Genome-wide data for effective conservation of manta and devil ray species

Jane Hosegood[1,2,11] *, Emily Humble[2,3], Rob Ogden[3,4], Mark de Bruyn[1,5], Si Creer[1], Guy Stevens[2], Mohammed Abudaya[6], Kim Bassos-Hull[7], Ramon Bonfil[8], Daniel Fernando[2,9,10], Andrew D. Foote[1], Helen Hipperson[11], Rima W. Jabado[12], Jennifer Kaden[13], Muhammad Moazzam[14], Lauren Peel[2,15,16,17], Stephen Pollett[2], Alessandro Ponzo[18], Marloes Poortvliet[19], Jehad Salah[20], Helen Senn[13], Joshua Stewart[2,21], Sabine Wintner[22,23] and Gary Carvalho[1]

[1] Molecular Ecology and Fisheries Genetics Laboratory, Bangor University, Bangor, LL57 2UW, UK
[2] The Manta Trust, Catemwood House, Norwood Lane, Dorset, DT2 0NT, UK
[3] Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, Easter Bush Campus, EH25 9RG, UK
[4] TRACE Wildlife Forensics Network, Edinburgh, EH12 6LE, UK
[5] The University of Sydney, School of Life and Environmental Sciences, Sydney 2006, NSW, Australia
[6] National Research Center, Gaza City-Palestine
[7] The Center for Shark Research, Mote Marine Laboratory, 1600 Ken Thompson Parkway, Sarasota, FL 34236, USA
[8] Océanos Vivientes A. C. Cerrada Monserrat 9, Col. La Candelaria. CDMX 04380 Mexico.
[9] Department of Biology and Environmental Science, Linnaeus University, SE 39182 Kalmar, Sweden
[10] Blue Resources Trust, Colombo, Sri Lanka
[11] NERC Biomolecular Analysis Facility, Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, S10 2TN, UK
[12] Gulf Elasmo Project, P.O. Box 29588, Dubai, UAE
[13] RZSS WildGenes Lab, Royal Zoological Society of Scotland, Edinburgh EH12 6TS, UK
[14] WWF-Pakistan, 46-K, PECHS, Block 6, Karachi 75400, Pakistan
[15] School of Biological Sciences, University of Western Australia, Crawley, WA 6009, Australia
[16] The Australian Institute of Marine Science, Crawley, WA 6009, Australia
[17] Save Our Seas Foundation – D'Arros Research Centre, CH-1201, Geneva, Switzerland
[18] Large Marine Vertebrates Research Institute Philippines, Cagulada compund, Jagna, Bohol, Philippines
[19] Tolheksbos 57, 2134 GH Hoofddorp, the Netherlands
[20] Ministry of Agriculture Directorate General of Fisheries, Palestine
[21] Scripps Institution of Oceanography, La Jolla, CA, United States
[22] KwaZulu-Natal Sharks Board, Private Bag 2, Umhlanga Rocks 4320, South Africa
[23] School of Life Sciences, University of KwaZulu-Natal, Durban 4000, South Africa

*Corresponding author: bsp420@bangor.ac.uk or jane@mantatrust.org

Abstract

Practical biodiversity conservation relies on delineation of biologically meaningful units, particularly with respect to global conventions and regulatory frameworks. Traditional approaches have typically relied on morphological observation, resulting in artificially broad delineations and non-optimal species units for conservation. More recently, species delimitation methods have been revolutionised with High-Throughput Sequencing approaches, allowing study of diversity within species radiations using genome-wide data. The highly mobile elasmobranchs, manta and devil rays (*Mobula* spp.), are threatened globally by targeted and bycatch fishing pressures resulting in recent protection under several global conventions. However, a lack of global data, morphological similarities, a succession of recent taxonomic changes and ineffectual traceability measures combine to impede development and implementation of a coherent and enforceable conservation strategy. Here, we generate genome-wide Single Nucleotide Polymorphism (SNP) data from among the most globally and taxonomically representative set of mobulid tissues. The resulting phylogeny and delimitation of species units represents the most comprehensive assessment of mobulid diversity with molecular data to date. We find a mismatch between current species classifications, and optimal species units for effective conservation. Specifically, we find robust evidence for an undescribed species of manta ray in the Gulf of Mexico and show that species recently synonymised are reproductively isolated. Further resolution is achieved at the population level, where cryptic diversity is detected in geographically distinct populations, and indicates potential for future traceability work determining regional location of catch. We estimate the optimal species tree and uncover substantial incomplete lineage sorting, where standing variation in extinct ancestral populations is identified as a driver of phylogenetic uncertainty, with further conservation implications. Our study provides a framework for molecular genetic species delimitation that is relevant to wide-ranging taxa of conservation concern, and highlights the potential for genomic data to support effective management, conservation and law enforcement strategies.

Introduction

The Anthropocene has been characterised by unprecedented human exploitation of natural resources, resulting in global threats to biodiversity and extinction events across diverse taxa (Dirzo et al. 2014). Effective measures for biodiversity conservation require understanding and characterisation of diversity within and among species. The field of conservation genetics focuses on quantifying diversity across space and time (Allendorf et al. 2010), facilitated by increasingly powerful genome-wide data. Such genomic approaches also have applications in investigating the evolutionary processes generating biodiversity (Seehausen et al. 2014), providing further knowledge towards mitigating declines.

Biodiversity conservation is enacted through global conventions and regulatory frameworks implemented through legislation at the species level. Examples include the Convention on the International Trade in Endangered Species of Wild Fauna and Flora (CITES), and the Convention on the Conservation of Migratory Species of Wild Animals (CMS). In practice however, conservation initiatives and enforcement of regulations typically occur at a more local scale. Species therefore have two important impacts on conservation implementation; as units for inclusion in international conventions designed to coordinate conservation efforts, and representing identifiable targets against which conservation actions are directed and measured (Mace, 2004). Effective wildlife protection, management and law enforcement therefore depend on unambiguous classification of diversity into biologically relevant species units. Recent examples of proposed taxonomic revisions having far-reaching consequences for conservation include giraffe (Fennessy et al. 2016) and African elephant (Roca et al. 2001), where genetic research underpins possible reclassification and changes to the legal status of these megafauna.

Consequently, species delimitation, the process by which individuals are grouped into reproductively isolated and separately evolving units, is a fundamental application of genomic data to biodiversity conservation, with numerous methods available (Carstens et al. 2013; Grummer et al. 2014; Leache

et al. 2014; Rannala 2015). Traditional approaches typically relied upon morphological observation, often resulting in artificially broad delineations arising from difficulties detecting and identifying cryptic species (Frankham et al. 2012), and impeding conservation efforts. More recently, DNA sequencing has allowed genetic data to be utilised for species delimitation, although interpretation may be challenging in recently diverged groups with substantial incomplete lineage sorting (Maddison & Knowles, 2006). Species delineations should minimise ambiguity by defining species units on the basis of reproductive isolation associated with limited gene flow and a lack of shared alleles (Frankham et al. 2012) and may therefore be optimised with evaluation of genomic data (Shafer et al. 2015). Genome-wide multi-locus approaches have increased the resolution of species delimitation studies, clarified contentious relationships and phylogenies (Leache et al. 2014; Herrera & Shank, 2016), disclosed previously unknown diversity (Pante et al. 2014) and elucidated evolutionary processes (Foote & Morin, 2016; Campbell et al. 2018). In addition, there are further applications in characterisation of Conservation Units and Evolutionary Significant Units to further enhance conservation efforts (Funk et al. 2012).

