For the virtual issue, "Red Sea and Western Indian Ocean Biogeography"

Original Article

LRH: M. A. Priest et al.

RRH: Barrier to larval dispersal in a range-limited grouper

A bridge too far: dispersal barriers and cryptic speciation in an Arabian Peninsula grouper (Cephalopholis hemistiktos)

Mark A. Priest¹,²*, Joseph D. DiBattista²,³, Jennifer L. McIlwain³, Brett M. Taylor⁴, Nigel E. Hussey⁵ and Michael L. Berumen²

¹Marine Spatial Ecology Lab, School of Biological Sciences, University of Queensland, St Lucia, QLD 4072, Australia, ²Red Sea Research Center, Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia, ³Department of Environment and Agriculture, Curtin University, GPO Box U1987, Perth, WA, Australia, ⁴James Cook University, College of Marine and Environmental Science, Townsville, QLD 4811, Australia, ⁵Great Lakes Institute for Environmental Research, University of Windsor, 401 Sunset Avenue, Windsor, ON N9B 3P4, Canada

*Correspondence: Mark A. Priest, Marine Spatial Ecology Lab, University of Queensland, St. Lucia, QLD 4072, Australia

Email: m.priest@uq.edu.au
ABSTRACT

Aim We use genetic and age-based analyses to assess the evidence for a biogeographical barrier to larval dispersal in the yellowfin hind, *Cephalopholis hemistiktos*, a commercially important species found across the Arabian Peninsula.

Location Red Sea, Gulf of Aden, Gulf of Oman and Arabian Gulf.

Methods Mitochondrial DNA (COI) and nuclear DNA (S7) sequences were obtained for *C. hemistiktos* sampled throughout its distributional range. Phylogeographical and population-level analyses were used to assess patterns of genetic structure and to identify barriers to dispersal. Age-based demographic analyses using otoliths were also used to determine differences in growth and longevity between regions.

Results Our analyses revealed significant genetic structure congruent with growth parameter differences observed across sampling sites, suggesting cryptic speciation between populations in the Red Sea and Gulf of Aden versus the Gulf of Oman and Arabian Gulf.

Main Conclusion Our results indicate historical disruption to gene flow and a contemporary dispersal barrier in the Arabian Sea, which *C. hemistiktos* larvae are unable to effectively traverse. This provides yet another example of a (cryptic) species with high dispersive
potential whose range is delimited by a lack of suitable habitat between locations or an inability to recruit at the range edge.

Keywords

Serranidae, coral reef fish, population genetics, demography, mitochondrial DNA, nuclear DNA, phylogeography

INTRODUCTION

For many coral reef fishes, the pelagic larval phase presents the only opportunity for dispersal amongst fragmented and spatially patchy adult habitats (Kritzer & Sale, 2004). On a demographic level, the process of dispersal determines the persistence and structure of populations (Hixon et al., 2002; Burgess et al., 2014). This is especially important when considering harvested species, which may rely on external replenishment if localized spawner biomass is depleted due to overfishing. Thus, understanding patterns of connectivity can help inform resource managers to preserve fishery resources and make conservation efforts more effective. For example, the success of spatial management techniques such as marine reserve networks relies on successful self-seeding and larval export (demographic connectivity) to adjacent fished areas (Botsford et al., 2003; Gaines et al., 2010). Conversely, the absence of connectivity between locations will define management units and result in isolated populations that over time generate distinct genetic signatures (Moritz, 1994), and ultimately can lead to speciation.

Genetic techniques are now routinely used to measure connectivity on multiple temporal and spatial scales. Recent technological and methodological advances allow larvae or recruits to be assigned back to their parents, or population of origin (e.g. Jones et al., 2005;
Harrison et al., 2012; Almany et al., 2013). This process, however, is challenging as it involves intensive sampling of both adults and recruits, and large numbers of genetic markers. Traditionally, population genetics has been used to infer levels of connectivity and estimate the number of migrants exchanged between locations, but these estimates require a number of simplifying assumptions (Whitlock & McCauley, 1999) and may reflect historical rather than contemporary dispersal (Hedgecock et al., 2007). Thus, the absence of genetic differentiation among sites does not necessarily confirm effective connectivity on demographic scales, but the presence of genetic structure does suggest some level of prolonged isolation (Hellberg, 2007; Hellberg, 2009).

