Genetic Consequences of Anthropogenic Pressures on Green Sea Turtles (*Chelonia mydas*) in the Eastern Mediterranean Sea

Arizona State University

LSC 586: Capstone II in Biol Data Sci (2025 Spring B)

Group 1: Daniel Bihnam, Noelle Eschen, Lucas Smith

May 2, 2025

Introduction

Green sea turtles (*Chelonia mydas*), also known as the black sea turtle or Pacific green sea turtle, are endangered aquatic reptiles that play a vital ecological role in maintaining healthy marine ecosystems (Wyneken et al., 2013). Being primarily herbivores, green sea turtles graze on seagrass meadows, which helps maintain seagrass meadows and promotes biodiversity (Oliver Ridley Project, n.d.). Green sea turtles can be found in tropical and subtropical regions worldwide, with two distinct populations located in the Atlantic and Pacific oceans (Álvarez-Varas et al., 2021). Although green sea turtles play a crucial role in maintaining marine ecosystems, their populations are threatened by climate change, pollution, and habitat destruction (National Oceanic and Atmospheric Administration [NOAA], n.d.). These anthropogenic pressures have profound implications for green sea turtle life cycles and population dynamics, and there is rising concern about their long-term survival. A distinct feature of green sea turtle biology is temperature-dependent sex determination (TSD), where the temperature of nesting sites influences the sex ratio of hatchlings (Jensen et al., 2020). When eggs are incubated at lower temperatures (27 to 29 °C), hatchlings will be male, and at higher temperatures (29 to 31 °C), hatchlings will be female (National Ocean Service, n.d). When the sands on nesting beaches are warmer during the middle third of the incubation period, there is a higher ratio of female turtles (Tezak et al., 2020). As global temperatures rise, sex ratios among green sea turtles could become skewed and jeopardize population viability (Broderick et al., 2001).

Understanding the genetic consequences of anthropogenic pressures is critical for implementing effective conservation strategies and for monitoring current conservation programs. One method of identifying genetic changes in green sea turtle populations is through the detection of single-nucleotide polymorphisms (SNPs). The presence and frequency of SNPs in green sea turtle genomes can highlight genetic variation amongst populations, which will help to identify genomic abnormalities caused by a variety of environmental stressors, including climate change and pollution (Brauer et al., 2023). By investigating how climate change and environmental stressors affect the genetic health of green sea turtles, this research aims to shed light on green sea turtle genetic diversity and the effectiveness of current conservation strategies. To better understand the implications of anthropogenic pressures on green sea turtle populations, our study primarily focuses on populations of green sea turtles in the Mediterranean Sea due to genetic and environmental data availability.

There are several conservation strategies currently in place for green sea turtles in the Mediterranean Sea including the “Action plan for the conservation of marine turtles in the Mediterranean” (Action Plan)(RAC/SPA–UNEP/MAP, 2022) whose main goals are to protect marine habitats, reduce anthropogenic mortality, enhance legal protections, and coordinate and collaborate among Mediterranean countries for consistent and effective conservation measures. Since being implemented in 1989, the Action Plan has been updated and revised in 1999, 2007, and 2013. In 2022, a comprehensive evaluation of the Action Plan was conducted to review its implementation on regional and national levels. This evaluation highlighted several areas of progress, including increased knowledge of turtle population dynamics and enhanced protection of nesting sites. The evaluation also identified ongoing challenges such as increased mortality due to bycatch, habitat degradation, and loss. While the Action Plan has facilitated significant advancements in marine turtle conservation, the evaluation indicates that persistent threats require

ongoing, adaptive, and collaborative efforts to ensure the long-term survival of these species. We hypothesize that human-driven environmental changes are having detectable changes in the genome of green sea turtles across the eastern Mediterranean. By understanding both environmental stressors as well as genetic responses, this study aims to assess the effectiveness of current conservation strategies, as well as inform the development of new improved strategies to ensure long-term population survival.

Methods

This study used a quantitative approach to compare biogeochemistry and sea surface temperature data to SNPs to investigate whether anthropogenic influences are potentially pressuring green sea turtle populations in the eastern Mediterranean Sea. Biogeochemistry and sea surface temperature were acquired using Python 3.11 (Python Software Foundation, 2022) via the Copernicus API, Copernicus (n.d.). The following parameters were supplied to the API: the geographic region was refined with a longitudinal range between 23°E and 37°E and a latitudinal range between 30°N and 37°N, and the time frame was narrowed to between 2000-01-01 and 2025-02-20. The data products downloaded using Python and processed using R v4.3.3 developed by the R Core Team, 2025 were High-Resolution L4 Sea Surface Temperature Reprocessed, which provided the monthly averages for sea surface temperature at a spatial resolution of 0.05° × 0.05°, and Mediterranean Sea Biogeochemistry Reanalysis that contains the monthly averages for nitrate, phosphate, ammonium, phytoplankton carbon biomass, chlorophyll, net primary production, dissolved oxygen, pH, dissolved inorganic carbon, and alkalinity at a spatial resolution of 0.042° × 0.042°, both of which were sourced from Copernicus Marine Service, 2025. The files were downloaded in netCDF format and were converted to a data frame using the Stars library (Pebesma & Bivand, 2023). Tidyverse was then used to remove any incomplete data sets by removing NAs and filtering the depth of the readings to 13.32 meters, an increment provided by Copernicus used as a cutoff (Wickham et al., 2019). This selection highlights the changes happening at shallower depths because green sea turtles spend much of their time at shallow depths, often not exceeding 7.9m during foraging (Southwood et al., 2009). The data came in a format where each coordinate in the selected latitudinal and longitudinal range was provided with a monthly mean value at that location. To get the general health and trend of the region of the Mediterranean Sea, our study focused on the mean of all the monthly means, removing the coordinates from the equation, and providing the overall trend of the ecosystem on a macroscale and then computed linear models using R's base linear model function using variables that resulted in less than five variable inflation factor (VIF) using the car package (Fox, Weisberg, & Price, 2023; Fox & Weisberg, 2019; James et al., 2021). The macro trend was visualized using ggplot2 (Wickham, 2016), and Pearson's correlation coefficients were visualized using ggcorrplot (Kassambara, 2019) and retrieved using the Hmisc package (Harrell & Dupont, 2024). This method of eliminating coordinates limits the scale to a broader vision of the ecosystem, removing the possibility of seeing a finer resolution of the ecosystem; however, since the analysis is primarily investigating the trends across the region, reducing the dataset by this method doesn't reduce the quality of our dataset or the trends inferred from it.

The genomic data required for SNP identification was sourced from a publicly available sequencing project on the National Center of Biotechnology Information (NCBI) GenBank (NCBI,

2024). This study was conducted across various beaches and institutes in Israel and has amassed double-digest restriction site-associated DNA sequencing (ddRAD-seq) data from 268 distinct green sea turtles. ddRAD-seq is useful for studies on species such as the green sea turtle due to its lower cost and DNA requirements, with the drawback being its limited representation when compared to whole genome sequencing techniques. This sequencing technology is commonly used to detect SNPs. As this data is publicly available on NCBI GenBank, it can be requested and downloaded to a specified cloud location. A reference genome was also required for this project and was obtained from NCBI GenBank. The most recent green sea turtle genome assembly was published in 2021 (NCBI, 2021), and was constructed with the Burrows-Wheeler Alignment Tool (BWA) to prepare for analysis (Li & Durbin, 2009), and indexed with SAMtools (Li et al., 2009). This genome is required to identify SNPs present in each individual sample.

For the SNP analysis, the ddRAD-seq data was available as raw sequencing reads in fastq format. This data was then processed in the command line with bash scripts on the Arizona State University (ASU) Sol HPC system (Jennewein et al., 2023), utilizing batch processing techniques such as job scheduling arrays to parallelize these computationally heavy operations. Compression and file merging were handled using pigz (Adler, 2022) and htslib (Bonfield et al., 2024). For this data preparation, a full genomic bioinformatics pipeline was implemented. Quality control (QC) of the raw reads was necessary to assess the sequencing quality using the tool FastQC (Andrews, 2010). Following this step, low-quality bases, and adapter sequences leftover from the sequencing library prep were trimmed using Trimmomatic (Bolger et al., 2014). Next, the trimmed raw data were aligned to the previously identified reference genome using BWA. The output of this tool was converted to a BAM file and indexed using SAMtools. At this point, the trimmed sequencing reads were aligned to the reference genome, and proper alignment was verified using the Interactive Genome Visualizer (IGV) software (Robinson et al., 2011). SAMtools was also used to generate flagstat reports of the read content and quality of the resulting BAM file. Following the BAM QC, SNP variants were called using BCFtools (Danecek et al., 2021). The output of this is a raw VCF file of both SNPs and insertions and deletions (indels) for each sample. SNPs were then filtered with BCFtools based on sequence quality, depth, and overall coverage before proceeding to further analysis. Downstream analyses required using a merged VCF file containing the SNPs of all samples in the cohort. Before creating this file, samples with < 70% properly mapped reads were excluded, reducing the cohort size from 268 to 237. All sample metadata were downloaded from NCBI GenBank along with the sequencing data. This provided the geographic source of each genetic sample.

Site Frequency Spectrum (SFS) allowed for the determination of the minor allele frequency distribution across the cohort, as well as in select regions. This aided in determining the growth behavior of the population and showed any evolutionary stress on the population (Rasmussen, M. S., 2022). To calculate this, VCFtools (Danecek et al., 2011) was used to count the occasions of each SNP detected across the cohort, and measure the minor allele frequency. Ggplot2 was used to visualize the data.

Principal Component Analysis (PCA) is a dimensionality reduction tool used to capture and visualize genetic variation across the different geographic locations of the cohort. Trends in genetic variation can help show the history and evolution of population structure (McVean, 2009). To complete this analysis, VCFtools was used to capture the observed genetic variation from a

merged VCF file containing the SNPs of all 237 samples and reduce it into ten principal components. Plotly (Plotly Technologies Inc., 2015) was used to visualize the first three principal components in a three-dimensional PCA plot.

Minimum Spanning Networks (MSNs) are useful in genomic studies to visualize genetic distance and differences between populations. For this analysis, a merged VCF file of SNPs was used to create the network. R packages such as vcfR (Knaus & Grünwald, 2016), adegenet (Jombart, 2008), poppr (Kamvar et al., 2014), dartR (Gruber et al., 2018), and igraph (Csárdi et al., 2025) were used to create the MSN. This visualization is important for characterizing the relationship and genetic overlap amongst the different geographical green sea turtle populations.

Mutational signature analysis is incredibly useful in studying the mutagenic processes that drive the accumulation of mutations. This is popularly used in somatic mutations acquired in human cancer and focuses on the chemical cause of the molecular changes in the DNA (Cosmic, 2020). However, using this approach to study heritable mutations in green sea turtles will provide a glimpse into the mutational history of these populations, and what external environmental factors caused them. Python was used to extract the mutations and their trinucleotide (nucleotide before and after the SNP) context from the VCF files and reference genome using pandas (The pandas development team, 2024) and pyfaidx (Shirley et al., 2015). These results were then grouped by mutation type, and plotted in R with ggplot2. To further assess the composition of this overall signature, non-negative matrix factorization (NMF) was employed in R with the NMF package (Gaujoux & Seoighe, 2010). NMF was used to decompose the overall signature into four smaller contributing signatures with 1,000 NMF iterations to determine if there are unique mutagenic processes occurring in each region. A variety of signature extractions were performed, and four signatures were determined to be the most well-fit to the data, with minimal stochastic effects from the sample size or sequencing quality.

A generalized workflow of the processing pipeline from raw sequencing reads to data visualizations is depicted below (Fig.1). This outlines the progression from raw sequencing reads sourced from all 268 samples into VCF files containing all SNPs observed across the study.

