Molecular Phylogenetics and Evolution 107 (2017) 415–430

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

# Restructuring higher taxonomy using broad-scale phylogenomics: The living Ophiuroidea

Timothy D. O'Hara [a,*], Andrew F. Hugall [a,*], Ben Thuy [b], Sabine Stöhr [c], Alexander V. Martynov [d]

[a] Marine Invertebrates, Sciences, Museum Victoria, GPO Box 666, Melbourne, Victoria 3000, Australia
[b] Section Paléontologie, Musée National d'Histoire Naturelle du Luxembourg, 24 Rue Münster, 2160, Luxembourg
[c] Swedish Museum of Natural History, Department of Zoology, Box 50007, 104 05 Stockholm, Sweden
[d] Zoological Museum, Moscow State University, Bolshaya Nikitskaya Str. 6, Moscow 125009, Russia

ARTICLE INFO

*Article history:*
Received 2 July 2016
Revised 5 December 2016
Accepted 7 December 2016
Available online 8 December 2016

*Keywords:*
Phylogenomics
Exon capture
Temporal-banding
Ophiuroidea
Invertebrates
Marine
Deep sea

ABSTRACT

The power and throughput of next-generation sequencing is instigating a major transformation in our understanding of evolution and classification of life on our planet. The new trees of life are robust and comprehensive. Here we provide a landmark phylogeny of the living ophiuroids and use it as the basis for a major revision of the higher classification of this class of marine invertebrates. We used an exon-capture system to generate a 1484 exon (273 kbp) data-matrix from DNA extracted from ethanol-preserved museum samples. We successfully obtained an average of 90% of our target sequence from 576 species spread across the known taxonomic diversity. The topology of the major lineages was robust to taxon sampling, exon-sampling, models and methods. However, estimates of node age were much less precise, varying by about a quarter of mean age. We used a combination of phylogenetic distinctiveness and temporal-banding to guide our revision of the family-level classification. Empirically, we determined that limiting family crown age to 110 ± 10 Ma (mid Cretaceous) selected phylogenetically distinct nodes while minimising disruption to the existing taxonomy. The resulting scheme of 32 families and six orders considerably expands the number of higher taxa. The families are generally longitudinally widespread across the world's oceans, although 17 are largely confined to temperate and equatorial latitudes and six to relatively shallow water (less than 1000 m depth).



## 1. Introduction

The Ophiuroidea are diverse and abundant in modern seas, from the equator to polar regions, from coastal shores to the hadal trenches (Stöhr et al., 2012). They can dominate seafloor assemblages (Gage and Tyler, 1991), playing an important role in enriching the benthic ecosystem by transferring primary production into the sediment (Josefson and Conley, 1997). They are the most speciose class of echinoderms, with ~2100 described species (Stöhr et al., 2016), and are one of the only 22 living classes of deuterostomes. The known species richness is almost certainly a large underestimate, as phylogeographic studies have identified suites of cryptic lineages for many morpho-species examined (O'Hara et al., 2004, 2014a; Stöhr et al., 2009; Hoareau et al., 2013; Naughton et al., 2014). Although complete articulated fossils are rare, there are abundant isolated ophiuroid ossicles in palaeosediments, some of which are distinctive enough to inform phylogenetic (O'Hara et al., 2014b) and palaeoecological studies (Thuy et al., 2012). However, there has been insufficient effort to integrate the majority of Mesozoic or Cenozoic fossils into a modern phylogenetic scheme.

