Journal of Crustacean Biology Advance Access published 2 November 2018

The Crustacean Society

Journal of Crustacean Biology

Journal of Crustacean Biology 38(6), 780–793, 2018. doi:10.1093/jcbiol/ruy084

DNA barcoding as a tool to facilitate the taxonomy of hermit crabs (Decapoda: Anomura: Paguroidea)

Jannes Landschoff[1,] and Gavin Gouws[2,]

[1]*Department of Biological Sciences and Marine Research Institute, University of Cape Town, Rondebosch, South Africa; and* [2]*National Research Foundation – South African Institute for Aquatic Biodiversity, Grahamstown, South Africa*

Correspondence: J. Landschoff: email: jannes@landschoff.net

(Received 7 June 2018; accepted 26 August 2018)

ABSTRACT

DNA barcoding has become an important aspect in biodiversity research, but is ineffective without comprehensive databases that can relate molecular sequences to scientific names. Hermit crabs, or paguroids, are an important component of the marine benthos, and are vastly understudied in southern Africa, a globally important bioregion. The study provides 194 cytochrome *c* oxidase subunit I (COI) barcode sequences (> 550 bp) of 43 nominal and 12 putative species from 19 genera and four families of Paguroidea. Preliminary COI-based phylogenetic analysis suggested some overall pattern of monophyly in the Parapaguridae and *Paguristes* Dana, 1851 and *Dardanus* Paul´son, 1875. Large groups in Diogenidae, however, such as *Calcinus* Dana, 1851 and *Diogenes* Dana, 1851, were recovered as poly- or paraphyletic. Testing the dataset for barcoding efficacy revealed that the technique can reach identification rates near 100% success. A barcoding gap was apparent for almost all but one problematic taxon. The optimum threshold range to discriminate between species for this set of data was calculated at 3.7–3.9%, although the overall mean intraspecific distance was well below 1%. Furthermore, glaucothoe larvae of two deep-water paguroids were matched to names in the new molecular library. The COI marker alone, however, did not provide clear species identifications in three problematic species, *Areopaguristes* cf. *jousseaumei* (Bouvier, 1892), *Goreopagurus poorei* Lemaitre & McLaughlin, 2003, and *Pagurus cuanensis* Bell, 1846. Despite the dataset having considerable species coverage of 55 total taxa, it is still estimated to be only about 50% complete in terms of the South African fauna. Future work should therefore increase coverage to other species in South Africa and expand geographic coverage for widely distributed species.

Key Words: COI, integrated taxonomy, molecular barcoding

INTRODUCTION

Although the usefulness of single-gene DNA fragments as tools to aid classical taxonomy has created great controversy (Ebach & Holdrege, 2005; Hickerson *et al.*, 2006; Rubinoff, 2006; Carvalho *et al.*, 2007; Roe & Sperling 2007; Ebach, 2011), much of the earlier debate was fuelled by the promise that sequencing all life forms on Earth alone would be sufficient to decipher the many problems taxonomists are confronted with (Sperling, 2003; Meier *et al.*, 2006). The initially theoretical discussion over the principal prospects of DNA taxonomy (Tautz *et al.*, 2003) using DNA barcoding (Hebert *et al.*, 2003) has streamlined into more empirical evaluations revealing both the benefits and limitations of the barcoding approach (Meier *et al.*, 2006; Costa *et al.*, 2007; Hajibabaei *et al.*, 2007; Ward, 2009; Robe *et al.*, 2012; Shackleton & Rees, 2015). DNA barcoding should be understood as a complementary item in taxonomy (Stevens *et al.*, 2011) and, as such, it primarily remains a mechanism to assist in species identification (DeSalle, 2006; Collins & Cruickshank, 2012). Barcoding may, however, fall short depending on the taxon and study design (see Prendini, 2005). Because the technique is vastly ineffective without barcode databases of sufficient species coverage and that are founded on verified, curated specimens (Funk & Omland, 2003; Meyer & Paulay, 2005), biologists are calling for increased integrative taxonomic knowledge (Will *et al.*, 2005; Padial & De La Riva, 2007; Boero, 2010; Padial *et al.*, 2010, Goldstein & DeSalle, 2011; Cristescu, 2014). Global-scale databases that contain adequate taxonomic and molecular coverage remain scarce (Kvist, 2013). They are, however, particularly needed for species-rich,

taxonomically-ambiguous groups. In a carefully designed study and for taxa in which DNA barcoding achieves sufficient resolution, the tool can then support and help to refine taxonomic hypotheses on species delineations, as applied in anomuran crustaceans (e.g. Puillandre *et al.*, 2011).

Crustaceans comprise one of Earth's most diverse animal groups and are especially diverse in the oceans in habitats that are explored the least (Costello *et al.*, 2010). Decapods such as crabs, shrimps, lobsters, and hermit crabs are particularly large and recognisable, and form an ecologically- and economically-important group known to contain about 15,000 extant species (De Grave *et al.*, 2009). Decapod crustaceans have been subject to successful barcoding studies (Lefébure *et al.*, 2006; Costa *et al.*, 2007; Radulovici *et al.*, 2009; Matzen da Silva *et al.*, 2011; Puillandre *et al.*, 2011; Robe *et al.*, 2012; Meyer *et al.*, 2013; Bilgin *et al.*, 2014; Raupach *et al.*, 2015), showing the great potential of the technique for taxonomic and biodiversity research. Compiling sufficient regional information that would enable analysis of decapod diversity on a global scale, however, remains an enormous challenge. Matzen da Silva *et al.* (2011) estimated that only 5.4% of all decapod species were represented in the available barcoding databases. Another problem is that voucher specimens that correspond to barcoding libraries are often not identified by taxonomists, and that they are geographically-biased towards regions where biodiversity research is better funded, and which are consequently better known. This leaves major gaps, particularly in marine habitats in the developing world, which include some important biodiversity regions. Data are particularly scarce across Africa. This study attempts to fill this gap by compiling and analysing a taxonomically-comprehensive barcoding dataset for the group of hermit crabs predominantly from South Africa.

Most paguroids inhabit empty marine gastropod shells, and their bodies are evolutionarily-adapted to this unusual habitat (Hazlett, 1981). The posterior parts of the body are almost entirely soft, forming few calcified characters that could be used to inform taxonomy. Yet most species of hermit crabs frequently utilize a suite of different types of shells (see Barnes, 2003; McLaughlin, 2015) with varying three-dimensional properties. This behaviour is known to influence and change growth rates in hermit crabs in different housings (Bertness, 1981), resulting in and enhancing intraspecific variation (Blackstone, 1985). Having half their body parts soft and being subject to high rates of intraspecific variation has contributed to the currently poorly understood taxonomy of many paguroids. Barcoding could provide additional resolution needed to identify and delineate species, particularly among closely-related species of difficult paguroid groups. Sequencing of the mitochondrial cytochrome *c* oxidase subunit I (COI) molecular marker has already been applied to paguroids, for example, to inform the taxonomy of species of *Clibanarius* Dana, 1852 and *Coenobita* Latreille, 1829 at a regional scale (Negri *et al.*, 2014; Rahayu *et al.*, 2016), as well on the entire genus *Calcinus* Dana, 1851 in an extensive global study (Malay & Paulay, 2010). No COI barcoding dataset of paguroids has nevertheless been tested across a larger range of genera and species, or even families.

We aimed to generate a DNA barcoding reference library for identified and curated hermit crab specimens from South Africa, which reciprocally informs and supports morphology-based identifications in an ongoing nation-wide biodiversity assessment. Such database was tested for efficacy as an identification tool for hermit crabs on a broad scale. The results from the divergence analyses could also be used to refine knowledge and ideas of speciation in Paguroidea, or in Decapoda as a whole. Some preliminary phylogenetic considerations are also provided given the limitations of the single mitochondrial marker. Aspects of previously unrecognised diversity, such as the presence of cryptic species, are discussed.

MATERIALS AND METHODS

Sampling, specimen identifications, and species concept

Hermit crabs were collected along the Atlantic and Indian oceans coasts of South Africa from August 2014 to April 2017. Sampling by hand, diving, dredging, and trawling yielded about 550 specimens from all possible habitats. Except for some glaucothoe larvae, all specimens were identified to species level by morphological examinations following the literature (Barnard, 1950; Forest, 1954; McLaughlin, 1998; McLaughlin & Forest, 1999; Landschoff & Lemaitre, 2017a, b; Landschoff, 2018). Specimens were also frequently reidentified, and in some instances morphology- and barcoding-based identifications were also reciprocally informed as sequences became available. This was the case for some juvenile specimens for which initial identifications were uncertain, and for four sequences obtained from glaucothoe larvae. These sequences, matched to names from the preliminary database, were thus included in the final analysis. Morphological separation of the putative species of *Diogenes* and *Paguristes*, which could not be done using existing taxonomic literature, was primarily based on colour patterns and the shape and armature of chelipeds and ambulatory legs. Species entities in the text are referred to as 'species' when a certain level of morphological evidence justified the assignment of a name from the literature, or are referred to as 'putative species' when no morphological identification to an existing taxon name was possible. Species entities were regarded and tested following the concept of 'species-like-units' (Collins & Cruickshank, 2012).

Where possible, up to five specimens of each taxon were selected for barcoding; additional specimens were selected from different localities when a taxon was sampled over a wide range, or when specimens were of particular interest as in colour variants or juveniles. Fresh tissue was usually taken from unpreserved specimens after freezing/defrosting. Tissue was preferably taken as eggs from ovigerous females or as muscle tissue, most often from the merus of the larger cheliped. In some small-size species, tissue was retrieved by dissecting one walking leg and using the coxa for barcoding. Larger body parts or even entire specimens were used in abundant, particularly small species. Some tissue was also taken from museum material following similar protocols as for fresh material. Most tissue, placed in 96% ethanol, was sent for barcoding at the National Research Foundation – South African Institute for Aquatic Biodiversity (NRF-SAIAB), Grahamstown, South Africa. Other tissues were submitted to the Canadian Centre for DNA Barcoding (CCDB), Biodiversity Institute of Ontario, University of Guelph (Guelph, Canada). Most specimens are deposited in the Iziko South African Museum, Cape Town, South Africa, a few in the National Museum of Natural History, Smithsonian Institution, Washington, D.C., USA. Specimens included from the Lee Kong Chian Natural History Museum, Singapore are from South Africa, but older than the timeframe indicated above. Barcodes were also generated for some specimens from Mozambique, southern Africa, Marion Island, non-continental South Africa, Australia (Museum Victoria, Melbourne) and New Zealand (now in the Iziko South African Museum (Supplementary material Table S1).