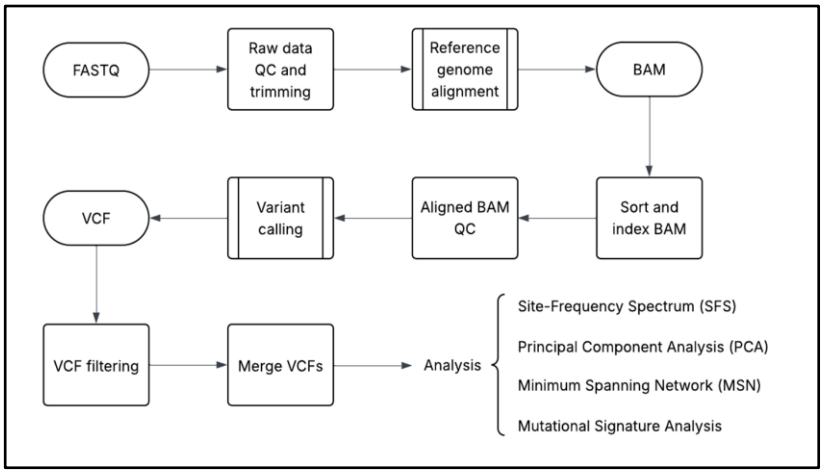


Fig. 1: Overview of ddRAD-seq bioinformatics pipeline used to generate SNP data and perform downstream analysis of *Chelonia mydas* samples. Circled steps correspond to the data file type (FASTQ,

BAM, VCF), steps contained in a single square represent processes modifying a file (QC, filtering, trimming), and steps contained in squares with the double lines correspond to major steps resulting in a change of file format (reference genome alignment and variant calling). The figure was created in Lucid (lucid.co).

Results

Assessing environmental trends reveals how ecological systems are adapting due to environmental pressures, which may drive negative or positive selection over time. This study will investigate the transformations happening in the eastern Mediterranean Sea using biogeochemistry and sea surface temperature data through correlation coefficients, linear regressions, and scatter plots. To investigate ecological changes, dissolved inorganic carbon (DIC), pH, alkalinity, sea surface temperature (SST), and surface CO_2 flux were selected from the Copernicus database. These variables were selected as potential influences on pH as it is a vital chemical balance that affects marine life and its food supply. Their overall trends across time can be seen in Fig. 2.

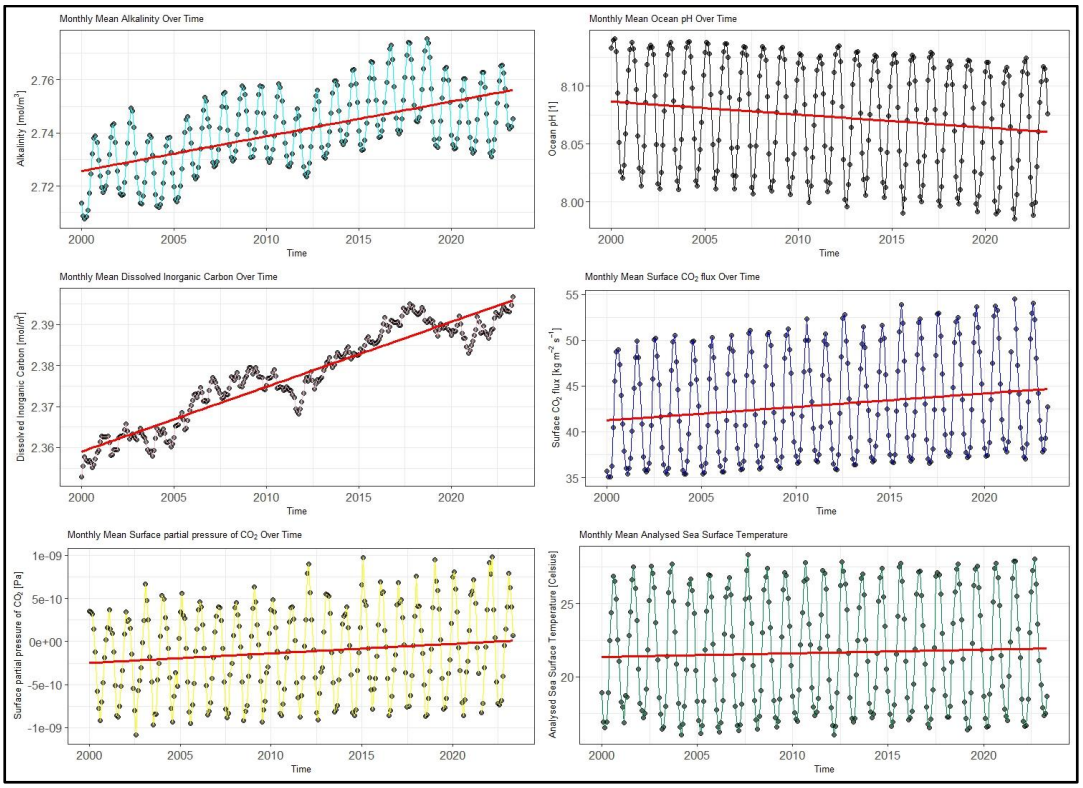


Fig. 2: Monthly mean values for alkalinity, pH, surface partial pressure CO_2, dissolved inorganic matter, surface CO_2 flux, and sea surface temperature.

Alkalinity emerged as a significant variable due to its role in buffering capacity. Typically, alkalinity and pH are measured together, as pH measures water's acidity, while alkalinity measures water's ability to buffer changes in acidity (Montana State University Extension, n.d.; Middelburg, Boyd, & Ilyina, 2020). Therefore, the pH should remain stable given the adjustments in the ecosystem's alkalinity. However, pH has been on a significant (F = 7.453, p = 0.007) negative

trend over time in the eastern Mediterranean Sea, as shown in Supp. 1, even though alkalinity has increased, suggesting external influences beyond buffer capacity. These three variables were found to be highly significant ($p < 2e^{-16}$) in linear models of pH using alkalinity, DIC, and surface CO_2 flux. For every 1 mol/m^3 increase of DIC, pH increases by 0.288 units, which is the opposite of what was noticed in the study by Merlivat et al. 2018 conducted in the Mediterranean, likely meaning this is a region-specific trait. The increase of DIC and increase of pH suggests bicarbonate and carbonate ions are actively buffering acidification (Cole & Prairie, 2014). The correlation yielded an r = -0.134 (p = 0.0246), which is counter to what the model produced, meaning the simple correlation was unable to identify the same trend. This association could be influenced by other variables, although the VIF was quite low for the model. For every 1 mol/m³ increase in alkalinity, the pH decreases by 0.251 units. The correlation between pH and alkalinity in this region was r = -0.648 ($p < 2e^{-16}$), reinforcing alkalinity is increasing as pH is decreasing in this region. This is not a typical association and suggests a complex interaction where buffers are unable to mitigate the constant change in acidity. Finally, for every 1 kg m^{-2} s^{-1} increase in surface CO_2 flux, pH decreases by 0.008 units, suggesting more CO_2 is being fluxed into the sea than leaving it. As shown in Fig. 3, there is a strong negative correlation between pH and surface CO_2 flux, with a correlation coefficient of r = -0.998 ($p < 2e^{-16}$). Therefore, as surface CO_2 flux increases, pH decreases, verifying that these two variables are closely connected. While these three variables provided a multiple R^2 value of 0.9985, explaining 99.85% of the variance in pH, it is essential to take the extremely high significance with caution, as monthly variability could be inflating the model's perceived strength.

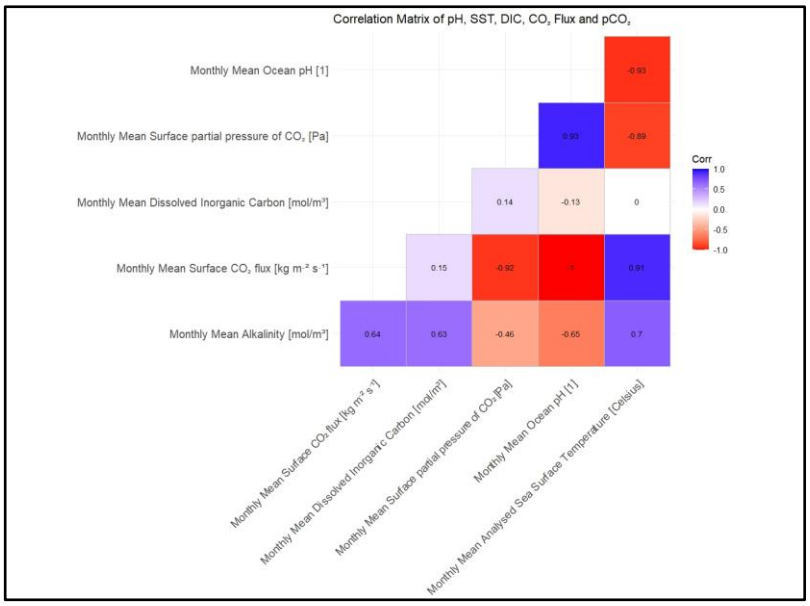


Fig. 3: Correlation matrix with correlation coefficients of monthly mean values for alkalinity, pH, surface partial pressure CO_2, dissolved inorganic carbon, and surface CO_2 flux.

SST and pCO_2 are not included in this linear model for pH because of their collinearity with other variables, indicated by a variance inflation factor value greater than 10, resulting in potential inflation of values making it difficult to interpret the model properly. Given that the model without SST and pCO_2 was highly significant ($p < 2.2e^{-16}$), with an F = 63,260, it was determined to be better without them.

DIC and alkalinity were highly significant ($p < 2.2e^{-16}$) with an $F = 188.3$. When 1 mol/m^3 of DIC increases, alkalinity increases by 0.826 mol/m^3. An explained variance of 40.3% (Multiple $R^2 = 0.403$) and a correlation of $r = 0.635$ ($p < 2e^{-16}$) support the idea that DIC enhances the ability of seawater to buffer acidification, resulting in increased alkalinity measurements (Middelburg, Boyd, & Ilyina, 2020).

This analysis found that surface CO_2 flux was of interest because CO_2 has been a global concern as an indicator of climate change, and it has a negative association with pH. In a study by Turi et al. (2023), it was suggested that the increase in temperatures has decreased CO_2 solubility, allowing CO_2 to be released more easily. Between 2000 and 2019, it was noticed that CO_2 uptake into the ocean had reduced by 13% (Turi et al., 2023). Regardless, the eastern region of the Mediterranean Sea is still taking more CO_2 as temperatures increase than releasing based on the linear model, $F = 1{,}311$ and $p = < 2.2e^{-16}$, where for every 1-degree Celsius increase, CO_2 flux increases by 0.620 kg m^{-2} s^{-1}, suggesting that the temperature directly affects CO_2 flux within this region and allowing CO_2 accumulation to occur. With this in mind, the increase in CO_2 flux between the atmosphere and the sea, resulting in higher CO_2 concentrations, can be directly related to the acidification of the eastern Mediterranean Sea.

Given the increasing acidification linked to surface CO_2 flux, it is important to understand the effects it has on the broader food web. The acidification of the eastern Mediterranean Sea is hazardous to the green sea turtle, as it has a limited diet consisting of algae, seagrasses, and seaweed (SEE Turtles, n.d.). This is important to note that as phytoplankton, a key species in the base food web, provides essential nutrients to the entire marine food web and is dependent on CO_2 concentrations. However, the quality of these nutrients is affected by acidification, as noted in a study by Peng, Hutchins, and Gao (2020). There are also other key variables, such as net primary productivity and chlorophyll, that can shed light on how acidification is altering the resources available to the green sea turtle in the Mediterranean Sea.

