american physiological society
PHYSIOLOGICAL GENOMICS

Physiol Genomics 54: 177–185, 2022.
First published April 20, 2022; doi:10.1152/physiolgenomics.00002.2022

RESEARCH ARTICLE

Gut microbiota of wild fish as reporters of compromised aquatic environments sleuthed through machine learning

John W. Turner, Jr.,* Xi Cheng,* Nilanjana Saferin,* Ji-Youn Yeo, Tao Yang, and Bina Joe
Department of Physiology and Pharmacology, University of Toledo College of Medicine and Life Sciences, Toledo, Ohio

Abstract

Human-generated negative impacts on aquatic environments are rising. Despite wild fish playing a key role in aquatic ecologies and comprising a major global food source, physiological consequences of these impacts on them are poorly understood. Here we address the issue through the lens of interrelationship between wild fish and their gut microbiota, hypothesizing that fish microbiota are reporters of the aquatic environs. Two geographically separate teleost wild-fish species were studied (Lake Erie, Ohio, and Caribbean Sea, US Virgin Islands). At each geolocation, fresh fecal samples were collected from fish in areas of presence or absence of known aquatic compromise. Gut microbiota was assessed via microbial 16S-rRNA gene sequencing and represents the first complete report for both fish species. Despite marked differences in geography, climate, water type, fish species, habitat, diet, and gut microbial compositions, the pattern of shifts in microbiota shared by both fish species was nearly identical due to aquatic compromise. Next, these data were subjected to machine learning (ML) to examine reliability of using the fish-gut microbiota as an ecomarker for anthropogenic aquatic impacts. Independent of geolocation, ML predicted aquatic compromise with remarkable accuracy (>90%). Overall, this study represents the first multispecies stress-related comparison of its kind and demonstrates the potential of artificial intelligence via ML as a tool for biomonitoring and detecting compromised aquatic conditions.

environment; fish; machine learning; microbiota; stress

INTRODUCTION

The scientific evidence for ongoing global climate change is extensive, and aquatic habitats and species therein are experiencing chronic compromises due to these alterations (1–3). The driving forces are both natural and anthropogenic, with the latter influencing the former and locally compounding the global impacts (4–6). Moreover, for 21 countries, fish is the primary source of protein and ~3 billion people rely on wild-caught and aquaculture-raised seafood for 20% or more of their animal protein (7, 8). It has thus become essential to biomonitor aquatic environments and species therein to assure a viable future for these valuable and sustaining resources.

Historically, evidence of environmental derangement and deterioration was based on loss of habitat, decreased species diversity, and reductions in within-species numbers (9). Although the hormone cortisol has been employed as a more precise and earlier-warning biomarker of chronic stress in fish (10, 11), this biomarker requires an extended (weeks) sampling period to rule out influences of acute stress events.

Eukaryotes including fish are holobionts, which comprise microbiota residing within the macroorganism as commensals. Recent technological advances in studies of gut microbiota in humans and other mammals have enabled the assessment of significant rearrangements in microbial composition or "dysbiosis" associated with stress and disease (12–15). However, such microbial studies in fish have been primarily limited to the field of commercial aquaculture (16, 17).

Artificial intelligence (AI) is a branch of computer science to imitate human thinking and learning competence, and enhance human-machine interfaces, such as digital virtual assistants (e.g., Siri) and autonomous driving technology. Specifically, machine learning (ML), one of the most popular concepts in AI, has been widely used to comprehend, learn, and recognize patterns within data using various complex algorithms. Recently, we have published seminal articles demonstrating the power of AI via fecal microbiome-based ML models to diagnose various human diseases (18, 19). The current study, albeit based on fish, is on this background and the scientific premise that gut bacteria are indicators of surrounding host environment. As fish are experiencing worsening aquatic environments worldwide due to various local and anthropogenic factors, we first hypothesized that fish microbiota can serve as reporters of the environment surrounding the fish. To test this hypothesis, we selected two geographically separated locations on the planet representing both saltwater and freshwater habitats (Fig. 1*A*). Gut microbiota of the wild fish from these locations was profiled and cataloged. Next, within each of these locations, we identified anthropogenically induced

*J. W. Turner, Jr., X. Cheng, and N. Saferin contributed equally to this work.
Correspondence: J. W. Turner, Jr. (John.Turner@utoledo.edu); B. Joe (bina.joe@utoledo.edu).
Submitted 5 January 2022 / Revised 13 April 2022 / Accepted 13 April 2022

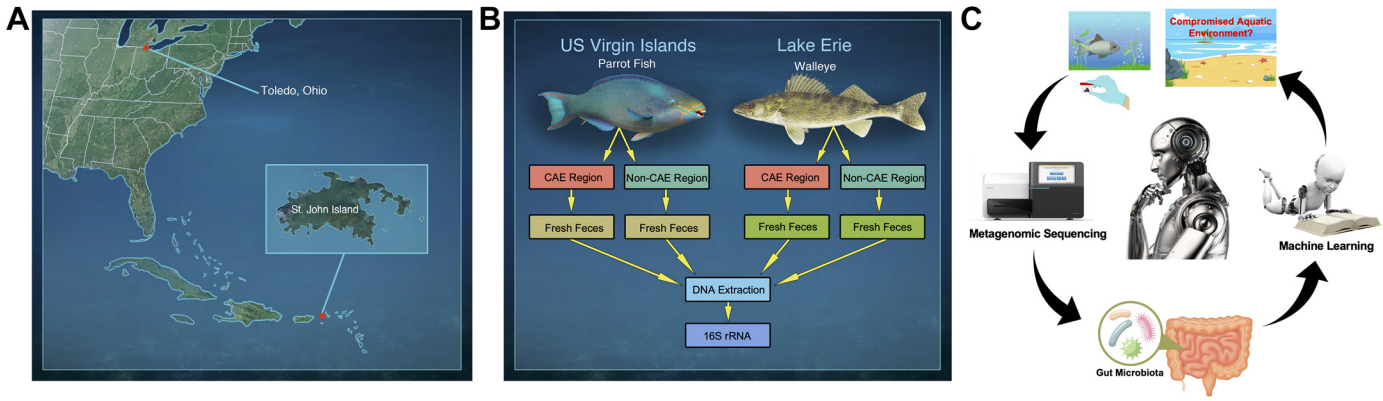

Figure 1. A novel fish gut microbiome-based machine learning approach for detecting compromised aquatic environment. *A*: study sites. *B*: wild fish species. *C*: study workflow.

compromised aquatic environments (henceforth termed CAE) and collected wild-fish samples for comparing their gut microbiota with those of samples from respective uncompromised (non-CAE) sites (Fig. 1*B*). Remarkably significant differences were observed between fish gut microbiota of non-CAE and CAE areas. Encouraged by these observations, we further expanded our study to couple data we obtained from our microbial profiling with ML as a possible tool to detect CAE (Fig. 1*C*).

Overall, this study *1*) reports gut microbial composition from saltwater wild fish (parrotfish); *2*) reports gut microbial composition from freshwater wild fish (walleye); *3*) compares and contrasts gut microbiota from non-CAE and CAE saltwater environments; *4*) compares and contrasts gut microbiota from uncompromised and compromised freshwater environments non-CAE and CAE; *5*) identifies the commonality of gut microbiota shifts of both fish species in their respective non-CAE and CAE; and *6*) applies and develops ML models for assessment of water quality using fish gut microbiota as reporters.