DNA extractions and sequencing

DNA was extracted from tissues using the protocol of Sunnucks & Hales (1996), followed by the amplification of the barcoding (Hebert *et al.*, 2003) fragment of the COI gene by polymerase chain reaction (PCR), using the universal invertebrate primers (LCOI-1490 and HCOI-2198) of Folmer *et al.* (1994) or their degenerate variants (dgLCOI490 and dgHCO2198; Meyer, 2003). PCR recipes and conditions followed Meyer (2003) and Gouws *et al.* (2015), with annealing performed at 48 °C for the latter. Successful amplification was determined by visualising products under UV light, following electrophoresis in 1% agarose gels, stained with

ethidium bromide, in a TBE buffer. PCR products were purified with an Exonuclease I - Shrimp Alkaline Phosphate (Exo/SAP, ThermoFisher Scientific, Waltham, MA, USA) protocol (Werle *et al.*. 1994), sequenced in both the forward and reverse directions using BigDye v3.1 (Applied Biosystems, Austin, TX, USA) terminator chemistry and analysed on an ABI-Hitachi 3500 Genetic Analyser (Applied Biosystems, Foster City, CA, USA). The resulting sequences were tested against their chromatograms for misreads and sequencing errors using ChromasLITE (Technylesium, South Brisbane, Australia). Sequences were aligned, edited, and the consensus DNA barcode compiled using Lasergene SeqMan Pro 9 (DNASTAR, Madison, WI, USA). Barcodes and trace files were uploaded to the South African SeaKeys (SEAKY) project, and to a global general decapod (GMDEC) project on BOLD (www.boldsystems.org; Ratnasingham & Hebert, 2007), using the BOLD inbuilt quality control functions for detecting stop-codons and annotating ambiguous bases. Sequences shorter than 550 basepairs (bp) were removed from the dataset. Specimens barcoded by the CCDB followed that institution's protocols, with the data being uploaded to the SeaKeys project on BOLD, as above.

Genetic analysis

Alignment and tree-building. Alignment of the COI barcodes was conducted in ClustalX 2.1 (Larkin *et al.*, 2007). Neighbour-joining (NJ) tree building (Saitou & Nei, 1987) of the aligned sequences was conducted in MEGA7 (Kumar *et al.*, 2016), using Kimura's 2-parameter (K2P) model of sequence evolution (Kimura, 1980), and with nodal support assessed using 1,000 bootstrap replicates (Felsenstein, 1985). Final trees were drawn in FigTree v1.4.3 (tree.bio.ed.ac.uk/software/figtree) and edited in Inkscape 0.91 (inkscape.org).

Distance analysis. All following analyses were conducted in the *R* 3.3.1 and *R* 3.4.3 statistical environments (R Core Team, 2016) using the packages *spider* (Brown *et al.*, 2012) and *sidier* (Muñoz-Pajares, 2013) within the Rstudio interface (RStudio Team, 2016). Inter- and intraspecific pairwise distances were calculated, corrected using the K2P model, and mean, minimum and maximum distances retrieved from comparisons of sequence pairs. Distances were also computed within or between congeneric and intrafamiliar taxa to calculate the overlap of distances on species-genus and genus-family levels. The taxonomic resolution ratio (TRR) was calculated as the quotient between congeneric interspecific and intraspecific divergences (Costa *et al.*, 2007) to provide a general idea of taxon resolution on generic level. The difference between intra- and interspecific distances, also known as the barcoding gap (BG) (Meyer & Paulay, 2005), was derived by subtracting the minimum interspecific from the maximum intraspecific distance (Meier *et al.*, 2008; Robe *et al.*, 2012). A variation of the BG, based on subtractions of the smallest interspecific from the largest intraspecific K2P distance value (Brown *et al.*, 2012) was also explored instead of the overall minimum and maximum values for the species.

Testing barcoding efficacy. Along with an analysis to find the threshold that would best separate hermit crab taxa, the barcoding efficacy for the COI dataset was tested using both Meier's best close match (Meier *et al.*, 2006) and the threshold-based analysis functions *threshID* and *threshOpt* provided by the *spider* package in *R*. Species boundary thresholds, as proposed in the literature: 3% (Hebert *et al.*, 2003), 1.0% (Ratnasingham & Hebert, 2007), 2% (Ratnasingham & Hebert, 2013), and calculated for various sets of data: 1% (10× intraspecific divergence) (Hebert *et al.*, 2004), 4.6% (assumption-free statistical approach) (Lefébure *et al.*, 2006), 3.0% (95% rule) (Meier *et al.*, 2006), and 1.9% (localMinima method) (Brown *et al.*, 2012), were tested for the highest identification success rates. An optimal threshold range was also calculated and tested after Brown *et al.* (2012).

RESULTS

COI dataset and quality

A total of 194 COI barcode sequences (> 550 bp) of 43 nominal and 12 putative species from 19 genera and four families were retrieved and analysed. The majority of these taxa belonged to Diogenidae (39 taxa, 131 sequences), whereas the Paguridae (nine taxa, 40 sequences) and Parapaguridae (six taxa, 23 sequences) were less represented, and only one barcode of *Coenobita rugosus* H. Milne Edwards, 1837 (Coenobitidae) was retrieved (Table 1). Of the total 55 taxa, 16 were represented by a single sequence. On average, each taxon was represented by four barcode sequences. A maximum of 16 barcodes were sequenced for the widespread *Diogenes brevirostris* Stimpson, 1858 (Fig. 1). Quality control detected stop codons, probably indicating nuclear mitochondrial pseudogenes (numts), in two sequences of *Calcinus guamensis* Wooster, 1984 (Fig. 2). The vast majority of the 177 sequences consisted of the normal length of 658 bp and had no ambiguities, but some were shorter and only one was below 600 bp (Figs. 2–4). Three sequences were derived from glaucothoe larvae of *Sympagurus dimorphus* Studer, 1883, and one of a glaucothoe of *Parapagurus bouvieri* Stebbing, 1910 (Fig. 4). Six sequences from three taxa (*Pagurus cuanensis* Bell, 1846, *Clibanarius virescens* (Krauss, 1843), and *Paguristes* sp.1) had > 2 (i.e., 4–7 n) ambiguous nucleotide bases (Fig. 1). Exploratory analyses including or excluding these sequences of lesser quality led to the same or very similar results. These sequences were thus included in the final dataset analysed, but respective data quality flags were retained (Figs. 2–4 and Supplementary material Table S1).

Phylogenetic aspects

The NJ tree of the 55 taxa supported monophyly at the species level (but see below). Paguridae and Parapaguridae were recovered as monophyletic, whereas Diogenidae was poly- or paraphyletic (Fig. 1). Family-level support, however, was low in this single

Table 1. Worldwide and South Africa diversity of Paguroidea; *worldwide figures after McLaughlin & Türkay (2018); **South African figures after Landschoff (2018); list of genera and species in this study include some species from outside South Africa.

Family	Genera (species) worldwide*	Genera (species) South Africa**	Genera (species) this study
Coenobitidae Dana, 1851	2 (18)	1 (3)	1 (1)
Diogenidae Ortmann, 1892	21 (464)	10 (39)	10 (39)
Paguridae Latreille, 1802	88 (587)	6 (19)	5 (9)
Parapaguridae Smith, 1882	10 (91)	5 (10)	3 (6)
Pylochelidae Bate, 1888	10 (47)	1 (1)	0 (0)
Pylojacquesidae McLaughlin & Lemaitre, 2001	2 (2)	0 (0)	0 (0)
Total	**133 (1209)**	**23 (72)**	**19 (55)**

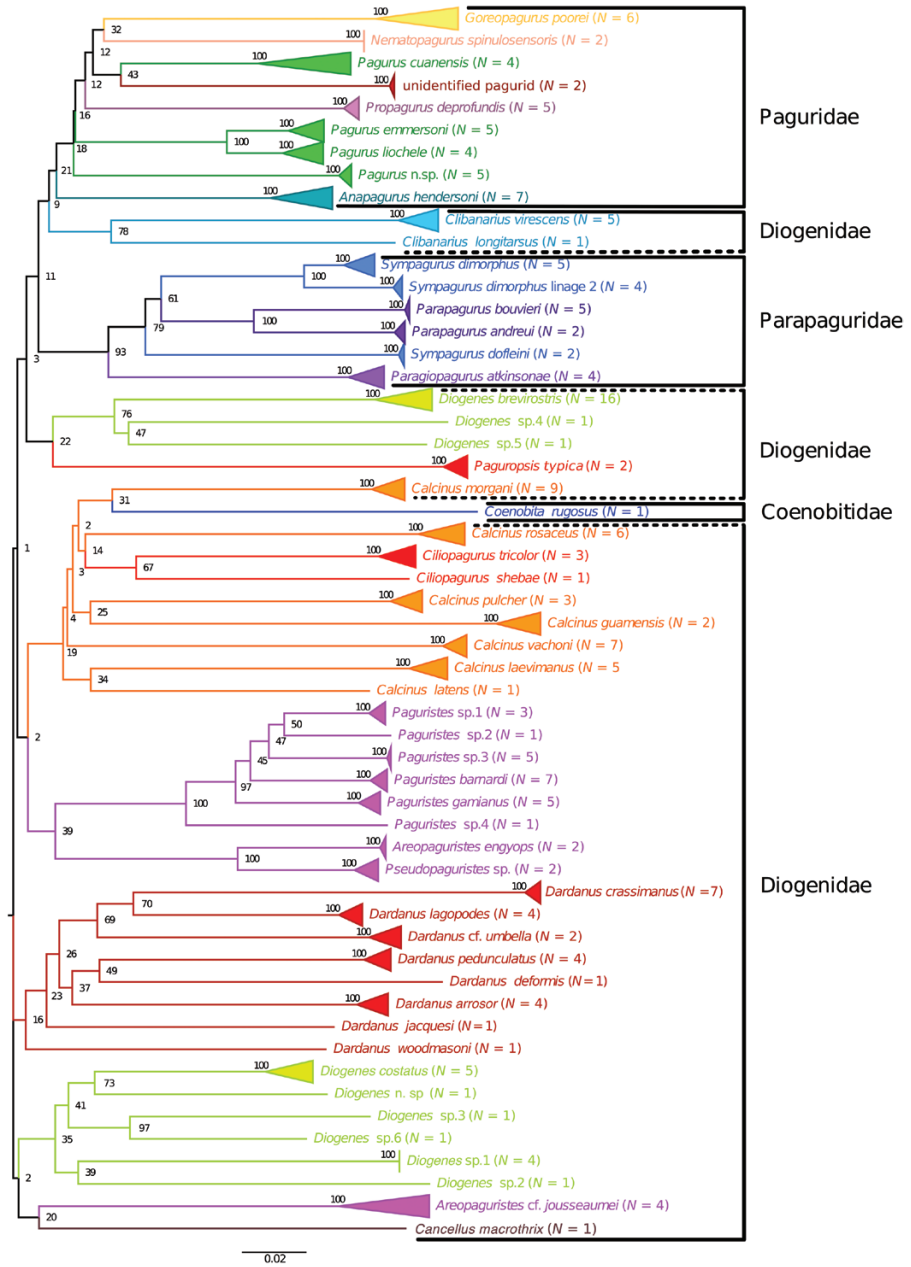

Figure 1. Neighbour-joining tree of distance relationships between hermit crabs, predominantly from South Africa, constructed using the cytochrome *c* oxidase I (COI) barcoding fragment and Kimura 2-parameter (K2P) distances. Bootstrap values based on 1000 replicates are indicated next to nodes, K2P divergences are shown by scale bar (%); *N*, sample size for each species. This figure is available in colour at *Journal of Crustacean Biology* online.

marker analysis although Parapaguridae had 93% bootstrap support. Nodal support for congeneric relationships was low within Paguridae except for the sister-taxa relationship of *P. emmersoni* McLaughlin & Forest, 1999 and *P. liochele* (Barnard, 1947), which was well supported (100%). The polyphyly of *Pagurus* Fabricius, 1775 was apparent (Figs. 1, 4). Nodal support for congeneric relationships within Parapaguridae was higher, with *Paragiopagurus* Lemaitre, 1996 appearing basal in this family. With only few taxa represented, *Parapagurus* was recovered as monophyletic, clustering within the poly- or paraphyletic *Sympagurus* Smith, 1883.

