Review of clypeasteroids phylogeny and a case study of

*Sinaechinocyamus mai* (Taiwanasteridae)

Lo-Yu Chang\(^1\,\,^2\), Shyh-Jye Lee\(^3\), Mong-Hsun Tsai\(^4\) and Jih-Pai Lin\(^2\)*

\(^1\)College of Medicine, National Taiwan University, No. 1, Sec. 1, Ren-Ai Road, Taipei 100, Taiwan

\(^2\)Department of Geosciences, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 106, Taiwan

\(^3\)Department of Life Science, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 106, Taiwan

\(^4\)Institute of Biotechnology, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 106, Taiwan

This manuscript is submitted to “Terrestrial, Atmospheric and Oceanic Sciences”

Running title: A review of clypeasteroids phylogeny

Three Key Points:

- Breakthrough in sequencing technology provides insights for paleontology.
- *Sinaechinocyamus mai* yields precious data given by sequencing technologies.
- A new transcriptome-based phylogeny of clypeasteroids is reconstructed.

* E-mail addresses: alexjplin@ntu.edu.tw, xyloplax1@gmail.com

Phone: +886-2-3366-2928

This is a PDF file of an unedited manuscript which has been accepted for publication. But the manuscript will undergo copyediting, typesetting, pagination, and proofreading process that might lead to differences between this version and the final version of publication.
ABSTRACT

Paleontology in the 21st century is a field with high interdisciplinary cooperation. Aside from the traditional morphology analysis, modern paleontologists and taxonomists can also leverage novel technical breakthroughs in molecular biology to peek into the evolutionary past. Irregular echinoids, which comprise the extant members of heart urchins (Atelostomata), sand dollars (Scutelloida), and sea biscuits (Clypeasteroida), is a subgroup of echinoids exhibiting a secondary bilateral symmetry body plan. This clade's evolution history and development remain a realm of intensive investigation that benefits profoundly from the progress of Next Generation Sequencing (NGS) technology. In this study, we reviewed the methodology and theoretical basis of molecular evolution and conducted a case study on the Taiwanese endemic sand dollar species *Sinaechinocyamus mai*. We exemplified the value of sequence analysis in phylogeny study and the prospect of incorporating molecular data in future investigations to shed light on the enigmatic phylogenetic relationship and body plan establishment of Scutelloida and Clypeasteroida.

Key words: Paleobiology, Scutelloida, Clypeasteroida, molecular evolution, monophyly
1. INTRODUCTION

1.1 Sequencing Technology

The development of evolutionary theory and sequencing technology are coupled with each other. Since 2005, next-generation sequencing (NGS) is one of the most significant technological breakthroughs in the field of molecular biology (Goodwin et al., 2016). With the aid of NGS, the availability of non-model species DNA sequences significantly increases. Therefore, the new field “phylogenomic,” using the genomic scale sequence data to reconstruct the phylogenetic relationship between species, has emerged and improved both the depth (more evolutionary distant clades, such as the divergence between kingdoms) and resolution (able to resolve the relationship between closely related groups) of the tree of life (Chan and Ragan, 2013).

In spite of the flourish of NGS technology, the traditional gene markers used in molecular evolution do not entirely step down from the stage. Due to various reasons (i.e., cost, difficulties in data analysis, sample quality), single gene markers remain useful in small-scale analyses. The massive quality of data produced by whole-genome sequencing provides detailed information but is cumbersome in analysis (Cannon and Kocot, 2016). One way to do so is by only sequencing the coding region of the genome. As shown in Table 1, transcriptome analysis represents a balance between the phylogeny resolution whole-genome sequencing (WGS) and the cost and convenient advantage of gene markers (Cannon and Kocot, 2016; Young and Gillung, 2019).

Modern biologists can facilitate paleontology studies in various ways. Through experimental models of modern species, developmental biologists can answer how certain structures possibly formed in fossil species and the potential gene regulatory networks responsible for the evolutionary origin and development of these characters (Thompson et al., 2017). The relationship among fossil species with extant representatives in their corresponding clades can be supported by reconstructing the phylogeny based on sequencing data of extant species. Furthermore, sequencing data is also able to provide evidence of divergence time and correlate with geological events (Coppard and Lessios, 2017). Analyzing the genetic material inside fossil specimens is a field under active development as well (Paabo et al., 2004). Under adequate conditions, ancient DNA (aDNA) can preserve up to more than one million years (van der Valk et al., 2021). Through direct sequencing of DNA of ancient species, researchers can be more certain about the population genetic dynamic and the relationship between fossil and extant species (Orlando and Cooper, 2014). Furthermore, the genetic basis provides an independent dataset for phylogeny inference other than traditional phenotype comparison alone (Orlando and Cooper, 2014). Therefore, the advent of sequencing technology has revolutionized paleontology by adding a new dimension to the study material.
1.2 Phylogeny of Irregular Echinoids

Echinoidea, commonly known as sea urchins, forms a conspicuous and essential element of many marine benthic communities. They exploit a wide array of marine habitats and contribute the greatest levels of biodiversity in shallow shelf areas. Arisen since the Middle Ordovician in 460 million years ago (Mya), there are more than 1000 living species in the clade of Echinoidea (Kroh, 2020; Kroh and Smith, 2010). The robust endoskeleton structure, known as stereom (Lin et al., In press), which is composed of multi-plated magnesium calcite, allows echinoids to leave a relatively rich fossil record: in total, there are about 10000 fossil species (Kroh, 2020). From paleontology to developmental biology, the extant species and the rich fossil record make the echinoids excellent candidates as valuable study models.

