

Characterizing A-to-I RNA Edits in Nervous Tissue of *Octopus rubescens* in response to
Ocean Acidification

by
Ricky Orion Wright

A THESIS

submitted to

WALLA WALLA UNIVERSITY

in partial fulfillment of
the requirements for the
degree of

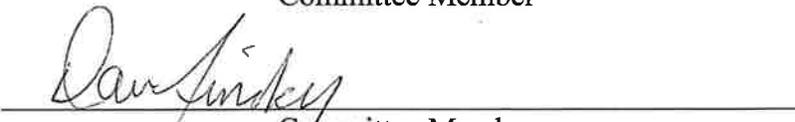
MASTER OF SCIENCE

June 2024

This thesis for the Master of Science degree
has been approved by the Department of Biological Sciences
and the Office of Graduate Studies
Walla Walla University


Major Professor

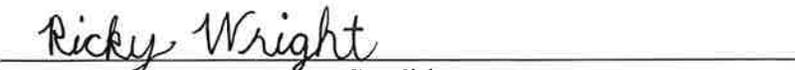

Committee Member


Committee Member

Committee Member

Dean of Graduate Studies


Observer of the Process - Graduate Representative


Candidate

Date

ABSTRACT

RNA editing can diversify expression beyond the genome. Adenosine deaminases that act on RNA (ADARs) edit single ribonucleotides in mRNA transcripts from adenosine to inosine, which is then read as guanosine by ribosomes. Selectivity for this reaction is not well understood, but its effects have been demonstrated in response to temperature. I tested nervous tissue from six specimens of *Octopus rubescens* for RNA editing in response to ocean acidification and found high levels of editing, particularly in zinc finger proteins, while editing overall was suppressed in response to acidification. Sanger verification of selected edits was unsuccessful, likely due to gDNA contamination. A preliminary genome assembly was drafted using PacBio HiFi long reads.

TABLE OF CONTENTS

ABSTRACT	3
TABLE OF CONTENTS	4
FIGURES AND TABLES	5
INTRODUCTION	7
Ocean Acidification	7
Effects of Ocean Acidification on Marine Life	7
RNA Editing	9
METHODS	14
Collection and Care of Octopuses	14
Extraction and sequencing	15
Transcriptome Assembly	15
Detection of Editing Sites	16
Comparison of Editing Levels	17
Sanger Verification	18
Genome Assembly	18
Data Availability	18
RESULTS	20
DISCUSSION	23
LITERATURE CITED	28

FIGURES AND TABLES

Figure 1. Trends in cephalopod abundance	32
Figure 2. Acid secretion in cephalopod gill epithelium	33
Figure 3. Frequencies and editing levels of synonymous (syn) and nonsynonymous (nonsyn) A-to-G editing in humans.	34
Figure 4. Mammalian set of editing sites.	35
Figure 5. Extensive recoding is conserved across coleoid cephalopods.	36
Figure 6. Signs for positive selection of recoding by editing.	37
Figure 7. Closing kinetics and editing levels.	38
Figure 8. Cold-induced editing in <i>Octopus bimaculoides</i> .	39
Figure 9. Summary of effect of intracellular acidification on ADAR editing efficiency.	40
Figure 10. Abstract representation of mismatch detection by alignment.	41
Figure 11. Yields of raw DNA reads from whole genome Illumina sequencing for each octopus (left). Total DNA after pooling and QC (right).	42
Figure 12. Raw RNA reads.	43
Figure 13. RNA reads after QC. No more than 400,000 reads were removed from each sample.	44
Figure 14. RNA reads mapped to assembled transcriptome per octopus.	45
Figure 15. Numbers of RNA and DNA reads after QC (left), and numbers of RNA and DNA reads mapped to the assembled transcriptome (right).	46
Figure 16. Zinc finger editing proportions.	47
Figure 17. Total editing levels in synonymous and nonsynonymous editing sites by pCO_2 level of experimental treatment.	48
Figure 18. Levels of editing only among significantly differentially edited sites with respect to pCO_2 .	49
Figure 19. Change in %-editing from low to high pCO_2 .	50
Figure 20. Designations within example ORF of editing site and primers for amplification and sanger sequencing.	51
Figure 21. Example of sanger results for a putative edit from an octopus exposed to high pCO_2 .	52
Figure 22. Snail plot of improved Phase Assembler (IPA) genome assembly using BlobToolkit.	53
Figure 23. Fractions of mismatch types of different cephalopod species.	54
Table 1. Average measurements for elevated and control pCO_2 experiments for the 10-13 days following acclimation and preceding harvest.	55
Table 2. Editing percentages detected by our pipeline vs Sanger results for a single putative A-to-G edit across six octopuses.	56

Table 3. Comparison of assembly statistics across Octopus genomes available on NCBI.

57

INTRODUCTION

Ocean Acidification

At least 30% of all atmospheric CO₂ produced by humans is absorbed by the ocean. When CO₂ dissolves in water, it becomes hydrated to carbonic acid (H₂CO₃). This acid immediately dissociates to bicarbonate (HCO₃⁻) and a proton (H⁺), reducing the pH of the solution. Before the industrial revolution, atmospheric CO₂ was ~281ppmv; today it has risen above 400ppmv, with an accompanying drop in average surface pH from 8.2 to 8.1 across the entire ocean (Doney et al., 2009; “IPCC,” 2021). This may not sound like much, but it represents a 26% rise in H⁺ concentration. By the year 2100, it is predicted that atmospheric CO₂ will rise to at least 420ppmv and ocean pH will drop by another 0.14 units (a 44% increase in H⁺ concentration from present) even if strong measures are taken; if no action is taken on climate change atmospheric CO₂ may exceed 1000ppmv and cause a 0.43 unit drop in ocean pH (a 260% increase in H⁺ concentration from present) (IPCC, 2019; Doney et al., 2009). Leaving aside the other massive oceanic changes from anthropogenic climate change, this pH change is unprecedented within the past 800,000 years (Lüthi et al., 2008), and such a dramatic change will certainly impact our biosphere, from individual organisms to entire ecosystems.

Effects of Ocean Acidification on Marine Life

There have been several demonstrated effects of near-future ocean acidification (OA) on cephalopods. The common squid *Loligo vulgaris* experienced lengthened embryonic development and slower growth rate resulting in an increase in premature larvae and reduced thermal tolerance (Rosa et al., 2014). Similarly, in the cuttlefish *Sepia officinalis*, OA resulted in increased embryonic mortality, premature hatching, and an overall reduction in embryonic metabolism (Rosa et al., 2013). In our own lab at Rosario Beach Marine Laboratory, *Octopus rubescens* demonstrated a short term increase in routine metabolic rate (RMR) and critical oxygen pressure (P_{crit}) in response

to near-future OA. While RMR returned to normal within one week, P_{crit} remained elevated after 5 weeks, indicating lingering hypoxia intolerance (Onthank et al., 2021). A later experiment under similar conditions measured immune responses of *O. rubescens*, and found increased hemocyte counts and an increase in total phagocytosis (Culler-Juarez and Onthank, 2021). However, findings regarding *Octopus rubescens* metabolic rates and immune responses to OA don't suggest imminent disaster, but instead highlight how minimal the effects of OA on cephalopods may be. Cephalopod populations have been on the rise for the past 60 years (Figure 1; Doubleday et al., 2016). Could they be adapting to the changing ocean in ways that other marine species cannot?

Marine cephalopods, much like teleosts, experience internal pCO_2 ranging from 0.1 to 0.4 kPa (1000 to 4000 ppm) (Melzner et al., 2007), and are already adapted to compensate for this (Hu et al., 2014), but this does not mean they are unaffected by rising external pCO_2 (hypercapnia). Studies on acid-base regulatory ability of *Sepia officinalis* showed that environmental hypercapnia was buffered internally by increased bicarbonate concentration, and despite this buffering, intracellular pH dropped slightly (Gutowska et al., 2010). The primary mechanism for maintaining ionic homeostasis likely occurs by ion transport within gill epithelia, and this mechanism relies on a sufficient pH/ pCO_2 gradient between the organism and the environment (Figure 2; (Hu et al., 2011). Recent studies in cephalopods have clarified possible pathways for H^+ excretion by tying it to elimination of nitrogenous waste (Gutowska et al., 2010; Hu et al., 2015, 2014).

In *Sepioteuthis lessoniana*, the principal transport proteins involved in acid/base homeostasis — Na⁺/K⁺ ATPases, anion exchangers, etc. — show changes in transcript levels within 28 hours of exposure to hypercapnia, but these changes disappear within 5-7 days (Hu et al., 2014). Cephalopods are affected by – and responding to – ocean acidification, but the mechanisms which allow them to maintain homeostasis haven't been well characterized.

RNA Editing

Editing of mRNA transcripts by adenosine deaminases that act on RNA (ADARs) changes adenosine to inosine (A to I) by hydrolytically replacing the amino group at C6 of the purine ring with a carbonyl. Inosine is read by ribosomes and polymerases as guanosine (G) (Bass, 2002). ADARs act on double-stranded RNA (dsRNA) by way of an evolutionarily acquired dsRNA binding domain (dsRBD) (Bass, 2002). Meaningful edits are known to occur in mammals. For example, glutamate receptor channels in mouse neurons contain an editing site which changes nearly 100% of transcripts to read arginine instead of glutamine at this codon. This edit is made possible by an intronic sequence with imperfect complementarity to the exon containing the edited adenosine; following transcription, the intron anneals to the exon, creating dsRNA and allowing the dsRBD of mouse ADAR to bind at this locus and deaminate the adenosine. When a recombinant mouse genome lacking this intron is generated, complementarity is removed as well, making editing by ADAR impossible. In response, mice exhibit epilepsy, hyperexcitability, and do not survive past 3 weeks (Seeburg et al., 2001). However, even when completely knocking out ADAR function, directly mutating the

genomic base to match the edited base completely rescues the wild type phenotype, giving strong indication that this is the *only* essential edit in mice and possibly other mammals. In humans, the majority of edits in coding regions are synonymous, and editing percentages in nonsynonymous sites are lower than synonymous sites, suggesting that editing in humans is generally nonadaptive (Figure 3; Xu and Zhang, 2014). Additionally, most editing sites in humans are within non-coding repeat sequences specific to primates known as *Alu* elements, and are thus not widely conserved among mammals (Figure 4; Pinto et al., 2014).

Beyond constitutive editing, in *Drosophila*, it was found that mRNA editing by ADAR was reduced at elevated temperature (Rieder et al., 2015). In coleoid cephalopods, RNA editing is far more prevalent. While there are only 59 conserved edits between mice and humans, coleoid cephalopods share 705 A-to-G recoding sites with $\geq 10\%$ editing across 393 proteins (Figure 5; (Liscovitch-Brauer et al., 2017a).

In contrast to nonsynonymous editing rates in humans, coleoid cephalopods demonstrate nonsynonymous edits at a much higher rate than simple randomness would predict, and the more highly edited a site is, the more likely it is to be nonsynonymous (Figure 6). While the bulk of editing in humans and mice takes place in noncoding regions, coleoid cephalopods demonstrate a prodigious diversity of recoding sites as yet unmatched in any other clade (Albertin et al., 2022; Liscovitch-Brauer et al., 2017a; Xu and Zhang, 2014).

Much research is yet to be done to uncover RNA editing in response to environmental change; aside from the previously mentioned work on temperature dependent ADAR activity in *Drosophila*, compelling evidence for adaptive editing in correlation with environmental differences has been found in octopus neural tissue. Voltage-gated K^+ -channels in neural tissue from coleoid cephalopods ranging from a cold water octopus in genus *Pareledone* to a tropical species, *Octopus vulgaris*,

showcase differential editing. Over 80% of the cold-adapted octopus's K⁺-channel transcripts contained an A-to-G edit, causing amino acid 321 to change from an isoleucine to a valine. This single substitution destabilized the open state of the channel, resulting in a steeper repolarization curve (Figure 7). Without this edit, repolarization of neurons in cold water would be slower and reduce the possible firing frequency of action potentials. Eight different species of octopus, including *Octopus rubescens*, were collected from water with different temperatures ranging from tropical to antarctic; they were assayed for editing levels at this codon and found to increase editing percentage as average ambient water temperature decreased (Figure 7; Garrett and Rosenthal, 2012).

More compellingly than correlation, *Octopus bimaculoides* demonstrates temperature-dependent RNA editing. Out of 62,661 editing sites, 20,850 were cold-induced, while 789 were warm-induced (Figure 8; Birk et al., 2023). At colder temperatures, a kinesin protein was more highly edited, and the edit caused the protein's function to be less temperature dependent. Cold-induced editing also resulted in a synaptotagmin protein binding its first calcium with lower affinity while leaving subsequent calcium binding unaffected. It is important to note that the overwhelming majority of temperature-induced edits were cold-induced; editing suppression was observed at higher temperatures (Birk et al., 2023).

If RNA editing can correlate to, and even be dependent upon, temperature changes, can the same be said for changes in ocean pH? In human embryonic kidney cells, when catalyzing the deamination of adenosine to inosine, ADAR binds to dsRNA, then flips the adenosine out of the RNA duplex and into its active site. The absence of the adenosine from the duplex is stabilized by a conserved glutamic acid which hydrogen bonds to the complementary guanosine during catalysis. This temporary H-bond is dependent upon the glutamic acid existing in its protonated form. With higher

H⁺ concentration within the cell, the population of glutamic acids existing in their protonated form will be higher as well, yielding a higher overall probability of successful deamination (Figure 9; (Malik et al., 2021).

