universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

The role of cephalopod-specific microsyntenies in the evolution of novel gene regulation

verfasst von / submitted by

Gözde Yalçın

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna, 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 877
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Genetik und Entwicklungsbiologie
Betreut von / Supervisor:	Assoz. Prof. Dr. Oleg Simakov

Acknowledgements

This master thesis has been an enriching experience, and I owe my appreciation to the exceptional persons and groups that have aided in my academic and personal development through this journey.

First and foremost, I would like to express my deepest gratitude to my mentor, Thea Rogers. Her continuous encouragement, generosity, patient mentorship, and vast knowledge have been indispensable in shaping the direction of my research. Without her guidance and encouragement, I would not have achieved this milestone.

I am equally indebted to Oleg Simakov for giving me the opportunity to be part of his lab and for his valuable advice and support. Exploring the fascinating world of cephalopod genomics with him and his team has been a life-changing experience.

I would like to extend my sincere thanks to the whole Simakov group for their friendly attitude and willingness to help have made my study experience not just scientifically rewarding but also enjoyable. I could not be part of a better team.

Lastly, I want to thank my family and my friends for their support. I am grateful to my mother, and brother for their unwavering belief in my abilities. I cannot express enough gratitude to my beloved grandma (Suna cilingir), whose love and encouragement have been invaluable. And heartfelt thanks to my partner for his day-to-day encouragement and emotional support.

Table of contents

1. Abstract

Coleiod cephalopods (octopus, squid, cuttlefish) have many prominent evolutionary innovations. They have the largest invertebrate brain, as well as very sophisticated neurological and sensory systems that allow for complex behavioral patterns and rapid reactions. Previous studies indicate that compared to other extant mollusks, the genome of coleoid cephalopods substantially rearranged, and as the outcome of this occurrence hundreds of tightly related and evolutionarily distinct gene clusters (microsyntenies) have appeared. In addition to this, many cephalopod-specific traits emerged afterward and this has been hypothesized to be the result of a number of genomic mechanisms, such as tandem gene duplications, repetitive element expansions and the emergence of novel genes. However less is known about the role of gene regulation in the evolution of cephalopod phenotypic complexity. Previous research using Hawaian bobtail squid (*Euprymna scolopes*) identified MACIs (microsyntenies associated with cephalopod inventions) and ancestral, metazoan microsyntenies. Based on this, by assessing the genomic structure of these microsyntenies in *E. scolopes* using ATAC-seq data and investigating the 3D genome organization of microsyntenies in *E. scolopes* and the Californian two-spot octopus (*Octopus bimaculoides*) using high resolution genome topology data (Micro-C), we expand on prior studies and gain further insight into how regulation influences coleoid cephalopod traits. Manual examination of MACIs revealed two types of microsyntenies that appeared distinct from metazoan microsyntenies. Statistical analysis showed that metazoan microsyntenies have significantly larger gene deserts compared to MACIs, but inversely gene deserts of MACIs exhibit higher ATAC peak coverage, suggesting their regulatory architecture is substantially different. Furthermore, contact maps of genome structure showed that MACIs are more frequently located on the TAD borders compared with metazoan microsyntenies, indicating that they are different in their 3D organization and suggesting that selective constraints at the regulatory level may differ between each microsynteny type. Taken together, our results allude to a coleoid cephalopod-specific ‘mode’ of gene regulation evolution.

1. Die Zusammenfassung

Die Kopffüßer (Oktopus, Tintenfisch, Sepia) weisen viele herausragende evolutionäre Innovationen auf. Sie haben das größte Gehirn der Wirbellosen und verfügen über sehr ausgeklügelte neurologische und sensorische Systeme, die komplexe Verhaltensmuster und schnelle Reaktionen ermöglichen. Frühere Studien deuten darauf hin, dass sich das Genom der coleoiden Kopffüßer im Vergleich zu dem anderer Weichtiere erheblich verändert hat, was zur Entstehung von Hunderten eng verwandter und evolutionär unterschiedlicher Gencluster (Mikrosyntenien) geführt hat. Darüber hinaus haben sich viele Kopffüßer-spezifische Merkmale herausgebildet, und man nimmt an, dass dies das Ergebnis einer Reihe von genomischen Mechanismen ist, wie z. B. Tandem-Genduplikationen, Ausdehnung repetitiver Elemente und das Auftreten neuer Gene. Über die Rolle der Genregulation bei der Entwicklung der phänotypischen Komplexität von Kopffüßern ist jedoch weniger bekannt. Bei früheren Untersuchungen an Hawaianischen Bobtail-Tintenfischen (*Euprymna scolopes*) wurden MACIs (Mikrosyntenien, die mit Cephalopoden-Erfindungen in Verbindung gebracht werden) und angestammte, metazoische Mikrosyntenien identifiziert. Darauf aufbauend haben wir die genomische Struktur dieser Mikrosyntenien in E. scolopes mit Hilfe von ATAC-seq-Daten bewertet und die 3D-Genomorganisation der Mikrosyntenien in *E. scolopes* und dem kalifornischen Zweipunktkraken (*Octopus bimaculoides*) mit Hilfe von hochauflösenden Genomtopologiedaten (Micro-C) untersucht, um frühere Studien zu erweitern und weitere Erkenntnisse darüber zu gewinnen, wie die Regulierung coleoide Kopffüßermerkmale beeinflusst. Die manuelle Untersuchung der MACIs ergab zwei Arten von Mikrosyntenien, die sich von den Mikrosyntenien der Metazoen zu unterscheiden scheinen. Die statistische Analyse zeigte, dass Mikrosyntenien von Metazoen im Vergleich zu MACIs deutlich größere Genwüsten aufweisen, während umgekehrt die Genwüsten von MACIs eine höhere ATAC-Peakabdeckung aufweisen. Darüber hinaus zeigten Kontaktkarten der Genomstruktur, dass MACIs im Vergleich zu Mikrosyntenien von Metazoen wesentlich häufiger an TAD-Grenzen zu finden sind, was darauf hindeutet, dass sich MACIs in ihrer regulatorischen Architektur unterscheiden und dass selektive Einschränkungen auf der regulatorischen Ebene zwischen MACIs und Mikrosyntenien von Metazoen unterschiedlich sein könnten. Zusammengenommen deuten unsere Ergebnisse auf einen coleoiden, cephalopodenspezifischen "Modus" der Genregulationsevolution hin.

2. Introduction

2.1. Coleoid cephalopods as model organisms for studying genome evolution

Cephalopoda is a class of marine invertebrates of the phylum Mollusca (Figure 1). Cephalopods have a long evolutionary history, first diverging from other molluscan clades 530 MYA (Kröger et al., 2011). Two major clades of cephalopods exist today: Nautiloidea, which includes several species of Nautilus, and Coleoidea, which comprises the majority of cephalopod species (Albertin & Simakov, 2020). Among the coleoid cephalopods, there are two more suborders, the Decapodiformes (squid and cuttlefish) and the Octopodiformes (octopuses and vampire squid) (Albertin & Simakov, 2020) (Figure 1). There have been a number of phylogenetic studies investigating the relationships among cephalopods, both at species level and genus level (Allcock et al., 2015; Sanchez et al., 2018; Albertin & Simakov, 2020). Although several evolutionary scenarios have been proposed for some clades, particularly the Decapodiforms, the phylogenetic branching within them remains largely unresolved (Lindgren & Anderson, 2018). However, phylogenetic analyses using genomic data as well as paleontological data indicate that the nautiloid–coleoid split occurred ~420 MYA (Zhang et al., 2021) (Figure 1) and Octopodiform–Decapodiform divergence occurred around 270 MYA (Kröger et al., 2011; Albertin & Simakov, 2020) (Figure 1).

Coleiod cephalopods have many clade- and species-specific innovations. For example, they possess the largest invertebrate brain and have highly complex nervous and sensory systems which enable complex behavioral patterns and fast responses (Albertin & Simakov, 2020; Packard, 1972; O'Brien et al., 2018). Furthermore, they have evolved suckered tentacles, camera-like eyes, color-changing skin, and complex learning behavior as well as novel organs such as the accessory nidamental gland (ANG) and the light organ in bobtail squids (Belcaid et al., 2019; Nyholm & McFall-Ngai, 2021), providing excellent adaptations for predation, locomotion, disguise, and communication (Albertin & Simakov, 2020; Packard, 1972).

Coleoid cephalopods have been used as model organisms for several decades, primarily due to their unique morphological and behavioral features (Xavier et al., 2015). Moreover, the Hawaiian bobtail squid (*Euprymna scolopes*) has frequently been used in symbiosis research due to the relationship of symbiotic bacteria housed in the light organ and ANG of the squid host (Albertin & Simakov, 2020). In particular, the small size, brief lifespan, rapid growth, and availability all year round of *E. scolopes* makes it a valuable cephalopod model species (Lee et al., 2009). Therefore, many molecular techniques have been created to examine the development of *E. scolopes* (Lee et al., 2009), meaning it is a suitable candidate for studying the evolution of cephalopod novel traits.

