A-TO-I RNA EDITING IN *OCTOPUS RUBESCENS* IN RESPONSE TO OCEAN ACIDIFICATION

by

JAYDEE SEREEWIT

A THESIS

submitted to

WALLA WALLA UNIVERSITY

in partial fulfillment of
the requirements for the
degree of

MASTER OF SCIENCE

11 Nov 2022

This thesis for the Master of Science degree

has been approved by the Department of Biological Sciences

and the Office of Graduate Studies

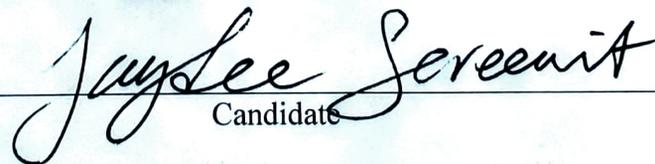Walla Walla University

_____
Major Professor

_____
Committee Member

_____
Committee Member

_____
Committee Member

_____
Dean of Graduate Studies

_____
Observer of the Process - Graduate Representative

_____
Candidate

11-28-22
_____
Date

Abstract

A-to-I RNA editing, the most common type of RNA editing in animals, is facilitated by a family of enzymes called adenosine deaminase acting on RNA (ADAR). A-to-I editing exists in many organisms, but A-to-I edits are significantly more common in coleoid cephalopods, especially nonsynonymous edits that alter the amino acid encoded, and potentially gene product function. In one case, A-to-I editing has shown to be involved in temperature adaptation of octopuses. As ocean acidity rises due to increasing anthropogenic carbon dioxide emission, research on how ocean acidification affects RNA editing in cephalopods has been nonexistent. In this study I examined the effects of ocean acidification on A-to-I editing in a common cephalopod species *Octopus rubescens* held for two weeks in elevated or control pCO_2. To identify potential mRNA editing sites, I adapted an approach of finding editing sites by aligning gill tissues mRNA and gDNA of individual organisms to their consensus transcriptome. Results indicate that A-to-I editing in the gill tissues of *Octopus rubescens* is not as prevalent as in the nervous tissues of other coleoid cephalopods. However, A-to-I editing does occur in the gill tissues of *Octopus rubescens*, as verified by poisoned primer extension assays and Sanger sequencing. More importantly, there are putative editing sites that exhibit differential editing levels due to increased pCO_2, suggesting the need for more research into those putative editing sites on whether they play an important role in how *Octopus rubescens* responds to ocean acidification.

Table of Contents

List of Figures

List of Tables

Introduction

RNA editing is a posttranscriptional modification of an RNA transcript at specific

nucleotides that is different from other RNA processing events such as intron splicing, 5'

capping, and 3' polyadenylation (Gott & Emeson, 2000). RNA editing exists in the forms

of insertion/deletion editing and substitution editing. Enzymes involved in

insertion/deletion editing include endonucleases, TUTases, exonucleases, RNA ligases,

and viral polymerase (Brennicke et al., 1999). Substitution editing includes

cytidine-to-uridine (C-to-U) editing and adenosine-to-inosine (A-to-I) editing. C-to-U

editing is carried out by the apolipoprotein B editing complex (APOBEC) family of

cytidine deaminases, whereas A-to-I editing is mediated by a family of enzymes called

adenosine deaminase acting on RNA (ADAR) (Brennicke et al., 1999). In animals A-to-I

editing is the most abundant form of RNA editing (Eisenberg & Levanon, 2018).

ADAR enzymes exist in all animals. All ADAR enzymes contain at least one

double-stranded RNA (dsRNA) domain in the N-terminus and a deaminase domain in the

C-terminus (Yablonovitch et al., 2017). Figure 1 shows structural diagrams of different

ADAR isoforms in mammals, fruit flies, and cephalopods. Adenosine is an adenine

attached to a five-carbon sugar in RNA whereas an inosine is a hypoxanthine attached to

a five-carbon sugar. ADAR converts an adenosine to an inosine through hydrolytic

deamination (Figure 2) (Vesely & Jantsch, 2021). More specifically, this mechanism

comprises three steps. First, ADAR binds to the dsRNA. Then the targeted adenosine is

flipped out of the RNA duplex and into the ADAR active site. The hydrolytic

deamination at the position 6 of the purine ring takes place and the adenosine is

converted to an inosine (Malik et al., 2021). A-to-I editing is manifested as A-to-G when

comparing genomic and cDNA sequences because of the cytidine to inosine base-pairing (similar to cytidine to guanosine base-pairing) during reverse transcriptase-mediated first-strand cDNA synthesis (Rueter et al., 1999). Many cellular machinery also recognize inosine as guanosine (Paz-Yaacov et al., 2010).

A-to-I RNA editing is a complex and dynamic process that is regulated by many factors, including cis-acting and trans-acting regulatory elements as well as internal and external physiological stimuli (Vesely & Jantsch, 2021). From the cis-acting element perspective, the underlying RNA sequence and the ability for different regions of the RNA molecule to base pair and form double stranded structure is paramount for RNA editing because ADAR binds to double-stranded RNA (Eisenberg & Levanon, 2018). Unsurprisingly, repetitive genomic sequences with inverted repeats in humans have been shown to be heavily edited as inverted repeats are major sites of recruitment for ADARs (Kim et al., 2004). The specific sites in the RNA sequence and their upstream and downstream nucleotides also influence editing activities. More specifically, G depletion is observed in one nucleotide upstream of the editing sites and G enrichment is observed in one nucleotide downstream of the editing sites. These editing preferences have been observed in humans, mice, fruit flies, and cephalopods (Picardi et al., 2015; Yablonovitch et al., 2017).

Trans-acting regulatory elements like ADARs and other RNA binding proteins (RBPs) can impact editing in numerous ways. Protein modifications of ADARs such as SUMOylation of ADAR1 at lysine 418, phosphorylation of specific sites on ADARp110 and ADAR2, and ubiquitination of ADARp110 upon interferons stimulation have all been shown to decrease editing efficiency (Vesely & Jantsch, 2021). The competition

between ADAR and other RNA binding proteins as well as other ADAR proteins within the family also affect editing levels. This is demonstrated in the brain where ADAR3 competes with ADAR1 and ADAR2 to bind to RNA and lower editing levels (Tan et al., 2017). Many RNA binding proteins that affect editing have also been identified. Some impact editing at a more global level while others regulate editing in a clear site-specific manner. Among them, NF90, NF90/ILF2 and NF90/ILF3 are RBPs that have been shown to negatively influence editing (Quinones-Valdez et al., 2019) while Zn72D has been demonstrated to enhance editing (Sapiro et al., 2020).

A-to-I editing responds to some environmental stimuli and physiological changes. In *Drosophila*, researchers found that changes in temperature outside of normal physiological range alter ADAR expression levels and editing patterns that are conserved across different *Drosophila* species (Rieder et al., 2015). In a human embryonic kidney cell line coexpressing 5HT2C, GluA2, Gli1 transcripts and ADAR1 and ADAR2, intracellular acidification enhanced editing activity by as much as 40%. The enhancement in editing is explained by increased protonation of a conserved glutamate residue in the ADAR base-flipping loop. When this glutamate residue is protonated to glutamine in acidified condition, it better stabilizes the flipped-out conformation by occupying the space vacated by the flipped-out adenosine and hydrogen bonding with the complementary-strand orphaned base (Malik et al., 2021).

The impact of RNA editing is many fold. RNA editing can affect RNA folding and its binding affinity to proteins (Wang et al., 2005). Edited RNA can produce a non synonymous change in the amino acid sequence and its encoded proteins, allowing the diversification of protein products from RNA transcripts in a way that is not encoded in

the genome (Walkley & Li, 2017). In addition, RNA editing has been found to be critical in maintaining the proper functioning of an organism in many examples, particularly in the nervous systems. In mice, A-to-I RNA editing in AMPA (GluR-B) glutamate receptors change the corresponding glutamine codon to an arginine codon and influence the ion flow properties in glutamate-gated channels (Sommer et al., 1991). Underediting in the AMPA receptors of ADAR2 mutant mice causes postnatal seizure and early death (Higuchi et al., 2000). In *Drosophila*, dADAR mutants suffer neurological abnormalities such as time-sensitive paralysis, locomotor incoordination, and tremors (Palladino et al., 2000).

Even though RNA editing has been observed in all domains of life, cephalopods have been a particularly interesting study subject. In the nervous tissues of coleoid cephalopods alone, researchers found more than 70,000 A-to-I recoding sites, which are editing sites that produce changes in the amino acid sequence, by far the most known in any animal; In contrast, about 3,000 recoding sites have been found in humans and 1,000 in *Drosophila* in total (Yablonovitch et al., 2017). About 60% of the ≈12,000 open reading frames (ORFs) in the nervous tissues of the common squid *Doryteuthis pealeii* are A-to-I edited, and most of the transcripts contain multiple recoding sites (Alon et al., 2015). Not only are recoding sites more common in coleoid cephalopods compared to other species, some of the recoding sites have been shown to facilitate environmental adaptation. For example, Garrett and Rosenthal discovered that one instance of A-to-I editing caused a recoding event in the delayed rectifier K^+ channel gene and this recoding site was extensively edited in Antarctic and Arctic octopuses but not in tropical octopuses. The opening of K^+ channels is responsible for the falling phase of action

potentials and the neuron membrane's return to resting potential, and this process is particularly sensitive to temperature. Cold temperature slows down the gating kinetics of K^+ channels, making action potentials disproportionally broad and severely limiting the repetitive firing of actions potentials. This recoding event enables Antarctic and Arctic octopuses to adapt to colder temperature by changing an isoleucine to a valine in the K^+ channel's pore and accelerating its gating kinetics (Garrett & Rosenthal, 2012).

In May of 2021, atmospheric carbon dioxide concentrations averaged 419ppm, the highest since accurate measurements started in 1958 (US Department of Commerce, n.d.), and 50% higher than preindustrial concentrations (Gingerich, 2019). Approximately 30% of atmospheric carbon dioxide is absorbed by the ocean while the rest stays in the atmosphere or gets taken up by the terrestrial biosphere (Gruber et al., 2019). When the ocean absorbs carbon dioxide, it reacts with water to form carbonic acid, which then dissociates into bicarbonate and hydrogen ions. The release of hydrogen ions increases acidity and decreases oceanic pH (Doney et al., 2020). Atmospheric carbon dioxide has been increasing at an average pace of 5% per year, resulting in a surface oceanic pH decrease from 8.2 to 8.1 over the last 100 years (Orr et al., 2005). A 0.1 drop in pH might seem insignificant, but it represents a 26% increase in hydrogen ion concentration because pH is calculated in logarithmic scale. By 2100, a further pH decrease of 0.5 in some places has been projected (Barford, 2013). This is alarming because the U.S. Environmental Protection Agency recommends a fluctuation of less than 0.2 units for marine aquatic life (Zeebe et al., 2008).

The impact of ocean acidification is significant and widespread, ranging from changes in marine organism physiology and population dynamics to altered communities

and ecosystems (Doney et al., 2020). For example, ocean acidification shifts the seawater acid-base balance, which poses challenges particularly for calcifying marine organisms such as corals and some plankton to maintain their external calcium carbonate skeletons (Orr et al., 2005). Ocean acidification also hinders the development of internal calcified structures such as statoliths in cephalopods and otoliths in fish, which are vital in sensing gravity and movement (Kaplan et al., 2013; Munday et al., 2011).

Although cephalopods are weak osmoregulators, they have strong acid-base regulation capacity, only having minor decrease in intracellular pH when exposed to elevated seawater pCO_2 (Gutowska et al., 2010). The mechanism of cephalopod's acid-base regulation is not fully understood, but their gills are the primary organ for acid-base regulation (Hu et al., 2015). During exposure to elevated pCO_2, cephalopods experience respiratory acidosis and show a minor drop in blood pH and an increase in pCO_2. CO_2 diffuses across biological membranes and is hydrolyzed to form bicarbonates and protons. To regulate this acid-base disturbance, the protons are exported across the apical membrane by the Na^+/H^+-exchanger 3 while the bicarbonates are transported by electrogenic Na^+/HCO_3^- cotransporter across the basolateral membrane and into the bloodstream to buffer the lowering of pH caused by excess protons (Figure 3) (Hu et al., 2015).

Besides acid-base regulation, limited studies have investigated cephalopod responses to ocean acidification. Most work have focused on the effects of ocean acidification on early development (Lacoue-Labarthe et al., 2011; Rosa et al., 2014), metabolism (Birk et al., 2018; Onthank et al., 2021) and immune response (Culler-Juarez & Onthank, 2021). No one has looked at the effects of elevated environmental pCO_2 on

RNA editing in cephalopods. This, combined with the facts that cephalopods harbor more recoding sites than any other animals and that RNA editing has been shown to facilitate environmental adaptation in cephalopods, spurred the interest of this study. The purpose of this study is to examine if A-to-I RNA editing in the gill of *Octopus rubescens* is impacted by short-term elevated pCO_2. This purpose was divided into two primary aims: 1) to determine if there is an overall A-to-I editing response to elevated pCO_2, and 2) to determine if editing frequency of specific sites changes in response to elevated pCO_2.

Methods

The bioinformatics portion of this research used an established method with some modifications that utilizes matched gDNA and mRNA sequencing data from the same individual octopuses and a consensus transcriptome assembled from mRNA obtained from all octopuses in the study to identify A-to-I editing sites (Alon et al., 2015). The workflow for editing sites screening and differential edits detection and verification are outlined in Figure 4. In brief, mRNA reads from the gill tissues of each octopus and gDNA reads of all octopuses were aligned to the consensus transcriptome constructed from the mRNA reads of all octopuses. Mismatches between the consensus transcriptome and homozygous gDNA sites were considered potential weak editing sites, sites where less than 50% of the reads were edited (Figure 5). A binomial test was applied to distinguish editing events from sequencing errors. Loci at which all gDNA reads differed from the consensus transcriptome were characterized as potential strong editing sites (Figure 5). The editing level at each site for each octopus was calculated and a permutation t-test was applied to determine whether editing levels are significantly different between octopuses in elevated pCO_2 and control pCO_2 treatments. To verify these potential editing sites. A separate, similar experiment was conducted where a group of octopuses were subjected to elevated pCO_2 and control pCO_2 treatments. Their extracted RNA were used to verify editing sites with differential editing levels between the elevated pCO_2 and control pCO_2 groups via poisoned primer extension assays and Sanger sequencing.