This biochemistry was explored in human cells, not cephalopods. However, in *Sepia officinalis* it has been shown that environmental hypercapnia can lead to significantly reduced intracellular pH, allowing for the possibility of a similar pH dependent interaction to take place in cephalopods (Gutowska et al., 2010). This is only one possible mechanism by which RNA editing might be affected by rising acidification, and until 2019, this connection had not been explored.

To test cephalopod editing in response to OA, Liscovitch-Brauer et al.'s bioinformatic methodology was replicated to detect editing in *Octopus rubescens* gill tissues. It was found that the number of editing sites in *rubescens* gill (122,596) was comparable to nervous tissue from *Octopus bimaculoides* (105,975), but the percentages at which those sites were edited were much lower. Despite overall lower editing percentages, 136 A-to-G recoding sites were found which are significantly differentially edited in octopuses exposed to environmental OA compared to those exposed to control seawater (Sereewit, 2022). Importantly, nonsynonymous A-to-G editing sites were edited at higher levels in octopuses subjected to OA vs control. This could suggest that octopuses upregulate editing of several genes or pathways in response to low pCO_2 , and could help explain why cephalopod populations at large seem unaffected by the changing ocean. Editing in gill tissue might play a role in adaptation to OA, but we know that editing levels in neural tissue are far higher (Liscovitch-Brauer et al., 2017a). I repeated this experiment using nervous tissue instead of gill.

Detecting editing without a genome assembly is straightforward, but requires aligning both gDNA and mRNA reads to a transcriptome, as was done by

Liscovitch-Brauer et al. This necessitates acquiring gDNA sequencing for every individual, substantially adding to cost. With an already assembled genome, editing detection requires only RNASeq data to align to the genome. Many editing detection pipelines for cephalopods rely on this method, including Birk et al.'s work showing temperature dependent RNA editing in *Octopus bimaculoides* (Birk et al., 2023). I began laying the foundation for us to utilize the genome assembly dependent method by producing a preliminary genome assembly for *Octopus rubescens*. While the pipeline was built for detecting edits broadly, my project puts a greater focus on verifying selected edits using sanger sequencing, and further characterizing the types of genes edited in response to a change in pCO_2 .

METHODS

Collection and Care of Octopuses

Twelve *Octopus rubescens* were collected at Driftwood Park on Whidbey Island and taken to RBML and placed in 113.5L seawater flow-through tanks. Temperature and $p\text{CO}_2$ of each tank were controlled with Open Acidification Tank Controllers (Onthank et al., 2023) attached to a three-wire PT-100 temperature probe and a single junction glass pH electrode, both inserted through holes drilled in the tops of the tank lids. Temperature was set to match the collection site, and temperature probes were calibrated weekly against a high-precision NIST-traceable alcohol thermometer. pH electrodes were calibrated approximately every other day to the water in the tank it was monitoring by taking a spectrophotometric pH measurement of the tanks and using a 1-point calibration to that pH value. Salinity was measured daily by Vernier probe (SAL-BTA), and total alkalinity was measured weekly by open-cell titration. Spectrophotometer measurements, salinity, and alkalinity were used to determine setpoint pH to achieve 1500 $\mu\text{atm } p\text{CO}_2$ using the seacarb package (V3.2.14) in R.

Following a one week acclimation period after collection, octopuses were randomly selected to be in control or treatment groups for 10 to 12 days. The experiment was run as long as possible given multiple collection dates while facing a single harvest date of August 23. Control groups were kept at a target $\sim 750 \mu\text{atm } p\text{CO}_2$ (approximately the $p\text{CO}_2$ of the location where they were collected) (Table 1). Experimental $p\text{CO}_2$ treatment groups were kept at $\sim 1500 \mu\text{atm } p\text{CO}_2$ in accordance with end of century $p\text{CO}_2$ based on business-as-usual emissions projections, ("IPCC," 2021). After 7 days, octopuses were euthanized by placing them in 2.5% ethanol in seawater until the octopuses were sedated, then bringing the ethanol concentration to 10% to euthanize. After octopuses were euthanized, optic lobe tissue was harvested (Butler-Struben et al., 2018).

Extraction and sequencing

RNA was extracted from optic lobe tissue immediately after tissue was harvested using a Zymo Research Quick-RNA™ MiniPrep Plus Kit (R1057), while DNA was extracted using a Zymo Research Quick-DNA™ Miniprep Plus Kit (D4068). Six RNA and six DNA samples were sent to Novogene for Illumina sequencing. RNA was processed according to Novogene's mRNASeq package including: total RNA QC, poly-A enrichment, library prep, and illumina sequencing, targeting 20M paired-end reads per sample. DNA was sequenced to a depth of 9X. The quality of gDNA reads was assessed using FastQC V0.11.9, and trimmed using Trim Galore V0.6.7 (Felix Krueger, n.d.; Simon Andrews, n.d.). gDNA reads from all six octopuses were combined for maximal coverage when aligning to the transcriptome.

Transcriptome Assembly

Quality control was performed on RNAseq data using Fastp V0.20.1 with default settings (Chen et al., 2018). Rcorrector V1.0.5 was used to correct errors or label uncorrectable errors to prevent them from interfering with transcriptome assembly (Song and Florea, 2015). Labeled uncorrectable errors were removed using a custom python script from HarvardInformatics qc pipeline (Freedman, Guan, 2024). To account for poly-adenylation of rRNA molecules (Slomovic et al., 2006), RNA reads were compared to *Octopus rubescens* rRNA sequences assembled from our raw read data. The *O. rubescens* rRNA sequences were indexed from a single fasta file using Bowtie2 V2.4.4 (Langmead and Salzberg, 2012). Our RNA reads were mapped to the fasta file using Bowtie2's local alignment function set to "very sensitive", with all other parameters default. Only reads that did not align to the rRNA sequences were used to assemble the transcriptome.

Trinity V2.0.2 was used to perform *de novo* transcriptome assembly on all of the mRNA derived from our six samples (Grabherr et al., 2011). Parameters were default

except for library type, which was set to "--SS_lib_type R", using the argument 'R' for paired end reads. As I was only interested in edits that would change the amino acid composition of the final protein, I only looked for nonsynonymous edits within open reading frames (ORFs). Therefore I restricted the assembled transcriptome to ORFs which were identified within the transcriptome using the NCBI ORFfinder tool. ORFs which fell completely within another (referred to as "nested" by ORFfinder) were ignored. The identified ORFs were blasted against the UniProt database (UniProt Consortium, 2021) to assign putative functions. ORFs not matched with a BLASTX e-value $<1e-6$ were ignored.

Detection of Editing Sites

Using Bowtie2, gDNA reads from all six octopuses were aligned to the transcriptome ORFs using the local alignment setting. Read pairs only required alignment on forward or reverse, not both, as we would with single-end reads. Reads that did not achieve primary alignment, the highest quality mapping score within Bowtie2, were ignored.

Putative editing sites were identified by finding mismatches between gDNA and mRNA using two methods. In the first, mismatches are defined as positions where all of the gDNA reads disagree with the transcriptome; since the transcriptome consensus is based on the majority of mRNA reads, these positions show more than 50% of the reads are edited. However, it is possible that all gDNA reads may differ from the transcriptome if the octopus is heterozygotic at a particular base and only one chromosome is sampled.

In the second method to locate mismatches, both gDNA and RNA reads were aligned to the transcriptome. If some of the RNA reads disagreed with the transcriptome at a particular locus, and if all gDNA reads were in agreement with one another, that locus was treated as a mismatch (Figure 10). Using only sites with complete gDNA

agreement reduces the probability of mistaking heterozygous loci or SNPs as editing events. Mismatches identified by both methods were combined into a single dataset.

At mismatch locations, a binomial test was used to test the probability that mismatches were more common than could likely be randomly generated by sequencing error. The binomial test was based on the number of mismatches at a given position, the total number of mRNA reads aligned to that position, and the sequencing error probability. Illumina sequencing probability was estimated to be 0.1%, therefore only nucleotides with a sequence quality score greater than 30 (sequencing score = $-10_{\log}(\text{sequencing error probability})$) were used (Stoler and Nekrutenko, 2021). The p-value for each locus was corrected for multiple testing using a Benjamini-Hochberg false-discovery rate of 10%. Sites with a corrected p-value < 0.05 were considered significant; it was assumed that the discrepancy between mRNA reads and the transcriptome was unlikely to be a result of sequencing errors, and therefore more likely a result of RNA editing.

Comparison of Editing Levels

To compare editing rates between octopuses in normal and high $p\text{CO}_2$, a permutation t-test compared the editing proportions at each locus between the two groups of octopuses. This was done using a custom R script (Sereewit, 2022). Benjamini-Hochberg p-value correction for multiple comparisons was used to adjust p-values.

To narrow the edits to those for further investigation we filtered the edits to only include nonsynonymous A-to-I edits with an adjusted p-value < 0.005 on the permutation t-test between treatments, evidenced by at least 50 mRNA edited reads, with a mean proportion editing difference between treatments of at least 20%.

Sanger Verification

To prepare samples for Sanger verification, RNA extracted from the other six octopuses (3 treatment, 3 control) was processed using Protoscript II First Strand Synthesis Kit (NEB E6560) to generate a cDNA library. Using a 2 step nested primer design, edited ORFs were amplified from cDNA using Hot Start *Taq* 2X Master Mix (NEB M0496S), purified with Monarch PCR & DNA Cleanup Kit (NEB T1030S), then resolved by electrophoresis on a 1% agarose gel. Successful amplicons were then amplified with nested primers, purified, and resolved on a 1% agarose gel. Final amplicons were sent to Lone Star Labs for Sanger sequencing. Editing levels in Sanger results were quantified using EditR.

Genome Assembly

Separately, a male *Octopus rubescens* was purchased from Monterey Abalone Company, which had been collected in Monterey Bay, California. Arm tissue was harvested from this individual, flash frozen in liquid nitrogen, and sent to the University of California, Davis for extraction and PacBio HiFi sequencing. Raw bam results were converted to fastq using bam2fastq. Assembly was performed using Improved Phase Assembler (IPA V1.8.0). Genome quality was assessed using BUSCO with the mollusca database, Odb10. We also generated a snail plot of the assembly using BlobToolkit.

Data Availability

Raw data is stored locally on the WWU Biology Department genomics computer. This is backed up locally on a network-attached storage device on WWU's network. Raw sequencing data is being uploaded to genbank SRA. All R code used in this data analysis and to create all original figures in this thesis is available on GitHub at

https://github.com/RickyW94/Octopus_rubescens_RNA_Editing_Pipeline and
https://github.com/RickyW94/Rubescens_PacBIO_HiFi_assembly.

RESULTS

In order to identify editing sites in the *Octopus rubescens* transcriptome I used Illumina HiSeq sequencing to sequence gDNA and mRNA for six octopuses, three in each treatment. Whole genome sequencing yielded >150 million DNA reads for each of the six octopuses, (Figure 11). Pooling and QC resulted in a total of ~1.2 billion DNA reads (Figure 11). Of these, only ~62 million DNA reads mapped to the transcriptome (Figure 14).

mRNASeq yielded >40 million raw RNA reads per octopus (Figure 12). QC removed no more than 400k reads per sample (Figure 13). About half of the post-QC RNA reads from each octopus mapped to the transcriptome (Figure 14).

Considerably more of the RNA reads mapped to the transcriptome than did DNA reads (Figure 15).

I constructed a *de novo* transcriptome from mRNA reads, aligned mRNA and gDNA reads to that transcriptome, and compared mRNA to gDNA to identify mismatches, which, after further filtering, could be assigned as putative editing sites. I found 279,235 total editing sites in the *Octopus rubescens* transcriptome. Of those, 11,999 of those editing sites were found to be responsive to environmental pCO_2 (significantly increased or decreased in the elevated pCO_2 treatment over control). Surprisingly, zinc finger proteins were highly represented in the pool of genes containing pCO_2 responsive editing sites. Zinc fingers make up 5.3% of *Octopus bimaculoides* genes, yet account for 16.3% of genes containing pCO_2 responsive edits (Figure 16). I filtered the pCO_2 responsive edits to get a smaller pool of “high priority” edits, which I defined as A-to-G nonsynonymous edits with a mean editing level difference of at least 20% between treatments and were evidenced by at least 10 mRNA reads in each

octopus and at least 10 gDNA reads from all octopuses combined, and had a permutation t-test adjusted p-value of less than 0.005, yielding 51 such editing sites. Of these edits, 51% are contained within zinc finger proteins.

Across all mismatches, synonymous edits were not significantly different between control and elevated $p\text{CO}_2$ treatments (Figure 17). Nonsynonymous edits were significantly more highly edited in elevated $p\text{CO}_2$ treatment. However, when considering only mismatches which were significantly differentially edited between treatment and control, both synonymous and nonsynonymous edits were very significantly suppressed (Figures 18,19).

To prioritize verification, I filtered $p\text{CO}_2$ responsive edits as described above. Of those 51 editing sites, I identified 4 that were contained within genes that I subjectively deemed as interesting and selected for verification. Nested primer PCR was performed on the four most promising highly significant edits on the other six octopuses which had not been subjected to RNASeq/whole genome sequencing (Figure 20). Our detection protocol detected at least 14% average editing among the original six octopuses, and at least one octopus demonstrated at least 68% editing for each of the 4 editing sites. To test verification procedures, I amplified a gene containing a high priority edit of interest from cDNA in the same six octopuses for which editing was quantified by Illumina mRNA-seq, and editing level quantified from Sanger sequencing chromatogram using EditR (Figure 21). Editing levels arrived at by Sanger sequencing verification did not approximate editing levels arrived at by Illumina mRNA-seq (Table 2). I did not have the time nor resources to troubleshoot the verification procedure, and therefore did not pursue verification further.