A linear model consisting of net primary production as the predictor and chlorophyll and phytoplankton carbon biomass as influencers resulted in $F = 572.3$ and $p < 2.2e^{-16}$ with an explained variance of 80.46% (Multiple $R^2 = 0.805$). It should be noted that pH was excluded from the model due to its collinearity with other variables and was instead run individually with each variable to understand the influence pH had on each variable. As chlorophyll increases by 1 mg/m^3, net primary productivity decreases by 65.907 mg/d/m^3. However, this value is not ecologically realistic because the variables are captured in different units at a vastly different scale. After making adjustments for the scaling issue by dividing by 100, the results suggest that for every 0.01 mg/m^3 increase in chlorophyll, the net primary productivity decreases by 0.659 mg/d/m^3 ($p < 2e^{-16}$). This adjustment aligns more with the expected ecological variations seen. The correlation coefficient, $r = -0.611$ ($p < 2e^{-16}$), supports the model that net primary productivity decreases and chlorophyll increases. An increase in chlorophyll has typically been associated with increased net primary productivity, which is opposite what has been seen in this linear model. However, it is crucial to note that in other studies, such as Dvoretsky et al. (2023), complex relationships between the two were found, where they did not always trend together in certain regions. This could be a trend to watch in the future. As phytoplankton carbon biomass increases by 1 mmol/m^3, net primary productivity increases by 13.792 mg/d/m^3. Again, this is not a realistic ecological scale, as seen in Fig. 4. To adjust for the scaling issue, the values were divided by 10. After making

adjustments for the scaling issue, the results suggest that for every phytoplankton carbon biomass 0.1 mmol/m³ increase, net primary productivity increases by 1.379 mg/d/m³ ($p < 2e^{-16}$). The correlation coefficient ($r = 0.110$) suggested a weak positive association, but it was not statistically significant ($p = 0.067$). This suggests the model is accounting for other influencing variables that the simple correlation does not capture. Although the correlation is not significant, the linear model provides a to-be-expected outcome: phytoplankton carbon biomass is of great significance on net primary production in the eastern Mediterranean Sea, considering phytoplankton makeup a large portion of the marine food web.

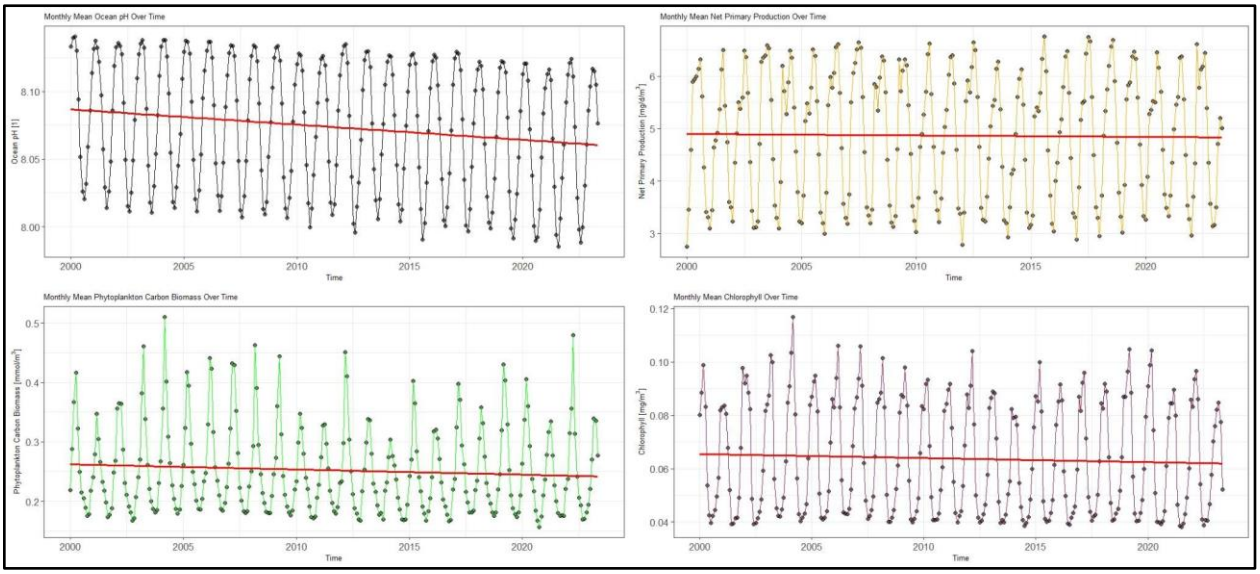


Fig. 4: Monthly mean values for net primary production, pH, phytoplankton carbon biomass, and chlorophyll.

Understanding how chlorophyll and phytoplankton influence net primary productivity helps better understand how the effects of acidification are impacting the nutrients available to primary producers, such as algae and seagrass. The pH and chlorophyll were highly significant ($p < 2.2e^{-16}$) with an $F = 1{,}416$. This shows that for every 1-unit increase in pH, chlorophyll increases by 0.411 mg/m³ and explains 83.54% (Multiple $R^2 = 0.8354$) of the variance in chlorophyll. For this study, this means that chlorophyll is affected by the acidification of the eastern Mediterranean Sea, supported by a correlation coefficient of $r = 0.914$ ($p = 2.2e^{-16}$) as shown in Fig. 5, indicating a strong positive association with high significance. The decline in chlorophyll, an indicator of primary productivity, is likely due to a decrease in the abundance of aquatic plants and/or phytoplankton, which is also declining due to acidification. Similarly, pH and phytoplankton carbon biomass had a high significance ($p < 2.2e^{-16}$) to each other, with an $F = 241.9$. However, pH only explained 46.44% (Multiple $R^2 = 0.4644$) of the variance in phytoplankton carbon biomass, meaning there is a greater influence outside of pH. The correlation coefficient supports a moderate association, $r = 0.681$ ($p < 2.2e^{-16}$), suggesting that they move together, but only some of the time. pH and net primary production were highly significant ($p < 2.2e^{-16}$) with the model producing an of $F = 187.4$ and only explaining 40.18% of the variance seen (Multiple $R^2 = 0.4018$) . This shows that for every 1 unit increase in pH, net primary production decreases by 16.137 mmol/m³ ($p < 2e^{-16}$). This valuation is again skewed due to differences in units. When divided by 10, the result is that for every 0.1 unit increase in pH, net primary production decreases by 1.614 mmol/m³, which is a more realistic measure of change. The correlation coefficient, $r = -0.634$ (p

$< 2e^{-16}$), demonstrates that as pH decreases, net primary production increases but this does not fully explain why net primary production is increasing. This model produced a somewhat unexpected result as both phytoplankton carbon biomass and chlorophyll indicated positive associations with pH and, while the original linear modeling of them with net primary production indicated that they could explain roughly 80% of the variance, it just happens to be that there is likely another primary source of explanation for net primary production not included within this data set.

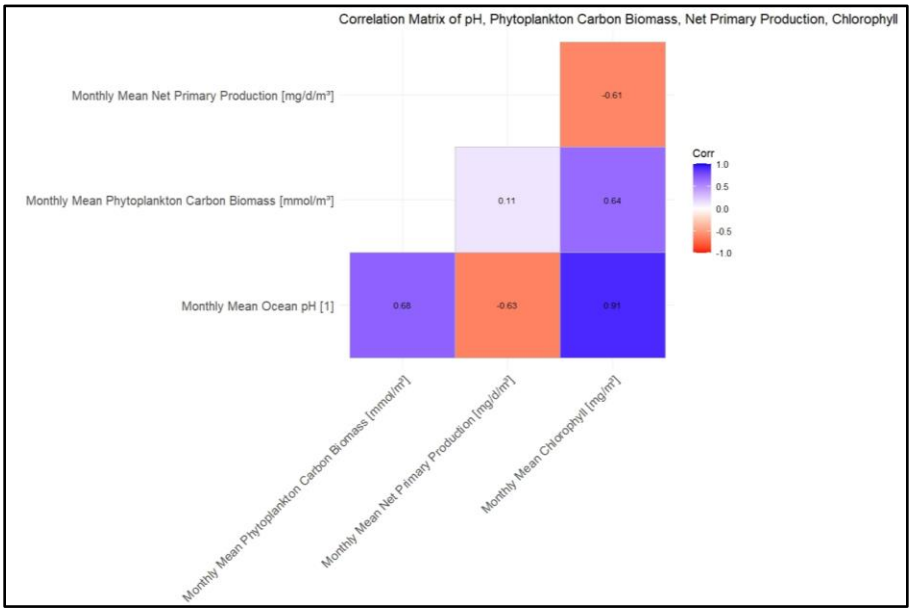


Fig. 5: Correlation matrix with correlation coefficients of monthly mean values for net primary production, pH, phytoplankton carbon biomass, and chlorophyll.

Although net primary production moves in the opposite direction in this data set, which as mentioned, can be attributed to other environmental factors that thrive in reduced pH levels, key nutrient readings like chlorophyll and phytoplankton carbon biomass can help indicate the available food resources for marine life. This means there could be a decline of algae, seagrasses, and seaweed in the eastern Mediterranean due to acidification, indicating a reduction of food availability for marine life at higher levels in the food web. The acidification can be attributed to the increase in CO_2 flux into the eastern region of the sea, which is attributed to increased anthropogenic CO_2 in the atmosphere. Being able to track increased anthropogenic CO_2 emissions helps understand the environmental risk factors that may be threatening the genetic health of green sea turtles in the eastern Mediterranean.

After processing and filtering the genomic sequencing data, a series of population genetics and statistical methods were applied to the data to discover the patterns of genetic variation across various nesting sites. These results provided insight into allele frequency distribution, population structure, and mutagenic processes which can serve as a window into the way a population is reacting to a changing ecosystem.

By analyzing the genomic sequencing data of green sea turtles, trends in genetic variation and population health can be inferred and connected back to environmental stressors. The raw

VCF data contained 1,372,864 variants detected across all samples. Stringent filtering was required to remove SNPs from low-quality regions that may have gotten less sequencing coverage due to the sparse nature of ddRAD-seq, as well as any low-quality samples. After excluding any samples with < 70% properly mapped reads to the reference genome and filtering for only SNPs with a Phred-scaled quality score > 30 (99.9% confidence in SNP call), sequencing depth > 30 reads, and a fraction of missing genotypes < 10% to avoid regions that were not consistently sequenced, the final data contained 109,227 SNPs in 237 samples. This level of stringency allows for the confidence to make meaningful inferences from the data about real biological signals, and not technical artifacts.

Initial analysis included calculating the folded (does not require inference of the reference allele) SFS of the *Chelonia mydas* population across all observed SNPs. The SFS provided context on the genetic variation and population history of green sea turtles across all regions. Rare alleles, which have low minor allele counts (k), are reflective of rare SNPs and are present only in a few individuals across the entire population. A heavily right-skewed SFS like the one observed (Fig. 6) shows an abundance of rare alleles, which may represent that a population has recently undergone rapid expansion (Keinan & Clark, 2012). Results like this may imply that previous bottlenecks or stressors are easing, and the population is rapidly growing back to healthy levels.