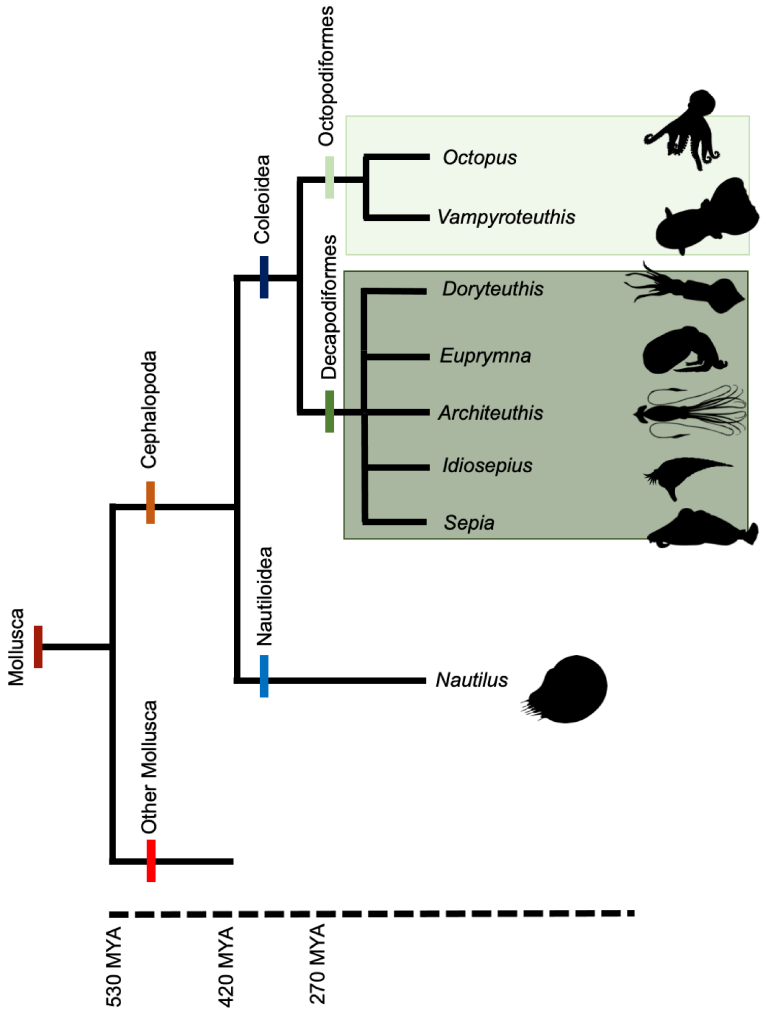

Figure 1. Phylogeny of cephalopods and their divergence times. Pale green indicates the Octopodiformes (octopus and vampire squid) and dark green indicates the Decapodiformes (squid and cuttlefish), which together make up the coleoid cephalopods.

More recently, genomic and transcriptomic data has emerged for cephalopods, including the publishing of several chromosome-scale genomes (Schmidbaur et al., 2022 ; Albertin et al., 2022), allowing for researchers to begin to untangle the genomic mechanisms underlying the evolution of cephalopod novelties. It was revealed that compared with other extant mollusks, the genomes of soft-bodied (coleoid) cephalopods are highly rearranged, indicating an early, intense burst of genome remodeling (Albertin et al., 2022). In addition to this, coleoids have a multi-megabase tandem array of genes associated with brain development and innovations specific to cephalopods (Albertin et al., 2022, Schmidbaur et al., 2022). Another interesting characteristic of cephalopod genomes is the rapid expansion of their intergenic regions by repetitive elements, as the conserved syntenic regions of *E. scolopes* possess much larger intergenic distances than their orthologs found in other mollusks or even vertebrates (Albertin & Simakov, 2020). However, the functional

significance of this dramatic rearrangement and subsequent genomic and regulatory architecture of cephalopod genomes remains unclear (Albertin & Simakov, 2020).

2.2. Coleoid cephalopod gene regulation

A synteny (or syntenic locus) is any pair of genes located on the same chromosome regardless of whether they are genetically linked (Passarge et al., 1999). The conservation of homologous chromosomes is known as macrosynteny, while the conservation of local gene content and order is known as microsynteny (Li et al., 2022). Previously, it was revealed that the *E. scolopes* genome displayed key genomic transitions toward cephalopod genomic architecture. It has been found that more than half of the local gene linkages (microsynteny) that were conserved across many non-cephalopod bilaterian genomes were lost in both *O. bimaculoides* and *E. scolopes*. In addition to this, a large number of gene linkages were found to be shared between bobtails and octopuses that were not present in other metazoans, supporting a genomic reorganization in the coleoid cephalopod ancestor (Belcaid et al., 2019). These large-scale reconstruction events in coleoid cephalopod genomes bought together many previously unlinked genes due to the mixing of ancestral chromosomes (Albertin et al., 2022). The consequence of this has been the emergence of new chromosomes and microsyntenies (Belcaid et al., 2019; Schmidbaur et al., 2022). Genes in these novel microsyntenies are expressed in highly organized organ systems (e.g., the central nervous system) as well as symbiotic organs from *E. scolopes* (Belcaid et al., 2019; Schmidbaur et al., 2022). Moreover, it has been demonstrated that microsyntenies correspond to topological compartments with a distinct regulatory structure and contribute to complex expression patterns (Schmidbaur et al., 2022).

2.3. The role of chromatin organization in gene regulation

Understanding gene regulation in cephalopods requires a thorough studying of their chromatin organization, particularly in the context of their unique microsyntenies, since coleoid-cephalopod specific microsyntenies exhibit distinct genomic properties compared to metazoan microsyntenies. For example, they have smaller average sizes and differences in locality of ATAC peaks within microsyntenies, suggesting they have a distinct mode of gene regulation (Schmidbaur et al., 2022). 3C techniques such as Hi-C and also Micro-C provide valuable insights into chromatin organization. In many organisms, at the scale of multiple megabases, gene-rich transcriptionally active areas of chromatin interact more frequently, while gene-poor heterochromatic regions interact less frequently, resulting in A and B compartments. Topologically associating domains (TADs) are formed at a smaller scale (on average 2.5 Mb in *E. scolopes*, Schmidbaur et al. 2022). These are chromatin areas with a high contact frequency that are largely separated from neighboring regions. TADs and chromatin loops are produced by strong connections between certain genomic areas, such as CCCTC-binding factor (CTCF) and enhancer-promoter loops (Bonev & Cavalli, 2016; Rada-Iglesias et al., 2018; Rowley & Corces, 2018 as cited in Tena & Santos-Pereira, 2021). By

facilitating interactions between cis-regulatory elements (CREs) and target promoters, it believed TADs ensure coordinated gene expression (Galupa & Heard, 2017; Franke & Gomez-Skarmeta, 2018; Furlong & Levine, 2018 as cited in Tena & Santos-Pereira, 2021). Micro-C produces higher resolution chromatin interaction maps compared to Hi-C (up to 1Kb resolution compared to approximately 100Kb in the traditional HiC method). The differences in methodology being that Hi-C uses restriction enzyme to cut particular sequences, and Micro-C instead uses Micrococcal nuclease (MNase) to break down cross-linked DNA in areas of the genome where proteins are not stably attached (Lee et al., 2022). ATAC-seq is a complementary method used to study gene regulation. ATAC-seq identifies accessible chromatin regions involved in the upregulation of gene expression (Buenrostro et al., 2013). By combining ATAC-seq data with 3C methods, we can begin to untangle how chromatin accessibility and higher-order chromatin structures relate, in order to shed light on gene regulation in cephalopods with unique microsyntenies.

2.4 Aim of thesis

Many genomic features emerging after the genome rearrangement have been implicated to underlie the evolution of cephalopod novelties, but little is known the role that gene regulation plays in this. Previous work of Schmidbaur et al., (2022) generated Hi-C, RNA-seq and ATAC-seq data for *E. scolopes* and identified microsyntenies associated with cephalopod innovations (MACIs) and ancestral, metazoan microsyntenies in order to gain the first insights into gene regulation in a coleoid cephalopod. To build upon this previous research and understand how regulation shapes coleoid cephalopod traits, my thesis has three main aims. Firstly, to examine how the genomic structure of cephalopod-specific microsyntenies differs from metazoan microsyntenies using ATAC-seq data. Secondly, assess how 3D genome structure differs between these regions using high resolution genome topology data (Micro-C). Lastly, I extend the study beyond *E. scolopes*, by exploring the 3D organization of microsyntenies on gene regulation in the Californian two-spot octopus, (*Octopus bimaculoides*), to assess whether mechanisms influencing gene regulation are conserved across distantly related coleoid species. The genomic structure and spatial organization of microsyntenies influences interactions between regulatory elements and target genes and therefore these analyses will provide crucial information on cephalopod gene regulation. Overall, by combining these analyses, my thesis contributes to the research field of cephalopod evolution and the study of gene regulation in driving the evolution of complex phenotypes.

3. Material and Methods

3.1. Characterization of microsyntenies in coleoid cephalopods

Microsyntenies were previously detected by Schmidbaur et al., (2022). Schmidbaur et al., (2022) defined a microsyntenic block as three or more orthologous genes appearing in close proximity, with up to five intervening genes with no restrictions on the degree of collinearity. In contrast to just detecting syntenic regions that underwent local rearrangement and expansion, this definition produces fewer false-positive blocks (compared to just detecting pairs of genes). Microsynteny shared by *E. scolopes* and at least one octopus species were classified as MACIs. All syntenic blocks shared by at least seven additional species out of a set of twenty species were designated as metazoan syntenies. Metazoan syntenies date back to at least the last common bilaterian ancestor. MACIs were found to have complex expression patterns connected to a specific set of cephalopod neural tissues and other novel organs (Schmidbaur et al., 2022).

Cephalopod-specific and metazoan microsyntenies were examined by manually using the *Euprymna* genome browser which can be found at: http://metazoa.csb.univie.ac.at:8000/euprymna/jbrowse. This is based on the most recent and chromosomal-scale genome assembly (Schmidbaur et al., 2022) based on HiC scaffolds of the publicly available assembly (Belcaid et al., 2019) and its annotation. ATAC-seq data from two samples of stage 29 *E. scolopes* embryos from Schmidbaur *et al*. (2022) was also inspected on the genome browser.

For the examination of microsynteny regions, details were recorded on gene deserts, ATAC peak regions, gene numbers, neighbor gene numbers, to give insight into whether the genomic structure of MACIs and metazoan microsyntenies differed. In addition to this, genes which appeared to be misannotations were noted down to assist in future analyses. Representative figures for microsynteny types were created using Adobe Illustrator (Adobe Inc., 2022).

3.2. Gene desert analysis

Using the previously generated list of microsyntenies (Schmidbaur et al., 2022) and the Belcaid et al. (2019) gene annotation, several custom Python3 (Van Rossum and Drake, 2009) scripts for analyzing gene deserts were made. Firstly, the Belcaid et al. (2019) gene annotation file was modified to remove genes over 1Mb that were clearly misannotations. These long genes were spanning more than the length of the whole microsynteny and/or they were not annotated on the two most recent, improved gene annotations (“PASA” and “BRAKER'' annotations from *Euprymna* genome browser) and therefore it was possible to confirm these were most likely not real genes. Also, overlapping genes in the Belcaid et al.

