The Anne Arundel Community College

Journal of Emerging Scholarship

Volume 1
May 2022

COVER

Deborah Judy

AACC Visual Arts Student

Emergence, 2022

Oil on panel, 17" x 22"

The Anne Arundel Community College

Journal of Emerging Scholarship

Volume 1
May 2022

Dear reader,

It is with great excitement that we present here the first edition of the Anne Arundel Community College Journal of Emerging Scholarship. The goal of this journal is to provide an outlet for peer-reviewed publication of collegiate undergraduate student research. The multidisciplinary approach of this endeavor is to reach as broad of an audience as possible and help students develop critical skills of scientific inquiry.

While we hope that our transfer-bound students who begin research projects at AACC will be inspired to continue this path at four-year institutions, all AACC students benefit from the skills gained by engaging in research. In addition to developing technical skills specific to a particular field of study, engaging in research develops other skills which are widely in demand by nearly all industry employers, such as effective team collaboration, writing, presentation, analysis, and critical thinking. Participation in research also fosters information literacy and helps students become critical consumers of the data they encounter in their daily lives.

We want to extend gratitude to the students, mentors, reviewers, and partners who have dedicated countless hours to creating the original work contained within this volume.

Sincerely,
The 2021–2022 Editorial Board

JASON BURKHOLDER, B.S.

Evaluating the Utility of Enterococcus Specific Primers

KEY WORDS

fecal indicator bacteria

microbial source tracking

Enterococcus

multi-locus sequence typing

FACULTY MENTOR

Tammy Domanski, Ph.D.
Professor, Biology Department
Director, Environmental Center at
Anne Arundel Community College

ABSTRACT

Enterococci are the preferred fecal indicator bacteria (FIB) for monitoring the safety of recreational beaches. A reliable and cost-effective method to identify the species of origin for enterococci-contaminated rivers is essential for decreasing the risk to human health. In this study human and canine fecal samples were analyzed in polymerase chain reaction (PCR) studies with primers reported to amplify targets specific to enterococcal species with the goal of identifying the fecal source. While the primers successfully amplified the target sequences in many samples, amplification in non-target species made identifying one, or a small set of primers, that reliably discriminate between fecal source species more challenging. Alignment and comparison of PCR product sequences were conducted with the goal of designing novel primers with increased specificity. Analysis of multi-locus sequence typing (MLST) data suggested that specific nucleotide variations within loci found in species-specific enterococcal strains might be exploited to determine the source of contamination in local waterways. To this end, primers for two target loci were designed specifically for nucleotide sequences more frequently isolated from canine enterococcal samples and initial screening assays were conducted to optimize conditions and discriminate between source DNA without success. Collection of additional species-specific bacterial samples and additional control type strains are needed to better distinguish between the species of interest in this study.

INTRODUCTION

Contamination of recreational waters with bacteria from fecal contamination poses a significant health risk to humans (Cabelli et al. 1979). Increasing water temperatures driven by climate change, increased incidence and severity of rain events bringing more runoff, larger impervious surfaces resulting in less absorption of runoff before entering waterways, and the aging sewer infrastructure, all contribute to more frequent occurrences of beach closures due to high bacterial concentrations (Rose et al. 2001). To develop programs that decrease contamination and to better understand the risk to humans, it is essential to not only quantify the bacterial load in water, but to identify the relative contribution from different contributing species.

Enterococcus sp. are prevalent in bird, mammal and to some extent, insect and reptile fecal material, and comprise approximately 1% of the bacteria in the human large intestine (Dubin and Pamer, 2014). Other species contain a similarly complex and varied array of bacteria (Layton et al. 2010; Harwood et al. 2014).

The correlation between levels of fecal bacteria and illness in humans has long been recognized, and the EPA has identified enterococci as fecal indicator bacteria (FIB), the measurement of which are used to determine the safety of recreational swimming beaches and seafood harvesting waters (Cabelli et al. 1979; US EPA, 2012). High levels in recreational waters can result in beach closures and halt fish and oyster harvests. The standard method for tracking FIB levels utilizes selective media and direct colony counting (US EPA, 2009). Monitoring for all possible pathogens that may be in contaminated water is an impossibility, so the use of FIB has made it possible to track a common set of organisms, compare many locations and set thresholds for safety (Leclerc et al. 2001).

Microbial source tracking (MST) has previously been used to identify the source of enterococci and other bacteria associated

with fecal contamination found in environmental waters (Leclerc et al. 2004) and has been used to identify the source in bacterial infection outbreaks from sources including food and water (McRobb et al 2015). Identification of the contamination source is necessary for developing plans to eliminate the source, such as repairing leaks, upgrading septic systems and educating the public on pet waste clean-up. MST methods, such as restriction analysis, quantitative polymerase chain reaction (qPCR), and DNA sequencing of one or several loci, have been used with varying success (Foley et al. 2009; Homan et al. 2002; Ruiz-Garbajosa, 2006). Polymerase chain reaction (PCR) potentially provides an inexpensive way to identify the source of fecal contamination. Many target organisms have been proposed for PCR-based MST (Harwood et al. 2014). However, methods that target species other than *Enterococcus* require processing of the sample without initially quantifying the level of contamination, adding cost and wasted effort. A method that first screens for enterococcal contamination followed by MST, would be more efficient. To this end a project was initiated to identify or design primer sets that discriminate between fecal source species responsible for *Enterococcus* contamination.

METHODS

Fecal sample collection

Canine fecal samples were obtained from local veterinarians (D samples) and dog owners (S samples). Each D sample contained fecal matter combined from 4 to 10 dogs (n=17). Individual human samples were obtained from anonymous volunteers (n=3; P002, P003, P004), and sewage samples were provided by several Anne Arundel County Water Reclamation Facilities (WRF) (n=8). *Enterococcus faecalis* NCTC 775, a positive control for *Enterococcus faecalis*-specific primers, was obtained from Biomerieux. *Enterococcus* faecium 700221, a positive control for *E. faecium*-specific primers, was obtained from American Type Culture Collection. Environmental

samples were collected from local waterways that contained high concentrations of enterococci (over 1000 bacteria/100 mL, approximately 10 times above the acceptable threshold).

Enterococcus isolation and genomic DNA isolation

Approximately 1 mL of liquid WRF influent or 0.1 mg of fecal matter suspended in sterile water and passed through a sterile 0.45-micron filter. Filters were placed on mE agar (Difco) selecting for *Enterococcus sp*. After incubation at 41 degrees Celsius for 24 hours, colonies with a blue halo were scraped, combined and suspended in sterile water. The Amresco Cyclo-Prep Genomic DNA Isolation kit was used for all DNA extractions (Avantor).

Primer selection and Polymerase Chain Reaction conditions

The primers chosen, their reported specificity and references are shown (Table 1). Primers were purchased from Integrated DNA Technologies (Coralville, USA).

TABLE 1

Primer target and specificity.

Primer Set	target	Reported Specificity	reference
Ent 376	16s rRNA	Enterococcus species	Ryu et al.2012
Ent	16s rRNA	*Enterococcus faecalis*	Ryu et al.2012
Cium	16s rRNA	*Enterococcus faecium*	Ryu et al. 2012
IS16	Insertion Sequence 16	nosocomial human *Enterococcus faecium*	Werner at al. 2011
esp	Enterococci Surface Protein gene	human *Enterococcus faecium*	Ahmed at al. 2008
psts11	*psts* gene fragment	canine *Enterococcus faecium*	This study
atpa15	*atpa* gene fragment	canine *Enterococcus faecium*	This study

Amplification reactions included 1 unit of Taq polymerase (New England Biolabs), 1X buffer, 300 nM dNTPs, 1.5 mM MgCl2, 1mM forward primer, 1mM reverse primer, 2 µl of the

template in a final volume of 50 µl. Samples were placed in a thermocycler and run for 30 cycles. Each cycle incubated samples for 60 sec at 94°C, 60 sec at an annealing temperature specific for a given primer set (Table 2), and 60 sec at 74°C.

TABLE 2

Primer sequences, conditions and predicted product size.

Primer set name	Annealing temp (°C)	Product Length (bp)	Forward Primer	Reverse Primer
esp	51	680	TAT GAA AGC AAC AGC ACA AGT T	ACG TCG AAA GTT CGA TTT CC
IS16	56.3	547	CAT GTT CCA CGA ACC AGA G	TCA AAA AGT GGG CTT GGC
ENT	56.3	229	TGC ATT AGC TAG TTG GTG	AGT TAC TAA CGT CCT TGT TC
ENT376	61	220	GGA CGM AAG TCT GAC CGA	TTA AGA AAC CGC CTG CGC
CIUM	57	512	TGC TCC ACC GGA AAA AGA	TTA AGA AAC CGC CTG CGC
psts11	56	181	CAA AGA TAC AGG TGT CAA AGA TAT CAC A	TAT AGG TGT GGC ACC ATC TAA
atpa15	56	248	GCC AAT CGG ACG CGG A	ATG GTG CGA TAT AAA GTA ATG GT

Analysis of PCR products

Aliquots of reactions were analyzed on 1.5% agarose alongside a 100 base pair standard (Amresco EZ-vision) and stained with ethidium bromide to estimate amplification product size. Samples that resulted in amplification of a product of the expected size were classified as positives. A sample agarose gel in Figure 1 highlights the expected product sizes. Samples that did not result in amplification, therefore no band on the agarose gel, were classified as negatives, and those with multiple bands were placed into a separate group. Once samples were verified they were sent out to Genewiz for sequencing to further verify that target sequences were amplified and to compare sequences.

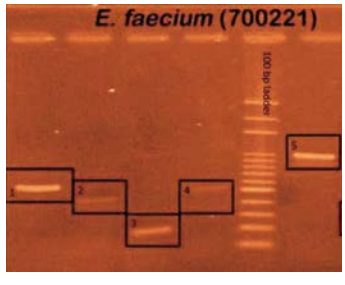


FIGURE 1

Agarose gel illustrating expected amplification product sizes. Each reaction contained Enterococcus faecium strain 700221 genomic template with a different primer set: 1-Esp (680 bp); 2-CIUM (512 bp); 3-ENT376 (220 bp); 4-IS16 (547 bp); 5-VanA (1029bp).