METHODS

Two studies differing in a number of characteristics (e.g., geographic location, study site, habitat, fish species, and collection/handling methods) are reported. The methods for these are therefore presented separately as *Study Site 1* and *Study Site 2*. However, the DNA extraction, 16S-rRNA sequencing and analysis, and machine learning procedures were identical for both studies. This study was performed under approved university IACUC 100679.

Study Site 1: Stoplight Parrotfish (*Sparisoma viride*)

Fish.

Parrotfish (phylogenetic family *Scaridae*) are present in tropical marine waters worldwide and are considered a keystone herbivore in coral reef habitats, functioning in part as an algae grazer that limits light-blocking algal cover on coral. The target study species is the adult stoplight parrotfish, which is 30–40 cm in length. It is a protogynous hermaphrodite, exhibiting marked sex differences in bright, distinct coloration. Males are territorial and both sexes generally restrict themselves to one bay or reef system.

Sampling sites.

The study habitat was marine fringing coral reef in the Lesser Antilles of the Caribbean Sea. The specific sites (Fig. 1*B*) are associated with the island of St. John, US Virgin Islands (18.3368 N, 64.7281 W) and a nearby undeveloped/protected island [Flanagan Island, <0.1 km Bay (pop. 2711)] on the west end of the island. The CAE site was the reef of Gallows Point at the town of Cruz Bay (pop. 2711) on the west end of the island and serving as its major commercial/tourism access, with regular ferry, barge, and private boat traffic to and from St. Thomas and an adjacent boat harbor. The non-CAE site was reef of Flanagan Island (18.1932 N, 64.3903 W), which is ~1.8 km east of St. John in open sea and is uninhabited and protected.

Sample collection.

The specific collection sites were chosen by the following criteria: *1*) presence of adult parrotfish; *2*) fringing coral reef extending <50 m offshore, *3*) minimum reef area of 300 m^2; *4*) reef depth 3–12 m; *5*) subjectively similar reef physical characteristics (e.g., rugosity, %coral cover, habitat type). Between February 1 and 11, 2019 (via SCUBA access to the reef), sampled parrotfish were carefully followed and observed until defecation, and 2–3 g of the deposited feces were immediately drawn into a modified sterile syringe, capped, and placed into a pouch. All study parrotfish samples were collected by seasoned parrotfish observer J. Turner. Fresh parrotfish feces are larger than those of most reef-fish species and are readily differentiable by color (yellow-green) and shape (tubular rough surface) from previously deposited feces (grey and broken down) nearby. Sample "*n*" ranged from 6–8 per site on a given collection day, and sexes were sampled randomly. On return to land the samples were stored at −18°C until transported to the laboratory on cold-packs and stored at −40°C for later extraction and analysis.

Stress reference base.

A reference base for quantifying chronic stress via plasma and/or fecal cortisol has previously been established for *Sparisoma viride* in the laboratory and in the wild, and the measurement of cortisol across weeks on fringing reef areas exposed to or absent of local anthropogenic stress have been documented (11). Known stressed versus unstressed local sites validated as

such in conjunction with cortisol data were employed to assess possible stress impacts on gut microbiome.

Study Site 2: Walleye (*Sander vitreus*)

Fish.

The walleye is a ray-finned freshwater fish of the family *Percidae* and is native to Lake Erie, which forms a portion of the United States/Canada border. This species is prized by fishermen and is the most commercially valuable fish in the Great Lakes. The greatest walleye density is in Lake Erie, where the sport-fishing industry yields >$1 billion annually. Walleye are carnivorous, eating zooplankton, insects, amphibians, and other fish. Adults are 45–60 cm with females consistently larger than males. They travel considerable distances in the spring and fall (especially females) but many are preferentially more local in preferred areas in the summer months.

Sampling sites.

The study habitat is the relatively shallow (3–13 m) western basin of Lake Erie (Fig. 1*B*), which is the southernmost of the five Laurentian Great Lakes. The lake is 388 km in length and is centered at 41.2 N, 81.2 W, along the northern land edge of the US states of Ohio and New York. Sample collections were done in the months of May, July, and August 2019. The lake study-area habitat was similar, with water depth of 5–13 m, summer temperature 20°C –24°C, and primarily mud/sand bottom. The elevated summer temperatures in the relatively shallow water in concert with extensive agricultural runoff encourage massive eutrophication in the westernmost parts of the lake. This leads to intense toxic algal blooms and low dissolved oxygen in the water, creating stressful environment for flora and fauna, especially in July and August (20, 21). The CAE site was in the area of heaviest algal blooms. The non-CAE site was selected ~10 km east of the algae-intensive area, where algal presence was minimal and water clarity was high. The samples from May were obtained from walleye in the Maumee River during the spring walleye spawn. This river empties into the southwest corner of the lake and served in this study as a water-condition reference for a nonalgal period and location. The river is cleanest in the Spring, and the sample collections there preceded agricultural runoff that partly seeds the harmful algal bloom.

Sample collection.

Samples for each site were obtained from fresh fisherman-caught walleye brought in for filleting. Both sexes were sampled. The hindgut portion of the gastrointestinal tract was removed, bagged, and placed on ice. In the laboratory, each tract exterior was rinsed and dried, and a fecal sample (~1 g) from the portion of each tract proximal to the anus was placed in a sterile vial and frozen (Note: both the fully formed material in the hindgut and the expelled material are defined as "feces").

Stress reference base.

Although fecal cortisol has not been reported for walleye, increased plasma cortisol in response to various stressors has been shown in walleye (22, 23) and in other freshwater and marine fish, including carnivorous species, like walleye (24). Although the degree of cortisol response to both acute and chronic stress varies across fish species and conditions, the increase in cortisol is universal. In addition, Lupica and Turner (10) reported similarity of fecal and plasma cortisol responses to a CAE condition (elevated nitrate) in captive freshwater koi fish. Based on these prior reports, walleye cortisol was not measured in this study. This does not ignore potential for cortisol measurement in study-site-specific fish.

Cortisol Extraction/Assay

Approximately 2 g of thawed raw fecal sample was desiccated at 40°C, and 0.2 g of the dried material was shaken for 1 h in methylene chloride, then centrifuged (10 min at 1,000 *g*) and mixed with 1 M NaOH. Four milliliters of the solvent layer was evaporated to dryness in a heated centrifuge (40°C at 1,000 *g*, 60 min), redissolved in 50-μL increments of 95% EtOH with vortexing until volume was 500 μL. This cortisol sample source was applied in duplicate in appropriate volume into a DetectX cortisol enzyme immunoassay kit (Arbor Assays, Ann Arbor, MI).