The importance of judiciously defined species or management units is particularly apparent in fisheries management (Reiss et al. 2009). Overexploitation of marine fisheries is a global problem (Agnew et al. 2009) resulting in loss of genetic diversity and bottlenecks in many species (Hauser et al. 2002; Pinsky & Palumbi, 2014). One group of heavily targeted fishes are the manta and devil rays (*Mobula* spp.; collectively, mobulids). Despite substantial economic value through tourism (O'Malley et al. 2013), these highly-mobile, circumglobally distributed megafauna are threatened by intense targeted and bycatch fishing pressure driven by demand for gill plates (Couturier et al, 2012; O'Malley et al. 2017). Consumptive exploitation of manta and devil rays is considered unsustainable due to slow life history traits, hindering recovery from fishing impacts (Dulvy et al. 2014; Croll et al. 2016). To alleviate threats, all mobulid species are listed on CITES Appendix II to regulate international trade, and on CMS Appendices I and II to coordinate protection and implement conservation efforts. These fish are poorly studied however, and marked homogeneity in morphology among species, a lack of representative

global samples and population-level data, ongoing taxonomic debate, and ineffectual traceability measures constrain classification of optimal species units for conservation (Stewart et al. 2018). Understanding of evolutionary history and diversification in the Mobulidae derives from few studies, which indicate secondary contact and introgression among lineages may further impede efforts to delimit species boundaries (Kashiwagi et al. 2012; Poortvliet et al. 2015).

Recent evaluation of eleven previously recognised mobulid species across two genera recognised eight species, and called for the genus *Manta* (manta rays; *Manta alfredi* and *Manta birostris*) to be subsumed into *Mobula* (devil rays) (White et al. 2017). Other recent taxonomic changes include the resurrection of *Manta alfredi*; recognising two species of manta ray (Marshall et al. 2009; Kashiwagi et al. 2012), yet evidence remains of historic (Kashiwagi et al. 2012) and modern (Walter et al. 2014) hybridisation. In addition, a third putative species of manta ray is hypothesised to occur in the Caribbean (Marshall et al. 2009; Hinojosa-Alvarez et al. 2016). To date however, studies have relied on morphological observation (Notarbartolo Di Sciara 1987; Marshall et al. 2009; White et al. 2017) and/or been limited to evaluation of a handful of genetic markers, with heavy reliance on mitochondrial DNA (Kashiwagi et al. 2012; Hinojosa-Alvarez et al. 2016). Previous studies have also been geographically restricted and reliant on few samples (White et al. 2017), resulting in classifications that fail to encapsulate the extent of diversity within the group and compromise the effectiveness of conservation efforts.

Here, we generate double-digest Restriction-site Associated DNA sequence (ddRAD) data from the largest and most comprehensive set of mobulid tissue samples available. We demonstrate utility in delimiting informative species units for conservation, detecting cryptic diversity, and improving our understanding of associated evolutionary processes in a global radiation of socio-economically important marine megafauna.

Methods

Sampling and Sanger sequencing

Tissue samples were obtained from existing collections and sampling initiatives of researchers and organisations worldwide, yielding samples representing all mobulid species from a broad geographical range (Figure 1 and Supplementary Table 1), including *Mobula japanica, Mobula eregoodootenkee* and *Mobula rochebrunei,* currently considered junior synonyms of *Mobula mobular*, *Mobula kuhlii* and *Mobula hypostoma,* respectively (White et al. 2017), and an outgroup, *Rhinoptera bonasus.* Samples were identified to species based on characteristics described by Stevens et al. (2018), using original species names assigned and valid at the time of collection.

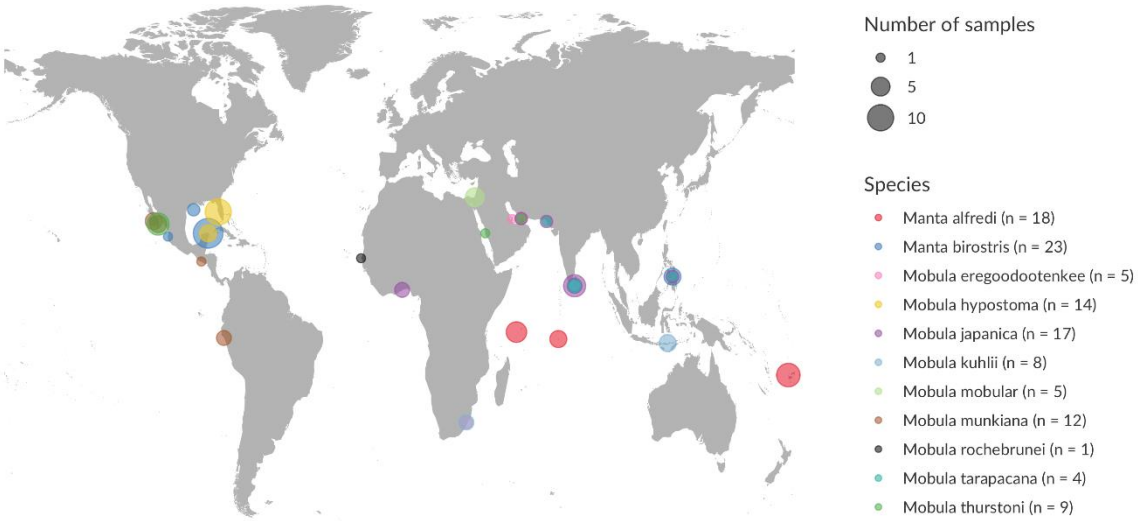


Figure 1: Sampling locations. Species are represented by coloured points, scaled for sample size. Total numbers of samples for each species provided in the key. Further details in Supplementary Table 1. Species names are those assigned at time of collection, some now considered invalid (White et al. 2017).

161 Genomic DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit and DNA yield measured
162 using a Qubit 3.0 Broad Range Assay. Extracts were quality assessed on 1% agarose gels stained with
163 SafeView. The single sample of *Mobula rochebrunei* was from a museum specimen stored in formalin
164 and yielded no detectable DNA.

165 To evaluate traditional markers for mobulid species delimitation, PCR amplification of an
166 approximately 650bp portion of the Cytochrome Oxidase Subunit I (COI) gene was performed using
167 universal Fish primers (Ward et al. 2005). Where these primers failed (for *M. munkiana* and *M.*
168 *hypostoma*), primers MunkF1 (GGGATAGTGGGTACTGGCCT) and MunkR1
169 (AGGCGACTACGTGGGAGATT) were designed using Primer-BLAST (Ye et al. 2012). 15 µl PCR reactions
170 consisted of: 5.6 µl nuclease-free water, 7.5 µl ReddyMix PCR Master Mix (ThermoFisher), 0.45 µl of
171 each primer and 1 µl DNA. PCR cycling conditions were: 95°C for 2 min, 35 cycles of 94°C for 30s, 54°C
172 for 30s and 72°C for 1 min and final extension of 72°C for 10 mins. Sanger sequencing was conducted
173 by Macrogen Europe. Data was aligned using ClustalW and the alignment checked for stop codons in
174 MEGA7 (Kumar et al. 2016). The HKY+G model was identified as optimal for our COI dataset using the
175 Find Best Model option in MEGA7. A Maximum Likelihood tree was built with 1,000 bootstrap
176 replicates.

177

178 **ddRAD library preparation and sequencing**

179 ddRAD libraries were prepared using a modified version of the original protocol (Peterson et al. 2012;
180 see Palaiokostas et al. 2015) with restriction enzymes *SbfI* and *SphI* (NEB). Unique P1 and P2 barcode
181 combinations were ligated to resulting DNA fragments, which were then size-selected between 400-
182 700bp using gel electrophoresis and PCR amplified. A pilot ddRAD library was sequenced on Illumina
183 MiSeq at the Institute of Aquaculture, University of Stirling. Subsequent ddRAD libraries were
184 sequenced by Edinburgh Genomics© on Illumina HiSeq High Output v4, 2 x 125PE read module.

185

186 **Data quality control and filtering**

187 Data quality was assessed with FastQC (Andrews, 2010), and processed in Stacks version 1.46 (Catchen
188 et al. 2011). The process_radtags.pl module in Stacks was used to demultiplex the data, filter for
189 adaptor sequences (allowing two mismatches), remove low quality sequence reads (99% probability)
190 and discard reads with any uncalled bases. To minimise linkage disequilibrium in the SNP data, only
191 forward reads were retained for subsequent analyses. Short fragments not removed through size-
192 selection were filtered with a custom bash script (8.5% of reads).