(2019) annotation file were merged using the merge function from BEDtools v2.30.0 (Quinlan & Hall, 2010) to avoid problems with calculations in downstream analyses. The distance between genes within the microsynteny regions, including intervening genes, was then calculated using the first Python script.

A second Python script was used to output files of gene deserts of given sizes and calculate the percentage of MACIs and metazoan microsyntenies over the specified gene desert size that had gene deserts. Gene desert sizes were defined as the following thresholds: at least 300Kbp, 500Kbp and 700Kbp in size. The minimum gene desert size threshold was defined as 300Kbp because the *E. scolopes* genome is 5Gbp in size and contains 34,026 genes (according to Belcaid et al. 2019). This means that genes are spaced out 147Kbp on average. Therefore, a spacing of 300Kbp is quite unusual. Furthermore, the following relative sizes were used to classify gene deserts: at least 10%, 30% and 50% of the microsynteny size, in order to somewhat correct for the large sizes differences between MACIs and metazoan microsyntenies (on average 300Kbp and 700Kbp respectively, Schmidbaur et al. 2022).

To calculate the coverage and density of ATAC peaks, a third Python script was used which utilized the file of called ATAC-seq peaks of whole stage 29 (sample ID 97309) *E. scolope*s embryos from Schmidbaur et al. (2022) and the files of gene deserts with the thresholds 300 Kbp, 500 Kbp and 700 Kbp. With this Python script, any overlaps between ATAC peaks and gene deserts were checked. If there was an overlap, it calculated the percentage of the gene desert covered by ATAC peaks. As a result, the number and percentage of gene deserts that have ATAC peak coverage above or equal to the specified threshold were calculated and a file of ATAC peak coverage over the specified thresholds was produced. The thresholds were defined as at least 0.1%, 0.05%, and 0.02% for every category of gene desert. Again, the proportions of ATAC peak coverage were consistent across thresholds for each microsynteny type, so downstream analyses were done on all gene deserts with more than 0% percent ATAC coverage that were at least 300Kbp and at least 10% of the microsynteny size in order to maximize sample size and therefore increase power of the statistical tests. The output files of all Python scripts were manually checked on the genome browser to ensure no errors were made in the coding.

Statistical tests, boxplots and scatterplots were made using Rstudio v1.4.1717 (RStudio Team, 2021.). Boxplots were made using ggboxplot() function from the ggpubr package (Kassambara, 2023) and scatterplots were made using using plots(), points() and legend() functions to show the relationship between gene desert size and ATAC peak coverage for each microsynteny type. Shapiro-Wilk normality tests were applied to check if the datasets followed a normal distribution. None of the datasets followed a normal distribution ($p < 0.05$), and so unpaired two-sample Wilcoxon tests were performed to analyze the significance of the difference in gene desert size and ATAC-peak coverage for each microsynteny type.

3.3. Application of high resolution genome topology data (Micro-C) to evaluate 3D chromatin organization

Previously generated Micro-C data was used to assess the spatial location of microsyntenies and estimate what regions are important for cephalopod-specific gene regulation, and if and how cephalopod-specific regions of the genome differ from ancestral, metazoan microsyntenies. Micro-C provides an advantage over Hi-C data when studying genome topology as it has a higher resolution (up to 1Kb as opposed to up to ~100Kb in HiC) and an improved signal to noise ratio. Micro-C was carried out on flash-frozen tissues from ten, pooled stage 29 *E. scolopes* whole embryos and one whole stage 20 *O. bimaculoides* embryo. A Dovetail® Micro-C kit was used according to the manufacturer's protocol. Briefly, the steps were as follows: first chromatin was fixed, crosslinked, and digested with MNase, then the cells were lysed. Next, the chromatin ends were repaired and ligated to a biotinylated bridge, followed by proximity ligation of adapter-containing ends. After this, crosslinks were reversed, DNA was purified and lastly, libraries were prepared for sequencing (https://dovetailgenomics.com/products/micro-c-product-page/).

3.4. Generation of micro-C plots and documentation of microsynteny chromatin topology

Micro-C data was used to create contact maps of genome structure for *E. scolopes* and *O. bimaculoides*. Reads were previously mapped to the reference genomes (Belcaid et al., 2019; Albertin et al., 2022) using Hi-CPro v3.1.0 (Servant et al., 2015). When generating Micro-C contact maps, HicPlotter v0.8.1 was used (Akdemir & Chin, 2015).

HicPlotter was run with default parameters, along with the “-tri 1” parameter to output a triangular plot, ”-ptr 1” parameter to activate the plotting rotated half matrix format, “-hmc 5” parameter to choose heatmap color code: BlueToRed(5), “-sn 1” parameter to clean the noise in data, “-mm 5” parameter to set the interaction matrix heatmap scale upper-limit,”-ext pdf “ parameter to choose output file format, ”-s” parameter to arrange the starting bin (default :0),”-e” parameter to arrange the end bin (default : length of matrix), ”-r” parameter to arrange resolution of bins. In total, 34 plots were created for *E. scolopes*; four of the largest MACIs (between 1.85 and 3Mbp), five of the largest metazoan microsyntenies (from 3.7 to 4.96Mbp), six small MACIs (from 200 to 450Kbp), nineteen small metazoan microsyntenies (from 200 to 570Kbp). In *O. bimaculoides* 24 plots were created in total; four MACIs orthologous to the largest MACIs plotted in *E. scolopes* (from 200 to 450Kbp), two metazoan microsyntenies orthologous to the largest metazoan microsyntenies in *E. scolopes* (200Kbp and 480Kbp size), thirteen metazoan microsyntenies orthologous to the small metazoan microsyntenies in *E. scolopes* (from 69 to 230Kbp except one was just 21Kbp) and five MACIs orthologous to the small MACIs in *E. scolopes* (from 55 to 200Kbp). For microsynteny IDs, refer to Table 3. For the exact sizes of microsynteny regions, see

Supplementary Data 1 in Schmidbaur et al. (2022). Plots were created for selected MACIs and metazoan microsyntenies in *E. scolopes* with three resolutions 100K, 40K, and 20K, and in *O. bimaculoides* with three resolutions 20K, 10K, and 5K. These were the best resolutions given the same read coverage in each species because of their genome size differences (5.1Gbp and 2.7Gbp respectively).

The table was created by classifying microsynteny positions with respect to TADs within 5Mbp sized contact maps for *E. scolopes* at 100Kbp, 40Kbp, 20Kbp and for *O. bimaculoides* at 20Kbp, 10Kbp, 5Kbp for each microsynteny. The position defined as “on the TAD border” classifies microsyntenies overlapping the border of the TAD, “near the edge of the TAD” refers to the portion that does not cover the central 30% of the base pairs in the TAD, “overlapping a small TAD (< 1Mb)” correspond to the microsynteny overlapping with the TAD smaller than 1 Mbp. “in the middle of the TAD” classifies the microsynteny in the middle 70% of the TAD, lastly “spanning multiple TADs” suggest that it is not just a small part of a second TAD but rather it covers a significant portion of multiple TADs. The percentage calculation was then made for each condition for each species and each microsynteny type.

4. Results

4.1. Genomic structure of microsyntenies

To characterize the genomic structure of MACIs and metazoan microsyntenies in *E. scolopes*, they were first examined by eye on the *Euprymna* genome browser. For each frequently observed genomic pattern (see examples Figure 2A-C) representative figures were created (Figure 2D-F). Two distinct types of MACI were frequently observed. The first type of MACI (Figure 2D) had clustered genes at the edges of the microsynteny, gene deserts in the center and some ATAC peaks. The second type of MACI (Figure 2E) also showed huge regions of gene deserts in the center with lots of ATAC peaks and fewer genes present at the edges of microsynteny. This was in contrast to the microsyntenies present across metazoans. Ancestral microsyntenies did not show a certain clear pattern but they tended to have genes spanning the whole length of them and fewer ATAC peaks than MACIs, however they still frequently had gene deserts.

Figure 2. Images from genome browser and representative figures of MACI and metazoan microsynteny types. These screenshots (A-C) and schematics (D-F) show commonly observed microsynteny types as seen on the *E. scolopes* genome browser. Figures A, B, D and E show MACIs, displaying ATAC peaks and genes overlapping with these regions. Figure C and F highlights metazoan microsynteny with ATAC peaks and overlapping genes. The first type (A and D) of frequently observed MACI show less ATAC peaks and clustered genes on either side of the microsynteny region. The second type (B and E) shows fewer genes clustered either side of the microsynteny and more ATAC peaks. Lastly, the metazoan microsynteny (F and C) shows some gene deserts, but genes spanning the whole length of the microsynteny and the fewest ATAC peaks.

4.2. Results of gene desert analysis

Next, the observed pattern of microsynteny structure was examined statistically. Multiple thresholds were used to classify a region as a gene desert. A large number of gene deserts were found in both MACIs and metazoan syntenies (>75% of syntenies over the given desert size had gene deserts over 300Kb, 500Kb and 700Kb that were over 10% or 30% of the total microsynteny size, Table 1). However, proportions of gene deserts in each microsynteny type were broadly consistent among different thresholds, with metazoan syntenies having between 0 to 12% more gene deserts than MACIs (Table 1). Therefore the next analyses were done on gene deserts that were at least 300Kbp and 10% of microsynteny size. For most gene desert size thresholds, gene deserts in MACIs had higher ATAC-peak coverage than gene deserts of metazoan microsyntenies (Table 2). Additionally, the pattern of MACIs having a higher ATAC peak coverage is mostly consistent regardless of the ATAC peak threshold given.

Table 1. Numbers of gene deserts classified with different size limits in MACI and metazoan microsyntenies of *E. scolopes*.

