

# **Genome-wide sequencing uncovers cryptic diversity and mito-nuclear discordance in the *Octopus vulgaris* species complex.**

**Michael D. Amor<sup>1,2\*</sup>, Stephen R. Doyle<sup>3</sup>, Mark D. Norman<sup>4</sup>, Alvaro Roura<sup>1</sup>, Nathan E. Hall<sup>5</sup>, Andrew J. Robinson<sup>6</sup>, Tatiana S. Leite<sup>7</sup>, Jan M. Strugnell<sup>8,1</sup>**

<sup>1</sup>Department of Ecology, Environment and Evolution, La Trobe University, Kingsbury Drive, Melbourne, Victoria, 3086, Australia

<sup>2</sup>Science Department, Museum Victoria, 11 Nicholson Street, Carlton, Victoria, 3053, Australia

<sup>3</sup>Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom

<sup>4</sup>Parks Victoria, Level 10, 535 Bourke Street, Melbourne, Victoria, 3001, Australia

<sup>5</sup>Department of Animal, Plant and Soil Sciences, La Trobe University, Kingsbury Drive, Melbourne, Victoria, 3086, Australia

<sup>6</sup>Genomics Research Platform, La Trobe University, Kingsbury Drive, Melbourne, Victoria, 3086, Australia

<sup>7</sup>Dept de Oceanografia e Limnologia, Universidade Federal do Rio Grande do Norte (UFRN), Natal, Brazil

<sup>8</sup>Centre for Sustainable Tropical Fisheries and Aquaculture, James Cook University, Townsville, Queensland, 4810, Australia

\*[michael.amor88@gmail.com](mailto:michael.amor88@gmail.com)

## ABSTRACT

Many marine species once considered to be cosmopolitan are now recognised as cryptic species complexes. Mitochondrial markers are ubiquitously used to address phylogeographic questions, and have been used to identify some cryptic species complexes; however, their efficacy in inference of evolutionary processes in the nuclear genome has not been thoroughly investigated. We used double digest restriction site-associated DNA sequencing (ddRADseq) markers to quantify species boundaries in the widely distributed and high value common octopus, *Octopus vulgaris*, comparing genome-wide phylogenetic signal to that obtained from mitochondrial markers. Phylogenetic analyses, genome-wide concordance and species tree estimation based on 604 genome-wide ddRADseq loci revealed six species within the *O. vulgaris* group. Divergence time estimates suggested modern-day species evolved over the last 2.5 ma, during a period of global cooling. Importantly, our study identified significant phylogenetic discordance between mitochondrial and nuclear markers; genome-wide nuclear loci supported *O. vulgaris* sensu stricto and Type III (South Africa) as distinct species, which mtDNA failed to recognise. Our finding of conflicting phylogenetic signal between mitochondrial and nuclear markers has broad implications for many taxa. Improved phylogenetic resolution of *O. vulgaris* has significant implications for appropriate management of the group and will allow greater accuracy in global fisheries catch statistics.

## Introduction

Cryptic diversity is common throughout the animal kingdom, occurring almost homogeneously among metazoan taxa<sup>1</sup>. The introduction of molecular sequencing has greatly improved our understanding of cryptic diversity<sup>2</sup> and it is clear that DNA barcoding and the use of mitochondrial DNA (mtDNA) markers have been invaluable in revealing many cryptic species complexes<sup>3-5</sup>. However, mtDNA markers have been shown to both underestimate and overestimate species boundaries, particularly among invertebrates<sup>6</sup>. Coalescence estimates based on mtDNA markers are typically poorly correlated with phylogenetic divergence at nuclear genes<sup>7</sup>, and estimates of species diversity based on mtDNA can lead to mischaracterisation of recently diverged species<sup>8</sup>. Although mtDNA barcoding has the potential to accurately delimit species<sup>5</sup>, one study reported a success rate of <30% when using this method<sup>9</sup>. This suggests that mtDNA-based inferences of species-level boundaries among broad ranging taxa may benefit from the inclusion of additional markers. In fact, many authors already recommend that DNA-based species delimitation should employ multiple loci to avoid such biases<sup>8,10,11</sup>.

Several studies have reported discordance between a single or small number of nuclear and mtDNA markers when attempting to delimit cryptic species<sup>12</sup> (see reference for review). However, few studies have compared species boundaries inferred from genome-wide nuclear markers to those estimated using mitochondrial markers. Genome-wide measures of genetic diversity are becoming increasingly common due to advances in high-throughput Next Generation Sequencing (NGS) technologies, and rapid genotyping of hundreds to thousands of markers from virtually any genome is now possible<sup>13</sup>, even with little or no previously available genetic information<sup>14</sup>. NGS approaches such as restriction site-associated DNA sequencing (RADseq)<sup>15</sup> are well

suited to phylogenetic studies, due to the low cost and high sample throughput relative to other NGS approaches<sup>16-18</sup>. For example, genome-wide data obtained via RADseq have revealed greater phylogenetic resolution of cichlids<sup>19</sup> and swordtails<sup>20</sup>, when compared to studies using mtDNA markers. This suggests that phylogenetic analyses of a large number of genome-wide markers may be well suited to resolve cryptic species complexes. This is important, as the misidentification of ecologically, medically and/or economically valuable species can have serious consequences for pest control, pathogen resistance, conservation of biodiversity and fisheries management<sup>21</sup>.

*Octopus vulgaris* is a highly-valued fisheries target throughout its range, and its north-west African population constitutes the largest single-species octopus fishery in the world<sup>22</sup>. Based on indistinguishable morphology, *O. vulgaris* was historically considered to have a cosmopolitan distribution, occurring in the Mediterranean Sea and the shallow tropical waters surrounding Australasia, Europe, Africa, Asia and the Americas<sup>23</sup>. However, recent mtDNA-based studies have provided some evidence for the existence of species boundaries among populations of *O. vulgaris*<sup>24-32</sup>. Such mtDNA evidence suggests *O. vulgaris* comprises a complex of morphologically similar but genetically distinct *O. vulgaris*-like species (the '*Octopus vulgaris* species complex'), which is composed of several hypothesised species 'Types'<sup>33</sup> (Fig. 1). A recent morphology-based investigation of this species complex (and close relatives) supported mtDNA-derived phylogenetic structure<sup>32</sup>. However, no studies of this group have attempted to validate these findings using nuclear data.

The poor state of octopus taxonomy is recognised as a major limitation of the global catch statistics reported by the FAO. Of the estimated 100 species of octopus

harvested, only four (*O. vulgaris*, *O. maya*, *Eledone cirrhosa* and *E. moschata*) are listed in global statistics<sup>34</sup>. Global production of octopuses exceeds 350,000 tonnes and has a total export value of US\$1.07 billion<sup>22</sup>, which surpasses many valuable finfish fisheries. However, the majority of octopus fisheries world-wide are in decline, in part due to overharvesting<sup>34</sup>. Improved taxonomy and identification of cryptic species boundaries is required to improve the reliability of catch estimates and inform sustainable management strategies of this global resource.

Current estimates of diversity within the *O. vulgaris* species complex are likely to underestimate cryptic speciation in this group. In this study, we have employed double digest (dd)RADseq<sup>16</sup> together with the most extensive global-scale sampling effort to date to investigate the phylogenetic relationships within the *O. vulgaris* species complex and its close relatives. Using genome-wide loci, we have characterised the degree of cryptic diversity within the purportedly widely distributed *O. vulgaris*, and estimated the number of extant species within this group. We contrast our findings with mtDNA-based phylogenies to determine the power of both approaches for resolving phylogenetic relationships. Finally, we use this new phylogenetic information to estimate divergence times among species.

## Results

### Data filtering and assembly

Sequencing of our ddRADseq library resulted in 17.25 million paired-end reads that contained a maximum of a single error per 1,000 bases (Phred quality score of 30). Of

the quality paired-end reads, 15.49 million (89.8%) paired reads sufficiently overlapped and therefore were merged into 'single end data'. Only read one of non-merged data was retained. Unmerged read two sequences were discarded as they were considered to be linked to their associated read one sequence and were lesser quality. Merged and non-merged (read one only) reads were demultiplexed according to their barcodes, which resulted in 13.8 and 1.7 million reads being retained, respectively (90.2% of raw reads). Approximately 9% of the total reads were classified to a microbial database and were discarded; therefore, 91% of the demultiplexed reads were retained. Our shared sequencing run contained 46 samples, 34 were attributed to our study, while the remaining 12 were for an independent project. On average, we obtained 307,316 reads per individual (standard deviation (SD) = 130,692). The 28 *O. vulgaris* group individuals (excluding *O. insularis*) had an average of 336,892 reads per individual (SD = 111,675).