193 The denovomap.pl program in Stacks was used to assemble loci and call SNPs. The three main
194 parameters for assembly were those generating the largest number of new polymorphic loci shared
195 across 80% of individuals, following Paris et al. (2017). Four identical reads were required to build a
196 stack (-m), stacks differing by up to four nucleotides were merged into putative loci (-M) and putative
197 loci across individuals differing by up to five nucleotides were written to the catalog (-n), giving an
198 average coverage of 105x. We then used the populations.pl program in Stacks to generate two VCF
199 files containing SNPs present in at least 10 and 90 individuals, respectively. To remove paralogous loci
200 and mitigate for allele dropout (Arnold et al. 2013; Gautier et al. 2013), loci sequenced at greater than
201 twice or less than one-third the standard deviation of coverage, respectively, were identified and
202 excluded using VCFtools (Danecek et al. 2011). The remaining loci were assessed for excess
203 heterozygosity using VCFtools, and those exhibiting a significant probability of heterozygote excess
204 were excluded. Finally, since Stacks ignores indels, SNPs in the last five nucleotide positions were
205 assumed erroneous and excluded. The remaining loci and SNPs were written to a whitelist and filtered
206 for a single random SNP per locus to minimise linkage using populations.pl. This resulted in two final
207 SNP matrices, p10 and p90, with 7926 and 1762 SNPs and 47.1% and 14% missing data, respectively
208 (Supplementary Table 2).

209

210 **Monophyly and clustering**

211 Relationships among individuals were inferred through Maximum Likelihood phylogenetic analysis
212 using RAxML version 8.2.11 (Stamatakis 2014). Both ddRAD datasets were analysed since missing data
213 may influence aspects of phylogenetic inference (Leaché et al. 2015). The GTRGAMMA model of rate
214 heterogeneity was implemented following assessment of best fit models in jModelTest (Darriba et al.
215 2015) and support assessed with 1,000 bootstrap replicates.

216 RAxML identified four highly supported clades separated by long branches. To assess how individuals
217 cluster within these clades, dataset p10 was divided by clade (Supplementary Table 3) and Principal
218 Components Analysis (PCA) performed on each using the R package Adegenet (Jombart 2008). After
219 assessment of ten axes, three were retained in all cases. Populations.pl was used to calculate F_{ST} values
220 among inferred clusters.

221

222 **Bayes Factor Delimitation**

223 Bayes Factor Delimitation (Leache et al. 2014) was conducted using the modified version of SNAPP
224 (Bryant et al. 2012), implemented as a plug-in to BEAST version 2.4.8 (Bouckaert et al. 2014). The
225 method allows for direct comparison of Marginal Likelihood Estimates (MLEs) for alternative species
226 delimitation hypotheses, hereafter models, under the multispecies coalescent. Path sampling involved
227 10 steps (1,000,000 MCMC iterations, 20% burnin), implementing the log-likelihood correction. Since
228 MLEs are affected by improper prior distributions, a gamma distribution was implemented on the
229 lambda (tree height) parameter. To assess the effect of priors on the ranking order of models, models
230 were also assessed retaining the default 1/X distribution on lambda, implementing upper and lower
231 bounds (10,000 and 0.00001 respectively), for a proper prior. Bayes Factors ($2\log_e BF$) were calculated
232 from the MLE for each model for comparison (Kass & Raftery 1995; Leache et al, 2014), as follows:

233 $$2\log_e BF=2*(MLE_{null}-MLE_{test})$$

234 Positive $2\log_e BF$ values indicate support for the null model (<10 is decisive; Leache et al. 2014),
235 negative values favour the tested model.

236 Due to computational constraints, dataset p90 underwent Bayes Factor Delimitation and the data
237 were split by clade, as previously described, but including four random individuals from a sister species
238 to evaluate support for interaction from higher phylogenetic levels. Alternative models were informed
239 by the literature and analyses herein (Supplementary Tables 4-7). Models randomly assigning
240 individuals to two or three species were assessed for each clade. Null models matched species defined
241 by White et al. (2017).

242

243 **Species tree inference**

244 Relationships among the Mobulidae were estimated through Maximum Likelihood phylogenetic
245 analysis of both ddRAD datasets as above with RAxML (Stamatakis 2014). Consensus sequences for
246 each species unit were ascertained using populations.pl in Stacks, providing a population map
247 assigning individuals to optimal species units based on our previous analyses.

248 To test tree topology and evaluate uncertainty due to incomplete lineage sorting, species trees were
249 additionally evaluated with SNAPP (Bryant et al. 2012), allowing each SNP to have its own history
250 under the multispecies coalescent, whilst bypassing the need to sample individual gene trees. Due to
251 the computational capacity required to run SNAPP, three individuals per species were randomly
252 selected from dataset p90 whilst maximising geographical coverage within species. Random sampling
253 of individuals with replacement was repeated a further three times, resulting in four subsampled
254 alignments (Supplementary Table 8). MCMC chains consisted of 5,000,000 iterations, sampling every
255 1,000 and retaining default priors on lambda and theta for each independent analysis. Convergence
256 to stationary distributions were observed after 20% burnin in TRACER (Rambaut et al. 2018), the
257 distribution of trees visualised in DensiTree (version 2.2.6; Bouckaert 2010) and maximum clade

258 credibility (MCC) trees drawn using TreeAnnotator (version 2.4.7; Bouckaert et al. 2014). Alternative
259 prior combinations produced highly concordant results.

260 Multispecies coalescent based approaches assume that any discordance of topologies among loci
261 results from incomplete lineage sorting, and do not consider introgression as a source of discordance.
262 TreeMix (Pickrell & Pritchard, 2012) was applied to dataset p10 to evaluate evidence for significant
263 introgression events within the Mobulidae by investigating the extent to which variation between
264 user-defined groups is explained by a single bifurcating tree. Given uncertainty identified using SNAPP,
265 specifically regarding the placement of *M. mobular*, the three-population test (Reich et al. 2009) was
266 additionally used to test for 'treeness' between clades. Similar to TreeMix, the three-population test
267 estimates covariance of allele frequencies among groups, but is simpler and less parameterised;
268 potentially more powerful for identifying introgression. In addition to *M. mobular, M. alfredi* and *M.*
269 *thurstoni* were selected randomly to represent their respective clades.

270

271 Results

272 **Monophyly and clustering**

273 Maximum Likelihood phylogenetic trees based on two genome-wide SNP matrices were highly
274 congruent (Figure 2 and Supplementary Figure 1). Species groups formed well-supported clades
275 separated by long branches. Principal Components Analyses (PCA) within each clade mirrored patterns
276 in phylogenetic trees (Figure 3). Putative species, including recently synonymised species *Mobula*
277 *kuhlii* and *Mobula eregoodootenkee* formed both reciprocally monophyletic groups with high
278 bootstrap support (Figure 2) and tight clusters separated along axes explaining large portions of
279 variance (63%-74%; Supplementary Figure 2). Two reciprocally monophyletic groups were detected
280 within *Manta birostris*; an Atlantic and a global group, respectively (Figure 2), visible as clusters
281 through PCA (Figure 3A). One individual was equally well, albeit poorly, placed with each clade in the

282 two phylogenetic analyses (Fig 2 and Supplementary Figure 1) and in an intermediate position through
283 PCA (Figure 3A). *Mobula japanica* and *Mobula mobular* formed a single monophyletic group with 100%
284 bootstrap support (Figure 2), with no clear separation through PCA (Figure 3C-D). Whilst the first axis
285 provides limited evidence to suggest a clustering of individuals into Indo-Pacific and Atlantic (including
286 Mediterranean) groups, this explained only 8.6% of variance (Supplementary Figure 2E), with minimal
287 differentiation between these two clusters (F_{ST} = 0.061). Geographically separated populations of
288 *Manta alfredi* and *Mobula kuhlii* formed highly-supported monophyletic groups (Figure 2) and were
289 demarcated clearly through PCA (Figure 3B; Figure 3F), showing a high degree of differentiation (F_{ST} =
290 0.16 and F_{ST} = 0.32, respectively). COI sequences failed to achieve resolution sufficient to discriminate
291 putative species, and phylogenetic analysis showed several multifurcating nodes (Supplementary
292 Figure 3).