Fecal DNA Extraction

Isopropanol DNA extraction method was modified with QIAamp PowerFecal DNA kit and DNeasy Kit (Qiagen, Hilden, Germany) for wet fish fecal samples (25). Each 100 μL fecal samples stored at −40°C was thawed at room temperature and spun at 3,000 *g* for 5 min. Excess water was removed and 800 μL of lysis buffer (500 mM NaCl, 50 mM Tris-HCl, 50 mM EDTA, 4% SDS, pH 8) was added. Dry bead was also added to the same vial. These samples were mechanically disrupted using BioShake XP mixer (Bulldog Bio, Inc., NH) for 3 min with 2,000 rpm, followed by incubation at 70°C for 20 min with a periodic vortex. Vials were then centrifuged for 5 min at 5000 g followed by transferring the supernatant to a sterile 1.5 mL tube. The 200 μL of 10 mM ammonium acetate was added. A brief vortex was followed by incubating the samples on ice for 5 min. The vials were centrifuged at 5,000 *g* for 5 min followed by a second transfer of the supernatant to sterile 1.5 mL tube. One volume of chilled isopropanol was added to the sample and incubated on ice for 30 min. Centrifugation of the tubes at 16,000 *g* was done at a constant temperature of 4°C for 15 min. The resulting DNA pellet was washed with 70% ethanol and finally resuspended in 150 μL of TE buffer (10 mM Tris, 1 mM EDTA). The 15 μL of proteinase K and 200 μL of AL buffer from the DNeasy Kit were added to the DNA. The DNA was further purified by following the DNeasy kit manufacturer's instructions. The final purified product was eluted in 50 μL of TE buffer. Each DNA concentration was measured by NanoDrop (Thermo Fisher Scientific, Waltham, MA) and diluted to be 5 ng/μL.

16S-rRNA Sequencing and Analysis

Standard methods for assessment of microbial communities were employed to generate the database of this report. As previously described (26, 27), 2.5 μL of 5 ng/μL DNA was used for PCR library preparation of 16S ribosomal RNA (rRNA) gene sequencing. Briefly, the V3 to V4 regions of the 16S-rRNA gene were amplified following the Illumina User Guide: 16S-rRNA gene Metagenomic Sequencing Library Preparation-Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq

System (Illumina, Part no. 15044223 Rev. B; Illumina, San Diego, CA). The 10 pM denatured and diluted libraries were mixed with 15% PhiX control and then loaded on an Illumina MiSeq V3 flow cell kit with 2 × 300 cycles. The raw sequencing data were stored in NCBI Sequence Read Archive (SRA) with the accession number as PRJNA813513. Raw paired-end reads of the 16S-rRNA sequencing data were merged to create consensus sequences and then quality filtered using USEARCH (27). Chimeric sequences were identified and filtered using the Quantitative Insights Into Microbial Ecology (QIIME) software package (version 1.9.1) (27) and USEARCH. Operational taxonomic units (OTUs) were subsequently picked using QIIME and USEARCH. Taxonomy assignment was performed using Greengenes (28) as the reference database. Linear discriminant analysis effect size (LEfSe: https://huttenhower.sph.harvard.edu/galaxy/) was used to identify differentially abundant taxonomic features (29). Principal coordinate analysis Bray-Curtis to evaluate the dissimilarity was generated using Calypso based on the OTUs (30).

Supervised Machine Learning

Supervised ML algorithms, including random forest (RF), support vector machine (SVM), gradient boosting (GB), and neural networks (NN), were implemented using Scikit-learn (31) with default parameters for each algorithm. In each ML experiment, fish fecal samples with their gut microbiota data through taxonomy assignment were assigned into training (70%) and testing (30%) data sets after the whole data set was shuffled. In the testing phase, prediction accuracy and confusion matrix were used to assess ML performance. The entire process, comprising of data shuffling, data splitting of training and testing sets, training, and testing, was performed for 100 independent iterations to avoid training and testing biases due to one-time specific assignment of training and testing sets and also comprehensively assess the capability of ML models to detect CAE in the Caribbean Sea and the Ohio Lake Erie across different sites and times. The values of mean and standard deviation of all the performance parameters were computed from the 100 independent iterations of ML modeling. Confusion matrix was generated to summarize the ML classifier's predictive performance (predicted labels vs. true labels).

RESULTS

Profiling Gut Microbiota of Parrotfish (Saltwater Teleost)

This study was carried out in the US Virgin Islands in locations identified as CAE and non-CAE sites (Fig. 2*A*). Fecal cortisol values (stress indicator) of parrotfish had been obtained in the same February time frame from 2014 to 2019, yielding an average cortisol value (and SE) across those 6 yr as a reference for the CAE (presence of chronic intensive human activity) and non-CAE (minimal human activity) sites. The 6-yr average cortisol level was 4.75-fold greater at the CAE site than the non-CAE site (Fig. 2*B*). In the present study, a total of 21 parrotfish fecal samples, collected from both the CAE (n = 11) and non-CAE (n = 10) sites, were processed for 16S-rRNA metagenomics analysis. Significant differences in gut microbial composition were observed between the CAE and non-CAE groups (Fig. 2, *C* and *D*). β-Diversity, which is commonly used to measure similarity or dissimilarity of two communities, showed two distinct clusters representing overall gut microbial differences between the CAE and non-CAE groups (Fig. 2*C*). The plot of linear discriminant analysis effect size (LEfSe) showed a significant number of differential bacterial taxa between the CAE and non-CAE groups (Fig. 2*D*). For example, the phyla Fusobacteria, Firmicutes, and Actinobacteria were more abundant in the CAE group, and the phylum Proteobacteria was more abundant in the non-CAE group (Fig. 2*D*). At the bacterial genus level, *Epulopiscium* and *Cetobacterium* were more abundant in the CAE group, and *Synechococcus* and *Robiginitalea* were more abundant in the non-CAE group (Fig. 2*D*).

Profiling Gut Microbiota of Walleye (Freshwater Teleost)

This study was carried out in Lake Erie, Ohio, where a major CAE problem is seasonal harmful algae blooms. These blooms jeopardize walleye in Lake Erie and also pose a risk to public health due to microcystins, a cyanobacteria-produced liver toxin found in lake water and in walleye (20, 21). We collected fecal samples of 25 individual walleye from the CAE site and 33 individual walleye from a non-CAE (minimal algae bloom) site in May, July, and August 2019 (Fig. 3*A*). The 16S-rRNA sequencing analysis demonstrated distinct gut microbial differences between walleye captured from CAE and non-CAE sites (Fig. 3, *B* and *C*). The β-diversity analysis showed two distinct clusters representing overall gut microbial differences between the CAE and non-CAE groups (Fig. 3*B*). The LEfSe analysis also showed a significant number of differential bacterial taxa between the CAE and non-CAE groups (Fig. 3*C*). For example, the phyla Firmicutes and Fusobacteria were more abundant in the CAE group, and the phylum Proteobacteria was more abundant in the non-CAE group (Fig. 3*C*). In the bacterial genus level, *Cetobacterium*, *Plesiomonas*, and *Clostridium* were more abundant in the CAE group, and *Agrobacterium*, *Delftia*, *U114*, *Pseudomonas*, *Phaeospirillum*, *Ralstonia*, and *Jeotgalicoccus* were more abundant in the non-CAE group (Fig. 3*C*). The findings of similar wild-fish gut microbiome alterations due to the CAE conditions in the Caribbean Sea and Lake Erie support further assessment of wild fish gut microbiota as a possible tool for flagging, monitoring, and pinpointing aquatic ecosystem dysfunction. Thus, we examined the potential of processing fish-gut microbiome data via supervised ML models for reliable detection of CAE.