Among the clades in Echinoidea, Irregularia, including species are commonly known as sand dollars (Scutelloida) and sea biscuits (Clypeasteroida), distinguish themselves from irregular echinoids through the unique body plan (Saucède et al., 2007). One of the characteristic features of the Echinodermata phylum is the pentameric symmetry body plan. The larvae of echinoderms are bilaterally symmetrical, while upon metamorphosis, the larval structures typically are lost and yield a pentameric juvenile (Peterson et al., 2000). However, irregular echinoids share a secondary adult bilateral symmetry body plan (Fig. 1), and the establishment of body axis is still largely unknown and intriguing. Furthermore, the phylogeny study regarding the relationship among irregular echinoids suffers from the discrepancy between morphology and molecular evidence (Kroh and Smith, 2010; Mongiardino Koch et al., 2018). Due to the fact that the taxonomy of Irregularia is still a topic under active discussion, in this article, the definition of Clypeasteroida is used in congruence with Kroh (2020) and Mongiardino Koch et al. (2020) to prevent ambiguity, which does not include the traditional suborder Scutellina. While the term “clypeasteroids” is used to denote the traditional clade consists of Clypeasterina and Scutellina.

Based on the recent total-evidence dated analysis, clypeasteroids possibly originated during the Cretaceous (145-66 Mya) (Mongiardino Koch and Thompson, 2020). By the Middle Eocene (49-37 million years ago), this group already presented in the fossil record worldwide (Pawson, 2007). The principal characters in echinoid fossil morphological analysis are the structure of test, spine, and Aristotle’s lantern (Kroh and Smith, 2010; Ziegler et al., 2015). In particular, the lantern structure of clypeasteroids is highly modified compared to other echinoids, and this modification has been hypothesized to be crucial for the epifaunal inhabitant of clypeasteroids (Mooi, 1990a). On the contrary of abundant morphological studies, molecular efforts have lagged on echinoids phylogeny studies. Early attempts to use molecular data to resolve the deep-sea urchin phylogeny relied on just a few single-gene markers, usually ribosomal RNA-coding ones (Smith et al., 2006). The lack of broad sampling of loci across the genome limits the robustness of these phylogenetic analyses due to the genes used might be the outliers of incomplete lineage sorting. When an ancestral species undergoes
rapid speciation in succession, ancestral polymorphisms may not be fully resolved into different monophyletic lineages, and the individual gene trees may be incongruent with the species tree (Galtier and Daubin, 2008). Other types of biases can also lead to inaccurate phylogenetic topology even in the absence of incomplete lineage sorting. For instance, species exhibiting significantly higher genetic distances to other clades tend to be erroneously clustered together. This phenomenon, known as long-branch attraction, is commonplace in molecular evolution inference, and it may arise from the heterogeneity of substitution rates, inadequate taxon sampling, and base compositional heterogeneity (Qu et al., 2017; Susko and Roger, 2021).

Therefore, although numerous studies have tried to investigate the phylogeny of Irregularia, the relationships among sea biscuits, sand dollars, and other irregular echinoids remain not well understood due to the contradicting results yielded by molecular and morphological data (Kroh and Smith, 2010; Mongiardino Koch et al., 2018). For example, most morphological phylogenies strongly supported the monophyly of clypeasteroids and their origin from a paraphyletic sister group collectively known as “cassiduloids”, including extant clades of Echinolampadoida Kroh & Smith, 2010, Cassiduloida Claus, 1880, and other extinct lineages (Kroh and Smith, 2010). On the other hand, the transcriptomic analyses refute the monophyly of Clypeasteroida (Mongiardino Koch et al., 2018). This mismatch remains somewhat unresolved, and further investigation is needed to answer this perplexing inquest.

1.3 Body Plan

One of the most conspicuous characteristics of extant echinoids is their pentameral symmetry in the adult stage. Based on the observation of extant echinoderms and the fossil record, it seems the early echinoderms were bilateral in both the larval and adult life stages while the extant species developed secondary pentameral symmetry in adulthood (Peterson et al., 2000).

Among many essential genes controlling development, one crucial important player is the homeobox genes, also known as Hox genes. They are a family of transcription factors regulating body plan and axis establishment in embryo development (Lacalli, 2014). A primary recognition criterion for these genes is a sequence of an 180 nucleotides, highly conserved “homeobox,” which translates into a sequence of 60 amino acids (the homeodomain motif) that binds to specific DNA or RNA sequences (Mooi and David, 2008). Hox genes have been found in animals, plants, fungi, and numerous other eukaryotes, and they are taxonomically widespread and highly conserved, suggesting a basal origin on the phylogenetic aspect. Almost all developmental gene networking involves the Hox gene family in metazoans. One of the most conspicuous phenotypic of the Hox cluster is the patterning along the posterior (A/P) axis of bilaterian animals (David and Mooi, 2014). Hox genes are generally arranged into three groups, or classes, on the chromosome: anterior (Hox1 and Hox2, with Hox3 possible categorized into an independent group); medial (Hox4 to Hox8); and posterior (Hox9 to Hox13) (Mooi and David, 2008). In echinoderms, the Hox 9/10 and Hox...
11/13a, Hox 11/13b, and Hox 11/13c were designated due to the homologous analysis did not yield a clear one-to-one correlation between echinoderms genes and their chordates counterparts. The anterior class is closer to the 3’ end and the posterior closer to the 5’ end of the DNA strand. This remarkable linearity in the organization of the genes along chromosomes is the basis for a concept known as “collinearity” (Mooi and David, 2008).

There are two kinds of collinearity observed in Hox gene cluster: temporal and spatial. The genes of the anterior class (3’ end) are activated first and expressed earlier in ontogeny than those of the medial class, and the medial ones are expressed earlier than those of the posterior class at the 5’ end, and this is temporal collinearity (Mooi and David, 2008). In organisms that grow from anterior to posterior, such as arthropods, temporal collinearity is in turn transcribed into spatial collinearity, in which the first expressed genes at the 3’ end of the cluster code for developmental events in the anterior part of the organism, and the last ones at the 5’ end for events in the posterior part (Mooi and David, 2008). The schematic diagram for temporal and spatial collinearity of Hox gene cluster is shown in Figure 2.