Approximately 69% of our reads mapped to the *O. bimaculoides* genome<sup>35</sup> across the three datasets (RAD1: all individuals including *O. insularis* (n = 34), RAD2: two to three individuals with highest read number per species (n = 15) and RAD3: only *O. vulgaris* group individuals (n = 26)). The remaining 31% were assembled de novo. This combined approach produced 357,536, 189,678 and 284,924 total putative pre-filtered loci, respectively. Across all three datasets, an average of 3% of the total putative loci were filtered out as they did not meet our set quality criteria and an average of 99% of loci were filtered as they were not present in  $\geq 46\%$  of individuals. The minimum read depth (number of identical reads) per individual of each dataset was 4. Across all loci, the average read depths per individual for the RAD1, RAD2 and RAD3 datasets were 5.9 (SD = 4.4), 4.2 (SD = 3.4) and 5.2 (SD = 4.7), respectively. This

resulted in 298 loci with 2,896 informative sites for the RAD1 dataset, 1,060 loci with 7,688 informative sites for the RAD2 data set and 604 loci with 2,976 informative sites for the RAD3 dataset.

### **Analysis of mitochondrial DNA derived ddRAD loci**

A partial ND2 sequence alignment was obtained after mapping all demultiplexed reads to the mtDNA genome of *O. vulgaris* (Accession: NC\_006353.1). Sequence data was obtained for *O. vulgaris* s. s., *O. vulgaris* Type II, *O. vulgaris* Type III, *O. sinensis* and *O. tetricus* which all had a single EcoRI cut site within their mtDNA genome. All *O. insularis* (Brazil, Ascension Island and St Helena) and *O. cf. tetricus* individuals were missing data for this region, which may be the result of mutation(s) within the five base pair EcoRI cut site. Maximum Likelihood (ML) analysis of *O. vulgaris* species group individuals placed the five taxa into four distinct monophyletic clades (Bootstrap value [BS] =  $\geq 77$ ; Fig. S1), which is consistent with a previous COI based phylogeny of the group<sup>29</sup>. Each taxon was placed into a distinct clade with the exception of *O. vulgaris* s. s. and *O. vulgaris* Type III which composed a single monophyletic clade.

### **Phylogenetic inference using genome-wide ddRAD loci**

Maximum likelihood analysis of the *O. vulgaris* group and *O. insularis* individuals using the RAD2 data set (1060 loci) supported seven phylogenetic clades (Fig. 2). *Octopus insularis* formed a highly supported monophyletic clade separated from the *O. vulgaris*

group (BS = 100) by a large branch length. *Octopus vulgaris* Type II (Brazil) was placed as sister taxon to the remaining *O. vulgaris* group members and was therefore selected to root the phylogeny for subsequent analyses including only *O. vulgaris* group individuals.

Maximum likelihood analysis of the *O. vulgaris* group individuals (excluding *O. insularis*) based on the RAD3 data set (604 loci) resulted in six highly supported monophyletic clades (Fig. 3). Five of the six phylogenetic clades corresponded to the species previously identified based on differences in morphology<sup>32</sup>. These included *O. vulgaris* Type II (BS = 100), *O. vulgaris* s. s. (BS = 100), *O. sinensis* (BS = 100), *O. tetricus* (BS = 100) and *O. cf. tetricus* (BS = 100). Furthermore, individuals from South Africa (*O. vulgaris* Type III) also formed a distinct monophyletic clade (BS = 100).

### **Phylogenetic concordance among genome-wide loci and hypothesis testing of alternate topologies**

To investigate the support for sub-optimal trees among the 604 loci from the RAD3 data set, we performed 'Partitioned RAD' analysis<sup>36</sup> using a nearest-neighbour interchange (NNI) approach. The RAxML best tree was shown to be the most reliable representation of the RAD3 data-matrix as it was favoured by the greatest number of loci, and disfavoured by fewer than 10 loci. Our phylogeny (Fig. 3) contrasted previous mtDNA-based topologies, therefore, we performed topological comparisons via an Approximately Unbiased (AU) test, which showed that our RAxML best tree was the highest ranked topology (Posterior Probability [PP] = 1; Table 1). The first constrained



topology forced *O. sinensis* to be sister taxon to *O. tetricus* and *O. cf. tetricus*, as previously reported<sup>30</sup>. This tree was ranked second and could not be rejected as a significantly worse topology ( $p = 0.051$ ), however, the support for this tree was far lower compared to the first ranked tree ( $PP = <0.001$ ). The tree constraining the monophyly of *O. vulgaris* Type III and *O. vulgaris* s. s. (based on mtDNA analyses) was ranked third. This tree was significantly less likely ( $p = 0.004$ ) than our optimal topology, and the probability supporting it was substantially lower ( $PP = <0.001$ ). Forcing the monophyly of *O. vulgaris* Type III and *O. vulgaris* s. s. resulted in a sister taxon relationship of two distinct clades, and did not result in the same single clade, 'conspecific' relationship obtained via mtDNA-based analyses.

### **Species tree estimation**

Species tree estimation supported the presence of six species within the *O. vulgaris* group (Fig. 4). Each species was highly supported ( $BS = >99.1$ ), with *O. vulgaris* s. s., *O. vulgaris* Type II, *O. vulgaris* Type III, *O. cf. tetricus*, *O. tetricus* and *O. sinensis* supported as distinct species.

### **Divergence time estimation**

Divergence times were estimated among the six species identified in the present study; *O. vulgaris* s. s., *O. vulgaris* Type II (Brazil), *O. vulgaris* Type III (South Africa) *O.*

*tetricus*, *O. cf. tetricus* and *O. sinensis*. The *O. vulgaris* group was estimated to have evolved within the past 2.5 million years, throughout the Pleistocene (Fig. 5).

## Discussion

Here, we demonstrate the successful application of ddRADseq for resolving phylogenetic and species-level questions among cryptic taxa. Despite lacking an available reference genome, this approach enabled us to sample hundreds of loci in a random, yet repeatable, manner from across the genome of each individual. Our resulting data-matrix is the most comprehensive assembled so far to investigate species diversity within the *O. vulgaris* species complex. In comparison to previous attempts, which relied solely on mtDNA, our analyses yielded greater phylogenetic resolution and higher clade support. We recommend future studies investigating cryptic diversity use genome-wide approaches such as ddRADseq because of these advantages.

We analysed genome-wide RAD loci to investigate the phylogenetic and species-level relationships within the highest value fisheries target species of octopus in the world. Our study provides evidence for the existence of three cryptic species within the *O. vulgaris* species complex, which has traditionally been considered a single taxonomic unit; *O. vulgaris* s. s. (Mediterranean and north-east Atlantic), *O. vulgaris* Type II (southern Brazil) and *O. vulgaris* Type III (South Africa). We also support that *O. insularis* is distinct from the *O. vulgaris* species group, as previously published<sup>26</sup>. In accordance with previous studies, phylogenetic relationships and species tree estimation in our study support the distinct species status of *O. tetricus*, *O. cf. tetricus*

and *O. sinensis*, and confirm that New Zealand's *O. gibbsi* is a junior synonym of east Australia's *O. tetricus*<sup>30</sup>.

Our findings based on nuclear DNA provide greater resolution than previous studies using mtDNA alone to investigate phylogenetic relationships within the *O. vulgaris* group. Although these previous studies have shown some phylogenetic structure among species<sup>24,25,28-31</sup>, our finding of clear species-level differentiation between *O. vulgaris* s. s. and *O. vulgaris* Type III is a novel discovery. The recent morphological investigation of this species group<sup>32</sup> did not include *O. vulgaris* Type III individuals. Our study provides evidence that this taxon is a distinct species and, therefore, a comparison of its morphology with other members of the *O. vulgaris* group is required for taxonomic purposes.

Based on analyses of mtDNA, *O. sinensis* has the widest distribution of any taxa within the *O. vulgaris* group. A previous analysis of mtDNA suggests individuals from Asia and the Kermadec Islands are conspecifics<sup>31</sup>, despite allopatric distributions in the North and South Pacific Ocean. Preliminary mtDNA-based analyses also suggest that individuals from the Caribbean/Gulf of Mexico (Type I) and South Brazil (Type II) are conspecifics<sup>37</sup>, which also appears to be true for *O. insularis* from the Caribbean and eastern North Brazil<sup>38,39</sup>. Both of the above examples suggest these taxa can maintain gene-flow despite connectivity requiring trans-tropical dispersal. Since we did not include samples from the Caribbean/Gulf of Mexico and Kermadec Islands in our study, and given that our analyses find that mtDNA underestimates cryptic species diversity in this species complex, these findings require further validation and complimentary morphological analyses.