Minimum gene desert size (bp)	Relative percentage of microsynteny size	Number of gene deserts (MACI)	Number of gene deserts (metazoan)
300000	10	94 (100%)	148 (100%)
	30	56 (80%)	70 (79.8%)
	50	21 (32.3 %)	33 (39.2%)
500000	10	52 (100%)	102 (100%)
	30	35 (82%)	50 (80%)
	50	13 (33%)	27 (45%)
700000	10	28 (100 %)	70 (100%)
	30	23 (95%)	39 (80.4%)
	50	8 (38%)	23 (50%)

Note. As mentioned in the 3.2, percentage result only reflects the proportion of microsyntenies out of the total microsyntenies that meet the base pair size criteria for gene deserts, rather than all of the microsyntenies in Schmidbaur et al. (2022).

Table 2. ATAC peak coverage in different gene desert sizes for MACI and metazoan microsyntenies of *E. scolopes.*

Minimum gene desert size (bp)	Relative percentage of microsynteny	Minimum ATAC peak coverage (%)	Number and percentage of gene deserts with ATAC peak coverage (%) over threshold (MACI)	Number and percentage of gene deserts with ATAC peak coverage (%) over threshold (metazoan)
300000	10	0.02	68 (72.4%)	100 (67.6 %)
		0.05	55 (58.6%)	74 (50 %)
		0.1	36 (38.3%)	50 (33.8 %)
	30	0.02	39 (69.7%)	49 (70 %)
		0.05	31 (55.4%)	35 (50 %)
		0.1	17 (30.4%)	23 (32.9 %)
	50	0.02	15 (71.5%)	22 (66.7 %)
		0.05	11 (52.4%)	15 (45.5 %)
		0.1	0	0
500000	10	0.02	39 (75%)	70 (68.7 %)
		0.05	28 (53.9%)	48 (47.06%)
		0.1	17 (32.7%)	33 (32.4%)
	30	0.02	26 (74.3%)	37 (74%)
		0.05	18 (51.5%)	25 (50%)
		0.1	10 (28.6%)	17 (34%)
	50	0.02	11 (84.7%)	19 (70.4%)
		0.05	7 (53.9%)	14 (51.9%)
		0.1	5 (38.5%)	10 (37.04%)
700000	10	0.02	22 (78.6%)	47 (67.2%)
		0.05	15 (53.6%)	31 (44.3%)
		0.1	8 (28.6%)	20 (28.6 %)
	30	0.02	18 (78.3%)	27 (69.3%)
		0.05	12 (52.2%)	18 (46.2%)
		0.1	7 (30.5%)	11 (28.3%)
	50	0.02	7 (87.5%)	15 (65.3%)
		0.05	4 (50%)	10 (43.5%)
		0.1	4 (50%)	6 (26.1%)

Boxplots of gene desert size (bp) (Figure 3A) and ATAC peak coverage (%) (Figure 3B) were made for the comparison between MACI and metazoan microsynteny. The boxplots of gene desert size (Figure 3A) suggest that metazoan microsyntenies exhibit a wider range of gene desert sizes and a higher median desert size compared to MACIs. The ATAC peak coverage per gene desert boxplot (Figure 3B) shows the inverse result, as MACIs had a much wider range of ATAC peak coverage and a higher median coverage than metazoan microsyntenies. Overall, metazoan syntenies have significantly larger gene deserts (Wilcoxon test, $p = 0.004$,

Figure 3A), whereas MACIs have marginally non-significant higher ATAC peak coverage (p = 0.05, Figure 3B). The scatterplot (Figure 3C) shows that as gene desert size increases, the percent ATAC peak coverage decreases.

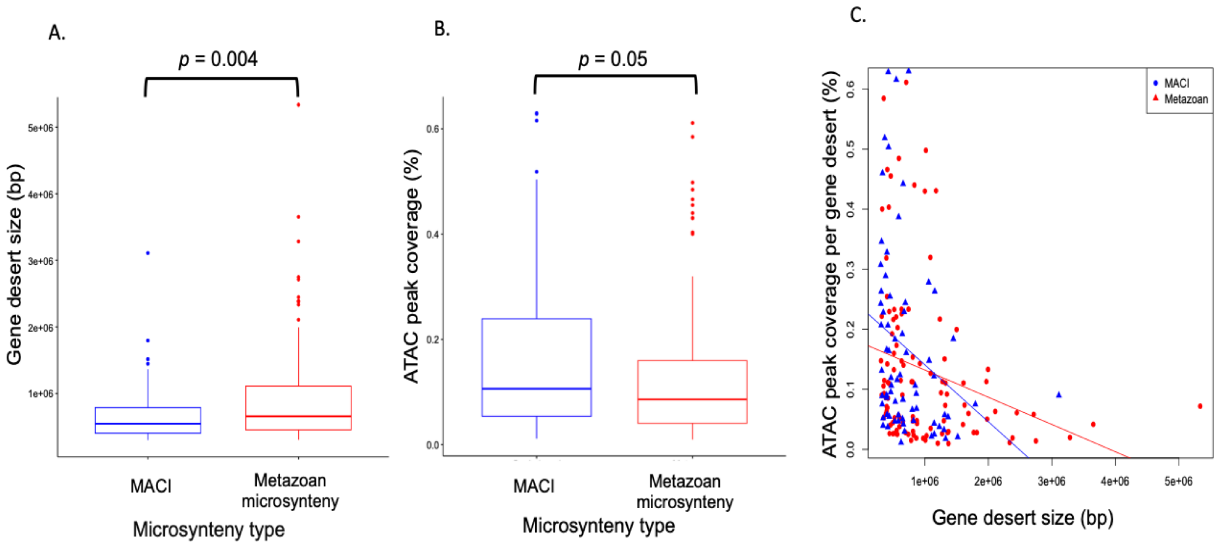

Figure 3. Comparative analysis of gene deserts and ATAC peaks in MACI and metazoan microsyntenies. (A) Boxplots of gene desert size and (B) ATAC peak coverage in gene deserts that have > 0% ATAC peak coverage for each microsynteny type. The p values displayed are for unpaired Wilcoxon tests. (C) Scatterplot of gene desert size against ATAC peak coverage (> 0%). Note in all three plots, gene deserts are defined as at least 300Kb and at least 10% of the microsynteny size.

4.3. Genomic location of microsyntenies in squid and octopus

In order to gain a better understanding of the chromatin structure of MACIs and the impact of this on cephalopod gene regulation, contact maps of genome structure spanning microsynteny regions in *E. scolopes* and *O. bimaculoides* were made using previously generated Micro-C data.

Micro-C data for *E. scolopes* performed sufficiently up to 20Kb when zoomed in, with the best resolution with adequate coverage achieved at 40Kb (Figure 4A). On the other hand, the data for *O. bimaculoides* showed good results when zoomed in up to 5Kb, while the best resolution given the coverage is found at 10Kb (Figure 4B). MACIs were frequently found on the TAD borders in *E. scolope*s (Figure 4A, 5B, Table 3). Furthermore, a similar pattern was seen in *O. bimaculoides* (Figure 4B, 5B, Table 3). Plots of *E. scolopes* metazoan microsyntenies and their orthologs in *O. bimaculoides* (Figure 6, Table 3) generally did not show the same patterning as the MACIs. In both species, metazoan microsyntenies were often observed in TADs but this pattern was not consistent across the majority of the metazoan syntenies checked, and so no clear pattern appeared.