The Hox gene cluster in Echinodermata lost the collinearity, as shown in Figure 3. Among all echinoderms, the echinoid cluster is by far the most thoroughly investigated, as it has been studied in several species belonging to quite disparate clades. In the Strongylocentrotus purpuratus, one of the most important model species of echinoids with its whole genome sequenced and annotated, the Hox cluster is almost complete besides Hox4, leaving a total of 11 genes (Li et al., 2020). The peculiarity of echinoid Hox clusters is not so much the absence of a member in the series but that the order of the genes themselves along the chromosome is rearranged. Hox1 to Hox3 are located at the 5’ end of the cluster in reverse order. This topology is the result of a chromosomal translocation and inversion event.

This observation of unique rearrangement inspired the hypothesis of the body plan of echinoids being a result of a disorganized Hox cluster (Mooi and David, 2008). Since the S. purpuratus is the first sequenced echinoid and the first in the whole phylum Echinodermata, it served as a model species for this highly diverse clade initially (Sea Urchin Genome Sequencing Consortium et al., 2006). However, some recent studies question the significance of the loss of collinearity in Echinodermata species due to a better understanding of the genomic landscape of other Echinodermata species (Byrne et al., 2016). As shown in Figure 3., sequenced samples from the Crinoidea (sea lily), Asteroidea (starfish), and Holothuroidea (sea cucumber) clades show no signs of the same translocation/inversion of S. purpuratus (Li et al., 2020).

1.4 Sinaechinocyamus mai

Here, we introduce a recent study to exemplify the importance of incorporating molecular methods in modern taxonomy research. Among 15 extant clypeasteroids recorded in Taiwan shown in Table 2., Sinaechinocyamus mai (Wang, 1984) is one of the most studied sand
dollars (Figure 4). The species was initially described by Chai-Chin Wang in 1984, and the species name is in honor of the Ting-Ying Ma, one of the eminent pioneer geologists in Taiwan and China (Wang, 1984). The original species assignment was Taiwanaster mai. However, Mooi examined the description of the feature of Taiwanaster and found that Taiwanaster is the same genus as Sinaechinocyamus Liao, 1979. Therefore, Taiwanaster is an invalid junior synonym, and Taiwanaster mai was reassigned as S. mai (Mooi, 1990b).

Aside from the paleontology research on the fossil record, investigation of the living populations of S. mai began in the early 1990s. The ecology habitat of S. mai is in the intertidal zone along the west coast of Taiwan, from Hsinchu to the mouth of the Tsengwen river (Lee et al., 2019). Since S. mai possesses several characters that do not fit any other clypeasteroids family, Wang proposed a novel superfamily, Taiwanasteritida Wang, 1984, and a new family Taiwanasteridae Wang, 1984 to place S. mai (Wang, 1984). Mainly based on the observation of miniaturized body size, Wang proposed that Sinaechinocyamus was related to other micro-echinoids in the Asia-Pacific region, such as Fibulariella acuta (Yoshiwara, 1898). On the other hand, Mooi concluded that Sinaechinocyamus is a derived scutelline sand dollar (Mooi, 1990b). More recent morphology-based phylogeny also supports the hypothesis that S. mai is a derived scutelline and places the family Taiwanasteridae as incertae sedis within Scutelliformes (Kroh and Smith, 2010). Based on the morphological study, the extant sister group may be Scaphechinus mirabilis, a widely distributed sand dollar in the northwestern Pacific (Mooi, 1990b). Adult S. mirabilis generally has a body diameter of approximately 7 cm in size, while adult S. mai nearly never exceeds 1 cm (Chen and Chao, 1997). Miniaturization does not only represent smaller body size but also contributes to essential physiology, behavior, development, and reproductive changes according to the size. There are two main models of paedomorphosis: “progenesis”, which miniaturized species has a similar growth rate but growth period ends earlier, and “neoteny”, which is characterized by a similar growth period with a slower growth rate (McNamara, 1986). In the case of S. mai, it seems that the predominant model might be the neoteny due to the homogenous decrease of growth rate to 19% of S. mirabilis with a comparable growth period (Chen and Chao, 1997). Further investigation of comparing the developmental process of S. mirabilis and S. mai might give valuable insight into the mechanism of miniaturization of S. mai. Fossil record and paleoclimatology may provide a possible evolutionary explanation of miniaturization.

In this study, we aimed to utilize the transcriptome-based approach to resolve the phylogenetic placing of S. mai, and investigate the monophyly of clypeasteroids through incorporating public available transcriptomic datasets of irregular echinoids.
2. METHODOLOGY

2.1 Sample and library preparation

Live specimens of *Sinaechinocyamus mai* were collected in their native habitat in Miaoli County, Taiwan (120°39′E; 24°29′N) in May 2018. Total RNA of the sample was extracted using Tri Reagent (Ambion, USA). RNA concentration and quality were determined using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, USA) and an Agilent 2100 Bioanalyzer (Agilent Technologies, USA), which calculates an RNA integrity number (RIN). Total RNA with A260/A280 = 1.8-2.0 and RIN >8.0. Normalize intact RNA samples to 200 ng/µl with DEPC-treated H2O. Total 1~2 ug RNA was purified by using poly-T oligo-attached magnetic beads. Following purification, the mRNA was fragmented into small pieces using divalent cations under elevated temperatures. The cleaved RNA fragments were reverse transcribed into first-strand complementary DNA (cDNA) using reverse transcriptase and random primers. This was followed by second-strand cDNA synthesis using DNA Polymerase I and RNase H. These cDNA fragments then went through an end repair process, adding of a single “A” base, and then ligating the adapters. The products were then purified and enriched with PCR to create the final cDNA library. Paired-end 75 nucleotide reads from each mRNA library were obtained using NextSeq500 (Illumina Inc., CIC bioGUNE, Spain).