Biogeographic inconsistencies between mtDNA and nuclear DNA typically rule out lineage sorting as a mechanism contributing to discordance<sup>12</sup>. If nuclear DNA shows greater levels of geographic structure than mtDNA analyses, this discordance is often the result of isolation followed by secondary contact and hybridisation. We therefore suggest that the absence of any mtDNA-based phylogenetic signal between *O. vulgaris* s. s. and South African *O. vulgaris* Type III may be the result of hybridisation following secondary contact. Previously, two distinct mtDNA lineages of *O. vulgaris* Type III were identified along the South African coastline<sup>40</sup>. The dominant lineage was abundant along the entire South African coastline, whilst the rarer lineage was reported in two individuals from the south-eastern extreme of the known distribution (Durban). If our hypothesis is accurate, there may be a selective advantage for the foreign (*O. vulgaris* s. s.) mtDNA throughout South Africa as it appears to be either at, or near, fixation.

The relationships we identified among *O. sinensis*, *O. tetricus* and *O. cf. tetricus* using nuclear DNA contrasts previous mtDNA-based analyses. Analyses of Amor *et al.*<sup>30</sup> showed *O. sinensis* (treated as Asian *O. vulgaris*) was sister taxon to a clade composed of *O. tetricus* and *O. cf. tetricus*. Our study instead suggests that *O. cf. tetricus* is sister taxon to a clade containing *O. tetricus* and *O. sinensis*, since this topology was significantly more probable. Given that our results are based on genome-wide evidence, we consider our current findings to be a more reliable representation of the true relationships among these taxa.

Our phylogenetic reconstruction also showed overall improvements in support values throughout the tree when compared to previous mtDNA-based topologies. In particular, previous mtDNA-based studies found poor support for the clade containing

*O. vulgaris* Type II individuals<sup>29,41</sup>. Datasets with multiple unlinked nuclear DNA loci have a greater ability to resolve phylogenetic relationships in comparison to mtDNA alone<sup>10,11</sup>. Our findings highlight the advantages of using multiple genome-wide loci when investigating species-level relationships in the *O. vulgaris* group, with important implications for future investigations of cryptic diversity among other taxa. Despite these advantages for resolving phylogenetic relationships among closely related taxa, a short-coming of RADseq studies is that an increase in phylogenetic distance also increases locus loss due to mutations at restriction sites<sup>18,36</sup>. However, complimenting ddRADseq with target capture sequencing has shown promise for addressing allelic dropout<sup>42</sup>.

Throughout the evolutionary history of the *O. vulgaris* group, the global geography was relatively similar to the present, however there have been several changes to the global climate<sup>43</sup>. Our estimates of divergence times suggest that modern day *O. vulgaris* group taxa evolved from their MRCA during the last 3 Ma. Over this time, the climate has changed from relatively warm with a reduced equatorial-polar temperature gradient to relatively cool with an increase in Antarctic ice volume and the establishment of the Arctic ice sheet<sup>43,44</sup>. Present day *O. vulgaris* taxa have sub-tropical to temperate distributions in both hemispheres, or inhabit the deeper and cooler water in tropical regions<sup>40,45</sup>. Our study suggests these taxa arose from a common ancestor that inhabited shallow equatorial waters. Our estimates of divergence time suggest modern anti-tropical distributions originated during a period of global cooling, rather than the poleward range shift of each species being the result of increasingly warm and less suitable equatorial temperatures. Anti-tropical divergence in this group may be explained in part by the weakening of ocean currents at this period<sup>46</sup>, which would likely

have restricted gene-flow of previously connected populations by limiting larval dispersal via ocean currents.

Our divergence time estimates predict that the *O. vulgaris* group evolved during the last 2.5 Ma. This time period follows warming of global oceans, where temperatures increased between 2-3°C at mid-latitudes and 5-10°C in higher latitudes from 6-3.2 Ma<sup>46,47</sup>. Such temperature increases may have created unsuitably warm climates at low to mid-latitudes, potentially driving a poleward range shift and the isolation of northern and southern hemisphere populations. A poleward shift in distribution may have impeded the transport of larvae between continents via strong equatorial currents. Oceanic gyres in the Indian, southern Atlantic and northern Atlantic Oceans transport larvae among continents, but are strong barriers to planktonic dispersal between hemispheres<sup>48</sup>. Despite the planktonic larvae of *O. vulgaris* s. s. being primarily oceanic<sup>49</sup>, our phylogenetic results support previous studies that suggest *O. vulgaris* group taxa can only maintain gene-flow across distances of 2000-3000 km<sup>29,30</sup>, and appear unable to maintain gene-flow among major continents. *Octopus insularis* inhabits shallow tropical waters throughout the West Atlantic (Espírito Santo state (southern Brazil) to the Gulf of Mexico<sup>37,38,39</sup> and mid-Atlantic oceanic islands<sup>26,29,50</sup>). As this tropical species is a distinct phylogenetic lineage from the *O. vulgaris* group, how the above mentioned historical temperature increases and fluctuation in current strength influenced its distribution requires future attention.

We have provided compelling evidence for cryptic speciation within the *O. vulgaris* species complex, which has increased our understanding of the geographic boundaries of this species complex and its close relatives. Our study has revealed that

octopuses currently being treated as a single *O. vulgaris* species actually represent three distinct species (*O. vulgaris* s. s., *O. vulgaris* Type II and *O. vulgaris* Type III). Furthermore, our findings support that three other taxa belonging to the *O. vulgaris* group (*O. tetricus*, *O. cf. tetricus* and *O. sinensis*) and *O. insularis* are valid species. Our novel finding that analyses based on mtDNA does underestimate species diversity among octopuses shows that future phylogenetic studies require nuclear markers and/or be complimented with morphology-based evidence. Octopuses being exported globally under the name *O. vulgaris* are of extremely high market value and profile<sup>33</sup>. Aquaculture and captive growing of wild caught juveniles are receiving increasing profile and funding, particularly in China. Our data has, therefore, significant implications for the naming, marketing, value and documentation of commercially harvested octopuses in this species complex, and for the appropriate management and conservation of this highly valued fisheries resource.

## **Methods**

No live animals were experimented on during this study. Tissue samples were obtained via donation from existing museum or university collections, or taken from whole deceased animals that were purchased from fish markets.

## **Data accessibility**

Raw, unassembled RADseq data is hosted in DRYAD (doi:00.0000/dryad). Assembled

RADseq data and the final alignments (RAD1 – 3) are also available (doi:00.0000/dryad).

## Sampling

Tissue samples of sub-mature and mature individuals belonging to the *O. vulgaris* species group (*O. vulgaris* s. s. (Mediterranean and north-east Atlantic), *O. vulgaris* Type II (southern Brazil) and *O. vulgaris* Type III (South Africa), *O. tetricus* (east Australia), *O. cf. tetricus* (west Australia) and *O. sinensis* (Asia)) and close relative, *O. insularis* were obtained from 16 localities around the world (Fig. 1, Table 2), and were stored in ~90% ethanol at -80°C until processing.

## Library preparation and sequencing

Genomic DNA was extracted from mantle or arm tissue (~1 mm<sup>3</sup>) using a DNeasy Blood and Tissue Kit (QIAGEN) according to the manufacturer's instructions, except for the final elution which was repeated twice in a single aliquot of 57°C 'low TE buffer' to increase DNA yield. Where possible, skin was trimmed from tissue as a noticeable decrease in PCR efficiency was observed when skin was included with the tissue sample in the initial DNA extraction. Our ddRAD library was prepared using a modified version of the Peterson *et al.*<sup>16</sup> protocol (available at [michaelamor.com/protocols](http://michaelamor.com/protocols)). Samples were randomly allocated to a position in a 96-well, round bottom PCR plate to minimise biases in sample preparation.



Genomic DNA was digested for 18 hours in 30  $\mu\text{L}$  reactions composed of 20 units each of EcoRI-HF and ClaI restriction enzymes (New England Biolabs), 3  $\mu\text{L}$  CutSmart buffer (New England Biolabs), 20  $\mu\text{L}$  DNA solution (50 ng total DNA) and 4  $\mu\text{L}$   $\text{H}_2\text{O}$ . The enzyme ClaI is known to be sensitive to CpG methylation, which may result in locus dropout. However, octopuses (1.2%) and other molluscs (2.0%) display relatively low methylation levels<sup>51</sup>. Furthermore, methylation has been shown to range from having an important effect in early development of *O. vulgaris* to having no effect in adulthood<sup>52</sup>. To minimise this impact, we sampled only sub-mature and mature individuals from each locality.