Machine learning modeling.

As shown in Fig. 4*A*, gut bacterial taxa of parrotfish and/or walleye (Supplemental Tables S1–S3; all Supplemental material is available at https://doi.org/10.5281/zenodo.5823363) were randomly split into training and testing samples. In the training phase, four different supervised ML models, including random forest (RF), support vector machine (SVM), gradient boosting (GB), and neural network (NN), were trained on training samples labeled CAE or non-CAE for the machine to learn how to use gut microbiome data to identify and classify them as belonging to the CAE or non-CAE group (Fig. 4*A*). After the training was completed, the trained ML models were tested (similar to students taking a test) on unlabeled testing samples to use their gut microbiome data to predict their labels (CAE or non-CAE) after which we compared the ML models' predicted labels with the actual labels (Fig. 4*A*). As

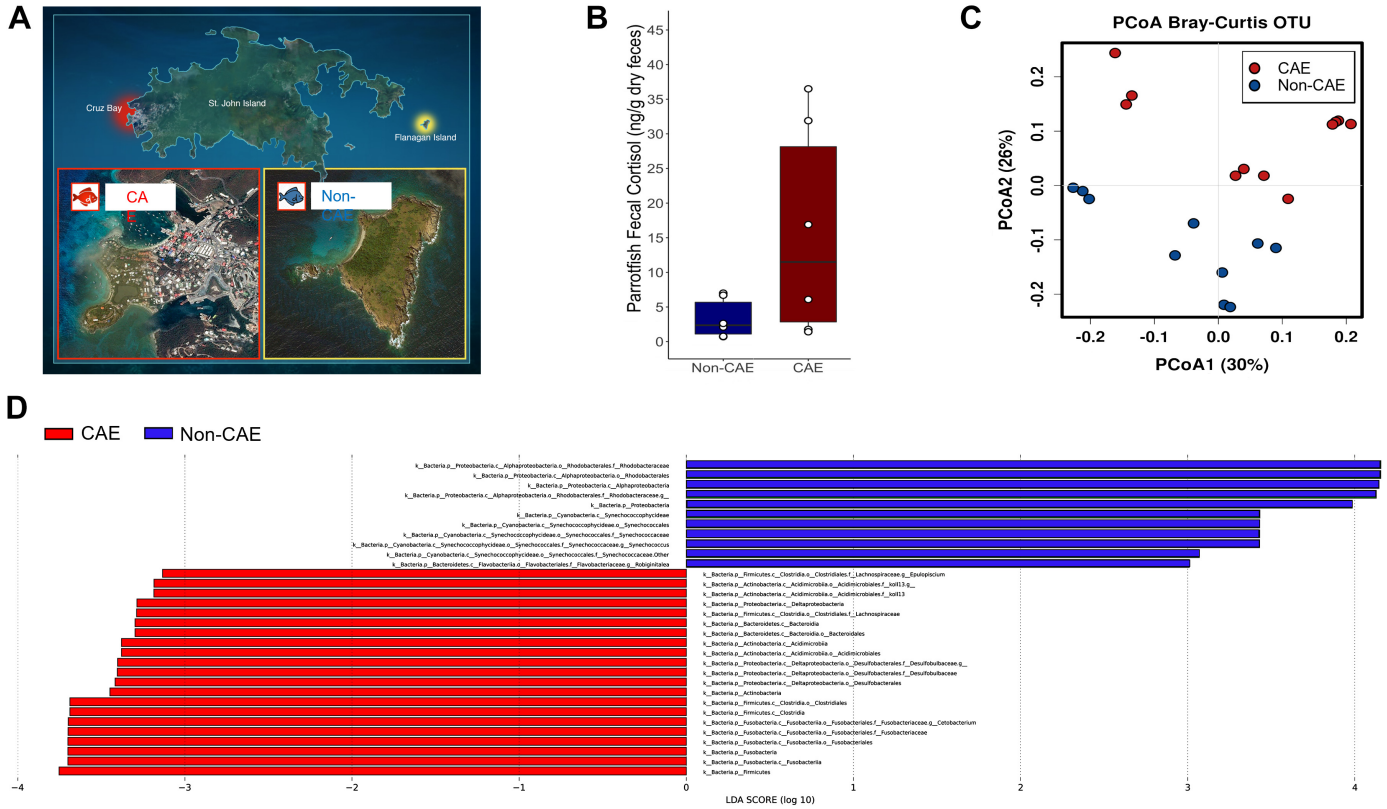

Figure 2. Notable gut microbiota differences of parrotfish in response to the CAE and non-CAE conditions in the Caribbean Sea. *A*: sample collection sites. *B*: the 6-yr average cortisol level (n = 5 or 6/group). *C*: quantitative Bray–Curtis β-diversity of gut microbiota (ANOSIM P = 0.002; n = 10 or 11/group). *D*: LEfSe plot showing enriched bacterial taxa in different groups (n = 10 or 11/group). Bacterial taxa with negative linear discriminant analysis (LDA) scores (red) are enriched in the CAE group, whereas bacterial taxa with positive LDA scores (blue) are enriched in the non-CAE group. Bacterial taxa, that have a score of linear discriminant analysis (LDA) greater than 3.0, were deemed significant. CAE, compromised aquatic environments; LEfSe, linear discriminant analysis effect size.

the samples were collected at different times and sites, we performed 100 independent iterations of random assignment of training and testing samples for ML training and testing to comprehensively assess the capability of ML models to detect CAE in the Caribbean Sea and the Ohio Lake Erie across different sites and times.

In ML *experiment 1*, a total of 21 parrotfish fecal samples, collected from the CAE (n = 11) and non-CAE (n = 10) locations in the Caribbean Sea, were used for ML training and testing. As shown in Fig. 4*B*, the models of random forest (RF), gradient boosting (GB), and neural networks (NN) performed well with >80% prediction accuracies (both median and mean accuracies among 100 independent iterations). The NN model performed best with ~96% mean and 100% median prediction accuracies (Fig. 4*B*) and indicated that most of the independent iterations returned 100% prediction accuracy (Fig. 4*B*).

In ML *experiment 2*, a total of 58 individual walleye fecal samples, collected from both CAE (n = 25) and non-CAE (n = 33) sites in Lake Erie, Ohio, were used for ML training and testing. As shown in Fig. 4*C*, all the four supervised ML models performed well with >80% prediction accuracies (both median and mean accuracies among 100 independent iterations). The GB model performed best with ~92% mean and ~94% median prediction accuracies (Fig. 4*C*) and suggested that most of the independent iterations returned >90% prediction accuracy. Remarkably, in most testing scenarios for the GB model, mislabeling occurred only in one case out of 18 testing cases (Fig. 4*C*).