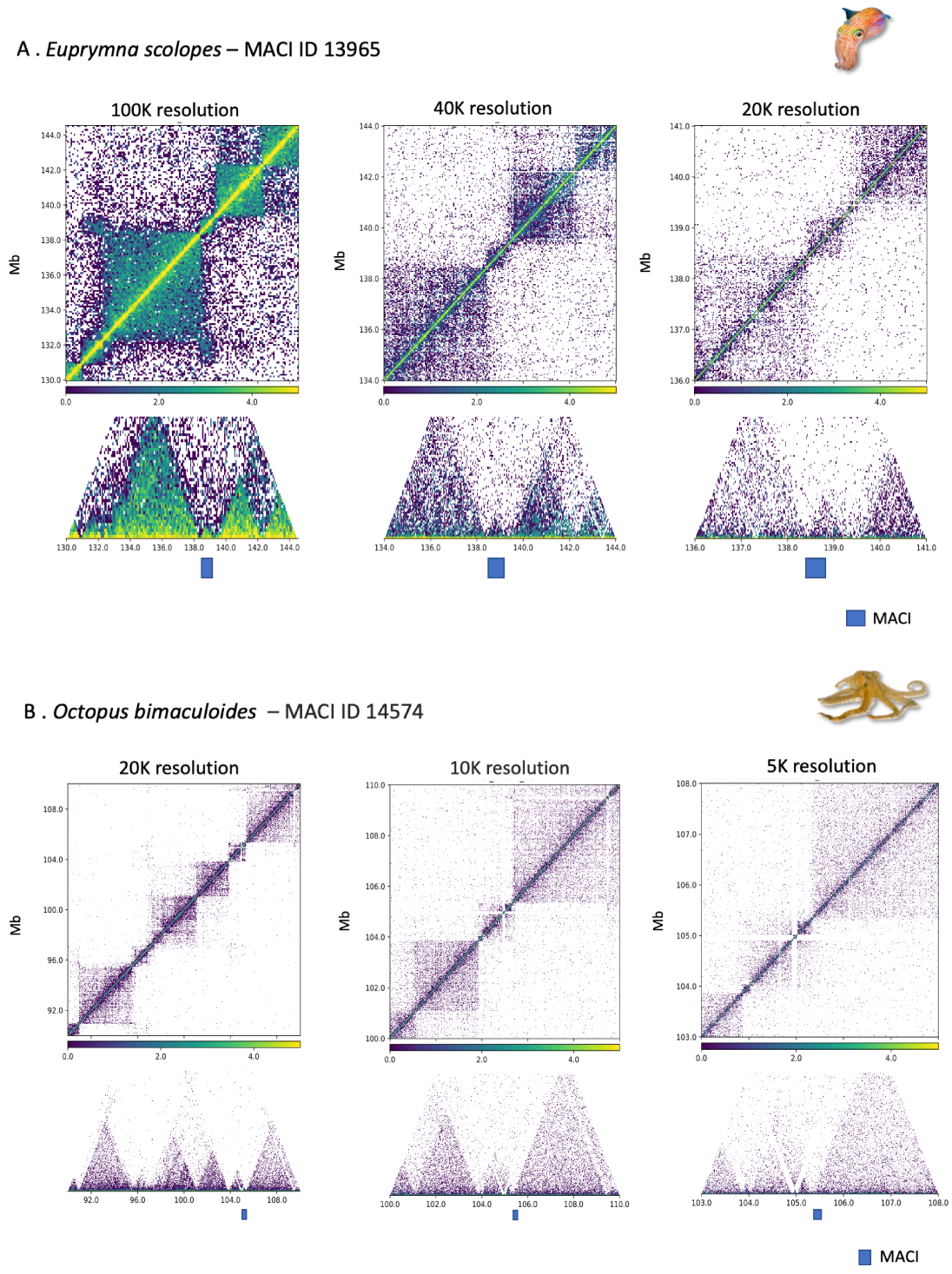

Figure 4. Contact maps that clearly show the quality of various resolutions of Micro-C data for *E. scolopes* and *O. bimaculoides.* (A) Micro-C plots for a region on *E. scolopes* chromosome 10 at 100Kb, 40Kb and 20Kb resolution. MACI 13965 (Schmidbaur et al. 2022 is labeled. (B) Micro-C plots for a region of *O. bimaculoides* chromosome 7 at 20Kb, 10Kb and 5Kb resolution. MACI 14574 (Schmidbaur et al. 2022) is labelled. Given the genome size difference and resulting read coverage in both species, 20Kb and 5Kb is approximately the resolution limit for the *E. scolopes* and *O. bimaculoides* Micro-C data respectively.

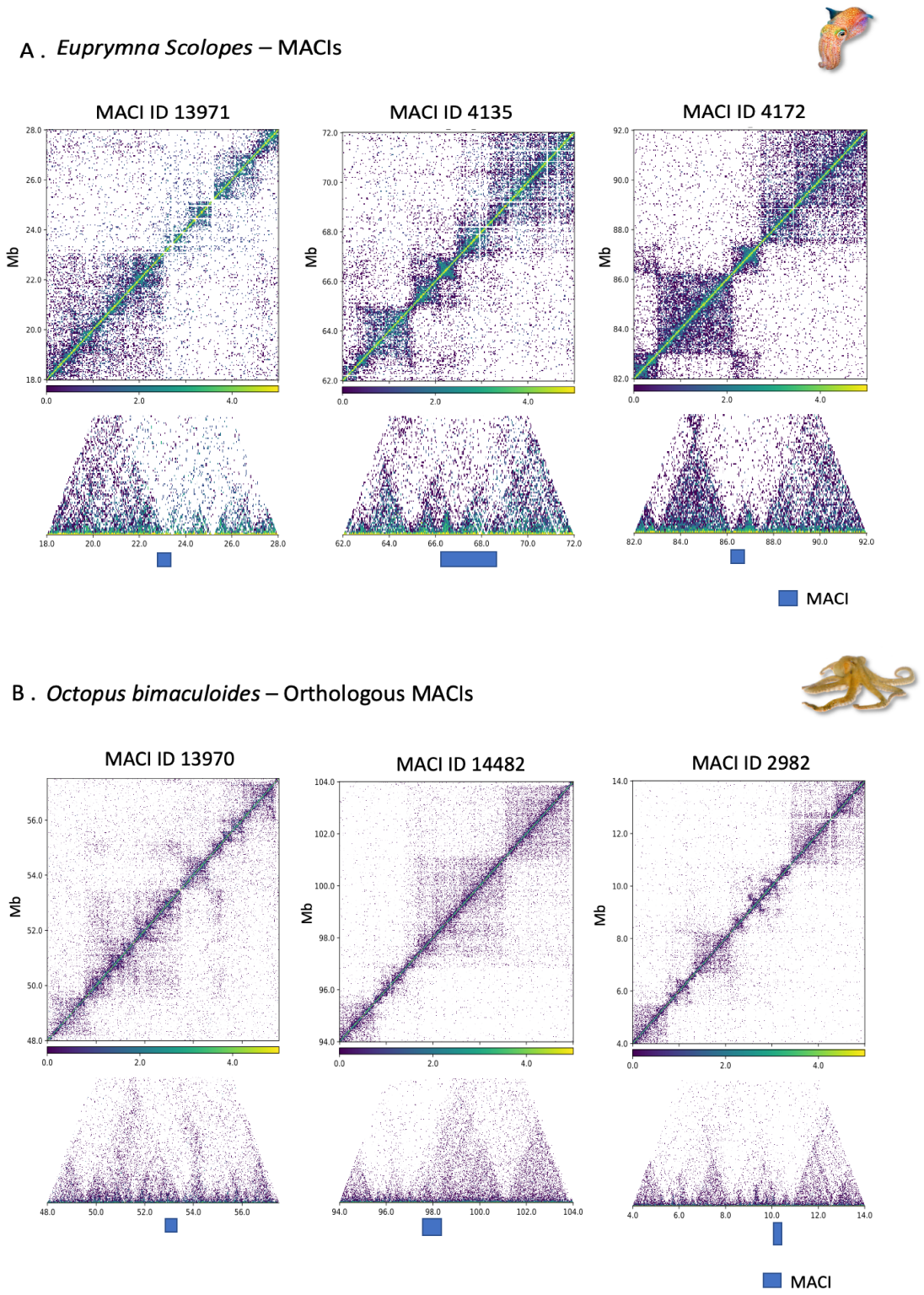

Figure 5. Contact maps of selected MACIs in *E. scolopes* and their orthologs in *O. bimaculoides*. (A) Micro-C plots of *E. scolopes* MACI 13971 on chromosome 43, MACI 4135 on chromosome 14 and MACI 4172 on chromosome 3, all at 40Kb resolution. (B) Micro-C plot of orthologous MACIs in *O. bimaculoides* MACI 13970 on chromosome 16, MACI 14482 on chromosome 7, MACI 2982 on chromosome 13, all at 10Kb resolution. The MACIs above and below each other are orthologous. MACI IDs from Schmidbaur et al. (2022).

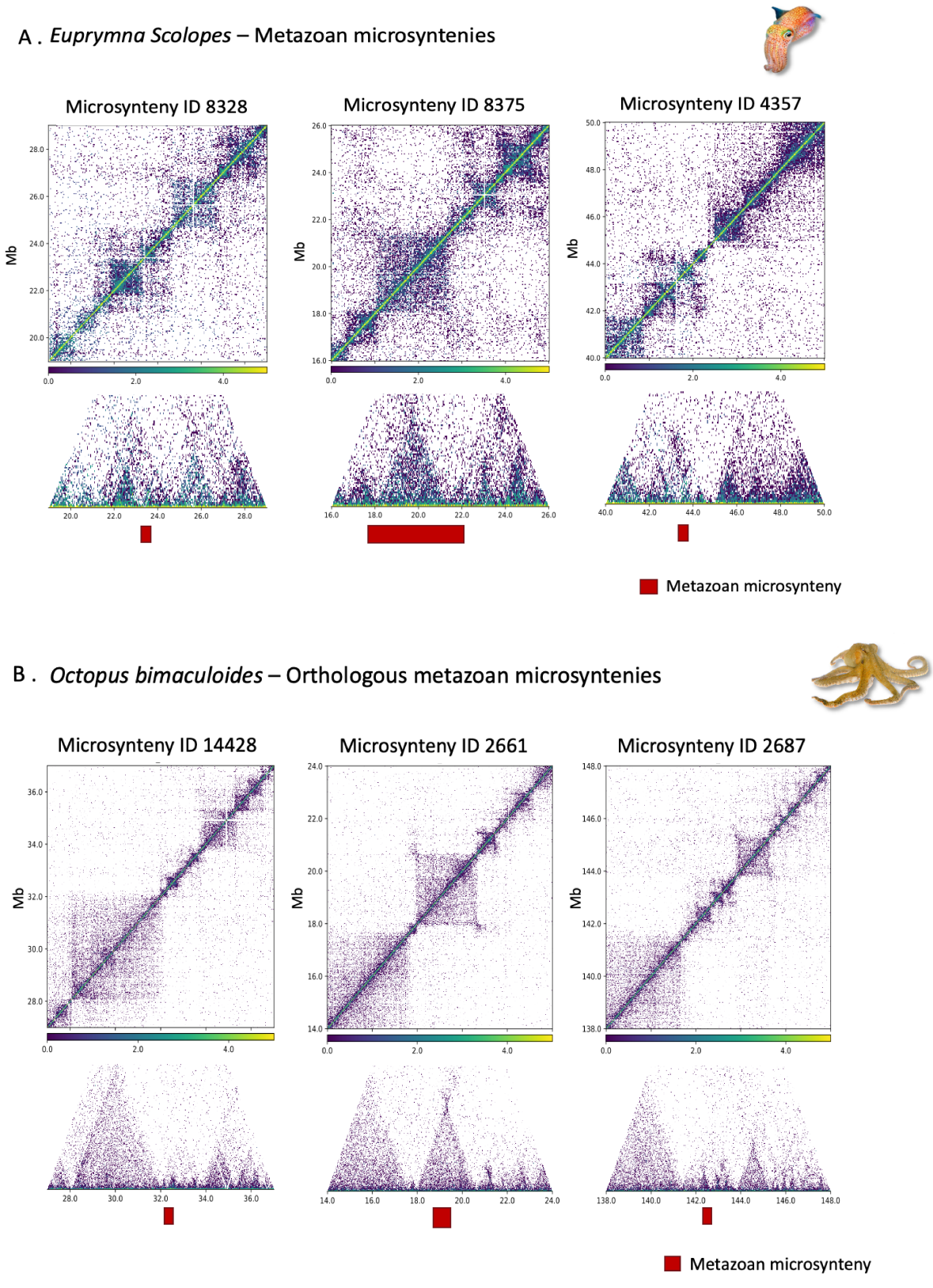

Figure 6. Contact maps of selected metazoan microsyntenies in *E. scolopes* and their orthologs in *O. bimaculoides*. (A) Micro-C plots of *E. scolopes* metazoan microsynteny 8328 on chromosome 26, microsynteny 8375 on chromosome 30 and microsynteny 4357 on chromosome 15, all at 40Kb resolution. (B) Micro-C plot of orthologous metazoan microsyntenies in *O. bimaculoides* metazoan microsynteny 14428 on chromosome 21, microsynteny 2661 on chromosome 1, microsynteny 2687 on chromosome 3, all at 10Kb resolution. The metazoan microsyntenies above and below each other are orthologous. Metazoan microsynteny IDs from Schmidbaur et al. (2022).