Barcodes/adapters were ligated to digested DNA fragments in 40  $\mu\text{L}$  reactions composed of 30  $\mu\text{L}$  DNA digestion solution, 1  $\mu\text{L}$  T4 DNA ligase, 4  $\mu\text{L}$  T4 DNA ligase buffer (New England Biolabs), 1  $\mu\text{L}$   $\text{H}_2\text{O}$ , 2  $\mu\text{L}$  common anti-sense adapter (2  $\mu\text{M}$  stock concentration) and 2  $\mu\text{L}$  sense adapter (2  $\mu\text{M}$  stock concentration) which was unique to each sample. Ligation solutions were incubated at 16-18°C for 1 hour then 2  $\mu\text{L}$  EDTA (0.5 M) was added to stop the reaction. Non-ligated adapters were removed using a 0.7x (DNA solution volume) AMPure XP (Agencourt) magnetic bead purification. The purified DNA pellets were suspended in 12  $\mu\text{L}$  40°C  $\text{H}_2\text{O}$ .

PCR reactions were performed using 0.5  $\mu\text{M}$  of each indexed sense and anti-sense primer and (10  $\mu\text{M}$  stock concentration), 10  $\mu\text{L}$  KAPA HiFi Real Time PCR master mix (KAPA Biosystems) and 9  $\mu\text{L}$  size selected DNA solutions in a total reaction volume of 20  $\mu\text{L}$ . PCR cycle conditions included a single initial denaturing step (98°C for 2 minutes) and 18 cycles of denaturing (98°C for 15 seconds), annealing (60°C for 30 seconds) and extension (72°C for 30 seconds). Amplified DNA solutions were purified

using AMPure XP beads/PEG 6000 solution (0.7× DNA solution volume), quantified using a Qubit® 2.0 Fluorometer (Invitrogen) and DNA was pooled in equal proportions (15.4 ng per sample). The pooled library was electrophoresed on a 1.5% agarose gel, after which fragments between 300-350 base pairs (bp) were excised and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). The size selected library was diluted to 12 pM and sequencing was performed using a 600 cycle (paired-end) v3 MiSeq Reagent Kit on an Illumina MiSeq with 10% PhiX spiked into the run.

### **Quality filtering and bioinformatics pipeline**

Raw paired-end reads were merged using PEAR v0.9.4<sup>53</sup>. Merged and unmerged (read one only) reads were demultiplexed into individual sample read-sets based on their corresponding ligated inline barcode and indexed adapter using the 'process\_radtags' feature of STACKS v1.2.7<sup>54</sup>. This step was also used to trim all reads to 250 bp as a noticeable decrease in read quality was recorded after 250 bp; any additional low quality reads were discarded (based on phred 30 quality score). The remaining good quality reads were then filtered to exclude microbial contamination using Kraken v0.10.4<sup>55</sup> and unclassified, non-microbial reads were retained for phylogenetic analyses. Unmerged read two data were not included in the subsequent analyses as these reads did not meet the assumption that all reads are unlinked.

### **Mapping to the mitochondrial genome**

Demultiplexed filtered reads for each sample were mapped using Bowtie v2.2.2<sup>56</sup> to the Japanese *O. vulgaris* mtDNA genome (Genbank accession NC006353). BAM files were generated using Samtools v0.1.19<sup>57</sup>. Mapped reads were imported into CLC genomics Workbench v7.0.4 (<https://www.qiagenbioinformatics.com>). A depth of five reads per individual surrounding restriction sites was required before consensus sequences were obtained manually. Multiple sequence alignment was performed on consensus sequences using the Muscle algorithm<sup>58</sup>. jModelTest v0.1.1<sup>59</sup> was used to carry out statistical selection of best-fit models of nucleotide substitution of the alignment. The most appropriate substitution model (GTR+G) was selected based on 'goodness of fit' via the Akaike Information Criterion (AIC<sup>60</sup>).

Maximum likelihood analyses were performed using RAxML v8.0.19<sup>61</sup>. Strength of support for internal nodes of ML construction was measured using 1,000 rapid BS replicates. Bayesian Inference (BI) marginal PP were calculated using MrBayes v3.2.5<sup>62</sup>. Model parameter values were treated as unknown and were estimated. Random starting trees were used and analysis was run for 15 million generation sampling the Markov chain every 1,000 generations. After removing the initial 10% of samples, split frequencies within MrBayes as well as the program Tracer v1.6<sup>63</sup> were used to ensure independent Markov chains had converged and reached stationarity.

### **Assembly of RAD loci**

To assemble our loci we used a combined reference mapping and de novo assembly approach implemented in ipyrad v0.7.28<sup>64</sup>. Indexing and mapping to the available *O.*

*bimaculoides*<sup>35</sup> genome was performed using the Burrows-Wheeler Aligner (BWA)<sup>65</sup> software package. To ensure our reads mapped to the *O. bimaculoides* genome, we allowed for up to 15% sequence difference at each locus. All unmapped reads were then assembled de novo at 85% similarity. Further sequence quality filtering was performed to convert base calls with a phred Q score of <33 (1 error per 1,000 bases) into Ns, whilst excluding reads with >5 Ns. A minimum of five reads per individual was required for clustering of putative loci, and those with fewer than five reads were excluded. Putative loci containing more than two alleles were excluded as potential paralogs. Any locus containing one or more heterozygous site across more than three samples was excluded. In cases where individuals were missing a given locus, gaps were replaced with Ns in the multiple sequence alignment output.

Three datasets were generated and analysed. For each comparison, RAD loci were filtered to ensure that for each locus, a genotype was present in at least 44-47% of the individuals. This level of missing data was chosen as it maximised the number of loci retained without influencing the phylogenetic position of individuals (details below). The first (RAD1, 298 loci, n = 34) contained all samples (*O. vulgaris* species complex, *O. tetricus*, *O. cf. tetricus* and *O. insularis*) and required 15 individuals per locus. The second data set (RAD2, 1,060 loci, n = 7/15) included only the two individuals with the highest number of reads per taxa (*O. insularis* included three individuals to include North Brazil and Ascension) and required seven individuals per locus. The final data set (RAD3, 604 loci, n = 26) included only individuals belonging to the *O. vulgaris* species complex, *O. tetricus* and *O. cf. tetricus* and *O. sinensis* and required 12 individuals required per locus. Maximum likelihood phylogenies were constructed using the

GTRGAMMA model in RAxML v8.0.19<sup>61</sup>. Strength of support for internal nodes of ML construction was measured using 1,000 rapid BS replicates. We visualised the correlation between missing data and phylogenetic position to ensure that missing data did not influence the position of individuals throughout our phylogenies. This analysis was performed using the 'plot.locus.dist' function in the RADami<sup>66</sup> package, which was implemented in R v3.1.0<sup>67</sup>. Finally, we screened our datasets for outlier loci using BayeScan v2.1<sup>68</sup>. The chain sample size was set to 50,000 with a 5,000 step burn-in pilot chain. All downstream analyses were performed on the entire dataset as no outliers were detected.

### **Phylogenetic concordance among genome-wide loci**

Support for sub-optimal topologies was investigated via the 'partitioned RAD phylogenetic analysis' approach<sup>36</sup> using the RADami package<sup>66</sup> in R v3.1.0<sup>67</sup>. To visualise the number of loci supporting the optimal tree relative to neighbouring sub-optimal trees generated from the RAD3 data-matrix, a candidate pool of 250 trees was generated via NNI for comparisons with the RAxML optimal tree (i.e. 'best tree'). A set of unique trees for each locus was then generated by pruning the 251 trees to only those tips present in each locus. Site likelihoods for each locus-tree were calculated in RAxML v8.0.19<sup>61</sup> under the GTR+G model. The likelihood of each tree was then plotted against the number of loci favouring and disfavouring each tree.