Octopus Collection and Acidification Treatments

In summer 2016 six octopuses were collected from Admiralty Strait near Driftwood Park on Whidbey Island, transported to Rosario Beach Marine Lab, and each placed into individual 27.5 liter enclosures. Elevated pCO_2 (n=3) and control pCO_2 (n=3) treatment enclosures were supplied with seawater from 415 liter chilled mixing aquaria held at 11 °C. Mixing aquaria were equipped with a pH-stat system to maintain aquaria at a controlled pH corresponding to the desired pCO_2. The pH-stat system consisted of a Vernier pH glass electrode connected to a laptop through a National Instruments/Vernier SensorDAQ. Custom software opened a solenoid valve on a CO_2 regulator connected to standard aquarium bubbler airline and bubble stone in the aquarium if the measured pH rose above a specified threshold. Glass pH electrodes were calibrated daily against NIST buffers.

Seawater alkalinity of the mixing aquaria and unmodified seawater system outflow was determined daily by open-cell titration (Dickson et al., 2007) and alkalinity values were calculated from titration data using the 'seacarb' package in R (Gattuso et al., 2021). Measured alkalinity was used to calculate target pH to maintain desired pCO_2. The pH of seawater samples was independently measured immediately after collection from aquaria for alkalinity measurements by glass electrode. This measurement was used to verify the measurements of the pH-stat system and was also used to calculate pCO_2 achieved in the aquaria.

The pCO_2 of all systems was initially set to match the pCO_2 of the seawater where the octopuses were collected, which was approximately 700 μatm. Octopuses were kept in their native pCO_2 for at least one week before experimental pCO_2 were introduced.

Aquaria were brought to target pCO_2 stepwise over the course of 24 hours, which remained 700 μatm for control animals, and was 1500 μatm for treatment animals. Three octopuses were kept in closed aquaria with an actual measured average pCO_2 of 1517±254 μatm, and the other three octopuses kept in closed aquaria with an average pCO_2 of 735±47 μatm. The octopuses were treated for two weeks starting on August 3, 2016 and were taken out of treatments on August 17, 2016 when those octopuses were sacrificed using ethanol and their gill and other tissues were collected. Ethanol euthanization protocol consisted of placing the octopus into 1 liter of seawater, and gradually raising the ethanol concentration to 2.25% over 10 minutes, then, when the octopus was fully sedated, raising the ethanol concentration rapidly to 10%.

For the verification experiment, a different control system was used (Culler-Juarez & Onthank, 2021). In summer 2020, six octopuses were collected from Admiralty Strait near Driftwood Park on Whidbey Island and each placed into individual 113.5L insulated cooler tanks. The pCO_2 and temperature of each tank were controlled by custom tank control hardware, which received input from a three-wire PT-100 temperature probe and a single junction glass pH electrode inserted into the water through holes drilled in the cooler tank lids. The temperature of each cooler tank was set to match the temperature where the octopuses were collected at 10.8 °C. Each temperature probe was calibrated using a high precision NIST-traceable calibrated alcohol thermometer, and the pH electrodes were each calibrated by a spectrophotometric measurement of the seawater sample pH in their respective tanks. The alkalinity of the seawater sample was determined by open-cell titration (Dickson et al., 2007). The salinity was measured with a Vernier salinity probe. The spectrophotometric measurements,

alkalinity, salinity, and temperature of the seawater samples were used to calculate the target pCO_2 using the seacarb package in R (Gattuso et al., 2021). Carbonate chemistry data and calculation can be found at Sereewit & Onthank, 2021.

For two weeks, three octopuses were kept in the cooler tanks at an average pCO_2 of 1460±151 µatm, and the other three octopuses kept at an average pCO_2 of 757±81 µatm. Following treatments, the octopuses were anesthetized using 2.25% ethanol without euthanization, departing from the protocol described above, and their gill tissues were collected. Octopuses were allowed to recover immediately after tissue collection in a well-oxygenated 2 liter seawater bowl, then returned to their holding tank after ~15 min. Immediately after tissue collection, part of their gill tissues was used to extract RNA using a Qiagen RNEasy Minikit following the manufacturer's instructions and kept frozen at -20 °C while the rest of their gill tissues were stored in a -80 °C freezer.

RNA Sequencing

RNA extracted from the gill tissues of the octopuses in the 2016 experiment was sent to the Genomics and Cell Characterization Core Facility at University of Oregon for further processing and sequencing. There the integrity of the isolated RNA was determined by Advanced Analytical Fragment Analyzer. Kapa Biosystems Stranded mRNA-seq kits were used to isolate mRNA and prepare cDNA for sequencing. Single-end 100bp reads were obtained for each of the six samples on a single lane of an Illumina HiSeq 4000 sequencer. From the three octopuses in control pCO_2 treatment 69.1 million, 68.1 million, 50.6 million RNA reads were obtained, and 42.9 million, 59.0 million, and 59.1 million RNA reads were obtained from the other three octopuses in elevated pCO_2 treatment.

RNA-seq Data Quality Control, Error Correction, and rRNA Removal

Quality control on RNA-seq data was done using Fastp with default settings (Chen et al., 2018). Because sequencing errors can impact transcriptome assembly quality, Rcorrector was used for error corrections and labeling uncorrectable errors on RNA-seq data (Tables 1) (Song & Florea, 2015). Unfixable reads were removed using a custom Python script FilterUncorrectabledSEfastq.py (Sereewit & Onthank, 2021). Poly-A selection was used in the RNA-prep protocol to select for mRNA and remove other types of RNA. However, some small number of rRNA molecules are polyadenylated (Slomovic et al., 2006) so this strategy typically does not remove all rRNA. To remove remaining rRNA sequences, the RNA-seq reads were mapped to *Octopus vulgaris* 18s rRNA sequences and *Octopus cyanea* 28s rRNA sequences as published on the SILVA rRNA database (Glöckner et al., 2017). The *Octopus vulgaris* 18s rRNA sequences and *Octopus cyanea* 28s rRNA sequences were combined in a single fasta file for which a Bowtie2 index was built. The RNA-seq reads were mapped to the combined rRNA fasta file using Bowtie2's very sensitive local alignment while keeping the other parameters default. Only reads that did not align to the rRNA sequences were used for transcriptome assembly and further analysis. Table 2 shows the number of rRNA sequences removed from each octopus.

Genomic DNA sequencing and Quality Control

gDNA were extracted from leftover tissue samples from the octopuses in the 2016 experiment using Thermo Scientific GeneJET Genomic DNA Purification Kit following the manufacturer's instructions. These gDNA samples were sent to Novogene and sequenced on the Illumina PE150 (paired-end 150bp) platform. Each octopus gDNA

sample, with exception to one in control pCO_2 treatment that did not have leftover

tissues, was sequenced at a 6X coverage. 85.3 million, 82.2 million, and 83.6 million

reads were obtained from three octopuses in control pCO_2 treatment. 100.7 million and

84.3 million reads were obtained from two of the three octopuses in elevated pCO_2

treatment. Quality of gDNA reads was assessed using FastQC and Trim Galore and was

used for quality and adapter trimming. Surprisingly, Trim Galore did not change the

number of reads in each sample, most likely because Novogene had already done quality

trimming on the gDNA reads. All gDNA reads were combined (436.2 million paired-end

reads) to create pooled gDNA reads that would be used for alignment to the

transcriptome. The purpose of combining the gDNA reads from different octopus

individuals was to increase coverage.

Consensus Transcriptome Assembly and Sequence Alignment

Consensus transcriptome was assembled using Trinity with processed mRNA

reads from all six octopuses in the 2016 experiment. Because these mRNA reads are

single end, the read orientation parameter was set to "--SS_lib_type R". Other parameters

were kept as default. Only the ORFs of the transcriptome were used to focus on

nonsynonymous editing sites (i.e. editing sites that alter the amino acid). The ORFs of

Octopus rubescens consensus transcriptome was obtained using the NCBI ORFfinder

command line tool. Nested ORFs, those completely placed within another, were ignored.

The identified ORFs were then matched to the Swiss-Prot database (The UniProt

Consortium, 2021) to focus on known annotated proteins. Only ORFs significantly

matched to the Swiss-Prot database were retained (Blastx e-value < 1e-6) to focus on

finding editing sites within known annotated proteins. Alignment of pooled gDNA reads

and mRNA reads to consensus transcriptome was performed using Bowtie2 with local alignment setting. Paired-end gDNA reads were treated as single-end when mapping to the consensus transcriptome. Only reads that were primary alignment, meaning those with the highest mapping quality score, were used. Table 3 shows the number of primary mRNA-ORF alignments for each octopus individual and the number of pooled gDNA-ORF alignments.

Editing Sites Detection

Editing sites were identified by finding weak editing sites and strong editing sites separately. Weak editing sites are loci where portions of the mRNA reads do not match the consensus transcriptome, and because the transcriptome is a consensus sequence of the aligned mRNA reads, these sites are positions at which less than 50% of the mRNA reads are modified. Strong editing sites are loci where all the gDNA reads do not match the transcriptome. Since the transcriptome is the consensus sequence of the aligned mRNA reads, this means more than 50% of the mRNA reads were modified. For identifying weak editing sites, all loci at which all gDNA reads were not uniform were not considered to focus on only homozygous sites and reduce the probability of mistaking heterozygous loci or SNPs as editing events. At loci where some of the mRNA reads did not match the transcriptome, a binomial test was used to test the probability that any mismatches were more common than could likely be randomly generated by sequencing errors. The binomial test used the number of mismatches at a given position, the total number of mRNA reads aligned to that position and the sequencing error probability. The expected sequencing error probability was estimated to be 0.1% because only nucleotides with a quality score greater than 30 (sequencing quality score = -10log(sequencing error

probability) were used (Ewing & Green, 1998). 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 less than 0.05 were considered significant, and that the discrepancy between the mRNA reads and the transcriptome was unlikely to be a result of sequencing errors, and therefore more likely a result of RNA editing. Strong editing sites were loci where all gDNA reads showed a different base than the consensus transcriptome, and by definition, a majority of the mRNA reads. However, all sequenced gDNA reads may differ from the consensus transcriptome if the octopus is heterozygotic at that base, and only one chromosome is sampled in the gDNA sequencing. The probability of mistaking genetic polymorphism as a strong editing site is $\frac{1}{2}^{\text{number of gDNA reads}}$ multiplied by a genetic polymorphism probability of 0.001 (Alon et al., 2015).

It is possible for the strong editing site protocol and the weak editing site protocol to find overlapping sites. The strong editing site protocol compares only the gDNA reads and the transcriptome, whereas the weak editing site protocol compares the mRNA reads and the transcriptome as well as the gDNA reads and the transcriptome. Locations where all the gDNA reads differ from the transcriptome are considered strong editing sites. The weak editing site protocol only considers locations that have uniform gDNA reads, which can be the same or different from the transcriptome, to focus on only homozygous sites. Then it looks at whether there are mismatches between the mRNA reads and the transcriptome at those locations. It is possible that a location in the transcriptome has uniform gDNA reads but they are all different from the transcriptome, making it a strong editing site as well as a weak editing site candidate. If at the same time at that location, there are mismatches between the mRNA reads and the transcriptome that are likely not a

result of sequencing errors as determined by a binomial test, then that location will also be identified by the weak editing sites protocol. For example, the strong editing sites in Figure 5 would also be identified by the weak editing sites screening workflow. However, because all weak editing sites and strong editing sites were consolidated into one dataset and overlapping editing sites were eliminated before downstream analyses (Figure 4), this minor mistake in the workflow does not impact my findings.

To test the accuracy of my editing sites screening workflow, I ran my pipeline on the mRNA reads and gDNA reads from the gill tissues and the nervous tissues of the common squid *Doryteuthis pealeii* from Alon et al. (2015) and publicly available at the NCBI Sequence Read Archive SRP044717.

Differential RNA editing between treatments was determined by comparing editing levels at editing sites between the control pCO_2 group to the elevated pCO_2 group. A permutation t-test comparing editing proportions between treatments at each editing site was used to determine whether the editing levels of octopuses in elevated pCO_2 treatment are significantly different from those in control pCO_2 treatment using a custom R script Significant_edits.Rmd (Sereewit & Onthank, 2021). Benjamani-Hochberg p-value correction for multiple comparisons was used to adjust p-values.

Global Editing Patterns Analyses

Two-tailed Wilcoxon rank sum tests were used to determine if there is any bias between the editing levels of putative edits that result in recoding events (i.e. nonsynonymous edits) versus those that do not (i.e. synonymous edits) in the gill tissues of *Octopus rubescens* subjected to elevated pCO_2 and control pCO_2 treatments. Additionally, Two-tailed Mann-Whitney tests were used to see if such bias occurs in the

gill tissues of *Octopus rubescens*, and the gill tissues and nervous tissues of *Doryteuthis pealeii* overall. The synonymous and nonsynonymous editing levels in *Octopus rubescens* used in the two-tailed Mann-Whitney test were the average of the editing levels in all *Octopus rubescens* individuals in treatments.

Since A-to-I RNA editing sites exhibit preferences in upstream and downstream nucleotides (Picardi et al., 2015; Yablonovitch et al., 2017), I also analyzed the nucleotides composition one nucleotide upstream and one nucleotide downstream of the putative editing sites in the gill tissues of *Octopus rubescens*.

RNA Editing Verification

After tentative editing sites and differential edits were identified using the gDNA and mRNA reads from octopuses in the 2016 experiment, a few A-to-I differential edits were selected for verification. Selection of sites to be verified favored putative editing sites with low adjusted p values from the permutation t-test, high reads coverage, greater differences between the average of editing levels of elevated pCO_2 and control pCO_2 treatment groups, and annotated functions involved in acid-base ion regulation and stress regulation. To verify the selected editing sites, Protoscript II First Strand Synthesis Kit was used to generate the cDNA library of the RNA of each octopus from the 2020 verification experiment. Using those cDNA as templates, a section of the cDNA containing the editing site(s) was amplified using the NEB Q5 High-Fidelity PCR Kit. PCR products were visualized using 1% agarose gel electrophoresis. If successfully amplified, PCR amplicons were gel extracted using the Thermo Fisher Scientific GeneJET Gel Extraction Kit.