Analysis of MLST

MLST is a technique introduced in the early 1990s utilizing a limited number of short sequences from several loci within bacterial genomes capable of assigning a sample to a specific strain (Maiden et al. 1998). Sequence data from the Public Database for Molecular Typing and Microbial Genomic Diversity (pubMLST) suggested that specific nucleotide variations within loci found in all enterococcal strains might be exploited to determine the source of enterococci in contaminated local waterways. The MLST strain typing method typically employs sequence comparisons at seven loci to assign a sample to a specific strain. In database searches of sequences from many sources, it was found that the sequences of two loci, *psts* and atpa, were aligned and compared, and primers were designed specifically for nucleotide sequences more frequently isolated from canine enterococcal samples. Sequences in the database were aligned to look for individual nucleotide differences prevalent between loci amplified from bacteria from different host species. The analysis was performed in Ugene (Okonechnikov et al. 2012).

RESULTS

PCR results

Amplification results from assays performed to evaluate species selectivity of primer sets were promising. Type strain controls, *E. faecalis* NCTC775 and *E. faecium* 700221, performed as expected with each primer set (Table 3). NCTC775 is a non-virulent strain that does not contain the *esp* gene, while *E. faecium* 700221 is known to contain both the *esp* and *IS16* locus. The number of bacterial DNA samples from individuals was very small in this study (n=3), and none of the samples obtained were from clinical settings. Two of the three samples showed amplification with ENT, ENT376 and CIUM primers, as would be expected.

Sewage samples collected from waste reclamation facilities

around Anne Arundel County contain fecal matter from large populations, so reflect the complexity of bacterial populations in humans. One sewage sample, PAT, was negative for amplification by the ENT376 primer set, although those primers are the most inclusive, reported to amplify sequences from a variety of *Enterococcus* species. The sequence targets associated with potentially more virulent enterococcal species, *esp* and *IS16*, were found in 57% and 86% of sewage samples, respectively (Table 3).

TABLE 3

Summary of PCR Results.

Source Species	Sample Name	Primer set Name				
		esp	*IS16*	ENT	ENT 376	CIUM
E. faecalis	Type strain NCTC775					
E. faecium	Type strain 700221					
Sewage	SP1, SP2, MAYO					
Sewage	BROAD, COX					
Sewage	PAT					
Sewage	COX2					
Human	P002, P003					
Human	P004					
Dog	D1, D2					
Dog	D3					
Dog	D4					
Dog	S1					
Dog	S3					
Dog	S4					
Dog	S5, S8, S15, S16					
Dog	S6, S14					
Dog	S9, S13, S17					
Dog	S12					
Environmental	CG					
Environmental	SS, EGO					
Environmental	BH					
Environmental	HW					
Environmental	HC					
Environmental	CG 2					
KEY	AMPLIFICATION OF TARGET		NO AMPLIFICATION OF TARGET		AMPLIFICATION & extra bands	

The amplification results with bacterial DNA from dog fecal samples, representing over 35 individual dogs, were encouraging

for several reasons. All of the samples from dog fecal material successfully amplified the ENT376 target, confirming the reports that ENT376 is the least selective of the primer sets used. Over 88% of the samples from dogs were positive for the *E. faecium* target (CIUM primer set), while only 69% were positive for *E. faecalis* (ENT primer set). Of note, only 13% of the samples obtained from dog feces were positive for the presence of the *esp* gene, and 29% were positive for *IS16*.

Environmental samples collected from area waterways on days associated with high concentrations of *Enterococcus sp.* were analyzed and compared to look for patterns that might suggest the species responsible for the contamination. Six of the seven samples were positive for amplification with ENT376, suggesting the presence of at least one species of *Enterococcus*. The two CG samples, CG and CG2, were collected on different days. Both were positive for ENT376, while CG was positive for ENT, suggesting the presence of *Enterococcus* faecalis, and CG2 was positive for CIUM, suggesting the presence of E. faecium. None of the environmental samples were positive for amplification of *esp*, and only two, SS and EGO were positive for *IS16*.

Sequence Analysis

To further analyze and compare DNA targets that were amplified, PCR products from a sampling of reactions were sequenced, and compared to sequences with the National Library of Medicine's National Center of Bioinformatics (NLM NCBI) database to confirm that the correct targets were amplified (Table 4). In each case the expected product was amplified with near 100% identity to predicted sequences (Table 2), with one caveat. The *IS16* primer set was designed to recognize human pathogenic, clinical E. faecium strains, but bacterial DNA template from both dog and human fecal samples resulted in amplification of identical products with the highest similarity to a human isolate, with a very

close second match to a dog isolate.

Sample	Primer Set	Product length (bp)	Top BLAST hit/s	% identity	Correct target gene?	Correct target species?
CG	ENT	186	*E. faecalis* strain AaR12 16S ribosomal RNA gene (soil)	100	yes	yes
p004	ENT376	178	*E. faecium* strain g4 16S ribosomal RNA (fish)	100	yes	yes
Dog1 SP2 (identical)	*IS16*	489	1-*E. faecium* strain NMVRE-001 plasmid p1 (human clinical isolate) 2- *E. faecium* strain V13-21-E11-012-001 plasmid pK21EFM001(canine)	99	yes	yes
SP1	*esp*	616	*E. faecium* strain VVEswe-R (human clinical sample)	100	yes	yes

TABLE 4

Sample PCR products sequenced.

Sequence data from a subset of amplification products were aligned and compared to each other. Of the 11 amplification products analyzed, only S1 and S15 products with ENT primers contained nucleotide variations. More variation was observed when comparing the sequences from amplification with the ENT376 primer set. Of 15 samples that were sequenced 5 of them contained at least one nucleotide difference. Only two samples from amplifications with *ESP* primers were sequenced and the sequences were identical. Both of these samples were from sewage effluent, MAYO and SP1. From the *IS16* primer set there were 4 samples sequenced. These had variations in at least two of the four samples, but because of low quality sequence data confidence in the variations was also low.

MLST database alignments and primer design

Alignment of a portion of the E. faecium *psts* locus revealed that of the 105 *psts* alleles in the pubMLST, alleles 11 and 7 were most frequently associated with bacterial DNA from canine sources, while allele 1 was more often associated with bacterial DNA

from human sources (Jolley et al. 2018). Nucleotide differences were used to design primers able to specifically amplify DNA from *psts* allele 11 (Figure 2 and Table 2). One such primer is indicated with yellow highlighting. Initial assays involved varying PCR conditions, specifically using different annealing temperatures that would affect the stability of primer binding. Higher annealing temperatures require a perfect match between primer and target and lower temperatures, potentially allow binding and amplification even if there are mismatches between the primer and the target. In amplification reactions comparing templates from sewage samples, SP1 and Mayo, and dog samples, D3 and S5, an annealing temperature that was able to differentiate between sources, therefore allowing amplification from templates of one species but not the other, was not found (data not shown).

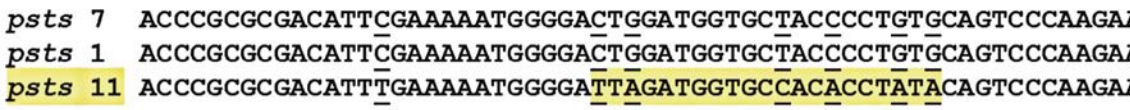


FIGURE 2

Comparison of psts *allele sequences from the pubMLST E. faecium database. Sequence differences are underlined and the sequence chosen for a potential species-specific primer is highlighted.*

In much the same way that the *psts* locus was analyzed, multiple *atpa* sequences from the pubmlst *E. faecalis* database were aligned to identify alleles frequently associated with canine sources. MLST allele 15 was more often associated with bacteria obtained from dogs than humans. Therefore, primers were designed that would target only allele 15. In PCR reactions containing the *atpa*-specific primers and DNA template from sewage samples, SP1 and Mayo, and canine samples, D3 and S5, varying annealing temperatures either resulted in successful amplification in all reactions or no amplification in all reactions. Similar to the results

observed in the *psts* assays, nucleotide differences were either not present in the template or not significant enough to cause temperature-dependent differential annealing of primers at the target sites (data not shown).

FIGURE 3

Sample of the atpa sequence alignment of E. faecium amplified using the atpal15 primer set and sequenced by Genwize. Nucleotide variations are underlined and highlighted.

Sequencing of a subset of amplification products revealed that only the product from bacterial sample P002 with primer set atpa15 contained a nucleotide difference (Figure 3). Interestingly P002 was also the only one of the 7 sequences amplified with the psts11 primer set that contained nucleotide differences (data not shown).

DISCUSSION

Determining bacterial concentration in a water sample, important to determining safety for recreational use, does not provide information on the source of contamination. Consequently, considerable effort has been made developing MST methods (Meays et al. 2004). *Enterococcus sp.* have emerged as the recommended FIB for both fresh and brackish waters, making them a convenient target for this study since samples identified as having high

FIB concentration can be targeted for MST without the need for collection of an additional sample, without the need for collection of a larger sample, and without the wasted effort of processing a sample that is later found to lack contamination.

Starting with primers previously reported to have specificity for one, or a subset of FIB species (Table 1), studies were undertaken to assess the feasibility of similar studies with samples from local sources including *Enterococcus* bacterial DNA from human, canine, sewage treatment facilities, and local rivers. The ENT and ENT376 primer sets target the 16s rRNA gene in *Enterococcus* faecalis and multiple *Enterococcus* species, respectively. The CIUM primer set is specific for the 16s rRNA gene in *Enterococcus* faecium, while the *esp* and *IS16* primer sets target sequences originally associated with virulence genes in virulent strains of E. faecium, but also present in some *E. faecalis* strains. In addition, reports utilizing *esp* and *IS16* primers relied on the association of their targets with bacterial samples from human clinical settings, both of which have been linked to vancomycin resistance (Werner et al. 2011; Willems et al. 2001).