Finally, in ML *experiment 3*, we tested the hypothesis that ML models could be trained on gut microbiome data of different fish species for detecting the occurrence of CAE across different aquatic ecosystems. To test this hypothesis, we combined the gut microbiome data of both walleye and parrotfish in the previous experiments to train and test the ML models for their capabilities of detecting the occurrence of CAE with different contributing factors in both Lake Erie and Caribbean Sea. Surprisingly, as shown in Fig. 4*D*, the models of RF, GB, and NN still performed comparably well with >85% prediction accuracies (both median and mean accuracies among 100 independent iterations). For example, the RF model achieved ~92% median prediction accuracy, indicating that misprediction occurred only in two cases out of 24 testing cases as shown in a representative testing scenario in Fig. 4*D*.

DISCUSSION

Climate change is a threat to the existence of all life forms on earth but has been viewed more critically from an anthropocentric view. Fish, like all eukaryotes, are holobionts, which comprise the host macroorganism with many commensal microorganisms living in it in an ecologically symbiotic relationship. In recent years, commensal microbial signatures of humans and a variety of terrestrial animals and plants have been reported (32). Such studies have established that commensal microbiota adapts to the environmental conditions of

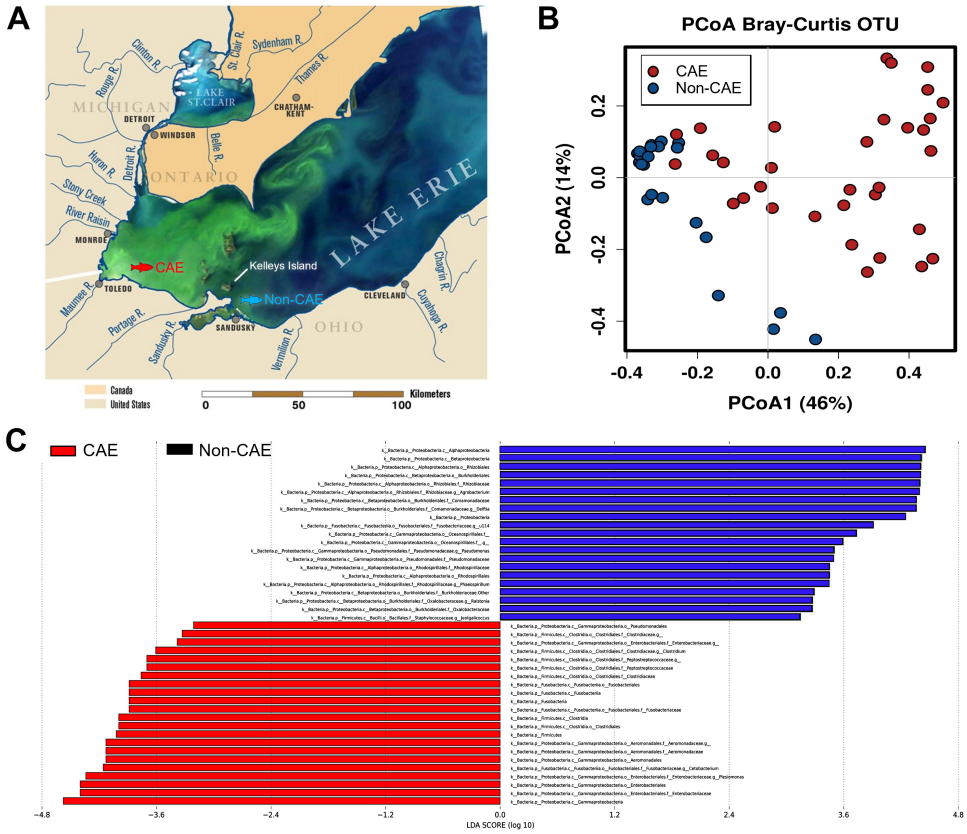

Figure 3. Notable gut microbiota differences of walleye in response to the CAE and non-CAE conditions in Lake Erie. *A*: fecal sample collection sites. *B*: quantitative Bray–Curtis β-diversity of gut microbiota (ANOSIM $P = 0.001$). *C*: LEfSe plot showing enriched bacterial taxa in different groups. Bacterial taxa with negative linear discriminant analysis (LDA) scores (red) are enriched in the CAE group, whereas bacterial taxa with positive LDA scores (green) are enriched in the non-CAE group. Bacterial taxa, that have a score of linear discriminant analysis (LDA) greater than 3.0, were deemed significant. $n = 25–33$/group. CAE, compromised aquatic environments; LEfSe, linear discriminant analysis effect size.

the host and contributes importantly to host health. In particular, environmental stressors on land are reported to reshape gut microbiota both in humans and in other terrestrial animals (33, 34). Since environmental stressors affect our water bodies too, we hypothesized that commensal microbiota of aquatic wildlife is impacted by these stressors. To test our hypothesis, we collected fecal samples from wild fish inhabiting areas with differences in their levels of environmental stressors. Two taxonomically different teleost fish, one species from the Caribbean Islands and another from Lake Erie, were collected for microbial composition analysis by metagenomics sequencing. This is the first report that *1*) details comparison between

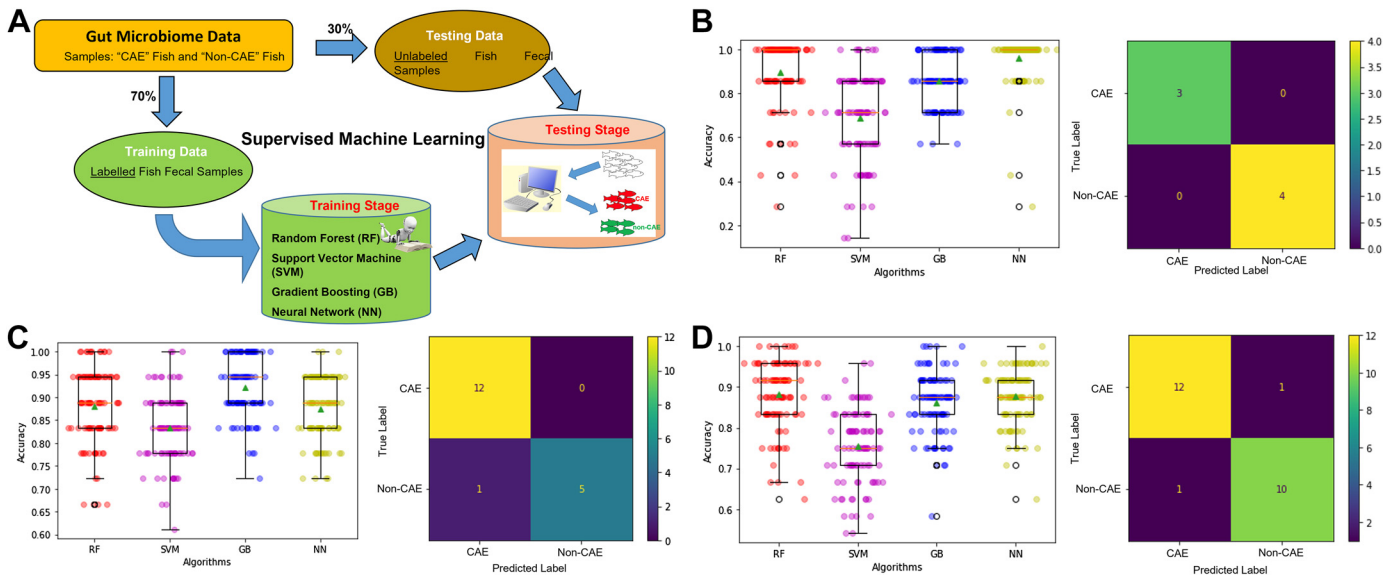

Figure 4. Wild-fish gut microbiome-based ML for detecting CAE in aquatic ecosystems. *A*: ML workflow. *B*: parrotfish gut microbiome-based ML performance of detecting compromised environment in the Caribbean Sea. *C*: walleye gut microbiome-based ML performance of detecting compromised environment in Lake Erie. *D*: walleye and parrotfish gut microbiome-based ML performance of detecting compromised environment in Lake Erie and Caribbean Sea. CAE, compromised aquatic environments; ML, machine learning.