2.2 RNA sequencing

We quantified the cDNA with Life Technologies Qubit and Agilent Technologies Tapestation 4200 and used 0.625 nanograms of size-selected cDNA for the sample preparation and the barcoded library prepared by the Chromium Gel bead and library kit (10X Genomics, USA). Barcoded library sequencing was performed with Illumina Novaseq 6000 sequencer with 2 x 151 paired-end reads. All read-pairs contain a 16-base barcode.

2.3 Bioinformatic analysis

The list of the species incorporated in this study is shown in Table 3. The raw reads from the RNA sequencing of *Sinaechinocyamus mai* in this study and published data in the Sequence Read Archive (SRA) repository underwent quality control and adapter removal using Trimmomatic (Bolger et al., 2014) with default parameters. Reads passed quality control was assembled by Trinity (Haas et al., 2013). Transcriptome-based phylogenetic analysis was then conducted through the Agalma pipeline (Dunn et al., 2013), which sequentially annotated the transcriptome assembly with BLAST (Camacho et al., 2009), aligned gene clusters with MACSE (Ranwez et al., 2011), trimmed alignments with GBLOCKS (Talava and Castresana, 2007), and provided preliminary maximum likelihood (ML) phylogeny based on the supermatrix with RAxML (Stamatakis, 2014).
Further ML phylogeny analysis was separately conducted with partition of homologues based on the Agalma pipeline output and without partition in RAxML v. 8.2.12 with 100 replicates, respectively. The substitution model was selected automatically by RAxML (-m PROTGAMMAAUTO).

3. RESULTS

3.1 Transcriptome-based phylogeneomics

The analysis included 22 echinoid species, and *Eucidaris tribuloides* was designated as the outgroup. In total, the Agalma pipeline yielded 9630 loci comprised 2,574,725 amino acid positions with an actual gene occupancy of 39.2%. The concatenated supermatrix dataset was then used in the transcriptome-based phylogenetic analysis. Both partitioned and unpartitioned dataset was used to infer the phylogeny, and the topology was identical. The phylogeny is shown in Figure 5. The result is congruent with Koch et al. (2018) and Koch et al. (2020). The result revealed that the “cassuloids” *Conolampas sigsbei* is a sister group to the Scutelloida clade. Therefore, the monophyly of traditional clypeasteroids was not supported. As for the newly sequenced *S. mai*, the result suggests that *S. mirabilis* is its sister taxa in the sampled dataset, supporting previous studies drawn on morphological analysis.

4. DISCUSSION

4.1 Monophyly of clypeasteroids

The overall cladogram comparison between previous studies and the result from this work is shown in Figure 6. Mongiardino Koch et al. (2018) firstly used transcriptome-based phylogeny to challenge the morphology-based monophyly of clypeasteroids. In this study, the “cassiduloids” included in the analysis, *Conolampas sigsbei*, intertwined in the phylogeny relationship of Clypeasteroida and Scutelloida.

In our previous study (Lin et al., 2020), we sequenced the whole mitochondria genome of *S. mai* and inferred the phylogenetic relationship through the data. The availability of a mitogenome for *S. mai* sets the stage for exploring the systematic position of this taxon. According to the mitogenome-based phylogeny, the clades of Echinolampadoida and Scutelloida both receive 100% bootstrap support (Lin et al., 2020). Consistent with the hypothesis raised by Mooi, this result suggests that the presence of Aristotle’s lantern in adults was firstly lost in the basal Irregularia, then the re-appeared in the Scutelloida clade (Mooi, 1990a). However, due to the fact that mitogenome data from the Scutelloida clade was unavailable, and this work was unable to support nor refute the monophyly of clypeasteroids.

On the other hand, one analysis published recently tried reconstructing the phylogeny of Echinoidea using the total-evidence dated approach incorporating the fossil morphology evidence and transcriptomic data of extant species (Mongiardino Koch and Thompson, 2020).
By combining the morphological fossil record and molecular data (transcriptome dataset was used for phylogenetic inference in this study), the monophyly of clypeasteroids was refuted again. Therefore, under this phylogenetic theory, the morphological characters shared between Clypeasteroida and Scutelloida (e.g., extremely flattened test and presence of Aristotle’s lantern) result from convergent evolution rather than true synapomorphies (Mongiardino Koch and Thompson, 2020).

The transcriptome-based phylogeny analysis presented in this article was consistent with Mongiardino Koch and Thompson (2020), which did not support the monophyly of clypeasteroids. The reason for this inconsistency between mitogenome- and transcriptome-based phylogeny might be primarily due to the limited sampling. As the mitogenome-based analysis did not include Clypeaster spp. from Clypeasteroida, the possible polyphyly status of clypeasteroids cannot be revealed.

4.2 Future Perspective

Despite the fact the seemingly solid evidence provided in the transcriptome-based phylogeny, the only representative of “cassiduloids” clade was the extant species Conolampas sigsbei which was resolved as the sister group to Scutelloida. Therefore, one should be careful when interpreting this result, and further investigation sampling more species in the “cassiduloids” clade is still needed for a deeper understanding of the conundrum of the monophyly of clypeasteroids. Nevertheless, these recent works showcase the potential of using molecular data to aid phylogenetic research in the field.