Looking more closely at the amplification results in Table 3, the control type strain *E. faecalis* NCTC 775 illustrated the expected pattern of primer specificity, positive for amplification by primers specific for *Enterococcus* faecalis and multiple *Enterococcus* species, ENT and ENT376, respectively, and lacking amplification of the *esp* and *IS16* targets, associated with bacteria from clinical human samples (Mohamed et al. 2018; Scott et al. 2005; Werner et al. 2011). E. faecium 700221 genomic template resulted in amplification of the CIUM target, specific for E. faecium, and the virulence specific *esp* and *IS16* targets as expected (Table 3 and Figure 1).

The low number of individual human fecal samples (n=3) in this study complicates statistical analysis of the results. While the sewage effluent (n=8) provided a larger human population, the

material entering treatment plants does not only contain human fecal matter. Sewage influent potentially contains animal feces and chemicals that may remove some bacterial species of study. To be confident in correlations between primer specificity and human fecal sources, future studies will require additional human samples from both community and clinical settings.

Although the *ESP* and *IS16* primer sets were not able to distinguish between human and canine fecal sources with 100% selectivity, this finding is not entirely surprising. First, work by Ahmed (2008) evaluating sensitivity of the *ESP* primer set, demonstrated that about 91% of sewage and septic samples were *esp* positive with sensitivity between 67% and 100% depending on the type of sample. The findings in this study showed *esp*-positive results in 60% of human and sewage samples tested, a value not significantly lower than earlier results. Second, a recent study reported that 29% of *Enterococcus* from canine fecal samples were *esp*-positive (Stępień-Pyśniak et al. 2021). In this study 13% of canine samples were *esp*-positive. These findings suggest that *esp*-carrying *Enterococcus* strains are moving from human clinical settings to human and animal populations outside of clinical settings, which will adversely affect the success of using *esp* as a species-selective target. An increase in genetic similarities in the bacteria found in humans and pet hosts will continue to rise as we live in close proximity to each other (Song et al. 2013). Results with *IS16* were similar. In this study 60% of bacterial samples from human and sewage samples were positive for *IS16*, and 29% of samples from dogs were positive for *IS16*. While a study by Werner reported 100% sensitivity in over 100 samples obtained from humans in a clinical setting, less than 5% of samples collected outside of hospitals were positive for *IS16* (Werner et al. 2011). In another study evaluating a transposon related to the *IS16* sequence in samples from dogs, researchers proposed exchange between humans and dogs to explain canine samples positive for the transposon (Simjee

et al. 2002).

As reported in several other studies, some samples collected from canine and human fecal samples for this study demonstrated similar amplification patterns when using primers that were designed to discriminate between species (Song et al. 2013, Stępień-Pyśniak et al. 2021). These findings may make some molecular MST methods invalid in the coming years. Consequently, additional nucleotide differences need to be identified, using alignments such as those performed in this study with *atpa* and *psts* alleles (Figures 2 and 3). Future studies will include larger sample sizes to better analyze the specificity of the atpa15 and psts11 primer sets and a wider range of annealing temperatures to identify allele-specific amplification conditions. Other methods such as exploiting known single nucleotide polymorphisms (SNPs), which are changes in a single nucleotide, will be explored. Building from a study using known SNPs (Rathnayake et al. 2011), a series of primers could be designed to recognize SNPs specific to *Enterococcus* from a single species.

ACKNOWLEDGMENTS

The authors would like to thank communities around Anne Arundel County that support the Operation Clearwater monitoring program and provided funding to develop molecular methods for quantifying and identifying the source of contamination in local rivers. Thanks also goes to the AACC Biology laboratory technical staff that has supported our efforts by providing valuable assistance in finding reagents, setting up equipment and troubleshooting issues.

REFERENCES

Ahmed MO, Baptiste KE. 2018. Vancomycin-resistant enterococci: a review of antimicrobial resistance mechanisms and perspectives of human and animal health. Microb Drug Resist. 24(5):590–606. doi:10.1089/mdr.2017.0147

Cabelli VJ, Dufour AP, Levin MA, McCabe LJ, Haberman PW. 1979. Relationship of

microbial indicators to health effects at marine bathing beaches. Am J Public Health. 69(7):690–696. doi:10.2105/ajph.69.7.690

Dubin K, Pamer EG. 2014. Enterococci and their interactions with the intestinal microbiome. Microbiol Spectr. 5(6). doi: 10.1128/microbiolspec.bad-0014-2016

Foley SL, Lynne AM, Nayak R. 2009. Molecular typing methodologies for microbial source tracking and epidemiological investigations of Gram-negative bacterial foodborne pathogens. Infect Genet Evol. 9(4):430–440. doi: 10.1016/j.meegid.2009.03.004

Homan WL, Tribe D, Poznanski S, Li M, Hogg G, Spalburg E, Embden JDA van, Willems RJL. 2002. Multilocus sequence typing scheme for *enterococcus* faecium. J Clin Microbiol. 40(6):1963–1971. doi:10.1128/JCM.40.6.1963-1971.2002

Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 3:124. doi:10.12688/wellcomeopenres.14826.1

Layton, BA, Walters, SP, Lam, LH, Boehm, MB. 2010. *Enterococcus* species distribution among human and animal hosts using multiplex PCR. J Appl Microbiol. 109(2):539-547. doi: 10.1111/j.1365-2672.2010.04675.x

Leclerc, H, Mossel, DAA, Edberg, SC, Strujik, CC. 2001. Advances in the bacteriology of the coliform group: their suitability as markers of microbial water safety. Annu Rev Microbiol. 55: 201-234. doi: 10.1146/annurev.micro.55.1.201

Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. P Natl Acad Sci USA. 95(6):3140–3145. doi:10.1073/pnas.95.6.3140

McRobb E, Sarovich DS, Price EP, Kaestli M, Mayo M, Keim P, Currie BJ. 2015. Tracing melioidosis back to the source: using whole-genome sequencing to investigate an outbreak originating from a contaminated domestic water supply. J Clin Microbiol. 53(4):1144–1148. doi:10.1128/jcm.03453-14

Meays CL, Broersma K, Nordin R, Mazumder A. 2004. Source tracking fecal bacteria in water: a critical review of current methods. J Environ Manage. 73(1): 71-79. doi:10.1016/j.jenvman.2004.06.001

Okonechnikov K, Golosova O, Fursov M, UGENE team. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012;28(8):1166–1167. http://dx.doi.org/10.1093/bioinformatics/bts091

Rathnayake IU, Hargreaves M, Huygens F. 2011. Genotyping of *Enterococcus* faecalis and *Enterococcus* faecium isolates by use of a set of eight single nucleotide polymorphisms. J Clin Microbiol. 49(1):367–372. doi:10.1128/jcm.01120-10

Rose JB, Epstein PR, Lipp EK, Sherman BH, Bernard SM, Patz JA. 2001. Climate variability and change in the United States: potential impacts on water- and foodborne diseases caused by microbiologic agents. Environ Health Perspect. 109(suppl 2):211–221. doi:10.1289/ehp.01109s2211

Ruiz-Garbajosa P, Bonten MJM, Robinson DA, Top J, Nallapareddy SR, Torres C, Coque TM, CantónR, Baquero F, Murray BE, et al. 2006. Multilocus sequence typing

scheme for *enterococcus* faecalis reveals hospital-adapted genetic complexes in a background of high rates of recombination. J Clin Microbiol. 44(6):2220–2228. doi:10.1128/jcm.02596-05

Ryu H, Henson M, Elk M, Toledo-Hernandez C, Griffith J, Blackwood D, Noble R, Gourmelon M, Glassmeyer S, Santo Domingo JW. 2013. Development of quantitative PCR assays targeting the 16S rRNA genes of *enterococcus* spp. and their application to the identification of *enterococcus* species in environmental samples. Appl Environ Microbiol. 79(1):196–204. doi:10.1128/aem.02802-12

Scott TM, Jenkins TM, Lukasik J, Rose JB. 2005. Potential use of a host associated molecular marker in *enterococcus* faecium as an index of human fecal pollution. Environ Sci Technol. 39(1):283–287. doi:10.1021/es035267n

Simjee S, White DG, McDermott PF, Wagner DD, Zervos MJ, Donabedian SM, English LL, Hayes JR, Walker RD. 2002. Characterization of Tn1546 in vancomycin-resistant *Enterococcus* faecium isolated from canine urinary tract infections: evidence of gene exchange between human and animal enterococci. J Clin Microbiol. 40(12):4659-65. doi:10.1128/JCM.40.12.4659-4665.2002

Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, Caporaso JG, Knights D, Clemente JC, Nakielny S, et al. 2013. Cohabiting family members share microbiota with one another and with their dogs. eLife 2:e00458 doi: 10.7554/eLife.00458

Stępień-Pyśniak D, Bertelloni F, Dec M, Cagnoli G, Pietras-Ożga D, Urban-Chmiel R, Ebani VV. 2021. Characterization and comparison of *Enterococcus* spp. isolates from feces of healthy dogs and urine of dogs with UTIs. Animals. 11(10):2845. doi:10.3390/ani11102845

[US EPA] US Environmental Protection Agency. 2009. Method 1600.1: Enterococci in water by membrane filtration using membrane-*Enterococcus* indoxyl-β-D-glucoside agar (mEI). Washington (DC): US Environmental Protection Agency. Report No.: EPA-821-R-09-016

[US EPA] US Environmental Protection Agency. 2012. Recreational water quality criteria. Washington (DC): US Environmental Protection Agency. Report No.: 820-F-12-058

Werner G, Fleige C, Geringer U, van Schaik W, Klare I, Witte W. 2011. IS element, IS16, as a molecular screening tool to identify hospital-associated strains of Enterococcus faecium. BMC Infect Dis. 11:80-87. doi:10.1186/1471-2334-11-80

Willems RJ, Homan W, Top J, van Santen-Verheuvel M, Tribe D, Manzioros X, Gaillard C, Vandenbroucke-Grauls CM, Mascini EM, van Kregten E, et al. 2001.Variant esp gene as a marker of a distinct genetic lineage of vancomycin-resistant Enterococcus faecium spreading in hospitals. Lancet. 357(9259):853-5. doi:10.1016/S0140-6736(00)04205-7

Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J Comput Biol. 7(1-2):203–214. doi:10.1089/10665270050081478.