So that I could assemble a high-quality draft *Octopus rubescens* genome, I sent an *Octopus rubescens* arm tissue sample to DNA Technologies Core at UC Davis Genome Center for PacBio long-read sequencing. PacBio sequencing generated 11.6

million HiFi reads. IPA generated an assembly with 4173 total scaffolds, with 98.37% of the genome residing in scaffolds of > 50Kbp (kilo base-pairs). The scaffold N50 was 5.77 Mbp, with the longest scaffold reaching 24.7 Mbp (Figure 22). Benchmarking Universal Single Copy Ortholog (BUSCO) results showed 4,636 complete BUSCOs or 87.5% complete, with 103 complete and duplicated, 258 fragmented, and 401 missing BUSCOs out of a total 5295 BUSCO groups searched. The final genome length is ~3.8gbp.

DISCUSSION

In this thesis I replicated the original 2019 experiment examining changes in gill RNA editing with changing environmental $p\text{CO}_2$ levels on nervous tissue, revealing substantially more RNA edits that are responsive to environmental $p\text{CO}_2$ changes (approximately 12,000) and novel editing patterns in response to elevated $p\text{CO}_2$. Additionally, I created a preliminary genome assembly for *Octopus rubescens*, laying the groundwork for future genomic studies.

The draft *Octopus rubescens* genome I produced compares favorably to other octopus genomes currently available. N50 is a measure of how fragmented an assembly is, and for our purposes is expressed as a length in base pairs. For any given contig N50, “at least half of the nucleotides in the assembly belongs to contigs with the N50 length or longer” (Videvall, 2017). If ultra-long read or 3D conformational data are available, contigs without direct overlap can be further grouped and ordered into scaffolds; the sizes of gaps between contigs are estimated and filled with “N” for unknown nucleotides, but since this number is unknown, the number of Ns filled may be wrong, and overly aggressive scaffolding may harm the accuracy of the assembly (Luo et al., 2021). Our genome assembly created from PacBio HiFi reads achieved a contig N50 of 5.768Mbp, but without other data types to assist in scaffolding, our scaffold N50 is 5.77Mbp, showing that virtually no scaffolding is present in this genome assembly. Compared to the annotated chromosome-level assemblies for *Octopus bimaculoides* (*bimac*) and *Octopus sinensis* (*sinensis*), with 96.9Mb and 105.9Mb scaffold N50s respectively, ours is relatively fragmented. Thanks to our acquisition of HiFi long-read data, our contig N50 is substantially higher than any other *Octopus* assembly on NCBI (Table 2). Also, while the assembly remains fragmented due to lack of scaffolding, we

are confident in its accuracy and completeness due to its high BUSCO score. Benchmarking universal single copy orthologs (BUSCOs) are genetic sequences determined to be present in at least 90% of species for a given clade; in our case we used the BUSCO database for molluscs (Manni et al., 2021). BUSCO score is the percent of the BUSCOs found to be present in a genome. If a genome assembly has a very low BUSCO score, then it has likely incorrectly assembled the universal orthologs which must be present, casting doubt on the assembly as a whole. The BUSCO score for our assembly is 87.5%, which is comparable to the two available chromosome-level reference assemblies for *Octopus* on NCBI, *bimac* (90.1%), and *sinensis* (89.3%). There are other published octopus genomes based on short read Illumina sequencing, several of which achieve far higher N50s than ours by scaffolding their contigs. These short read assemblies present difficulties during scaffolding which are apparent when considering their low BUSCO scores, which are all 70% or lower. For example, the published *rubescens* genome on NCBI by Iridian Genomes boasts a scaffold N50 of 23Mbp, but the proportion of universal single copy orthologs found in this assembly is only 57% (Table 2). While our current editing detection methodology doesn't utilize a genome assembly, but rather aligns gDNA reads to a *de novo* assembled transcriptome, replacing gDNA reads with an accurate genome assembly would remove the necessity to sequence gDNA with every experiment, thereby significantly reducing costs and effort needed for future RNA editing experiments. For these purposes, it is much more important to have an accurate and complete assembly (high BUSCO score) than to achieve low fragmentation (high N50).

While PacBio HiFi reads exhibit a very low error rate, assembly errors are possible in especially challenging regions of the genome such as homopolymers, which code for highly repetitive amino acid sequences, and make up ~0.4% of eukaryotic genomes. Correction of single nucleotide substitutions, insertions, or deletions at these points is accomplished programmatically using short read Illumina sequencing data, which has a much higher sequencing accuracy than long-read sequencing methods, in a process known as “genome polishing” (Becerra et al., 2022; Hu et al., 2024). Our genome has only been assembled to the level of contiguous segments (contigs) according to overlaps of long-read PacBio HiFi reads; scaffolding these contigs involves assigning gaps between contigs and their placements within putative chromosomes. This would make the genome even less fragmented. Scaffolding would be greatly assisted with the acquisition of Hi-C sequencing data. Hi-C involves crosslinking chromatin such that DNA fragments are bound to nearby fragments up or downstream on the same chromosome, then sequencing those pairs. Algorithmically comparing the Hi-C sequence overlaps to our contigs correlates them in 3D space, allowing us to scaffold them to help map across the many long repeat regions I observed in *Octopus rubescens*. Alongside improving the assembly itself, annotation of the genome, which is the process of assigning identity and function to specific loci of the genome, would allow us to easily add newly discovered edits to a library of edits without blasting the identity of the containing gene every single time. Annotation can be accomplished preliminarily through homology, but would be assisted by the acquisition of mRNASeq data for the individual originally subjected to HiFi sequencing, (an *O. rubescens* specimen named Aristotle). Tissue from Aristotle has been harvested and stored at -80C in anticipation of future progress.

Taken together, the existence and further improvement of this draft *Octopus rubescens* genome will be an invaluable tool for further work on RNA editing in this species.

While other coleoid cephalopod species show highly disproportionate A-to-G editing, *rubescens* nervous tissue and gill tissue both show relatively fewer A-to-G edits compared to others, though still elevated (Figure 23; (Liscovitch-Brauer et al., 2017b). Despite not finding the expected A-to-G editing ratio, the overall number of significant edits in nervous tissue was 11,999, far higher than the 744 edits found in gill tissue. This was expected given the established pattern of coleoid cephalopod nervous tissue displaying far higher editing levels than other tissues (Liscovitch-Brauer et al., 2017).

Zinc fingers were dramatically overrepresented at high priority edit sites (Figure 16). Coleoid cephalopods show an expansion of zinc finger genes (Albertin et al., 2022). Zinc fingers are short alpha helices with particular amino acids imparting specificity for a short nucleotide sequence. They are commonly found working in tandem, with multiple domains present in a single protein. The classic C2H2-like zinc fingers serve as DNA-binding domains for any protein which needs to bind a specific DNA sequence, such as a transcription factor, repressor, nuclease, etc. (Ichikawa et al., 2023; Persikov et al., 2015). Their broad functions and complex binding profiles make it difficult to describe the zinc fingers we have identified as highly edited. Further exploration of the exact roles these zinc fingers likely play in *O. rubescens* will be undertaken by successors to this research. Future graduate students should utilize protein structure prediction software I-TASSER alongside gene ontology using COFACTOR to identify the structure and function of highly edited proteins (Zhou et al., 2022).

It was unfortunate that our samples failed to show editing through sanger sequencing, however, this was likely due to gDNA contamination, as the original RNA extractions weren't treated with DNase. Follow up research will start by treating existing

RNA samples with DNase prior to regeneration of cDNA and amplification of edited sequences.

Among editing sites that are significantly differentially edited with respect to OA and control, both synonymous and nonsynonymous editing is suppressed in OA (Figure 18). This parallels previous research results on temperature induced editing wherein octopuses subjected to elevated temperatures showed editing suppression (Birk et al., 2023). In their experiment, only 1% of the edits were upregulated during elevated temperatures, while the remaining 33% of temperature-responsive edits were downregulated. .

As previously mentioned, coleoid cephalopods show distinctly higher editing levels than other cephalopods. Coleoids diverged from other cephalopods approximately 240 mya, and at that point in earth's history global temperatures and CO₂ levels were much higher than in the past few thousand years (López-Córdova et al., 2022; Shaun A Marcott, Jeremy D Shakun, Peter U Clark, Alan C Mix, 2013). Atmospheric CO₂ levels may rise to the same level they were around 200 mya by the year 2250 AD. If editing suppression occurs in octopuses in response to rising ocean temperatures and CO₂, it may resemble an ancestral state similar to when coleoid cephalopods first split.

LITERATURE CITED

- Albertin, C.B., Medina-Ruiz, S., Mitros, T., Schmidbaur, H., Sanchez, G., Wang, Z.Y., Grimwood, J., Rosenthal, J.J.C., Ragsdale, C.W., Simakov, O., Rokhsar, D.S., 2022. Genome and transcriptome mechanisms driving cephalopod evolution. *Nat. Commun.* 13, 2427. <https://doi.org/10.1038/s41467-022-29748-w>
- Bass, B.L., 2002. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* 71, 817–846. <https://doi.org/10.1146/annurev.biochem.71.110601.135501>
- Becerra, A., Muñoz-Velasco, I., Aguilar-Cámara, A., Cottom-Salas, W., Cruz-González, A., Vázquez-Salazar, A., Hernández-Morales, R., Jácome, R., Campillo-Balderas, J.A., Lazcano, A., 2022. Two short low complexity regions (LCRs) are hallmark sequences of the Delta SARS-CoV-2 variant spike protein. *Sci. Rep.* 12, 936. <https://doi.org/10.1038/s41598-022-04976-8>
- Birk, M.A., Liscovitch-Brauer, N., Dominguez, M.J., McNeme, S., Yue, Y., Hoff, J.D., Twersky, I., Verhey, K.J., Sutton, R.B., Eisenberg, E., Rosenthal, J.J.C., 2023. Temperature-dependent RNA editing in octopus extensively recodes the neural proteome. *Cell* 186, 2544–2555.e13. <https://doi.org/10.1016/j.cell.2023.05.004>
- Butler-Struben, H.M., Brophy, S.M., Johnson, N.A., Crook, R.J., 2018. In Vivo Recording of Neural and Behavioral Correlates of Anesthesia Induction, Reversal, and Euthanasia in Cephalopod Molluscs. *Front. Physiol.* 9.
- Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Culler-Juarez, M.E., Onthank, K.L., 2021. Elevated immune response in Octopus rubescens under ocean acidification and warming conditions. *Mar. Biol.* 168, 137. <https://doi.org/10.1007/s00227-021-03913-z>
- Doney, S.C., Fabry, V.J., Feely, R.A., Kleypas, J.A., 2009. Ocean acidification: the other CO₂ problem. *Annu. Rev. Mar. Sci.* 1, 169–192.
- Doubleday, Z.A., Prowse, T.A., Arkhipkin, A., Pierce, G.J., Semmens, J., Steer, M., Leporati, S.C., Lourenço, S., Quetglas, A., Sauer, W., others, 2016. Global proliferation of cephalopods. *Curr. Biol.* 26, R406–R407.
- Felix Krueger, n.d. TrimGalore [WWW Document]. URL <https://github.com/FelixKrueger/TrimGalore/tree/master> (accessed 5.22.24).
- Freedman, Guan, 2024. Harvard Informatics [WWW Document]. GitHub. URL <https://github.com/harvardinformatics/TranscriptomeAssemblyTools/tree/master/utilities> (accessed 1.16.24).
- Garrett, S., Rosenthal, J.J.C., 2012. RNA editing underlies temperature adaptation in K⁺ channels from polar octopuses. *Science* 335, 848–851. <https://doi.org/10.1126/science.1212795>
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29, 644–652. <https://doi.org/10.1038/nbt.1883>
- Gutowska, M.A., Melzner, F., Langenbuch, M., Bock, C., Claireaux, G., Pörtner, H.O., 2010. Acid–base regulatory ability of the cephalopod (*Sepia officinalis*) in response to environmental hypercapnia. *J. Comp. Physiol. B* 180, 323–335.