Table 3 displays the summary of outcomes of the contact maps made for orthologous MACI and metazoan syntenies in *E. scolopes* and *O. bimaculoides.* To conduct the analysis, the percentage of microsyntenies that matched specific conditions were calculated and each condition was evaluated separately. Importantly, each microsynteny region can match multiple conditions, which results in a detailed understanding of the data.

All *E. scolopes* MACIs checked were located “on the TAD border” and a substantial proportion were located “spanning multiple TADs” (40%). Also, a small proportion were located “overlapping a small TAD” (10%) (Table 3). However, none of them were located “in the middle of the TAD” or “near the edge of the TAD”. For orthologous MACIs in *O. bimaculoides*, “on the TAD border” (55.5%) also accounts for a substantial portion of microsyntenies. Also, the second most common occurrence was “near the edge of the TAD” (44.4%) and a considerable portion of microsyntenies were found “overlapping a small TAD” (22.3%). None were found “in the middle of the TAD” or “spanning multiple TADs”. For *E. scolopes*, most metazoan microsyntenies were located either “near the edge of the TAD” (37.5%), and secondly “in the middle of the TAD” (25.0%). After this, they were located “on the TAD border” (20.8%) and “spanning multiple TADs” (20.8%), then least commonly they were located “overlapping a small TAD” (12.5%). When their metazoan orthologs in *O. bimaculoides* were considered Table 3 showed that a similar distribution occurred for “on the TAD border” (33.4%), “in the middle of the TAD” (33.4%) and “near the edge of the TAD” (33.4%), followed by “overlapping a small TAD” (26.7%), while table showed that fewer instances was found “spanning multiple TADs” (6.7%).

Overall, the results show that the arrangement of microsyntenies in TAD-associated areas varies between MACI and metazoan microsynteny. The strongest difference was for the condition located “on the TAD border” between MACI and metazoan microsynteny. Additionally *E. scolopes,* MACIs and their orthologs in *O. bimaculoides* showed mostly similar patterns, but this was not observed for *E. scolopes* metazoan microsyntenies and their orthologs in *O. bimaculoides.*

Table 3. Positions of microsyntenies in contact maps considering TAD structures. The percentages in the last two rows of the table are calculated individually for each condition out of total MACI and metazoan microsyntenies checked in each species. More than one condition could be matched for each microsynteny. Crosses indicate there is no orthologous microsynteny in *O. bimaculoides,* only *Callistoctopus minor* for which micro-C data is currently unavailable for. Conditions were checked using 5Mbp contact maps in *E. scolopes* at resolutions 100Kb, 40Kb and 20Kb and for *O. bimaculoides* at resolutions 20Kb, 10Kb and 5Kb. For more details on microsynteny regions such as the size, chromosomal location and gene content, refer to Supplementary Data 1 in Schmidbaur et al. (2022).

Synteny type	Species	*E. scolopes* microsynteny ID	Species of orthologous microsyntenies	*O. bimaculoides* microsynteny ID	On the TAD border	Near the edge of the TAD	Overlapping a small TAD (< 1Mb)	In the middle of the TAD	Spanning multiple TADs
MACI	*E. scolopes*	4135	*O. bimaculoides*	14482	✓	✓			✓
	E. scolopes	4338	*O. bimaculoides*	3075	✓✓				✓
	E. scolopes	13984	*O. bimaculoides*	13983	✓	✓			✓
	E. scolopes	14023	*O. bimaculoides*	14022	✓✓				✓
	E. scolopes	4172	*O. bimaculoides*	2982	✓✓		✓		
	E. scolopes	4436	*O. bimaculoides*	X	✓				
	E. scolopes	13863	*O. bimaculoides*	13864	✓	✓			
	E. scolopes	13965	*O. bimaculoides*	13966	✓	✓	✓		
	E. scolopes	14573	*O. bimaculoides*	14574	✓✓		✓		
	E. scolopes	13971	*O. bimaculoides*	13970	✓✓				
Metazoan	*E. scolopes*	4420	*O. bimaculoides*	2849	✓		✓		
	E. scolopes	4382	*O. bimaculoides*	X	✓	✓			
	E. scolopes	8317	*O. bimaculoides*	X					✓
	E. scolopes	8330	*O. bimaculoides*	8155		✓			✓
	E. scolopes	8352	*O. bimaculoides*	X	✓				
	E. scolopes	8375	*O. bimaculoides*	2661	✓			✓	✓
	E. scolopes	6897	*O. bimaculoides*	X				✓	
	E. scolopes	1208	*O. bimaculoides*	X			✓		
	E. scolopes	9056	*O. bimaculoides*	2634			✓	✓	
	E. scolopes	14950	*O. bimaculoides*	X		✓			
	E. scolopes	12183	*O. bimaculoides*	X				✓	
	E. scolopes	9076	*O. bimaculoides*	X		✓			
	E. scolopes	13921	*O. bimaculoides*	3551		✓		✓	
	E. scolopes	6905	*O. bimaculoides*	X				✓	
	E. scolopes	8328	*O. bimaculoides*	14428	✓	✓	✓		✓
	E. scolopes	13835	*O. bimaculoides*	13070	✓			✓	
	E. scolopes	4341	*O. bimaculoides*	3945		✓✓			
	E. scolopes	4357	*O. bimaculoides*	2687	✓✓			✓	✓
	E. scolopes	6892	*O. bimaculoides*	6597		✓		✓	
	E. scolopes	9044	*O. bimaculoides*	8944	✓	✓	✓		
	E. scolopes	1204	*O. bimaculoides*	14577		✓		✓	
	E. scolopes	4234	*O. bimaculoides*	2532		✓✓			✓
	E. scolopes	4442	*O. bimaculoides*	3296	✓		✓	✓	
	E. scolopes	14190	*O. bimaculoides*	14189		✓	✓		
				Total MACI	**100% ;55.5%**	**0% ; 44.4%**	**10% ; 22.3%**	**0% ; 0%**	**40% ; 0%**
				Total metazoan	**20.8% ;33.4%**	**%37.5 ;33.4%**	**12.5% ;26.7%**	**25% ; 33.4%**	**20.8% ; 6.7%**

5. Discussion

5.1. Gene deserts in MACIs may act as novel regulatory regions

Previously, it was found that novel cephalopod microsyntenies differ from ancestral, metazoan microsyntenies based on size, enrichment of functional categories and their genomic properties (Schmidbaur et al., 2022). For instance, in MACIs, ATAC-seq peaks were revealed to be located more frequently in the introns of neighboring genes within the same microsyntenic cluster compared to metazoan microsyntenies. Also, weaker co-expression of genes in MACIs, compared to the metazoan microsyntenies was revealed. These findings

lead to the hypothesis that cephalopod-specific regions of the genome evolve as genomic regulatory blocks (GRBs), or more specifically, they evolve under the gene bystander model (Schmidbaur et al., 2022), which has also been found in *Drosophila* and fish genomes (Krefting et al., 2018; Dimitrieva & Bucher, 2012; Engström et al., 2007). Our study builds upon these previous findings and further explores the role of MACIs in shaping the evolution of novel gene regulation.

Manual examination of the genomic structure of microsyntenies on the genome browser revealed two commonly occurring MACI 'types' that were distinct from metazoan microsynteny, providing further evidence that cephalopod-specific genomic regions might be regulated differently from ancestral regions of the genome (Figure 2). Statistical analysis also revealed a large proportion of both microsynteny types had gene deserts (Table 1), but metazoan syntenies had a higher proportion of and significantly larger gene deserts than MACIs (Table 1, Figure 3A). It should be noted that the reason that we observed overall larger gene deserts in metazoan microsyntenies is likely because metazoan microsyntenies are significantly larger than MACIs but tend to have a similar gene density, as is known from the previously published research (Schmidbaur et al., 2022). Another reason why metazoan microsyntenies have a higher proportion of gene deserts and larger gene deserts could be because metazoan syntenies are older (at least 600 MYA as opposed to 270 MYA for MACIs, Irimia et al., 2012, Schmidbaur et al, 2022). Therfore they may have undergone more repetitive element expansion, which is known to be one of the primary reasons for genome expansion in cephalopods (Albertin & Simakov, 2020; Albertin et al., 2015, 2022). Despite ancestral microsyntenies having larger and proportionally more gene deserts, MACIs had more ATAC peak coverage within gene deserts (Table 2, Figure 3B, C). It is possible that in metazoans, as gene deserts become larger and regulatory elements move farther away, gene expression regulation can become increasingly difficult. The remote positioning of regulatory elements may result in decreased accessibility and interaction with target genes, making it more difficult for transcription factors and other regulatory proteins to bind to distant enhancers. This diminished contact may result in lower ATAC peak counts in metazoan gene deserts compared to more closely positioned regulatory elements in MACIs (Engström et al., 2007). The presence of a large gene deserts that are conserved and contain many open chromatin regions as shown by the presence of many ATAC peaks, suggests that MACIs are under regulatory constraint and may play essential roles in transcription. Consistent with this, studies in humans show that some gene deserts include regulatory sequences that affect the expression of nearby genes over long distances (Nobrega et al., 2003; Ovcharenko et al., 2004).