HANNAH CLAGGETT

The Restoration of Submerged Aquatic Vegetation in the Chesapeake Bay

KEY WORDS

Submerged Aquatic Vegetation

restoration

turbulator

seed processing

FACULTY MENTOR

Susan Lamont, Ph.D.
Professor, Biology Department

ABSTRACT

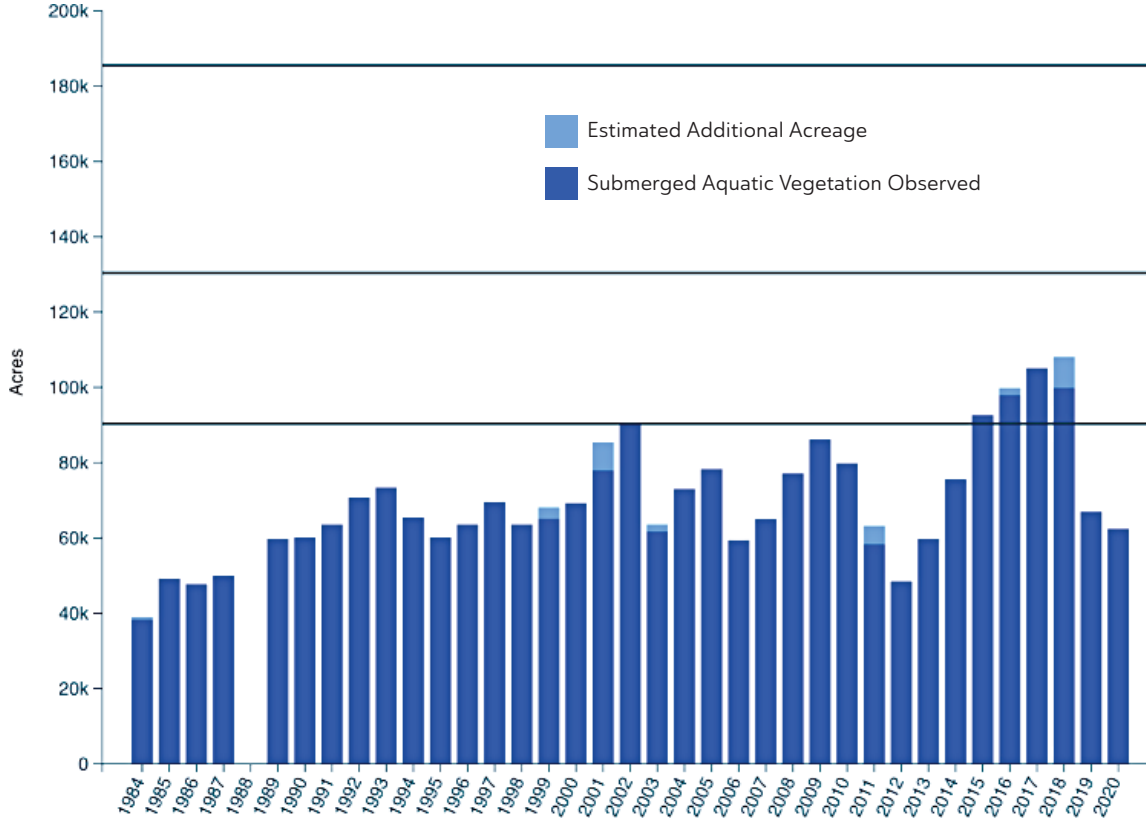
Submerged Aquatic Vegetation (SAV) is critical to maintaining water quality and providing food and shelter for numerous estuarine organisms. As part of a larger project to restore SAV in the Chesapeake Bay, the goals of this research project were to identify healthy beds for seed harvesting, harvest seeds from four native SAV species and refine the seed so that it could be stored until dispersal for restoration purposes. Through collaboration between Shore Rivers, Maryland Department of Natural Resources and the Anne Arundel Community College Environmental Center, seeds collected in the summer of 2021 will be dispersed in 2022, with the ultimate goal of restoring one acre of SAV. Four types of native aquatic plants, *Ruppia maritima* (widgeon grass), *Zannichellia palustris* (horned pondweed), *Stuckenia pectinata* (sago pondweed), and *Potamogeton perfoliatus* (redhead grass) were collected into 20 baskets per species and then processed in a turbulator to separate the seed. After turbulating, the plant material was further processed through a series of screens to refine the pure seed, which was later isolated and placed into jars with a salt solution. Over the winter, seeds were stored in the jars until they will be mixed with sand and dispersed into the bay for future restoration projects. More than 1,000,000 seeds were collected this summer from all four species combined, and over 100 hours of volunteer time went into the seed processing/refining process.

INTRODUCTION

Submerged Aquatic Vegetation (SAV) plays a crucial role in maintaining the health of the bay ecosystem. SAV is composed of a diverse collection of plant species that are located beneath the water surface and are entirely submerged except during low tide. There are over 20 species of SAV located in the Chesapeake Bay watershed (Chesapeake Bay Program, 2020). SAV beds help to absorb excess nutrients and trap particulate matter such as sand and silt that often cloud the water, suffocating and killing marine life (Chesapeake Bay Program, 2020). These beds provide shelter, habitat, and a food source for many organisms, especially waterfowl (Chesapeake Bay Program, 2020). SAV beds serve as a general indicator of the overall health of the Chesapeake Bay due to their sensitivity to water quality changes (Blankenship, 2021). When water quality improves, the abundance and quality of the

FIGURE 1

Abundance of SAV 1984–2020 (Chesapeake Bay Progress).



aquatic vegetation beds are affected positively and tend to improve (Blankenship, 2021).

After several years of continual growth in acreage of SAV beds in the Chesapeake Bay, total acreage of SAV declined 7% in 2020 – the second consecutive year of SAV decline since peaking three years ago (Blankenship, 2021) (Fig. 1). However, the presence of underwater grasses often shows trends of a boom and bust cycle, as some grasses are more sensitive to changes in water quality than others and will rapidly decline one year, but flourish the next year such as *Ruppia maritima* (Blankenship, 2021). According to Brooke Landy, a biologist with the Maryland Department of Natural Resources, "It's important to keep in mind that last year's decrease, and the decrease in 2019, didn't represent a loss of a long-term abundance and distribution, it was a decrease from a relatively recent expansion" (Blankenship, 2021). This emphasizes the importance of protecting and maintaining stable underwater grass populations.

In the Chesapeake Bay, SAV restoration planting efforts began in 1978 with whole *Zostera marina* plants, using sods, cores, or bare-root plants (Ailstock & Shafer, 2006). In the 1980's whole plant cuttings, seeds, and tubers of *Vallisneria americana* and several other low-salinity species were planted in the upper Chesapeake Bay, and in 1985 whole plants of *R. maritima* were transplanted in the mid-bay Choptank River (Ailstock & Shafer, 2006). In the past, it was most common to restore underwater grasses by harvesting the plants from suitable donor beds and transplanting them into the bay as individual shoots, shoot bundles, or sods (Ailstock & Shafer, 2006). This caused SAV restoration to be limited to small projects, typically on a scale of tens or hundreds of square meters due to the high costs and logistical constraints of this method (Ailstock & Shafer, 2006). In addition, approximately 40,500 additional hectares of SAV were needed to reach the restoration goals established by the Chesapeake Bay Program in 2003, therefore a

new restoration method had to be identified in order to establish plants at such a scale (Chesapeake Executive Council, 2003).

In 2003, the U.S. Army Corps of Engineers (USACE) Engineer Research and Development Center (ERDC) and the National Oceanic and Atmospheric Administration (NOAA) Chesapeake Bay Office began to plan and implement their respective research programs to promote the development of innovative tools and techniques for the large-scale restoration of SAV (Marion & Orth, 2010). This program represented the first coordinated interagency effort to develop, evaluate, and refine protocols suitable for large-scale SAV restoration (Shafer & Bergstrom, 2010). Since this research initiative began, an average of 13.4 ha/year of SAV has been planted in the Chesapeake Bay, compared to an average rate of 3.6 ha/year during the previous 21 years (1983–2003) (Shafer & Bergstrom, 2010). The new techniques and technologies allow submerged aquatic plants to be planted at scales that would have been unattainable with existing technologies only a few years ago (Busch, 2010). Furthermore, the costs of conducting these plantings declined with increased understanding of the limiting factors and new advances in technology development (Ganassin & Gibbs, 2008).

The most effective approach involves directly sowing seeds into suitable planting areas, a method that emerged as a viable means of planting and restoring large areas of the seagrass, *Zostera marina* (Ailstock & Shafer, 2006). Once an existing healthy, viable underwater seagrass bed is identified, fruiting plants are collected into baskets and then later processed through a turbulator to essentially "shake" the seeds off of them. After turbulating, the plants are processed and refined through a series of mesh screens until just the pure seed is left. After storing the pure seed in various containers under brackish conditions in a cold room over the winter months, the seeds are mixed with sand and redistributed into areas where SAV beds used to be prominent in Chesapeake Bay

regions throughout the spring.