RNA editing was verified using Sanger sequencing and poisoned primer extension (PPE). PPE is a polymerase extension technique that can distinguish edited/unedited transcripts by using the appropriate extension terminators such as dideoxynucleotides (ddNTP) or acyclonucleotides (acyNTP) that are complementary to the putative edited base in the assay template (Roberson & Rosenthal, 2006). Using the PCR amplicons as templates, if a transcript is edited, the polymerase terminates at the edited base when the chain terminator complementary to the edited base is incorporated. Otherwise, the polymerase terminates at the next base that is the same as the edited base. Therefore when a transcript is partially edited, the PPE products contain fragments of two different lengths and their amounts correspond to the editing level. When a transcript is 100% edited or not edited at all, the PPE products contain fragments of only one length. An example of PPE products using unedited transcripts versus edited transcripts as templates is shown in Figure 6. Our PPE assay used the NEB Vent(exo-) polymerase and followed a modified protocol for a routine Vent (exo-) PCR (2.5 ul reaction buffer, 5 ul PPE primers, 10 ul PPE templates, 0.5 ul Vent(exo-) polymerase, 0.5 ul of each three dNTPs , 0.5 ul of one acyNTP, 0.5 ul MgSO4, and 4.5 ul nuclease-free water). PPE products were gel electrophoresed at 300V for 2.5 hours using a 15% acrylamide gel (19.35 ml DEPC H_2O, 14 ml 40% acrylamide, 3.75 ml 10x TBE, 360 ul 10% APS, 36 ul TEMED in a 37.5 ml gel mixture) in a vertical slab gel system. After electrophoresis, the gel was stained with the Thermo Fisher Scientific SYBR Safe DNA Gel Stain for 30 minutes and viewed on a 302 nm transilluminator.

PCR amplicons were sent to Lone Star Labs for Sanger sequencing. EditR (Kluesner et al., 2018) was used to visualize the chromatograms of the amplicons and

estimate any discrepancy between the amplicons and the transcriptome at the editing sites. For amplicons that yielded quality chromatograms for all octopus samples such as the methionine adenosyltransferase 2 subunit beta-like gene, editing proportion was calculated from the .ab1 chromatograms using EditR and a two-tailed Wilcoxon rank sum test was used to determine whether differential editing levels exist between the elevated pCO_2 treatment and control pCO_2 treatment groups.

Results

Transcriptome and ORFs Metrics

Because the *Octopus rubescens* consensus transcriptome was used as a reference

for mapping mRNA and gDNA reads, the quality of the transcriptome was important for

downstream analysis. To evaluate the quality of the consensus transcriptome, I compared

the transcriptome statistics to those of other cephalopods, such as *Octopus vulgaris* and

Octopus bimaculoides, and BUSCO completeness. The *Octopus rubescens* transcriptome

statistics indicating total number of mRNA transcripts found in the transcriptome was

similar to other octopus transcriptomes used for assessing RNA editing that have

previously been assembled (Table 4). However, the transcriptome statistics indicating

mRNA transcript completeness, such as contig 50 (N50), median contig length, and

average contig length were slightly lower. BUSCO (Benchmarking Universal

Single-Copy Orthologs) was used to evaluate transcriptome assembly completeness

based on the concept that single-copy orthologs should be conserved among closely

related species (Manni et al., 2021). The BUSCO evaluation found that our *Octopus*

rubescens transcriptome is 81.5% complete, 2.4% fragmented and 16.1% missing in

terms of expected gene content when compared to the mollusca_odb10 dataset (Appendix

A).

To focus on the editing sites in regions of the consensus transcriptome that

translate to known proteins and to avoid repetitive elements, only the ORFs of the

consensus transcriptome that significantly matched to the Swissprot database (Blastx

e-value < 1e-6) were used. Overall I found 38,890 ORFs that significantly matched

14,603 proteins in the Swissprot database. Those ORFs have a mean length of 1193 nt

and a median length of 807 nt. The ORFs metrics are similar to those of *Octopus vulgaris* and *Octopus bimaculoides* (Table 4).

Putative Editing Sites Characterization

Here I use 'editing' to describe something that has been verified to be a true A-to-I editing event by either PPE or Sanger sequencing. 'Mismatch' is used to describe the discrepancy between the nucleotide composition in the mRNA reads and the consensus transcriptome and that discrepancy has not been verified by either PPE or Sanger sequencing. Sites that are identified by the editing sites detection pipeline but not verified by PPE or Sanger sequencing are called 'putative editing sites' or 'mismatch sites.'

Overall, I found 19,302 putative strong editing sites and 113,111 putative weak editing sites, with 9,817 sites found in both categories. Excluding those duplicates, 122,596 putative editing sites were found. 55,646 putative editing sites have mRNA reads coverage for all six octopus samples in the 2016 experiment. Because inosine is recognized as guanosine (G) by biological processes, A-to-I editing is manifested as A-to-G mismatches. Of those 55,646 sites, 9,094 or 16.3% of them were A-to-G mismatches while the other 46,552 were non A-to-G mismatches (Figure 7). 744 out of the 55,646 sites had statistically significant differential mismatch levels between the control pCO_2 treatment and elevated pCO_2 treatment, and 136 of those 744 sites were A-to-G mismatches (false discovery rate < 0.1). The mismatch levels in nonsynonymous A-to-G mismatch sites are higher in octopuses in elevated pCO_2 treatment than in octopuses in control pCO_2 treatment (two- tailed Wilcoxon rank sum test, p<0.0001, Figure 8).

To test the accuracy of my editing sites protocols, I ran my pipeline using the mRNA reads and gDNA reads from the nervous tissues of the common squid *Doryteuthis pealeii*, which has previously been used by other studies to detect editing sites (Alon et al., 2015), and found 60.9% of mismatches in the giant fiber lobe and optic lobe of squid were A-to-G mismatches. To investigate whether a low level of A-to-G mismatch occurs in the gill tissues of the common squid as in *Octopus rubescens*, mRNA reads and gDNA reads from the same squid featured in Alon et al. (2015) were analyzed using my pipeline. Only 9,164 (16.4 %) out of 55,828 mismatch sites were A-to-G mismatches, similar to what I observed in the gill tissues of *Octopus rubescens* (Figure 7).

In the gill tissues of *Octopus rubescens* 30,886 out of 55,646 mismatch sites result in recoding events, compared to 22,668 out of 55,828 in the gill tissues of *Doryteuthis pealeii* and 205,425 out of 315,411 in the nervous tissues of *Doryteuthis pealeii*, which is substantially more than the previously reported 70,000 recoding events (Alon et al., 2015; Yablonovitch et al., 2017). In the gill tissues of *Octopus rubescens* and *Doryteuthis pealeii*, the A-to-I mismatch levels are significantly higher in synonymous sites than in nonsynonymous sites, whereas in the nervous tissues of *Doryteuthis pealeii*, the mismatch levels are significantly higher in nonsynonymous sites (Figure 9).

Editing Sites Verification

Poisoned primer extension (PPE) and Sanger sequencing were used to verify one of the differential A-to-G mismatches in the methionine adenosyltransferase 2 subunit beta-like gene, and Sanger sequencing alone was used to verify another mismatch in the thioredoxin-like protein 4A gene. Methionine adenosyltransferase 2 subunit beta-like gene's PCR amplicons electrophoresis results shows that the PPE template of 636 bp

(Appendix B) was successfully amplified for all six octopus samples (Figure 10). Figure

11 shows a gel of the PPE assay on the methionine adenosyltransferase 2 subunit

beta-like gene for all 6 octopuses. In the gel photo, there was only one band visible

besides the primer for each sample at the position expected for edited amplicons,

indicating that the methionine adenosyltransferase 2 subunit beta-like gene might be

close to 100% edited or unedited in all octopuses. These bands showing the PPE products

are quite faint because the 302 nm wavelength used by our transilluminator only excites

the SYBR safe pigment at an approximate 12.5% relative intensity compared to the

optimal excitation wavelength of 502 nm (*SYBR Safe - DNA Gel Stain - US*, n.d.).

However, bands appear to be near 18 nt, which does not match either expected locations

of edited or unedited amplicons. The bands are faintly visible in the gel image (Figure

11), but were much brighter to the naked eye. Although results of the PPE assay are

uninterpretable, Sanger sequencing results show that close to 100% of the nucleotides at

the edited site are guanosines for all six octopuses, indicating nearly 100% editing

(Figure 12, 13) (only one example chromatogram is shown), and not statistically different

between treatments (two-tailed Wilcoxon rank sum test, p=0.4795).

Gel electrophoresis shows that three out of six octopus samples had PCR

amplicons (Figure 14) using primers set targeting the thioredoxin-like protein 4A gene

(Appendix B). Of the three samples that were sent for Sanger sequencing, only one had a

chromatogram where there is baseline separation between peaks (Figure 15). Because the

reverse primer was used for this Sanger sequencing, the sequencing results were the

reverse complement of the PCR amplicons and A-to-I edits are shown as T-to-C

mismatches. The chromatogram shows a 24% editing level in the one octopus that had its

thioredoxin-like protein 4A gene successfully amplified and sequenced. The editing

levels in thioredoxin-like protein 4A gene determined from RNA-seq data versus from

Sanger sequencing are shown in Figure 16.

Discussion

This research project is the first to investigate the effects of ocean acidification on RNA editing in a marine organism. Overall 55,646 mismatch sites that have mRNA-seq coverage for all six *Octopus rubescens* samples were found, of which 16.3% of them are A-to-G mismatches (Figure 7). More importantly, there are 136 sites that have differential A-to-G mismatch levels between the control pCO_2 treatment and elevated pCO_2 treatment groups. Because previous work has shown that the vast majority of RNA edits are A-to-I edits, and the ratio of G-to-A mismatches has been used to estimate the noise level (Liscovitch-Brauer et al., 2017; Xu & Zhang, 2014), the low ratio of A-to-G mismatches to other mismatch types I found signifies a relatively low level of RNA editing in the gill tissues of *Octopus rubescens* compared to the nervous tissues of other coleoid cephalopods (Figure 7). The A-to-G mismatch ratios are at least 60% in the nervous system tissues of *Doryteuthis pealeii, Sepia, Octopus vulgaris, and Octopus bimaculoides*. Interestingly, the A-to-G mismatch and G-to-A mismatch ratios found in the gill tissues of *Doryteuthis pealeii* resembles those in the gill tissues of *Octopus rubescens* (Figure 7). This suggests RNA editing might not be as common in the gill tissues in cephalopods in general compared to in the nervous tissues. In addition, the synonymous A-to-G mismatch levels are higher than the nonsynonymous A-to-G mismatch levels in the gill tissues of *Octopus rubescens* and *Doryteuthis pealeii*, whereas the opposite is observed in the nervous tissues of *Doryteuthis pealeii* (Figure 9). Previous studies have used lower mismatch levels in nonsynonymous mismatches than synonymous mismatches in humans as evidence that A to I editing is generally nonadaptive in humans (Xu & Zhang, 2014). From an evolutionary perspective, my

findings might indicate that the protein products of the recoding A-to-G mismatches in the gill tissues of *Octopus rubescens* and *Doryteuthis pealeii* are not as useful in aiding environmental adaptation and therefore not favored by natural selection compared to recoding events in the nervous tissues of cephalopods. However, I only analyzed the gill tissues of two cephalopod species and further analyses on the gill tissues of other species are needed to reach these conclusions.

The non-A-to-G mismatches in the gill tissues of *Octopus rubescens* are likely results of a combination of transcriptome assembly problems, undetected single nucleotide polymorphism, somatic mutations, and systemic misalignment (Alon et al., 2015). I attempted to address the misalignment issue by applying a mapping quality filter to the mRNA and pooled gDNA reads aligned to the transcriptome. The mapping quality score (MAPQ) generated by the Bowtie2 aligner is calculated by $-10\log_{10}$(an estimate of the probability that the alignment does not correspond to the read's true point of origin) (Langmead & Salzberg, 2012). A MAPQ of 10 indicates that probability of 10%, a MAPQ of 20 indicates 1%, and a MAPQ of 30 indicates 0.1%. When those mapping quality filters were applied, a majority of the reads were eliminated (Table 5). Because my mismatch sites detection methods require both gDNA and mRNA coverage, a major loss in reads resulted in a significant reduction in the number of mismatch sites that have both gDNA and mRNA coverage. In addition, because I intentionally looked for sites where there are mismatches between the transcriptome and mRNA or gDNA, and mismatches inherently lower the MAPQ score, applying MAPQ filters would naturally lower the number of mismatch sites. Therefore, my mismatch sites detection pipeline

does not include applying mapping quality filters and is consistent with the established

procedures by Alon et al. (2015).

The level of false positive mismatches because of the reasons mentioned above is

a challenge to finding true A-to-I editing events. However, the low ratio of A-to-G

mismatches is low in the gill tissues of *Octopus rubescens* does not indicate all the

A-to-G mismatches were not actual A-to-I edits. One of the signatures of A-to-I RNA

editing facilitated by ADAR is patterns of G depletion one nucleotide upstream of the

editing sites and G enrichment one nucleotide downstream of the editing sites (Picardi et

al., 2015; Yablonovitch et al., 2017). If the gDNA/mRNA mismatches I found were all

due to misalignment, one would not expect ADAR's upstream and downstream bias in

the data. Instead, one would expect the upstream and downstream nucleotides

compositions to closely resemble the overall nucleotide composition in the ORFs.

However, in my findings G is under-represented in the upstream nucleotide and

over-represented in the downstream nucleotide. That bias is greater in the 9,094 A-to-G

mismatches than the complete set of 55,646 mismatches, and greater still in the 136

differential A-to-G mismatches (Figure 17, 18). This strongly suggests that a substantial

portion of the differential A-to-G mismatches are likely true A-to-I RNA edits because

the upstream and downstream patterns are consistent with ADAR editing biases.

In fact, two genes containing differential A-to-G mismatches--methionine

adenosyltransferase 2 subunit beta-like gene and thioredoxin-like protein 4A gene--were

verified to be A-to-I RNA edits by Sanger sequencing. Methionine adenosyltransferase 2

subunit beta-like gene encodes the regulatory beta subunit of methionine

adenosyltransferase. Methionine adenosyltransferase 2 catalyzes the formation of

S-adenosylmethionine which synthesizes polyamines that have been shown to have neuroprotection function and is also critical in melatonin synthesis (Hunsberger et al., 2005). Thioredoxin are small proteins that protect cells and tissues against oxidative stress (Das & Das, 2000). Although these verifications show actual A-to-I RNA editing in the gill tissues of *Octopus rubescens*, they failed to confirm that differential edits occur in *Octopus rubescens* when they are subjected to elevated pCO_2 treatments (Figure 12, 16). The methionine adenosyltransferase 2 subunit beta-like gene was close to 100% edited in all octopuses in the verification experiment from 2020 compared to the differential editing levels seen in the RNA-seq data from the 2016 experiment (Figure 12). The reasons could be experiment replication issues or false positives produced by permutation t-test when screening for differential edits. The thioredoxin-like protein 4A gene was verified to be edited, but only in one octopus (Figure 15, 16). In other octopuses, either the gene wasn't successfully amplified for Sanger sequencing or when it did the quality of the Sanger sequencing chromatogram was inadequate for interpretation. From my experience of using Sanger sequencing and PPE assays to verify RNA editing, Sanger sequencing has shown to be a better approach in terms of time and labor costs as well as results interpretability.