Limits to dispersal, and thus gene flow, in the marine environment are not always reflective of larval dispersal ability (Weersing & Toonen, 2009; but see Riginos & Victor, 2001) and may be driven by numerous factors such as oceanographic patterns (White et al., 2010; Simpson et al., 2014), larval behaviour (Leis, 2002), ecological requirements (Rocha et al., 2002) and biogeographic barriers (Rocha et al., 2007; Briggs & Bowen, 2012). One area that may provide conditions limiting for dispersal are the seas surrounding the Arabian Peninsula (Fig. 1). This area is at the intersect of several biogeographic provinces (Schils & Coppejans, 2003), and contains a diverse range of habitats experiencing a wide range of environmental conditions. During Pleistocene glacial cycles, the Red Sea experienced repeated periods of sea level fluctuation, causing major deviations in temperature and salinity, as well as putative faunal extinctions (Siddall et al., 2003; DiBattista et al., in reviewA). Such events are thought to have driven the high levels of diversity and endemism recorded in the Red Sea (14% in fishes; Randall, 1994; also see DiBattista et al., in reviewB). Whilst the shallow and narrow opening to the Gulf of Aden has historically restricted oceanic water and propagule exchange, an ecological barrier of turbid water in the southern Red Sea has been described as a contemporary constraint to dispersal (Ormond & Edwards, 1987; Roberts et al.,...
Outside the Gulf of Aden, the southern coast of Oman bordering the Arabian Sea is subjected to seasonal cold-water upwelling events generated by southwest monsoons, providing an additional isolating force (Savidge et al., 1990). This results in large annual fluctuations in surface water temperatures (16 – 28 °C) and impedes the development of coral reefs, creating large tracts of coastline devoid of suitable settlement habitat for reef fish larvae. Extensive coral reefs return in northern Oman and extend into the Gulf of Oman, where the effects of upwelling events are less frequent (Coles & Wilson, 2001). The Arabian Gulf (also known as the Persian Gulf), whilst containing somewhat similar benthic communities to the Gulf of Oman (Feary et al., 2010), has vastly different environmental parameters. The Arabian Gulf is shallow and has limited water exchange with the Gulf of Oman. Consequently, sea surface temperatures vary dramatically (12 – 36 °C) with consistently high seasonal maximums along with elevated salinity (Sheppard et al., 1992).

To investigate how these diverse bio-physical environments promote or impede larval dispersal, we investigated the connectivity and demography of the yellowfin hind, *Cephalopholis hemistiktos* (Serranindae; Rüppell 1830), a commercially important fishery species (Hashim, 1993; Gladstone, 2003) restricted to the Arabian Peninsula region. *Cephalopholis hemistiktos* possesses a disjunctive distribution and is absent from the western Arabian Sea (southern Oman; Fig. 1) (Craig et al., 2011). Additionally, previous reports have suggested that these two populations differ morphologically; in pectoral fin size and ray count, oblique scale rows and asymptotic size (Randall & Ben-Tuvia, 1983; see Appendix S1 in Supporting Information); these differences suggests isolation on evolutionary timescales. Given the spatial distribution and physical differences between locations, we conducted both genetic and age-based demographic assessments of *C. hemistiktos* by sampling throughout the entire species range. We used one mitochondrial and one nuclear genetic marker to evaluate
gene flow, dispersal barriers and connectivity among populations, as well as otolith-based age estimates to explore divergence in life-history characteristics.

MATERIALS AND METHODS

Sampling

*Cephalopholis hemistiktos* were collected from ten sites encompassing the majority of the range of this species (Fig. 1 and Table 1). Individuals were sampled between 2005 and 2014 using hand spears whilst snorkeling or SCUBA diving (Gulf of Aqaba, Red Sea, Gulf of Aden and Arabian Sea), or from local fish markets (Gulf of Oman and Arabian Gulf). Efforts were made to sample individuals representative of the whole size range available at each site. Measurements of total length (TL) were taken to the nearest millimeter. Sagittal otoliths were extracted, cleaned in ethanol and stored dry until sectioning. Tissue samples for genetic analysis were stored in 70% ethanol or in a saturated salt-DMSO buffer (Seutin *et al.*, 1991) and stored at room temperature. As samples were collected over many years for multiple research projects, we did not obtain both genetic and otolith samples from every site. Thus, some sites have only genetic or otolith data (see Fig. 1 and Table 1 for details).

Mitochondrial DNA analysis

Genomic DNA was extracted using the ‘HotSHOT’ protocol (Meeker *et al.*, 2007) and stored at -20°C. Samples were sequenced for a 629 base pair (bp) region of the mtDNA cytochrome-c oxidase subunit-I (COI) gene using the primers FishF2 and FishR2 (Ward *et al.*, 2005). Polymerase chain reactions (PCRs) were carried out in 12 μl reaction volumes using Qiagen
Multiplex PCR kits (Qiagen, Valencia, CA, USA) according to the manufacturer’s instructions. Each reaction contained 1 μl gDNA and 0.5 μl of each forward and reverse primers (10 μM). PCR cycling conditions consisted of an initial denaturation step at 95°C for 15 min, followed by 35 cycles of 94°C denaturing for 30 sec, 50°C annealing for 60 sec and 72°C extension for 60 sec, with a final 72°C extension for 10 min. PCR products were purified using exonuclease I and shrimp alkaline phosphatase (ExoSAP; USB, Cleveland, OH, USA), sequenced in the forward direction with fluorescently labeled dye terminators (BigDye version 3.1, Applied Biosystems, Foster City, CA, USA) and analysed using an ABI 3130xl Genetic Analyzer (Applied Biosystems). All haplotypes were deposited in GenBank (Accession numbers: XXX-XXX). Sequences were aligned, trimmed and edited using Geneious Pro vers. 7.0.6 (Drummond et al., 2009). The GTR + I model for nucleotide substitution was selected using JMODELTEST 2.0 vers. 0.1.1 (Guindon & Gascuel, 2003; Darriba et al., 2012) based on the Akaike Information Criterion test.