two distinctly different taxonomic families of wild fish representing both the fresh water and oceanic habitats; *2*) identifies that within each of these habitats, significant reshaping of gut microbiota occurred in response to environmental stressors; *3*) documents the surprising finding that despite the differences between the two taxonomically and geographically different wild species, there is a remarkably significant similarity in the gut microbiome compositional response to chronic stress; and *4*) demonstrates a clear link between the rearrangement of microbiota in response to a stressor with the stress experienced by the host, suggesting commensal microbiota as a novel factor influenced by the damaging consequences of environmental stressors on aquatic wildlife. The consistency of such alterations in microbial compositions regardless of the species or aquatic site on the planet portends that environmental stressors are contributing to evolutionary pressure on the selective ecological adaptations of gut microbiota residing in fish, possibly governed by Darwin's theory of the survival of the fittest.

Although some studies have been performed in both fish and mammals to document gut microbiome content and impacts of specific variables on gut microbiome patterns, far less has been devoted to studies in the wild (16, 35). Although farmed-fish yield presently exceeds wild-caught fish yield annually, 30%–50% of the mass of the latter is used to feed the former (species dependent) (36). Thus, wild fish flourishing in a healthy natural environment continues to be a critical goal for the future. The present report is the first to delineate full taxonomic information on the gut microbiome of these two important wild species, i.e., *Sparisoma viride* (a keystone coral-reef species) and *Sander vitreus* (a commercially valuable freshwater species). Parrotfish and walleye experience an array of notable differences in their taxonomy and their respective ecologies (geographic location, water type, climate, diet, habitat, and niche). In addition, the type of chronic anthropogenic stressor in this report differed for these two fish species (fouled reef/water vs. toxic algal bloom, respectively). Despite this plethora of differing input variables, the gut microbiota pattern in fish occupying stressed environs versus unstressed environs was remarkably similar for both species. For example, at the bacterial phylum level, Firmicutes and Fusobacteria were more abundant in the CAE group in both Lake Erie and Caribbean Sea; Proteobacteria was more abundant in the non-CAE group of both locations (Figs. 2 and 3). At the bacterial class level, Fusobacteriia and Clostridia were more abundant in both CAE groups, and Alphaproteobacteria was more abundant in the non-CAE group (Figs. 2 and 3). In addition, the bacterial orders Fusobacteriales and Clostridiales, the bacterial family *Fusobacteriaceae*, and the bacterial genus *Cetobacterium* were more abundant in both CAE groups (Figs. 2 and 3). These findings demonstrate a clear link between the rearrangement of microbiota of aquatic wildlife in response to a worsening environmental condition and support further assessment of wild fish gut microbiota as a possible tool for flagging, monitoring, and pinpointing compromised aquatic environment in an aquatic ecosystem. Future laboratory studies could be performed to more closely examine individual species type/density shifts across the gut microbiota in response to known agents that cause CAE (i.e., toward gut-microbe fingerprinting to identify CAE cause). Toward this perspective, we further examined the potential of supervised ML models to use fish gut microbiome data for detecting aquatic compromise.

Four different supervised ML models, including random forest (RF), support vector machine (SVM), gradient boosting (GB), and neural network (NN), were trained on fish gut microbiome samples labeled with CAE or non-CAE for the machine to learn how to use gut microbiota data to identify and classify them as CAE or non-CAE (Fig. 4*A*). After the training was completed, the trained ML models were tested on unlabeled samples to use their gut microbiota data to predict their labels. Next, the predicted labels were compared with the actual labels to assess each model's classification capabilities (Fig. 4*A*). After the training was completed, the trained ML models were tested on unlabeled samples to use their gut microbiota data to predict their labels, after which we compared the predicted labels with the actual labels (Fig. 4*A*). As the wild-fish fecal samples were collected across different times and sites in the Caribbean Sea and the Ohio Lake Erie, we performed 100 independent iterations of random assignment of training and testing samples for ML training and testing to comprehensively assess spatiotemporal performance of ML models to detect CAE. Supervised ML achieved high prediction accuracies (>90%) to detect different anthropogenic stressors in the Caribbean Sea and the Ohio Lake Erie, respectively (Fig. 4, *B* and *C*). We further hypothesized that ML models could be trained on gut microbiome data of different fish species for detecting the occurrence of compromised environments across different aquatic ecosystems. To test this hypothesis, an independent experiment was performed combining the gut microbiota data of all walleye and parrotfish fecal samples to independently train and test the ML models for their capabilities to detect the occurrence of compromised environment in the presence of the different contributing factors of both Lake Erie and Caribbean Sea. Surprisingly, the RF model still achieved >90% prediction accuracy (Fig. 4*D*). These results demonstrate the exciting feasibility of constructing a consistent ML model trained on gut microbiome data for accurately detecting different compromised environmental conditions across different aquatic ecosystems, such as lakes and oceans. It is noteworthy that the sample sizes of both walleye and parrotfish in these experiments were less than optimal for training and assessing ML models and that ML predictive precision could be maximized with larger sample numbers. Future studies should also include additional fish species to consolidate ML as a reliable predictive tool for identifying compromised aquatic conditions.

Although environmental compromise at the global level is already present and highly challenging, the addition of "local" stressors can add further imbalance to those local environments (37, 38). Early detection of such local impacts is best as human activities continue to expand. Further consideration of gut microbiome assessment as a potential tool in the arena of environmental-stress monitoring appears warranted, for both aquatic and terrestrial wildlife. In addition to being unaffected by acute stressors, gut microbiome assessment relative to cortisol requires one-tenth of the sample volume, allows simpler extraction, is lower in cost and labor, provides more information that can be categorized and may offer differentiation of specific stressor types. The fact that we compared two different species (differing in

geography, water type, climate, habitat, food, and type of stress) and observed similar gut microbiome shifts for both fish species for CAE versus non-CAE conditions is promising for consistency and could serve as evidence for potential broader, species-independent applications.

DATA AVAILABILITY

The data that support the findings of this study are openly available in the National Center for Biotechnology Information Sequence Read Archive (SRA) BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/), accession number PRJNA813513.

SUPPLEMENTAL DATA

Supplemental Tables S1–S3: https://doi.org/10.5281/zenodo.5823363.

ACKNOWLEDGMENTS

Our thanks to the following individuals for field assistance: (USVI; V. Powell, T. Kelley, R. Meyers) and (Lake Erie; A. Ankenbrandt, P. Pacholski).