Despite their evolutionary and ecological importance, there have been comparatively few morphological or molecular phylogenetic studies focusing on ophiuroids. Ljungman (1867) defined many of the existing family-level taxa and Matsumoto (1915) erected the first modern classification for the group, qualitatively constructing a 'tree' of ophiuroid families (Matsumoto, 1917). This scheme largely persisted until it was modified for the Treatise on Invertebrate Paleontology (Spencer and Wright, 1966). The only class-wide cladistic study was undertaken by Smith et al. (1995) who compiled and analysed a subfamily-level morphological dataset, implicitly assuming the monophyly of the subfamilies. The inclusion of some ophiuroid samples in echinoderm-wide molecular phylogenies did little to inform classification, except to show that the aberrant *Ophiocanops* was not a representative of an otherwise extinct Palaeozoic group (Janies, 2001; Janies et al.,

* Corresponding author at: Museum Victoria, GPO Box 666, Melbourne, Victoria 3000, Australia.
*E-mail addresses:* tohara@museum.vic.gov.au (T.D. O'Hara), ahugall@museum.vic.gov.au (A.F. Hugall).

http://dx.doi.org/10.1016/j.ympev.2016.12.006
1055-7903/

2011) as had been previously hypothesised (Fell, 1963). A recent class-wide phylogeny of 39 species based on ribosomal (16S, 18S) genes showed several traditional families to be polyphyletic but failed to resolve most basal nodes (Hunter et al., 2016). The pioneering phylogenomic study of O'Hara et al. (2014b) constructed a robust phylogeny of the class for the first time, using a dataset of 425 genes (306 kbp) derived from 52 de novo ophiuroid transcriptomes. This phylogeny differed from all historical classification schemes and the existing higher-level (genus and above) taxa were shown in many cases to be poly- or paraphyletic. Clearly many of the characters traditionally used to differentiate these taxa were plesiomorphic, convergent or not informative. On the other hand, there was a great deal of similarity between the topology and emerging morphological analyses based on the microstructure of modern and fossil ophiuroid ossicles, particularly the morphology of the lateral arm plates including the arm spine articulation surface (Martynov, 2010a; Thuy and Stöhr, 2011). The latest morphological phylogeny (Thuy and Stöhr, 2016) also showed great congruence with O'Hara et al. (2014b) and suggested synapomorphies for all major branches. However, O'Hara et al. (2014b) refrained from naming the new higher-level nodes, citing the need for more comprehensive taxon sampling.

Our goal here is to present a comprehensive family-level phylogeny of the extant Class Ophiuroidea using our next-generation sequence-capture methodology (Hugall et al., 2016) to hybridise, capture and map 1552 exons (285 kb) from DNA extracted from ethanol-preserved museum samples. We used the transcriptome data of O'Hara et al. (2014b) as the basis from which to design sets of consensus probes that could capture the target exons from across the genetic diversity of the group.

The establishment of higher level taxa has generally resulted from either (1) traditional taxonomic revisions that identify groups based on sets of shared morphological characters (similarity by resemblance), or (2) the naming of significant nodes on a phylogenetic tree (similarity through ancestry) (Avise and Johns, 1999). However, established higher ranks of many groups in the Linnaean system are known to be highly variable with respect to time of origin, phenotypic divergence, number of contained species or any other intrinsic biological variable (Avise and Liu, 2011). Taxonomic ranks are not quantitatively equivalent between (or even within) higher groups and it has been argued that they should not be used in cross-taxa comparisons (Pleijel and Rouse, 2003; Avise and Liu, 2011). One response has been to develop the PhyloCode (Cantino and de Queiroz, 2010), an alternative system of nomenclature that dispenses with ranks, restricting taxonomic names to nodes on a phylogenetic tree. However, the traditional Linnaean nomenclatural system has proved to be resilient and continues to be widely used (Avise and Liu, 2011).