- <https://doi.org/10.1007/s00360-009-0412-y>
 Hu, J., Wang, Z., Liang, F., Liu, S.-L., Ye, K., Wang, D.-P., 2024. NextPolish2: A Repeat-aware Polishing Tool for Genomes Assembled Using HiFi Long Reads. *Genomics Proteomics Bioinformatics* qzad009.
<https://doi.org/10.1093/gpbjnl/qzad009>
- Hu, M.Y., Guh, Y.-J., Stumpp, M., Lee, J.-R., Chen, R.-D., Sung, P.-H., Chen, Y.-C., Hwang, P.-P., Tseng, Y.-C., 2014. Branchial NH₄⁺-dependent acid–base transport mechanisms and energy metabolism of squid (*Sepioteuthis lessoniana*) affected by seawater acidification. *Front. Zool.* 11, 55.
<https://doi.org/10.1186/s12983-014-0055-z>
- Hu, M.Y., Hwang, P.-P., Tseng, Y.-C., 2015. Recent advances in understanding trans-epithelial acid-base regulation and excretion mechanisms in cephalopods. *Tissue Barriers* 3, e1064196. <https://doi.org/10.1080/21688370.2015.1064196>
- Hu, M.Y., Tseng, Y.-C., Stumpp, M., Gutowska, M.A., Kiko, R., Lucassen, M., Melzner, F., 2011. Elevated seawater Pco₂ differentially affects branchial acid-base transporters over the course of development in the cephalopod *Sepia officinalis*. *Am. J. Physiol.-Regul. Integr. Comp. Physiol.* 300, R1100–R1114.
<https://doi.org/10.1152/ajpregu.00653.2010>
- Ichikawa, D.M., Abdin, O., Alerasool, N., Kogenaru, M., Mueller, A.L., Wen, H., Giganti, D.O., Goldberg, G.W., Adams, S., Spencer, J.M., Razavi, R., Nim, S., Zheng, H., Gionco, C., Clark, F.T., Strokach, A., Hughes, T.R., Lionnet, T., Taipale, M., Kim, P.M., Noyes, M.B., 2023. A universal deep-learning model for zinc finger design enables transcription factor reprogramming. *Nat. Biotechnol.* 1–13.
<https://doi.org/10.1038/s41587-022-01624-4>
- IPCC, 2021. . IPCC 2021 Clim. Change 2021 Phys. Sci. Basis Contrib. Work. Group Sixth Assess. Rep. Intergov. Panel Clim. Change Masson-Delmotte V P Zhai Pirani SL Connors C Péan Berg. N Caud Chen Goldfarb MI Gomis M Huang K Leitzell E Lonnoy JBR Matthews TK Maycock T Waterfield O Yelekçi R Yu B Zhou Eds Camb. Univ. Press Press.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>
- Liscovitch-Brauer, N., Alon, S., Porath, H.T., Elstein, B., Unger, R., Ziv, T., Admon, A., Levanon, E.Y., Rosenthal, J.J.C., Eisenberg, E., 2017a. Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods. *Cell* 169, 191-202.e11. <https://doi.org/10.1016/j.cell.2017.03.025>
- Liscovitch-Brauer, N., Alon, S., Porath, H.T., Elstein, B., Unger, R., Ziv, T., Admon, A., Levanon, E.Y., Rosenthal, J.J.C., Eisenberg, E., 2017b. Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods. *Cell* 169, 191-202.e11. <https://doi.org/10.1016/j.cell.2017.03.025>
- López-Córdova, D.A., Avaria-Llautureo, J., Ulloa, P.M., Braid, H.E., Revell, L.J., Fuchs, D., Ibáñez, C.M., 2022. Mesozoic origin of coleoid cephalopods and their abrupt shifts of diversification patterns. *Mol. Phylogenet. Evol.* 166, 107331.
<https://doi.org/10.1016/j.ympev.2021.107331>
- Luo, J., Wei, Y., Lyu, M., Wu, Z., Liu, X., Luo, H., Yan, C., 2021. A comprehensive review of scaffolding methods in genome assembly. *Brief. Bioinform.* 22, bbab033.
<https://doi.org/10.1093/bib/bbab033>
- Lüthi, D., Le Floch, M., Bereiter, B., Blunier, T., Barnola, J.-M., Siegenthaler, U., Raynaud, D., Jouzel, J., Fischer, H., Kawamura, K., Stocker, T.F., 2008. High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* 453, 379–382. <https://doi.org/10.1038/nature06949>
- Malik, T.N., Doherty, E.E., Gaded, V.M., Hill, T.M., Beal, P.A., Emeson, R.B., 2021.

- Regulation of RNA editing by intracellular acidification. *Nucleic Acids Res.* 49, 4020–4036. <https://doi.org/10.1093/nar/gkab157>
- Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M., 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* 38, 4647–4654. <https://doi.org/10.1093/molbev/msab199>
- Melzner, F., Bock, C., Pörtner, H.-O., 2007. Coordination between ventilatory pressure oscillations and venous return in the cephalopod *Sepia officinalis* under control conditions, spontaneous exercise and recovery. *J. Comp. Physiol. [B]* 177, 1–17. <https://doi.org/10.1007/s00360-006-0104-9>
- Onthank et al., 2023. Tank Controller Overview [WWW Document]. Open Acidif. Proj. URL <https://open-acidification.github.io/docs/11-overview/> (accessed 1.16.24).
- Onthank, K.L., Trueblood, L.A., Schrock-Duff, T., Kore, L.G., 2021. Impact of Short- and Long-Term Exposure to Elevated Seawater Pco₂ on Metabolic Rate and Hypoxia Tolerance in *Octopus rubescens*. *Physiol. Biochem. Zool.* 94, 1–11. <https://doi.org/10.1086/712207>
- Persikov, A.V., Wetzel, J.L., Rowland, E.F., Oakes, B.L., Xu, D.J., Singh, M., Noyes, M.B., 2015. A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. *Nucleic Acids Res.* 43, 1965–1984. <https://doi.org/10.1093/nar/gku1395>
- Pinto, Y., Cohen, H.Y., Levanon, E.Y., 2014. Mammalian conserved ADAR targets comprise only a small fragment of the human editosome. *Genome Biol.* 15, R5. <https://doi.org/10.1186/gb-2014-15-1-r5>
- Rieder, L.E., Savva, Y.A., Reyna, M.A., Chang, Y.-J., Dorsky, J.S., Rezaei, A., Reenan, R.A., 2015. Dynamic response of RNA editing to temperature in *Drosophila*. *BMC Biol.* 13, 1. <https://doi.org/10.1186/s12915-014-0111-3>
- Rosa, R., Trübenbach, K., Pimentel, M.S., Boavida-Portugal, J., Faleiro, F., Baptista, M., Dionísio, G., Calado, R., Pörtner, H.O., Repolho, T., 2014. Differential impacts of ocean acidification and warming on winter and summer progeny of a coastal squid (*Loligo vulgaris*). *J. Exp. Biol.* 217, 518–525.
- Rosa, R., Trübenbach, K., Repolho, T., Pimentel, M., Faleiro, F., Boavida-Portugal, J., Baptista, M., Lopes, V.M., Dionísio, G., Leal, M.C., Calado, R., Pörtner, H.O., 2013. Lower hypoxia thresholds of cuttlefish early life stages living in a warm acidified ocean. *Proc. R. Soc. Lond. B Biol. Sci.* 280, 20131695. <https://doi.org/10.1098/rspb.2013.1695>
- Seeburg, P.H., Single, F., Kuner, T., Higuchi, M., Sprengel, R., 2001. Genetic manipulation of key determinants of ion flow in glutamate receptor channels in the mouse. *Brain Res.* 907, 233–243. [https://doi.org/10.1016/S0006-8993\(01\)02445-3](https://doi.org/10.1016/S0006-8993(01)02445-3)
- Sereewit, J., 2022. A-TO-I RNA EDITING IN OCTOPUS RUBESCENS IN RESPONSE TO OCEAN ACIDIFICATION 23.
- Shaun A Marcott, Jeremy D Shakun, Peter U Clark, Alan C Mix, 2013. A reconstruction of regional and global temperature for the past 11,300 years 339, 1198–201.
- Simon Andrews, n.d. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [WWW Document]. URL <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed 5.5.24).
- Slomovic, S., Laufer, D., Geiger, D., Schuster, G., 2006. Polyadenylation of ribosomal RNA in human cells. *Nucleic Acids Res.* 34, 2966–2975. <https://doi.org/10.1093/nar/gkl357>
- Song, L., Florea, L., 2015. Rcorrector: efficient and accurate error correction for Illumina

- RNA-seq reads. *GigaScience* 4, 48. <https://doi.org/10.1186/s13742-015-0089-y>
- Stoler, N., Nekrutenko, A., 2021. Sequencing error profiles of Illumina sequencing instruments. *NAR Genomics Bioinforma.* 3, lqab019. <https://doi.org/10.1093/nargab/lqab019>
- UniProt Consortium, 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
- Videvall, E., 2017. What's N50? [WWW Document]. *Mol. Ecol.* URL <https://www.molularecologist.com/2017/03/29/whats-n50/> (accessed 5.15.24).
- Xu, G., Zhang, J., 2014. Human coding RNA editing is generally nonadaptive. *Proc. Natl. Acad. Sci.* 111, 3769–3774. <https://doi.org/10.1073/pnas.1321745111>
- Zhou, X., Zheng, W., Li, Y., Pearce, R., Zhang, C., Bell, E.W., Zhang, G., Zhang, Y., 2022. I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. *Nat. Protoc.* 17, 2326–2353. <https://doi.org/10.1038/s41596-022-00728-0>

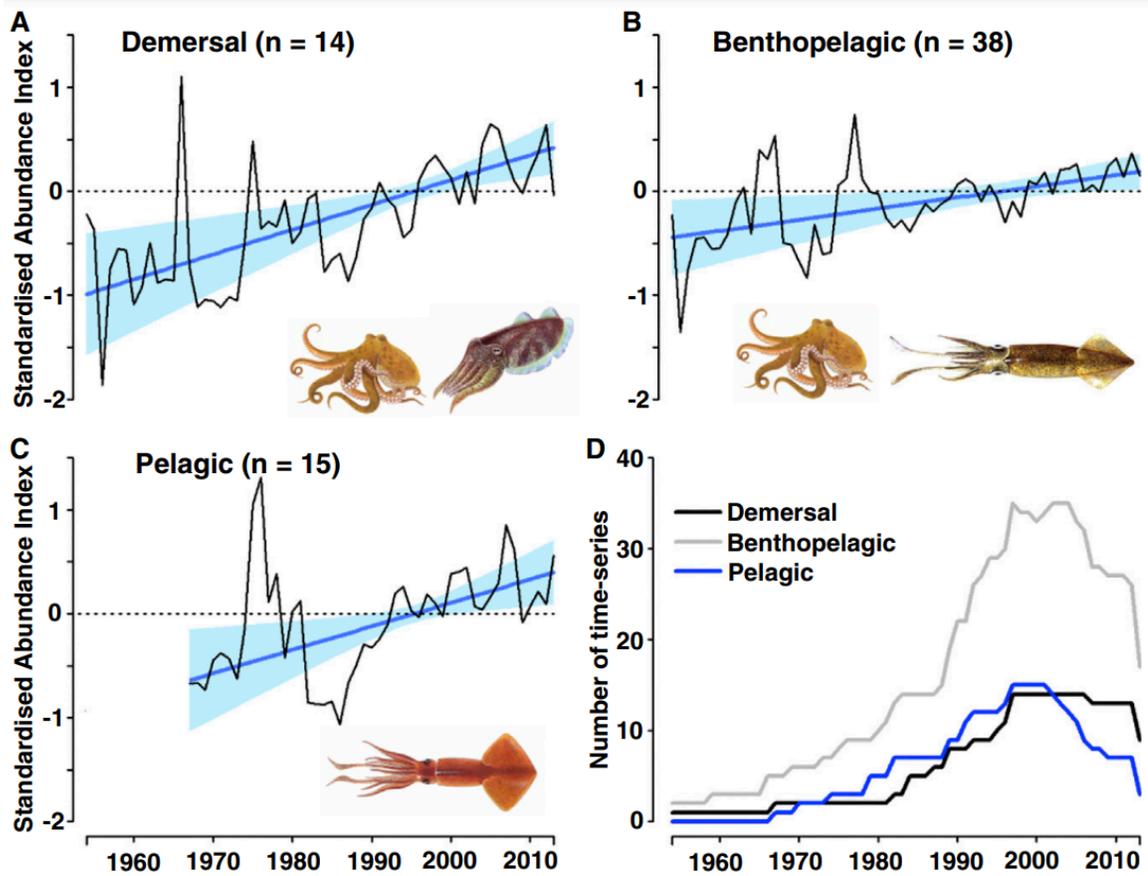


Figure 1. Trends in cephalopod abundance. Demersal, benthopelagic, and pelagic cephalopods show overall increasing abundance according to fisheries and survey data from 1953 to 2013. Reprinted from (Doubleday et al., 2016).