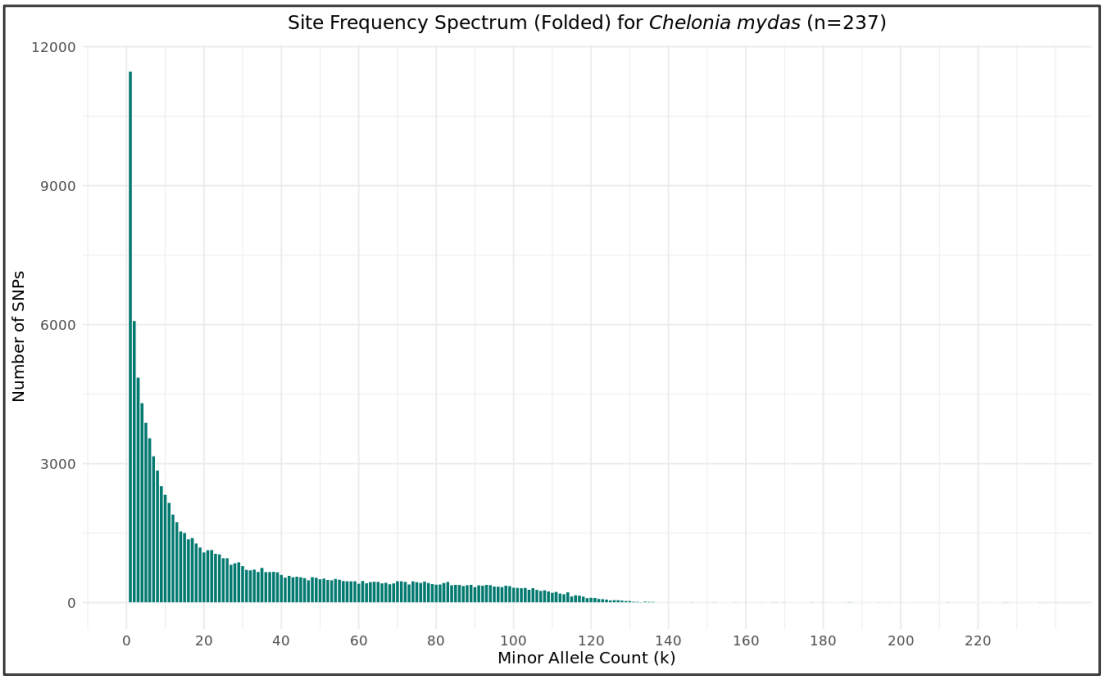


Fig. 6: Folded SFS of all 237 *Chelonia mydas* individuals. A heavily right-skewed curve shows bias for rare alleles.

Initial SFS analysis prompted the need to explore individual nesting sites to better understand the region-specific mechanisms of genetic variation and population dynamics. Using a PCA for dimensionality reduction of all the SNP variance, then connecting it with the geographic source metadata for each SNP provides a glimpse at the differences among all the different locations. The source of each sequencing sample is tied back to a specific nesting site on a beach, so for the sake of grouped analysis, the locations were binned together by beach/general location into 13 distinct locations, including captive-bred turtles as a control. The first ten principal components were extracted from the data, then from that, the first three were plotted in a three-

dimensional PCA plot (Fig. 7). This gave a glimpse into the genetic variance across the cohort, showing regions such as Netanya, Caesarea, and Ashkelon as possible outlier regions to investigate further. With Netanya and Caesarea separating in opposite directions along PC3, and Ashkelon separating along PC2. As expected, the captive-bred samples separated from the rest of the samples, indicating non-random gene flow. This can be seen separated along PC1. The remainder of the samples showed a high degree of genetic similarity to each other. Overall, the visualization shown explains 44.33% of all variance within the samples. It is possible that any remaining variance could be influenced by more subtle factors such as finer-scale population structure, stochastic genetic drift, and minor environmental pressures not captured by this study.

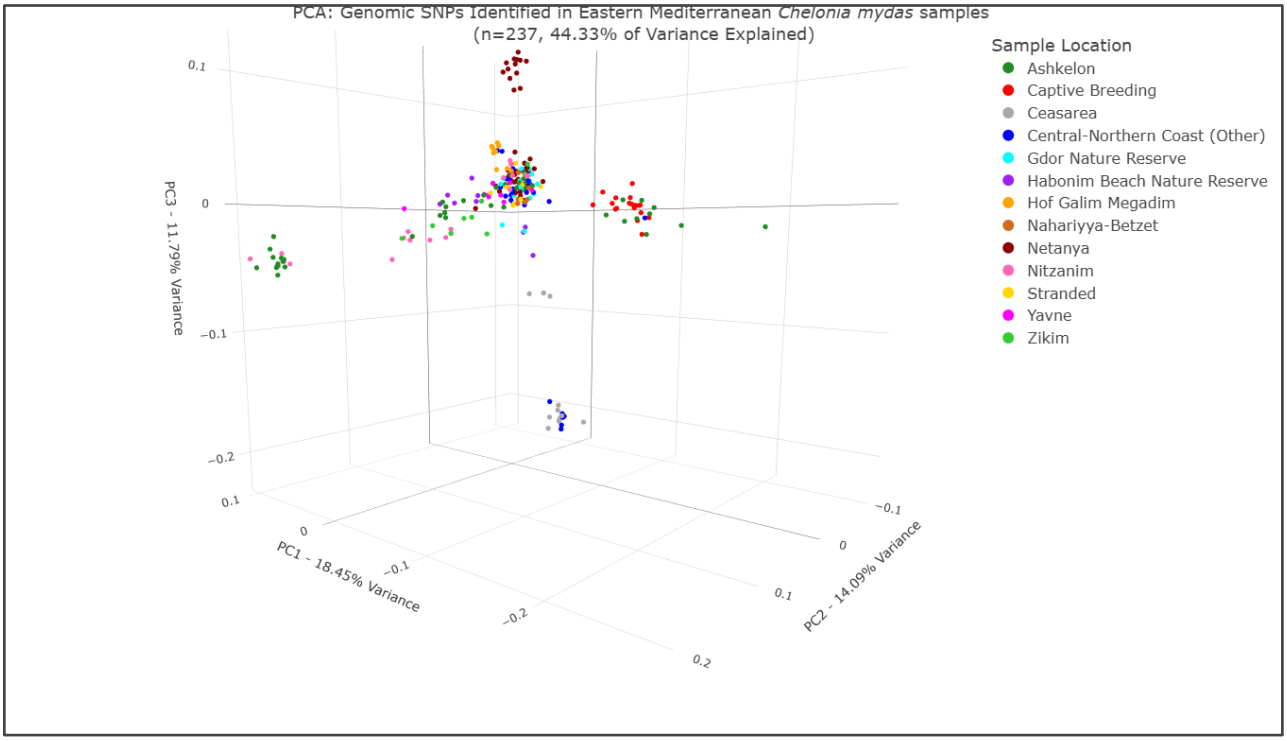


Fig. 7: Three-dimensional PCA plot of the first three extracted PCs from filtered SNP data. PC1 explains 18.45% of the variance, PC2 explains 14.09% of the variance, and PC3 explains 11.79% of the variance, for a total of 44.33% explained by the plot. Each dot represents a single sample and is colored to match its geographic location of origin.

To examine these observed patterns further and visualize finer relationships between samples based on genetic distance, an MSN was constructed from the same data (Fig. 8). Where PCA specializes in maximizing the major variations within the data, MSN is more useful for showing direct relationships between individual data points. This is calculated by connecting each sample based on the minimum genetic distance between them. This data shows similar trends to the PCA but allows for more accurate visualization of samples on an individual level. When comparing the variance shown by the MSN to a real map of the region, the interrelationships begin to make more sense. For instance, samples taken from Ashkelon and Nitzanim show close genetic relations, which is further supported by the short geographic distance between them. Insights like this allow for a better understanding of interbreeding between populations at different nesting sites and give confidence that they are less likely to be impacted by geographical barriers.

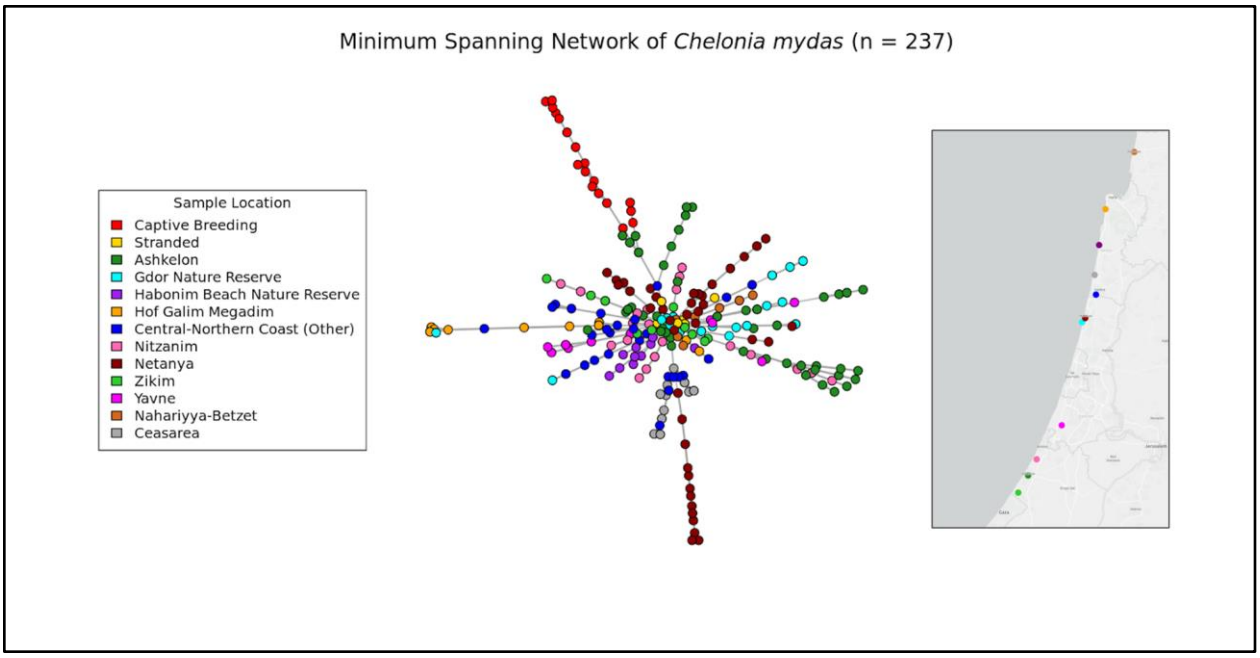


Fig. 8: MSN displaying the genetic distance between *Chelonia mydas* samples. Each dot represents an individual sample and is colored based on its geographic location. On the right is a map representing the geographic location of these nesting sites.

After using both MSN and PCA as a means of identifying outlier populations, SFS was performed based on individual regions. Five outliers were chosen for further exploration: captive-bred (Fig. 9), Ashkelon, Caesarea, Hof Galim Megadim, and Netanya (Fig. 10). The rest of the samples were also analyzed (Supp. 3). The captively bred samples show a relatively bell-shaped curve. In the wild, this may indicate a bottlenecked population, but in captivity, it may indicate the artificial nature of population management. This sample serves as a control in the analysis. Interestingly, samples from Hof Galim Megadim hint at a minor peak at intermediate allele frequencies (bins 3-5) through the curve. This could be indicative of slightly reduced gene flow and diversity when compared to the full cohort, or even small population bottlenecking. Samples from Caesarea exhibit right-skew, but to a lesser degree than the full dataset. This may be indicative of a stable and slowly growing population. Similarly to the full cohort, samples from Netanya are heavily right-skewed, which would suggest recent population expansion. However, even though Netanya's SFS curve followed the trend of the full study, its samples tended to cluster separately from the rest, insinuating that there may be other genetic or demographic factors influencing the separation. When looking at samples from Ashkelon, it follows a similar right-skewed trend as seen in Netanya. However, there is a slight shoulder in the graph in bins 2-5, which may indicate some small population substructure. Taking into account the genetic and geographic closeness between Ashkelon and Nitzanim (Fig. 7), there is a possibility that this trend is caused by interbreeding between the two regions. Altogether, this site-specific SFS analysis managed to highlight the genetic diversity between the nesting sites and supports the presence of region-specific population histories and genetic pressures.