The four types of native SAV that are of interest in the local region due to their frequency, abundance and diversity of tolerances and habitat value are *Ruppia maritima* (widgeon grass), *Zannichellia palustris* (horned pondweed), *Stuckenia pectinata* (sago pondweed), and *Potamogeton perfoliatus* (redhead grass). *R. maritima* tolerates a wide range of salinity, from the slightly brackish upper and mid-Bay tributaries through near-seawater salinity in the lower Bay (Maryland DNR, n.d.). *R. maritima* is notorious for disappearing in large quantities when water quality declines but tends to quickly reappear a few years later if conditions are healthy again (Maryland DNR, n.d.). *R. maritima* is most common in areas with sandy substrates, although it occasionally grows on soft, muddy sediments (Maryland DNR, n.d.). *Z. palustris* is found in every state in the continental United States, as well as in Europe and South America (Maryland DNR, n.d.). *Z. palustris* is widely distributed in Chesapeake Bay, growing in fresh to moderately brackish waters, in muddy and sandy sediments (Maryland DNR, n.d.). *Z. palustris* seems to grow most abundantly in very shallow water but may grow to depths of 5m if it receives enough light (Maryland DNR, n.d.). *S. pectinata* is widespread in the Chesapeake Bay, growing in fresh non-tidal to moderately brackish waters as well as in some lakes (Maryland DNR, n.d.). It can tolerate high alkalinity and grows on silty-muddy sediments (Maryland DNR, n.d.). Lastly, *P. perfoliatus* is typically found in fresh to moderately brackish and alkaline waters (Maryland DNR, n.d.). *P. perfoliatus* grows best on firm, muddy soils and in quiet water with slow-moving currents (Maryland DNR, n.d.).

Methods

The first step of SAV restoration was to identify large-scale vegetated beds in the Chesapeake Bay that were healthy enough to be harvested. Potentially viable beds were identified using satellite

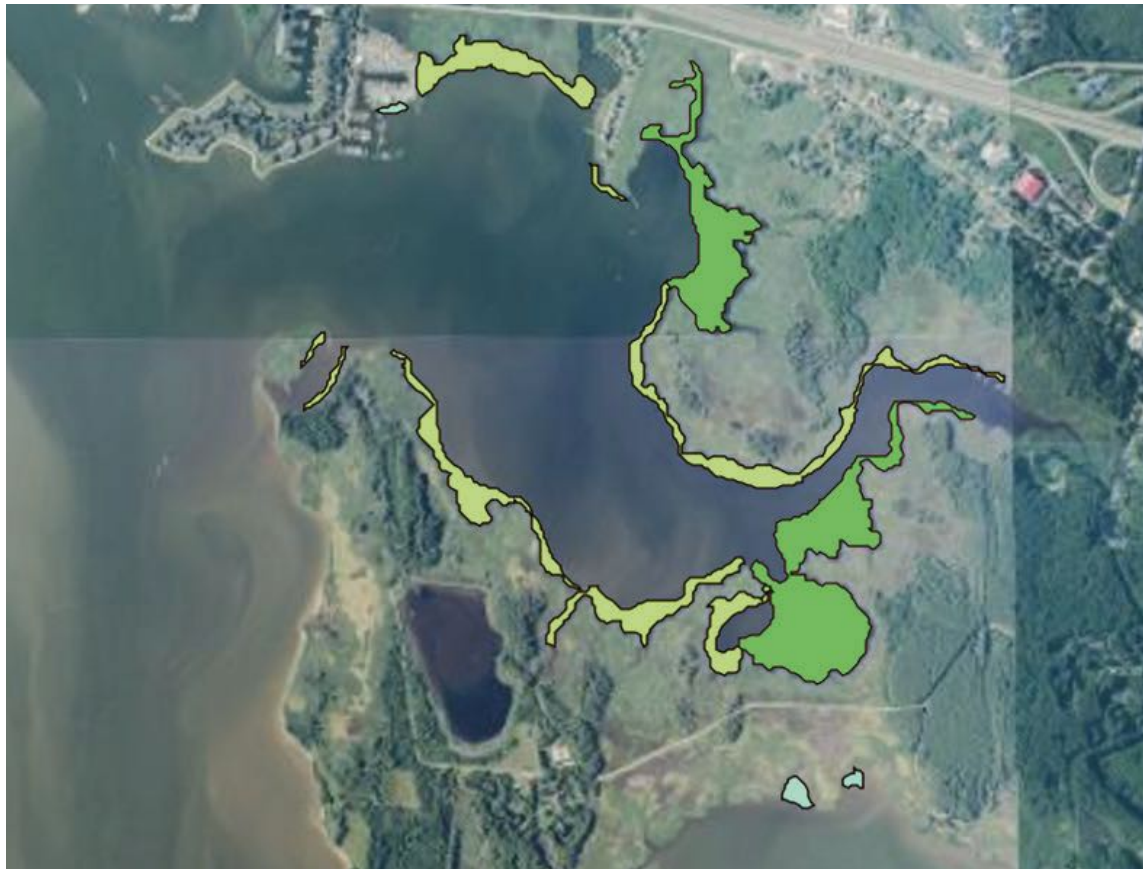


FIGURE 2

2020 Satellite Image depicting high bed density (the dark green area) found in Marshy Creek, MD (Virginia Institute of Marine Science). Light green shows lower-density beds.

imagery from the Virginia Institute of Marine Science (VIMS), and locations of nearby boat launches were recorded (Fig. 2). It was important to identify SAV beds that had high bed density because this ensures the greatest chance of finding an adequately-sized and healthy donor site.

Once a suitable donor bed was identified and an accessible nearby boat launch was found, kayaks were used to gain access to the sites to monitor the growth stage of plants in those beds (Fig. 3). Beds were deemed appropriate for collection when the majority of plants were in fruit (which contain the seeds).

When the plants were ready to collect, volunteers from Anne Arundel Community College (AACC), Maryland Department of Natural Resources, and Shore Rivers visited the identified locations by motorboat and hand collected the plants by removing the

FIGURE 3

Ripe fruits of Ruppia maritima.

upper third of viable stems and placing them into 17" round by 14-1/2" high plastic crab baskets (Fig. 4). After collecting about 20 crab baskets worth of plants, they were stored indoors in an unlit space in the AACC shed for seven days in large bins to achieve after-ripening. The plants were kept moist throughout the seven days and were occasionally churned with a metal rake in order to prevent rotting. All four species were processed twice: seven days after they were harvested, and then again after 14 days.

A turbulator was used to separate the seeds from the stems (Fig. 5). There are three turbulators in the state of Maryland and AACC has two of them. The turbulator is a large six-foot by six-foot round tank that has a series of PVC pipes with vacuums attached to run CO2 through the water and create a "jet-like" effect. These jets help to churn the plants and shake the seeds off the plants. Plants were turbulated in water for 15 minutes and then the tank was drained into a mesh bag to collect the separated seeds. Typically, 14 days after collection, seeds were processed a second

FIGURE 4

Choptank Riverkeeper, Matt Pluta, harvesting bushels of Stuckenia pectinata (Sago Pondweed) from Broad Creek.

FIGURE 5

Volunteers from Anne Arundel Community College, Shore Rivers and Submerged Aquatic Vegetation Watchers use the turbulator to separate seed from stems of Potomogeton perfoliatus (Redhead grass).

time to collect any seeds remaining on the plants after the first process.

Following the seed processing stage, seeds were refined outdoors to get them into a storage-ready state. The seeds needed to be as clean as possible with little detritus attached, as they were stored in multiple clear gallon-sized plastic jars in a walk-in refrigerator at AACC and Shore Rivers Offices. When seeds have too much extra material on them, they often begin to decompose and can easily become contaminated, and then cannot be dispersed back into the bay for restoration. Storage conditions must provide an environment that allows seeds to remain viable and dormant, since embryo death or premature germination will negate their use for restoration. Aeration during storage was also important for retaining the viability of stored seeds. Research has shown that seeds stored at 4°C with aeration have the highest germination rates (Ailstock & Shafer, 2006).

To refine the seeds, the mesh bag full of seed and detritus collected from the turbulator was emptied gradually onto a series of wire screens with decreasing mesh sizes (Fig. 6). A hose was used to spray water and push the plant material through the screens to separate the detritus from the pure seed (Fig. 7). After refining, seeds were stored in a brackish condition with aeration in a cold room at AACC with the intention to mimic the estuarine environment. Fish tank aeration

FIGURE 6

Spherical seeds of Stuckenia pectinata (Sago pondweed) with detritus attached, ready to be processed.

FIGURE 7

Processing seed through the series of mesh screens to refine it.

FIGURE 8

AACC Faculty (Tammy Domanski, left) and student volunteers distribute seeds on the Magothy River.

pump devices were placed in each jar of seed in order to prevent bacteria and algae from growing in the jars during the storage period. The storage containers were gallon-sized clear plastic jars with a screw-on cap with a hole in it, in order to allow for the aeration pump to be placed inside. All seeds will remain in the dark cold room over the winter months and will be redistributed throughout the Bay in the Spring for restoration (Fig. 8).

DISCUSSION/RECOMMENDATIONS

In 2021, AACC, in partnership with Maryland Department of Natural Resources and Shore Rivers, collected all four native plants with a goal of restoring one acre of underwater seagrass with the seeds collected. Approximately 20 baskets of each of the following species were collected: *R. maritima* was collected from Broad Creek in Talbot County, *S. pectinata* was collected from Rock Hall in Kent County, *Z. palustris* was collected from Tilghman Creek and the Wye River in Talbot County, and *P. perfoliatus* was collected from Marshy Creek in Queen Anne's County.

This project could be improved by increasing monitoring of both previously restored beds and harvested beds. Post-restoration monitoring can be a strain on organizational resources, and therefore most volunteer restoration projects do not include follow-up monitoring to determine their long-term effectiveness (Chesapeake Bay Program Submerged Aquatic Vegetation Workgroup, 2020). In addition, no long-term data has been collected analyzing the health of the harvested beds, some of which have been harvested over several successive years. Monitoring of affected beds (both harvested and restored) is necessary to determine the success rate of restoration efforts and to ensure that healthy beds are not being

jeopardized in the restoration process. In conjunction with this effort, it would be helpful to monitor water quality in the areas of restored and harvested beds. Each of the four species of interest in this region have slightly different tolerance limits and growing conditions, therefore water quality data from harvested and restored sites would provide additional information to help explain restoration success rates.