My research shows that RNA editing does occur in the gill tissues of *Octopus rubescens*, albeit not as commonly compared to the nervous tissues of other coleoid cephalopods such as *Doryteuthis pealeii, Sepia, Octopus vulgaris, and Octopus bimaculoides*. I have also shown that overall editing levels in the gill transcriptome rise in elevated pCO_2 environments, consistent with previous work that has shown increased editing rates under intracellular acidification (Malik et al., 2021). Two differential A-to-G

mismatches in the methionine adenosyltransferase 2 subunit beta-like gene and the

thioredoxin-like protein 4A gene were confirmed to be actual A-to-I RNA edits using

Sanger sequencing; However, they were not verified to be differentially edited between

the elevated pCO_2 and control pCO_2 treatments. Of the 136 editing sites that are

differentially expressed between high and low CO_2 treatments, some could still be critical

to how *Octopus rubescens* responds to environmental acidification, and further research

is warranted. Using A-to-I editing in humans as an example, even though the number of

editing sites in humans has shown to be far fewer than in cephalopods and the majority of

the editing sites are non-recoding and non-adaptive, reduced editing in a couple specific

targets are reported to be associated with neuronal and CNS disorders, including

Alzhimer's disease and amyotrophic lateral sclerosis (Walkley & Li, 2017). To take this

research further, one of the directions is to confirm the differentially A-to-G mismatches

and see whether they play an important physiological role in how *Octopus rubescens* deal

with ocean acidification.

Acknowledgements

References

Alon, S., Garrett, S. C., Levanon, E. Y., Olson, S., Graveley, B. R., Rosenthal, J. J. C., &
Eisenberg, E. (2015). The majority of transcripts in the squid nervous system are
extensively recoded by A-to-I RNA editing. *ELife*, *4*.
https://doi.org/10.7554/eLife.05198

Barford, E. (2013). Rising ocean acidity will exacerbate global warming. *Nature*.
https://doi.org/10.1038/nature.2013.13602

Birk, M. A., McLean, E. L., & Seibel, B. A. (2018). Ocean acidification does not limit
squid metabolism via blood oxygen supply. *Journal of Experimental Biology*,
221(19), jeb187443. https://doi.org/10.1242/jeb.187443

Brennicke, A., Marchfelder, A., & Binder, S. (1999). RNA editing. *FEMS Microbiology
Reviews*, *23*(3), 297–316. https://doi.org/10.1111/j.1574-6976.1999.tb00401.x

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ
preprocessor. *Bioinformatics*, *34*(17), 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. *Marine Biology*,
168(9), 137. https://doi.org/10.1007/s00227-021-03913-z

Das, K. C., & Das, C. K. (2000). Thioredoxin, a Singlet Oxygen Quencher and Hydroxyl
Radical Scavenger: Redox Independent Functions. *Biochemical and Biophysical
Research Communications*, *277*(2), 443–447.
https://doi.org/10.1006/bbrc.2000.3689

Dickson, A. G., Sabine, C. L., Christian, J. R., Bargeron, C. P., & North Pacific Marine

Science Organization (Eds.). (2007). *Guide to best practices for ocean CO2 measurements*. North Pacific Marine Science Organization.

Doney, S. C., Busch, D. S., Cooley, S. R., & Kroeker, K. J. (2020). The Impacts of Ocean Acidification on Marine Ecosystems and Reliant Human Communities. *Annual Review of Environment and Resources*, *45*(1), 83–112. https://doi.org/10.1146/annurev-environ-012320-083019

Eisenberg, E., & Levanon, E. Y. (2018). A-to-I RNA editing—Immune protector and transcriptome diversifier. *Nature Reviews Genetics*, *19*(8), 473–490. https://doi.org/10.1038/s41576-018-0006-1

Ewing, B., & Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Research*, *8*(3), 186–194. https://doi.org/10.1101/gr.8.3.186

Garrett, S., & Rosenthal, J. J. C. (2012). RNA editing underlies temperature adaptation in K+ channels from polar octopuses. *Science (New York, N.Y.)*, *335*(6070), 848–851. https://doi.org/10.1126/science.1212795

Gattuso, J.-P., Epitalon, J.-M., Lavigne, H., Orr, J., Gentili, B., Hagens, M., Hofmann, A., Mueller, J.-D., Proye, A., Rae, J., & Soetaert, K. (2021). *seacarb: Seawater Carbonate Chemistry* (3.3.0). https://CRAN.R-project.org/package=seacarb

Gingerich, P. D. (2019). Temporal Scaling of Carbon Emission and Accumulation Rates: Modern Anthropogenic Emissions Compared to Estimates of PETM Onset Accumulation. *Paleoceanography and Paleoclimatology*, *34*(3), 329–335. https://doi.org/10.1029/2018PA003379

Glöckner, F. O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G.,

Yarza, P., Peplies, J., Westram, R., & Ludwig, W. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and tools. *Journal of Biotechnology*, *261*, 169–176. https://doi.org/10.1016/j.jbiotec.2017.06.1198

Gott, J., & Emeson, R. (2000). Functions and Mechanisms of RNA Editing. *Annual Review of Genetics*, *34*, 499–531. https://doi.org/10.1146/annurev.genet.34.1.499

Gruber, N., Clement, D., Carter, B. R., Feely, R. A., van Heuven, S., Hoppema, M., Ishii, M., Key, R. M., Kozyr, A., Lauvset, S. K., Lo Monaco, C., Mathis, J. T., Murata, A., Olsen, A., Perez, F. F., Sabine, C. L., Tanhua, T., & Wanninkhof, R. (2019). The oceanic sink for anthropogenic CO2 from 1994 to 2007. *Science*, *363*(6432), 1193–1199. https://doi.org/10.1126/science.aau5153

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. *Journal of Comparative Physiology B*, *180*(3), 323–335. https://doi.org/10.1007/s00360-009-0412-y

Higuchi, M., Maas, S., Single, F., Hartner, J., Rozov, A., Burnashev, N., Feldmeyer, D., Sprengel, R., & Seeburg, P. H. (2000, July 6). *Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2 | Nature*. https://www.nature.com/articles/35017558/

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*(4), e1064196. https://doi.org/10.1080/21688370.2015.1064196

Hunsberger, J. G., Bennett, A. H., Selvanayagam, E., Duman, R. S., & Newton, S. S. (2005). Gene profiling the response to kainic acid induced seizures. *Molecular*

Brain Research, *141*(1), 95–112.

https://doi.org/10.1016/j.molbrainres.2005.08.005

Kaplan, M. B., Mooney, T. A., McCorkle, D. C., & Cohen, A. L. (2013). Adverse Effects of Ocean Acidification on Early Development of Squid (Doryteuthis pealeii). *PLOS ONE*, *8*(5), e63714. https://doi.org/10.1371/journal.pone.0063714

Kim, D. D. Y., Kim, T. T. Y., Walsh, T., Kobayashi, Y., Matise, T. C., Buyske, S., & Gabriel, A. (2004). Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Research*, *14*(9), 1719–1725. https://doi.org/10.1101/gr.2855504

Kluesner, M. G., Nedveck, D. A., Lahr, W. S., Garbe, J. R., Abrahante, J. E., Webber, B. R., & Moriarity, B. S. (2018). EditR: A Method to Quantify Base Editing from Sanger Sequencing. *The CRISPR Journal*, *1*, 239–250. https://doi.org/10.1089/crispr.2018.0014

Lacoue-Labarthe, T., Réveillac, E., Oberhänsli, F., Teyssié, J. L., Jeffree, R., & Gattuso, J. P. (2011). Effects of ocean acidification on trace element accumulation in the early-life stages of squid Loligo vulgaris. *Aquatic Toxicology*, *105*(1), 166–176. https://doi.org/10.1016/j.aquatox.2011.05.021

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 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. (2017). Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods. *Cell*, *169*(2), 191-202.e11. https://doi.org/10.1016/j.cell.2017.03.025

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 Research*, *49*(7), 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. *Molecular Biology and Evolution*, *38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

Munday, P. L., Hernaman, V., Dixson, D. L., & Thorrold, S. R. (2011). Effect of ocean acidification on otolith development in larvae of a tropical marine fish. *Biogeosciences*, *8*(6), 1631–1641. https://doi.org/10.5194/bg-8-1631-2011

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. *Physiological and Biochemical Zoology*, *94*(1), 1–11. https://doi.org/10.1086/712207

Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G.-K., Rodgers, K. B., … Yool, A. (2005). Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature, 437*(7059), Article 7059. https://doi.org/10.1038/nature04095

Palladino, M. J., Keegan, L. P., O'Connell, M. A., & Reenan, R. A. (2000, August 18). *A-to-I Pre-mRNA Editing in Drosophila Is Primarily Involved in Adult Nervous*

System Function and Integrity: Cell.

https://www.cell.com/fulltext/S0092-8674(00)00049-0

Paz-Yaacov, N., Levanon, E. Y., Nevo, E., Kinar, Y., Harmelin, A., Jacob-Hirsch, J., Amariglio, N., Eisenberg, E., & Rechavi, G. (2010). Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(27), 12174–12179. https://doi.org/10.1073/pnas.1006183107

Picardi, E., Manzari, C., Mastropasqua, F., Aiello, I., D'Erchia, A. M., & Pesole, G. (2015). Profiling RNA editing in human tissues: Towards the inosinome Atlas. *Scientific Reports*, *5*(1), Article 1. https://doi.org/10.1038/srep14941

Quinones-Valdez, G., Tran, S. S., Jun, H.-I., Bahn, J. H., Yang, E.-W., Zhan, L., Brümmer, A., Wei, X., Van Nostrand, E. L., Pratt, G. A., Yeo, G. W., Graveley, B. R., & Xiao, X. (2019). Regulation of RNA editing by RNA-binding proteins in human cells. *Communications Biology*, *2*(1), Article 1. https://doi.org/10.1038/s42003-018-0271-8

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 Biology*, *13*, 1. https://doi.org/10.1186/s12915-014-0111-3

Roberson, L. M., & Rosenthal, J. J. C. (2006). An accurate fluorescent assay for quantifying the extent of RNA editing. *RNA*, *12*(10), 1907–1912. https://doi.org/10.1261/rna.166906

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). *Journal of Experimental Biology*, *217*(4), 518–525. https://doi.org/10.1242/jeb.096081

Rueter, S. M., Dawson, T. R., & Emeson, R. B. (1999). Regulation of alternative splicing by RNA editing. *Nature*, *399*(6731), Article 6731. https://doi.org/10.1038/19992

Sapiro, A. L., Freund, E. C., Restrepo, L., Qiao, H.-H., Bhate, A., Li, Q., Ni, J.-Q., Mosca, T. J., & Li, J. B. (2020). Zinc Finger RNA-Binding Protein Zn72D Regulates ADAR-Mediated RNA Editing in Neurons. *Cell Reports*, *31*(7), 107654. https://doi.org/10.1016/j.celrep.2020.107654

Sereewit, J., & Onthank, K. (2021). *RNA Editing Detection Pipepline* (0.1.0) [Python]. https://doi.org/10.5281/zenodo.5655824

Slomovic, S., Laufer, D., Geiger, D., & Schuster, G. (2006). Polyadenylation of ribosomal RNA in human cells. *Nucleic Acids Research*, *34*(10), 2966–2975. https://doi.org/10.1093/nar/gkl357

Sommer, B., Köhler, M., Sprengel, R., & Seeburg, P. H. (1991). RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell*, *67*(1), 11–19. https://doi.org/10.1016/0092-8674(91)90568-j

Song, L., & Florea, L. (2015). Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience*, *4*(1), 48. https://doi.org/10.1186/s13742-015-0089-y

SYBR Safe—DNA Gel Stain—US. (n.d.). Thermo Fisher Scientific. Retrieved November 7, 2021, from //www.thermofisher.com/us/en/home/life-science/dna-rna-purification-analysis/nu

cleic-acid-gel-electrophoresis/dna-stains/sybr-safe.html

Tan, M. H., Li, Q., Shanmugam, R., Piskol, R., Kohler, J., Young, A. N., Liu, K. I., Zhang, R., Ramaswami, G., Ariyoshi, K., Gupte, A., Keegan, L. P., George, C. X., Ramu, A., Huang, N., Pollina, E. A., Leeman, D. S., Rustighi, A., Goh, Y. P. S., … Li, J. B. (2017). Dynamic landscape and regulation of RNA editing in mammals. *Nature*, *550*(7675), Article 7675. https://doi.org/10.1038/nature24041

The UniProt Consortium. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. https://doi.org/10.1093/nar/gkaa1100

US Department of Commerce, N. (n.d.). *Global Monitoring Laboratory—Carbon Cycle Greenhouse Gases*. Retrieved May 4, 2022, from https://gml.noaa.gov/ccgg/trends/

Vesely, C., & Jantsch, M. F. (2021). An I for an A: Dynamic Regulation of Adenosine Deamination-Mediated RNA Editing. *Genes*, *12*(7), 1026. https://doi.org/10.3390/genes12071026

Walkley, C. R., & Li, J. B. (2017). Rewriting the transcriptome: Adenosine-to-inosine RNA editing by ADARs. *Genome Biology*, *18*(1), 205. https://doi.org/10.1186/s13059-017-1347-3

Wang, Q., Zhang, Z., Blackwell, K., & Carmichael, G. G. (2005). Vigilins bind to promiscuously A-to-I-edited RNAs and are involved in the formation of heterochromatin. *Current Biology: CB*, *15*(4), 384–391. https://doi.org/10.1016/j.cub.2005.01.046

Xu, G., & Zhang, J. (2014). Human coding RNA editing is generally nonadaptive.