### **Phylogenetic hypothesis testing**

To assess confidence of our study's optimal topology, relative to previously published mtDNA-based alternate topologies, the AU test<sup>69</sup> was performed using the software package CONSEL<sup>70</sup>. Three topologies were investigated; the optimal topology generated by analysis of the *O. vulgaris* group via the RAD3 data set, an alternate hypothesis whereby *O. sinensis* was sister taxon to a clade containing *O. tetricus* and *O. cf. tetricus* (mtDNA; Amor *et al.*<sup>30</sup>), and a second alternate hypothesis whereby *O. vulgaris* Type III and *O. vulgaris* s. s. form a monophyletic clade (mtDNA; Guerra *et al.*<sup>27</sup>; Amor *et al.*<sup>29</sup>). Constrained topologies were constructed using the concatenated RAD3 data set based on 608 loci and 28 individuals. Site based likelihoods for each alternate tree were estimated using RAxML v8.0.19<sup>61</sup> under the GTR+G evolutionary model. Significance values for the AU test and BS/PP values were calculated for each topology.

### **Species tree estimation**

Coalescent based species tree estimation on the RAD3 data-matrix was conducted following a quartet inference approach<sup>71</sup> using the SVDQuartet command in PAUP v4.0a146 for Unix/Linux<sup>72</sup> (Swofford 2003). The ML based analysis was performed using 100,000 randomly generated quartets and 1,000 BS replicates.

### **Combined Bayesian inference and divergence time estimation**

The consensus sequence for each species was obtained and aligned using Geneious v7.1.7<sup>73</sup>. Statistical selection of best-fit models of nucleotide substitution were undertaken using jModelTest v0.1.1<sup>59</sup>. The appropriate model (GTR+G) was chosen based on 'goodness of fit' via the AIC. Ultrametric topologies and divergence times among clades were co-estimated using BI analyses implemented in BEAST v2.3.1 and its associated LogCombiner and TreeAnnotator software<sup>74</sup>. All analyses were conducted using a log-normal relaxed clock<sup>75</sup> under the Yule-speciation process<sup>76</sup> with a normally distributed prior<sup>77</sup>. Topologies were estimated using random starting trees, gamma site models and log-normal relaxed clocks and were run for 750 million generations, sampling the Markov chain every 5,000. Three independent runs were performed for each evolutionary rate analysed and the resulting log and tree files were combined. Tracer v1.6<sup>63</sup> was used to ensure Markov chains had reached stationarity, effective sampling size (ESS) was adequate (>1,000) and to determine the correct 'burn-in' for the analysis. The 'maximum clade credibility tree' was then obtained.

To estimate divergence times within the *O. vulgaris* species group we used the rate of 0.0036 neutral substitutions per million years in cephalopods<sup>35</sup>. We also estimated divergence times a substitution rates based on the distantly related molluscan bivalve family Arcidae Lamarck, 1809, which was similarly used in a recent study investigating the historical population expansion of the giant squid genus *Architeuthis*<sup>78</sup>. Three rates of evolution have been estimated from bivalves, each calculated using the nuclear marker histone 3: (1) 0.14-0.2%, (2) 0.09-0.1% and (3) 0.02-0.03% per million years<sup>79</sup>. As we were interested in estimating divergence times among a group of closely related species, we used the latter rate which was based on the calibration point at the

split between the sub-genera *Anadara* s. s. and the sub-genus *Grandiarca*. Mean values at the mid-point of each rate range were calculated and the upper and lower range were set using a normally distributed prior. The two remaining bivalve rates of evolution were not used for our estimates as they were based on deeper phylogenetic calibration points; (1) the subfamilies Noetiinae and Striarcinae and (2) Anadarinae and the clade containing subgenera *Fugleria* and *Cucullaearca*.

## **Acknowledgements**

We thank CC Lu (Museum Victoria, Australia), Chih-Shin Chen and Chia-Hui Wang (National Taiwan Ocean University, Taiwan), Xiaodong Zheng and Yuanyuan Ma (Ocean University of China, China) and Eric Hochberg (Santa Barbara Museum of Natural History, USA) for project support. We also thank the following people for contributing tissue samples; Eduardo Almansa (Instituto Español de Oceanografía, Spain), Jorge Ramos Castillejos (University of Tasmania, Australia), Mandy Reid (Australian Museum, Australia) Ian Gleadall (Tohoku University, Japan), Xiaodong Zheng, Vladimir Laptikhovsky (Centre for Environment, Fisheries and Aquaculture Science, UK) and Warwick Sauer (Rhodes University, South Africa). A La Trobe University internal grant and a La Trobe Asia grant awarded to JMS funded experiments and travel. An Australian Biological Resources Study (ABRS) grant awarded to MDA assisted with the travel costs. A La Trobe University (School of Life Sciences) postgraduate publication grant awarded to MDA provided financial support during the writing of this manuscript. SRD was supported by an Illumina MiSeq grant. The Brazilian



Institute of Environment and Renewable Natural Resources (IBAMA/ICMBio), the Brazilian National Research Council (CNPq559863/2008-0; 481491/2013-9) and Ciencias do Mar-CAPES (043/2013) are acknowledged for their logistic and financial support to TSL. MDA thanks Julian Finn and the staff of the Department of Marine Invertebrates at Museum Victoria, Australia for assistance in carrying out this project and Shannon Hedtke (La Trobe University, Australia) for assistance with phylogenetic hypothesis testing.

### **Author contributions statement**

MDA, MDN, AR, TSL and JMS obtained tissue samples. MDA, SRD and JMS conceived the experiments. MDA conducted experiments with assistance from SRD. Bioinformatics was performed by MDA, NEH and AJR. MDA wrote the manuscript. All authors reviewed the manuscript.

### **Additional information**

### **Competing financial interests**

The authors declare no competing financial interests.

### **References**

- 1 Pfenninger, M. & Schwenk, K. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC evolutionary biology* **7**, 121 (2007).
- 2 Bickford, D. *et al.* Cryptic species as a window on diversity and conservation. *Trends in ecology & evolution* **22**, 148-155 (2007).
- 3 Hebert, P. D. N., Cywinska, A., Ball, S. L. & DeWaard, J. R. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences* **270**, 313-321 (2003).
- 4 Hebert, P. D., Penton, E. H., Burns, J. M., Janzen, D. H. & Hallwachs, W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 14812-14817 (2004).
- 5 Hebert, P. D. N., Stoeckle, M. Y., Zemplak, T. S. & Francis, C. M. Identification of birds through DNA barcodes. *PLoS Biol* **2**, e312 (2004).
- 6 Hurst, G. D. & Jiggins, F. M. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society of London Series B-Biological Sciences* **272**, 1525-1534 (2005).
- 7 Hudson, R. R. & Turelli, M. Stochasticity overrules the “three-times rule”: genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution* **57**, 182-190 (2003).
- 8 Moritz, C. & Cicero, C. DNA Barcoding: Promise and Pitfalls. *PLoS Biol* **2**, e354 (2004).
- 9 Johnson, N. K. & Cicero, C. New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. *Evolution* **58**, 1122-1130 (2004).
- 10 Edwards, S. V. *et al.* Speciation in birds: genes, geography, and sexual selection. *Proceedings of the National Academy of Sciences* **102**, 6550-6557 (2005).
- 11 Edwards, S. & Bensch, S. Looking forwards or looking backwards in avian phylogeography? A comment on. *Mol. Ecol.* **18**, 2930-2933 (2009).
- 12 Toews, D. P. L. & Brelsford, A. The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* **21**, 3907-3930 (2012).
- 13 Stapley, J. *et al.* Adaptation genomics: the next generation. *Trends in Ecology & Evolution* **25**, 705-712 (2010).
- 14 Davey, J. W. *et al.* Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499-510 (2011).
- 15 Baird, N. A. *et al.* Rapid SNP discovery and genetic mapping using sequenced RAD markers. *Plos One* **3** (2008).
- 16 Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double Digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One* **7** (2012).
- 17 Lemmon, E. M. & Lemmon, A. R. High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* **44**, 99-121 (2013).