The single sequence of *Coenobita rugosus* (Coenobitidae) clustered within the Diogenidae within *Calcinus*. *Diogenes* Dana, 1851 was recovered as polyphyletic (Figs. 1, 2), forming two distinct moderately (76%, *D. brevirostris* clade) and poorly (35%, *D. costatus* clade) supported clades. The well-represented *Calcinus* was found to be poly- or paraphyletic, containing a member of *Coenobita*, but also the closely-related *Ciliopagurus* Forest, 1995 (Figs. 1, 2). *Dardanus* Paul′son, 1875 was recovered as monophyletic (16%), whereas *Paguristes* Dana, 1851 was the only monophyletic genus with high (100%) bootstrap support within this family. Sister to the *Paguristes* clade, *Areopaguristes engyops* (Barnard, 1947) and *Pseudopaguristes* McLaughlin, 2002 formed a well-supported sister-taxa relationship, but the former was notably distant from the other species of *Areopaguristes* Rahayu & McLaughlin, 2010, *A.* cf. *jousseaumei* (Bouvier, 1892). The positions of *Paguropsis* Henderson, 1888 and *Cancellus* H. Milne Edwards, 1836 within Diogenidae had little support (Figs. 1, 2).

COI divergence analysis

Mean, minimum, and maximum intraspecific K2P distances for all taxa represented by more than one sequence are summarised in Supplementary material Table S2. At 0.0081 (± 0.0016 SE),

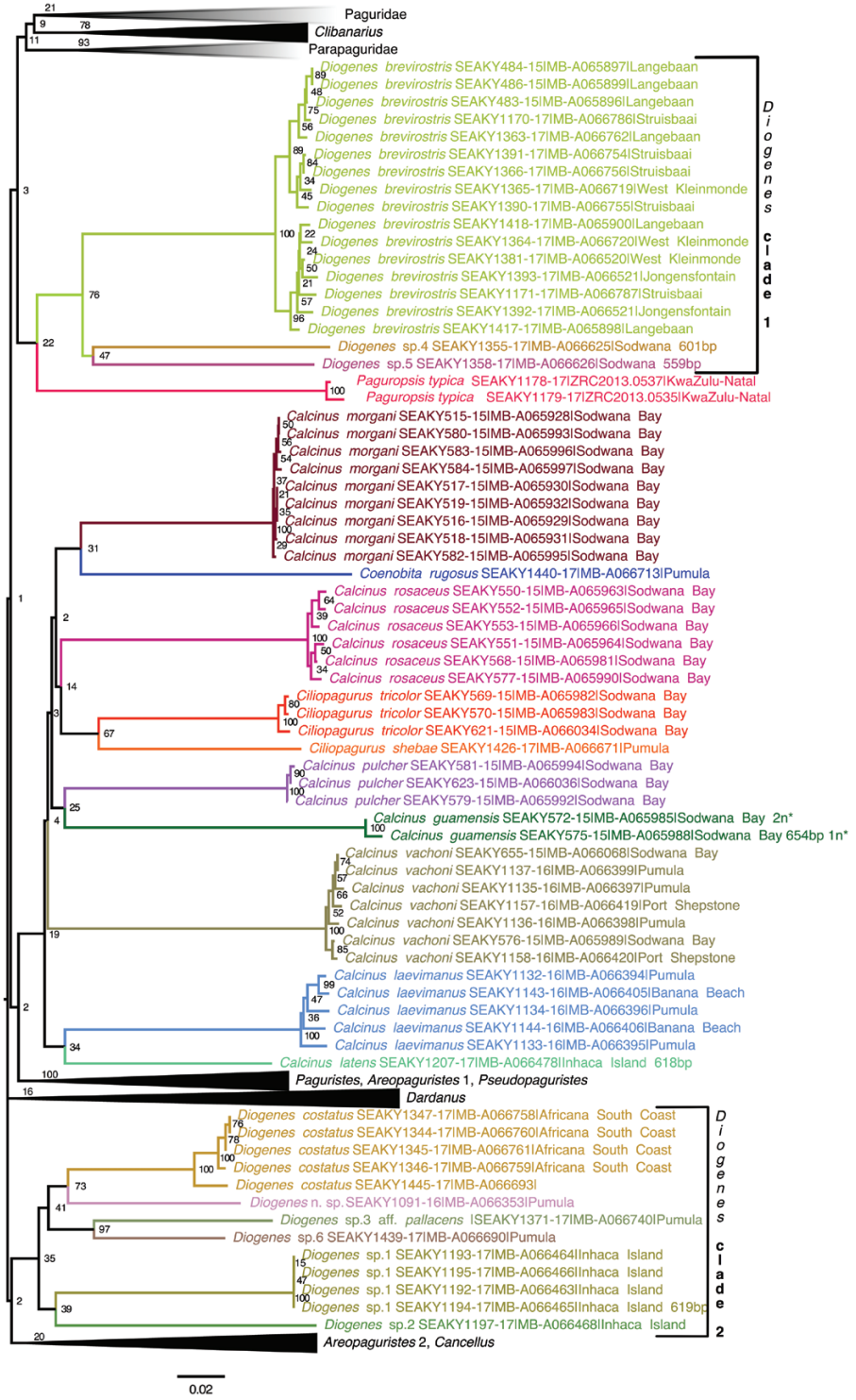

Figure 2. Neighbour-joining tree of distance relationships between South African hermit crabs (except where indicated), constructed using the cytochrome *c* oxidase I (COI) barcoding fragment and Kimura 2-parameter (K2P) distances. Detail of Diogenidae 1. Bootstrap values based on 1,000 replicates are indicated next to nodes, K2P divergences are shown by scale bar (%); *, stop codons; n, ambiguous bases. Sequence length 658 bp if not specified otherwise. This figure is available in colour at *Journal of Crustacean Biology* online.

the overall mean intraspecific distance was well below 1%, and 93.65% of the 441 intraspecific comparisons fell below 3%. The highest maximum distance resulted from intraspecific comparisons within *Anapagurus hendersoni* Barnard, 1947 (3.5%), *Diogenes brevirostris* (3.5%), *Areopaguristes* cf. *jousseaumei* (4.3%), *Goreopagurus poorei* Lemaitre & McLaughlin, 2003 (5.0%), and *Pagurus cuanensis* (5.1%). Congeneric distances varied from a minimum of 3.9%, between *Pagurus emmersoni* and *P. liochele*, to maximum values of 27–28% when comparing congeneric sequences of *Calcinus*, *Dardanus*, and *Diogenes* (Supplementary material Table S3). In the 1379 congeneric comparisons, 99.85% were above 4%. The overlap of intraspecific, congeneric interspecific, and intergeneric distances for the three main families is given in Figure 5. No overlap (or only partial in Paguridae) was observed between intraspecific

and interspecific (congeneric) divergences, but major overlap was apparent between divergences from congeneric and intergeneric comparisons in all three families; the latter being most apparent in Diogenidae. The mean congeneric and intergeneric distances were also notably higher for Diogenidae (about or > 20%) than for Paguridae or Parapaguridae (< 20%).

Following the BG of Brown *et al.* (2012), all genera exhibited a measurable gap. When averaging intra- and interspecific distances of each genus first, however, and then subtracting the minimum interspecific from the maximum intraspecific distance (Robe *et al.*, 2012), an overlap of 1.2% was found in the case of *Pagurus* (Supplementary material Table S3). Although present, the gap was small in *Sympagurus* (2.7%) and *Paguristes* (4.1%). It was close to, or larger than, 10% for all other genera examined. Generally reflecting the trends of the BG analyses, and highlighting taxonomic complexities (see discussion below), the TRR values were lowest in *Areopaguristes* (11.528) and *Sympagurus* (11.287), and highest in *Clibanarius* (113.125; Supplementary material Table S2).

Thresholds and dataset efficacy

Testing the dataset using various distance thresholds (1–8.1%) resulted in a range of success rates (2–39 cumulative errors) in false negative (congeneric species identified as conspecifics) or false positive (conspecifics regarded as congeneric species) identifications (Table 2). Meier's best close match (Meier *et al.*, 2006) and the spider functions threshID/threshOpt had the same results. Sequences of the 16 species with only a single sequence present could not be matched and hence were not positively identified using the dataset. The lowest threshold (1%) tested resulted in 20 false positive barcoding identifications, whereas the largest threshold (8.1%), corresponding to ten times the mean intraspecific divergence, resulted in 39 false negatives. The threshold derived from the assumption-free statistical approach of Lefébure *et al.* (2006) resulted in 13 cumulative misidentifications (Fig. 6). With only three false positive errors, the 3% threshold of Hebert *et al.* (2003), which also coincided with the 95% rule of Meier *et al.* (2006), was close to the optimum thresholds (3.7–3.9%) that had a minimum of only two false positives and 192 correctly-identified sequences. Applying this optimal threshold range therefore resulted in a barcoding efficacy of 98.96%.