Another area of further study could include analyzing the restoration success rate per species of interest. Low transplant survival and seedling establishment rates at the large-scale planting sites within Chesapeake Bay suggest that current site selection criteria are either not stringent enough or are incomplete, due to a lack of understanding of factors influencing both seed germination and seedling establishment (Shafer & Bergstrom, 2010). Ideally, a series of germination tests would be performed on each seed type collected in order to determine seed viability per species. In addition to lab-based germination tests using terrestrial substrate, an aquatic germination test should be conducted as well. The underwater planting environment differs substantially from terrestrial systems in that conditions such as light and nutrient availability and sediment stability are much less predictable (Koch, 2001). This would provide more information about the specific conditions that support high germination rates for each species.

Proper seed storage conditions also deserve further research, as there is a lack of data in regards to storing seeds with detritus attached. There are currently three methods used to store and disperse seeds for restoration projects involving all species (Ailstock & Shafer, 2006). Two require either no storage or temporary storage under the ambient conditions to which wild populations are generally exposed (Ailstock & Shafer, 2006). The third method focuses on long-term storage, which enables seed availability whenever they are needed (Ailstock & Shafer, 2006). With the possible exception of such plants as *Zostera marina* and *Thalasia testudinum*,

information on the variation in storage and germination requirements of the seeds of most underwater grasses is sparse (Ailstock & Shafer, 2006).

These questions remain unanswered because they require an immense amount of resources and volunteer time. In order to collect the necessary data, a large volunteer base is needed to consistently monitor and analyze beds as well as perform lab tests over multiple years. Collaborations between local nonprofits (eg. Shore Rivers), state agencies (eg. Maryland DNR), and academic institutions (eg. AACC Environmental Center faculty, staff and students) provide a great opportunity to seek the answers to these questions.

ACKNOWLEDGEMENTS

Jose Barrata, Coordinator of STEM Initiatives, provided the funding opportunity that enabled Hannah Claggett's participation in this project, through a Louis Stokes Alliances for Minority Participation (LSAMP) grant. Michael Norman, Lab Manager of the Biology Department, spearheaded AACC's involvement in the project via a grant from MDDNR and served as supervisor on the project. Eastern shore Riverkeepers, including Ellie Bassett, Zack Kelleher, Annie Richards and Matt Pluta, and DNR staff member Mark Lewandowski helped to organize, recruit volunteers for and participate in the seed harvesting and refining.

REFERENCES

Ailstock, Steve and Deborah Shafer. 2006. "Protocol for large-scale collection, processing, and storage of seeds of two mesohaline submerged aquatic plant species." *SAV Technical Notes Collection*: ERDC/TN SAV-06-3. https://apps.dtic.mil/sti/pdfs/ADA454247.pdf

Bergstrom, Peter. 1998. "SAV Hunter's Guide for Chesapeake Bay." *The Volunteer Monitor 10*(2): 17

Blankenship, Karl. 2021. "Chesapeake Bay grass beds declined for the second year in a row." *Bay Journal*, July 29, 2021. https://www.bayjournal.com/news/fisheries/chesapeake-bay-grass-beds-decline-for-second-year-in-a-row/article_e6f-097fa-e568-11eb-a573-9766d206b1a9.html

Busch, Katheryn, Rebecca Golden, Thomas Parham, and Lee Karrh. 2010. "Large-Scale *Zostera marina* (eelgrass) Restoration in Chesapeake Bay, Maryland, USA. Part I: A Comparison of Techniques and Associated Costs" Restoration 18, no 4: 490-500 DOI:10.1111/j.1526-100X.2010.00690.x

Chesapeake Bay Program. 2020. "Underwater grasses." Learn the Issues. https://www.chesapeakebay.net/issues/bay_grasses.

Chesapeake Bay Program. Submerged Aquatic Vegetation Workgroup. 2020. "Submerged aquatic vegetation (SAV)." Chesapeake Progress. https://www.chesapeakeprogress.com/abundant-life/sav

Chesapeake Executive Council. 2003. "Strategy to accelerate the protection and restoration of submerged aquatic vegetation in the Chesapeake Bay. United States Environmental Protection Agency Chesapeake Bay Program, Annapolis, MD. 18 pp. http://www.chesapeakebay.net/content/publication/cbp_12608.pdf.

Ganassian, C and P.J. Gibbs. 2008. "A review of seagrass planting as a means of habitat compensation following loss of seagrass meadow" Fisheries and Research Development Corporation (Australia) & New South Wales. Department of Primary Industries, Fisheries Final Report Series, No. 96. ISSN 1449-9967

Koch, Eva Maria, Steve Ailstock, Deborah Shafer, Dale Booth, and Dale Magoun. 2010. "The Roles of Current and Waves in the Dispersal of Submersed Angiosperm Seeds and Seedlings" 2003–2008." *Restoration Ecology* 18, no 4: 584-595 DOI:10.1111/j.1526-100X.2010.00698.x

Marion, Scott and Robert Orth. 2010. "Innovative Techniques for Large-scale Seagrass Restoration Using *Zostera marina* (eelgrass) Seeds" Restoration 18, no 4: 514-526 DOI:10.1111/j.1526-100X.2010.00692.x

Maryland Department of Natural Resources. n.d. "Submerged aquatic vegetation (SAV) identification key." Learning Resources. https://dnr.maryland.gov/waters/bay/Pages/sav/key.aspx?savname=Redhead+Grass.

Shafer, Deborah and Peter Bergstrom. 2010. "An Introduction to a Special Issue on Large-Scale Submerged Aquatic Vegetation Restoration Research in the Chesapeake Bay: 2003–2008." *Restoration Ecology* 18, no. 4: 481-489. https://doi.org/10.1111/j.1526-100X.2010.00689.x

U.S. Fish and Wildlife Service, Chesapeake Bay Estuary Program, 1992. *Field Guide to the Submerged Aquatic Vegetation of the Chesapeake Bay*, by Linda M. Hurley.

Virginia Institute of Marine Science. 2020. "SAV Program: Monitoring and Restoration." Research & Services. https://www.vims.edu/research/units/programs/sav/

Virginia Institute of Marine Science. n.d. "Interactive SAV map." Research & Services. https://www.vims.edu/research/units/programs/sav/access/maps/index.php?showLayers=SAV_Base_Layers_2504.

Mollie Crossman and Jeremy Snyder

Optimizing Quantitative PCR to Distinguish Between Human and Canine Bacterial Samples

KEY WORDS

fecal indicator bacteria

Bacteroides

16s rRNA

quantitative PCR

microbial source tracking

FACULTY MENTOR

Tammy Domanski, Ph.D.
Professor, Biology Department
Director, Environmental Center at
Anne Arundel Community College

ABSTRACT

High bacterial levels in recreational bodies of water can be a risk to human health, and significant effort and funding are invested in monitoring fecal indicator bacteria (FIB) levels. Despite the health risk, methods commonly utilized to determine bacterial concentrations provide no information about the source of contamination. This study assesses the feasibility of utilizing quantitative polymerase chain reactions (qPCR) to perform microbial source tracking (MST) that will identify the source of fecal contamination in rivers and streams in Anne Arundel County. Human and canine fecal bacterial DNA samples were analyzed using primer sets previously reported to target genes frequently identified in host-specific bacterial species. Primers specific for *esp*, encoding the enterococcal surface protein often associated with human fecal bacteria in clinical settings, and primers specific for the *Bacteroides* 16s rRNA genes, either specific to bacterial genomes from canine or human sources, were utilized. Quantitative polymerase chain reaction (qPCR) analysis demonstrated that, while the primer sets successfully amplified target sequences, there was some amplification of non-target sequences within the target host, and some amplification of genes in samples from non-target hosts, such as amplification of sequences in dog bacterial DNA by human bacterial-specific primers. Gel electrophoresis and DNA

sequencing of sample qPCR products were conducted and confirmed that target genes were amplified, although the identity of some of the off-target products remains to be determined. Primers reported to target sequences in bacteria from dog feces showed higher specificity, but still resulted in some off-target amplification. On-going work includes optimizing assay conditions and primer sequences to increase specificity and reducing potential sources of reaction contamination which may be contributing to some off-target results.

Introduction

Fecal contamination in environmental waters intensifies the human health risk of infection from waterborne pathogens. These pathogens originate not only from human fecal sources, but also from the feces of other mammals and some birds. Human-compatible pathogens are particularly prevalent in the feces of domestic pets, with one study estimating 39.1% of human pathogens being able to infect domestic animals (Green, White et al. 2014). Feces of domestic animals are also likely found in higher proportions than feces of wildlife in environmental waters, as the disposal of feces from pets and domestic animals are typically left to the owner discretion. This increased likelihood of cross infection and higher proportion of human-compatible infection sources accentuates the need to detect not only the presence of waterborne pathogens from fecal contamination, but also the source of feces.

Due to the wide variety of waterborne pathogens, particularly in environmental waters with fecal contamination, it is unreasonable and unrealistic to attempt monitoring all waterborne pathogens. As such, environmental water samples have historically been tested for one or more species of bacteria that are unlikely to be found in water absent of fecal contamination. For example, the Environmental Protection Agency (EPA) recommends testing recreational swimming waters for fecal coliforms, *Escherichia coli (E.*

coli) or *Enterococcus sp.*, with enterococcal testing more preferable in marine and brackish waters (US EPA, 2012). While enumeration of bacteria can be informative and guide decisions about the safety of contact with recreational waters, microbiological assays do not provide information about the specific source of the bacteria.