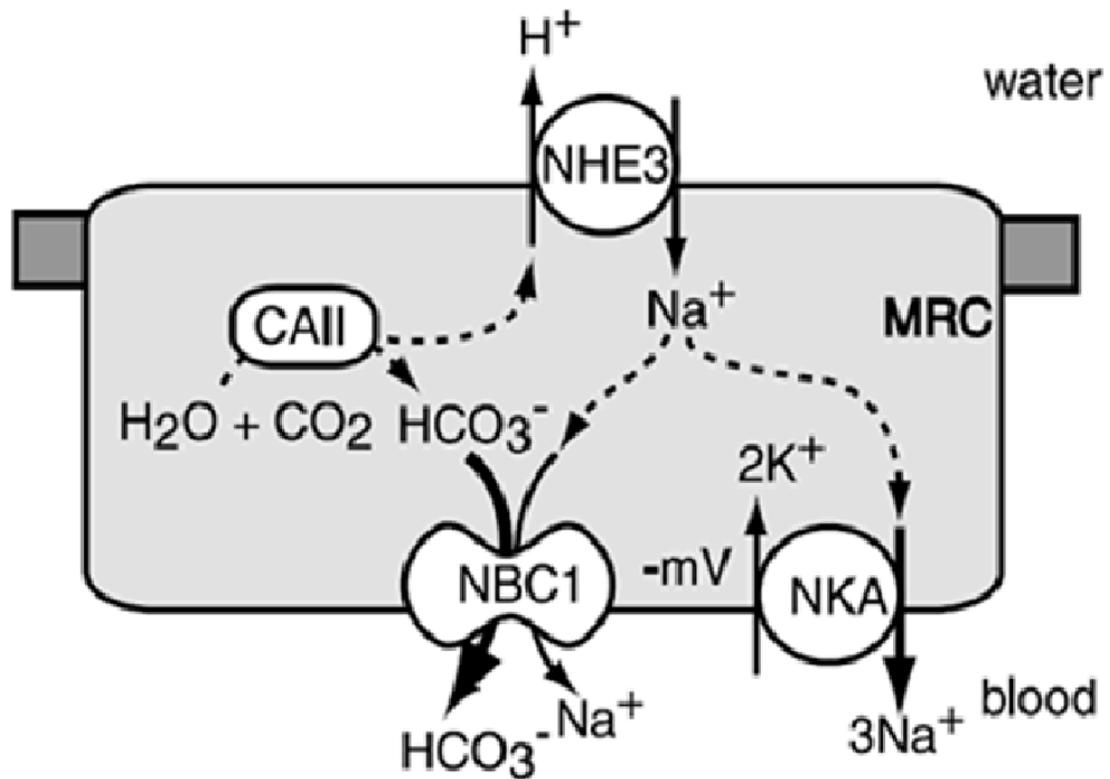


Figure 2. Acid secretion in cephalopod gill epithelium. Na^+/K^+ -ATPase (NKA) creates a negative potential inside the cell. This draws in sodium from the environment via a Na^+/H^+ exchanger (NHE3) while excreting H^+ , thereby raising intracellular pH. Increased intracellular pH leads to increased bicarbonate concentration via carbonic anhydrase (CAII), which is then transported into the blood alongside Na^+ via anion exchanger (NBC1), replenishing positive intracellular charge and pH gradient (Hu et al., 2011).

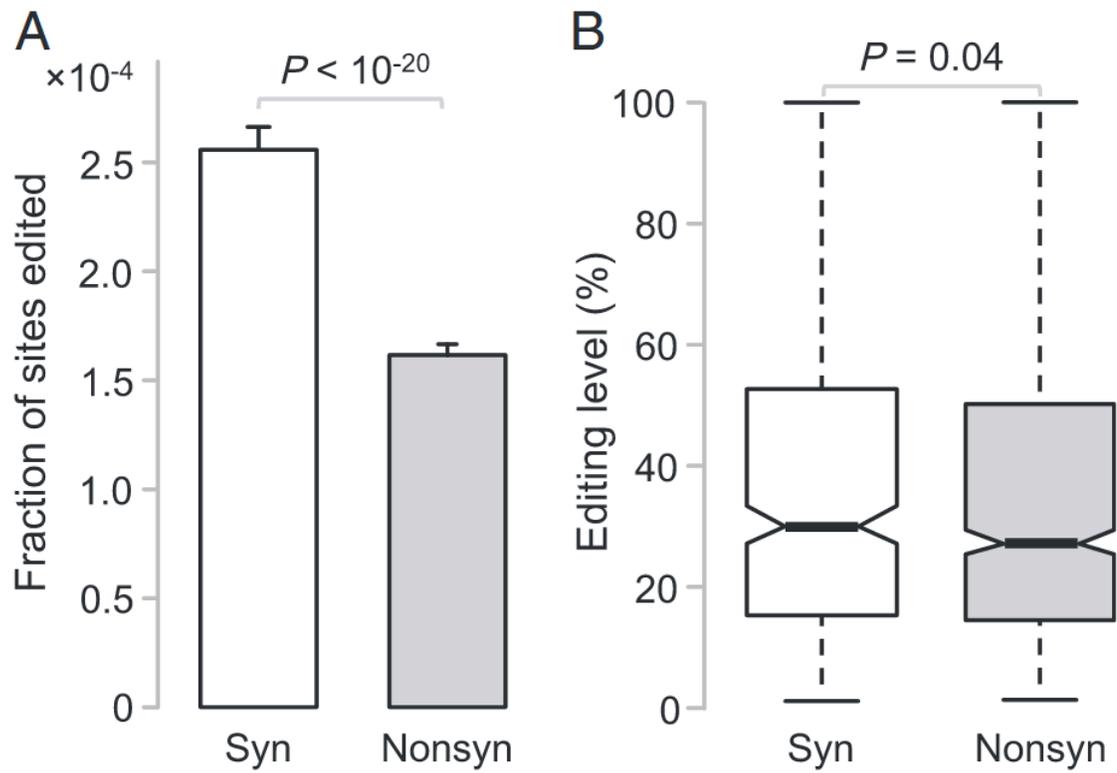


Figure 3. Frequencies and editing levels of synonymous (syn) and nonsynonymous (nonsyn) A-to-G editing in humans. (A) Out of all edited sites, the fraction which are synonymous is significantly greater than nonsynonymous. (B) Nonsynonymous sites are edited at a significantly lower level than synonymous sites. Reprinted from (Xu and Zhang, 2014).

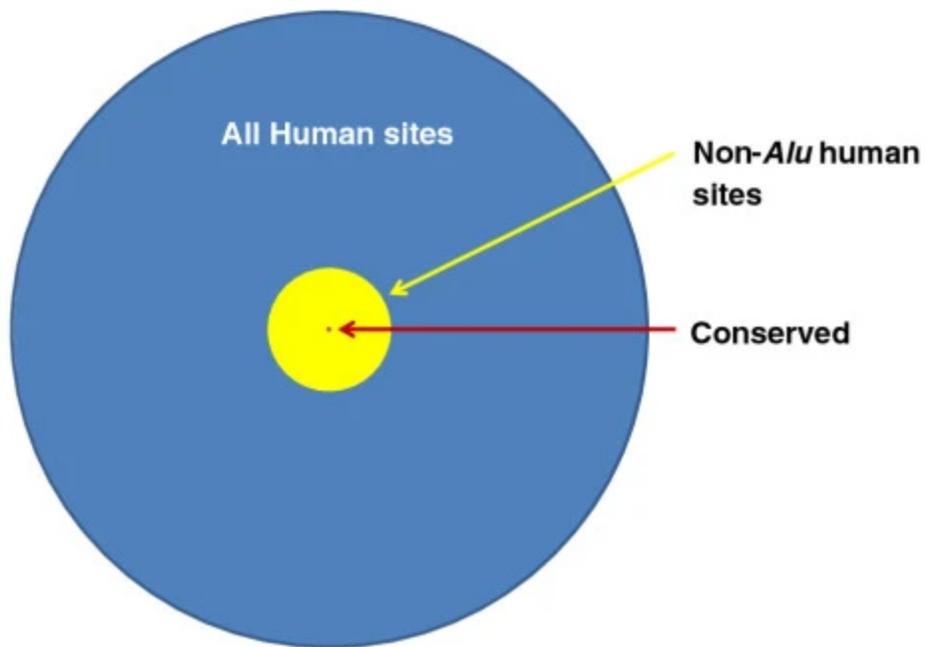


Figure 4. Mammalian set of editing sites. Out of 1,432,744 editing sites in humans (blue), non-*Alu* human sites (yellow) comprise 52,312 and only 59 are conserved between humans and mice. Reprinted from (Pinto et al., 2014).

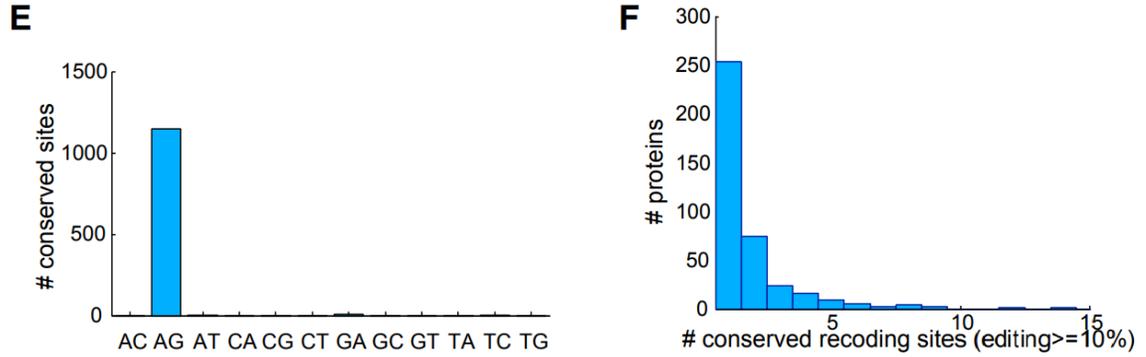


Figure 5. Extensive recoding is conserved across coleoid cephalopods. Out of 1,146 A-to-G modification sites conserved across coleoid cephalopods, 705 are edited at $\geq 10\%$ across 393 proteins. Reprinted from (Liscovitch-Brauer et al., 2017a).

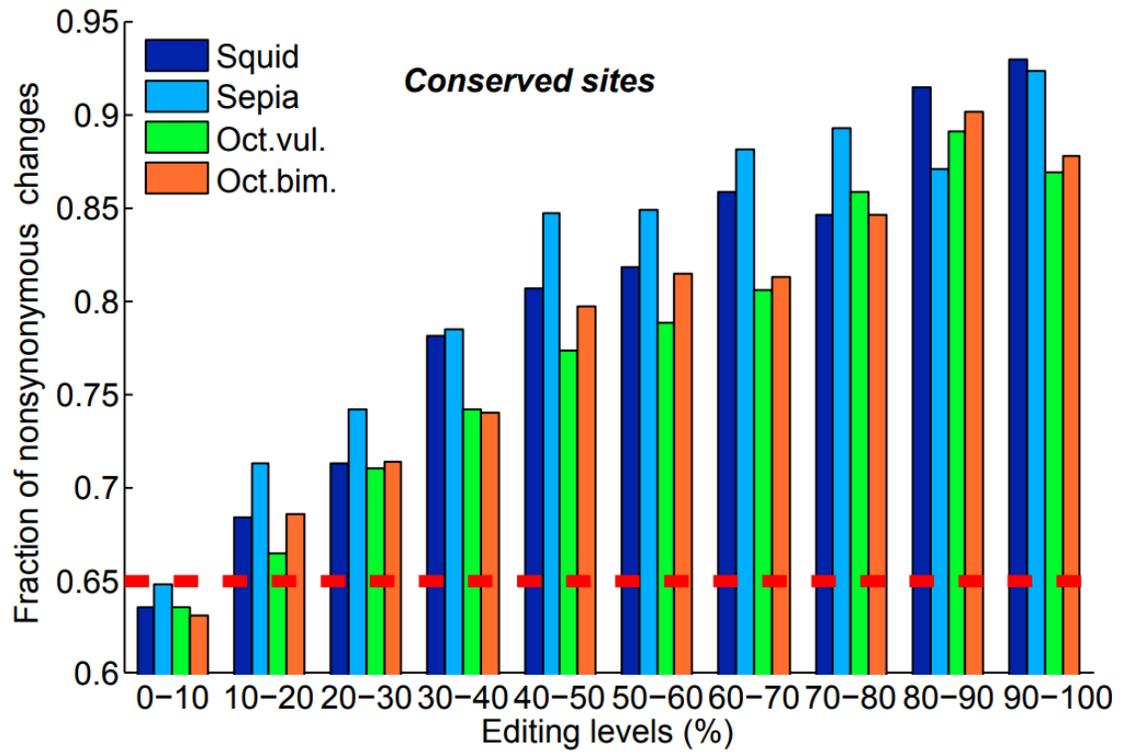


Figure 6. Signs for positive selection of recoding by editing. As editing levels increase, the fraction of sites which are recoding also increases. Red horizontal dashed line indicates the expected recoding fraction assuming neutrality. Reprinted from (Liscovitch-Brauer et al., 2017a).

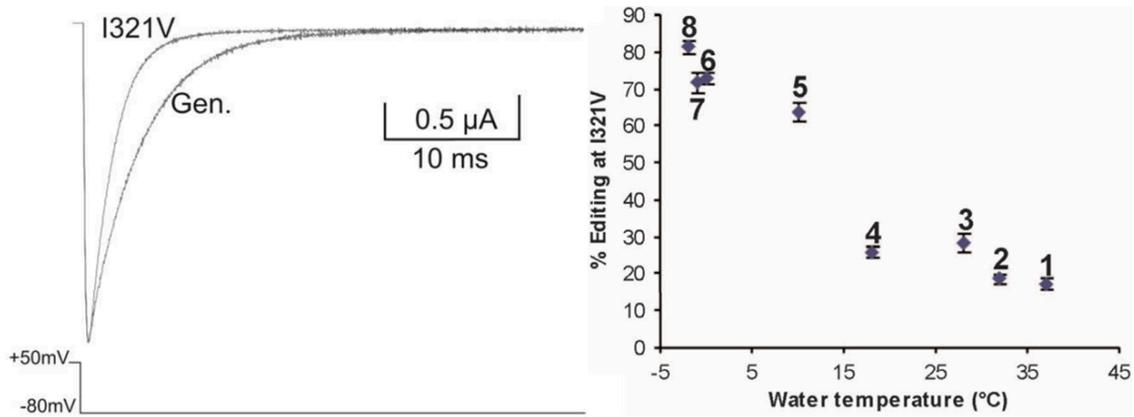


Figure 7. Closing kinetics and editing levels. (Left) Current trace of closing kinetics of genomic (Gen.) and edited (I321V) K⁺-channel from cold-adapted octopus. (Right) Editing percentages of eight octopus species by average temperature of water in which they were caught, assayed by poison primer extension. Reprinted from (Garrett and Rosenthal, 2012).