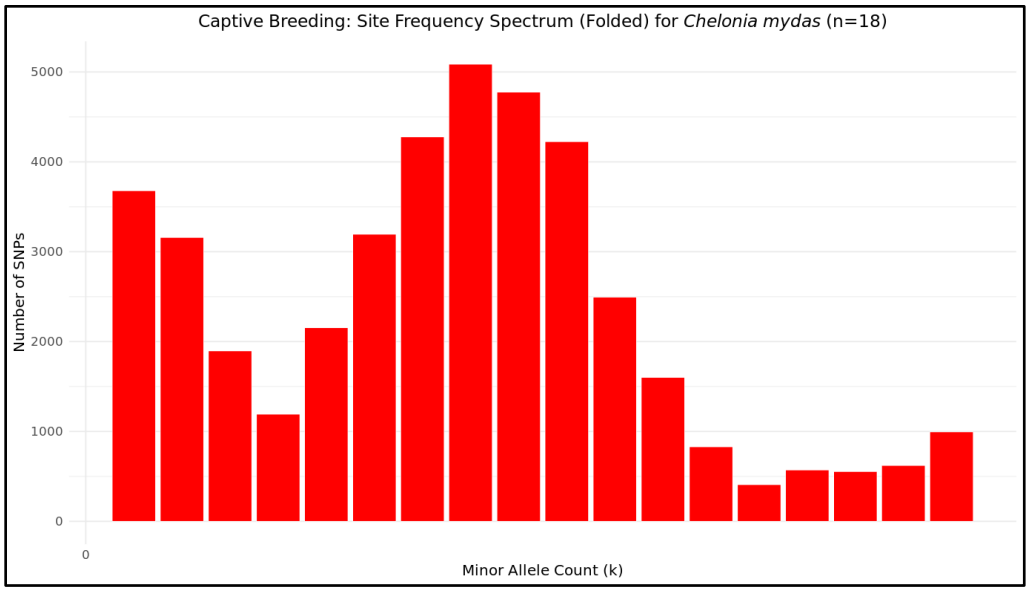


Fig. 9: Folded SFS for captive-bred samples. This sample appeared as an outlier in previous analysis and serves as a reference group for atypical population growth since captive breeding does not perfectly mimic true breeding patterns.

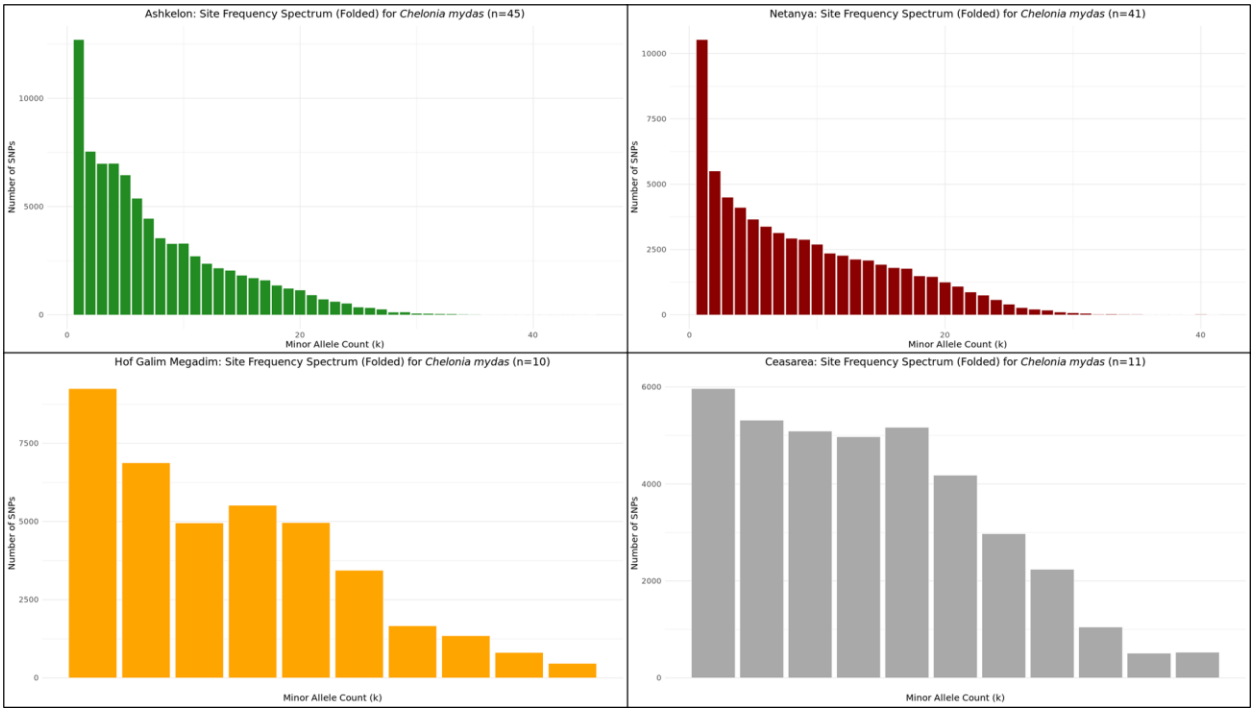


Fig. 10: Folded SFS for Ashkelon (top left), Caesarea (top right), Hof Galim Megadim (bottom left), and Netanya (bottom right). These samples appeared to be outliers from prior analyses and showed different SFS patterns than the rest of the cohort.

To further study the driving environmental factors causing regional genetic differentiation, mutational signature analysis was performed. Netanya was a primary candidate for mutational signature extraction, as its genetic separation from a population genomics level suggested perhaps a unique mutational process was occurring there. The individual nucleotide changes were used to group trinucleotide contexts with shared mutations. Adenine (A) and guanine (G) bases were converted to complementary cytosine (C) and thymine (T) bases. A large number of C>T and T>C

mutations were observed (Fig. 11). All other mutations fell below 1% of the total mutational burden of the region. Numerous studies have shown mutations following this trend are indicative of ultraviolet (UV) light exposure, and oxidative stress in response to environmental pollutants (Pfeifer et al., 2005; Valavanidis et al., 2013; Cadet & Wagner, 2013), and may be linked to previously characterized pressures such as ocean acidification. Due to the environmental context this signature outlines, all other regions were compared and showed very similar results (Supp. 4). The shared prevalence and intensity of these mutations across all regions aid in suggesting a shared mutagenic history amongst all the populations, even captively bred samples since these mutations have been shown to be heritable.

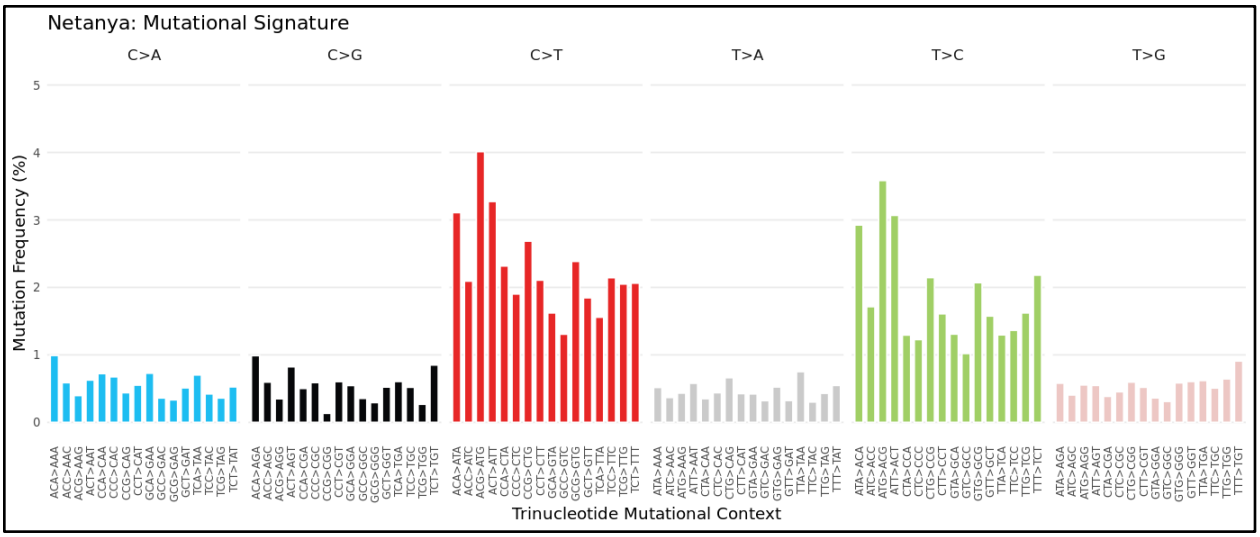


Fig. 11: Mutational signature of *Chelonia mydas* samples collected from Netanya. An overwhelming bias of observed SNPs fall into C>T or T>C transitions and seem to be elevated further within trinucleotide contexts beginning with an A.

Although overall regional mutational signatures looked similar, it was important to check for any unique mutational processes native to certain regions. This was accomplished by using NMF for mutational signature decomposition. Four signatures were selected after testing a range of signatures, and was determined to be the best fit to the data. In future studies, more signatures can be extracted if there is more data available. After extracting the four signatures from all regions, it was observed that the newly identified (de novo) signatures extracted from our dataset did not show overwhelmingly large amounts of variability within regions (Fig. 12). No signatures were shown to be exclusive to only one region, but there were minor changes in the proportion of mutations contributing to each signature per region. These variations may underlie real environmental events such as localized pollution causing slightly different patterns of mutations amongst regions. The presence of the same mutational signature components at comparable levels in captively bred samples also helps to confirm that these observed trends have been acquired in the past and inherited by recent generations.

Overall, genetic analysis of the green sea turtle in the eastern Mediterranean has shown similar mutagenic pressures on the genome across a range of regions. Despite this, some locations such as Netanya, Hof Galim Megadim, Ashkelon, and Caesarea exhibit unique trends in their population genetics. These results will allow researchers to make inferences about the genomic health of the green sea turtle in the eastern Mediterranean and devise targeted conservation

strategies to ensure the future health of the species. Future studies that contain more samples and higher amounts of overall genome representation will aid in further understanding the genetic diversity and health of the region.

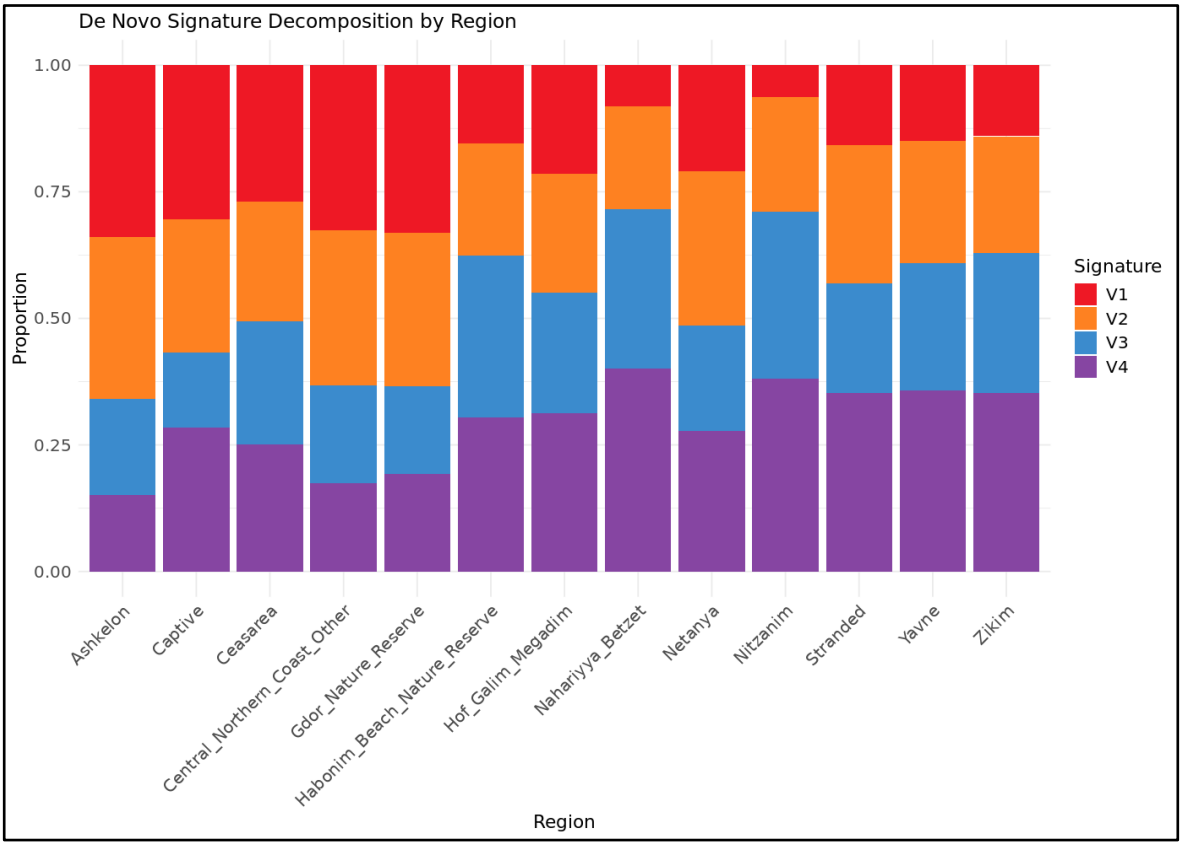


Fig. 12: NMF was used to extract de novo signatures across all regions. Each of the de novo signatures are labeled as V#. The data was normalized to one to focus on the proportional makeup of each region's mutational signature.