293

294 **Species Delimitation**

295 Species models were compared within clades using Bayes Factor Delimitation (Figure 2). Marginal
296 Likelihood estimates were unaffected by lambda priors, with no change in the rank order of models
297 (Supplementary Tables 4-7). We find decisive support for models recognising the Gulf of Mexico and
298 global *M. birostris* groups as separate species ($2\log_e BF$ = -775.82; hereafter '*Mobula* sp. 1' and '*M.*
299 *birostris*' respectively) and where individuals identified as *M. eregoodootenkee* belong to a separate
300 species to *M. kuhlii* ($2\log_e BF$ = -1007.04). Models splitting *M. mobular* and *M. japanica* based on
301 geographic origin marginally out-performed the null model. Geographically informed models involving
302 *M. alfredi* and *M. kuhlii* also performed well, achieving decisive support ($2\log_e BF$ = -1063.58 and -
303 1263.8, respectively). The null model was favoured within the *M. hypostoma* and *M. munkiana* clade.
304 Models assessing support for interaction from higher levels and testing random individual assignments
305 performed comparatively poorly (Supplementary Tables 4-7).

306

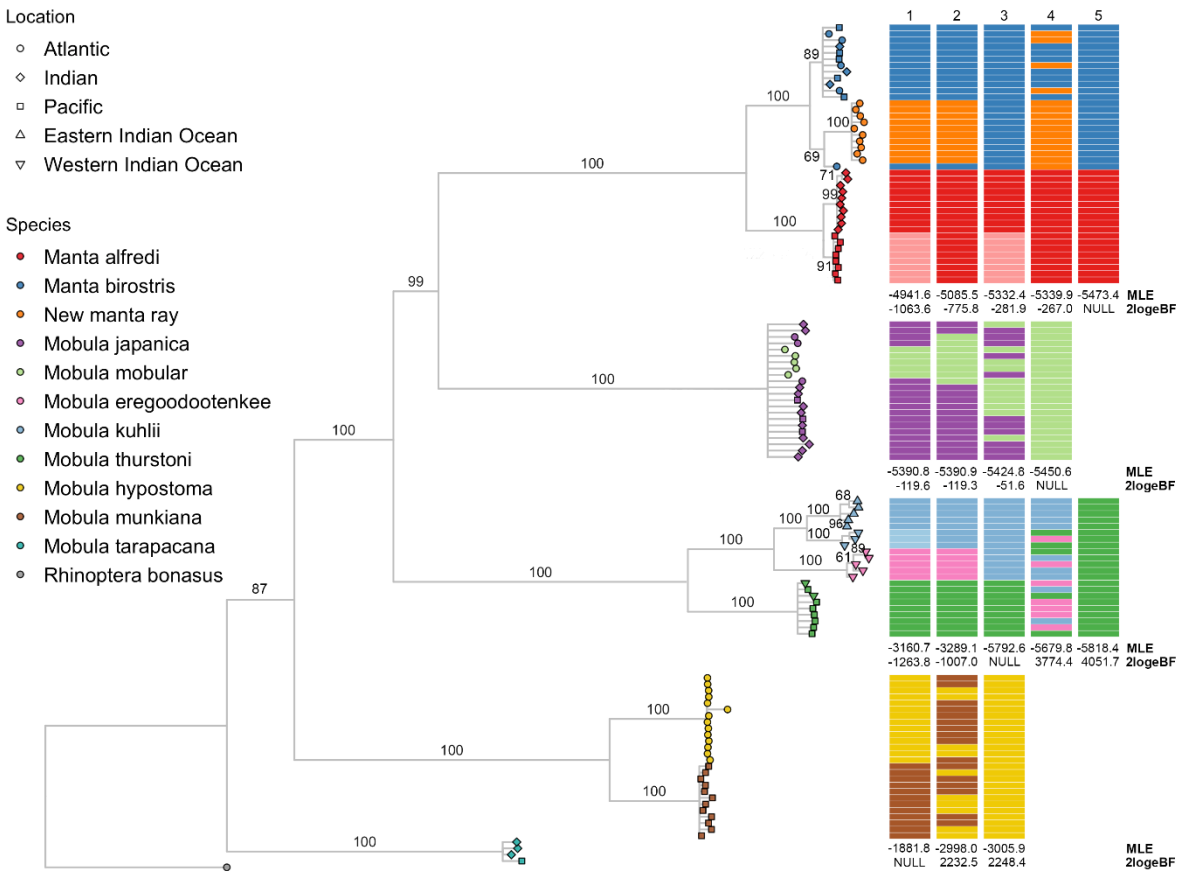


Figure 2: (Left) Maximum Likelihood Phylogenetic Tree of mobulid individuals based on 7926 SNPS (dataset p10). Coloured points indicate putative species, and shape indicates geographic origin of samples as specified in the key. Bootstrap values are shown on the branches and nodes with less than 50% support are collapsed. (Right) Bayes Factor Delimitation (BFD*) models with individuals assigned to species groups indicated by coloured bars are also presented, ranked in order of performance from left to right. Marginal Likelihood Estimates (MLEs) and Bayes Factors relative to the null model ($2\log_e BF$) are shown beneath each model for chains with a gamma prior on lambda. Models including individuals from a sister clade are not shown, as these consistently performed poorly. Species names are those assigned at time of collection, some now considered invalid (White et al. 2017).