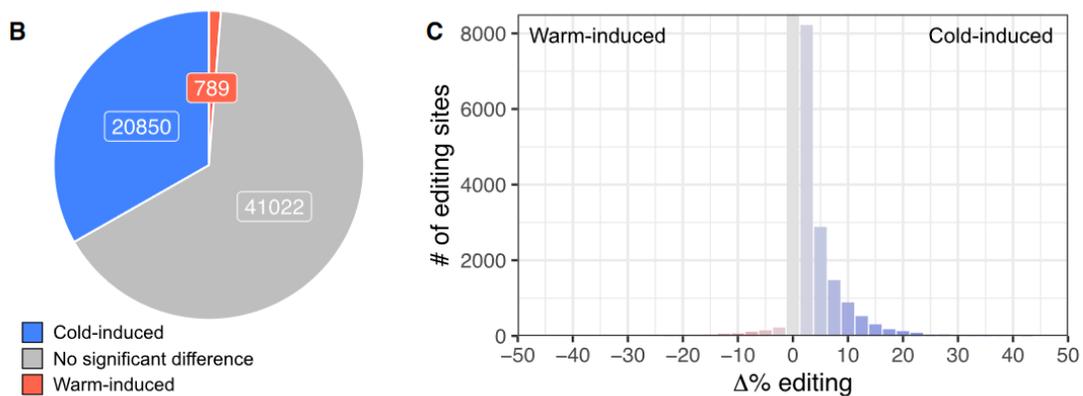


Figure 8. Cold-induced editing in *Octopus bimaculoides*. Reprinted from (Birk et al., 2023). Change in percent editing when comparing warm vs cold treatments. Sites which exhibit higher editing in cold water show as positive change, sites which exhibit higher editing in warm water show as negative change.

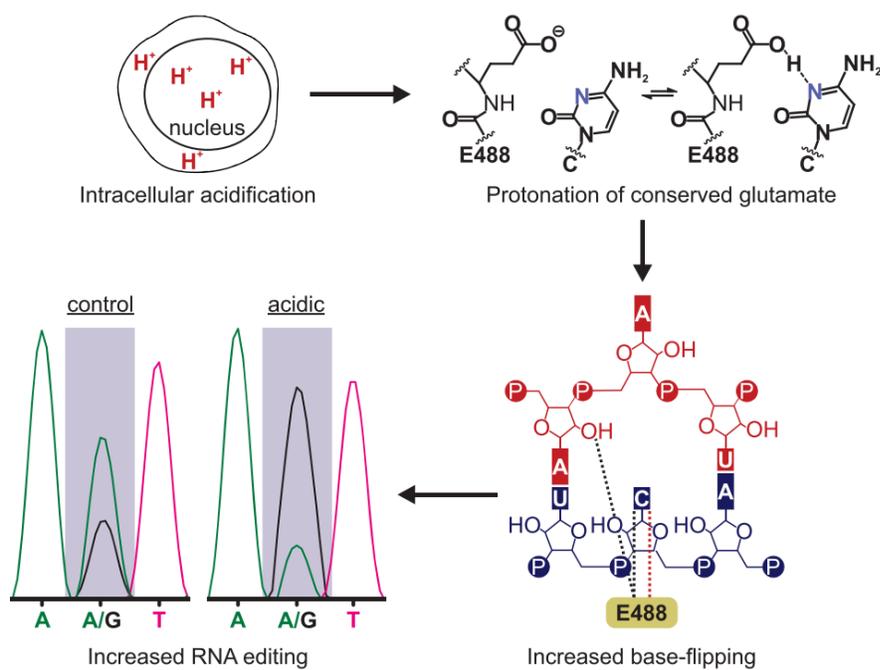


Figure 9. Summary of effect of intracellular acidification on ADAR editing efficiency. As intracellular pH drops, glutamic acid is more likely to be protonated, allowing base-flipping to occur and facilitating the completion of catalysis. Reprinted from (Malik et al., 2021).

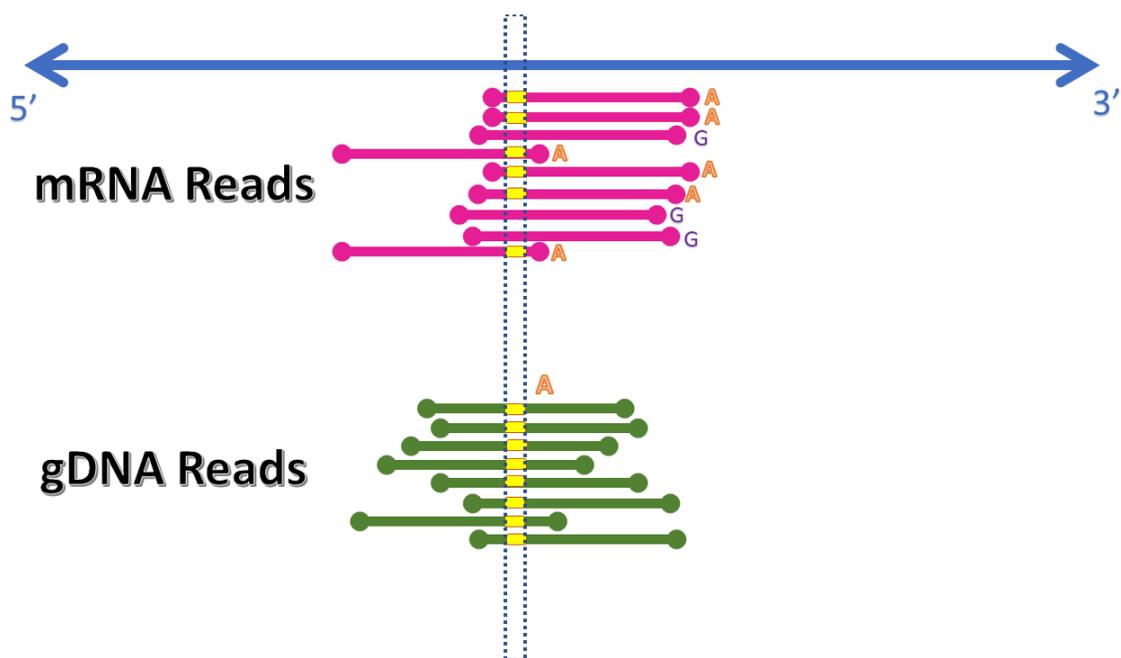


Figure 10. Abstract representation of mismatch detection by alignment. mRNA reads (pink) and gDNA reads (green) were aligned to the assembled transcriptome (blue). At loci where all gDNA reads agreed with one another, mRNA reads were compared to the transcriptome. Mismatches represent putative editing locations.

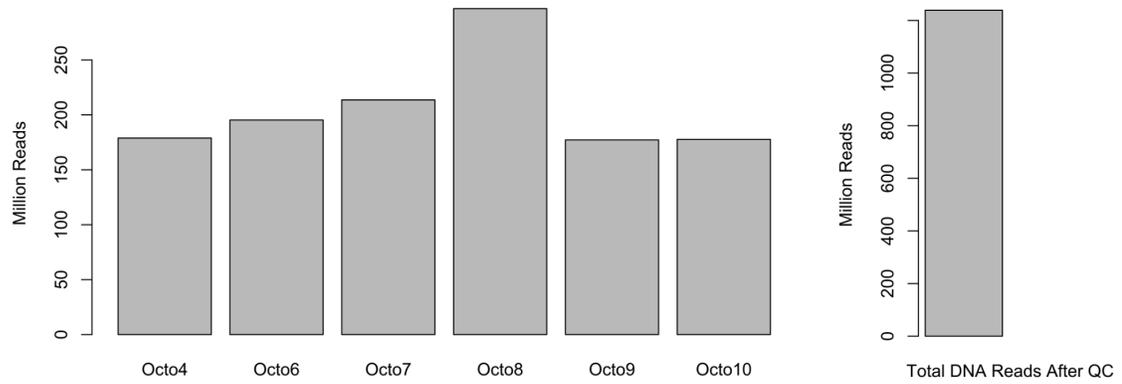


Figure 11. Yields of raw DNA reads from whole genome Illumina sequencing for each octopus (left). Total DNA after pooling and QC (right).

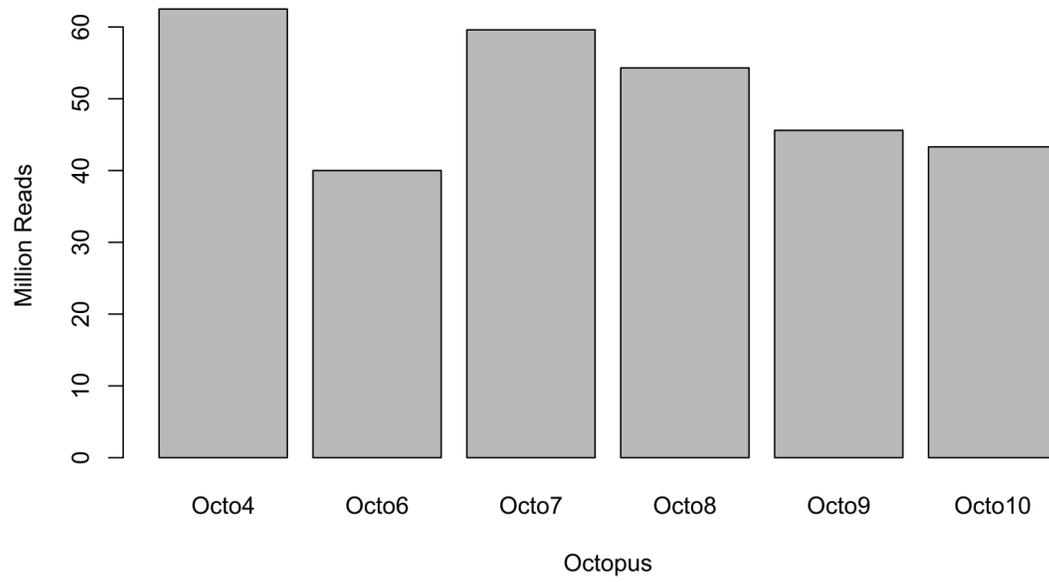


Figure 12. Raw RNA reads.

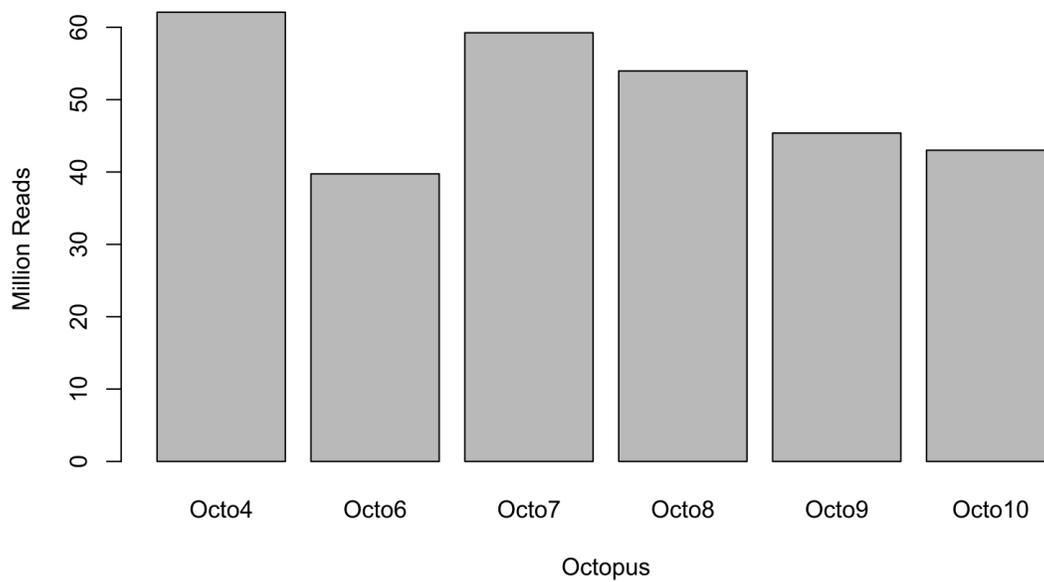


Figure 13. RNA reads after QC. No more than 400,000 reads were removed from each sample.