Proceedings of the National Academy of Sciences of the United States of America, *111*(10), 3769–3774. https://doi.org/10.1073/pnas.1321745111

Yablonovitch, A. L., Deng, P., Jacobson, D., & Li, J. B. (2017). The evolution and adaptation of A-to-I RNA editing. *PLoS Genetics*, *13*(11), e1007064. https://doi.org/10.1371/journal.pgen.1007064

Zeebe, R. E., Zachos, J. C., Caldeira, K., & Tyrrell, T. (2008). *Carbon Emissions and Acidification*. 2.
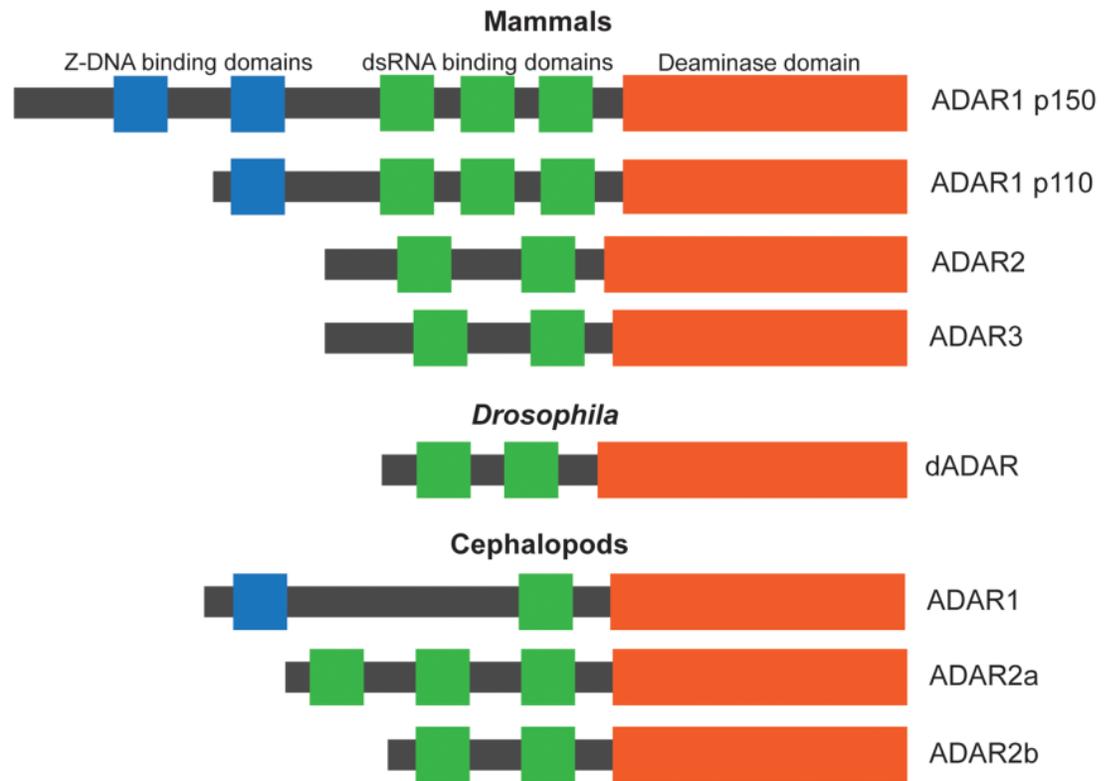
Figures

Figure 1. Diagrams from Yablonovitch et at. (2017) depicting ADAR isoforms in mammals, *Drosophila*, and cephalopods.
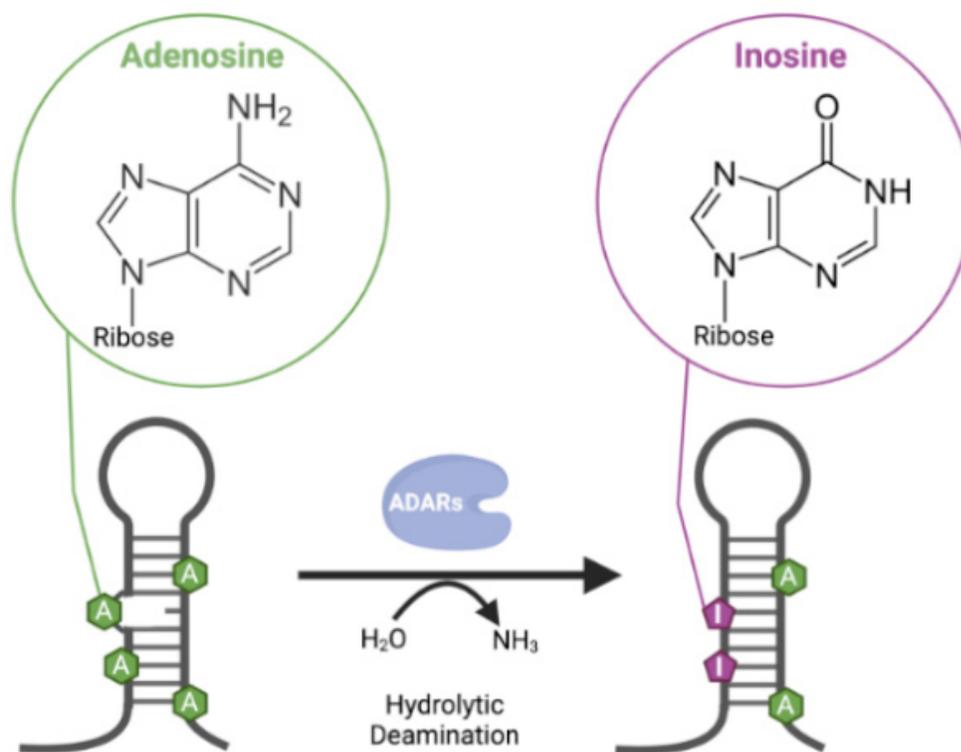
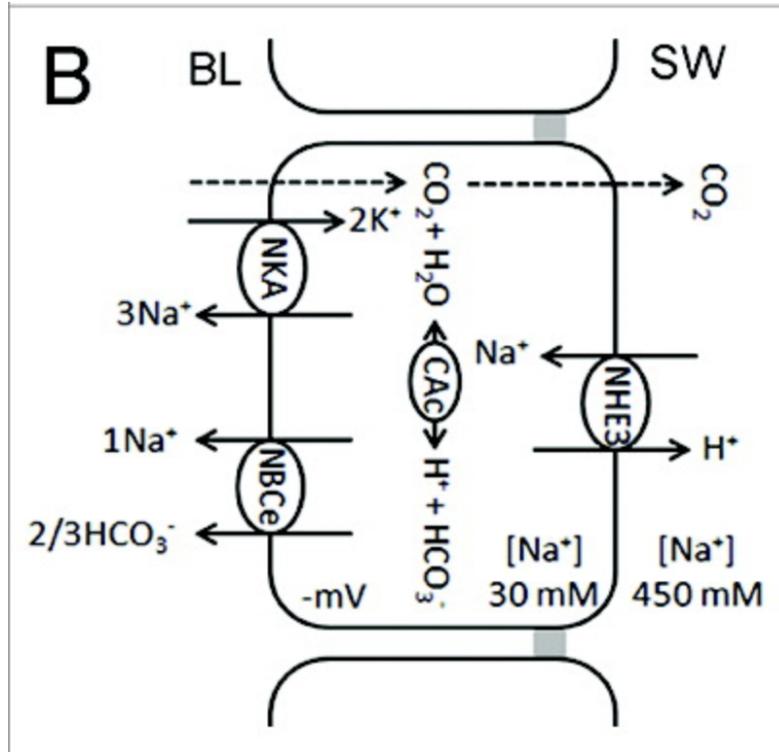**Figure 2.** Diagram from Malik et al. (2021) showing hydrolytic deamination of adenosine to inosine.

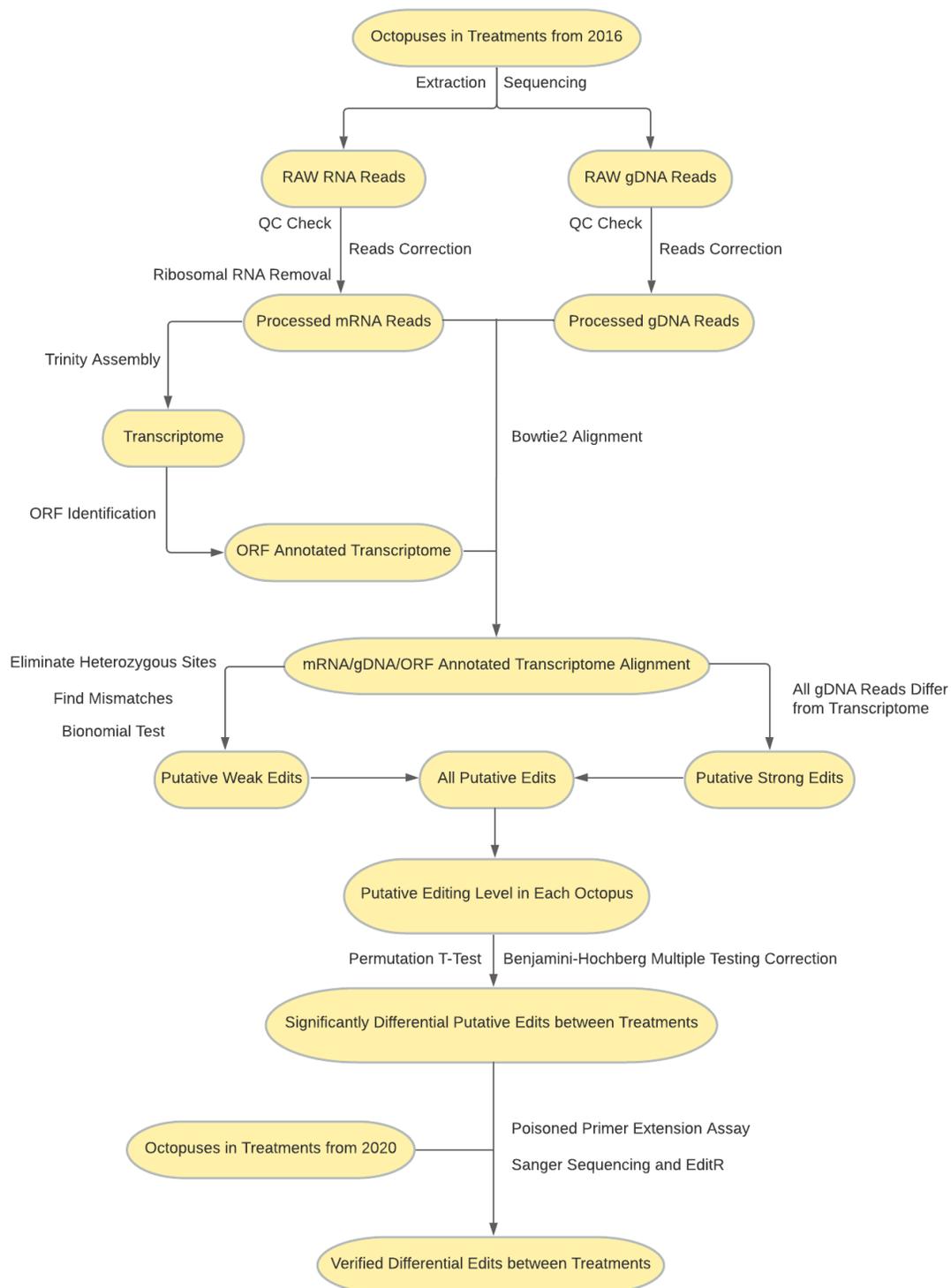**Figure 3.** Hypothetical acid-base regulation mechanism in cephalopod gills by Hu et al.

(2015).

Figure 4. Editing sites and differential edits detection workflow.

'strong' editing site

↓

```
ORF ..TTGATCGGATCAA....ACAATATGTGCACGGAGT..ORF
       TGATCGAATCAA        ACAATATGTGCACGGAGT
       TGATCGAAT           CAACATGTGCACGG
gDNA   TGATCGAATC          ACAATATGTGCACGGAGT   gDNA
      TTGATCGAATCA          AACATGTGCACGGAG
       TGATCGGATCA         ACAATATGTGCGCGGAGT
       TGATCGGATCAA        CAACATGTGCGCGGAGT
      GATCGAATCA           AATATGTGCACGGAGT
mRNA   TGATCGGATCAA        ACAACATGTGCACGGAGT   mRNA
      GATCGAATCA           ACAATATGTGCACGGAGT
ORF ..TTGATCGGATCAA....ACAATATGTGCACGGAGT..ORF
```

↑   ↑
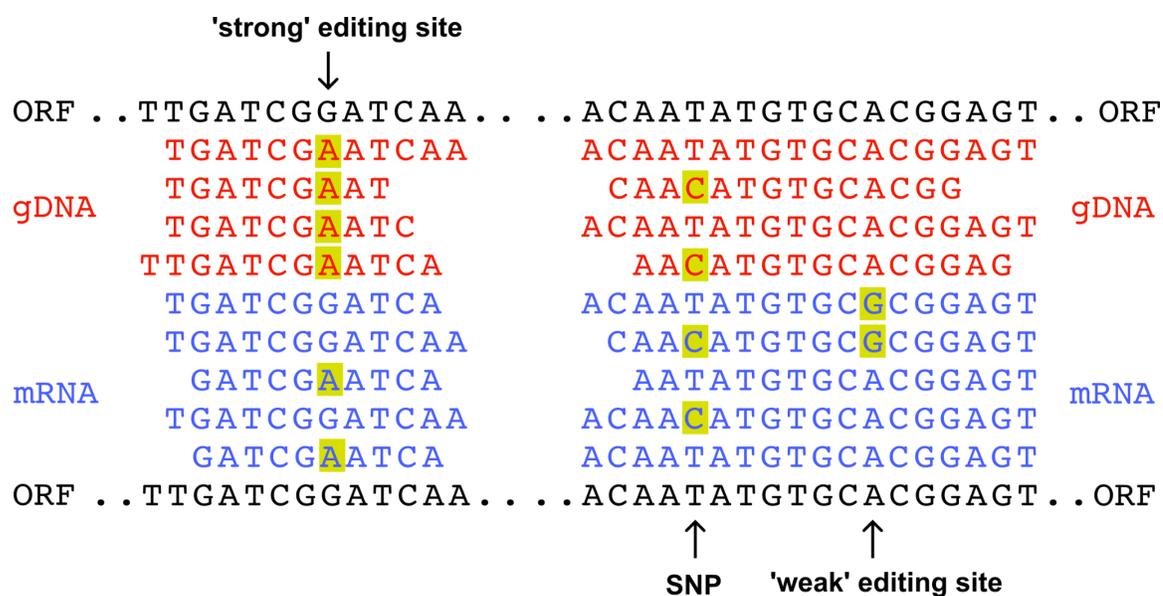
SNP   **'weak' editing site**

Figure 5. Editing sites are detected by comparing mRNA reads and gDNA reads to ORF.