Discussion

This study provides genomic and environmental evidence that anthropogenic pressures have measurably altered the ocean ecosystem in the eastern Mediterranean, and has negatively impacted the genetic health of the green sea turtle. From 2000 to 2025, the Mediterranean region suffered from ecological disasters caused by anthropogenic influences. In 2012, the northern Israeli coastal town of Nahariya had massive asbestos contamination, which Ronny Linder, a writer for Haaretz, likened to Chernobyl (Linder, 2012). In 2023, a tar spill was reported at the Gdor Sea Reserve along the coastline (Lidor, 2023). In 2021, a tanker carrying 90,000 tons of crude oil was dumped into the Mediterranean Sea, further hurting the marine ecosystem (JNS Staff, 2021). Another incident in 2016 resulted in roughly 15,000 tons of oil spilling into the eastern Mediterranean Sea (Glanz, 2006). It is essential to note the ecological impacts of oil and tar spills and how often this occurs in the eastern Mediterranean region. The implications of this are that coastlines that have been the natural nesting locations for green sea turtles in the area have been under constant duress and have put pressure on the species and other marine life over the past 25 years. Along with ecological disasters, there have been noticeable biogeochemical trends that have

had negative impacts on marine life, including pH (ocean acidification), declining chlorophyll levels, and a decrease in phytoplankton carbon biomass.

The data collected has provided invaluable insight that the eastern region of the Mediterranean is acidifying, potentially due to increased anthropogenic CO_2 emissions across the globe. The National Oceanic and Atmospheric Administration (NOAA) explains that we see acidification due to an increase in carbon dioxide in the sea, which has been linked to the global rise in atmospheric CO_2 resulting from increased human activity. Because roughly 30% of atmospheric CO_2 is absorbed by the ocean, it would be expected that CO_2 accumulation would rise (National Oceanic and Atmospheric Administration, n.d.). Additionally, ocean acidification poses significant threats to green sea turtles primarily through its impact on their primary food sources, such as seagrass and algae. Ocean acidification can hinder the growth and nutritional quality of these plants, potentially leading to reduced food availability. Ocean acidification can also degrade critical habitats such as coral reefs and seagrass meadows, which are essential for food production and shelter. It is essential that while there are strong correlations in the data noted, there are many factors that can influence the biogeochemistry of the ocean that we did not have data related to in this study, such as industrial waste, waste runoff, etc. Additionally, exposure to pollutants and habitat destruction has been linked to congenital malformations that threaten hatchling survival and reduced nesting sites (Martín-del-Campo et al., 2021). Both the ecological disasters and biogeochemical shifts observed threaten the stability of this marine ecosystem. Anthropogenic pressures threaten not only the viability of green sea turtles, but also the overall health of the ecosystem.

Ongoing conservation strategies in the region have been beneficial in helping sea turtle populations begin to rebound in the eastern Mediterranean. Genomic analysis reveals recent population expansions at nesting sites like Netanya and Ashkelon, indicated by a right-skewed frequency spectrum. This suggests that current conservation efforts may be supporting rapid recovery in some regions. However, the genetic heterogeneity of certain populations also underscores the importance of site-specific management strategies to maintain genetic diversity among green sea turtle populations to ensure long-term resilience. Natural disasters such as oil spills have detrimentally affected the overall marine ecosystem. Regions such as Netanya, Caesarea, and Hof Galim Megadim that have undergone environmental stress in recent years show genetically distinct population histories from each other. These population responses range from signs of genetic drift, population bottlenecks, and recent expansion in recovering regions, as shown across site-specific SFS analysis (Fig. 9). Although the genetic variability of populations is beginning to recover, mutagenic signatures caused by climate change and pollutant exposures have caused heritable modifications in the green sea turtle genome across all regions. Signs of pollutant and UV exposure were detected in mutational signature analysis by high levels of C>T and T>C transition mutations linked to oxidative stress (Fig. 11). This is concordant with the previously mentioned widespread environmental disasters that introduced hazardous pollutants directly into the green sea turtles' ecosystem, as well as ocean acidification causing increased oxidative stress in the turtles. When analyzing decomposed signatures, the Nahariyya-Betzet region had the highest proportion of de novo signature V4 compared to all other regions. Given the asbestos contamination in Nahariyya back in 2012, there is a chance signature V4 could be characterizing mutations caused by that event. However, due to a small sample size of only seven samples from that region, it is difficult to draw a definitive conclusion.

Another environmental trend that was observed is increasing sea surface temperature, which pose significant threats to green sea turtles in the Mediterranean, specifically when looking at TSD. For example, Laloë, Schofield, and Hays (2024) conducted a study that examined the sex ratio data and patterns of climate change over multiple decades for 64 nesting sites and found that temperatures increased at 40 sites and female-skewed hatching ratios occurred at 57 of the sites, 17 of which were skewed greater than 90% female. Although their study did not find a direct relationship between increasing temperatures and female sex ratios (suggesting that sex ratio skews are not simply a consequence of warming temperatures, but a combination of other factors), some sites were observed to have both a high female sex ratio and high warming temperatures. This suggests that these sites may be the first to require male-boosting interventions and continued monitoring to help identify when high incubation temperatures threaten population viability (Laloë, Schofield, & Hays, 2024). Blechschmidt, Wittmann, and Blüml (2020) found similar results in their study on TSD and increasing temperatures due to climate change. Blechschmidt et. al (2020) studied a population of green sea turtles at the northern Great Barrier Reef that showed a highly female-skewed sex ratio (overall sex ratio of 80% female, with 99% of non-adult turtles being female), and that the shortage of males might eventually cause population extinction without rapid conservation efforts. The researchers observed evolvable traits in their study (pivotal temperature, nest depth, and shading) and concluded that among the evolvable traits, nest depth was likely the most successful trait to target in species conservation, and that conservation strategies should promote population survival through trait evolution.

Beyond direct reproductive impacts, broader ecosystem indicators also highlight the vulnerabilities of the eastern Mediterranean sea environment. Signs of increased UV exposure in green sea turtles determined by mutational signature analysis, alongside biogeochemical evidence of ocean acidification, can point to severe downstream effects across the entire food web of the ecosystem. While the evidence of increased UV exposure was detected while studying just green sea turtle genomic samples, this exposure likely threatens most other organisms in the region. The combination of these two pressures have been shown to decrease the level of free calcium carbonate ($CaCO_3$) in the water, and thus hinders the calcification of phytoplankton, preventing them from forming shells (Peng et al., 2022). Their shells are necessary for staying buoyant and floating in their optimal ocean levels. The loss of this shell can decrease their survival, reproduction, and metabolic rates (Center for Biological Diversity, n.d.). These changes can cascade throughout the entire food web, as phytoplankton play an important role as a primary food source, and also play a major role in carbon fixation. Overall, this is indicative of a loss of life-sustaining resources in the Mediterranean Sea that are necessary for the survival of higher trophic levels, including the green sea turtle.

The cumulative impacts of anthropogenic activity over the past 25 years, including oil spills, rising sea surface temperatures, and ocean acidification, have significantly harmed the eastern Mediterranean marine ecosystem. Genomic evidence from green sea turtles reveals adaptations to environmental degradation showcasing species vulnerability to pollution and climate change. The effects of these stressors extend beyond green sea turtles, as phytoplankton and other keystone species experience reduced survival, reproduction, and metabolic function, threatening the entire food web. Encouraging signs of population recovery in some areas underscore the effectiveness of targeted conservation strategies. This can be primarily achieved through habitat cleanup, continuous monitoring of genetic health, and implementing protocols to contain environmental disasters like oil spills better before they cause widespread damage. These

insights play a key role in understanding conservation needs in the region and open the door for further research with more samples and sampling sites. Continued conservation efforts are critical to preserving biodiversity and ecosystem stability in the eastern Mediterranean. Ultimately, this research provides a roadmap for future studies and supports the dire need for proactive environmental conservation in the eastern Mediterranean to ensure the long-term survival of the green sea turtle.

References

Adler, M. (2022, August 19). *Madler/pigz: A parallel implementation of GZIP for modern multi-processor, multi-core machines.* GitHub. https://github.com/madler/pigz

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. *Babraham Bioinformatics.* https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Blechschmidt, J., Wittmann, M. J., & Blüml, C. (2020). Climate change and green sea turtle sex ratio—Preventing possible extinction. Animals, 10(6), 1045. https://doi.org/10.3390/ani10061045

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bonfield, J., Davies, R., Marshall, J., & Danecek, P. (2024, September 12). *Samtools/htslib: C library for high-throughput sequencing data formats*. GitHub. https://github.com/samtools/htslib

Cadet, J., & Wagner, J. R. (2013). DNA base damage by reactive oxygen species, oxidizing agents, and UV radiation. *Cold Spring Harbor Perspectives in Biology*, *5*(2). https://doi.org/10.1101/cshperspect.a012559

Center for Biological Diversity. (n.d.). Endangered oceans: Frequently asked questions. https://www.biologicaldiversity.org/campaigns/endangered_oceans/faq.html

Cole, J. J., & Prairie, Y. T. (2014). Dissolved CO_2 in freshwater systems☆. In *Reference module in Earth systems and environmental sciences*. Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.09399-4

Copernicus. (n.d.). *Copernicus: Europe's eyes on Earth*. Retrieved February 25, 2025, from https://www.copernicus.eu/en

Copernicus Marine Service. (Accessed 2025). *Mediterranean Sea Biogeochemistry Reanalysis product* (MEDSEA_MULTIYEAR_BGC_006_008). https://marine.copernicus.eu

Copernicus Marine Service. (Accessed 2025). *Mediterranean Sea - High Resolution L4 Sea Surface Temperature Reprocessed dataset (cmems_SST_MED_SST_L4_REP_OBSERVATIONS_010_021)*. https://marine.copernicus.eu

Cosmic. (2020, April 7). *Mutational signatures*. COSMIC | Mutational Signatures. https://cancer.sanger.ac.uk/signatures/

Csárdi G, Nepusz T, Traag V, Horvát Sz, Zanini F, Noom D, Müller K, Schoch D, Salmon M (2025). _igraph: Network Analysis and Visualization in R_. doi:10.5281/zenodo.7682609 <https://doi.org/10.5281/zenodo.7682609>, R package version 2.1.4.9035, <https://CRAN.R-project.org/package=igraph>.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

Dvoretsky, V. G., Vodopianova, V. V., & Bulavina, A. S. (2023). Effects of Climate Change on Chlorophyll a in the Barents Sea: A Long-Term Assessment. Biology, 12(1), 119. https://doi.org/10.3390/biology12010119

Fox J, Weisberg S (2019). *An R Companion to Applied Regression, Third edition.* Sage, Thousand Oaks CA. https://www.john-fox.ca/Companion/.