DISCUSSION

Data scrutiny and study limitations

Seventy-two nominal species of hermit crab are currently reported from South African mainland waters (Landschoff, 2018). Of these, 39 nominal taxa (54%) are represented in this dataset, with sequences of 12 more putative species all from South Africa. Sequences are also included for *Dardanus woodmansoni* (Alcock, 1905) and two unidentified or putatively new species, *Diogenes* sp.1 and sp.2, from southern Mozambique, neither of which are represented by a current South African record. The dataset also includes additional hermit crab specimens, regarded as a distinct taxon, *Sympagurus dimorphus* lineage 2, from New Zealand and the non-mainland South African waters of Marion Island.

Despite a few problematic sequences, the overall quality of the sequence data can be considered high. With six sequences having more than two ambiguous bases, two sequences having stop codons, potentially indicating numts, and 17 sequences being slightly shorter than the normal 658 bp, about 90% of the barcodes in this dataset have high sequencing quality. Future users of the data should consider that the sequences of lower quality might nevertheless influence barcoding identifications. Because general aspects of homologous comparison are violated in COI-like (i.e., numts or pseudogenes) fragments (Buhay, 2009), the two sequences of *Calcinus guamensis* representing potential numts will most likely lead to uncertain identifications, even when compared to sequences of conspecific specimens.

Table 1 gives an overview on the representation of the genera and species represented in this dataset compared to the known global and South African diversity. Of the 23 genera occurring in South African waters, 18 are represented in this study. The dataset, however, is inevitably biased by unequal taxon sampling. Some taxa are well covered, as they are common and widespread, whereas 16 taxa are represented by a single sequence only, as only one or a few individuals were available. The dataset is also limited in its geographic scope. While geographic coverage is generally better for taxa endemic to the South African region, almost all species from the subtropical east coast north of Durban have widespread Indo-West Pacific distributions. Future paguroid barcoding programmes in this region of the Indian Ocean should therefore extend the geographical taxon coverage, ideally to the maximum true occurrence of each species. The current study is, however, valuable in that it provides records of those widespread taxa from their southwestern-most geographical distribution in South Africa.

Phylogenetic considerations

While many phylogenetic aspects of Paguroidea are reasonably well-resolved at higher taxonomic levels or within the Anomura (Mclaughlin, 1983; Tudge, 1997; McLaughlin *et al.*, 2007; Tsang *et al.*, 2008, 2011; Bracken-Grissom *et al.*, 2013), intrafamiliar relationships of hermit crabs are generally poorly understood.

Table 2. Hermit crab COI barcoding efficacy evaluation using the *threshOpt* function provided in the *spider* package (Brown *et al.*, 2012) in *R* (R Core Team, 2016). True and false positives as well as negatives tested for thresholds proposed in the literature given, and optimised for the current dataset; *values calculated for this dataset.

Source/criterion	Threshold	True negative	True positive	False negative	False positive	Cumulative error
BOLD species identification	1.0%	16	158	0	20	20
Ratnasingham & Hebert (2013)	2.0%	16	173	0	5	5
Hebert *et al.* (2003) **and** Meier *et al.* (2006) **95% rule**	3.0%	16	175	0	3	3
Hebert *et al.* (2004) **10× intraspecific divergence**	8.1%	15	140	39	0	39
Lefébure *et al* (2006) **assumption-free statistical approach***	4.6%	16	165	12	1	13
Brown *et al.* (2012) ***localMinima* method***	1.9%	16	173	0	5	5
Brown *et al.* (2012) **threshold optimisation***	3.7–3.9%	16	176	0	2	2

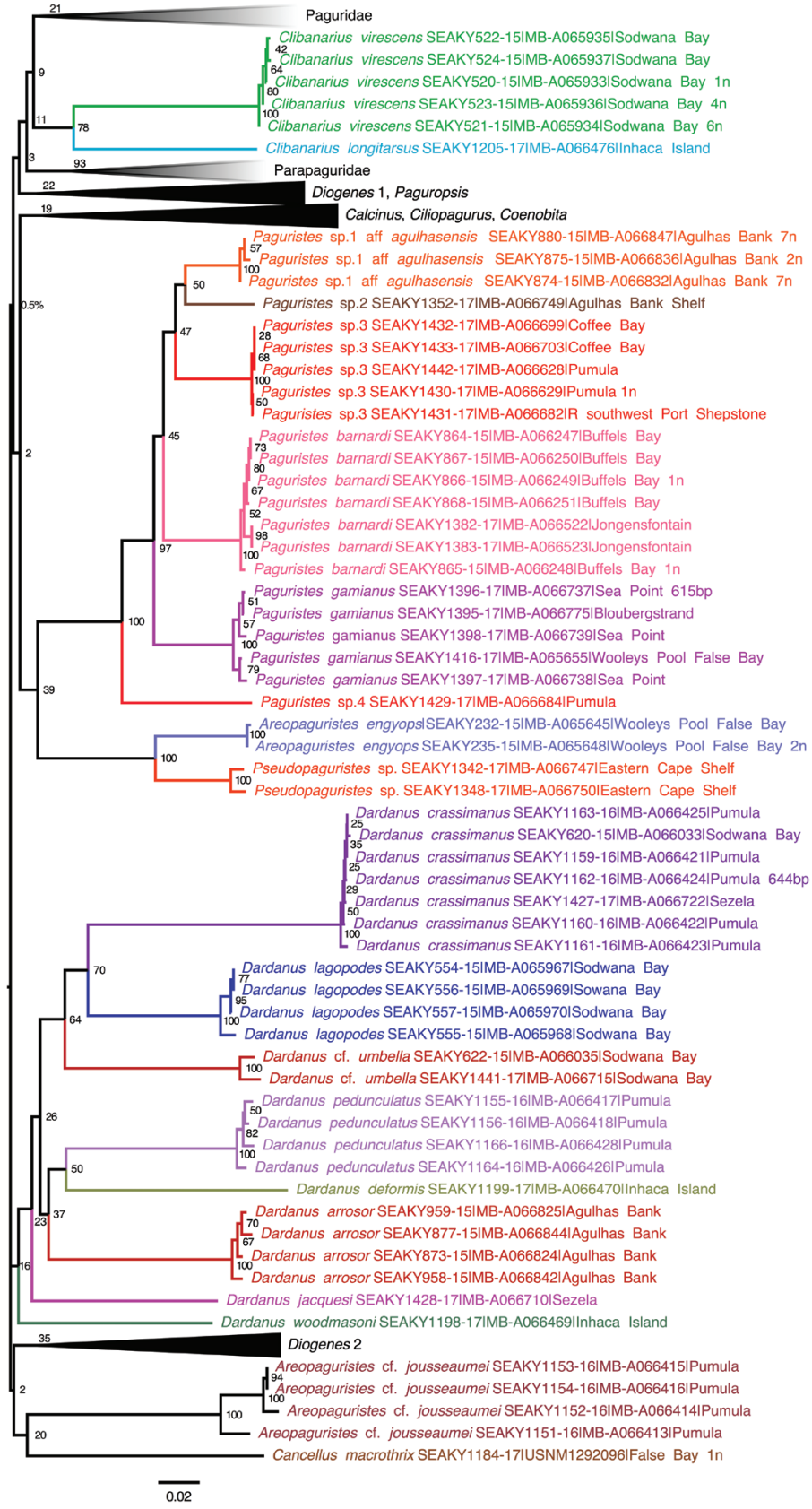

Figure 3. Neighbour-joining tree of distance relationships between South African hermit crabs (except where indicated), constructed using the cytochrome *c* oxidase I (COI) barcoding fragment and Kimura 2-parameter (K2P) distances. Detail of Diogenidae 2. Bootstrap values based on 1,000 replicates are indicated next to nodes, K2P divergences are shown by scale bar (%); *, stop codons; n, ambiguous bases. Sequence length 658 bp if not specified otherwise. This figure is available in colour at *Journal of Crustacean Biology* online.

Although mitochondrial DNA alone is unsuitable to resolve evolutionary relationships (Galtier *et al.*, 2009) and phylogenies derived from COI in Decapoda must be interpreted with caution (Chu *et al.*, 2009), this study may point towards some future avenues of phylogenetic investigations on Paguroidea. Our results once more confirm the para- or polyphyly of *Pagurus*, previously proposed (McLaughlin, 1974; Komai, 1998; Bracken-Grissom *et al.*, 2013). The very close genetic distance of *Pagurus emmersoni* to *P. liochele*, however, is supported of their close morphological similarities (McLaughlin and Forest, 1999). The other four genera tested in Paguridae, and the unidentified pagurid, were only represented by a single species and hence recovered as monophyletic. Relative to global diversity, the typically species-rich Paguridae was the least represented family in the study (Table 1).

The recovery of Parapaguridae as monophyletic is consistent with other molecular studies (Tsang *et al.*, 2011; Bracken-Grissom *et al.*, 2013), and the results further suggest that the monophyletic *Parapagurus* could be nested within *Sympagurus*. Our data reveal that *Paragiopagurus* and *Sympagurus* may not be monophyletic. This should be explored further as both genera are rather related by the absence of morphological characters than by synapomorphy (Lemaitre, 1996).

Based on nuclear, protein-coding genes (Tsang *et al.*, 2011) or the combination of mitochondrial and nuclear markers and morphological characters (Bracken-Grissom *et al.*, 2013), non-monophyletic relationships have been discovered for Diogenidae that are similar to our results (Figs. 1, 2). The position with moderately high support (78%) of *Clibanarius* between Paguridae and Parapaguridae outside Diogenidae, however, was surprising. Previous molecular studies (Tsang *et al.*, 2011; Bracken-Grissom *et al.*, 2013) usually found members of *Clibanarius* clustering well within Diogenidae. Our results further showed that *Dardanus* is likely to be monophyletic, even if the support for incorporating *D. woodmansoni*, which could be the basal generic taxon, was weak (16%; Figs. 1, 3). As a result of high diversity combined with the lack of species-delimiting morphological characters, *Paguristes* and