Consequently, we partially adopted the 'temporal banding' approach (Hennig, 1966; Avise and Johns, 1999; Holt and Jønsson, 2014), which names nodes on a phylogenetic tree but includes age as part of the rationale for grouping nodes into Linnaean "ranks" of classification. The original Hennig (1966) and Avise and Johns (1999) approaches are often considered too disruptive to existing widely-accepted classifications to be attempted (Avise and Liu, 2011). Instead we used elevated phylogenetic distinctness (similar to the 'discrete phylogenetic clustering' of Humphreys and Barraclough, 2014) as an empirical method of determining the age of family-level taxa. The existing classification of the Ophiuroidea is not congruent with recent phylogenetic analyses (O'Hara et al., 2014b; Hugall et al., 2016; Thuy and Stöhr, 2016) and needs to be replaced. Given the scale and strength of our dataset (416 genes, 273 kb, 576 taxa, 90% data complete), we took the opportunity in this paper to authoritatively establish a consistent family to order level hierarchy based on the topology and node age in our molecular phylogeny.

## 2. Materials and methods

### 2.1. Phylogenomic data

Supplementary material, large tables and trees (figures and tables with prefix S) are to be found in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.rb334 and http://dx.doi.org/10.5061/dryad.db339. The phylogenomic dataset used here was built on the exon-capture system described in Hugall et al. (2016), which in turn was derived from transcriptome data described in O'Hara et al. (2014b). Briefly, the ophiuroid transcriptome data supplied a 425 gene dataset, which was refined into a 418 gene, 1552 exon 285 kb exon-capture target.

For this paper we had a total of 645 exon-capture and 52 transcriptome samples, from which we chose 540 exon-capture and 36 transcriptomes to best represent 576 taxa (nominal species), listed in Table S1. This is an extension of Hugall et al. (2016), with additional data (196 species with 46.7 Mb) generated by the same exon-capture procedure: Illumina TruSeq library, microarray in-solution RNA target enrichment, and Illumina MiSeq 150 base pair-end sequencing. For this paper, all exon-capture samples were mapped using the Trinity Assembled Super-Reference method (TASR; scripts TASSER and TASSMAP, as described in Hugall et al. (2016) and http://dx.doi.org/10.5061/dryad.db339) with a minimum coverage limit of 5, followed by global deletion of 68 poorly-recovered or paralogous exons and per-sample exclusion of exons with excess polymorphism (script HEXER), to give a total aligned in-frame data-matrix of 140 million bases arranged as 272,952 sites in 1484 exons of 416 genes.

We also obtained approximately 1.4 kb of the mitochondrial gene COI from the majority of transcriptomic and target-capture samples (NCBI Nucleotide sequences KU894924 - KU895455). This was used to validate the taxonomic identity of sequenced samples and to identify contaminants, through comparison with available 'barcode-of-life' sequences (Hugall et al., 2016). However, these fast-evolving sequences were not included in the phylogenetic analyses.

Since one of the aims of our program was to integrate phylogenetic data into biogeographic analyses of the greater Australian region (O'Hara et al., 2011), our selection of samples was biased towards waters around Australia, New Zealand, Antarctica and the south-western Pacific Ocean (110°E to 170°W, 70°S to 0°S). However, we also opportunistically included samples from all ocean basins in order to boost phylogenetic representation (Table 1). The majority of samples were derived from the collection at Museum Victoria (Melbourne, MV). Other samples were obtained (in order of importance) from the Muséum National D'Histoire Naturelle, (Paris, MNHN), University of Florida (Gainesville, UF), National Institute of Water and Atmospheric Research (Wellington, NIWA), California Academy of Science (San Francisco, CAS), Western Australian Museum (Perth, WAM), Australian Museum (Sydney, AM), the South Australian Museum (Adelaide, SAM), the Swedish Museum of Natural History (Stockholm, SMNH) and the Natural History Museum (London, BMNH). The samples were collected largely from habitats on the continental shelf and upper slope (0–2000 m) with only a few samples from the lower slope and abyss, the latter predominantly collected from north-east of New Zealand and from around the Scotia Trench. Tissue samples were up to 15 years old.

### 2.2. Phylogenetic analyses

We conducted phylogenetic analyses using two sets of samples: (1) 576 taxa, and (2) a subset of 185 taxa used for computationally intensive analyses, selected to represent all major lineages while