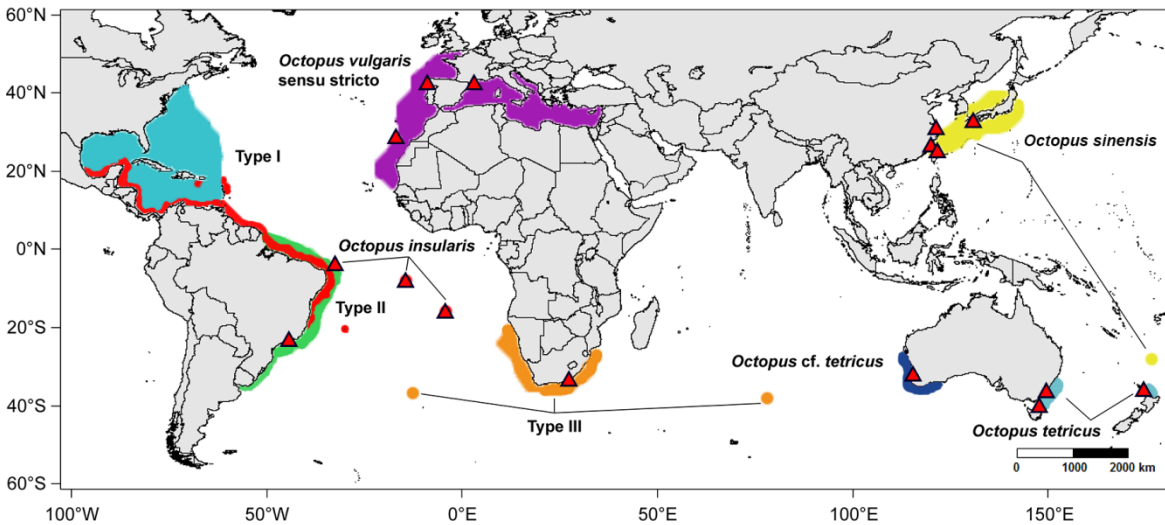
- 18 Ree, R. H. & Hipp, A. L. Inferring phylogenetic history from restriction site associated DNA (RADseq). *Next-Generation Sequencing in Plant Systematics*. (Eds E Hörandl, MS Appelhans), 181-204 (2015).
- 19 Wagner, C. E. *et al.* Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Mol. Ecol.* **22**, 787-798 (2013).
- 20 Jones, J. C., Fan, S., Franchini, P., Scharl, M. & Meyer, A. The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Mol. Ecol.* **22**, 2986-3001 (2013).
- 21 Bickford, D. *et al.* Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* **22**, 148-155 (2007).
- 22 Fisheries and Aquaculture Statistics 2010. *FAO Yearbook*. Food and Agriculture Organization (FAO) of the United Nations. Rome (2012).
- 23 Robson, G. C. *A Monograph of the Recent Cephalopoda, Part I: Octopodinae.*, British Museum of Natural History (1929).
- 24 Söller, R., Warnke, K., Saint-Paul, U. & Blohm, D. Sequence divergence of mitochondrial DNA indicates cryptic biodiversity in *Octopus vulgaris* and supports the taxonomic distinctiveness of *Octopus mimus* (Cephalopoda : Octopodidae). *Marine Biology* **136**, 29-35 (2000).
- 25 Warnke, K., Söller, R., Blohm, D. & Saint-Paul, U. A new look at geographic and phylogenetic relationships within the species group surrounding *Octopus vulgaris* (Mollusca, Cephalopoda): indications of very wide distribution from mitochondrial DNA sequences. *J. Zool. Syst. Evol. Res.* **42**, 306-312 (2004).
- 26 Leite, T. S., Haimovici, M., Molina, W. & Warnke, K. Morphological and genetic description of *Octopus insularis*, a new cryptic species in the *Octopus vulgaris* complex (Cephalopoda : Octopodidae) from the tropical South-Western Atlantic. *Journal of Molluscan Studies* **74**, 63-74 (2008).
- 27 Guerra, A. *et al.* Morphological and genetic evidence that *Octopus vulgaris* Cuvier, 1797 inhabits Amsterdam and Saint Paul Islands (Southern Indian Ocean). *Ices Journal of Marine Science* **67**, 1401-1407 (2010).
- 28 Acosta-Jofré, M. S., Sahade, R., Laudien, J. & Chiappero, M. B. A contribution to the understanding of phylogenetic relationships among species of the genus *Octopus* (Octopodidae: Cephalopoda). *Scientia Marina* **76**, 311-318 (2012).
- 29 Amor, M. D., Laptikhovskiy, V., Norman, M. D. & Strugnell, J. M. Genetic evidence extends the known distribution of *Octopus insularis* to the mid-Atlantic islands Ascension and St Helena. *Journal of the Marine Biological Association of the United Kingdom*, **97** (4), 753-758 (2017).
- 30 Amor, M. D., Norman, M. D., Cameron, H. E. & Strugnell, J. M. Allopatric speciation within a cryptic species complex of Australasian octopuses. *PLOS ONE* **9**, e98982. doi:98910.91371/journal.pone.0098982 (2014).
- 31 Reid, A. L. & Wilson, N. G. Octopuses of the Kermadec Islands: Discovery and description of a new member of the *Octopus 'vulgaris'* complex (*O. jollyorum*, sp. nov.) and the first description of a male *Callistoctopus kermadecensis* (Berry, 1914). *Bulletin of the Auckland Museum* **20**, 349-368 (2015).

- 32 Amor, M. D. *et al.* Morphological assessment of the *Octopus vulgaris* species complex evaluated in light of molecular-based phylogenetic inferences. *Zoologica Scripta* **46**, 275-288 (2017).
- 33 Norman, M. D., Finn, J. K. & Hochberg, F. G. in *Cephalopods of the world. An annotated and illustrated catalogue of cephalopod species known to date. Octopods and Vampire Squids. FAO Species Catalogue for Fishery Purposes. No. 4. Vol. 3* (eds P. Jereb, C. F. E. Roper, M. D. Norman, & J. K. Finn) 36-215 (Food and Agriculture Organization of the United Nations, 2014).
- 34 Norman, M. D. & Finn, J. K. in *Cephalopods of the World. An Annotated and Illustrated Catalogue of Cephalopod Species Known to Date. Octopods and Vampire Squids Vol. 3* (eds P. Jereb, C. F. E. Roper, M. D. Norman, & J. K. Finn) 9-21 (Food and Agriculture Organization of the United Nations, 2014).
- 35 Albertin, C. B., Simakov, O., Mitros, T., Wang, Z. Y., Pungor, J. R., Edsinger-Gonzales, E., Brenner, S.R., Ragsdale, C. W. & Rokhsar, D. S. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature*, **524** (7564), 220 (2015).
- 36 Hipp, A. L. *et al.* A framework phylogeny of the American oak clade based on sequenced RAD data. *PLoS ONE* **9**, e93975 (2014).
- 37 Lima, F. D., Berbel-Filho, W. M., Leite, T. S., Rosas, C. & Lima, S. M. Q. Occurrence of *Octopus insularis* Leite and Haimovici, 2008 in the tropical northwestern Atlantic and implications of species misidentification to octopus fisheries management. *Mar Biodiv* **47**, 723-734 (2017).
- 38 González-Gómez, R. *et al.* An integrative taxonomic approach reveals *Octopus insularis* as the dominant species in the Veracruz Reef System (southwestern Gulf of Mexico). *PeerJ*, **6** (2018).
- 39 Rosas-Luis, R., Jiménez Badillo, M. D. L., Montoliu-Elena, L., & Morillo-Velarde, P. S. Food and feeding habits of *Octopus insularis* in the Veracruz Reef System National Park and confirmation of its presence in the southwest Gulf of Mexico. *Marine Ecology* (2019).
- 40 Teske, P. R., Oosthuizen, A., Papadopoulos, I. & Barker, N. P. Phylogeographic structure of *Octopus vulgaris* in South Africa revisited: identification of a second lineage near Durban harbour. *Marine Biology* **151**, 2119-2122 (2007).
- 41 Sales, J. B. D. L. *et al.* Phylogeographical features of *Octopus vulgaris* and *Octopus insularis* in the south-eastern Atlantic based on the analysis of mitochondrial markers. *Journal of Shellfish Research* **32**, 325-339 (2013).
- 42 Souza, C. A. *et al.* Efficiency of ddRAD target enriched sequencing across spiny rock lobster species (Palinuridae: *Jasus*). *Scientific Reports* **7**, 6781 (2017).
- 43 Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686-693 (2001).
- 44 Sostdian, S. & Rosenthal, Y. Deep-sea temperature and ice volume changes across the Pliocene-Pleistocene climate transitions. *Science* **325**, 306-310 (2009).
- 45 Haimovici, M. & Perez, J. A. Coastal cephalopod fauna of southern Brazil. *Bulletin of Marine Science* **49**, 221-230 (1991).