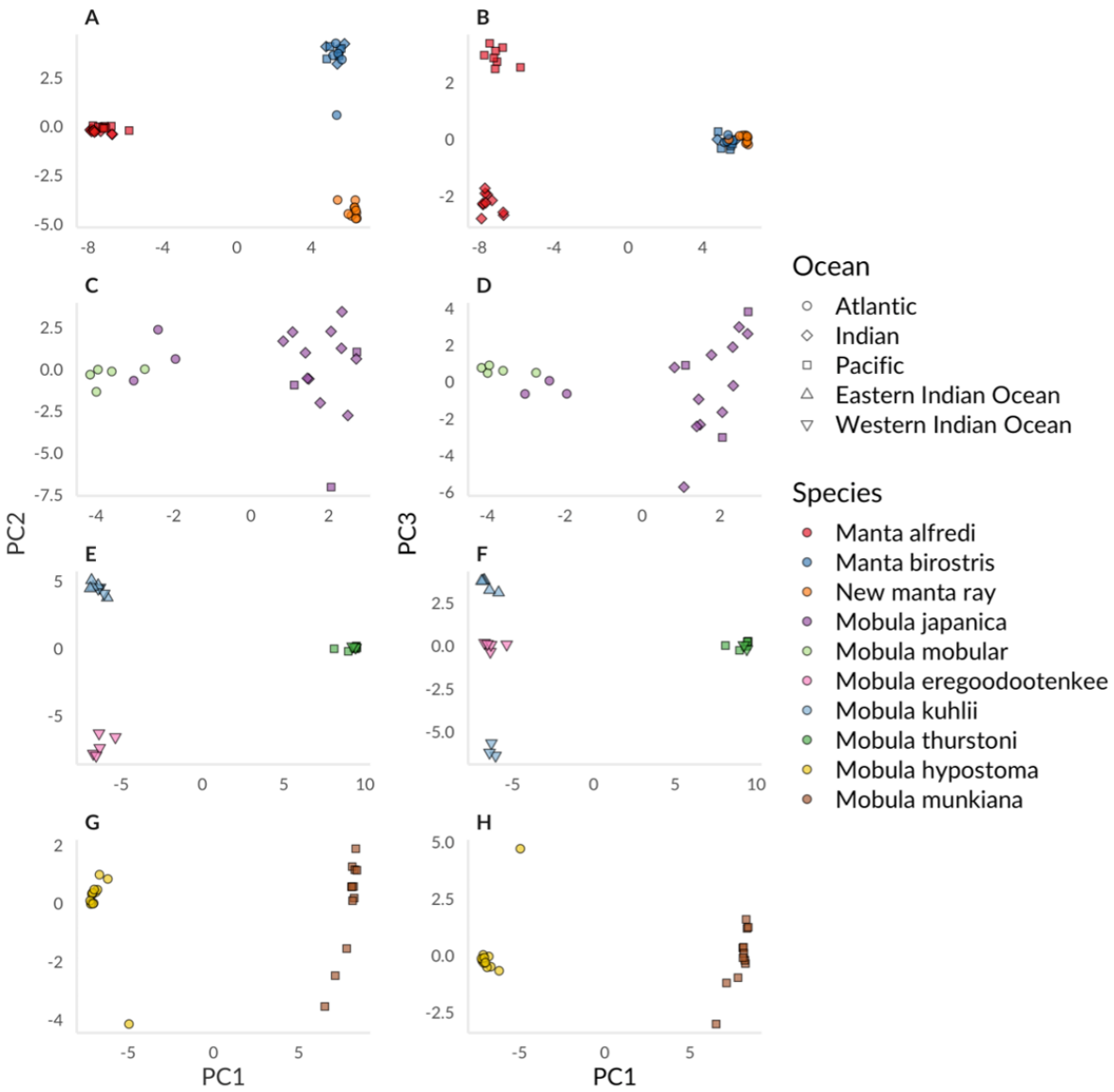


Figure 3: Principal Components 1-3 plotted for each mobulid clade. Individuals are represented by a point, colour indicates putative species, and shape indicates geographic origin of samples as specified in the key. Manta rays, A) PC1 and 2, and B) PC1 and 3; *M. mobular* and *M. japanica*, C) PC1 and 2, and D) PC1 and 3; *M. thurstoni*, *M. kuhlii* and *M. eregoodootenkee,* E) PC1 and 2, and F) PC1 and 3; *M. hypostoma* and *M. munkiana,* G) PC1 and 2, and H) PC1 and 3. Species names are those assigned at time of collection, some now considered invalid (White et al. 2017).

Relationships among species

Maximum Likelihood species trees based on two genome-wide SNP matrices were highly congruent (Figure 4 and Supplementary Figure 4). Consistent with White et al. (2017), manta rays were nested

within the genus *Mobula*, sister to *M. mobular* (≥98% bootstrap support) and hereafter all species of manta ray are referred to as *Mobula*. These trees strongly suggest that an undescribed third species of manta ray is sister to *M. birostris* (100% bootstrap support). *M. tarapacana* was tentatively placed on the group's oldest lineage (84% bootstrap support).

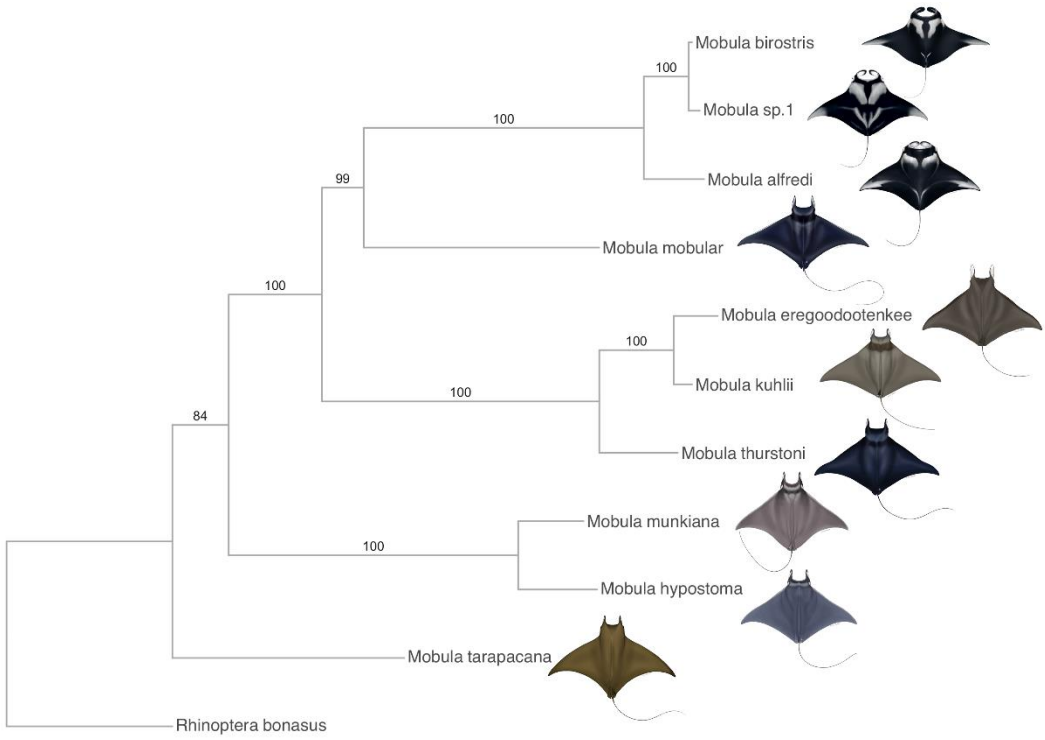
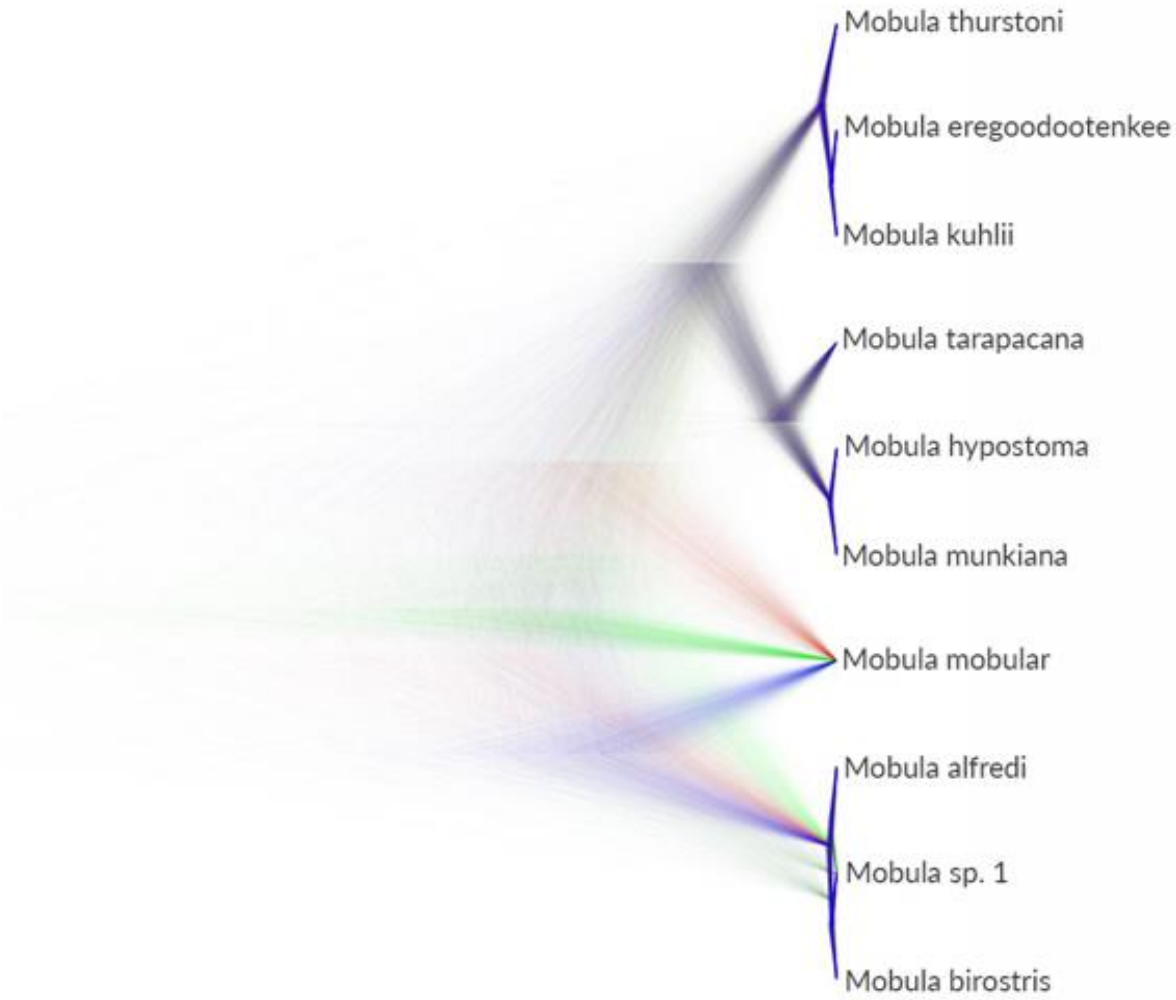