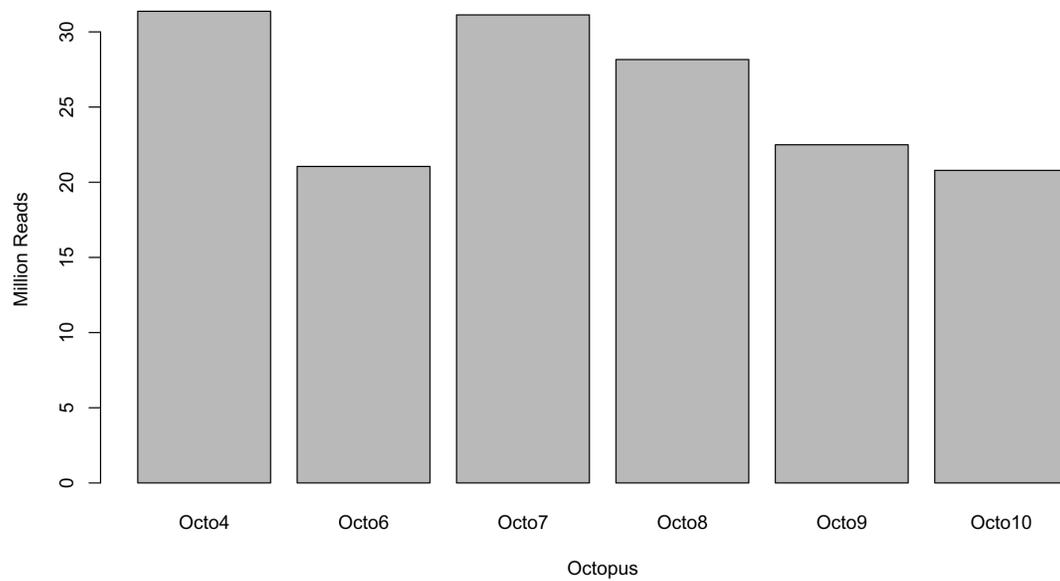


Figure 14. RNA reads mapped to assembled transcriptome per octopus.

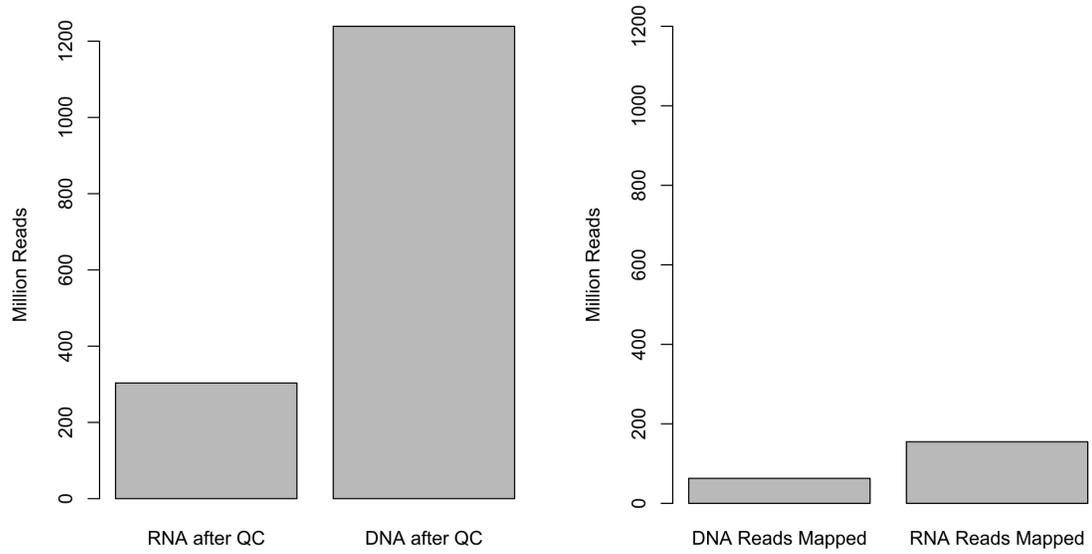


Figure 15. Numbers of RNA and DNA reads after QC (left), and numbers of RNA and DNA reads mapped to the assembled transcriptome (right).

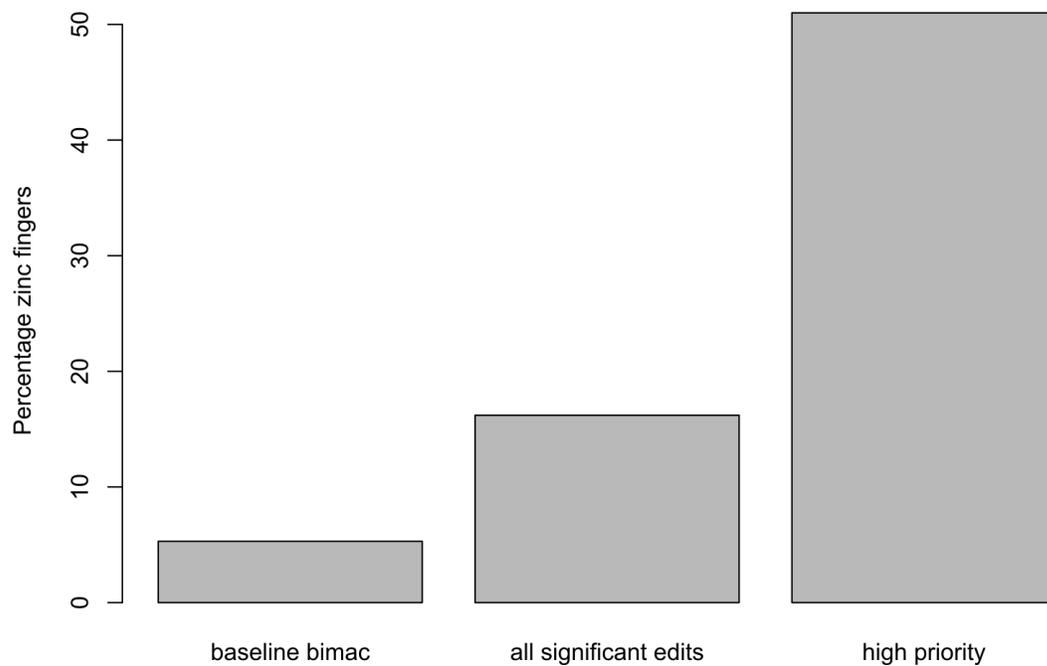


Figure 16. Zinc finger editing proportions. Zinc fingers are overrepresented in significantly differentially edited sites, and moreso in very highly edited sites.

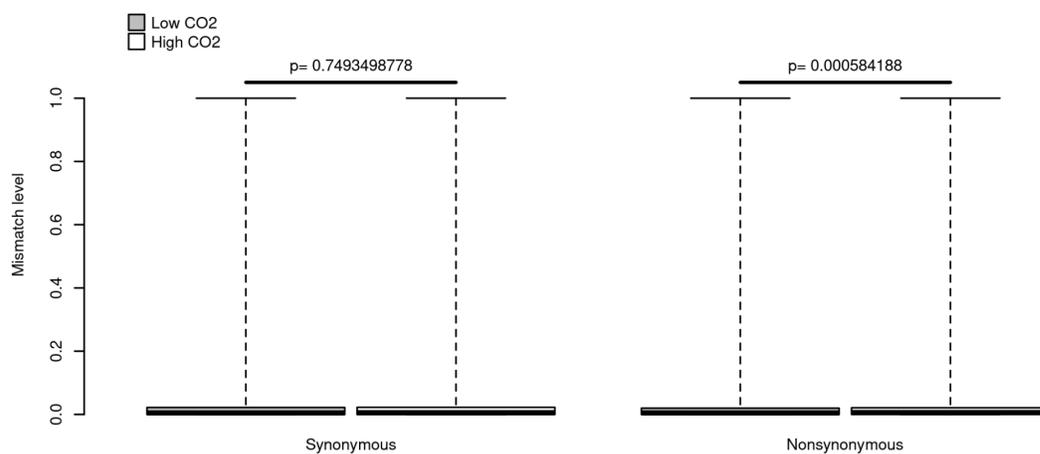


Figure 17. Total editing levels in synonymous and nonsynonymous editing sites by $p\text{CO}_2$ level of experimental treatment. Synonymous total editing levels were not significantly different between high and low $p\text{CO}_2$ treatments (Wilcoxon signed ranked test, $p=0.7493$). Nonsynonymous editing levels were significantly different between high and low $p\text{CO}_2$ treatments (Wilcoxon signed ranked test, $p=0.0005842$).

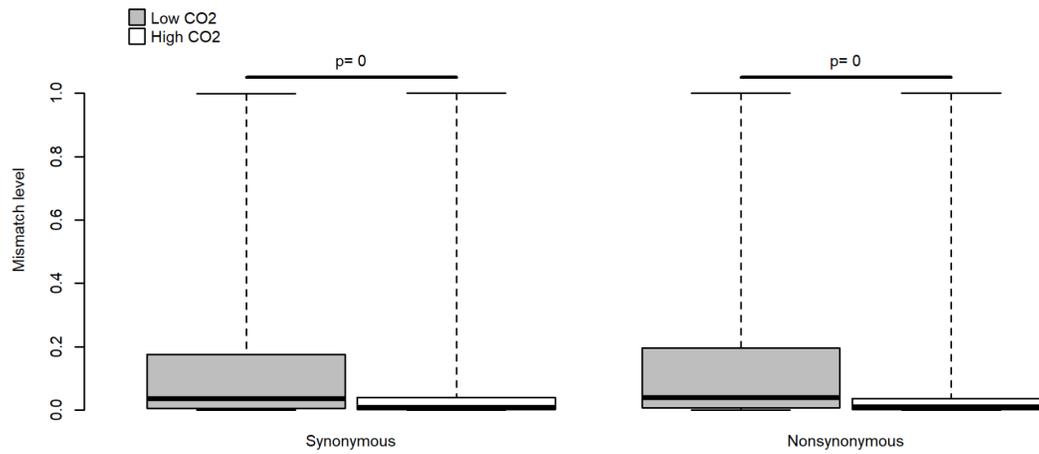


Figure 18. Levels of editing only among significantly differentially edited sites with respect to pCO_2 . Both synonymous and nonsynonymous edits were significantly suppressed in elevated pCO_2 compared to control. Wilcoxon signed rank test: synonymous $p < 2.2 \times 10^{-16}$, nonsynonymous $p < 2.2 \times 10^{-16}$.

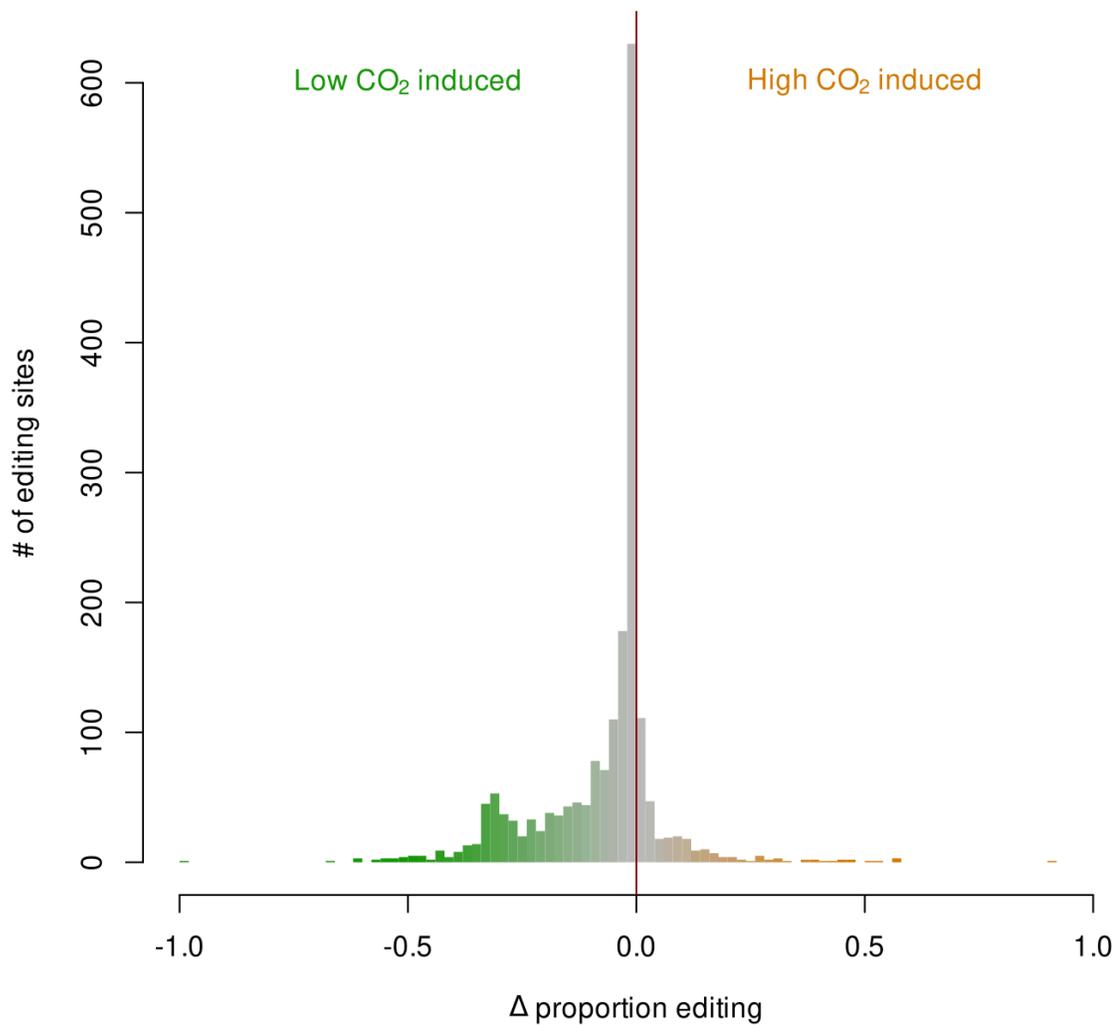


Figure 19. Change in %-editing from low to high pCO_2 . Sites which were edited more in elevated pCO_2 show a positive change in %-editing (orange, right). Sites which were edited less in elevated pCO_2 show a negative change in %-editing (green, left).