GRANTS

Funding is acknowledged for J.W.T. (PNC Foundation), B.J. (NIH; HL143082, AI164449), and T.Y. (American Heart Association; CDA852969.

DISCLOSURES

Xi Cheng is an editor of Physiological Genomics and was not involved and did not have access to information regarding the peer-review process or final disposition of this article. An alternate editor oversaw the peer-review and decision-making process for this article.

None of the other authors has any conflicts of interest, financial or otherwise, to disclose.

AUTHOR CONTRIBUTIONS

J.W.T.Jr., X.C., and B.J. conceived and designed research; J.W.T.Jr., X.C., N.S., J.-Y.Y., and T.Y. performed experiments; J.W.T.Jr., X.C., N.S., J.-Y.Y., and T.Y. analyzed data; J.W.T.Jr., X.C., N.S., T.Y., and B.J. interpreted results of experiments; J.W.T.Jr., X.C., N.S., and T.Y. prepared figures; J.W.T.Jr., X.C., and N.S. drafted manuscript; J.W.T.Jr., X.C., N.S., J.-Y.Y., T.Y., and B.J. edited and revised manuscript; J.W.T.Jr., X.C., N.S., J.-Y.Y., T.Y., and B.J. approved final version of manuscript.

REFERENCES

1. **Occhipinti-Ambrogi A.** Global change and marine communities: alien species and climate change. *Marine Pollut Bull* 55: 342–352, 2007. doi:10.1016/j.marpolbul.2006.11.014.
2. **Laidre KL**, **Stirling I**, **Lowry LF**, **Wiig Ø**, **Heide-Jørgensen MP**, **Ferguson SH.** Quantifying the sensitivity of Arctic marine mammals to climate-induced habitat change. *Ecol Appl* 18: S97–S125, 2008. doi:10.1890/06-0546.1.
3. **Hoegh-Guldberg O**, **Bruno JF.** The impact of climate change on the world's marine ecosystems. *Science* 328: 1523–1528, 2010. doi:10.1126/science.1189930.
4. **Ellis JI**, **Jamil T**, **Anlauf H**, **Coker DJ**, **Curdia J**, **Hewitt J**, **Jones BH**, **Krokos G**, **Kürten B**, **Hariprasad D**, **Roth F**, **Carvalho S**, **Hoteit I.** Multiple stressor effects on coral reef ecosystems. *Glob Chang Biol* 25: 4131–4146, 2019. doi:10.1111/gcb.14819.
5. **Danz NP**, **Niemi GJ**, **Regal RR**, **Hollenhorst T**, **Johnson LB**, **Hanowski JM**, **Axler RP**, **Ciborowski JJH**, **Hrabik T**, **Brady VJ**, **Kelly JR**, **Morrice JA**, **Brazner JC**, **Howe RW**, **Johnston CA**, **Host GE.** Integrated measures of anthropogenic stress in the U.S. Great Lakes basin. *Environ Manage* 39: 631–647, 2007. doi:10.1007/s00267-005-0293-0.
6. **Steinman AD**, **Cardinale BJ**, **Munns WR Jr**, **Ogdahl ME**, **Allan JD**, **Angadi T**, **Bartlett S**, **Brauman K**, **Byappanahalli M**, **Doss M**, **Dupont D**, **Johns A**, **Kashian D**, **Lupi F**, **McIntyre P**, **Miller T**, **Moore M**, **Muenich RL**, **Poudel R**, **Price J**, **Provencher B**, **Rea A**, **Read J**, **Renzetti S**, **Sohngen B**, **Washburn E.** Ecosystem services in the Great Lakes. *J Great Lakes Res* 43: 161–168, 2017. doi:10.1016/j.jglr.2017.02.004.
7. **Troell M**, **Naylor RL**, **Metian M**, **Beveridge M**, **Tyedmers PH**, **Folke C**, **Arrow KJ**, **Barrett S**, **Crépin A-S**, **Ehrlich PR**, **Gren A**, **Kautsky N**, **Levin SA**, **Nyborg K**, **Österblom H**, **Polasky S**, **Scheffer M**, **Walker BH**, **Xepapadeas T**, **de Zeeuw A.** Does aquaculture add resilience to the global food system? *Proc Natl Acad Sci USA* 111: 13257–13263, 2014. doi:10.1073/pnas.1404067111.
8. **Béné C**, **Barange M**, **Subasinghe R**, **Pinstrup-Andersen P**, **Merino G**, **Hemre G-I**, **Williams M.** Feeding 9 billion by 2050—putting fish back on the menu. *Food Sec* 7: 261–274, 2015. doi:10.1007/s12571-015-0427-z.
9. **Rogers CS**, **Garrison G**, **Grober R**, **Hillis Z-M**, **Franke MA.** Coral Reef Monitoring Manual for the Caribbean and Western Atlantic (Online). Virgin Islands National Park: National Park Service, Washington, DC, 1994. https://reefresilience.org.
10. **Lupica SJ**, **Turner JW Jr.** Noninvasive assessment of nitrate-induced stress in koi Cyprinus carpio L. by faecal cortisol measurement. *Aquacult Res* 41: 1622–1629, 2010. doi:10.1111/j.1365-2109.2010.02496.x.
11. **Turner JW Jr**, **Nemeth R**, **Rogers C.** Measurement of fecal glucocorticoids in parrotfishes to assess stress. *Gen Comp Endocrinol* 133: 341–352, 2003. doi:10.1016/S0016-6480(03)00196-5.
12. **Zeller G**, **Tap J**, **Voigt AY**, **Sunagawa S**, **Kultima JR**, **Costea PI**, **Amiot A**, **Böhm J**, **Brunetti F**, **Habermann N**, **Hercog R**, **Koch M**, **Luciani A**, **Mende DR**, **Schneider MA**, **Schrotz-King P**, **Tournigand C**, **Tran Van Nhieu J**, **Yamada T**, **Zimmermann J**, **Benes V**, **Kloor M**, **Ulrich CM**, **von Knebel Doeberitz M**, **Sobhani I**, **Bork P.** Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* 10: 766, 2014. doi:10.15252/msb.20145645.
13. **Frank DN**, **Amand ALS**, **Feldman RA**, **Boedeker EC**, **Harpaz N**, **Pace NR.** Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci USA* 104: 13780–13785, 2007. doi:10.1073/pnas.0706625104.
14. **Tang WW**, **Kitai T**, **Hazen SL.** Gut microbiota in cardiovascular health and disease. *Circ Res* 120: 1183–1196, 2017. doi:10.1161/CIRCRESAHA.117.309715.
15. **Mell B**, **Jala VR**, **Mathew AV**, **Byun J**, **Waghulde H**, **Zhang Y**, **Haribabu B**, **Vijay-Kumar M**, **Pennathur S**, **Joe B.** Evidence for a link between gut microbiota and hypertension in the Dahl rat. *Physiol Genomics* 47: 187–197, 2015. doi:10.1152/physiolgenomics.00136.2014.
16. **Ghanbari M**, **Kneifel W**, **Domig KJ.** A new view of the fish gut microbiome: advances from next-generation sequencing. *Aquaculture* 448: 464–475, 2015. doi:10.1016/j.aquaculture.2015.06.033.
17. **Egerton S**, **Culloty S**, **Whooley J**, **Stanton C**, **Ross RP.** The gut microbiota of marine fish. *Front Microbiol* 9: 873, 2018. doi:10.3389/fmicb.2018.00873.
18. **Aryal S**, **Alimadadi A**, **Manandhar I**, **Joe B**, **Cheng X.** Machine learning strategy for gut microbiome-based diagnostic screening of cardiovascular disease. *Hypertension* 76: 1555–1562, 2020. doi:10.1161/HYPERTENSIONAHA.120.15885.
19. **Manandhar I**, **Alimadadi A**, **Aryal S**, **Munroe PB**, **Joe B**, **Cheng X.** Gut microbiome-based supervised machine learning for clinical diagnosis of inflammatory bowel diseases. *Am J Physiol Gastrointest Liver Physiol* 320: G328–G337, 2021. doi:10.1152/ajpgi.00360.2020.
20. **Wituszynski DM**, **Hu C**, **Zhang F**, **Chaffin JD**, **Lee J**, **Ludsin SA**, **Martin JF.** Microcystin in Lake Erie fish: risk to human health and relationship to cyanobacterial blooms. *J Great Lakes Res* 43: 1084–1090, 2017. doi:10.1016/j.jglr.2017.08.006.
21. **Nieman CL**, **Oppliger AL**, **McElwain CC**, **Gray SM.** Visual detection thresholds in two trophically distinct fishes are compromised in algal compared to sedimentary turbidity. *Conserv Physiol* 6: coy044, 2018. doi:10.1093/conphys/coy044.
22. **Barton BA**, **Haukenes AH**, **Parsons BG**, **Reed JR.** Plasma cortisol and chloride stress responses in juvenile walleyes during capture,