Figure 4: Maximum Likelihood tree of inferred mobulid species units based on 7902 SNPs (dataset p10). Bootstrap values are shown on the branches. The drawing of *Mobula* sp. 1 is based on images of dozens of individuals off the Yucatan Peninsula, Gulf of Mexico. Illustrations © Marc Dando.

Consensus species trees estimated under the multispecies coalescent exhibited relatively consistent topologies and theta estimates across independent runs, suggesting no major effect of subsampling on species tree topology inferred with SNAPP. *M. tarapacana* was consistently sister to *M. hypostoma* and *M. munkiana* (highest posterior density (HPD) = 1.0). Topological uncertainty at other nodes is apparent with a cloudogram of gene trees sampled from the posterior distribution (Figure 5 and

348 Supplementary Figures 5-7). Relationships between sister species within clades remained consistent
349 in alternative topologies within the 95% HPD, but large discrepancies in the placement of *M. mobular*
350 (including *M. japanica)* relative to other clades were observed (Supplementary Table 9).

351



352

353 Figure 5: SNP phylogeny of 30 individuals assigned to ten species units based on 1242 SNPs (dataset
354 p90, individual subsample 1; Supplementary Table 8). Tree cloud of sampled trees produced using
355 DENSITREE (representing samples taken every 1000 MCMC steps from 5,000,0000 iterations) from
356 SNAPP analysis to visualise the range of alternative topologies.

357

358 TreeMix inferred an admixture graph similar to trees produced with RAxML (Supplementary Figure 8),
359 explaining 99.86% of variance, indicating mobulid species placement is unaffected by admixture. We
360 found no evidence of introgression between clades containing *M. alfredi, M. mobular* and *M.*
361 *thurstoni,* through three-population tests (Supplementary Table 10).

362

363 Discussion

364 Genome-wide SNP data provide unprecedented resolution in a group of conservation concern, and
365 our analyses produced the most extensive phylogeny for the Mobulidae to date. In contrast to
366 previous studies examining mobulid diversity, the global nature of our dataset allowed us to identify
367 reproductive isolation between lineages and distinguish between population and species units
368 (Sukumaran & Knowles, 2017). We find a mismatch between current classifications and species units
369 optimal for conservation, with implications for management and law enforcement. We provide robust
370 evidence for a new species of manta ray and demonstrate that individuals identified as recently
371 synonymised species *Mobula kuhlii* and *Mobula eregoodootenkee* are distinct and reproductively
372 isolated. We therefore recommend that such units coincide with enforceable protection (see
373 Appendix 1 for critical evaluation). In addition, we detect cryptic diversity between geographically
374 segregated populations of *Mobula alfredi* and *Mobula kuhlii*, which may merit independent
375 management.

376 These findings have international implications for practical conservation of the Mobulidae since
377 legislation applies to species units and can severely impact anthropogenic pressures on wildlife
378 populations. Our data suggest that the oceanic manta ray (*M. birostris*) and an undescribed species of
379 manta ray (*Mobula* sp. 1) occur in sympatry in the Gulf of Mexico, since samples collected within sites
380 fall into both groups, and provides evidence of hybridisation between these species (Figure 2; Fig 3A).
381 Management of these similar species as independent units will therefore be challenging, potentially

requiring blanket protection of all manta rays in regions where sympatry and/or hybridisation occur, and indeed such protection already exists in Mexico. Notwithstanding, *Mobula* sp. 1 is likely to occur over a broad geographic range, given patterns of distribution of its closest relatives. To establish effective conservation and traceability measures for this new species, it will therefore be necessary to formally describe *Mobula* sp. 1 and determine the extent of its range, which may extend into international waters or span areas with high fishing pressure lacking suitable protective measures.

Similarly, *Mobula eregoodootenkee* (as formerly recognised), shown here to be distinct from *M. kuhlii*, shares a geographic range with the latter across a region with intense fishing pressure (Notarbartolo di Sciara et al. 2017). Inference from related species suggests low reproductive output likely resulting in population sizes vulnerable to exploitation (Dulvy et al. 2014; Croll et al. 2016). It is therefore imperative that such units are managed separately. In contrast, species such as *M. mobular* may be of lower conservation priority given that *M. japanica* is a junior synonym (White et al. 2017; this study - see Appendix 1). Significant population structure in *M. alfredi* and *M. kuhlii* indicates potential for future traceability work to determine regional location of catch in these species (Appendix 1), which is increasingly required to comply with global obligations (Nielsen et al. 2012). Additional population-level studies will allow further assessment of stock structure within fisheries and delineation of mobulid conservation units for effective management.

We find substantial uncertainty in the placement of *M. mobular,* and trees within the 95% HPD where the manta rays (formerly genus *Manta*) are nested within *Mobula* are present in approximately equal proportions to trees where the former genera are reciprocally monophyletic (Supplementary Table 9). In groups that have undergone rapid speciation with large ancestral effective population size, the effects of incomplete lineage sorting on species tree estimation are particularly prominent (Lischer et al. 2014; Flouri et al. 2018). The Mobulidae have undergone recent rapid bursts of speciation (Poortvliet et al. 2015), and our estimates of mutation-scaled effective population size were larger on deeper branches of the tree, indicating large effective population size of extinct shared ancestral

407 species (Supplementary Figure 9). Thus, standing variation in ancestral populations of mobulid rays is
408 likely to drive uncertainty with respect to the validity of the genus *Manta*. Since we find no evidence
409 of admixture driving these patterns, this uncertainty can be attributed to incomplete lineage sorting.
410 Factors such as similarities in life history and difficulties distinguishing between related species in
411 trade can lead to whole genera being listed on international conventions such as CITES designed to
412 preserve biodiversity. Our data therefore demonstrates the importance of understanding the extent
413 and nature of incomplete lineage sorting for effective conservation of threatened groups.

414 Genomic approaches are increasingly informative for inferring phylogenetic relationships among
415 species. Results must, however, be interpreted with caution. Our Maximum Likelihood analysis
416 identified *M. tarapacana* as the oldest mobulid lineage, coincident with similar analyses of nuclear
417 data (White et al. 2017), yet our Bayesian analyses consistently placed *M. tarapacana* sister to *M.*
418 *hypostoma* and *M. munkiana*; a previously unreported phylogenetic placement. Analyses employing
419 mitochondrial data support *M. tarapacana* as sister to the manta rays and *M. mobular* (Poortvliet et
420 al. 2015; White et al. 2017), an observation we were unable to reproduce. Discordant trees in
421 phylogenomic studies may be attributed to few loci driven by positive selection resulting in convergent
422 evolution, or evolutionary processes such as incomplete lineage sorting (Shen et al. 2017). Coalescent-
423 based approaches, as applied here, account for the independent history of each gene tree and are
424 therefore less likely to be influenced by single genes, highlighting the suitability of genome-wide data
425 for the inference of species relationships.