Strong editing sites are sites where all gDNA reads differ from ORF. Weak editing sites

are loci where the minority of the RNA reads differ from ORF. Only homozygous sites,

sites where all gDNA reads are the same, are considered for weak editing sites screening

to avoid mistaking heterozygosity as editing events. Diagram adapted from Alon et al.,

2015.

```
Unedited transcript:    ACTGGACATTAGATCCATAGTTGACTTACTTGCGCTGGCTAC
Unedited PPE product:   TGACCTGTAATCTAGGTATCAACTGAATGAAC


Edited transcript:      ACTGGACATTAGATCCATAGTTGACTTGCTTGCGCTGGCTAC
Edited PPE product:     TGACCTGTAATCTAGGTATCAACTGAAC
```

red: editing site
blue: primer
green: ddNTP
orange: dNTP

Figure 6. Poisoned primer extension (PPE) products using unedited and edited

transcripts. ddNTP is the PPE terminator. ddNTP is chosen to be the complement of the

edited base (ddCTP in this illustration, because guanine (G) and cytocine (C) are

complement to each other) so PPE terminates at the edited base in edited transcripts or

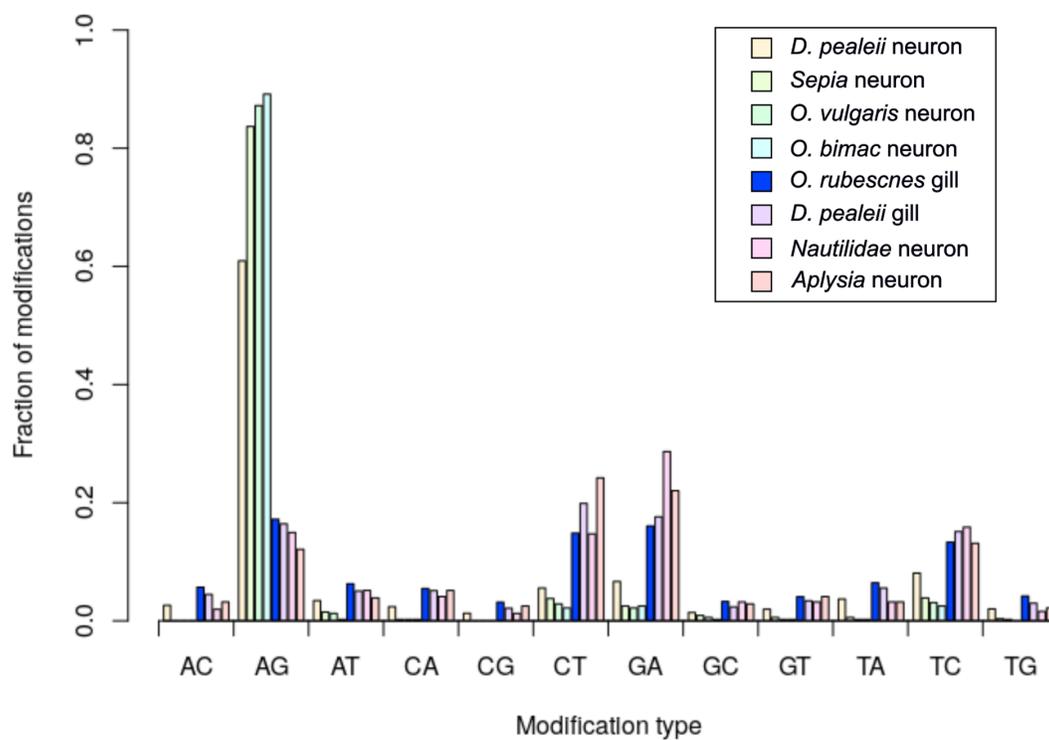the fist that is the same as the edited base in unedited transcripts.

Figure 7. Fractions of mismatch types in the nervous tissues and gill tissues of different cephalopod species. Squid neuron, squid gill and *Octopus rubescens* gill all analyzed with my pipeline. All others from Liscovitch-Brauer et al. (2017).
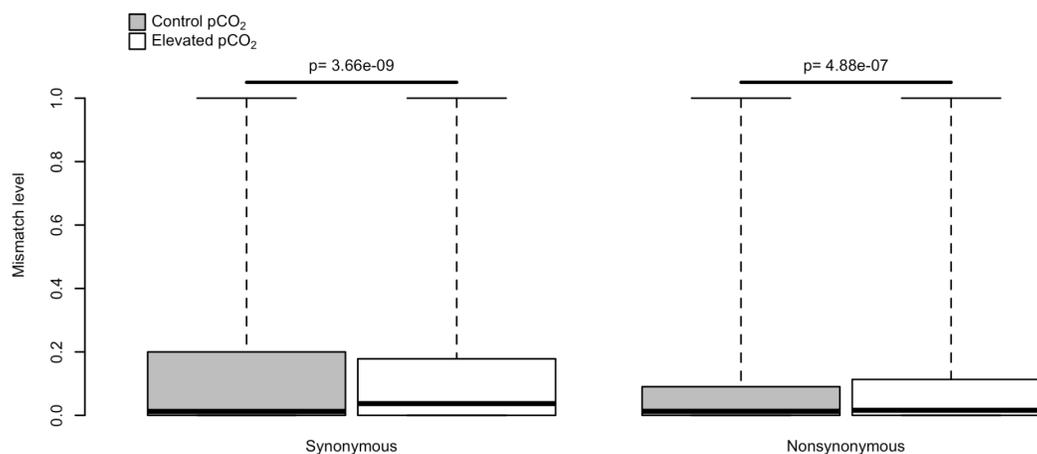
Figure 8. Distributions of synonymous and nonsynonymous A-to-G mismatch levels of *Octopus rubescens* in elevated pCO_2 (n=3) and control pCO_2 treatments (n=3). Mismatch level is significantly different between elevated and control CO_2 treatments for both synonymous (two-tailed Wilcoxon rank sum test, p=3.66×10^{-9}) and nonsynonymous changes (two-tailed Wilcoxon rank sum test, p=4.88×10^{-7}).
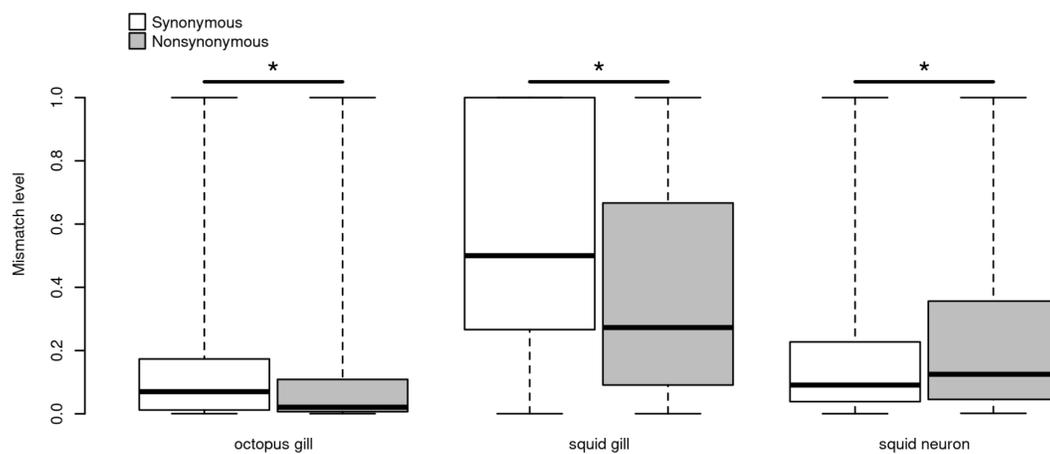
Figure 9. Synonymous and nonsynonymous A-to-G mismatch levels in the gill tissues of *Octopus rubescens* (two-tailed Mann-Whitney, $p<1x10^{-15}$) and in *Doryteuthis pealeii*'s gill (two-tailed Mann-Whitney, $p<1x10^{-15}$) and nervous tissues (two-tailed Mann-Whitney test, $p<1x10^{-15}$). The A-to-G mismatch levels in *Octopus rubescens* are the average of all octopuses (n=6) in both elevated pCO_2 and control pCO_2 treatments.
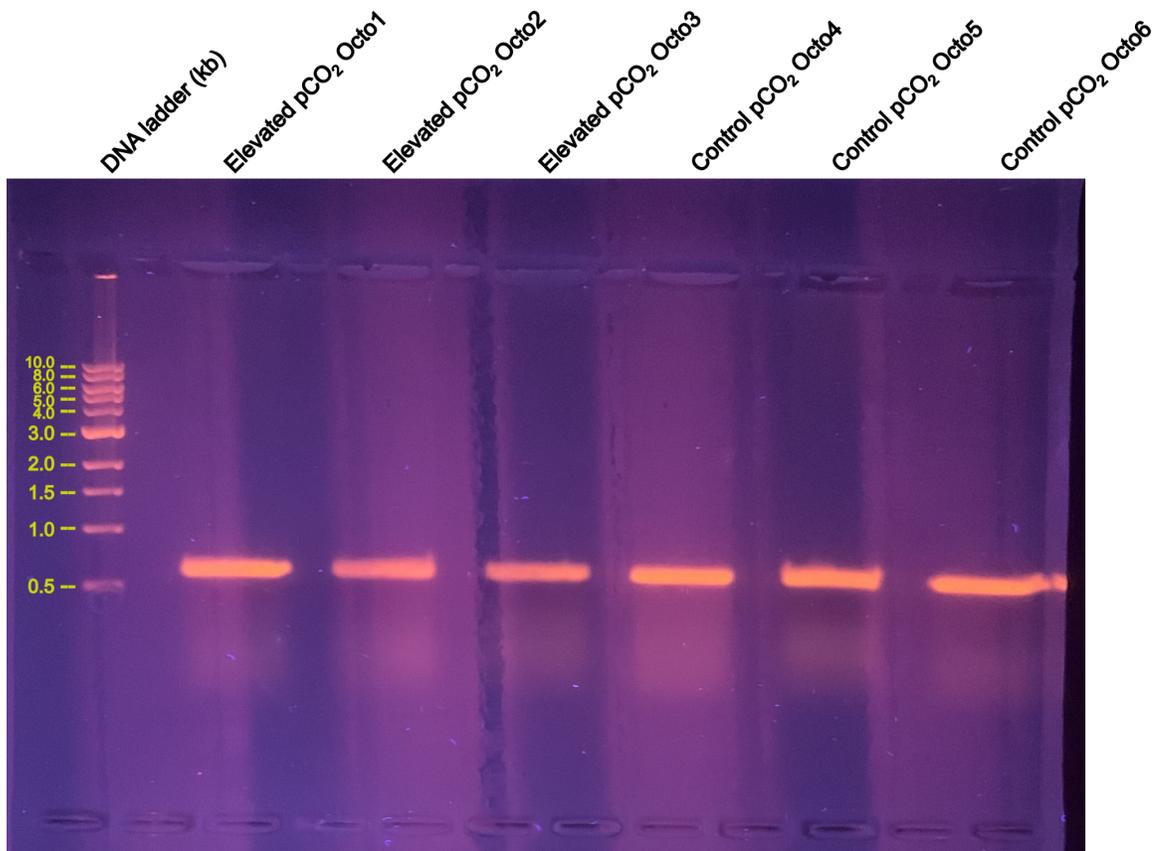
Figure 10. PCR amplicons of the methionine adenosyltransferase 2 subunit beta-like

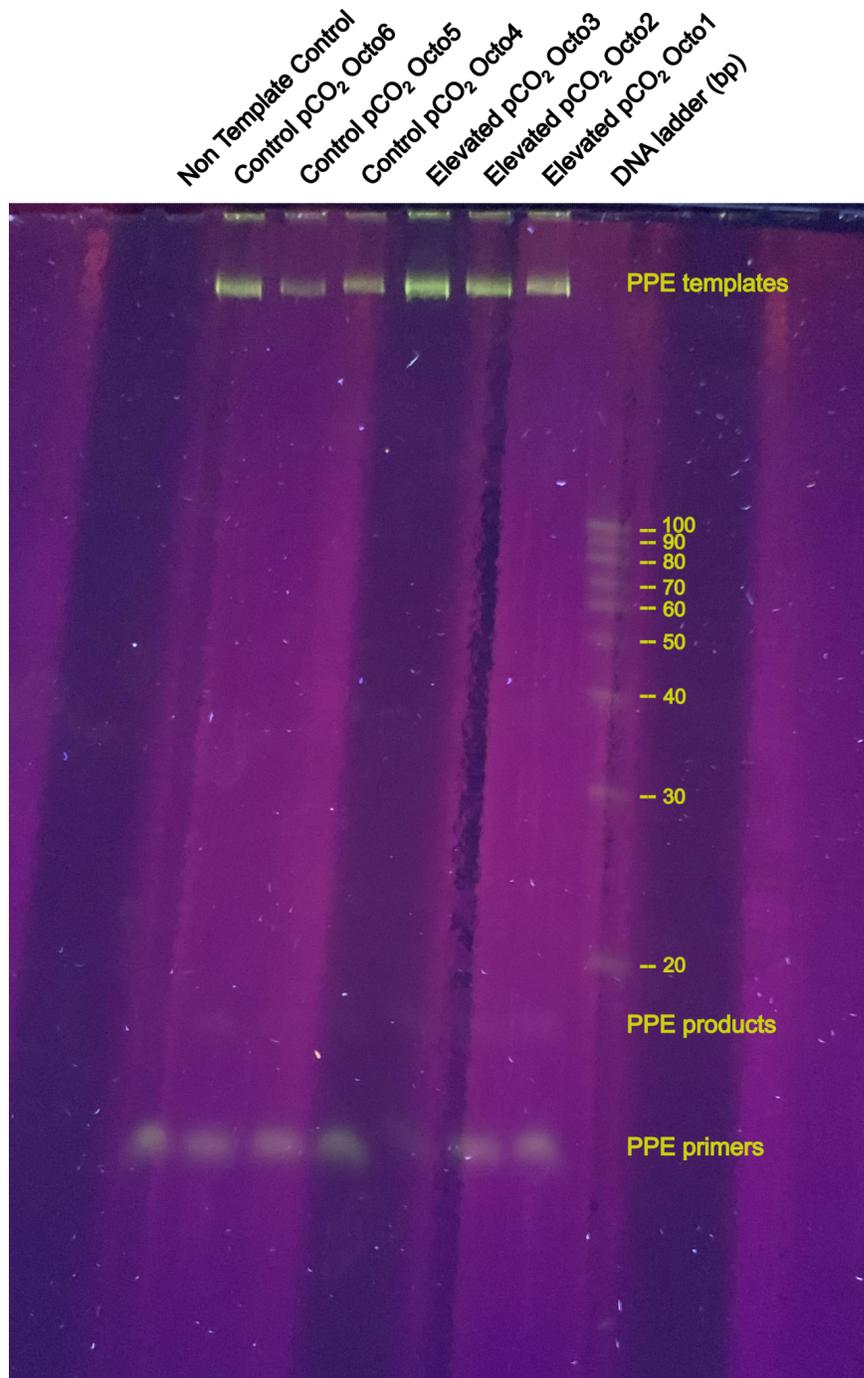gene from six *Octopus rubescens* in the 2020 verification experiment.