Fox, J., Weisberg, S., & Price, B. (2023). *car: Companion to Applied Regression* (Version 3.1-2) [R package]. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/package=car

Gaujoux, R., & Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*, *11*(1). https://doi.org/10.1186/1471-2105-11-367

Glanz, J. (2006, August 11). *Spills of war*. WIRED. https://www.wired.com/2006/08/spills-of-war/

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An r package to facilitate analysis of snp data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, *18*(3), 691–699. https://doi.org/10.1111/1755-0998.12745

Harrell, F. E., Jr., & Dupont, C. (2024). *Hmisc: Harrell Miscellaneous* (Version 5.1-1) [R package]. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/web/packages/Hmisc/index.html

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R* (2nd ed.). Springer. https://www.statlearning.com/

Jennewein, D. M., Lee, J., Kurtz, C., Dizon, W., Shaeffer, I., Chapman, A., Chiquete, A., Burks, J., Carlson, A., Mason, N., Kobawala, A., Jagadeesan, T., Basani, P. B., Battelle, T., Belshe, R., McCaffrey, D., Brazil, M., Inumella, C., Kuznia, K., … Yalim, J. (2023). The Sol Supercomputer at Arizona State University. *Practice and Experience in Advanced Research Computing*, 296–301. https://doi.org/10.1145/3569951.3597573

JNS Staff. (2021, February 22). *Toxic tar spill is one of the most serious ecological disasters in Israel's history*. Jewish News Syndicate. https://www.jns.org/toxic-tar-spill-is-one-of-the-most-serious-ecological-disasters-in-israels-history/

Jombart, T. (2008a). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*. https://doi.org/10.7717/peerj.281

Kassambara, A. (2019). *ggcorrplot: Visualization of a Correlation Matrix using 'ggplot2'*. R package version 0.1.3. https://CRAN.R-project.org/package=ggcorrplot

Keinan, A., & Clark, A. G. (2012). Recent explosive human population growth has resulted in an

excess of rare genetic variants. *Science*, *336*(6082), 740–743. https://doi.org/10.1126/science.1217283

Knaus, B. J., & Grünwald, N. J. (2016a). VcfR: A package to manipulate and visualize VCF format data in R. *Molecular Ecology Resources*. https://doi.org/10.1101/041277

Laloë, J. O., Schofield, G., & Hays, G. C. (2024). Climate warming and sea turtle sex ratios across the globe. Global Change Biology, 30(1), e17004. https://doi.org/10.1111/gcb.17004

Liao, H., Yang, Z., Dou, Z., Sun, F., Kou, S., Zhang, Z., Huang, X., & Bao, Z. (2019). Impact of Ocean Acidification on the Energy Metabolism and Antioxidant Responses of the Yesso Scallop (Patinopecten yessoensis). Frontiers in physiology, 9, 1967. https://doi.org/10.3389/fphys.2018.01967

Linder, R. (2012, March 26). *Nahariya is like Chernobyl*. Haaretz. https://www.haaretz.com/israel-news/business/2012-03-26/ty-article/nahariya-is-like-chernobyl/0000017f-e4d1-d568-ad7f-f7fb33400000

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lidor, C. (2023, July 12). *Tar contaminates Herzliya beach: 'Catastrophe for nature'*. The Jerusalem Post. https://www.jpost.com/environment-and-climate-change/article-749865

Merlivat, L., Boutin, J., Antoine, D., Beaumont, L., Golbol, M., & Vellucci, V. (2018). Increase of dissolved inorganic carbon and decrease in pH in near-surface waters in the Mediterranean Sea during the past two decades. *Biogeosciences*, 15(18), 5653–5662. https://doi.org/10.5194/bg-15-5653-2018

McVean, G. (2009). A genealogical interpretation of Principal Components Analysis. *PLoS Genetics*, *5*(10). https://doi.org/10.1371/journal.pgen.1000686

Middelburg, J. J., Boyd, P. W., & Ilyina, T. (2020). Ocean alkalinity, buffering and biogeochemical processes. *Reviews of Geophysics, 58*(3), e2019RG000681. https://doi.org/10.1029/2019RG000681

Montana State University Extension. (n.d.). *Understanding alkalinity, pH, and total dissolved solids*. MSU Extension Water Quality Program. Retrieved April 22, 2025, from https://waterquality.montana.edu/well-ed/interpreting_results/fs_alkalinity_ph_tds.html

National Center for Biotechnology Information. (2021, July 15). Chelonia mydas genome assembly rchemyd1.pri.v2. National Library of Medicine. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_015237465.2/

National Center for Biotechnology Information. (2024, August 8). Chelonia mydas (ID 1145648). National Library of Medicine. https://www.ncbi.nlm.nih.gov/bioproject/1145648

National Oceanic and Atmospheric Administration. (n.d.). *Ocean acidification*. NOAA. https://www.noaa.gov/education/resource-collections/ocean-coasts/ocean-acidification

Pebesma E, Bivand R (2023). *Spatial Data Science: With applications in R*. Chapman and Hall/CRC, London. doi:10.1201/9780429459016, https://r-spatial.org/book/.

Peng, J., Wan, J., Zhang, J., Overmans, S., Xiao, M., Ye, M., Dai, X., Zhao, J., Gao, K., & Xia, J. (2022). *Additive impacts of ocean acidification and ambient ultraviolet radiation threaten calcifying marine primary producers*. *Science of The Total Environment, 818*, 151782. https://doi.org/10.1016/j.scitotenv.2021.151782

Peng, J., Hutchins, D. A., & Gao, K. (2020). The impacts of ocean acidification on marine food quality and its potential food chain consequences. *Frontiers in Marine Science, 7*, Article 543979. https://doi.org/10.3389/fmars.2020.543979

Pfeifer, G. P., You, Y.-H., & Besaratinia, A. (2005). Mutations induced by ultraviolet light. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *571*(1–2), 19–31. https://doi.org/10.1016/j.mrfmmm.2004.06.057

Plotly Technologies Inc. Collaborative data science. Montréal, QC, 2015. https://plot.ly.

Python Software Foundation. (2022). *Python Language Reference, version 3.11.0*. Available at: https://www.python.org/downloads/release/python-3110/

R Core Team (2025). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Rasmussen, M. S., Garcia-Erill, G., Korneliussen, T. S., Wiuf, C., & Albrechtsen, A. (2022). Estimation of site frequency spectra from low-coverage sequencing data using stochastic EM reduces overfitting, runtime, and memory usage. Genetics, 222(4), iyac148. https://doi.org/10.1093/genetics/iyac148

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

SEE Turtles. (n.d.). *Sea turtle diet*. Retrieved April 24, 2025, from https://www.seeturtles.org/sea-turtle-diet

Shirley, M. D., Ma, Z., Pedersen, B. S., & Wheelan, S. J. (2015). Efficient “Pythonic” access to Fasta files using pyfaidx. *PeerJ Preprints*. https://doi.org/10.7287/peerj.preprints.970v1

Southwood, A. L., Andrews, R. D., Lutcavage, M. E., Paladino, F. V., West, N. H., George, R. H., & Spotila, J. R. (2009). Heart rates and dive behavior of leatherback sea turtles (Dermochelys coriacea) during the internesting interval. *Journal of Experimental Marine Biology and Ecology*, *374*(1), 26–31. https://doi.org/10.1016/j.jembe.2009.03.005

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., ... & Sabine, C. L. (2009). Climatological mean and decadal change in surface ocean pCO_2, and net sea–air CO_2 flux over the global oceans. *Deep Sea Research Part II: Topical Studies in Oceanography*, 56(8-10), 554-577. https://doi.org/10.1016/j.dsr2.2008.12.009

The pandas development team. (2024). pandas-dev/pandas: Pandas (v2.2.3). Zenodo. https://doi.org/10.5281/zenodo.13819579

Turi, G., Landschützer, P., Hauck, J., & Rödenbeck, C. (2023). Temperature-driven changes in oceanic pCO_2: Evidence for the thermodynamic control of surface CO_2 dynamics. *Geophysical Research Letters*, 50(23), e2023GL107030. https://doi.org/10.1029/2023GL107030

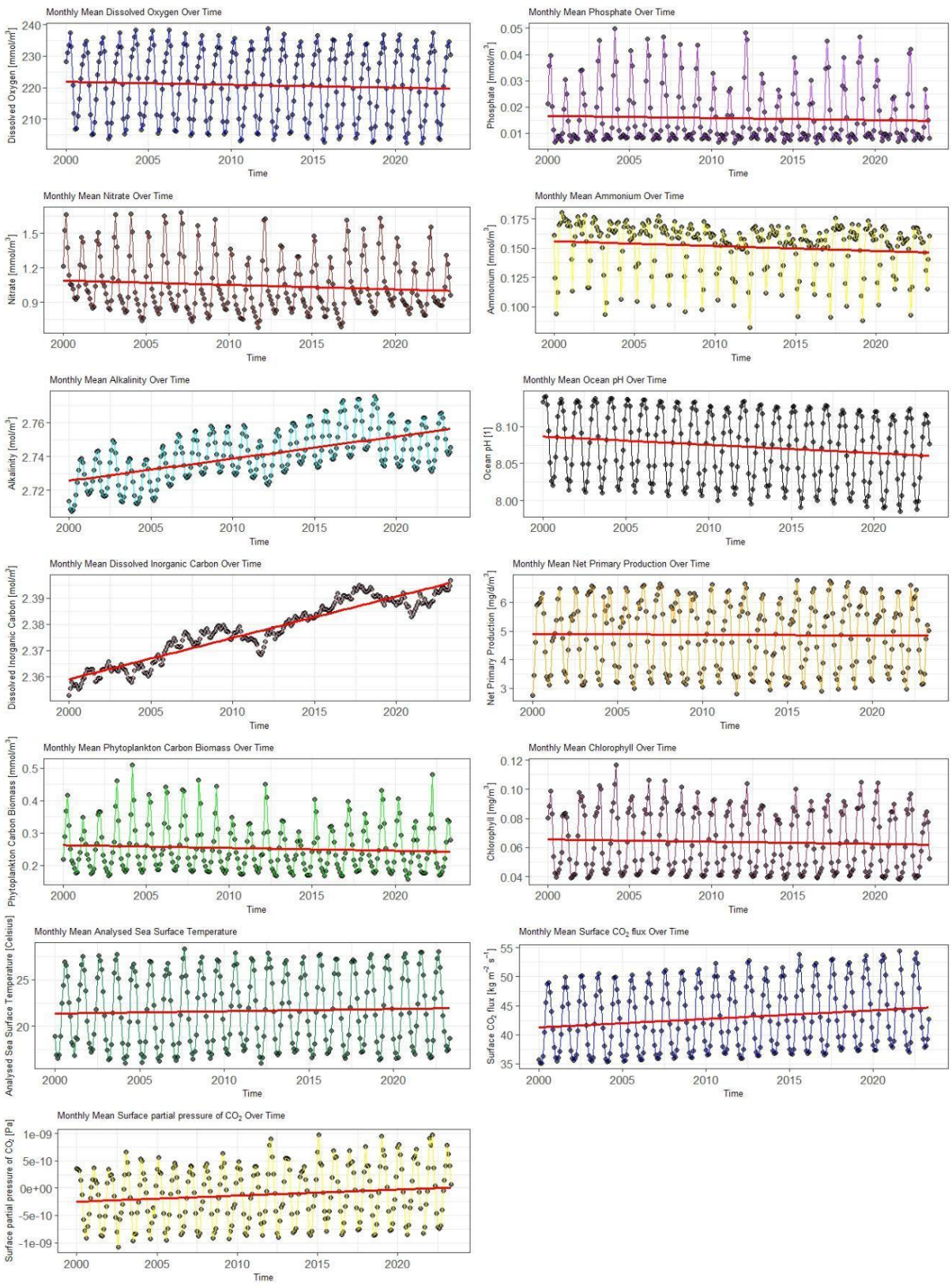
UNEP/MAP-RAC/SPA. (2007). Action plan for the conservation of marine turtles in the Mediterranean. Regional Activity Centre for Specially Protected Areas. https://rac-spa.org/sites/default/files/action_plans/p_a_tortues_en.pdf