ARLEQUIN vers. 3.5 (Excoffier & Lischer, 2010) was used to calculate haplotype (h) and nucleotide diversity (π), and to test for range-wide patterns of population structure for *C. hemistiktos*. Global $\Phi_{ST}$ was estimated using analysis of molecular variance (AMOVA, Excoffier et al., 1992), with deviations from null distributions tested using nonparametric permutations ($N = 9,999$). AMOVA was also used to test for significant genetic partitioning among regions. Additionally, pairwise $\Phi_{ST}$ statistics were generated to identify sites associated with genetic subdivision within regions.

To assess deviation from neutrality, Fu’s $F_S$ (Fu, 1997) was calculated for all sites individually, all sites grouped and also sites grouped into the two regions (Red Sea and Gulf of Aden [RS-GA]; Gulf of Oman and Arabian Gulf [GO-AG]) using ARLEQUIN with significance assessed with 9,999 permutations. Historical demography was explored using mismatch analyses for all sites together and for the two separate regions. Populations
experiencing recent or rapid expansion exhibit unimodal mismatch distributions and non-
significant raggedness indices (Harpending, 1994). The parameter $\tau$ was fitted to the two
regions and population ages were estimated using the equation $\tau = 2\mu t$ where $t =$ population
age in generations and $\mu =$ mutation rate per generation for the sequence (Rogers &
Harpending, 1992). We used a mutation rate previously recorded for coral reef fish of 1.55 %
per million years (Myr) (Lessios, 2008). Generation time was estimated from the equation $T =
(\alpha + \omega)/2$ (Pianka, 1978), where $\alpha =$ age at first reproduction and $\omega =$ the age at last
reproduction. We therefore obtained generation times of 13 years based on our existing age-
based demographic and reproductive data. Given the approximations involved, we interpret
these coalescence estimates with caution.

Additional estimates of historical demographics were reconstructed using Bayesian
skyline plots (BSP) in BEAST vers. 1.8.2 (Drummond et al., 2005; Drummond et al., 2012)
and visualised using Tracer vers. 1.6 (Rambaut et al., 2014). Markov Chain Monte Carlo
sampling was run for 20,000,000 iterations sampling every 20,000 iterations using the GTR +
I substitution model with a strict clock and a nucleotide substitution rate of 1.55 %. Due to the
large number of parameters estimated under the GTR + I substitution model, we specified
lognormal priors on substitution rates to achieve acceptable effective sample sizes. Each
region was analysed separately for five independent model runs after standardising sample
sizes between regions, with all runs being combined for the final BSP analyses.

Evolutionary relationships among COI haplotypes were constructed using a median joining
algorithm and default settings (as per Bandelt et al., 1999) with the program NETWORK
vers. 4.5.1.0 (www.Fluxus-engineering.com/network_terms.htm). The time of divergence
between regions was estimated in ARLEQUIN using the corrected pairwise sequence distance
between regions minus the corrected pairwise sequence distance within regions ($d_C$; Tamura
Divergence times were also estimated using a Bayesians approach in BEAST. The time to most recent common ancestor (TMRCA), for both regions and all samples combined, was estimated using the same model parameters used in the BSP analyses.

**Nuclear gene analysis**

Samples (Table 1) were sequenced for a 206 bp region of the first intron of the S7 ribosomal protein gene using the primers S7RPEX2F and S7RPEX2R (Chow & Hazama 1998) to provide an independent assessment of evolutionary relationships among lineages. PCR cycling conditions consisted of a 95°C denaturation step for 15 min, followed by 35 cycles of 94°C denaturing for 30 sec, 50°C annealing for 30 sec, 72°C extension for 90 sec and a final 72°C extension for 10 min. PCR products were cleaned and sequenced using the same protocols as outlined in the mtDNA analyses and all sequences were deposited in GenBank (Accession numbers: XXX-XXX). Sequences were aligned, trimmed and edited using Geneious Pro. Allelic states of nuclear sequences with more than one heterozygous site were estimated with the Bayesian program PHASE vers. 2.1 (Stephens & Donnelly, 2003) as implemented in the software DnaSP vers. 5.0 (Librado & Rozas, 2009). We conducted three independent runs in PHASE with a burn-in of 10,000, and 100,000 iterations. All runs returned consistent allele identities and PHASE was able to determine all alleles with > 96% probability. Differences in allele frequencies between regions and sites were assessed using AMOVA in ARLEQUIN following the mtDNA analyses. Evolutionary networks among alleles were constructed using a median joining algorithm and default settings with the program NETWORK.

**Age-based demographic analysis**

Individual fish ages were determined by otolith analysis. Otoliths were affixed by
thermoplastic glue (Crystalbond 509; Aremco, Valley Cottage, NY, USA) to a clear glass slide and ground along the longitudinal axis to the core. The ground otolith was then removed and fixed to a clean slide with the flat surface down, and then polished until a thin transverse section (~150 μm) was obtained. Age was determined by counting alternating opaque and translucent zones along a consistent axis on the sectioned otolith face. Blind readings were made on three separate occasions, one to two weeks apart, and the final age (years) of an individual was determined when two or more counts agreed. When agreement was not achieved after three counts, the otolith was excluded from further analysis. Site specific growth patterns were estimated using the von Bertalanffy growth function (VBGF), represented by \[ L_t = L_\infty \left( 1 - e^{-K(t-t_0)} \right) \], where \( L_t \) is the predicted TL (mm) at age \( t \) (years), \( L_\infty \) is the mean asymptotic LF, \( K \) is the coefficient used to describe the curvature of fish growth towards \( L_\infty \), \( t \) represents age (years) and \( t_0 \) is the theoretical age at which TL is equal to zero, as described by K. To enhance precision among sites and account for early growth trajectories (Kritzer et al., 2001; Berumen, 2005), models were constrained to a common TL at settlement (50 mm). Growth parameters were compared among sites using bivariate 95% confidence ellipses surrounding the \( K \) and \( L_\infty \) estimates (Kimura 1980). Plots of size and age frequencies were compared among regions. Mean length, age and length at ages one through five were compared using Student’s t-tests with Welch’s correction when unequal variance was observed.