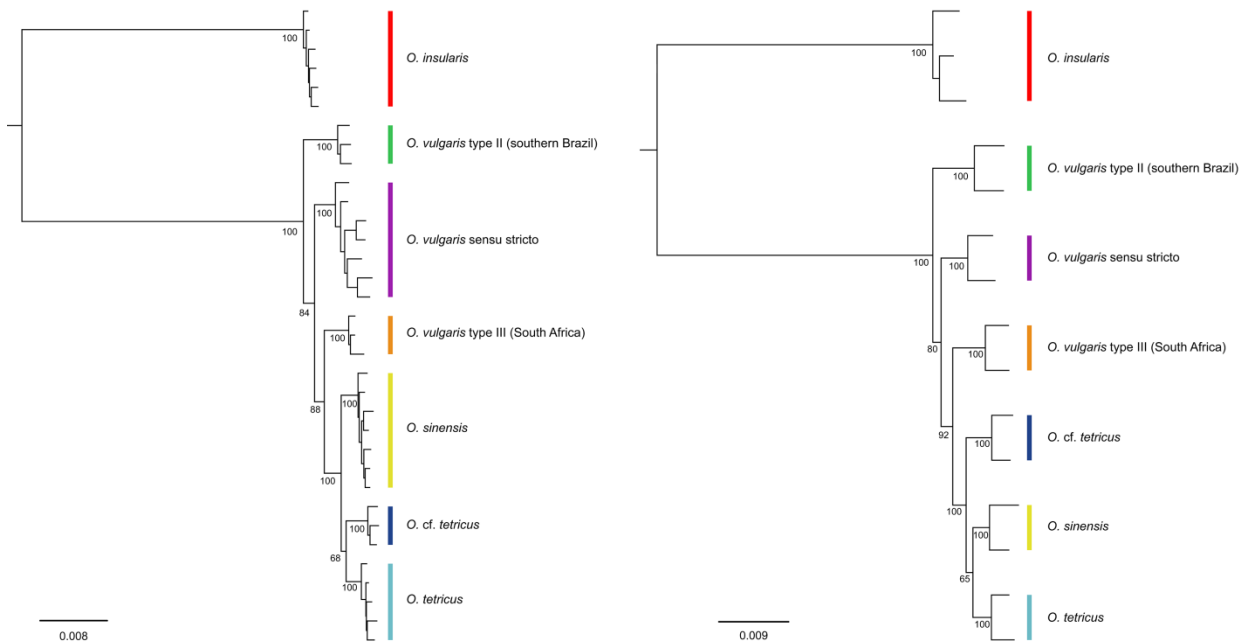
- 46 Haywood, A. M. & Valdes, P. J. Modelling Pliocene warmth: contribution of atmosphere, oceans and cryosphere. *Earth and Planetary Science Letters* **218**, 363-377 (2004).
- 47 Morley, J. J. & Dwoletzky, B. A. Evolving Pliocene-Pleistocene climate: A North Pacific perspective. *Quaternary Science Reviews* **10**, 225-237 (1991).
- 48 Goetze, E., Andrews, K. R., Peijnenburg, K. T., Portner, E. & Norton, E. L. Temporal stability of genetic structure in a mesopelagic copepod. *PLoS one* **10**, e0136087 (2015).
- 49 Roura, Á. *et al.* Life strategies of cephalopod paralarvae in a coastal upwelling system (NW Iberian Peninsula): insights from zooplankton community and spatio-temporal analyses. *Fisheries Oceanography* **25**, 241-258 (2016).
- 50 Leite, T. S., Haimovici, M., Mather, J. & Oliveira, J. E. L. Habitat, distribution, and abundance of the commercial octopus (*Octopus insularis*) in a tropical oceanic island, Brazil: Information for management of an artisanal fishery inside a marine protected area. *Fisheries Research* **98**, 85-91 (2009).
- 51 García-Fernández, P., García-Souto, D., Almansa, E., Morán, P., & Gestal, C. Epigenetic DNA methylation mediating *Octopus vulgaris* early development: effect of essential fatty acids enriched diet. *Frontiers in physiology*, **8** 292 (2017).
- 52 Díaz-Freije, E., Gestal, C., Castellanos-Martínez, S., & Morán, P. The role of DNA methylation on *Octopus vulgaris* development and their perspectives. *Frontiers in physiology*, **5**, 62 (2014).
- 53 Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614-620 (2014).
- 54 Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124-3140 (2013).
- 55 Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* **15**, R46-R46 (2014).
- 56 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359 (2012).
- 57 Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993 (2011).
- 58 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797 (2004).
- 59 Posada, D. jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256 (2008).
- 60 Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716-723 (1974).
- 61 Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
- 62 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574 (2003).
- 63 Rambaut, A., A., S. M., Xie, D. & Drummond, A. J. Tracer v1.6, Available from <http://beast.bio.ed.ac.uk/Tracer>. (2014).
- 64 Eaton, D. A. R. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* **30**, 1844-1849 (2014).



- 65 Li H. and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, **25** (2009)
- 66 Hipp, A. L. RADami: R package for phylogenetic analysis of RADseq data. R package version 1.0-3. url: <http://CRAN.R-project.org/package=RADami/> (2014).
- 67 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/> (2015).
- 68 Foll, M., & Gaggiotti, O. E. A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*. (2008)
- 69 Shimodaira, H. An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology* **51**, 492-508 (2002).
- 70 Shimodaira, H. & Hasegawa, M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**, 1246-1247 (2001).
- 71 Chifman, J. & Kubatko, L. Quartet inference from SNP data under the coalescent model. *Bioinformatics* **30**, 3317-3324 (2014).
- 72 Swofford, D. L. PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts. (2003).
- 73 Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649 (2012).
- 74 Bouckaert, R. *et al.* BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput Biol* **10**, e1003537 (2014).
- 75 Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88 (2006).
- 76 Aldous, D. J. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Statistical Science*, 23-34 (2001).
- 77 Ho, S. Y. W. & Phillips, M. J. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology* **58**, 367-380 (2009).
- 78 Winkelmann, I. *et al.* Mitochondrial genome diversity and population structure of the giant squid *Architeuthis*: genetics sheds new light on one of the most enigmatic marine species. *Proceedings of the Royal Society B-Biological Sciences* **280** (2013).
- 79 Marko, P. B. Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. *Molecular Biology and Evolution* **19**, 2005-2021 (2002).
- 80 Becker, R. A. & Wilks, A. R. maps: Draw Geographical Maps. R package version 3.1.1. R version by Ray Brownrigg. Enhancements by Thomas P Minka and Alex Deckmyn. <https://CRAN.R-project.org/package=maps> (2016).
- 81 Becker, R. A. & Wilks, A. R. mapdata: Extra Map Databases. R package version 2.2-6. R version by Ray Brownrigg. <https://CRAN.R-project.org/package=mapdata> (2016).

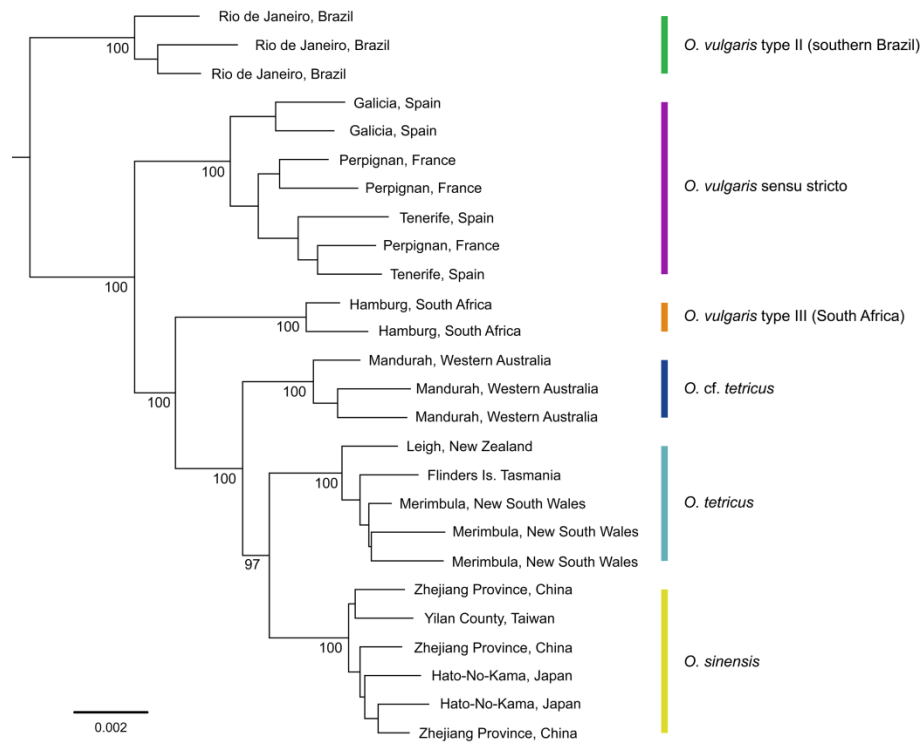


**Figure 1.** Sampling locations (red triangles) of *Octopus vulgaris* species group [*O. vulgaris sensu stricto* (Mediterranean and eastern North Atlantic), *O. vulgaris* Type II (southern Brazil), *O. vulgaris* Type III (South Africa), *O. tetricus* (east Australia), *O. cf. tetricus* (west Australia), *O. sinensis* (Asia)] and *O. insularis* individuals included in the present study. Distributions of *O. vulgaris sensu stricto* and species 'Types' are shaded as per Norman *et al.*<sup>33</sup>. Purple, *O. vulgaris* s. s.; blue, Type I (not sampled in our study - Caribbean/ Gulf of Mexico); green, Type II (Brazil); and orange, Type III (South Africa). Distributions of non-*O. vulgaris* species are shaded in yellow: *O. sinensis*, dark blue: *O. cf. tetricus*, light blue: *O. tetricus* and red: *O. insularis*. This map was generated using the maps v3.1.1<sup>80</sup> and mapdata v2.2-6<sup>81</sup> packages in R v3.1.1<sup>67</sup>.