On the other hand, although the role of Hox gene cluster rearrangement in establishing pentameric symmetry of Echinodermata has been questioned based on the due to several sequenced species from Crinoidea (sea lily), Asteroidea (starfish), and Holothuroidea (sea cucumber) clades do not present with rearranged clusters. Due to the lack of available data on irregular echinoids, it is still unclear if the structure of Hox gene cluster contributes to secondary bilateralism in the Irregularia. A further investigation of the high-quality genome assembly within this group will benefit the research of this topic.
5. CONCLUSION

The technology breakthroughs in the first ten years of the 21st century serve as an important foundation for further investigation in the field of genomics, and investigators of other specialties benefit significantly as well, including paleontologists. Both the traditional morphology-based paleontology and novel genetic methods have their own limitation. However, by combining these two paradigms, further investigation may shed light on the phylogenetic relationships, developmental processes, and the possible mechanism behind the secondary bilateral symmetry of clypeasteroids.
Acknowledgments

This study was funded by the Taiwan Ministry of Science and Technology (MOST 108-2116-M-002-014, MOST 109-2116-M-002-020 and MOST 110-2116-M-002-016 to JPL). We are grateful to the Computer and Information Networking Center, National Taiwan University for the support of high-performance computing facilities.
REFERENCES

Bolger, A. M., Lohse, M., and Usadel, B., 2014: Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114-2120, doi: 10.1093/bioinformatics/btu170

Byrne, M., Martinez, P., and Morris, V., 2016: Evolution of a pentameral body plan was not linked to translocation of anterior Hox genes: the echinoderm HOX cluster revisited. *Evol Dev*, 18, 137-143, doi: 10.1111/ede.12172

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L., 2009: BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 421, doi: 10.1186/1471-2105-10-421

Cannon, J. T., and Kocot, K. M., 2016: Phylogenomics Using Transcriptome Data. *Methods Mol Biol*, 1452, 65-80, doi: 10.1007/978-1-4939-3774-5_4

Chan, C. X., and Ragan, M. A., 2013: Next-generation phylogenomics. *Biol Direct*, 8, 3, doi: 10.1186/1745-6150-8-3

Chen, C. P., and Chao, C. M., 1997: Reduction of Growth Rate as the Major Process in the Miniaturization of the Sand Dollar Sinaechinocyamus mai. *Biol Bull*, 193, 90-96, doi: 10.2307/1542738

Coppard, S. E., and Lessios, H. A., 2017: Phylogeography of the sand dollar genus Encope: implications regarding the Central American Isthmus and rates of molecular evolution. *Sci Rep*, 7, 11520, doi: 10.1038/s41598-017-11875-w

David, B., and Mooi, R., 2014: How Hox genes can shed light on the place of echinoderms among the deuterostomes. *Evodevo*, 5, 22, doi: 10.1186/2041-9139-5-22

Dunn, C. W., Howison, M., and Zapata, F., 2013: Agalma: an automated phylogenomics workflow. *BMC Bioinformatics*, 14, 330, doi: 10.1186/1471-2105-14-330

Fell, H. B., Pawson, D. L., 1966, U. Echinodermata. 3, Boulder, CO, Geological Society of America, Treatise on Invertebrate Paleontology.

Galtier, N., and Daubin, V., 2008: Dealing with incongruence in phylogenomic analyses. *Philos Trans R Soc Lond B Biol Sci*, 363, 4023-4029, doi: 10.1098/rstb.2008.0144

Gaunt, S. J., 2015: The significance of Hox gene collinearity. *Int J Dev Biol*, 59, 159-170, doi: 10.1387/ijdb.150223sg

Goodwin, S., McPherson, J. D., and McCombie, W. R., 2016: Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*, 17, 333-351, doi: 10.1038/nrg.2016.49

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F.,
Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., and Regev, A., 2013: De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*, 8, 1494-1512, doi: 10.1038/nprot.2013.084

Horton, T., Kroh, A., Ahyong, S., Bailly, N., Boyko, C. B., Brandão, S. N., Gofas, S., Hooper, J. N. A., Hernandez, F., Holovachov, O., Mees, J., Molodtsova, T. N., Paulay, G., Decock, W., Dekeyser, Z., Poffyn, G., Vandepitte, L., Vanhoorne, B., Adlard, R., Agatha, S., Ahn, K. J., Akkari, N., Alvarez, B., Amorim, V., Anderberg, A., Anderson, G., Antic, D., Antonietto, L. S., Arango, C., Artois, T., Atkinson, S., Auffenber, K., Baldwin, B. G., Bank, R., Barber, A., Barbosa, J. P., Bartsch, I., Bellan-Santini, D., Bergh, N., Bernot, J., Berta, A., Bezerra, T. N., Bieler, R., Blanco, S., Blasco-Costa, I., Blazewicz, M., Bock, P., Bonifacio de León, M., Böttger-Schnack, R., Bouche, P., Boury-Esnault, N., Bouzian, R., Boxshall, G., Bray, R., Bruce, N. L., Bueno, V., Bueno-Villegas, J., Cairns, S., Calvo Casas, J., Carballo, J. L., Cárdenas, P., Carstens, E., Chan, B. K., Chan, T. Y., Cheng, L., Christenhusz, M., Churchill, M., Coleman, C. O., Collins, A. G., Collins, G. E., Corbari, L., Cordeiro, R., Cornils, A., Coste, M., Costello, M. J., Crandall, K. A., Cremonte, F., Cribb, T., Cutmore, S., Dahdouh-Guebas, F., Daly, M., Daneliya, M., Davvin, J. C., Davie, P., De Broyer, C., De Grave, S., de Lima Ferreira, P., de Mazancourt, V., de Voogd, N. J., Decker, P., Defaye, D., d'Hondt, J. L., Dippenaar, S., Dohrmann, M., Dolan, J., Domning, D., Downey, R., Dreyer, N., Ector, L., Eresende, U., et al., 2021, World Register of Marine Species (WoRMS), WoRMS Editorial Board.

Kroh, A., 2020, Phylogeny and classification of echinoids, Developments in Aquaculture and Fisheries Science, Volume 43, Elsevier, 1-17.