RESULTS

Mitochondrial DNA analysis

COI sequence data revealed 39 haplotypes with haplotype diversity ranging from 0.68 to 0.81 and nucleotide diversity ranging from 0.0020 to 0.0035 (Table 1). After grouping sites into
regions, AMOVA revealed that the majority of the genetic variability was explained by significant differentiation between these two regions (82\%, \( \Phi_{CT} = 0.827, P < 0.001 \); Table 2), with no shared haplotypes. The variance explained by the among-sites within-regions variance component was considerably smaller (0.1\%, \( \Phi_{SC} = 0.005, P = 0.224 \)) than that between regions and not significant. Pairwise \( \Phi_{ST} \) comparisons confirmed the regional patterns described above. There was no significant differentiation among sites within regions (pairwise \( \Phi_{ST} \) ranged from 0.00 to 0.046); however, comparisons among sites from different regions were all statistically significant (pairwise \( \Phi_{ST} \) ranged from 0.796 to 0.851).

We found limited evidence of population expansion or selective sweeps when considering sites individually. Despite negative values in seven of eight sites for Fu’s \( F_s \), only two values were statistically significant (Table 1). However, when sites were pooled by region, \( F_s \) values for both regions were significantly negative (RS-GA, \( F_s = -11.71, P < 0.001 \); GO-AG, \( F_s = -6.47, P = 0.007 \)). Pairwise mismatch distributions for the two regions were bimodal but did not show significant deviation from the sudden population expansion model (Fig. 2a and b; RS-GA, Harpending’s raggedness index, \( r = 0.14, P = 0.07 \); GO-AG, \( r = 0.06, P = 0.56 \)), suggesting that both regions have undergone recent population expansion. Time since most recent population expansion were different between the two regions; we estimated the RS-GA population (\( \tau = 2.502 \)) at 128,314 years, and the GO-AG population (\( \tau = 1.453 \)) at 74,516 years. BSP provided more recent estimates of population expansion (Fig. 2c and d), suggesting both regions experienced sudden population expansion approximately 18,000 years ago, with the GO-AG region expanding at a greater rate than the RS-GA region. However, given the overlapping confidence interval between the start and end of population reconstructions for both regions, it appears the overall magnitude of recent population expansion may be modest.
The median joining mtDNA haplotype network supports the genetic partitioning of *C. hemistiktos* into two distinct lineages separated by eight mutational steps, which correspond to the two geographically distinct regions (Fig. 3a). Additionally, the only previously reported COI sequence for *C. hemistiktos* from the Arabian Gulf (*i.e.*, Iran, Genbank accession number: HQ149822; Asgharian *et al.*, 2011) grouped within the GO-AG region haplotypes. Average corrected sequence divergence between these two lineages was large \((d_C = 0.0132)\) compared to genetic divergence within lineages (RS-GA, \(d_C = 0.0027\); GO-AG, \(d_C = 0.0028\)), suggesting a long period of isolation between them (~852,000 years based on a mutation rate of 1.55 % per Myr; Lessios, 2008). Bayesian analyses estimated the TMRCA as 505,700 years (95% highest posterior density [HPD] = 249,100-787,200 years) when considering both regions. Individually, TMRCA for the RS-GA region was 206,200 (95% HPD = 85,500-348,200 years), and 284,600 (95% HPD = 109,100-489,200 years) for the GO-AG region.

**Nuclear gene analysis**

Nuclear sequences for 261 individuals at the S7 locus returned seven variable sites with nine alleles. Median joining networks revealed only one shared allele between regions (Fig. 3b). AMOVA determined that allele frequencies differed significantly among regions \((F_{CT} = 0.287, \ P = 0.03)\) and among-sites within-regions \((F_{SC} = 0.287, \ P = 0.03)\), suggesting that this marker may be in the process of segregating by location, but at a slower rate than the mtDNA COI marker. Pairwise \(F_{ST}\) comparisons were congruent with COI data. Contrasts among sites from different regions were all statistically significant (pairwise \(F_{ST}\) ranged from 0.23 to 0.43), whereas comparisons within regions were all non-significant (pairwise \(F_{ST}\) ranged from -0.02 to 0.01) with the exception of comparisons between Muscat in the Gulf of Oman and Jubail in the Arabian Gulf \((F_{ST} = 0.23)\).
Demographic analysis