Methods have been developed, with varying success, to determine the bacterial source of contamination using library-dependent methods such as antibiotic resistances, bacteriophage sensitivity, pulsed-field gel electrophoresis and biophysical characteristics (Simpson et al. 2002; USEPA, 2005); however, many of those methods are labor-intensive, requiring twenty-four hours for bacterial culturing and collection of many samples from species known to most likely cause contamination in a given area (Harwood et al. 2013). Library-independent methods rely on the identification of sequences within a given species that have identifiable nucleotide variation dependent on the source organism. This process is referred to as molecular microbial source tracking (MST) and predominantly utilizes qPCR technology employing tagged primers and probes that are measured at each cycle in an amplification reaction. This makes it possible to compare the concentration of template DNA between samples and more precisely differentiate relative levels of contamination from specific sources. Some studies have suggested that *Bacteroides*, while not recommended as an indicator for contamination with microbiological methods due to the inability to culture the anaerobes, may be the best organism for molecular MST due to identified host-specific sequences (Layton et al. 2006).

While a number of studies have proposed species-specific primer sets, the method is far from standardized (Harwood et al. 2013). There are multiple organisms and gene targets proposed, some studies have not been field tested, and those that have been field tested have suggested that sequence specificity may also be region-specific (Harwood et al. 2013). In addition, the EPA has

released two methods for quantifying fecal contamination in water. One quantifies total *Enterococcus* concentration (US EPA, 2015) and the second quantifies human-specific fecal contamination (USEPA, 2019). Both procedures require costly, assay-specific reagents. This project assesses the feasibility of qPCR for molecular MST to not only determine if the source is human, but to identify other sources, specifically starting with pet waste, predicted to be responsible for up to 46% of fecal contamination in the waters around Anne Arundel County (TMDL Plan, 2017). Significant progress was made toward that goal, and further studies will expand use of additional primer sets and probe types.

Materials and Methods

DNA template samples

Canine fecal samples were obtained from local veterinarians and dog owners. Individual human samples were obtained from anonymous volunteers. Bacterial DNA from canine feces (BDCanine) (21D1 through 21D7; n=7) and bacterial DNA from human feces (BDHuman) (21p001 through 21p004, p003, p004; n=6) were isolated using a Zymo Quick Fecal/Soil Microbe kit from approximately 0.1 g of fecal matter. DNA concentrations were determined by absorbance at 260 nm and DNA quality was determined by 260/280 ratio.

Quantitative PCR

Assays were conducted utilizing SYBR green chemistry, specifically utilizing the qPCRBio SyGreen Blue Mix (PCR Biosystems). Unless otherwise stated, 20 µl reactions included 1X SyGreen Blue Mix (3 mM MgCl2 final concentration), 400 nM forward and reverse primers, and 10-73 ng DNA. Primer sets with previously reported specificity for host organisms were utilized (Table 1). Primers were obtained from Integrated DNA Technologies (Coralville, USA).

Primer 'Set'	Target gene	Fecal source species targeted	Forward primer	Reverse primer	Reference
Human 1	Bacteroides 16s rRNA gene	human	HF183F	BFDrev	Haugland et al. 2010
Human 2	Bacteroides 16s rRNA gene	human	HF183F	BacR287	Green, Haugland et al. 2014
Dog	Bacteroides 16s rRNA gene	dog	DG3F	DG3R	Green, White et al. 2014
Esp	enterococcus *esp* gene	Human from clinical setting	*Esp*F	*Esp*R	Ahmed et al. 2008

TABLE 1

Molecular microbial source-tracking (MST) primers utilized in this study.

Conditions for qPCR

The samples were analyzed in a mic PCR instrument (Biomolecular Systems, Sydney, Australia) unless otherwise stated. The conditions for the qPCR were as follows: hold Steps- hold at 95℃ for 3 min; cycling- 1) 95℃ for 10 sec 2) 60℃ for 30 sec acquiring on green; melt on green- hold at 95℃ for 15sec, hold at 60℃ for 60sec, melt from 65℃ at 0.15 C/sec. Threshold values were automatically assigned by the instrument for each assay. The Cq value is defined as the cycle at which a reaction's fluorescence reaches the threshold, and the lower the Cq value, the higher the number of DNA targets in the template sample.

Agarose gel electrophoresis

qPCR products were analyzed on 1.5% agarose and stained with ethidium bromide. Fragment size was estimated by comparison to two standards (100 bp and 1kb, EZvision, Amresco).

Gene sequencing and analysis

Select qPCR products were sequenced by Genewiz (genewiz.com) on both strands using the primers listed in Table 1. The raw

sequences from both strands were manipulated in DNA Subway (dnasubway.cyverse.org) to trim low quality ends, align the complementary strands from each sample, and trim each result to a consensus sequence. To assess specificity of each primer set NCBI Nucleotide BLAST (blastn) analysis was performed.

RESULTS

Host Specificity of primers

Initial qPCR assays were conducted utilizing primers previously reported to have specificity for BDCanine (DG3F and DG3R) and BDHuman (HF183F, HFDrev and HFBacR) using qPCR conditions recommended for the SyGreen Blue reagents.

Human-specific primers

Two reverse primers were used in the analysis of BDHuman in order to compare the validity of the results based on the qPCR practices available. The HF183F forward primer has been historically used with the HFDRev reverse primer to specifically amplify the 16s rRNA gene in the genus *Bacteroides* (human 1). A research paper reported that the HFBacR reverse primer improved specificity (human 2) (Green et al., 2014) so reactions were run separately with the human 1 and human 2 primer sets for comparison. While the average Cq value for reactions using BDHuman were almost identical when comparing human 1 and human 2 primers (Cq average = 23.0), the human 2 primer set resulted in more variation as determined with standard deviation (SD) calculations (SD= 3.7 and 9.0, respectively) (Figure 1). Consequently, the human 2 primers were not included in later assays.

When comparing the average Cq values in assays performed with the other primer sets, the *esp* and dog primer sets produced consistently higher Cq values than the human 1 primers (Figure 2). The average Cq value for all reactions with dog primers and BDHuman was 26.5 and the average with *esp* primers was Cq of

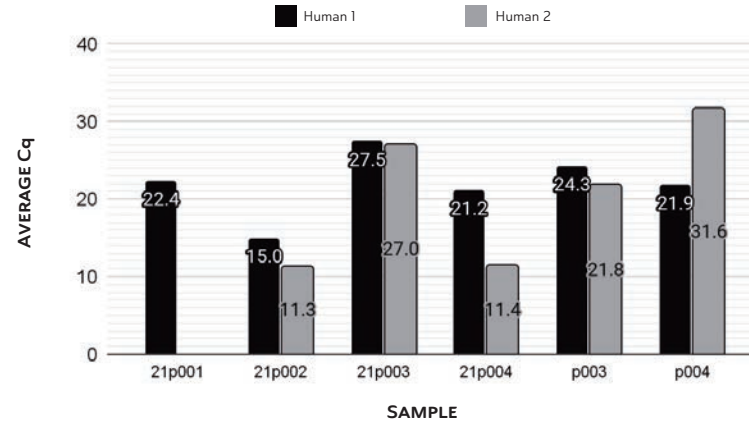


FIGURE 1

Average Cq value of each BDHuman with the human 1 primers versus the human 2 primers. Each sample was run in triplicate (n=3). No amplification was detected in the reaction with 21p001 and human 2 primers.

28.9. The SD with *esp* primers was 2.6, overlapping the average from the dog primers (SD= 3.8). Higher Cq values in assays with BDHuman and dog primers is encouraging. The dog primers' high Cq values indicate they are specific for something not found in the BDHuman. The *esp* primers resulted in Cq values similar to those with dog primers and BDHuman.

The largest SD from the triplicate samples using the dog primers was 0.5, which does not place the Cq values within the range of any of the data collected using the human 1 primers. The average Cq values of the dog and *esp* primers were similar, and were closest with the 21p002 sample (Figure 2). The dog primers exhibited much lower SD values, such as 0.2 for the 21p002 sample, which did not fall in the range of the *esp* results.

Overall, in assays with BDHuman, the human 1 primers resulted in a Cq average significantly lower than the Cq average with dog primers (p=0.0011 in two-tailed t-test) and significantly lower than the Cq average with *esp* primers (p=2.1 x 10^{-8}). Of note, the dog primer Cq average was significantly lower than the *esp* primer Cq average (p=0.022).

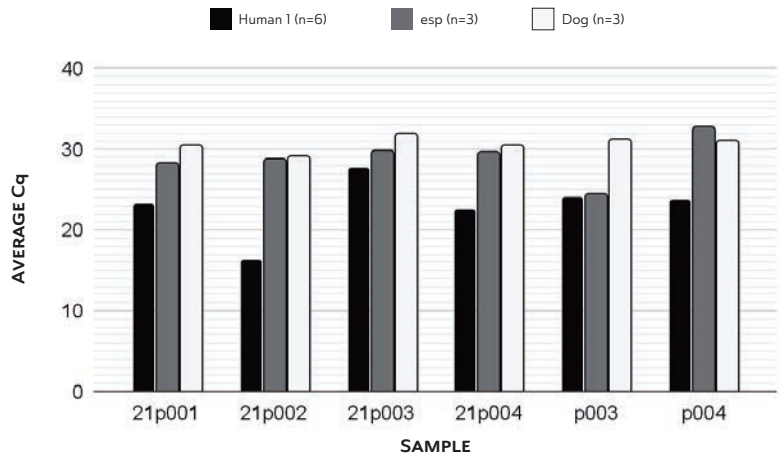


FIGURE 2

Average Cq values for each primer set. The human 1 and esp primers were specific for BDHuman and the dog primers were specific for BDcanine.