transport, and stocking procedures. *North Am J Aquacult* 65: 210–219, 2003. doi:10.1577/C02-030.

23. **Killen SS**, **Suski CD**, **Morrissey MB**, **Dyment P**, **Furimsky M**, **Tufts BL.** Physiological responses of walleyes to live-release angling tournaments. *North Am J Fish Manage* 23: 1238–1246, 2003. doi:10.1577/M02-164.
24. **Barton BA.** Stress in fishes: a diversity of responses with particular reference to changes in circulating corticosteroids. *Integr Comp Biol* 42: 517–525, 2002. doi:10.1093/icb/42.3.517.
25. **Hart ML**, **Meyer A**, **Johnson PJ**, **Ericsson AC.** Comparative evaluation of DNA extraction methods from feces of multiple host species for downstream next-generation sequencing. *PLoS One* 10: e0143334, 2015. doi:10.1371/journal.pone.0143334.
26. **Galla S**, **Chakraborty S**, **Cheng X**, **Yeo J**, **Mell B**, **Zhang H**, **Mathew AV**, **Vijay-Kumar M**, **Joe B.** Disparate effects of antibiotics on hypertension. *Physiol Genomics* 50: 837–845, 2018. doi:10.1152/physiolgenomics.00073.2018.
27. **Chakraborty S**, **Galla S**, **Cheng X**, **Yeo J-Y**, **Mell B**, **Singh V**, **Yeoh BSan**, **Saha P**, **Mathew AV**, **Vijay-Kumar M**, **Joe B.** Salt-responsive metabolite, β-hydroxybutyrate, attenuates hypertension. *Cell Rep* 25: 677–689.e4, 2018. doi:10.1016/j.celrep.2018.09.058.
28. **DeSantis TZ**, **Hugenholtz P**, **Larsen N**, **Rojas M**, **Brodie EL**, **Keller K**, **Huber T**, **Dalevi D**, **Hu P**, **Andersen GL.** Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72: 5069–5072, 2006. doi:10.1128/AEM.03006-05.
29. **Segata N**, **Izard J**, **Waldron L**, **Gevers D**, **Miropolsky L**, **Garrett WS**, **Huttenhower C.** Metagenomic biomarker discovery and explanation. *Genome Biol* 12: R60–R18, 2011. doi:10.1186/gb-2011-12-6-r60.
30. **Zakrzewski M**, **Proietti C**, **Ellis JJ**, **Hasan S**, **Brion M-J**, **Berger B**, **Krause L.** Calypso: a user-friendly web-server for mining and visualizing microbiome–environment interactions. *Bioinformatics* 33: 782–783, 2017.
31. **Pedregosa F**, **Varoquaux G**, **Gramfort A**, **Michel V**, **Thirion B**, **Grisel O**, **Blondel M**, **Prettenhofer P**, **Weiss R**, **Dubourg V**, **Vanderplas J**, **Passos A**, **Cournapeau D**, **Brucher M**, **Perrot M**, **Duchesnay E.** Scikit-learn: machine learning in Python. *J Mach Learn Res* 12: 2825–2830, 2011.
32. **Rocca JD**, **Simonin M**, **Blaszczak JR**, **Ernakovich JG**, **Gibbons SM**, **Midani FS**, **Washburne AD.** The microbiome stress project: toward a global meta-analysis of environmental stressors and their effects on microbial communities. *Front Microbiol* 9: 3272, 2018. doi:10.3389/fmicb.2018.03272.
33. **Nelson TM**, **Rogers TL**, **Brown MV.** The gut bacterial community of mammals from marine and terrestrial habitats. *PLoS One* 8: e83655, 2013 [Erratum in *PLoS One* 9: e99562, 2014]. doi:10.1371/journal.pone.0083655.
34. **Fackelmann G**, **Gillingham MAF**, **Schmid J**, **Heni AC**, **Wilhelm K**, **Schwensow N**, **Sommer S.** Human encroachment into wildlife gut microbiomes. *Commun Biol* 4: 800, 2021. doi:10.1038/s42003-021-02315-7.
35. **Wang AR**, **Ran C**, **Ringø E**, **Zhou ZG.** Progress in fish gastrointestinal microbiota research. *Rev Aquacult* 10: 626–640, 2018. doi:10.1111/raq.12191.
36. **Bell JD**, **Ganachaud A**, **Gehrke PC**, **Griffiths SP**, **Hobday AJ**, **Hoegh-Guldberg O**, **Johnson JE**, **Le Borgne R**, **Lehodey P**, **Lough JM**, **Matear RJ**, **Pickering TD**, **Pratchett MS**, **Gupta AS**, **Senina I**, **Waycott M.** Mixed responses of tropical Pacific fisheries and aquaculture to climate change. *Nature Clim Change* 3: 591–599, 2013. doi:10.1038/nclimate1838.
37. **Pecl GT**, **Araújo MB**, **Bell JD**, **Blanchard J**, **Bonebrake TC**, **Chen I-C** et al. Biodiversity redistribution under climate change: impacts on ecosystems and human well-being. *Science* 355: eaai9214, 2017. doi:10.1126/science.aai9214.
38. **Ojea E**, **Lester SE**, **Salgueiro-Otero D.** Adaptation of fishing communities to climate-driven shifts in target species. *One Earth* 2: 544–556, 2020. doi:10.1016/j.oneear.2020.05.012.