Figure 11. Gel scan of poisoned primer extension assays on the methionine adenosyltransferase 2 subunit beta-like gene from six *Octopus rubescens* in the 2020 verification experiment.
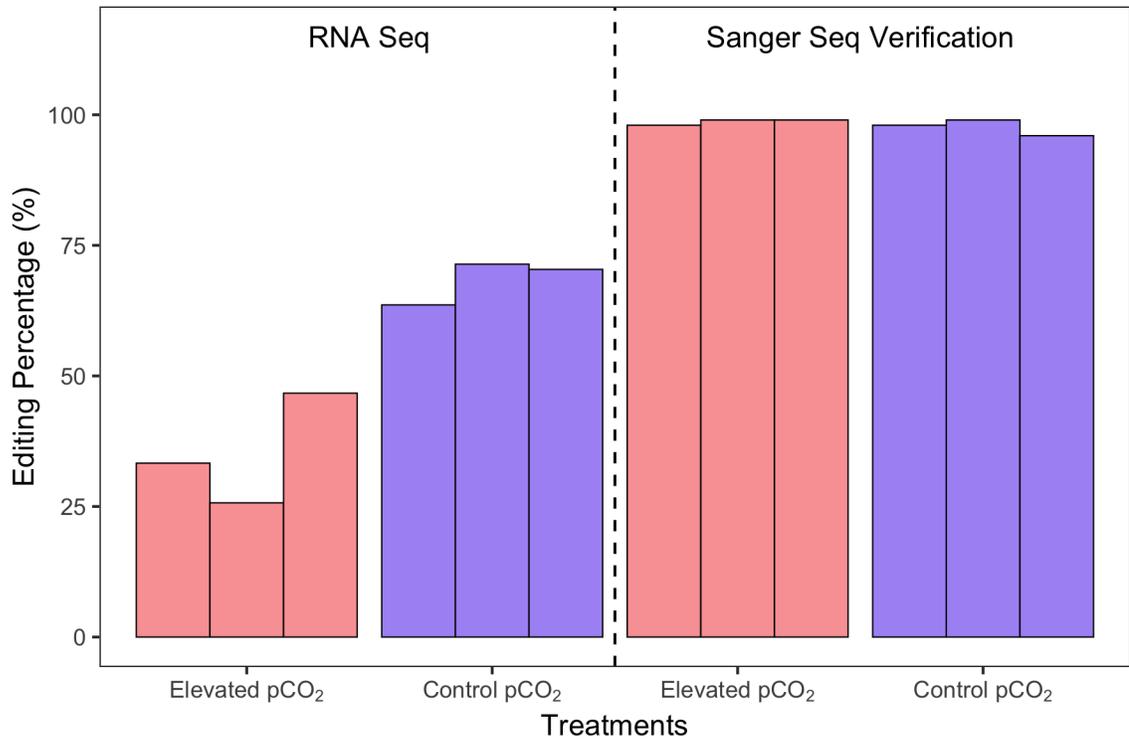
Figure 12. Editing percentages of the methionine adenosyltransferase 2 subunit beta-like

gene in *Octopus rubescens* in elevated pCO_2 and control pCO_2 treatments from

mRNA-seq data and Sanger sequencing verifications. Each bar represents the editing
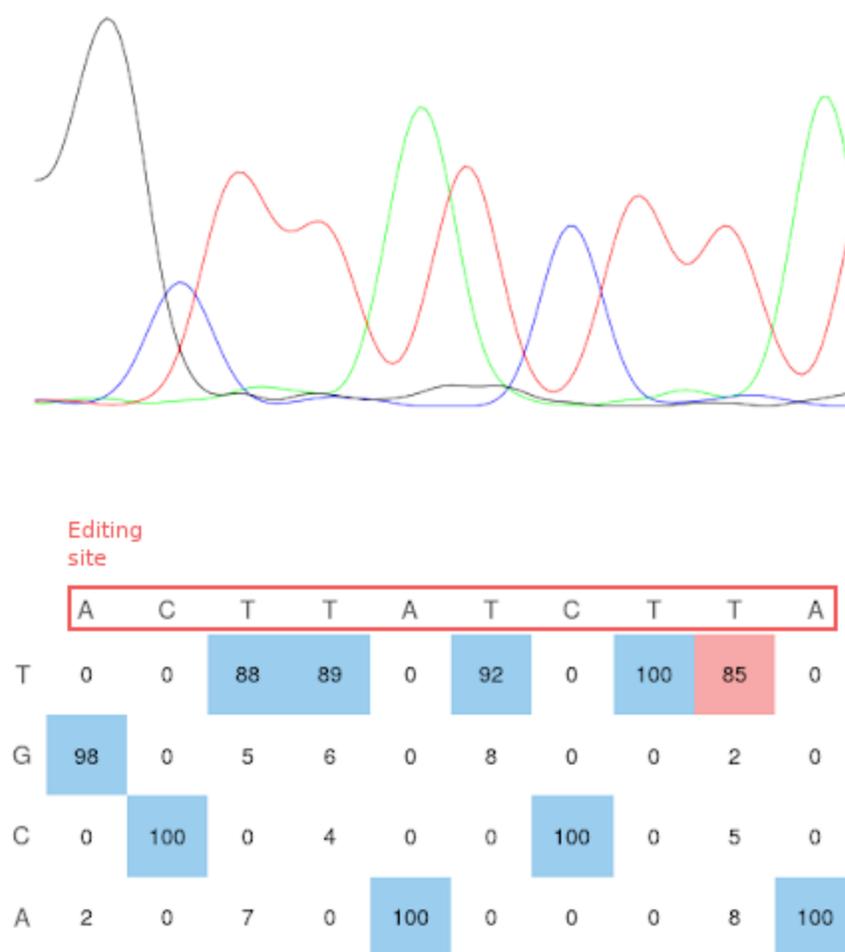
percentage in an octopus.

Figure 13. Sanger sequencing chromatogram of the methionine adenosyltransferase 2 subunit beta-like gene from one of the *Octopus rubescens* in control pCO_2 treatment. The red box highlights the segment of the transcriptome the editing site is in. 98% of the chromatogram signal at the editing site (1st position from left) corresponds to G versus A in the PCR amplicon sequence.
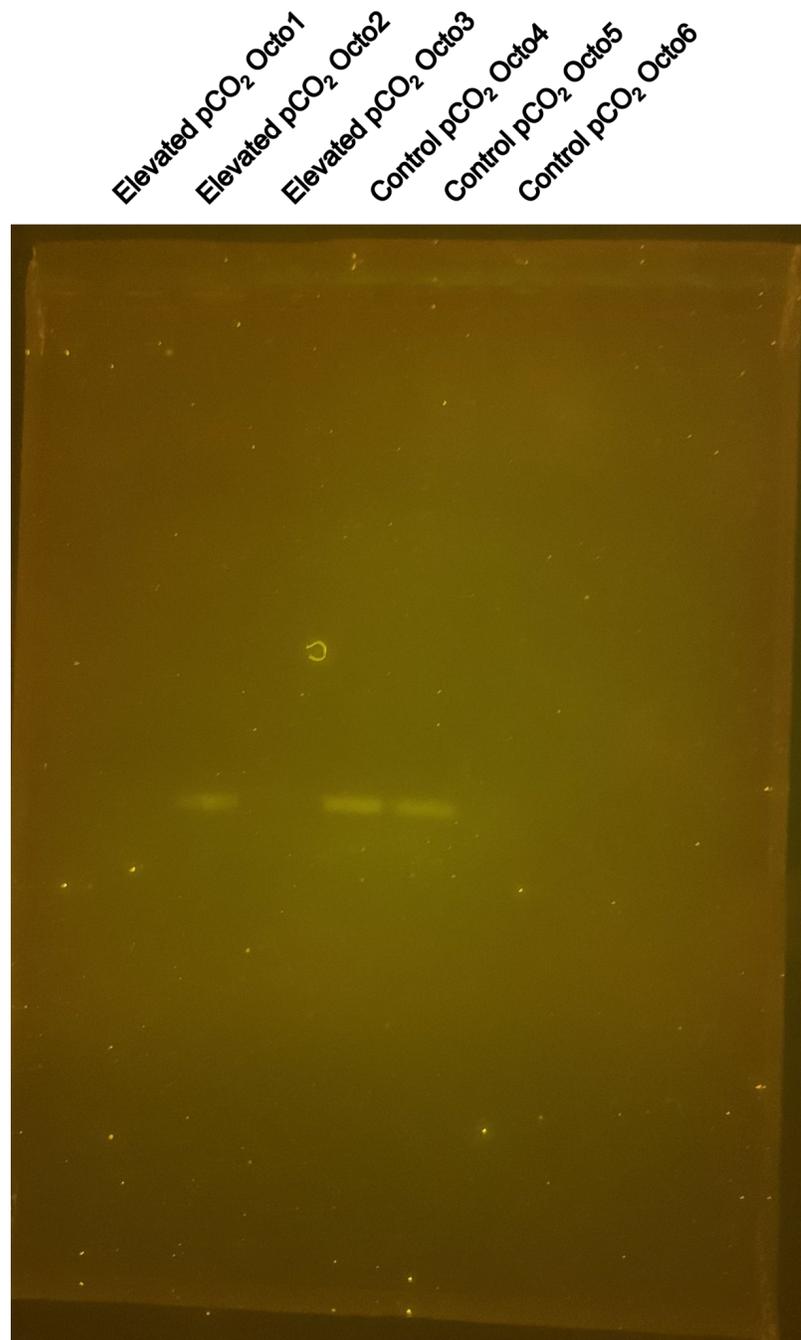
Figure 14. PCR amplification of six octopus samples using primers set targeting

thioredoxin-like protein 4A gene. Three out of six samples show amplification.
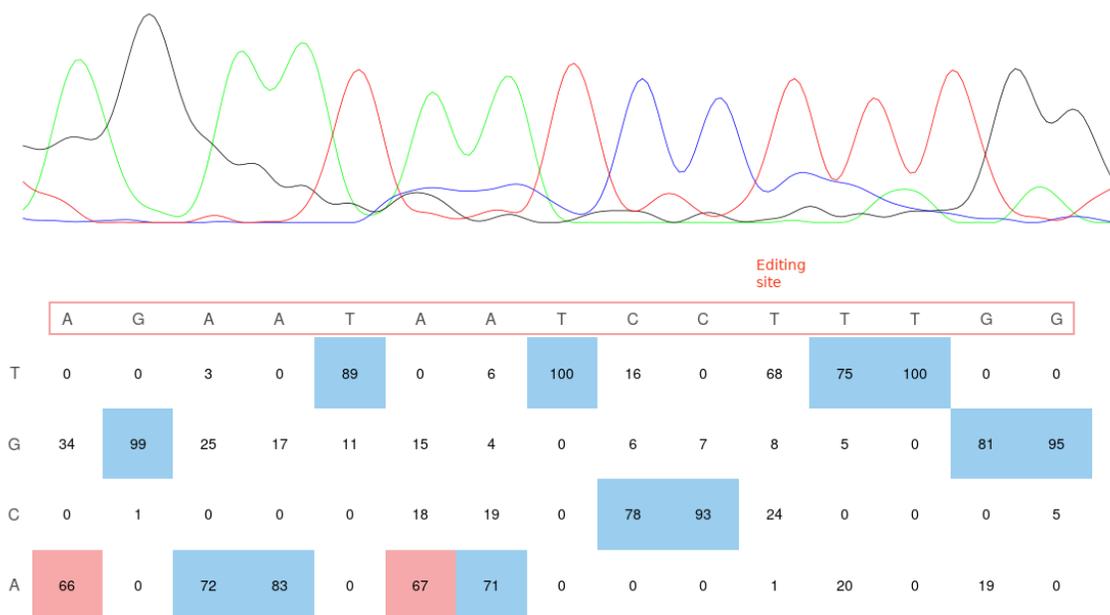
Figure 15. Sanger sequencing chromatogram of the thioredoxin-like protein 4A gene from one *Octopus rubescens* in control pCO$_2$ treatment. Red box highlights the segment of the transcriptome (reverse complemented) the editing site is in. The reverse complement of the edited strand was sequenced because the reverse primers were used, therefore A-to-I edit is shown as T-to-C.
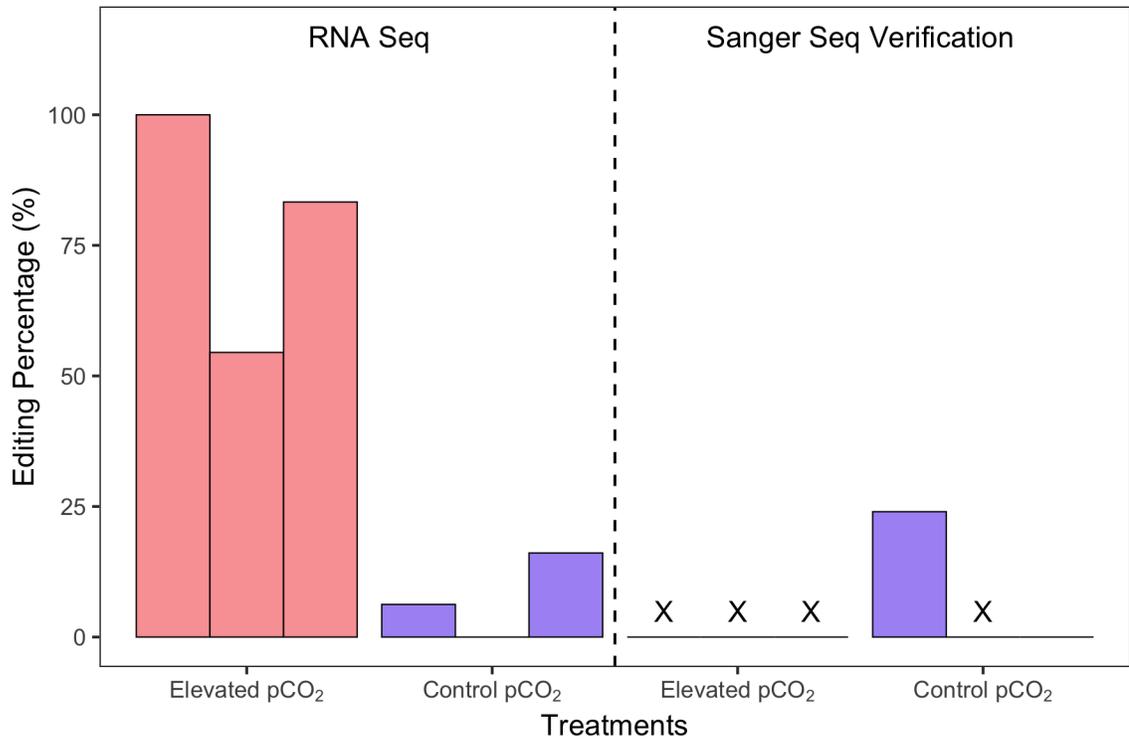
Figure 16. Editing percentages of the thioredoxin-like protein 4A gene in *Octopus rubescens* in elevated pCO_2 and control pCO_2 treatments from mRNA-seq data and Sanger sequencing verifications. Each bar represents the editing percentage in an octopus. "X" indicates missing data.