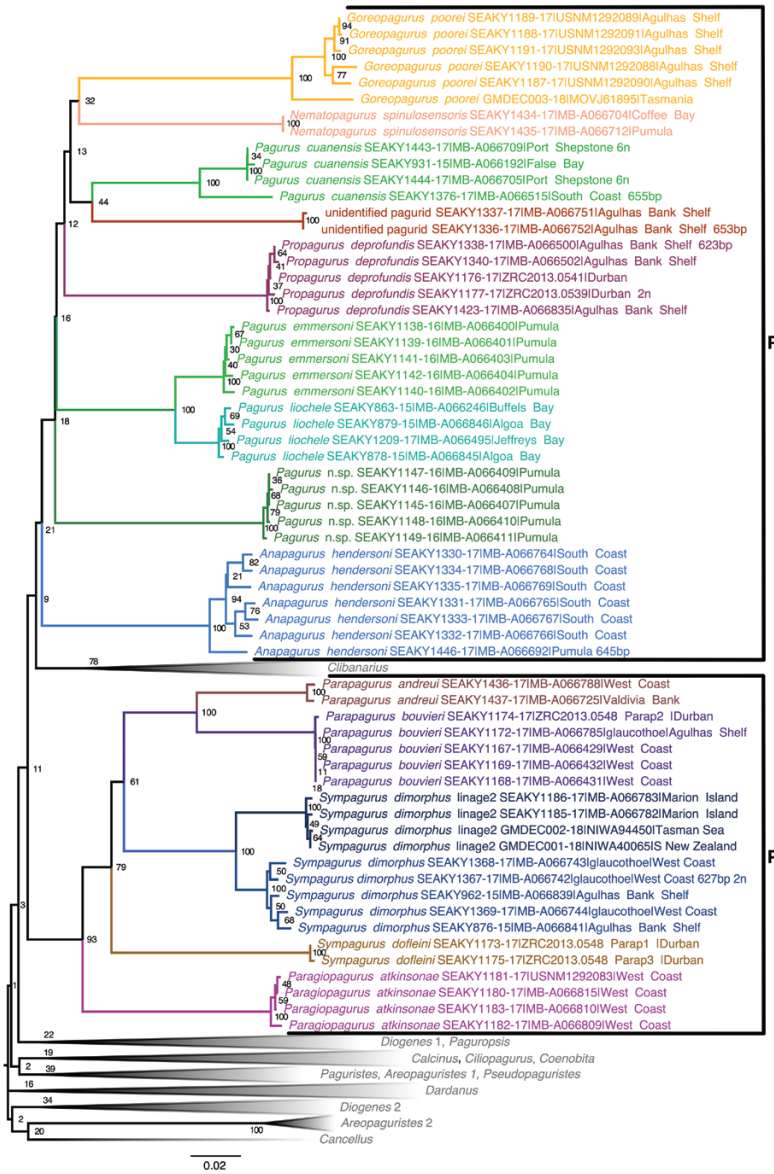

Figure 4. Neighbour-joining tree of distance relationships between South African hermit crabs (except where indicated), constructed using the cytochrome *c* oxidase I (COI) barcoding fragment and Kimura 2-parameter (K2P) distances. Detail of Paguridae and Parapaguridae. Bootstrap values based on 1000 replicates are indicated next to nodes, K2P divergences are shown by scale bar (%); *, stop codons; n, ambiguous bases. Sequence length 658 bp if not specified otherwise. This figure is available in colour at *Journal of Crustacean Biology* online.

its allies *Areopaguristes* and *Pseudopaguristes* form a taxonomically challenging group (Landschoff, 2018). In comparison to other clades, such as *Calcinus*, *Dardanus*, and *Diogenes*, branch lengths were much shorter in *Paguristes* (Fig. 1) and intraspecific distances were among the lowest (< 1% in *Paguristes*; Supplementary material Tables S2, S3), suggesting recent divergence and speciation. *Paguristes* was also monophyletic (100% bootstrap) and was sister-taxon to a well-supported clade formed by *Areopaguristes engyops* and *Pseudopaguristes* sp. (Figs. 1, 3). Future research might show that this clade may require a separate generic rank. The position of *Areopaguristes* cf. *jousseaumei* far from *A. engyops* further suggests that the taxonomic placement in this group requires detailed study. *Diogenes* forming two separate clades (Figs. 1–2), one of which was moderately supported to the exclusion of the other, is of some interest. It has been proposed that *Diogenes* is monophyletic (Forest, 1955), but no recent studies discuss phylogenetic relationships within this genus. To potentially resolve the evolutionary relationships of this genus, however, more detailed studies are needed than the single marker approach herein.

Species determination and taxonomic ambiguities

A main bias in assessments of divergence is incorrect or uncertain taxonomic classification (Matzen da Silva *et al.*, 2011; Collins & Cruickshank, 2012). The integrated taxonomic approach of this study led to the discovery of many unnamed, putative species in *Paguristes* and *Diogenes*, and to the finding of the distinctive Marion Island lineage 2 of *Sympagurus dimorphus*. Identifying four deep-water glaucothoe larvae further revealed the potential of barcoding to identify crustacean larval forms (Costa *et al.*, 2007).

Sympagurus dimorphus is widely-distributed in the southern hemisphere, but with considerable morphological variation (Lemaitre, 2004), and with different colour pattern (Landschoff & Lemaitre, 2017b). The two clusters of *S. dimorphus* herein, which were initially separated with some hesitation, were also assigned to different BINs in a BIN analysis (Ratnasingham & Herbert, 2013) performed on BOLD (not shown). Considering these two lineages as distinct reduced intraspecific divergence rates in the overall analysis. At slightly below 1%, the overall K2P intraspecific distance observed was consistent with results of other barcoding studies within Crustacea (Lefébure *et al.*, 2006; Costa *et al.*, 2007; Radulovici *et al.*, 2009; Matzen da Silva *et al.*, 2011; Puillandre *et al.*, 2011; Robe *et al.*, 2012). It seems justified to consider the two lineages as separate, with the taxonomy remaining to be resolved.

In contrast, a few species with taxonomic ambiguities, in which the separation into further defined species entities was not justifiable, explains notably increased maximum intraspecific divergences of up to 5.1%. In particular, three problematic species are represented in the dataset (*Goreopagurus poorei*, *Pagurus cuanensis*, and *Areopaguristes* cf. *jousseaumei*; Supplementary material Tables S2, S3) that potentially contain additional cryptic taxa. Even though the BIN analysis (not shown) allocated each of these individuals to a different BIN than its supposed conspecifics, available data are regarded as insufficient to make this determination at this stage. The range of *G. poorei* was recently extended from Tasmania to South Africa (Landschoff & Lemaitre, 2017a). The South African specimens were inclined to be regarded as a new species, but the only marginal morphological differences did not justify splitting these taxa. The maximum COI genetic distance of 5.0% between the newly available barcode sequence from the type locality (MOVJ61895, BOLD: GMDEC003-18) and the sequences of

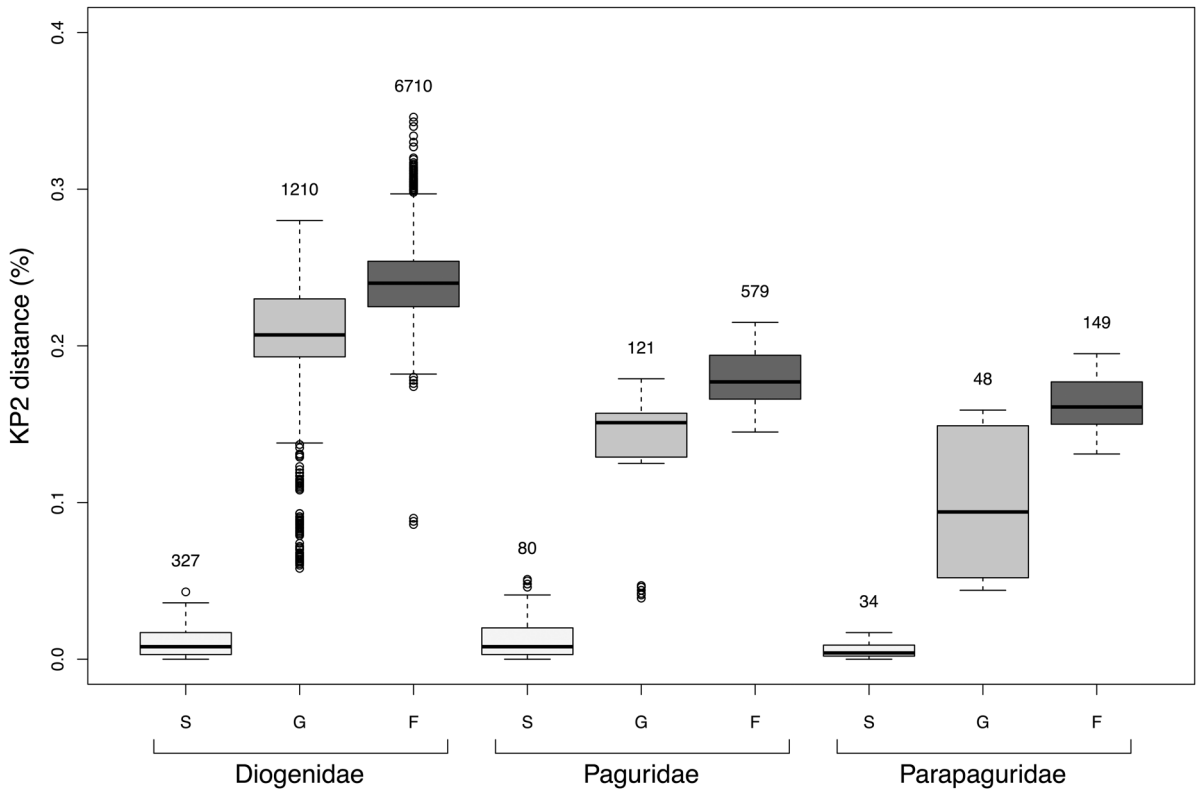

Figure 5. Boxplot distribution of intraspecific (S), intrageneric (G), and intrafamiliar (F) cytochrome *c* oxidase I (COI) Kimura 2-parameter (K2P) distances (%) for three families of hermit crabs. Plot giving median as central bar, upper and lower quartiles as central box, and extreme values of the data as dots. Numbers above individual boxplots correlate to the total number of pairwise comparisons of each group. Mean K2P distance (%) ± SE within taxa: Diogenidae, S = 0.0110 ± 0.0005, G = 0.1982 ± 0.0015, F = 0.2417 ± 0.0003; Paguridae, S = 0.0144 ± 0.0016, G = 0.1315 ± 0.0037, F = 0.1772 ± 0.0007; Parapaguridae, S = 0.0056 ± 0.0009, G = 0.0974 ± 0.0065, F = 0.1620 ± 0.0016.