In agreement with the observations of Randall & Ben-Tuvia (1983), the maximum lengths of fish sampled from the three sites in the GO-AG region were all greater than the maximum lengths of fish sampled from the RS-GA region (Table 1). Mean lengths were also significantly greater in the GO-AG region (163 vs. 282 mm; t = 41.77, P < 0.001; Fig. 4a), but mean age was not different between regions (6.1 vs. 6.2 years; t = 0.47, P = 0.64; Fig. 4b). Comparisons of mean length at age one through five (at which point fish from the RS-GA region were approaching asymptotic size) revealed that fish from the GO-AG region were consistently attaining a larger size for a given age (Table 3). Subsequently, plots of bivariate confidence ellipses surrounding growth parameters K and \(L_\infty\) confirmed that the fish from the GO-AG region achieved greater asymptotic size but had similar growth coefficients (Fig. 5). These plots also revealed significant differences in asymptotic size within regions. Muscat and Musandam fish exhibited similar growth parameters, possessing much larger \(L_\infty\) estimates than at Abu Dhabi. However, these values were all greater than estimates for sites within the RS-GA region, which showed markedly reduced \(L_\infty\) estimates for sites within the central Red Sea (Al Lith and Thuwal).

DISCUSSION

Our results indicate a clear and pervasive barrier to dispersal in *C. hemistiktos* between RS-GA and GO-AG regions. Moreover, the lack of shared haplotypes between locations at the COI marker, and allele frequency shift for the slower segregating nuclear marker, shows that these two regions have been isolated for considerable amounts of time. This historical and contemporary separation has several consequences for both this species’ classification and management.
The location of the genetic break, on Southern Oman’s Arabian Sea coastline, corresponds with an environmentally dynamic area characterized by seasonal upwelling (Savidge et al., 1990). This results in highly variable seawater surface temperatures, which severely restrict coral reef development along a 400 km stretch of coastline (Schils & Coppejans, 2003). For C. hemistiktos, either traversing this distance as larvae, or subsequent settlement and recruitment to establish a viable reproductive population, appears impossible. Recent studies investigating individual larval trajectories have shown that larval dispersal may be more restricted than first thought (e.g., Jones et al., 2005; Almany et al., 2007). Moreover, estimates of dispersal in another grouper species demonstrate that over 50% of larvae are retained within 14 km of the spawning site (Almany et al., 2013). However, we found no evidence of genetic differentiation between the east and west coasts of the Red Sea, despite the presence of a deep trench (> 2000 m) between coasts, indicating successful dispersal occurs across this ~200 km distance (also see Fernandez-Silva et al., in review). The absence of connectivity between regions may be due, in part, to the reproductive timing recorded for C. hemistiktos. In the Gulf of Aqaba spawning occurs in July and August (El-Etreby et al., 2003) and in Oman between May and July (see Appendix S2 in Supporting Information). This timing corresponds with the seasonal upwelling on the Arabian Sea coastline of Oman, when sea surface temperatures can drop to 16 °C, which is well below the physiological tolerance of most tropical fish larvae (McCormick & Molony, 1995; Sponaugle et al., 2006). This large environmental fluctuation may provide an explanatory mechanism for the continuing barrier to larval dispersal in this species. Contrasting results are found when considering patterns of connectivity in other coral reef fish species across the Arabian Peninsula, even within the Cephalopholis genus. For example, Cephalopholis argus shares haplotypes among samples from the Gulf of Oman and Red Sea (DiBattista et al., 2013), indicating some level of successful dispersal across the Arabian Sea, but then this species is found throughout the
Indo-Pacific. Furthermore, the Oman clownfish, *Amphiprion omanensis*, have larvae that are capable of traversing the same distance (Simpson et al., 2014) despite the reduced pelagic larval duration associated with this genus (~10 days; Thresher et al., 1989). Given that variability in connectivity patterns can be prevalent in closely related species with similar life-history characteristics (Gaither et al., 2010; DiBattista et al., 2012), it seems likely that the existing patterns reported here may have originated via species-specific responses to historical regional disturbances (DiBattista et al., 2013), but are maintained by the present upwelling barrier to dispersal in the Arabian Sea. Demographic reconstructions suggest that both regions underwent population expansion at the end of the last glacial maximum (15 to 20 kya), whilst our estimates of lineage divergence predate this period, indicating that both Red Sea and Arabian Gulf populations have survived through multiple palaeoclimatic challenges after initial separation. It is likely that sea-level fluctuations as a result of Pleistocene glacial cycles caused multiple periods of isolation between the two regions, and after one of these events the isolated populations failed to reconnect. Whilst our coalescence estimates are necessarily approximate, our Bayesian estimation of lineage divergence coincides with the largest Pleistocene glacial period (c. 450,000 years ago), during which the Red Sea endured major sea level and salinity disturbance (Siddall et al., 2003). Further work in this understudied region is necessary to fully understand the role of such peripheral habitats in a global biogeographic context (Berumen et al., 2013; Bowen et al., 2013).

**Management considerations**

The presence of genetic structure between RS-GA and GO-AG regions clearly shows the spatial limit to dispersal in *C. hemistiktos*, as only a small number of migrants per generation are needed to homogenize populations (Slatkin, 1993). Thus, the potential for external
recolonisation between regions after a localised extinction event is low, and implies each
region should be treated as a separate management unit. Within regions, the lack of genetic
structure suggests recent or ongoing gene flow between sites, and highlights potential
recolonisation pathways. However, the scale of any ongoing demographically relevant
connectivity between sites remains unresolved, and warrants further investigation given the
limited spatial ranges of the populations may put them a greater risk of extinction from natural
or anthropogenic disturbance (Hastings & Botsford, 2006).