5.2. MACIs located on the TAD border may be indicative of distinct regulatory patterns

Our analysis of chromatin organization provides valuable insights into the distribution and positioning of microsyntenies in the 3D genome. The occurrence or non-occurrence of specific conditions in each microsynteny type may impact the preservation of gene order and interactions between regulatory elements and genes throughout evolution. MACIs showed a specific pattern of 3D organization regardless of their size and this was generally conserved across *E. scolopes* and *O. bimaculoides* (Figure 4, 5, Table 3). For metazoan microsyntenies in both species our results reveal that there is no clear pattern, and they do not have the same pattern as MACIs (Figure 6, Table 3). The most repeated pattern that we observed for the MACIs was that they are on the border of the TAD, secondly, they were located near the edge of TADs, and thirdly, they spanned multiple TADs (Figure 4, 5, Table 3) (averaged across species). The reason for this might be that the ancestral regions are located in more conserved parts of the genome. There are some studies that show selection acts differently on TADs vs TAD boundaries. TAD structures are often conserved while the borders are under less selective constraint (Krefting et al., 2018; Harmston et al., 2017; Eres & Gilad, 2021; Preger-Ben Noon & Frankel, 2022). This could also be why metazoan syntenies are larger than MACIs and show larger gene deserts that are not broken down by recombination. Conversely, it is possible that while metazoan TADs are more deeply conserved, MACI TADs split into multiple TADs, and regulatory regions on the TAD borders gain access to genes in both neighboring TADs (Rogers & Simakov, *in review*; Rouressol et al., 2023). This pattern is consistent with the patterns we see of the putative regulatory regions (ATAC peaks) being located closer to the center of microsyntenies, whereas the genes are clustered at either ends (Figure 2A, B). Consistent with MACIs having important regulatory functions, TAD boundaries are known to be important during development, and their deletion not only affects genomic regions in the immediate vicinity but also the operation of genes and regulatory elements in neighboring TADs (Rajderkar et al., 2023). This pattern may also be consistent with the gene bystander model hypothesized for MACIs in the previous research (Schmidbaur et al., 2022) because in this model genes are not necessarily coexpressed and therefore may be present in different regulatory domains (TADs), but with one connecting regulatory region (in this case on the TAD border).

It is worth noting that we cannot effectively compare the differences in each condition found between *E. scolopes* and *O. bimaculoides* because of differences in read coverage and resulting data quality between species meaning that 3D location of each microsynteny was assessed at a different Kb resolution. It was possible to look more closely at the 3D organization of microsyntenies in *O. bimaculoides* and still have a clear picture of genome structure. Consequently, this could be the reason for the lower result of the most common patterns that we observed in *O. bimaculoides* compared to *E. scolopes.* However, the relative

abundance of each condition is generally consistent across species except for conditions “spanning multiple TADs” and “near the edge of the TAD”.

5.3. Selective constraints on regulatory evolution may be different for each microsynteny type

Since some studies that show selection acts differently on TADs vs TAD boundaries (Krefting et al., 2018; Harmston et al., 2017; Eres & Gilad, 2021), certain microsynteny types may have more conserved regulatory regions, whereas others may have experienced more evolutionary flexibility due to less selective constraints. As a result, MACIs may be under more positive selection or relaxed purifying selection (genetic drift) at the regulatory (transcriptional) level, and so reside in less conserved areas of the genome, such as the TAD boundaries. In accordance with the positive selection hypothesis, it is possible that the environmental changes or challenges that coleoid cephalopods have experienced may have caused them to undergo rapid adaptations under selective pressure, with neural complexity enabling these adaptations and flexibility (Amodio et al., 2019). Or, perhaps MACIs are simply under less purifying selection than the highly conserved metazoan syntenies that may play a crucial, core, ‘housekeeping’ role across the coleoids and beyond. This is consistent with the gene ontology results found in the previous research (Schmidbaur et al. 2022).

5.4. Future directions

In future investigations, obtaining ATAC-seq data from other coleiod cephalopods as well as more distantly related mollusks such as nautilus would help confirm if the patterns seen in Figure 2 and 3 are conserved across just the coleoids, helping us understand whether this is really a clade-specific mode of gene regulation. Additionally, tissue-specific ATAC-seq and Micro-C data could help further dissect regulatory evolution in the coleoids by identifying regions important in tissue-specific gene regulation that might be missed when carrying out investigations on whole embryos. The functional impact of MACIs and their effect on gene regulation could also be studied further using advanced genomic techniques, such as knockout experiments using CRISPR-Cas9. This technique was recently successfully applied to coleiod cephalopods (Crawford et al., 2020; Ahuja et al., 2023). Using this method, hypotheses related to microsyntenies located on the TAD borders, and the gene bystander model could be tested. For example, if we knocked out a MACI (gene or non-coding region) on a TAD border, we might expect to observe mutations which could affect gene expression in neighboring TADs associated with cephalopod innovations. If the gene bystander model for MACIs is correct, we might expect a knockout to affect expression of neighboring genes but not the expression of the gene where the regulatory element was located. Statistical modeling could also be applied to test for regulatory selection on each microsynteny type, however this would require transcriptomic data from many different cephalopod species (methods reviewed in Price et al., 2022).

5.5. Conclusion

Our findings showed that coleiod cephalopods have distinct 'types' of microsynteny that differ from ancestral, metazoan microsyntenies in their genomic structure and gene regulation. Additionally, cephalopod microsyntenies, unlike metazoan microsyntenies, tend to be situated on the TAD borders. Moreover, these differences in regulatory architecture may imply that selective constraints differ between MACIs and metazoan microsyntenies. Our findings suggest MACIs display a unique 'mode' of gene regulation compared to ancestral, metazoan microsyntenies, consistent with previous research, and provide an additional understanding of the evolutionary processes that shape coleoid cephalopod-specific novel traits.

6. References

Acemel, R. D., & Lupiáñez, D. G. (2023). Evolution of 3D chromatin organization at different scales. Current Opinion in Genetics Development, 78, 102019. https://doi.org/10.1016/j.gde.2022.102019

Adobe Inc. (2022). Adobe Illustrator. Retrieved from https://adobe.com/products/illustrator

Ahuja, N., Hwaun, E., Pungor, J. R., Rafiq, R., Nemes, S., Sakmar, T., Vogt, M. A., Grasse, B., Diaz Quiroz, J., Montague, T. G., Null, R. W., Dallis, D. N., Gavriouchkina, D., Marletaz, F., Abbo, L., Rokhsar, D. S., Niell, C. M., Soltesz, I., Albertin, C. B. & Rosenthal, J. J. C. (2023). Creation of an albino squid line by CRISPR-Cas9 and its application for in vivo functional imaging of neural activity. *Current Biology: CB*, *33*(13), 2774–2783.e5. https://doi.org/10.1016/j.cub.2023.05.066

Akdemir, K. C. & Chin, L. (2015). HiCPlotter integrates genomic data with interaction matrices. *Genome Biology*, *16*(1), 198. https://doi.org/10.1186/s13059-015-0767-1

Albertin, C. B., & Simakov, O. (2020). Cephalopod Biology: At the Intersection Between Genomic and Organismal Novelties. Annual Review of Animal Biosciences, 8(1), 71–90. https://doi.org/10.1146/annurev-animal-021419-083609

Albertin, C. B., Simakov, O., Mitros, T., Wang, Z. Y., Pungor, J. R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C. W. & Rokhsar, D. S. (2015). The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature*, *524*(7564), 220–224. https://doi.org/10.1038/nature14668

Albertin, C. B., Medina-Ruiz, S., Mitros, T., Schmidbaur, H., Sanchez, G., Wang, Z. Y., Grimwood, J.,Rosenthal, J. J., Ragsdale, C. W., Simakov, O., & Rokhsar, D. S. (2022). Genome and transcriptome mechanisms driving cephalopod evolution. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-29748-w

Allcock, A. L., Lindgren, A. & Strugnell, J. M. (2015). The contribution of molecular data to our understanding of cephalopod evolution and systematics: a review. *Journal of Natural History*, *49*(21-24), 1373–1421. https://doi.org/10.1080/00222933.2013.825342

Amodio, P., Boeckle, M., Schnell, A. K., Ostojíc, L., Fiorito, G. & Clayton, N. S. (2019). Grow Smart and Die Young: Why Did Cephalopods Evolve Intelligence? Trends in Ecology & Evolution, 34(1), 45–56. https://doi.org/10.1016/j.tree.2018.10.010

Belcaid, M., Casaburi, G., McAnulty, S. J., Schmidbaur, H., Suria, A. M., Moriano-Gutierrez, S., Pankey, M. S., Oakley, T. H., Kremer, N., Koch, E. J., Collins, A. J., Nguyen, H., Lek, S., Goncharenko-Foster, I., Minx, P., Sodergren, E., Weinstock, G., Rokhsar, D. S., McFall-Ngai, M., … Nyholm, S. V. (2019). Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. *Proceedings of the National Academy of Sciences*, *116*(8), 3030–3035. https://doi.org/10.1073/pnas.1817322116

Bonev, B. & Cavalli, G. (2016). Organization and function of the 3D genome. *Nature Reviews. Genetics*, *17*(11), 661–678. https://doi.org/10.1038/nrg.2016.112

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods,* *10*(12), 1213–1218. https://doi.org/10.1038/nmeth.2688

Crawford, K., Diaz Quiroz, J. F., Koenig, K. M., Ahuja, N., Albertin, C. B. & Rosenthal, J. J. C. (2020). Highly Efficient Knockout of a Squid Pigmentation Gene. *Current Biology: CB*, *30*(17), 3484–3490.e4. https://doi.org/10.1016/j.cub.2020.06.099

Dimitrieva, S., & Bucher, P. (2012). Genomic context analysis reveals dense interaction network between vertebrate ultraconserved non-coding elements. *Bioinformatics*, *28*(18), i395–i401. https://doi.org/10.1093/bioinformatics/bts400

Dovetail genomics Micro-c user guide. (2021, 19. August). Dovetail Genomics. https://dovetailgenomics.com/products/micro-c-product-page/

Engström, P. G., Ho Sui, S. J., Drivenes, Ø., Becker, T. S., & Lenhard, B. (2007). Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Research*, *17*(12), 1898–1908. https://doi.org/10.1101/gr.6669607

Eres, I. E., & Gilad, Y. (2021). A tad skeptic: Is 3D genome topology conserved? *Trends in Genetics*, *37*(3), 216–223. https://doi.org/10.1016/j.tig.2020.10.009