Figure 17. The proportions of nucleotides one nucleotide upstream of all the mismatch sites, A-to-G mismatch sites, and differential A-to-G mismatch sites in the gill tissues of *Octopus rubescens* in treatments from 2016 experiment. Dashed lines are the proportions of each nucleotide in the transcriptome.

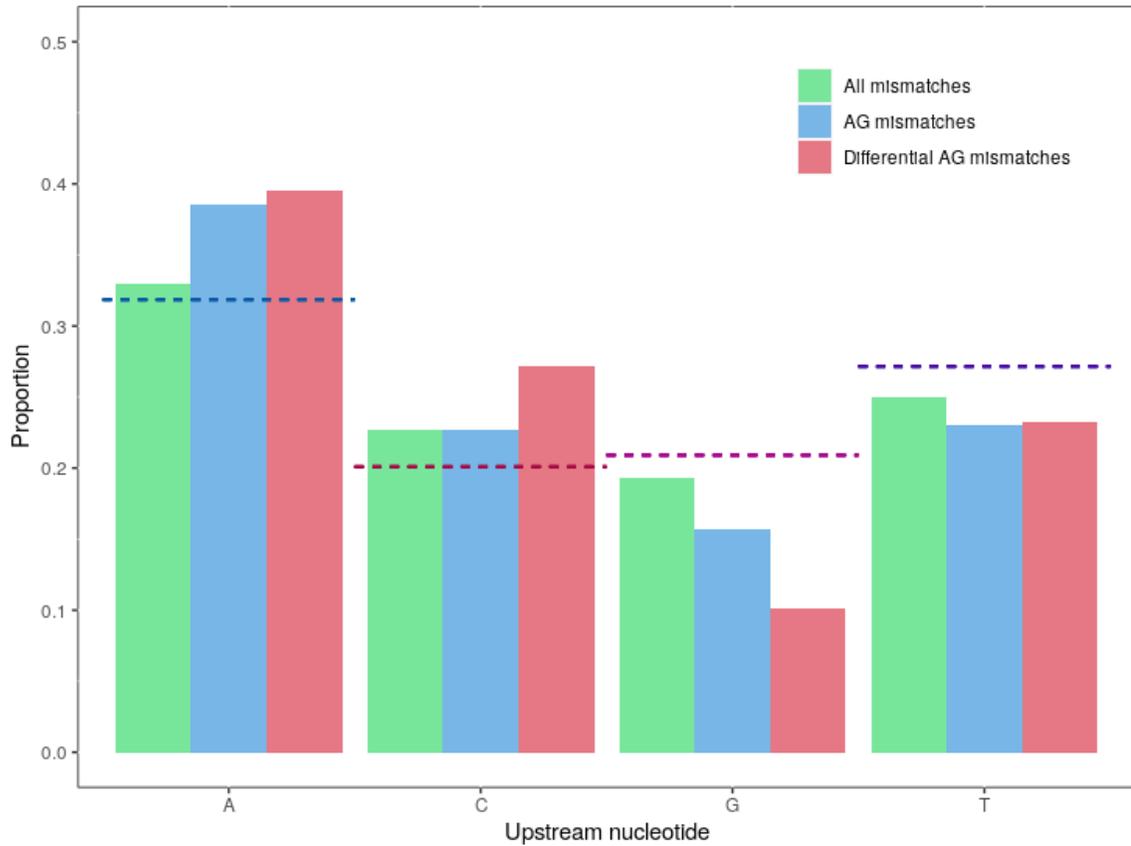**Figure 18.** The proportions of nucleotides one nucleotide downstream of all the mismatch sites, A-to-G mismatch sites, and differential A-to-G mismatch sites in the gill tissues of *Octopus rubescens* in treatments from 2016 experiment. Dashed lines are the proportions of each nucleotide in the transcriptome.
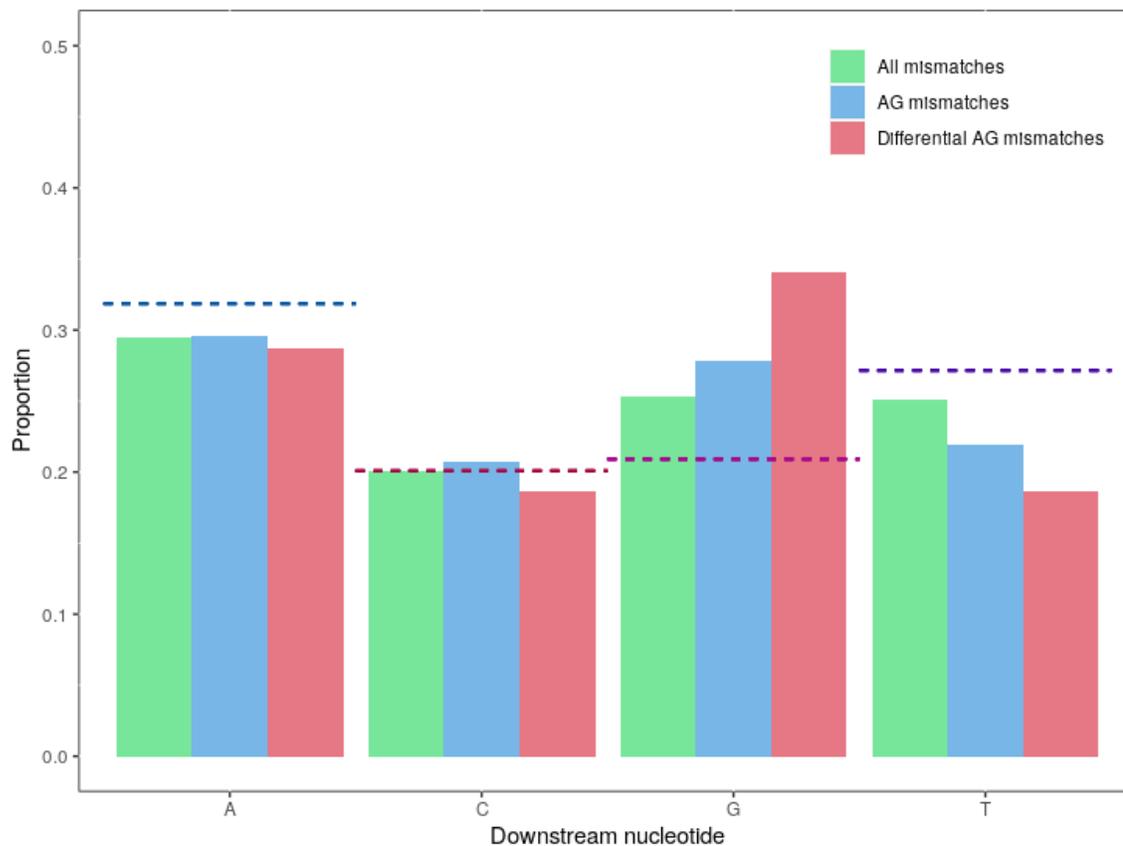
Tables

Table 1. Number of removed and corrected RNA reads from octopuses in 2016 experiment.

	Elevated pCO_2 Treatment			Control pCO_2 Treatment		
	Octopus 1	Octopus 2	Octopus 3	Octopus 4	Octopus 5	Octopus 6
Total RNA Reads	69,052,341	68,098,848	50,639,565	42,865,535	59,001,480	59,101,325
Reads removed by Fastp	138,093	205,305	235,915	238,919	142,811	120,941
Unfixable reads removed by Rcorrector	1,398,018	2,924,978	2,306,320	2,879,492	1,926,065	2,524,057
Reads corrected by Rcorrector	8,048,109	10,060,689	7,757,493	7,395,996	10,786,114	8,067,919

Table 2. The number of RNA sequences in each octopus individual from 2016 experiment that mapped to the combined *Octopus vulgaris* 18s rRNA and *Octopus cyanea* 28s rRNA sequences.

	Elevated pCO$_2$ Treatment			Control pCO$_2$ Treatment		
	Octopus 1	Octopus 2	Octopus 3	Octopus 4	Octopus 5	Octopus 6
Number of sequences	128,223	125,123	307,361	169,934	161,361,	152,670

Table 3. Alignment of gDNA and mRNA reads (from octopuses in the 2016 experiment) to swissprot open reading frames.

	Elevated pCO$_2$ Treatment			Control pCO$_2$ Treatment		
	Octopus 1	Octopus 2	Octopus 3	Octopus 4	Octopus 5	Octopus 6
Primary mapped mRNA reads	59,338,933	32,642,533	30,394,827	16,280,242	30,451,244	27,681,790
Pooled primary mapped gDNA reads	184,855,790					

Table 4. Transcriptome and Swissprot open reading frames metrics of *O. rubescens*, *O. vulgaris*, and *O. bimaculoides*. Data for *O. vulgaris* and *O. bimaculoides* taken from Liscovitch-Brauer et al. (2017).

	O. rubescens	*O. vulgaris*	*O. bimaculoides*
Total Trinity 'transcripts'	230,646	201,414	271,576
Total Trinity 'genes'	168,396	150,616	207,439
Contig 50 (nt)	623	784	749
Median contig length (nt)	323	334	340
Mean Contig length (nt)	525.49	579.98	574.31
Total assembled bases	88,490,825	87,354,553	119,133,693
Number of unique swissprot proteins	14,603	10,218	12,852
Mean ORF length (nt)	1,193	1,317	1,253
Median ORF length (nt)	807	969	882
Total ORF length (nt)	17,443,797	17,129,859	22,292,061

Table 5. Number of primary reads with different mapping quality filters in individual octopus mRNA-seq data and pooled gDNA-seq data.

	Octopus 1 mRNA	Octopus 2 mRNA	Octopus 3 mRNA	Octopus 4 mRNA	Octopus 5 mRNA	Octopus 6 mRNA	Pooled gDNA
Primary reads	59,338,933	32,642,533	30,394,827	16,280,242	30,451,244	27,681,790	184,855,790
>MAPQ10	12,061,071	15,244,652	9,401,499	7,675,213	12,055,872	12,673,169	41,686,836
>MAPQ20	11,362,518	14,608,456	8,780,612	7,134,540	11,384,796	11,963,434	31,622,965
>MAPQ30	10,475,239	13,585,864	7,916,346	6,426,345	10,220,344	11,090,019	9,260,549

Appendix A

***Octopus rubescens* Transcriptome BUSCO Output**:

The lineage dataset is: mollusca_odb10 (Creation date: 2020-08-05, number of species: 7, number of BUSCOs: 5295)
 3 # Summarized benchmarking in BUSCO notation for file Trinity.fasta
 4 # BUSCO was run in mode: transcriptome
 5
 6   ***** Results: *****
 7
 8   C:81.5%[S:55.0%,D:26.5%],F:2.4%,M:16.1%,n:5295
 9   4315   Complete BUSCOs (C)
10   2912   Complete and single-copy BUSCOs (S)
11   1403   Complete and duplicated BUSCOs (D)
12   128   Fragmented BUSCOs (F)
13   852   Missing BUSCOs (M)
14   5295   Total BUSCO groups searched

Appendix B

Methionine adenosyltransferase 2 subunit beta-like gene PCR amplicons sequence:
*GTGGTCATTGTATGTGGTGCTATC*ACCAATGTAAACTTGTGTGAATCAG
AGACAGAAAAG<u>A</u>CTT**ATCTTATTAACCGAGACA**GTATTGAACAGTTA
GGAAAACAAACCAAAAAGCGTGGTGCCAAATTGATCTTCTTCTCAACCAATTA
TATCTTCAGTGATCCTCTCAAACCAGATTTACCCCAAGATAGAATCGATATGTTA
GCTAATGATGTCGGCTTAAAGGAAGACGCACGTCCCAACTGTATGAACGTCTT
TGGTAAATCTAAGCTGGCCTCTGAGAAACTATTTCAAGATGATGACACCAATGT
CTTAATTATACGCACATCAGGTGTGTATGGACCTGATATCAAGAAAAGAATTT
TGTGTACCAAGTTATCGGGAATCTTCTAGCTGGGAATAAAATGACGCTGCTTAC
AGATGAAATTCAGTGTCCAATTTATACTGAGGATTTGTGTAGAGCGACACTTCA
ACTGATGGAGAAAATTGTTCAGGAATTTACCATGTTGTAGGACCAGAGGCAT
TCAACAAACACGCATTTGGTGTGAAGATTGCTGAAATCTTCAATCTTGATACTT
CGATGTTAATTCCGATGTCATCAGAAGAACTCAAAT*TAGCTGCCAAAAGGC*
CGAAT

Thioredoxin-like protein 4A gene PCR amplicons sequence:
*CTGTGTATCGAGGGGCAAGA*AAGGGACGTGGTCTTGTTGTCTCACCAA
<u>A</u>GGATTATTCTACTAAATACCGTTATTAAAATATTTTTTAATTAACTG
ATGAAAGAGAAGTAATTCCTTGGTGTGTCTAGAACAAGAATCCCTTTCCATATT
GCCCTCATTGATTTCACT*GTGCTCTCTCAGCTGGGTTTA*

PCR primers are *italicized.*
<u>Editing sites</u> are <u>underlined</u>.
Poisoned primer extension primers are **bolded**.