426 Here, genome-wide data considerably enhances delimitation of species units for the conservation of
427 manta and devil rays. These findings have profound implications for the practical conservation of a
428 group threatened by fishing, and are relevant to enforcement of CITES regulations by laying the
429 groundwork for species and regional traceability of parts in trade. Furthermore, we demonstrate the
430 ability of genomic data to resolve and identify diversity within organismal radiations and improve
431 understanding of evolutionary processes generating biodiversity. As such, this study provides a

framework for molecular genetic species delimitation which is relevant to other wide-ranging taxa of conservation concern, and highlights the potential for applied research in supporting conservation, management and law enforcement.

Acknowledgements

We are very grateful to the Save Our Seas Foundation (SOSF) and to The People's Trust for Endangered Species (PTES) for providing generous support for this work. JH is supported by a NERC CASE studentship through the ENVISION DTP (CASE partner - Royal Zoological Society of Scotland) and has received additional grants from the Fisheries Society of the British Isles (FSBI) and the Genetics Society. Data analysis was supported by the UK Natural Environment Research Council (NERC) Biomolecular Analysis Facility at the University of Sheffield.

The authors are very grateful to the following people and organisations for their help and support sourcing and collecting tissue samples; J. Spaet, A. Moore, R. Brittain, G. Phillips, J. Schleyer, F. Doumbouya, D. Bowling, H. Pacey, BD. Croll, K. Newton, H. Badar Osmany, S. Hinojosa, all LAMAVE staff and volunteers, field team Captains D. Dougherty, P. Hull, G. Byrd, K. Wilkinson, B. DeGroot and organisations Akazul, West Africa Musee de la mer a Dakar, the Barefoot Collection and Planeta Oceano. We would also like to thank all the staff at Atlantis-The Palm Dubai for giving access to specimens brought in by fishermen and for their valuable help with data collection and dissections.

Blue Resources Trust (BRT) would like to thank the Department of Wildlife Conservation and the Department of Fisheries and Aquatic Resources for support provided to the fieldwork carried out in Sri Lanka. BRT also acknowledges the generous support provided by the SOSF and the Marine Conservation and Action Fund (MCAF) that enabled fieldwork in Sri Lanka.

We thank Disney Conservation Fund, SOSF and Mote Scientific Foundation for supporting sample collection in Florida. Special thanks also to the Local Government Unit of Jagna, the Philippines Bureau of Fisheries and Aquatic Resources Region 7. The SOSF D'Arros Research Centre is a main affiliate of the Seychelles Manta Ray Project, funded by the SOSF. Sample collection in the Seychelles was approved by, and conducted with the knowledge of, the Ministry of Environment, Energy, and Climate Change.

The National Commission for Fisheries and Aquaculture of Mexico (CONAPESCA) allowed RB the collection of samples in Mexico through research permit PPF/DGOPA-091/15; the National

Commission for Natural Protected Areas (CONANP) of Mexico and authorities of the Biosphere Reserve of Whale Sharks kindly gave permission for work in the reserve. The SOSF and the MCAF provided funding for research in Mexico. The Perfect World Foundation generously funded RB for the replacement of a drone used to locate manta rays. The Mexican CITES authority, Secretary of Environment and Natural Resources (SEMARNAT) provided CITES export permit for tissue samples through permit MX 80544.

We also thank J. Taggart for his support with the ddRAD library preparation protocol, and for his help sequencing a pilot ddRAD library. G. Colucci assisted with DNA extractions and COI amplifications. In addition, we thank M. Dando for kindly agreeing for us to reproduce his illustrations.

AF was funded by the Welsh Government and Higher Education Funding Council for Wales through the Sêr Cymru National Research Network for Low Carbon, Energy and Environment, and from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 663830.

Author Contributions

JH, EH, GC, MdB, RO, SC and GS designed and conceived of the study and secured funding for consumables relating to laboratory work. EH, GS, DF, AP, MA, JS, SP, SW, RJ, MP, MM, KBH, RB, JS and LP were responsible for sourcing and collecting samples. JH, HS and JK carried out laboratory work. JH, EH, GC, MdB, RO, SC, HH, AF and HS contributed to analysis of genome-wide SNP data. Figures were designed by EH and JH and produced by EH. All authors contributed to writing and editing the manuscript.

References

Agnew DJ, Pearce J, Pramod G, Peatman T, Watson R, Beddington JR and Pitcher TJ. 2009. Estimating the worldwide extent of illegal fishing. PLoS ONE **4**(2):e4570.

Allendorf FW, Hohenlohe PA and Luikart G. 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics, **11**(10):697–709.

Andrews S. 2010. FastQC: A quality control tool for high throughput sequence data.

Arnold B, Corbett-Detig RB, Hartl D and Bomblies K. 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Molecular Ecology **22**:3179–3190.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A and Drummond

494 AJ. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Computational
495 Biology **10**(4):1–6.

496 Bouckaert RR. 2010. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics **26**(10),
497 1372-1373.

498 Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA and Roychoudhury A. 2012. Inferring species
499 trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis.
500 Molecular Biology and Evolution **29**(8):1917–1932.

501 Campbell CR, Poelstra JW and Yoder AD. 2018. What is Speciation Genomics? The roles of ecology,
502 gene flow, and genomic architecture in the formation of species. Biological Journal of the Linnean
503 Society **124**(4):561–583.

504 Carstens BC, Pelletier TA, Reid NM and Satler JD. 2013. How to fail at species delimitation. Molecular
505 Ecology **22**(17):4369–4383.

506 Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH and De Koning DJ. 2011. Stacks:
507 Building and Genotyping Loci De Novo From Short-Read Sequences. G3: Genes, genomes, genetics
508 **1**(3):171-182.

509 Couturier LIE, Marshall AD, Jaine FRA, Kashiwagi T, Pierce SJ, Townsend KA, Weeks SJ, Bennett MB
510 and Richardson AJ. 2012. Biology, ecology and conservation of the Mobulidae. Journal of Fish Biology
511 **80**(5):1075–1119.

512 Croll DA, et al. 2016. Vulnerabilities and fisheries impacts: the uncertain future of manta and devil
513 rays. Aquatic Conservation: Marine and Freshwater Ecosystems **26**(3):562–575.

514 Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics **27**(15):2156–2158.

515 Darriba D, Taboada GL, Doallo R and Posada D. 2015. jModelTest 2: more models, new heuristics and
516 high-performance computing. Nature Methods **9**(8):6–9.

517 Dirzo R, Young HS, Galetti M, Ceballos G, Isaac NJB and Collen B. 2014. Defaunation in the
518 Antrhopocene. Science **401**(6195)401–406.

519 Dulvy NK, Pardo SA, Simpfendorfer CA and Carlson JK. 2014. Diagnosing the dangerous demography
520 of manta rays using life history theory. PeerJ **2**:e400.

521 Fennessy J, Bidon T, Reuss F, Kumar V, Elkan P, Nilsson MA, Vamberger M, Fritz U and Janke A. 2016.
522 Multi-locus Analyses Reveal Four Giraffe Species Instead of One. Current Biology **26**(18):2543–2549.

523 Flouri T, Jiao X, Rannala B and Yang Z. 2018. Species Tree Inference with BPP Using Genomic
524 Sequences and the Multispecies Coalescent. Molecular Biology and Evolution **35**(10):2585-2593.

525 Foote AD and Morin PA. 2016. Genome-wide SNP data suggest complex ancestry of sympatric North
526 Pacific killer whale ecotypes. Heredity **117**(5):316.

527 Frankham R, et al. 2012. Implications of different species concepts for conserving biodiversity.
528 Biological Conservation **153**:25–31.

529 Funk WC, McKay JK, Hohenlohe PA and Allendorf FW. 2012. Harnessing genomics for delineating
530 conservation units. Trends in Ecology and Evolution **27**(9):489–496.

531 Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, Cornuet JM and Estoup A. 2013. The
532 effect of RAD allele dropout on the estimation of genetic variation within and between populations.
533 Molecular Ecology **22**:3165–3178.