Franke, M. & Gómez-Skarmeta, J. L. (2018). An evolutionary perspective of regulatory landscape dynamics in development and disease. *Current Opinion in Cell Biology*, *55*, 24–29. https://doi.org/10.1016/j.ceb.2018.06.009

Furlong, E. E. M. & Levine, M. (2018). Developmental enhancers and chromosome topology. *Science*, *361*(6409), 1341–1345. https://doi.org/10.1126/science.aau0320

Galupa, R. & Heard, E. (2017). Topologically Associating Domains in Chromosome Architecture and Gene Regulatory Landscapes during Development, Disease, and Evolution. *Cold Spring Harbor Symposia on Quantitative Biology*, *82*, 267–278. https://doi.org/10.1101/sqb.2017.82.035030

Harmston, N., Ing-Simmons, E., Tan, G., Perry, M., Merkenschlager, M., & Lenhard, B. (2017). Topologically associating domains are ancient features that coincide with metazoan clusters of extreme noncoding conservation. *Nature Communications,* *8*(1). https://doi.org/10.1038/s41467-017-00524-5

Irimia, M., Tena, J. J., Alexis, M. S., Fernandez-Miñan, A., Maeso, I., Bogdanovic, O., de la Calle-Mustienes, E., Roy, S. W., Gómez-Skarmeta, J. L. & Fraser, H. B. (2012). Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Research*, *22*(12), 2356–2367. https://doi.org/10.1101/gr.139725.112

Kassambara A (2023). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.6.0, https://rpkgs.datanovia.com/ggpubr/

Krefting, J., Andrade-Navarro, M. A., & Ibn-Salem, J. (2018). Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biology*, *16*(1). https://doi.org/10.1186/s12915-018-0556-x

Kröger, B., Vinther, J., & Fuchs, D. (2011). Cephalopod origin and evolution: A congruent picture emerging from fossils, development and molecules. *BioEssays*, *33*(8), 602–613. https://doi.org/10.1002/bies.201100001

Lee, B. H., Wu, Z., & Rhie, S. K. (2022). Characterizing chromatin interactions of regulatory elements and nucleosome positions, using hi-C, micro-C, and promoter capture micro-C. Epigenetics & Chromatin, 15(1). https://doi.org/10.1186/s13072-022-00473-4

Lee, P. N., Callaerts, P., De Couet, H. G. & Martindale, M. Q. (2003). Cephalopod Hox genes and the origin of morphological novelties. *Nature*, *424*(6952), 1061–1065. https://doi.org/10.1038/nature01872

Lee, P. N., McFall-Ngai, M. J., Callaerts, P., & de Couet, H. G. (2009). The Hawaiian Bobtail Squid (Euprymna scolopes): A Model to Study the Molecular Basis of Eukaryote-Prokaryote Mutualism and the Development and Evolution of Morphological Novelties in Cephalopods.*Cold Spring Harbor Protocols,* *2009*(11), pdb.emo135. https://doi.org/10.1101/pdb.emo135

Li, Y., Liu, H., Steenwyk, J. L., LaBella, A. L., Harrison, M.-C., Groenewald, M., Zhou, X., Shen, X.-X., Zhao, T., Hittinger, C. T., & Rokas, A. (2022). Contrasting modes of macro and Microsynteny Evolution in a eukaryotic subphylum. *Current Biology*, *32*(24). https://doi.org/10.1016/j.cub.2022.10.025

Lindgren, A. R. & Anderson, F. E. (2018). Assessing the utility of transcriptome data for inferring phylogenetic relationships among coleoid cephalopods. *Molecular Phylogenetics and Evolution*, *118*, 330–342. https://doi.org/10.1016/j.ympev.2017.10.004

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., … Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, *161*(5), 1012–1025. https://doi.org/10.1016/j.cell.2015.04.004

Nobrega, M. A., Ovcharenko, I., Afzal, V., & Rubin, E. M. (2003). Scanning human gene deserts for long-range enhancers. *Science, 302*(5644), 413–413. https://doi.org/10.1126/science.1088328

Nyholm, Spencer V., & McFall-Ngai, M. J. (2021). A lasting symbiosis: How the Hawaiian bobtail squid finds and keeps its bioluminescent bacterial partner. *Nature Reviews Microbiology*, *19*(10), 666–679. https://doi.org/10.1038/s41579-021-00567-y

O'Brien, C. E., Roumbedakis, K., & Winkelmann, I. E. (2018). The Current State of Cephalopod Science and Perspectives on the Most Critical Challenges Ahead From Three Early-Career Researchers. *Frontiers in Physiology, 9*. https://doi.org/10.3389/fphys.2018.00700

Ovcharenko, I., Loots, G. G., Nobrega, M. A., Hardison, R. C., Miller, W., & Stubbs, L. (2004). Evolution and functional classification of vertebrate gene deserts. *Genome Research*, *15*(1), 137–145. https://doi.org/10.1101/gr.3015505

Packard, A. (1972). Cephalopods and fish: The limits of convergence. *Biological Reviews of the Cambridge Philosophical Society*, *47*(2), 241–307. https://doi.org/10.1111/j.1469-185x.1972.tb00975.x

Passarge, E., Horsthemke, B., & Farber, R. A. (1999). Incorrect use of the term synteny. *Nature Genetics*, *23*(4), 387–387. https://doi.org/10.1038/70486

Preger-Ben Noon, E. & Frankel, N. (2022). Can changes in 3D genome architecture create new regulatory landscapes that contribute to phenotypic evolution? *Essays in Biochemistry*, *66*(6), 745–752. https://doi.org/10.1042/EBC20220057

Price, P. D., Palmer Droguett, D. H., Taylor, J. A., Kim, D. W., Place, E. S., Rogers, T. F., Mank, J. E., Cooney, C. R. & Wright, A. E. (2022). Detecting signatures of selection on gene expression. *Nature Ecology & Evolution*, *6*(7), 1035–1045. https://doi.org/10.1038/s41559-022-01761-8

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842

Rada-Iglesias, A., Grosveld, F. G. & Papantonis, A. (2018). Forces driving the three-dimensional folding of eukaryotic genomes. *Molecular Systems Biology*, *14*(6), e8214. https://doi.org/10.15252/msb.20188214

Rajderkar, S., Barozzi, I., Zhu, Y., Hu, R., Zhang, Y., Li, B., Alcaina Caro, A., Fukuda-Yuzawa, Y., Kelman, G., Akeza, A., Blow, M. J., Pham, Q., Harrington, A. N., Godoy, J., Meky, E. M., von Maydell, K., Hunter, R. D., Akiyama, J. A., Novak, C. S., … Pennacchio, L. A. (2023). Topologically associating domain boundaries are required for normal genome function. *Communications Biology*, *6*(1), 435. https://doi.org/10.1038/s42003-023-04819-w

Rogers, T., & Simakov, O. Emerging questions on the mechanisms and dynamics of 3D genome evolution in spiralians. *In review at Briefings in Functional Genomics.*

Rouressol, L., Briseno, J., Vijayan, N., Chen, G. Y., Ritschard, E. A., Sanchez, G., Nyholm, S. V., McFall-Ngai, M. J. & Simakov, O. (2023). Emergence of novel genomic regulatory regions associated with light-organ development in the bobtail squid. iScience, 26(7), 107091. https://doi.org/10.1016/j.isci.2023.107091

Rowley, M. J. & Corces, V. G. (2018). Organizational principles of 3D genome architecture. *Nature Reviews. Genetics*, *19*(12), 789–800. https://doi.org/10.1038/s41576-018-0060-8

RStudio Team version 1.4.1717. (2021). RStudio: Integrated Development Environment for R. Boston, MA. Retrieved from http://www.rstudio.com/

Sanchez, G., Setiamarga, D. H. E., Tuanapaya, S., Tongtherm, K., Winkelmann, I. E., Schmidbaur, H., Umino, T., Albertin, C., Allcock, L., Perales-Raya, C., Gleadall, I., Strugnell, J. M., Simakov, O. & Nabhitabhata, J. (2018). Genus-level phylogeny of cephalopods using molecular markers: current status and problematic areas. *PeerJ*, *6*, e4331. https://doi.org/10.7717/peerj.4331

Schmidbaur, H., Kawaguchi, A., Clarence, T., Fu, X., Hoang, O. P., Zimmermann, B., Ritschard, E. A., Weissenbacher, A., Foster, J. S., Nyholm, S. V., Bates, P. A., Albertin, C. B., Tanaka, E., & Simakov, O. (2022). Emergence of novel cephalopod gene regulation and expression through large-scale genome reorganization. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-29694-7

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., & Barillot, E. (2015). Hic-pro: An optimized and flexible pipeline for Hi-C Data Processing. *Genome Biology, 16*(1). https://doi.org/10.1186/s13059-015-0831-x

Tena, J. J., & Santos-Pereira, J. M. (2021). Topologically associating domains and regulatory landscapes in development, evolution and disease. *Frontiers in Cell and Developmental Biology, 9*. https://doi.org/10.3389/fcell.2021.702787

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace

Xavier, J. C., Allcock, A. L., Cherel, Y., Lipinski, M. R., Pierce, G. J., Rodhouse, P. G., Rosa, R., Shea, E. K., Strugnell, J. M., Vidal, E. A., Villanueva, R., & Ziegler, A. (2015). Future challenges in cephalopod research. *Journal of the Marine Biological Association of the United Kingdom, 95*(5), 999–1015.https://doi.org/10.1017/s0025315414000782

Zhang, Y., Mao, F., Mu, H., Huang, M., Bao, Y., Wang, L., Wong, N.-K., Xiao, S., Dai, H., Xiang, Z., Ma, M., Xiong, Y., Zhang, Z., Zhang, L., Song, X., Wang, F., Mu, X., Li, J., Ma, H., … Yu, Z. (2021). The genome of Nautilus pompilius illuminates eye evolution and biomineralization. *Nature Ecology & Evolution*, *5*(7), 927–938. https://doi.org/10.1038/s41559-021-01448-6