Our age-based demographic assessment provides evidence of significantly different
growth patterns, not only among regions, but also within regions. Marked differences in the
growth patterns of coral reef fishes have been demonstrated in other studies over relatively
small spatial scales (e.g., Gust et al., 2002; Kritzer, 2002; Taylor & McIwain, 2010). This
may be a response to a number of environmental or biological factors. For C. hemistiktos this
is especially pertinent given that the two regions we surveyed differ vastly in environmental
and biological conditions, as well as fishing pressure. Cephalopholis hemistiktos is commonly
found on open bottom and coral rubble reefs, and as a smaller-bodied Cephalopholis species
it is dominated by its congeners during interspecific interactions (Shpigel & Fishelson, 1989).
Given that the GO-AG region has less well developed, shallower reef systems (Coles, 2003)
and reduced fish diversity (Feary et al., 2010), we suspect that a combination of favourable
conditions, increased productivity and reduced interspecific competition may be driving the
differences in growth between regions. Additionally, levels of exploitation vary across
regions. In the RS-GA region C. hemistiktos is part of the artisanal catch but is generally not
targeted owing to its relatively small size (Gladstone, 2002; Roberts & Polunin, 1992),
whereas in the GO-AG region, where asymptotic size is larger, C. hemistiktos is actively
targeted as part of the commercial catch (Hashim, 1993), and in Oman this species accounts
for 42% of the total grouper catch (J. McIlwain, unpub. data). Whilst a full demographic
assessment of *C. hemistiktos* is beyond the scope of this study owing to a lack of histological
data, our results suggest that different management regimes may be needed within the spatial
boundaries demarcated by our genetic analyses.

**Taxonomic considerations**

Our genetic analyses reveal differences between populations that most likely represent cryptic
lineages of isolated populations, which were unable to reconnect during interglacial periods.
With at least five fixed mutational steps between regions (COI data), we estimate that these
populations have been separated for several hundred thousand years. Age-based otolith
analyses also confirm markedly different growth profiles between regions, matching previous
observation by Randall & Ben-Tuvia (1983), whom also reported morphological differences
in pectoral fin size and ray count, as well as oblique scale rows. Whilst we acknowledge reef
fish growth can be extremely plastic, limited pectoral fin ray counts during our sampling
study confirm Randall & Ben-Tuvia’s (1983) observations. Despite the weight of congruent
evidence for treating the two populations as recently diverged sister species, we found no
obvious differences in body colour between regions. These traits are frequently used as initial
indicators of species determination, but in reef fishes, many species exhibit stable colour
morphs that do not always represent genetically different lineages (e.g., Rocha *et al.*, 2004;
Messmer *et al.*, 2005; DiBattista *et al.*, 2012; DiBattista *et al.*, in reviewC), and conversely,
cryptic speciation with no obvious visual differences is also common (e.g., Drew *et al.*, 2008;
DiBattista *et al.*, 2011, 2013; Fernandez-Silva *et al.*, in review). Thus, we recommend that the
two allopatric *C. hemistiktos* populations be considered as distinct reciprocally monophyletic
species, with a formal description forthcoming.
Conclusions

Our results show species-level molecular divergence congruent with the morphological difference observed between two evolutionary distinct lineages of *C. hemistiktos*, signifying cryptic speciation among populations in the Red Sea and Gulf of Aden versus the Gulf of Oman and Arabian Gulf. Age-based demographic analyses also suggest differing management regimes may be needed across the species range to provide a buffer to future environmental disturbances. These findings underline the importance of molecular techniques, not only in the context of connectivity, but also as a taxonomic tool for understanding the origins and ongoing maintenance of biogeographic patterns within and among species.

ACKNOWLEDGEMENTS

This research was supported by the KAUST Office of Competitive Research Funds (OCRF) under award No. CRG-1-2012-BER-002 and baseline research funds to M.L.B., as well as National Geographic Society Grant 9024-11 to J.D.D. For support during sample collection trips we thank the KAUST Coastal and Marine Resources Core Lab, Amr Gusti, and Eric Mason at Dream Diver in KSA; the crew of the M/V Deli in Djibouti; the Wildlife Conservation General Administration of Sudan, Ministry of Agriculture and Fisheries in Oman, including Abdul Kareem; the Red Sea State Government and The Red Sea University in Port Sudan, as well as Equipe Cousteau support for the Sudan Shark and Ray Conservation and Management Program including Steven Kessel, Claudio Scarpellini, and Mohammed Younis. We are grateful to a crack team of sharpshooters for help with fish collections, including: Brian Bowen, Camrin Braun, Howard Choat, Richard Coleman, JP Hobbs, Gerrit Nanninga, Luiz Rocha, Tane Sinclair-Taylor and members of the Reef Ecology Lab at KAUST. Finally, we thank the KAUST Bioscience Core Laboratory for assistance with DNA
REFERENCES


Additional Supporting Information may be found in the online version of this article:

Appendix S1
Summary of morphometric differences between two populations of *C. hemistiktos* from Randall & Ben-Tuvia (1983).