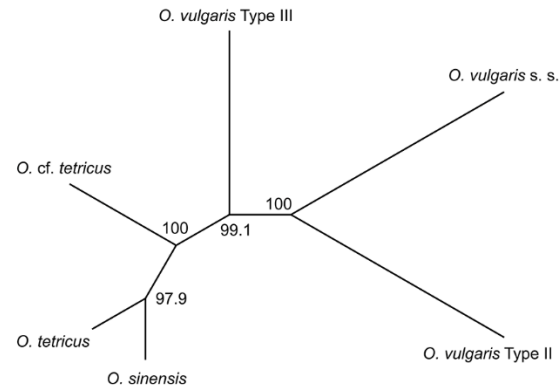


**Figure 2.** Maximum likelihood phylogenies depicting the relationships among members of the *Octopus vulgaris* species complex, *O. insularis*, *O. tetricus* and *O. cf. tetricus*. The phylogeny on the left is based on all 34 individuals sequenced and included 298 loci (RAD1 dataset). The phylogeny on the right is based on the 15 individuals, with 2-3 representatives with the highest number of reads sequenced from each species included. This phylogeny included 1060 loci (RAD02 dataset). Both analyses were conducted using the GTR+G evolutionary model in RAxML v8.0.19 (Stamatakis 2014). ML bootstrap values are displayed below major nodes. Differences among topologies include support at major nodes and the placement/relationships among clades including individuals representing *O. sinensis*, *O. tetricus* and *O. cf. tetricus* (although these clades are supported as being distinct (BS = 100). Colours correspond to Fig. 1.

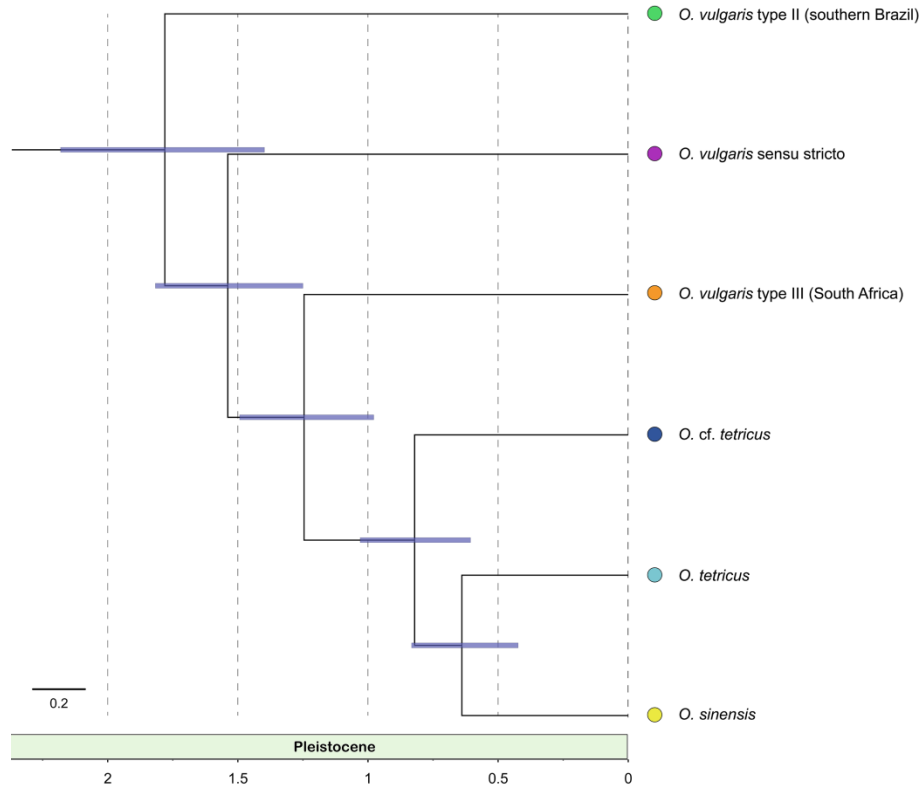




**Figure 3.** Maximum likelihood phylogeny depicting the relationships among members of the *Octopus vulgaris* species group. Analysis was conducted using 604 concatenated double digest RADseq loci using the GTR+G evolutionary model in RAxML v8.0.19<sup>57</sup>. ML bootstrap values are displayed below major nodes. Tip labels correspond to sampling region in Table 2. Colours correspond to Fig. 1.



**Figure 4.** Species tree estimation supporting six species within the *Octopus vulgaris* species group. Analysis was based on 604 concatenated RAD loci using the *SVDQuartets* algorithm<sup>69</sup> implemented in *PAUP v4.0a146 for Unix/Linux*<sup>70</sup>.



**Figure 5.** Divergence time estimation within the *Octopus vulgaris* group. The topology presented is based on combining three independent runs that were all calculated using the rate of evolution for cephalopods (0.0036 neutral substitutions per million years). Node bars represent 95% highest posterior densities for divergence estimates and estimate that the *O. vulgaris* species group evolved within the last 2.5 million years. Colours correspond to Fig. 1.

**Table 1.** Approximate Unbiased (AU) test results comparing our phylogenetic topology (Fig. 3) with contrasting results from previous studies. The Bayesian Posterior Probability (PP) calculated under the Bayesian Information Criterion (BIC) is also shown. Topologies are rejected with a p-value of <0.05.

Tree rank	Constraint	p-value	PP	Reference
1	No constraint	0.951	1	
2	<i>O. sinensis</i> as sister taxon to <i>O. tetricus</i> and <i>O. cf. tetricus</i>	0.051	1E-025	Amor <i>et al.</i> <sup>30</sup>
3	Forced monophyly of <i>O. vulgaris</i> s. s. and <i>O. vulgaris</i> Type III	0.004	1E-045	Amor <i>et al.</i> <sup>29</sup>

**Table 2.** Sampling location of *Octopus vulgaris* species group, and *O. insularis* individuals included in the present study.

Species	Location	n	Latitude	Longitude
<i>O. vulgaris</i> s. s.	Perpignan, France	3	42.48352	3.13145
<i>O. vulgaris</i> s. s.	Galicia, Spain	2	42.22719	-8.89403
<i>O. vulgaris</i> s. s.	Tenerife, Spain	2	28.29156	-16.62913
<i>O. vulgaris</i> Type II	Rio de Janeiro, Brazil	3	-23.15000	-44.23330
<i>O. vulgaris</i> Type III	Hamburg, South Africa	3	-33.29446	27.48093
<i>O. sinensis</i>	Zhejiang Province, China	2	30.71616	121.36597
<i>O. sinensis</i>	Zhejiang Province, China	1	26.17578	119.63425
<i>O. sinensis</i>	Yilan County, Taiwan	3	24.94063	121.90000
<i>O. sinensis</i>	Hato-No-Kama, Japan	3	32.60504	130.41029
<i>O. tetricus</i>	Merimbula, Australia	3	-36.89175	149.91045
<i>O. tetricus</i>	Flinders Is, Australia	1	-39.98364	148.05269
<i>O. tetricus</i>	Leigh, New Zealand	1	-36.29165	174.80965
<i>O. cf. tetricus</i>	Mandurah, Australia	3	-32.12540	115.75856
<i>O. insularis</i>	Natal, Brazil	3	-3.85381	-32.42379
<i>O. insularis</i>	Ascension Island	2	-7.94672	-14.35592
<i>O. insularis</i>	Georgetown, St Helena	1	-15.96501	-5.70892