4

>|cl|TRINITY_DN2727_c0_g3_i1:9-200 ORF1_TRINITY_DN2727_c0_g3_i1:8:199

edit:73 Atg -> Gtg Met25-Val25

PREDICTED: Octopus bimaculoides zinc finger protein 91 (LOC106873964), mRNA

CTTCTCTCCTGTATGAATA**C**GTTTGTGTTTTGTTGAGTTACTACTTGCAGAGAATGTA
 TTACCACAGATATCACAGTGGT**A**TGGGTTCTCTCCTGTATGAATATGTTTGTGATTG
 TTAAACTACCTCTTGCAGAGAATGATTTACCACAGATATCACAGTGATATGGTTTTTCT
 CCTGTATGAATGCGTTTGTGATTAGTTAAG**TGGTCACTTGAGGAAATGCTT**TACCA
 CAGATATCACATTGATATGGT**CTCTCTCCTGTATGAGTGCG**TTTGTGAGTA

Editing site highlighted in red

Outer Primers

Forward Primer: **OR73F** **C**GTTTGTGTTTTGTTGAGTTACTAC

Reverse Primer: **OR73N** **CGCACTCATA CAGGAGAGAG**

Inner Primers

Forward Primer: **OR73F** **CGTTTGTGTTTTGTTGAGTTACTAC**

Reverse Primer: **OR73R** **AAGCATTTCCTCAAAGTGACCA**

Sanger Primer: **OR73R** **AAGCATTTCCTCAAAGTGACCA**

Length: 208

Figure 20. Designations within example ORF of editing site and primers for amplification and sanger sequencing.

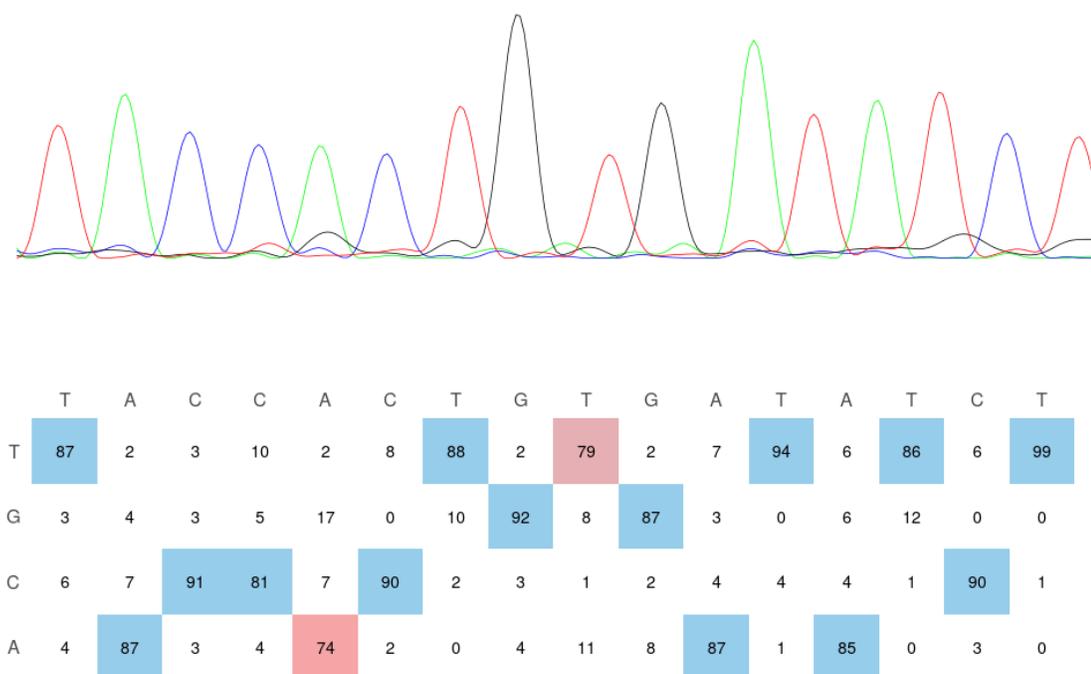


Figure 21. Example of sanger results for a putative edit from an octopus exposed to high pCO_2 . The genomic sequence at the leftmost position corresponded to an A/T pair, while RNASeq read as G. Sanger sequencing of the complementary strand should have shown some high percentage C after an A-to-G edit, but the overwhelming signal on the chromatogram reads T.

Record statistics

- Log10 record count (total 4.17k)
- Record length (total 3.82G)
- Longest record (24.7M)
- N50 length (5.77M)
- N90 length (1.11M)

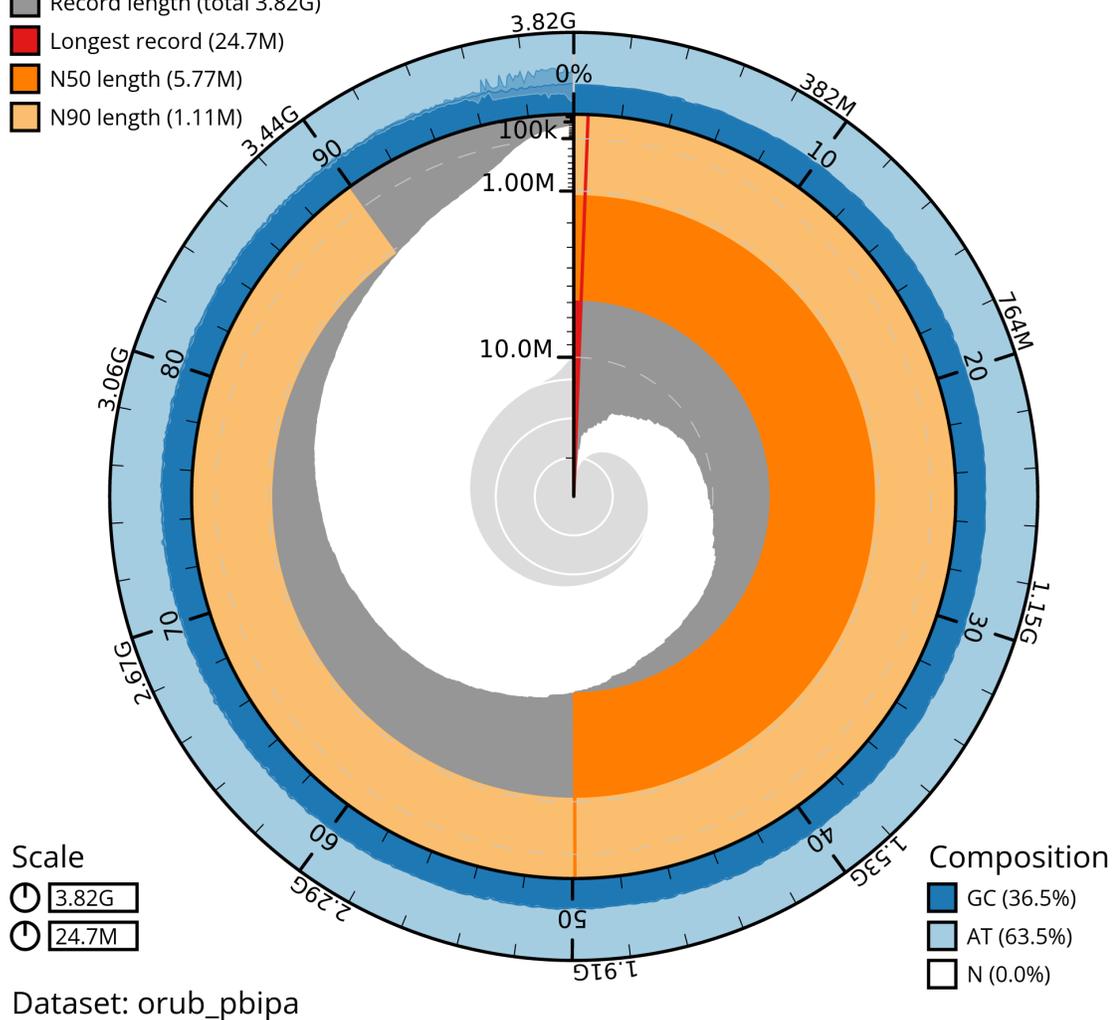
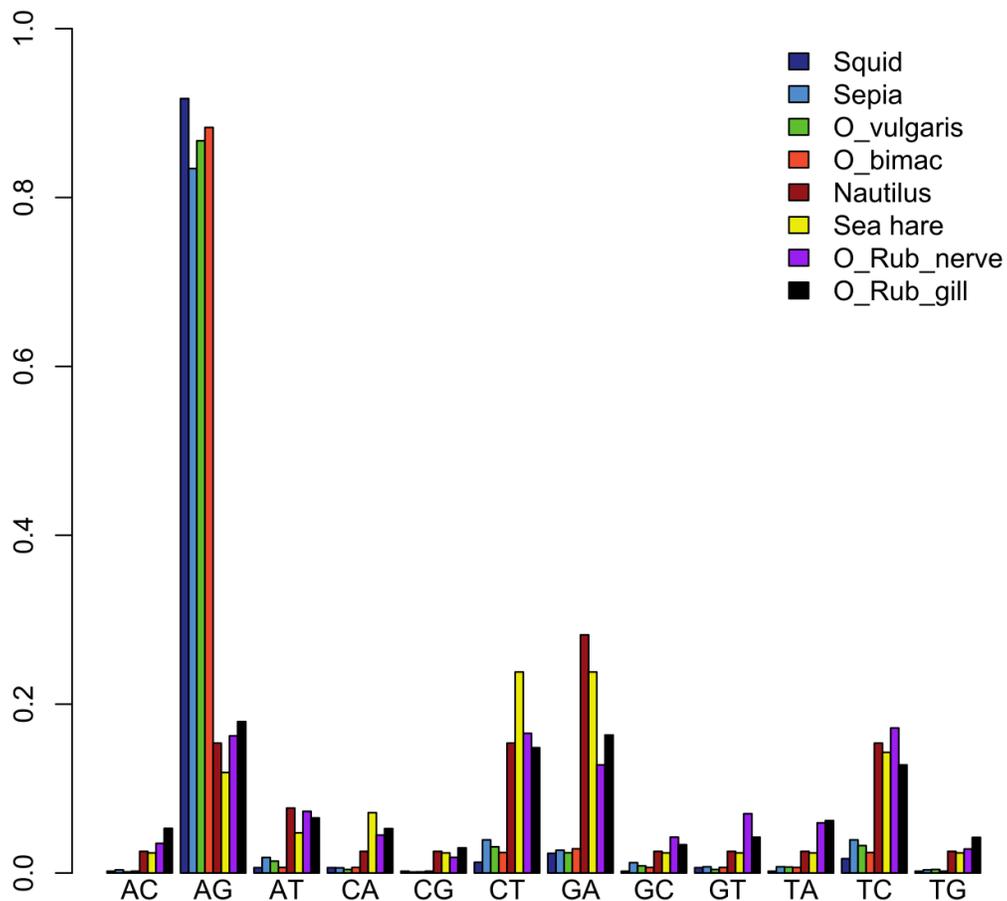


Figure 22. Snail plot of Improved Phase Assembler (IPA) genome assembly using BlobToolkit.



Figure

Figure 23. Fractions of mismatch types of different cephalopod species. All tissues are nervous except where specified. Squid and *Octopus rubescens* analyzed using Jaydee's pipeline, all others reprinted from (Liscovitch-Brauer et al., 2017b).

Table 1. Carbonate system parameters for elevated and control pCO₂ experiments for the 10-13 days following acclimation and preceding harvest. Values are means ± SD.

Treatment	pCO₂ (µatm)	pH	Alkalinity (µmol/kg)	Salinity (PSU)
Control	685 ± 91	7.753 ± 0.045	2092 ± 27	30.5 ± 0.3
Elevated CO ₂	1522 ± 253	7.495 ± .0.123	2080 ± 9	30.5 ± 0.4

Table 2. Editing percentages detected by our pipeline vs Sanger results for a single putative A-to-G edit across six octopuses.

	Octo 4	Octo 6	Octo 7	Octo 8	Octo 9	Octo 10
Illumina mRNA-seq	22.4	100	18.5	8.1	20	22.1
Sanger seq verification	0	2	2	4	4	1

Table 3. Comparison of assembly statistics across Octopus genomes available on NCBI.

Organism	Size (Mbp)	Level	Contig N50 (bp)	Scaffold N50 (bp)	Busco (single copy %)	Author
<i>Octopus bimaculoides</i>	2343	Chromosome	5,523	96,881,196	90.1	UC Berkeley
<i>Octopus sinensis</i>	2719	Chromosome	490,217	105,900	89.3	Yellow Sea Fisheries
<i>Octopus maya</i>	2342	Scaffold	2,361	89,813,523	65	Iridian
<i>Octopus mimus</i>	2057	Scaffold	1,141	80,823,707	25	Iridian
<i>Octopus rubescens</i>	2757	Scaffold	2,937	23,148,921	57	Iridian
<i>Octopus insularis</i>	1618	Contig	1,978			Rockefeller University
<i>Octopus jollyorum</i>	2567	Scaffold	1,973	103,274,698		Iridian
<i>Octopus americanus</i>	2535	Scaffold	1,553	100,565,213	42	Iridian
<i>Octopus vulgaris</i>	2800	Chromosome	3,577,857	118,905,486		KU Leuven
<i>Octopus vulgaris</i>	1773	Scaffold	3,040	265,914		Stazione Zoologica Anton Dohrn
<i>Octopus rubescens</i>	3820	Contig	5,768,000	5,770,000	87.5	Walla Walla University