Appendix S2
Phylogenetic

Appendix S3
Reproductive timing of *C. hemistiktos* sampled from the coast of Oman.
Biosketch

The authors’ interests are focused on illuminating the evolutionary processes that generate marine biodiversity. This study builds on recently completed range-wide phylogeographic surveys of over 20 Indo-Pacific reef fishes, providing a robust genetic foundation for subsequent investigations of isolated archipelagos and seas, including the evolution of the unique Red Sea fauna.

Author contributions: J.L.M conceived the initial idea for this study; all authors collected tissue samples; M.A.P and J.D.D produced DNA sequences and analysed the data; M.L.B and N.E.H. contributed reagents and materials; B.M.T and J.L.M produced and analysed age-based demographic data; M.A.P led the writing with contributions from all other authors.

Editor: Michelle Gaither
Table 1 Summary statistics for genetic and demographic analyses for *Cephalopholis hemistiktos* sampled from 11 sites around the Arabian Peninsula. Given are sample sizes (*n*), haplotype diversity (*h*), nucleotide diversity (*π*), maximum total length sampled (TL\(_{\text{MAX}}\); mm). Values in bold are statistically significant (*P* < 0.02; Fu, 1997).

<table>
<thead>
<tr>
<th>Site</th>
<th>Region</th>
<th>COI</th>
<th>h (\pm SD)</th>
<th>(\pi \pm SD)</th>
<th>Fu’s (F_3)</th>
<th>S7</th>
<th>Demography</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magna, KSA</td>
<td>Gulf of Aqaba</td>
<td>35</td>
<td>0.72 (\pm 0.04)</td>
<td>0.0027 (\pm 0.0018)</td>
<td>0.14</td>
<td>37</td>
<td>41</td>
</tr>
<tr>
<td>Thuwal, KSA</td>
<td>Red Sea</td>
<td>32</td>
<td>0.72 (\pm 0.05)</td>
<td>0.0027 (\pm 0.0018)</td>
<td>-0.89</td>
<td>40</td>
<td>57</td>
</tr>
<tr>
<td>Al Lith, KSA</td>
<td>Red Sea</td>
<td>13</td>
<td>0.69 (\pm 0.12)</td>
<td>0.0020 (\pm 0.0015)</td>
<td>-1.24</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>Sanganeb Atoll, Sudan</td>
<td>Red Sea</td>
<td>39</td>
<td>0.80 (\pm 0.06)</td>
<td>0.0032 (\pm 0.0021)</td>
<td><strong>-6.75</strong></td>
<td>33</td>
<td>29*</td>
</tr>
<tr>
<td>Farasan Islands, KSA</td>
<td>Red Sea</td>
<td>31</td>
<td>0.68 (\pm 0.06)</td>
<td>0.0023 (\pm 0.0016)</td>
<td>-1.34</td>
<td>27</td>
<td>36</td>
</tr>
<tr>
<td>Bay de Ghoubett, Djibouti</td>
<td>Gulf of Aden</td>
<td>49</td>
<td>0.73 (\pm 0.05)</td>
<td>0.0027 (\pm 0.0017)</td>
<td>-1.90</td>
<td>44</td>
<td>52</td>
</tr>
<tr>
<td>Muscat, Oman</td>
<td>Gulf of Oman</td>
<td>23</td>
<td>0.81 (\pm 0.06)</td>
<td>0.0035 (\pm 0.0022)</td>
<td>-1.59</td>
<td>20</td>
<td>493</td>
</tr>
<tr>
<td>Musandam, Oman</td>
<td>Gulf of Oman</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Abu Dhabi, UAE</td>
<td>Arabian Gulf</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Jubail, KSA</td>
<td>Arabian Gulf</td>
<td>52</td>
<td>0.73 (\pm 0.05)</td>
<td>0.0026 (\pm 0.0017)</td>
<td><strong>-3.79</strong></td>
<td>47</td>
<td>52*</td>
</tr>
</tbody>
</table>

KSA = Kingdom of Saudi Arabia; UAE = United Arab Emirates; * = not included in otolith age-based analyses
Table 2 Pairwise $F$-statistics for 8 populations of *Cephalopholis hemistiktos* from the Arabian Peninsula. Pairwise $\Phi_{ST}$ values based on COI data are below the diagonal and $F_{ST}$ values based on S7 data are above the diagonal. Values in bold are significant ($P < 0.05$) after correction for multiple tests using the false discovery rate (Benjamini *et al.*, 2006).