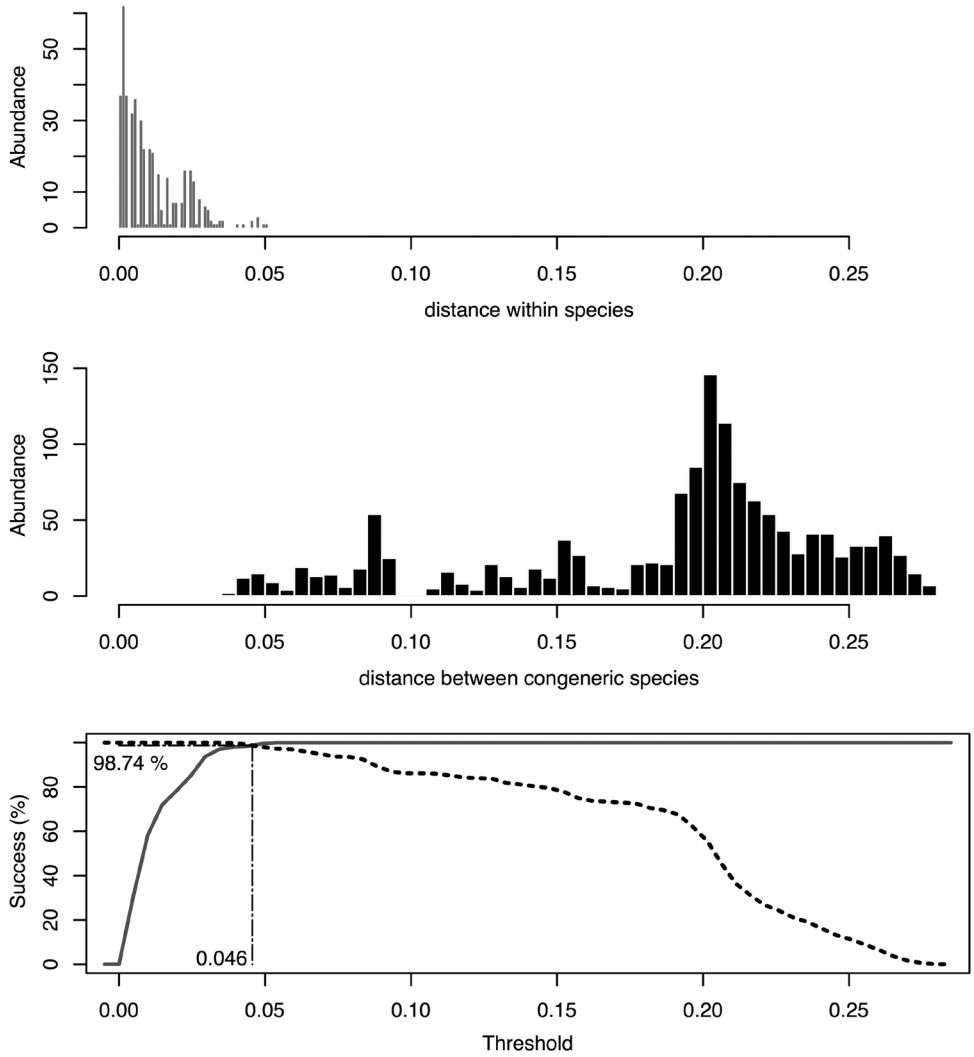

Figure 6. Overlap and threshold evaluation between conspecific (intra-species) and interspecific (but congeneric) Kimura 2-parameter (K2P) cytochrome *c* oxidase I (COI) distances (%) of all hermit crabs studied using the assumption-free statistical approach by Lefébure *et al.* (2006). Lines in bottom graph show percentage of samples that are below (distribution of distances within species, solid line) or above (distribution of distances between congeneric species, dotted line) the given range of thresholds. The percentage for each threshold value is thought to represent the chance of 'success' to discriminate samples from a distribution, where the best compromise, or 'threshold', is found where both success curves intersect.

specimens from South Africa was high compared to the overall average of < 1.0% in this dataset, and also compared to other values from the literature (see below). This result indicates that the two populations of *G. poorei* are most likely genetically distinct as well, but the high maximum distance of 2.0% among the South African specimens leaves some doubt (Fig. 3). Zhang *et al.* (2010) found that the ideal sample size for barcoding often greatly exceeds the usual numbers of 5–10 specimens. More sequences of *G. poorei*, especially from Tasmania, should be compared to contextualise the observed genetic variation.

A similar problem occurs in *P. cuanensis*, with maximum intraspecific divergences of up to 5.1%. Whether this is due to a cryptic lineage, however, needs to be validated with more data. Morphologically, no differences were detectable between these specimens of *P. cuanensis* that would fall outside levels of acceptable intraspecific morphological variation, as the species does show strong rates of intraspecific variation (McLaughlin & Forest, 1999; Landschoff, 2018). It therefore seems likely that *P. cuanensis* consists of a yet unrecognized species complex.

As a third problematic species, one sequence within *A.* cf. *jousseaumei* was 4.3% distant from one of the other three sequences, suggesting there could also be another cryptic lineage present (Figs. 1, 3). There are marked taxonomic difficulties in this and related species (Landschoff, 2018), and both morphological and molecular tools consistently fail to provide a reliable identification. Resolving the taxonomy of this species or species complex requires a combination of increased molecular and morphological effort.

It is noteworthy that, out of four species (*S. dimorphus, G. poorei, P. cuanensis*, and *A.* cf. *jousseaumei*) with taxonomic issues (Dechancé, 1963; McLaughlin & Forest, 1999; Lemaitre, 2004; Landschoff & Lemaitre, 2017a; Landschoff, 2018), only the case of *S. dimorphus* seems resolvable by molecular barcoding.

Sequence divergence in hermit crabs

The moderately low intraspecific divergences for this set of data find their maxima in those sequences that involve the problematic and unresolved cases. The overlaps analysis (Fig. 5) reveals that COI DNA sequences are highly efficient in discriminating between specific and generic divergences in hermit crabs, and this efficiency is even increased to zero overlaps (equal to 100% success) when these few sequences of the problematic species are removed. These findings are consistent with the overall pattern in Crustacea (Lefébure *et al.*, 2006), particularly Decapoda (Matzen

da Silva *et al.*, 2011). The overlaps between the genus and family levels, which are nearly complete for Diogenidae and about half for Paguridae and Parapaguridae (Fig. 5), show that COI divergences could indicate broad patterns, but lack the resolution to be an accurate model of evolution in hermit crabs at higher taxonomic levels.

Threshold evaluation and barcoding efficacy

The thorough sorting and identification of species resulted in a dataset with high taxonomic accuracy to be applied *a priori*. This high resolution of morphological taxonomy had a positive influence on threshold and BG values, because these values could be determined for well-defined taxonomic groups. Thresholds and success rates were most impacted by the *a priori* decision to treat the two lineages of *S. dimorphus* as distinct taxa. That the BG was present when assessed using each sequence, but absent for *Pagurus* when presenting the BG based on averaging intra- and interspecific distances of each genus (Supplementary material Table S3), shows that the gap was small and caused by the single problematic sequence in *Pagurus*. While the other problematic sequences (see above) decreased the barcoding gaps, they did not close them in those taxa.

An optimal threshold range of 3.5–3.7% to separate paguroid species in our dataset was closest to the originally-proposed barcoding threshold of 3.0% (Hebert *et al.*, 2003). The assumption-free statistical approach of Lefébure *et al.* (2006) had a high (98.74%) success rate for a 4.6% threshold (Fig. 6). Testing this threshold in the *spider* barcoding efficacy evaluations, however, the 4.6% value was less appropriate for this dataset (Table 2). The *spider* in-built localMinima function (Brown *et al.*, 2012), suggesting a threshold of 1.9% that led to five misidentifications, was closer to the optimal test result of only two misidentifications. It is not surprising that the taxonomically-problematic specimens discussed above were responsible for the two sequences that could not be matched to their *a priori* species identifications. It should be mentioned that removing the three problematic sequences would have resulted in an optimal threshold range of 2.7–3.7%, with a 100% success rate. The efficacy evaluations of nearly 99% success, however, were already extremely high for the entire dataset, highlighting how effective barcoding can be for use in an integrated taxonomic framework.

Prospects of DNA barcoding of southern Africa hermit crabs

The dataset analysed for South Africa marine habitats is one of the most comprehensive barcoding libraries of hermit crabs for any bioregion. Testing this set of data for barcoding efficacy revealed that, in an integrated taxonomic framework, the technique achieves barcoding identification rates of near 100% success. The prospects of an automated identification process through barcodes for paguroids nevertheless remain poor. Molecular barcoding alone is no substitute for morphology-based taxonomy. It has been speculated that the number of hermit crab species in South Africa may well exceed 100 (Landschoff, 2018). Barcoding greatly help inform morphological identifications, including the glaucothoe larval stages of two deep-water species and the identification of 12 putatively new species. The latter still require additional taxonomic study, as do the other three possible cryptic species lineages discovered. Despite considerable species coverage of 55 taxa, the dataset remains only about 50% complete, so that even with a near 100% identification success rate, only every second species of hermit crab from this region would be identifiable by matching its COI sequence to a taxon in this molecular library. Although this study has shown the benefits of barcoding for hermit crabs, the regional study design inevitably limits geographic and taxonomic coverage. It remains to be tested if the findings hold true when species coverage is extended to global species distributions.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of Crustacean Biology* online.

S1 Table. Specimen information spread sheet.

S2 Table. Kimura 2-parameter (KP2) distance values (%) from pairwise intraspecific comparisons of hermit crab species.

S3 Table. Kimura 2-parameter (KP2) distance values (%) from pairwise intraspecific comparisons of hermit crab genera.

ACKNOWLEDGEMENTS

Sampling of South African specimens was carried out within the SeaKeys Project, a collaborative effort overseen by the South African National Biodiversity Institute (SANBI), and during which numerous people contributed with the specimen collections, and curating. We particularly thank Charles Griffiths for guidance, funding, and for editing of the first draft of the manuscript. We are deeply indebted to Jose Christopher Mendoza for providing specimens from the Lee Kong Chian Natural History Museum, Joanne Taylor and Melanie Mackenzie for specimens from Australia, and Kareen Schnabel for specimens from New Zealand. We thank two anonymous reviewers for their comments on the manuscript.

REFERENCES

Alcock, A. 1905. *Catalogue of the Indian decapod Crustacea in the collection of the Indian Museum. Part II. Anomura. Fasciculus I. Pagurides.* Trustees of the Indian Museum, Calcutta [= Kolkata].

Barnard, K.H. 1947. Descriptions of new species of South African decapod Crustacea, with notes on the synonymy and new records. *Annals of the South African Museum*, **12**: 361–392.

Barnard, K.H. 1950. Descriptive catalogue of South African decapod Crustacea (crabs and shrimps). *Annals of the South African Museum*, **38**: 1–838.

Barnes, D.K.A. 2003. Local, regional and global patterns of resource use in ecology: hermit crabs and gastropod shells as an example. *Marine Ecology Progress Series*, **246**: 211–223.

Bell, T. 1844–1852. *A history of the British stalk-eyed Crustacea.* John van Voorst, London.

Bertness, M.D. 1981. The influence of shell-type on hermit crab growth rate and clutch size (Decapoda, Anomura). *Crustaceana*, **40**: 197–205.