Kroh, A., and Smith, A. B., 2010: The phylogeny and classification of post-Palaeozoic echinoids. *Journal of Systematic Palaeontology*, 8, 147-212, doi: 10.1080/14772011003603556

Lacalli, T., 2014: Echinoderm conundrums: Hox genes, heterochrony, and an excess of mouths. *Evodevo*, 5, 46, doi: 10.1186/2041-9139-5-46

Lee, H., Lin, J.-P., Li, H.-C., Chang, L.-Y., Lee, K.-S., Lee, S.-J., Chen, W.-J., Sankar, A., and Kang, S.-C., 2019: Young colonization history of a widespread sand dollar (Echinodermata; Clypeasteroida) in western Taiwan. *Quaternary International*, 528, 120-129, doi: 10.1016/j.quaint.2018.12.003

Li, Y., Omori, A., Flores, R. L., Satterfield, S., Nguyen, C., Ota, T., Tsurugaya, T., Ikuta, T., Ikeo, K., Kikuchi, M., Leong, J. C. K., Reich, A., Hao, M., Wan, W., Dong, Y., Ren, Y., Zhang, S., Zeng, T., Uesaka, M., Uchida, Y., Li, X., Shibata, T. F., Bino, T., Ogawa, K., Shigenobu, S., Kondo, M., Wang, F., Chen, L., Wessel, G., Saiga, H., Cameron, R. A., Livingston, B., Bradham, C., Wang, W., and Irie, N., 2020: Genomic insights of body plan transitions from
bilateral to pentameral symmetry in Echinoderms. *Commun Biol*, **3**, 371, doi: 10.1038/s42003-020-1091-1

Lin, J. P., Tsai, M. H., Kroh, A., Trautman, A., Machado, D. J., Chang, L. Y., Reid, R., Lin, K. T., Bronstein, O., Lee, S. J., and Janies, D., 2020: The first complete mitochondrial genome of the sand dollar Sinaechinocyamus mai (Echinoidea: Clypeasteroida). *Genomics*, **112**, 1686-1693, doi: 10.1016/j.ygeno.2019.10.007

Lin, Y.-J., Fang, J.-N., Chang, C.-C., Cheng, C.-C., and Lin, J.-P., In press: Stereomic microstructure of Clypeasteroida in thin section based on new material from Pleistocene strata in Taiwan. doi: 10.3319/TAO.2021.07.28.01

Mongiardino Koch, N., Coppard, S. E., Lessios, H. A., Briggs, D. E. G., Mooi, R., and Rouse, G. W., 2018: A phylogenomic resolution of the sea urchin tree of life. *BMC Evol Biol*, **18**, 189, doi: 10.1186/s12862-018-1300-4

Mongiardino Koch, N., and Thompson, J. R., 2020: A Total-Evidence Dated Phylogeny of Echinoidea Combining Phylogenomic and Paleontological Data. *Syst Biol*, doi: 10.1093/sysbio/syaa069

Mooi, R., 1989: Living and fossil genera of the Clypeasteroida (Echinoidea, Echinodermata): an illustrated key and annotated checklist. *Smithsonian Contributions to Zoology*, doi:

-, 1990a: Paedomorphosis, Aristotle's lantern, and the origin of the sand dollars (Echinodermata: Clypeasteroida). *Paleobiology*, **16**, 25-48, doi: 10.1017/s0094837300009714

Mooi, R., 1990b, Progenetic miniaturization in the sand dollar Sinachinocyamus: implications for clypeastroid phylogeny, *In Jangoux, M., ed., Echinoderm Research: Rotterdam, CRC Press, 137-143.*

Mooi, R., and David, B., 2008: Radial Symmetry, the Anterior/Posterior Axis, and Echinoderm Hox Genes. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 43-62, doi: 10.1146/annurev.ecolsys.39.110707.173521

Orlando, L., and Cooper, A., 2014: Using Ancient DNA to Understand Evolutionary and Ecological Processes. *Annual Review of Ecology, Evolution, and Systematics*, **45**, 573-598, doi: 10.1146/annurev-ecolsys-120213-091712

Paabo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., and Hofreiter, M., 2004: Genetic analyses from ancient DNA. *Annu Rev Genet*, **38**, 645-679, doi: 10.1146/annurev.genet.37.110801.143214

Pawson, D. L., 2007: Phylum Echinodermata. *Zootaxa*, **1668**, 749-764, doi: 10.11646/zootaxa.1668.1.31
Peterson, K. J., Arenas-Mena, C., and Davidson, E. H., 2000: The A/P axis in echinoderm ontogeny and evolution: evidence from fossils and molecules. *Evol Dev, 2*, 93-101, doi: 10.1046/j.1525-142x.2000.00042.x

Qu, X. J., Jin, J. J., Chaw, S. M., Li, D. Z., and Yi, T. S., 2017: Multiple measures could alleviate long-branch attraction in phylogenomic reconstruction of Cupressoideae (Cupressaceae). *Sci Rep, 7*, 41005, doi: 10.1038/srep41005

Ranwez, V., Harispe, S., Delsuc, F., and Douzery, E. J., 2011: MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One, 6*, e22594, doi: 10.1371/journal.pone.0022594

Saucède, T., Mooi, R., and David, B., 2007: Phylogeny and origin of Jurassic irregular echinoids (Echinodermata: Echinoidea). *Geological Magazine, 144*, 333-359, doi:

Sea Urchin Genome Sequencing Consortium, Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., Gibbs, R. A., Angerer, R. C., Angerer, L. M., Arnone, M. I., Burgess, D. R., Burke, R. D., Coffman, J. A., Dean, M., Elphick, M. R., Ettensohn, C. A., Foltz, K. R., Handoun, A., Hynes, R. O., Klein, W. H., Marzluff, W., McClay, D. R., Morris, R. L., Mushegian, A., Rast, J. P., Smith, L. C., Thorndyke, M. C., Vacquier, V. D., Wessel, G. M., Wray, G., Zhang, L., Elsik, C. G., Ermolaeva, O., Hlavina, W., Hofmann, G., Kitts, P., Landrum, M. J., Mackey, A. J., Maglott, D., Panopoulou, G., Poustka, A. J., Pruitt, K., Sapojnikov, V., Song, X., Souvorov, A., Solovyev, V., Wei, Z., Whittaker, C. A., Worley, K., Durbin, K. J., Shen, Y., Fedrigo, O., Garfield, D., Haygood, R., Primus, A., Satija, R., Severson, T., Gonzalez-Garay, M. L., Jackson, A. R., Milosavljevic, A., Tong, M., Killian, C. E., Livingston, B. T., Wilt, F. H., Adams, N., Belle, R., Carboneau, S., Cheung, R., Cormier, P., Cosson, B., Croce, J., Fernandez-Guerra, A., Genevire, A. M., Goel, M., Kelkar, H., Morales, J., Mulner-Lorillon, O., Robertson, A. J., Goldstone, J. V., Cole, B., Epel, D., Gold, B., Hahn, M. E., Howard-Ashby, M., Scally, M., Stegeman, J. J., Allgood, E. L., Cool, J., Judkins, K. M., McCafferty, S. S., Musante, A. M., Obar, R. A., Rawson, A. P., Rossetti, B. J., Gibbons, I. R., Hoffman, M. P., Leone, A., Istrail, S., Materna, S. C., Samanta, M. P., Stolc, V., et al., 2006: The genome of the sea urchin Strongylocentrotus purpuratus. *Science, 314*, 941-952, doi: 10.1126/science.1133609

Smith, A. B., Pisani, D., Mackenzie-Dodds, J. A., Stockley, B., Webster, B. L., and Littlewood, D. T., 2006: Testing the molecular clock: molecular and paleontological estimates of divergence times in the Echinoidea (Echinodermata). *Mol Biol Evol, 23*, 1832-1851, doi: 10.1093/molbev/msl039

Stamatakis, A., 2014: RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*, 1312-1313, doi: 10.1093/bioinformatics/btu033
Susko, E., and Roger, A. J., 2021: Long Branch Attraction Biases in Phylogenetics. *Syst Biol*, 70, 838-843, doi: 10.1093/sysbio/syab001

Talavera, G., and Castresana, J., 2007: Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol*, 56, 564-577, doi: 10.1080/10635150701472164

Thompson, J. R., Erkenbrack, E. M., Hinman, V. F., McCauley, B. S., Petsios, E., and Bottjer, D. J., 2017: Paleogenomics of echinoids reveals an ancient origin for the double-negative specification of micromeres in sea urchins. *Proc Natl Acad Sci U S A*, 114, 5870-5877, doi: 10.1073/pnas.1610603114

van der Valk, T., Pecnerova, P., Diez-Del-Molino, D., Bergstrom, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., Thomas, J. A., Dehasque, M., Saglican, E., Fidan, F. R., Barnes, I., Liu, S., Somel, M., Heintzman, P. D., Nikolskiy, P., Shapiro, B., Skoglund, P., Hofreiter, M., Lister, A. M., Gotherstrom, A., and Dalen, L., 2021: Million-year-old DNA sheds light on the genomic history of mammoths. *Nature*, 591, 265-269, doi: 10.1038/s41586-021-03224-9

Wang, C. C., 1984: New Classification Of Clypeasteroid Echinoids. *Proc Geol Soc China*, 27, 119-152, doi:

Young, A. D., and Gillung, J. P., 2019: Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, 45, 225-247, doi: 10.1111/syen.12406

Ziegler, A., Lenihan, J., Zachos, L. G., Faber, C., and Mooi, R., 2015: Comparative morphology and phylogenetic significance of Gregory’s diverticulum in sand dollars (Echinoidea: Clypeasteroida). *Organisms Diversity & Evolution*, 16, 141-166, doi: 10.1007/s13127-015-0231-9
Figure legends

Figure 1. Comparison of the body symmetry between regular and irregular echinoids. Modified from Fell (1966) and Mooi (1989).

Figure 2. The schematic diagram for the temporal and spatial collinearity of the Hox gene cluster. Data from Mooi and David (2008).

Figure 3. Phylogeny and Hox gene cluster arrangement among Echinodermata clades. Modified from Byrne et al. (2016) and Li et al. (2020). Arrows represent transcription direction, boxes represent unknown transcription direction, and dash lines represent discontinuous sequences.

Figure 4. The living individual of Sinaechinocyamus mai.

Figure 5. The maximum likelihood (ML) phylogram based on the analysis of unpartitioned supermatrix dataset, rooted on the outgroup Eucidaris tribuloides. The topology was identical between partitioned and unpartitioned analysis, and all nodes received maximum bootstrapping support value in both analyses.

Figure 6. Comparison among the morphology-, mitogenome-, and transcriptome-based phylogenetic analyses. Clades with incongruent positions across studies are highlighted in color. Data from Kroh and Smith (2010), Mongiardino Koch et al. (2018), Mongiardino Koch and Thompson (2020), and Lin et al. (2020)
Illustrations

Regular echinoids (aboral side)

Irregular echinoids (aboral side)

Ambulacrum
Anus
Interambulacrum

---

Symmetry axis

Figure 1.
Figure 2.
Figure 3.
Figure 5.
Figure 6.
### Tables

Table 1. Comparison of three strategies for implementing molecular data in phylogeny analysis