<table>
<thead>
<tr>
<th>Site</th>
<th>Magna</th>
<th>Thuwal</th>
<th>Al Lith</th>
<th>Sanganeb</th>
<th>Farasan Is.</th>
<th>Ghoubett</th>
<th>Muscat</th>
<th>Jubail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magna</td>
<td>-</td>
<td>-0.007</td>
<td>-0.019</td>
<td>-0.009</td>
<td>-0.014</td>
<td>-0.010</td>
<td>0.407</td>
<td>0.309</td>
</tr>
<tr>
<td>Thuwal</td>
<td>0.015</td>
<td>-</td>
<td>-0.017</td>
<td>0.001</td>
<td>-0.014</td>
<td>-0.004</td>
<td>0.366</td>
<td>0.287</td>
</tr>
<tr>
<td>Al Lith</td>
<td>0.038</td>
<td>0.090</td>
<td>-</td>
<td>-0.024</td>
<td>-0.023</td>
<td>-0.009</td>
<td>0.408</td>
<td>0.273</td>
</tr>
<tr>
<td>Sanganeb</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>-</td>
<td>-0.009</td>
<td>-0.001</td>
<td>0.399</td>
<td>0.299</td>
</tr>
<tr>
<td>Farasan Is.</td>
<td>0.001</td>
<td>0.005</td>
<td>0.004</td>
<td>0.000</td>
<td>-</td>
<td>-0.010</td>
<td>0.401</td>
<td>0.298</td>
</tr>
<tr>
<td>Ghoubett</td>
<td>0.000</td>
<td>0.046</td>
<td>0.018</td>
<td>0.000</td>
<td>0.018</td>
<td>-</td>
<td>0.427</td>
<td>0.331</td>
</tr>
<tr>
<td>Muscat</td>
<td><strong>0.809</strong></td>
<td><strong>0.808</strong></td>
<td><strong>0.822</strong></td>
<td><strong>0.796</strong></td>
<td><strong>0.796</strong></td>
<td><strong>0.815</strong></td>
<td>-</td>
<td><strong>0.233</strong></td>
</tr>
<tr>
<td>Jubail</td>
<td><strong>0.835</strong></td>
<td><strong>0.835</strong></td>
<td><strong>0.851</strong></td>
<td><strong>0.846</strong></td>
<td><strong>0.825</strong></td>
<td><strong>0.836</strong></td>
<td>0.000</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 3 Mean total length-at-age for age one through five for *Cephalopholis hemistiktos* among the two distinct regions. RS-GA = Red Sea and Gulf of Aden region; GO-AG = Gulf of Oman and Arabian Gulf region; $n =$ sample size; $t =$ test statistic; $d.f. =$ degrees of freedom; $P =$ probability value. Values in bold are significant ($P < 0.05$).

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Mean TL (mm) ± SE</th>
<th>n</th>
<th>RS-GA</th>
<th>n</th>
<th>GO-AG</th>
<th>t</th>
<th>d.f</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>120.9 ± 6.03</td>
<td>12</td>
<td>260.2 ± 14.49</td>
<td>5</td>
<td>153.2 ± 4.63</td>
<td>2.48</td>
<td>15</td>
<td>0.03</td>
</tr>
<tr>
<td>2</td>
<td>151.1 ± 4.63</td>
<td>18</td>
<td>226.8 ± 5.04</td>
<td>57</td>
<td>151.1 ± 4.72</td>
<td>11.06</td>
<td>56.9</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>3</td>
<td>150.3 ± 5.29</td>
<td>19</td>
<td>256.8 ± 4.11</td>
<td>105</td>
<td>150.3 ± 4.72</td>
<td>15.90</td>
<td>43.6</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>4</td>
<td>175.7 ± 5.62</td>
<td>23</td>
<td>262.2 ± 3.99</td>
<td>109</td>
<td>175.7 ± 5.62</td>
<td>12.54</td>
<td>47.5</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>5</td>
<td>176.5 ± 6.06</td>
<td>21</td>
<td>276.8 ± 4.23</td>
<td>99</td>
<td>176.5 ± 6.06</td>
<td>13.57</td>
<td>42.2</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Figure 1 Map of sampling locations for *Cephalopholis hemistiktos* across the Arabian Peninsula. Dark grey denotes extent of species range highlighting the disjunctive distribution pattern.

Figure 2 Mismatch distributions (a, b) and Bayesian skyline plots (c, d) based on mitochondrial COI sequence data (629 bp) for *Cephalopholis hemistiktos* from the Red Sea and Gulf of Aden (a, c), and Gulf of Oman and Arabian Gulf (b, d) regions. Grey lines on skyline plots represent 95% highest posterior densities.

Figure 3 Median-joining network based on (a) 629 base pairs of mitochondrial COI sequence data, and (b) a 206 base pairs of the S7 nuclear intron from *Cephalopholis hemistiktos* sampled across the Arabian Peninsula. Circles represent haplotypes and are sized in proportion to total frequency. Branches represent a single nucleotide change with crossbars indicating additional nucleotide changes. Unsampled haplotypes are unshaded and colours denote collection site as indicated by the key.

Figure 4 Length- (a) and age-frequency (b) distributions for *Cephalopholis hemistiktos* sampled from the Arabian Peninsula. Data are grouped into regions: Red Sea and Gulf of Aden (open bars), and Gulf of Oman and Arabian Gulf (black bars).

Figure 5 Comparison of growth parameters for *Cephalopholis hemistiktos* among eight sites using bivariate 95% confidence ellipses surrounding estimates of K (growth coefficient) and $L_\infty$ (mean asymptotic length).
(a) Red Sea and Gulf of Aden

\[ r = 0.14 \]
\[ p = 0.07 \]

(b) Gulf of Oman and Arabian Gulf

\[ r = 0.06 \]
\[ p = 0.56 \]

(c) \( N_e^* \pm 95\% \text{ HPD} \)

(d) \( N_e^* \pm 95\% \text{ HPD} \)