Bilgin, R., Utkan, M.A., Kalkan, E., Karhan, S.Ü. & Bekbolet, M. 2014. DNA barcoding of twelve shrimp species (Crustacea: Decapoda) from Turkish seas reveals cryptic diversity. *Mediterranean Marine Science*, **16**: 36–45.

Blackstone, N.W. 1985. The effects of shell size and shape on growth and form in the hermit crab *Pagurus longicarpus*. *Biological Bulletin*, **168**: 75–90.

Boero, F. 2010. The study of species in the era of biodiversity: a tale of stupidity. *Diversity*, **2**: 115–126.

Bouvier, E.-L. 1892. Étude sur les Paguriens recueillis par M. le Dr. Jousseaume sur les côtes de la Mer Rouge. *Bulletin de la Société Philomathique de Paris*, **8**: 50–55.

Bracken-Grissom, H.D., Cannon, M.E., Cabezas, P., Feldmann, R.M., Schweitzer, C.E., Ahyong, S.T., Felder, D.L., Lemaitre, R. & Crandall, K.A. 2013. A comprehensive and integrative reconstruction of evolutionary history for Anomura (Crustacea: Decapoda). *BMC Evolutionary Biology*, **13**: 1–28.

Brown, S.D.J., Collins, R.A., Boyer, S., Lefort, M.-C., Malumbres-Olarte, J., Vink, C.J. & Cruickshank, R.H. 2012. *Spider*: an *R* package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, **12**: 562–565.

Buhay, J.E. 2009. "COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. *Journal of Crustacean Biology*, **29**: 96–110.

de Carvalho, M.R., Bockmann, F.A., Amorim, D.S., Brandão, C.R.F., de Vivo, M., de Figueiredo, J.L., Britski, H.A., de Pinna, M.C.C.,

Menezes, N.A., Marques, F.P.L., Papavero, N., Cancello, E.M., Crisci, J.V., McEachran, J.D., Schelly, R.C., Lundberg, J.G., Gill, A.C., Britz, R., Wheeler, Q.D., Stiassny, M.L.J., Parenti, L.R., Page, L.M., Wheeler, W.C., Faivovich, J., Vari, R.P., Grande, L., Humphries, C.J., DeSalle, R., Ebach, M.C. & Nelson, G.J. 2007. Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. *Evolutionary Biology*, **34**: 140–143.

Chu, K.H., Tsang, L.M., Ma, K.Y., Chan, T.Y. & Ng, P.K.L. 2009. Decapod phylogeny: what can protein-coding genes tell us. In: *Decapod crustacean phylogenetics* (J.W. Martin, K.A. Crandall & D. Felder, eds), pp. 89–99. CRC Press, Boca Raton, FL, USA.

Collins, R.A. & Cruickshank, R.H. 2012. The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*, **13**: 969–975.

Costa, F.O., DeWaard, J.R., Boutillier, J., Ratnasingham, S., Dooh, R.T., Hajibabaei, M. & Hebert, P.D.N. 2007. Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Sciences*, **64**: 272–295.

Costello, M.J., Coll, M., Danovaro, R., Halpin, P., Ojaveer, H. & Miloslavich, P. 2010. A census of marine biodiversity knowledge, resources, and future challenges. *PLoS One*, **5**: e12110 [doi.org/10.1371/journal.pone.0012110].

Cristescu, M.E. 2014. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global diversity. *Trends in Ecology & Evolution*, **29**: 566–571.

Dana, J.D. 1851. Conspectus crustaceorum quae in orbis terrarum circumnavigatione, Carolo Wilkes e classe reipublicae foederatae duce, lexit et descripsit. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **5**: 267–272.

Dana, J.D. 1852. *Crustacea, part I. United States Exploring Expedition, during the years 1838, 1839, 1840, 1841, 1842, under the command of Charles Wilkes, U.S.N., vol. 13*. C. Sherman, Philadelphia, PA, USA.

Dechancé, M. 1963. Sur des *Paguristes* littoraux de l'ouest de l'océan Indien: *P. jousseaumei* Bouvier, *P. perspicax* Nobili et *P. abbreviatus* sp. nov. (Crustacea Decapoda Paguridea). *Bulletin du Muséum national d'Histoire naturelle*, série 2, **35**: 291–301.

DeSalle, R. 2006. Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. *Conservation Biology*, **20**: 1545–1547.

Ebach, M.C. 2011. Taxonomy and the DNA barcoding enterprise. *Zootaxa*, **68**: 67–68.

Ebach, M.C. & Holdrege, C. 2005. DNA barcoding is no substitute for taxonomy. *Nature*, **434**: 697–697.

Fabricius, J.C. 1775. *Systema entomologiae, sistens insectorum classes, ordines, genera, species, adiectis synonymis, locis, descriptionibus, observationibus*. Officina Libraria Kortii, Flensburg and Leipzig.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**: 783–791.

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**: 294–299.

Forest, J. 1954. Les *Paguristes* des côtes occidentales et méridionales d'Afrique. *Annals of the South African Museum*, **41**: 159–213.

Forest, J. 1955. Crustacés Décapodes, Pagurides. Expédition océanographique Belge dans les eaux côtiéres africaines de l'Atlantique Sud (1948–1949). *Résultats scientifiques*, **3**: 23–147.

Forest, J. 1995. Crustacea Decapoda Anomura: Révision du genre *Trizopagurus* Forest, 1952 (Diogenidae), avec l'établissement de deux genres nouveaux. In: *Résultats des Campagnes MUSORSTOM*, Volume **13**. (A. Crosnier, ed.), pp. 9–149. *Mémoires du Muséum national d'Histoire naturelle*, **Vol. 163**.

Funk, D.J. & Omland, K.E. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, **34**: 397–423.

Galtier, N., Nabholz, B., Glémin, S. & Hurst, G.D.D. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**: 4541–4550.

Goldstein, P.Z. & DeSalle, R. 2011. Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. *Bioessays*, **33**: 135–147.

Gouws, G., Peer, N. & Perissinotto, R. 2015. MtDNA lineage diversity of a potamonautid freshwater crab in KwaZulu-Natal, South Africa. *Koedoe*, **57**: 1–12.

De Grave, S., Pentcheff, N.D., Ahyong, S.T., Chan, T.-Y., Crandall, K.A., Dworschak, P.C., Felder, D.L., Feldmann, R.M., Fransen, C.H.J.M., Goulding, L.Y.D., Lemaitre, R., Low, M.E.Y., Martin, J.W., Ng, P.K.L., Schweitzer, C.E., Tan, S.H., Tshudy, D. & Wetzer, R. 2009. A classification of living and fossil genera of decapod crustacea. *Raffles Bulletin of Zoology*, Supplement **21**: 1–109.

Hajibabaei, M., Singer, G.A.C., Hebert, P.D.N. & Hickey, D.A. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics*, **23**: 167–172.

Hazlett, B.A. 1981. The behavioral ecology of hermit crabs. *Annual Review of Ecology and Systematics*, **12**: 1–22.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & DeWaard, J.R. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society*, **270**: 313–321.

Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S. & Francis, C.M. 2004. Identification of birds through DNA barcodes. *PLoS Biology*, **2**: e312 [doi.org/10.1371/journal.pbio.0020312].

Henderson, J.R. 1888. Report on the Anomura collected by H.M.S. *Challenger* during the years 1873–76. In: *Scientific Results of the Exploratory Voyage of H.M.S. Challenger, Zoology, 27*. Morrison and Gibb, Edinburgh.

Hickerson, M.J., Meyer, C.P., Moritz, C. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*, **55**: 729–739.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**: 111–120.

Komai, T. 1998. The taxonomic position of *Pagurus gracilipes* (Stimpson) and *Pagurus nipponensis* (Yokoya), and description of a new species of *Pagurus* from Japan (Decapoda: Anomura: Paguridae). *Zoosystema*, **20**: 265–288.

Krauss, F. 1843. *Die südafrikanischen Crustaceen. Eine Zusammenstellung aller bekannten Malacostraca, Bermerkungen über deren Lebensweise und geographische Verbreitung, nebst Beschreibung und Abbildung mehrer neuer Arten*. E. Schweizerbartsche Verlagsbuchhandlung, Stuttgart, Germany.

Kumar, S., Stecher, G. & Tamura, K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, **33**: 1870–1874.

Kvist, S. 2013. Barcoding in the dark?: a critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. *Molecular Phylogenetics and Evolution*, **69**: 39–45.

Landschoff, J. 2018. Contributions to the taxonomy of South African hermit crabs (Crustacea: Decapoda: Paguroidea) – integrating microCT scanning and barcoding. Ph.D. thesis, University of Cape Town, Cape Town, South Africa.

Landschoff, J. & Lemaitre, R. 2017a. Crossing the Indian Ocean: a range extension for *Goreopagurus poorei* Lemaitre & McLaughlin, 2003 (Crustacea: Decapoda: Paguridae). *Zootaxa*, **4306**: 271–278.

Landschoff, J. & Lemaitre, R. 2017b. Differentiation of three common deep-water hermit crabs (Crustacea, Decapoda, Anomura, Parapaguridae) from the South African demersal abundance surveys, including the description of a new species of *Paragiopagurus* Lemaitre, 1996. *ZooKeys*, **676**: 21–45.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. & Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**: 2947–2948.

Latreille, P.A. 1829. Crustacés, Arachnides et partie des Insectes. In: *Le règne animal distribué d'après son organisation, pour servir de base à l'histoire naturelle des animaux et d'introduction à l'anatomie comparée. Nouvelle édition, revue et augmentée* (G. Cuvier, ed.), pp. 1–39. Déterville, Paris.

Lefébure, T., Douady, C.J., Gouy, M., Gibert, J. 2006. Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution*, **40**: 435–447.

Lemaitre, R. 1996. Hermit crabs of the family Parapaguridae (Crustacea: Decapoda: Anomura) from Australia: species of *Strobopagurus* Lemaitre, 1989, *Sympagurus* Smith, 1883, and two new genera. *Records of the Australian Museum*, **48**: 163–221.

Lemaitre, R. 2004. A worldwide review of hermit crab species of the genus *Sympagurus* Smith, 1883 (Crustacea: Decapoda: Parapaguridae). In: *Tropical Deep-Sea Benthos* (B. Marshall & B. Richer de Forges, eds.), pp. 85–149. *Mémoires du Muséum national d'Histoire naturelle*, **Vol. 23**.