534 Grummer JA, Bryson RW and Reeder TW. 2014. Species delimitation using bayes factors: Simulations

and application to the sceloporus scalaris species group (Squamata: Phrynosomatidae). Systematic Biology **63**(2):119–133.

Hauser L, Adcock GJ, Smith PJ, Ramírez JHB and Carvalho GR. 2002. Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (Pagrus auratus). Proceedings of the National Academy of Sciences **99**(18):11742-11747.

Herrera S and Shank TM. 2016. RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. Molecular Phylogenetics and Evolution **100:**70–79.

Hinojosa-Alvarez S, Walter RP, Diaz-Jaimes P, Galván-Magaña F and Paig-Tran EM. 2016. A potential third Manta Ray species near the Yucatán Peninsula? Evidence for a recently diverged and novel genetic *Manta* group from the Gulf of Mexico. PeerJ **4:**e2586.

Jombart T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics **24**(11):1403–1405.

Kashiwagi T, Marshall AD, Bennett MB and Ovenden JR. 2012. The genetic signature of recent speciation in manta rays (Manta alfredi and M. birostris). Molecular Phylogenetics and Evolution **64**(1):212–218.

Kass RE, Raftery AE. 1995. Bayes factors. Journal of the american statistical association **90**:773-795.

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.

Leaché AD, Banbury BL, Felsenstein J, De Oca ANM and Stamatakis A. 2015. Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. Systematic Biology **64**(6):1032–1047.

Leache AD, Fujita MK, Minin VN and Bouckaert RR. 2014. Species delimitation using genome-wide SNP Data. Systematic Biology **63**(4):534–542.

Lischer HEL, Excoffier L and Heckel G. 2014. Ignoring heterozygous sites biases phylogenomic estimates of divergence times: Implications for the evolutionary history of Microtus voles. Molecular Biology and Evolution **31**(4):817-831.

Mace GM. 2004. The role of taxonomy in species conservation. Philosophical Transactions of the Royal Society B: Biological Sciences **359**(1444):711–719.

Maddison W and Knowles L. 2006. Inferring phylogeny despite incomplete lineage sorting. Systematic Biology **55**(1):21-30.

Marshall AD, Compagno LJV and Bennett MB. 2009. Redescription of the genus Manta with resurrection of Manta alfredi (Krefft, 1868) (Chondrichthyes; Myliobatoidei; Mobulidae). Zootaxa **2301**:1–28.

Nielsen EE, et al. 2012. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. Nature communications **3**:851.

Notarbartolo di Sciara GN, Fernando D, Adnet S, Cappetta H and Jabado RW. 2017. Devil rays (Chondrichthyes: Mobula) of the Arabian Seas, with a redescription of Mobula kuhlii (Valenciennes in Müller and Henle, 1841). Aquatic Conservation: Marine and Freshwater Ecosystems **27**(1):197-218.

Notarbartolo Di Sciara G. 1987. A revisionary study of the genus Mobula Rafinesque, 1810 (Chondrichthyes: Mobulidae) with the description of a new species. Zoological Journal of the Linnean

577 Society **91**(1):1–91.

578 O'Malley MP, Lee-Brooks K and Medd HB. 2013. The Global Economic Impact of Manta Ray
579 Watching Tourism. PLoS ONE **8**(5): e65051.

580 O'Malley MP, Townsend KA, Hilton P, Heinrichs S and Stewart JD. 2017. Characterization of the trade
581 in manta and devil ray gill plates in China and South-east Asia through trader surveys. Aquatic
582 Conservation: Marine and Freshwater Ecosystems **27**(2):394-413.

583 Palaiokostas C, Bekaert M, Khan MGQ, Taggart JB, Gharbi K, McAndrew BJ and Penman DJ. 2015. A
584 novel sex-determining QTL in Nile tilapia (Oreochromis niloticus). BMC Genomics **16**(1):1–10.

585 Pante E, Abdelkrim J, Viricel A, Gey D, France SC, Boisselier MC and Samadi S. 2014. Use of RAD
586 sequencing for delimiting species. Heredity **114**(5):450–459.

587 Paris JR, Stevens JR and Catchen JM. 2017. Lost in parameter space: A road map for stacks. Methods
588 in Ecology and Evolution **8**(10):1360-1373.

589 Peterson BK, Weber JN, Kay EH, Fisher HS and Hoekstra HE. 2012. Double digest RADseq: An
590 inexpensive method for de novo SNP discovery and genotyping in model and non-model species.
591 PLoS ONE **7**(5):e37135.

592 Pickrell JK and Pritchard JK. 2012. Inference of Population Splits and Mixtures from Genome-Wide
593 Allele Frequency Data. PLoS Genetics **8**(11):e1002967.

594 Pinsky ML and Palumbi SR. 2014. Meta-analysis reveals lower genetic diversity in overfished
595 populations. Molecular Ecology **23**(1):29-39.

596 Poortvliet M, et al. 2015. A dated molecular phylogeny of manta and devil rays (Mobulidae) based
597 on mitogenome and nuclear sequences. Molecular Phylogenetics and Evolution **83**:72–85.

598 Rambaut A, Drummond AJ, Xie D, Baele G and Suchard MA. 2018. Posterior Summarization in
599 Bayesian Phylogenetics Using Tracer 1.7. Systematic Biology **67**(5):901–904.

600 Rannala B. 2015. The art and science of species delimitation. Current Zoology **61**(5):846–853.

601 Reich D, Thangaraj K, Patterson N, Price AL and Singh L. 2009. Reconstructing Indian population
602 history. Nature **461**(7263):489.

603 Reiss H, Hoarau G, Dickey-Collas M and Wolff WJ. 2009. Genetic population structure of marine fish:
604 mismatch between biological and fisheries management units. Fish and Fisheries **10**(4):361-395.

605 Roca AL, Georgiadis N, Pecon-Slattery J and O'Brien SJ. 2001. Genetic evidence for two species of
606 elephant in Africa. Science **293**(5534):1473–1477.

607 Seehausen O, et al. 2014. Genomics and the origin of species. Nature Reviews Genetics **15**(3):176–
608 192.

609 Shafer ABA, et al. 2015. Genomics and the challenging translation into conservation practice. Trends
610 in Ecology and Evolution **30**:78–87.

611 Shen XX, Hittinger CT and Rokas A. 2017. Contentious relationships in phylogenomic studies can be
612 driven by a handful of genes. Nature Ecology and Evolution **1**(5):1–10.

613 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
614 phylogenies. Bioinformatics **30**(9):1312–1313.

615 Stevens G, Fernando D, Dando M, Notarbartolo Di Sciara GN. 2018. Guide to the Manta and Devil
616 Rays of the World. Princeton University Press.

617 Stewart JD, et al. 2018. Research Priorities to Support Effective Manta and Devil Ray Conservation.
618 Frontiers in Marine Science **5**:314.

619 Sukumaran J and Knowles LL. 2017. Multispecies coalescent delimits structure, not species.
620 Proceedings of the National Academy of Sciences **114**(7):1607–1612.

621 Walter RP, et al. 2014. First record of living Manta alfredi × Manta birostris hybrid. Marine
622 Biodiversity **44**(1):1–2.

623 Ward RD, Zemlak TS, Innes BH, Last PR and Hebert PDN. 2005. DNA barcoding Australia's fish
624 species. Philosophical transactions of the Royal Society of London B: Biological Sciences
625 **360**(1462):1847–1857.

626 White WT, Corrigan S, Yang L, Henderson AC, Bazinet AL, Swofford DL and Naylor GJP. 2017.
627 Phylogeny of the manta and devilrays (Chondrichthyes: Mobulidae), with an updated taxonomic
628 arrangement for the family. Zoological Journal of the Linnean Society **182**(1):50-75.

629 Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S and Madden TL. 2012. Primer-BLAST: a tool to
630 design target-specific primers for polymerase chain reaction. BMC bioinformatics **13**(1):134.

631