Valavanidis, A., Vlachogianni, T., Fiotakis, K., & Loridas, S. (2013). Pulmonary oxidative stress, inflammation and cancer: Respirable particulate matter, fibrous dusts and ozone as major causes of lung carcinogenesis through reactive oxygen species mechanisms. *International Journal of Environmental Research and Public Health*, *10*(9), 3886–3907. https://doi.org/10.3390/ijerph10093886

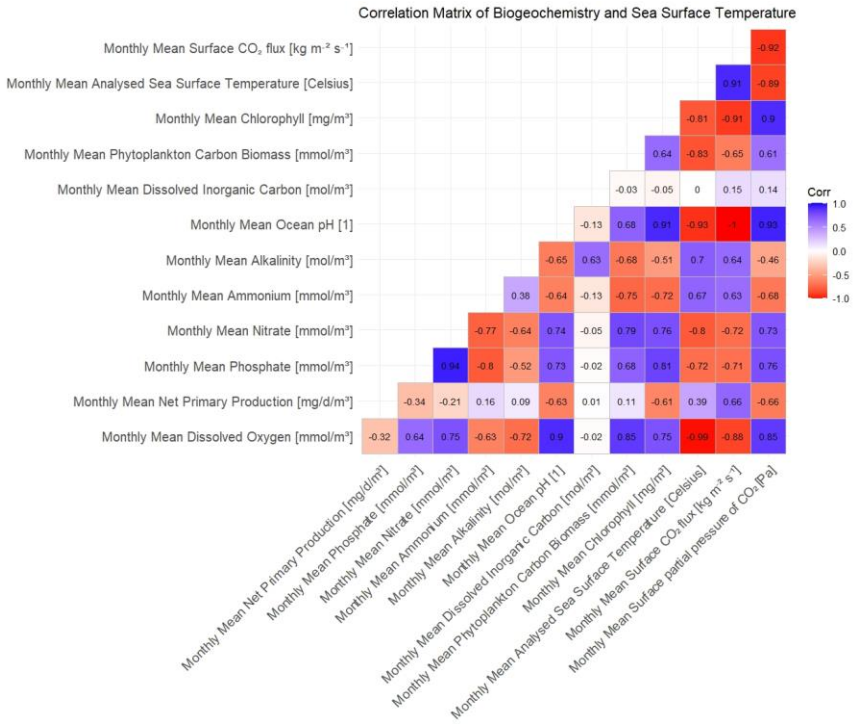
Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, **4**(43), 1686. doi:10.21105/joss.01686.

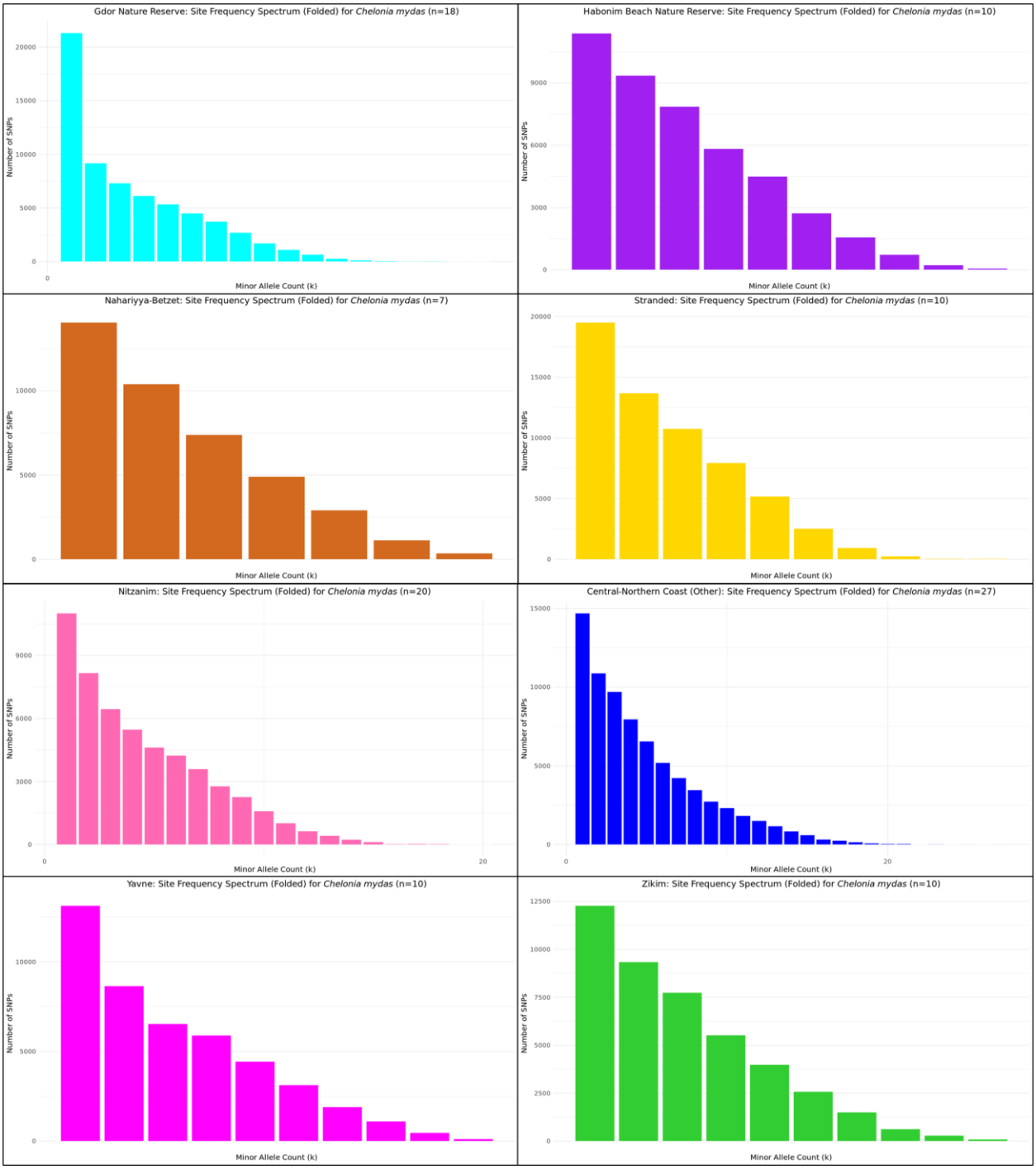
Supplemental Figures



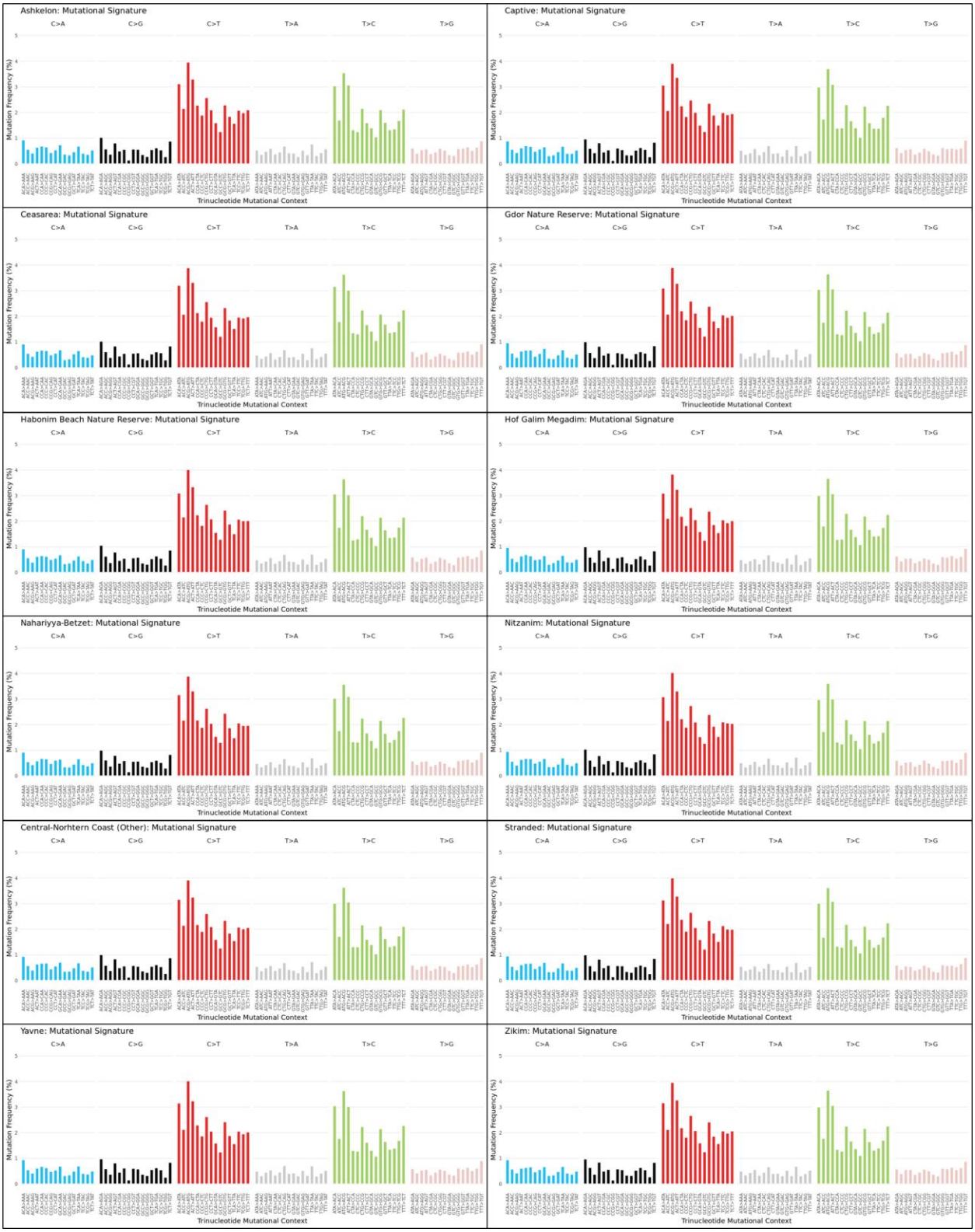
Supp. 1: Biogeochemistry and sea surface temperature with linear regression trend lines



Supp. 2: Correlation matrix of biogeochemistry and sea surface temperature.



Supp. 3: Folded SFS for all remaining regions excluded from Fig. 10. From left to right moving down the figure: Gdor Nature Reserve, Habonim Beach Nature Reserve, Nahariyya-Betzet, Stranded, Nitzanim, Central-Northern Coast (Other), Yavne, Zikim.



Supp. 4: Mutational signature of all regions not included in Fig. 11. From left to right moving down the figure: Ashkelon, Captive, Ceasarea, Gdor Nature Reserve, Habonim Beach Nature Reserve, Hof Galim Megadim, Nahariyya-Betzet, Nitzanim, Central-Northern Coast (Other), Stranded, Yavne, Zikim.