Lemaitre, R. & McLaughlin, P.A. 2003. New species of *Goreopagurus* (Decapoda: Anomura: Paguridae) from Tasmania and reevaluation of

sexual tubes in hermit crab systematics. *Memoirs of Museum Victoria*, **60**: 221–227.
Malay, M.C.M.D. & Paulay, G. 2010. Peripatric speciation drives diversification and distributional pattern of reef hermit crabs (Decapoda: Diogenidae: Calcinus). *Evolution*, **64**: 634–662.
Matzen da Silva, J., Creer, S., dos Santos, A., Costa, A.C., Cunha, M.R., Costa, F.O. & Carvalho, G.R. 2011. Systematic and evolutionary insights derived from mtDNA COI barcode diversity in the Decapoda (Crustacea: Malacostraca). *PLoS ONE*, **6**: e19449 [doi.org/10.1371/journal.pone.0019449].
McLaughlin, P.A. 1974. The hermit crabs (Crustacea, Decapoda, Paguridea) of northwestern North America. *Zoologische Verhandelingen*, **130**: 1–396.
Mclaughlin, P.A. 1983. Hermit crabs: are they really polyphyletic? *Journal of Crustacean Biology*, **3**: 608–621.
McLaughlin, P.A. 1998. Hermit crabs of the genus *Nematopagurus* (Crustacea: Decapoda: Paguridae) from southeastern South Africa and Madagascar: New records and new species. *Zoosystema*, **20**: 335–358.
McLaughlin, P.A. 2002. *Pseudopaguristes*, a new and aberrant genus of hermit crabs (Anomura: Paguroidea: Diogenidae). *Micronesica*, **34**: 185–199.
McLaughlin, P.A. 2015. Crustaceans in mobile homes. In: *The Natural history of the Crustacea*. Vol. 2, *Lifestyles and feeding biology* (M. Thiel & L. Watling, eds.), pp. 145–162. Oxford University Press, New York.
McLaughlin, P.A. & Forest, J. 1999. Hermit crabs of the genus *Pagurus* Fabricius (Crustacea, Decapoda, Paguridae) from south-eastern South Africa. *Annals of the South African Museum*, **105**: 297–344.
McLaughlin, P.A. & Türkay, M. 2018. Paguroidea Latreille, 1802. In: *World Paguroidea & Lomisoidea database. World Register of Marine Species* (R. Lemaitre & P. McLaughlin, 2018) [http://www.marinespecies.org/aphia.php?p=taxdetails&id=106687].
McLaughlin, P.A., Lemaitre, R. & Sorhannus, U. 2007. Hermit crab phylogeny: a reappraisal and its "fall-out." *Journal of Crustacean Biology*, **27**: 97–115.
Meier, R., Shiyang, K., Vaidya, G. & Ng, P.K.L. 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**: 715–728.
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**: 809–813.
Meyer, C.P. 2003. Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. *Biological Journal of the Linnean Society*, **79**: 401–459.
Meyer, C.P. & Paulay, G. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, **3**: e422 [doi.org/10.1371/journal.pbio.0030422].
Meyer, R., Weis, A. & Melzer, R.R. 2013. Decapoda of southern Chile: DNA barcoding and integrative taxonomy with focus on the genera *Acanthocyclus* and *Eurypodius*. *Systematics and Biodiversity*, **11**: 389–404.
Milne Edwards, H. 1836. Observations zoologiques sur les *Pagures* et description d'un nouveau genre de la tribu des Paguriens. *Annales des Sciences Naturelles*, série 2, **6**: 257–288, pls, 13, 14.
Milne Edwards, H. 1837. *Histoire naturelle des Crustacés, comprenant l'anatomie, la physiologie et la classification de ces animaux*. **Vols. 1, 2**. Librairie Encyclopédique de Roret. Paris.
Muñoz-Pajares, A.J. 2013. *SIDIER*: substitution and indel distances to infer evolutionary relationships. *Methods in Ecology and Evolution*, **4**: 1195–1200.
Negri, M., Lemaitre, R. & Mantelatto, F.L. 2014. Molecular and morphological resurrection of *Clibanarius symmetricus* (Randall, 1840), a cryptic species hiding under the name for the "thinstripe" hermit crab *C. vittatus* (Bosc, 1802) (Decapoda: Anomura: Diogenidae). *Journal of Crustacean Biology*, **34**: 848–861.
Padial, J.M. & De La Riva, I. 2007. Integrative taxonomists should use and produce DNA barcodes. *Zootaxa*, **68**: 67–68.
Padial, J.M., Miralles, A., De la Riva, I. & Vences, M. 2010. The integrative future of taxonomy. *Frontiers in Zoology*, **7**: 1–14.
Paul´son, O. 1875. *Studies on Crustacea of the Red Sea with notes regarding other seas, Part I, Podophthalmata and Edriophthalmata* (*Cumacea*). S.V. Kul´zhenko, Kiev [in Russian].
Prendini, L. 2005. Comment on "Identifying spiders through DNA barcodes." *Canadian Journal of Zoology*, **83**: 481–491.
Puillandre, N., MacPherson, E., Lambourdière, J., Cruaud, C., Boisselier-Dubayle, M.C. & Samadi, S. 2011. Barcoding type specimens helps to identify synonyms and an unnamed new species in *Eumunida* Smith, 1883 (Decapoda: Eumunididae). *Invertebrate Systematics*, **25**: 322–333.
R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria [http://www.r-project.org].
Radulovici, A.E., Sainte-Marie, B. & Dufresne, F. 2009. DNA barcoding of marine crustaceans from the Estuary and Gulf of St Lawrence: a regional-scale approach. *Molecular Ecology Resources*, **9**: 181–187.
Rahayu, D.L. & McLaughlin, P.A. 2010. *Areopaguristes*, a generic replacement name for *Stratiotes* Thomson, 1899 (Crustacea: Decapoda: Paguroidea: Diogenidae). *Zootaxa*, **2509**: 67–68.
Rahayu, D.L., Shih, H.T. & Ng, P.K.L. 2016. A new species of land hermit crab in the genus *Coenobita* Latreille, 1829 from Singapore, Malaysia and Indonesia, previously confused with *C. cavipes* Stimpson, 1858 (Crustacea: Decapoda: Anomura: Coenobitidae). *Raffles Bulletin of Zoology*, **34**: 470–488.
Ratnasignham, S. & Hebert, P.D.N. 2007. BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**: 355–364.
Ratnasingham, S. & Hebert, P.D.N. 2013. A DNA-based registry for all animal species: the barcode index number (BIN) system. *PLoS ONE*, **8**: e66213 [doi.org/10.1371/journal.pone.0066213].
Raupach, M.J., Barco, A., Steinke, D., Beermann, J., Laakmann, S., Mohrbeck, I., Neumann, H., Kihara, T.C., Pointner, K., Radulovici, A., Segelken-Voigt, A., Wesse, C. & Knebelsberger, T. 2015. The application of DNA barcodes for the identification of marine crustaceans from the North Sea and adjacent regions. *PLoS ONE*, **10**: e0139421 [doi: 10.1371/journal.pone.0139421].
Robe, L.J., Machado, S. & Bartholomei-Santos, M.L. 2012. The DNA barcoding and the caveats with respect to its application to some species of *Palaemonidae* (Crustacea, Decapoda). *Zoological Science*, **29**: 714–724.
Roe, A.D. & Sperling, F.A.H. 2007. Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution*, **44**: 325–345.
RStudio Team. 2016. *RStudio: Integrated development environment for R*. [http://www.rstudio.com/].
Rubinoff, D. 2006. Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biology*, **20**: 1026–1033.
Saitou, N. & Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**: 406–425.
Shackleton, M. & Rees, G.N. 2015. DNA barcoding Australian macroinvertebrates for monitoring programs: benefits and current short comings. *Marine and Freshwater Research*, **67**: 380–390.
Smith, S.I. 1883. Preliminary report on the Brachyura and Anomura dredged in deep water off the south coast of New England by the United States Fish Commission in 1880, 1881, and 1882. *Proceedings of the United States National Museum*, **6**: 1–57.
Sperling, F. 2003. DNA barcoding: Deus ex Machina (opinion page). *Newsletter of the Biological Survey of Canada* (*Terrestrial Arthropods*), **22**: 50–53.
Stebbing, T.R.R. 1910. General catalog of South African Crustacea (Part V. of S. A. Crustacea, for the Marine Investigations in South Africa). *Annals of the South African Museum*, **6**: 281–593.
Stevens, M.I., Porco, D., D'Haese, C.A. & Deharveng, L. 2011. Comment on "Taxonomy and the DNA barcoding enterprise" by Ebach (2011). *Zootaxa*, **88**: 85–88.
Stimpson, W. 1858. Prodromus descriptionis animalium evertebratorum, quae in expeditione ad oceanum Pacificum septentrionalem, a Republica Federate missa, Cadwaldaro Ringgold et Johanne Rodgers ducibus, observavit et descripsit. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **7**: 225–252.
Studer, T. 1883. Verzeichnisse der während der Reise S.M.S. Gazelle an der Westküste von Afrika, Ascension und dem Cap der guten Hoffnung gesammelten Crustaceen. *Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin*, Physikalische Abhandlungen, **2**1882]: 1–32, pls. 1, 2.
Sunnucks, P. & Hales, D.F. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I–II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution*, **13**: 510–524.
Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. & Vogler, A.P. 2003. A plea for DNA taxonomy. *Trends in Ecology & Evolution*, **18**: 70–74.

Tsang, L.M., Chan, T.Y., Ahyong, S.T. & Chu, K.H. 2011. Hermit to king, or hermit to all: multiple transitions to crab-like forms from hermit crab ancestors. *Systematic Biology*, **60**: 616–629.

Tsang, L.M., Ma, K.Y., Ahyong, S.T. & Chan, T.-Y. & Chu, K.H. 2008. Phylogeny of Decapoda using two nuclear protein-coding genes: origin and evolution of the Reptantia. *Molecular Phylogenetics and Evolution*, **48**: 359–368.

Tudge, C.C. 1997. Phylogeny of the Anomura (Decapoda, Crustacea): spermatozoa and spermatophore morphological evidence. *Contributions to Zoology*, **67**: 125–141.

Ward, R.D. 2009. DNA barcode divergence among species and genera of birds and fishes. *Molecular Ecology Resources*, **9**: 1077–1085.

Werle, E., Schneider, C., Renner, M., Volker, M. & Fiehn, W. 1994. Convenient single-step, one tube puri cation of PCR products for direct sequencing. *Nucleic Acids Research*, **22**: 4354–4355.

Will, K.W., Mishler, B.D. & Wheeler, Q.D. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**: 844–851.

Wooster, D.S. 1984. The genus *Calcinus* (Paguridae, Diogenidae) from the Mariana Islands including three new species. *Micronesica*, **18**: 121–162.

Zhang, A.B., He, L.J., Crozier, R.H., Muster, C. & Zhu, C.D. 2010. Estimating sample sizes for DNA barcoding. *Molecular Phylogenetics and Evolution*, **54**: 1035–1039.