|                        | Single gene markers or organelle genomics | Transcriptome | Whole-genome sequencing |
|------------------------|------------------------------------------|---------------|-------------------------|
| **Sample requirement** | Low                                      | High          | Variable.               |
|                        | Museum specimens preserved in alcohol are usually satisfactory. | RNA is unstable in most storage conditions, and special preserving measures like liquid nitrogen are needed. Thus, it is not suitable for decomposed specimens. | The desired reading depth, the average sequenced counts of individual bases in the genome, determines the sample requirement. Museum specimens preserved in alcohol are generally satisfactory. |
| **Acquired genomic data** | Predetermined target loci or organelle genomes (mitochondria or chloroplast) | Majority of coding sequences of the nuclear genome | Nearly complete genome, including coding and non-coding regions. |
| **Computation requirement** | Low, can be analyzed by personal computers | Middle-high, need to be analyzed by laboratory workstations | High, need to be analyzed by laboratory workstations or computing clusters |
| **Potential pitfalls** | Evolutionary trajectory of organelle genomes and single gene markers may be different from overall species, particularly under the circumstance of horizontal gene transfer. | Sampling from different tissues possibly yields different mRNA isoforms, rendering orthologue identification more challenging. | Poor quality assembly may introduce artifacts interfering with downstream phylogenetic inference. |
| **Cost**               | Low                                      | Middle        | High                    |
| **Phylogeny resolution** | Low                                      | Middle        | High                    |
| **Example**            | Smith et al. (2006)                       | Mongiardino Koch et al. (2018) | Allen et al. (2017) |
Table 2. Extant clypeasteroids species in Taiwan. Based on Lee et al. (2019) and Horton et al. (2021).

| Species                          | Author and Year |
|----------------------------------|-----------------|
| *Arachnoides placenta*           | Linnaeus, 1758  |
| *Astriclypeus mannii* Verrill    | 1867            |
| *Clypeaster reticulatus*         | Linnaeus, 1758  |
| *Clypeaster virescens* Döderlein | 1885            |
| *Echinocyamus crispus* Mazzetti  | 1893            |
| *Fibularia ovulum* Lamarck       | 1816            |
| *Fibulariella acuta* (Yoshiwara)| 1898            |
| *Jacksonaster depressum* (L. Agassiz)| 1841 |
| *Laganum decagonalis* Blainville| 1827            |
| *Laganum fudsiyama* Döderlein    | 1885            |
| *Lagignum lagnum* Leske          | 1778            |
| *Peronella lesueuri* (L. Agassiz)| 1841            |
| *Peronella orbicularis* Leske    | 1778            |
| *Sculpsitechinus aruitus* Leske  | 1778            |
| *Sinaechinocyamus mai* Wang     | 1984            |
Table 3. The list of species used in the transcriptome-based phylogenetic analysis in this study. The clade designation is based on Kroh (2020) and Horton et al. (2021).

| Clade                                      | Species                                                | NCBI_ID  | SRA        |
|--------------------------------------------|--------------------------------------------------------|----------|------------|
| Camarodonta Jackson, 1912                  | *Colobocentrotus (Podophora) atratus* (Linnaeus, 1758) | 39290    | SRR7513588 |
| Camarodonta Jackson, 1912                  | *Sphaerechinus granularis* (Lamarck, 1816)             | 39374    | SRR1139199 |
| Cidaroida Claus, 1880                      | *Eucidaris tribuloides* (Lamarck, 1816)                | 7632     | SRR2844625 |
| Clypeasteroida A. Agassiz, 1835            | *Clypeaster rosaceus* (Linnaeus, 1758)                 | 105356   | SRR7513591 |
| Clypeasteroida A. Agassiz, 1835            | *Clypeaster subdepressus* (Gray, 1825)                 | 105357   | SRR7513586 |
| Diadematoida Duncan, 1889                  | *Diadema setosum* (Leske, 1778)                       | 31175    | SRR7513577 |
| Echinolampadoida Kroh & Smith, 2010       | *Conolampas sigsbei* (A. Agassiz, 1878)                | 271677   | SRR7513579 |
| Echinothurioida Claus, 1880                | *Araeosoma leptaleum* A. Agassiz & H.L. Clark, 1909   | 105353   | SRR7513578 |
| Holasteroida Durham & Melville, 1957      | *Pilematechinus sp.*                                    | 2268254  | SRR7513585 |
| Scutelloida Mongiardino Koch et al., 2018  | *Echinarchniius parma* (Lamarck, 1816)                  | 869203   | SRR1139193 |
| Scutelloida Mongiardino Koch et al., 2018  | *Mellita tenuis* H.L. Clark, 1940                      | 1351943  | SRR7513583 |
| Scutelloida Mongiardino Koch et al., 2018  | *Dendraster excentricus* (Eschscholtz, 1831)           | 273752   | SRR2844623 |
| Scutelloida Mongiardino Koch et al., 2018  | *Scaphechinus mirabilis* A. Agassiz, 1864              | 262334   | DRR197471  |
| Scutelloida Mongiardino Koch et al., 2018  | *Sinaechinocyamus mai* (Wang, 1984)                    | 2661602  | NA         |
| Spatangoida (L. Agassiz, 1840)             | *Abatus agassizii* Mortensen, 1910                     | 1194296  | SRR1324765 |
| Spatangoida (L. Agassiz, 1840)             | *Abatus cordatus* (Verrill, 1876)                      | 471970   | SRR1324767 |
| Taxon                                      | Species Name                               | Accession Number | SRA Accession |
|-------------------------------------------|--------------------------------------------|------------------|--------------|
| Spatangoida (L. Agassiz, 1840)            | *Brissus obesus* Verrill, 1867             | 271724           | SRR7513590   |
| Spatangoida (L. Agassiz, 1840)            | *Echinocardium cordatum* (Pennant, 1777)   | 39298            | SRR1324913   |
| Spatangoida (L. Agassiz, 1840)            | *Echinocardium mediterraneum* (Forbes, 1844) | 471983           | SRR1324910   |
| Spatangoida (L. Agassiz, 1840)            | *Meoma ventricosa* (Lamarck, 1816)         | 39340            | SRR7513582   |
| Spatangoida (L. Agassiz, 1840)            | *Schizocosmus abatoides* (H.L. Clark, 1925) | 1502937          | SRR1325136   |
| Stomopneustoida Kroh & Smith, 2010        | *Stomopneustes variolaris* (Lamarck, 1816) | 7663             | SRR7513587   |