers and GenBank numbers of all collected specimens.
Download file (33.18 kb)

Suppl. material 5: Supplementary Figure S1a `doi`

Authors: Yin Cheong Aden Ip, Ywee Chieh Tay, Su Xuan Gan, Hui Ping Ang, Karenne Tun, Loke Ming Chou, Danwei Huang, Rudolf Meier
Data type: Cluster dendrogram
Brief description: Cluster dendrogram based on percentage pairwise differences in COI for all 632 specimens with COI barcodes. Values at the nodes represent the percentage pairwise difference between two specimens. Taxon names on the branches represent taxonomic identities, based on morphological identification.
Download file (97.33 kb)

Suppl. material 6: Supplementary Figure S1b `doi`

Authors: Yin Cheong Aden Ip, Ywee Chieh Tay, Su Xuan Gan, Hui Ping Ang, Karenne Tun, Loke Ming Chou, Danwei Huang, Rudolf Meier
Data type: Cluster dendrogram
Brief description: Cluster dendrogram based on percentage pairwise differences in COI for 304 specimens (excluding IP0303) with species-level identification. Values at the nodes represent the percentage pairwise difference between two specimens. Taxon names on the branches represent taxonomic identities, based on morphological identification.
Download file (59.51 kb)