Dog-specific primers

Assays performed using the dog primers with BDCanine samples resulted in Cq values consistently lower than assays containing human-specific primers with the same samples, indicating a greater specificity for BDCanine (Figure 3). Across assays using BDCanine with dog primers, the Cq average was about 18 (SD=2.5). Cq values for each BDCanine were also consistent across assays, with the SD per sample reaching a maximum of only about 2 cycles. Assays using human 1 primers with BDCanine and dog primers with BDHuman produced high average Cq values, 27 cycles and 31, respectively, confirming that each primer set did not efficiently amplify bacterial DNA from non-target species. In a two-tailed t-test the Cq average in assays with the dog primer set was significantly lower with BDCanine (Cq average=17.8) when compared with BDHuman (Cq average=30.8) ($p=1.14 \times 10^{-36}$).

Specificity of primers for target genes

Analysis of qPCR products. To examine the specificity of the primers in targeting specific genes, two 1.5% agarose electrophoresis

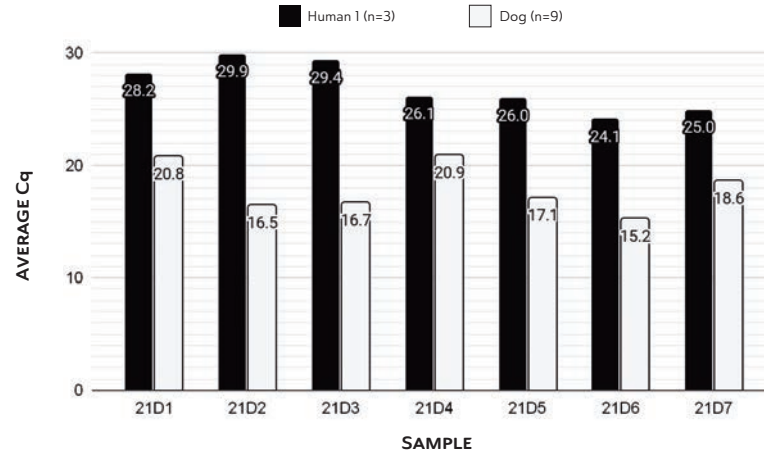


FIGURE 3

Average Cq values of each sample of BDCanine with human 1 primers versus dog primers. The human 1 data represents the results of one assay (n=3), while the dog data represent the results of three assays (n=9).

gels were run using the qPCR products from promising assay samples (Figure 4). Gel A contained qPCR products from assays with BDCanine template and gel B contained products from assays using BDHuman template. The single bands in lanes two and three of Figure 4a, containing the 21D6 product and its first dilution from a dilution series assay, indicate a highly specific reaction, while the additional less prominent bands in lanes four and five, containing the 21D2 and 21D4 products from the 15-Oct assay, indicate a much less specific reaction. This difference in specificity for gel A is to be expected, as the dilution series assay used canine-specific primers (dog) with BDCanine and the 15-Oct assay used human-specific primers (human 1) with BDCanine.

Similarly, in Figure 4b, lanes two and three contain 21p003 and P003 products from the 11-Nov assay using the *esp* primers. Both reactions resulted in two bands. The *esp* primers target the *esp* gene, previously reported to be present only in *Enterococcus* from human sources in clinical settings (Ahmed et al. 2008). The presence of multiple bands on the agarose gel suggests that the *esp*

primers are not necessarily specific to one target. Lanes four and five contain 21p002 and P004 products from the 29-Oct assay with human 1 primers. The results of the gel show that in lane four there is a band of the expected size (167 base pairs). Lane five contains multiple products, suggesting that the primers were not specific to one target. Lanes six and seven contain products from reactions containing 21p003 and 21p002 templates from the

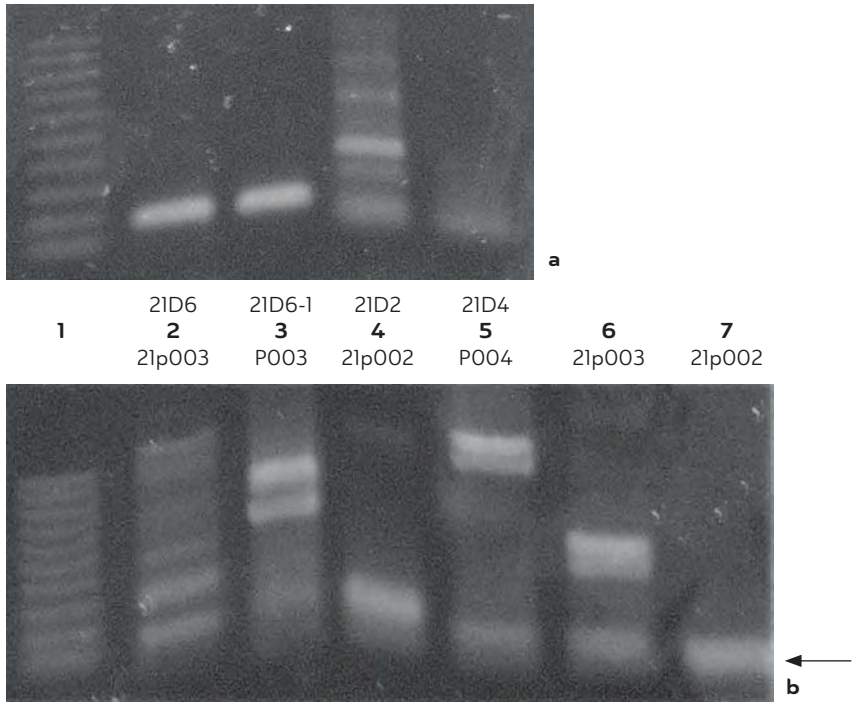


FIGURE 4

1.5% agarose gel electrophoresis. In both gels, lane 1 contains a 100-bp ladder, and the remaining lanes contain qPCR product samples. Gels were stained with ethidium bromide. 4a) *qPCR products from reactions containing BDCanine as indicated and dog primer set: lane 2 = 21D6, lane 3 = 21D6-1, lane 4 = 21D2, lane 5 = 21D4.* 4b) *qPCR products from reactions containing BDHuman as indicated and either esp primers (lanes 2 and 3), human 1 primer set (lanes 4 and 5), or human 2 (lanes 6 and 7): lane 2 = 21p003, lane 3 = P003, lane 4 = 21p002, lane 5 = P004, lane 6 = 21p003, lane 7 = 21p002. The arrow next to* 4b *indicates the 126 bp band.*

8-Oct assay with human 2 primers. The expected band size for DNA with these primers is 126 base pairs and was present in both lanes. Lane six also contained a larger band likely due to non-specific binding at non-target sequences.

Sequencing of amplification products from each primer set with each type of template was conducted (BDHuman and BDCanine) to confirm successful amplification of the target sequence. The qPCR products that displayed greater specificity in the agarose electrophoresis gel runs were sent for sequencing. Reactions containing multiple products were expected to produce poor quality sequence, confirmed by the poor quality of sequence obtained from the P003 with *esp* primers sample. Specifically, lanes two, three, and five from the gel in Figure 4a, containing reaction products from samples 21D6 and 21D6-1 with dog primers and sample 21D4 with the human 1 primers, and lanes four and seven from the gel in figure 4b, containing reaction products from sample 21p002 with the human 1 primers and with the human 2 primers. Lane three of the gel in figure 4b, containing sample P003 with the *esp* primers, was also sent for sequencing even though there were extra bands present.

The DNA sequences were manipulated and analyzed with Cyverse's DNA Subway. Poor quality reads near the ends were trimmed and the forward and reverse strands from each sample were paired to find a consensus of high confidence. Each consensus sequence underwent a BLAST search (blast.ncbi.nlm.nih.gov/) to determine the identity of the gene amplified in qPCR reactions. Sequence manipulation and an example outcome are illustrated in figure 5.

The product from the reactions with 21D6 templates, undiluted and -1 dilution, and canine specific DG3 were nearly identical, with the only difference being an additional three bases on the undiluted sequence. As such, the BLAST searches produced nearly identical results, matching *Bacteroides* sequences, coinciding

>21p002-HF183
NNNGNNTTTCGGTAGACGATGGGGATGCGTTCCATTAGATAGTAGGCGGGGTAACGGCCCACCTAGTCAACGA
TGGATAGGGGTTCTGAGAGGAAGGTCCCCCACATTGGAACTGAGACACGGTCCAAACTCCTACGACACGGTCC
AAACTCCTACGACACN
>21p002-HFDrev
NNNNNNNNNGNNNNGTTCGTCTGNNANCCCTATCCATCGTTGACTAGGTGGGCCGTTACCCCGCCTACTATCTA
ATGGAACGCATCCCCATCGTCTACCGGAAAATACGTTTAATCATGCGGACATGTGAACTCATGNN

a

HF183: CGGTAGACGATGGGGATGCGTTCCATTAGATAGTAGGCGGGGTAACGGCCCACCTAGTCAACGATGGATAGGG

HFDrev: CGGTAGACGATGGGGATGCGTTCCATTAGATAGTAGGCGGGGTAACGGCCCACCTAGTCAACGATGGATAGGG

Consensus
CGGTAGACGATGGGGATGCGTTCCATTAGATAGTAGGCGGGGTAACGGCCCACCTAGTCAACGAT
GGATAGGG

b

Bacteroides dorei strain 8642 16S ribosomal RNA gene, partial sequence

Sequence ID: MT464394.1 Length: 1416 Number of Matches: 1

Range 1: 179 to 251 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
135 bits(73)	3e-28	73/73(100%)	0/73(0%)	Plus/Plus

```
Query  1    CGGTAGACGATGGGGATGCGTTCCATTAGATAGTAGGCGGGGTAACGGCCCACCTAGTCA  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  179  CGGTAGACGATGGGGATGCGTTCCATTAGATAGTAGGCGGGGTAACGGCCCACCTAGTCA  238

Query  61   ACGATGGATAGGG  73
            |||||||||||||
Sbjct  239  ACGATGGATAGGG  251
```

c

FIGURE 5

Sequence manipulation and analysis. The amplification product from human fecal bacterial DNA amplified with the human 1 primer set was loaded into the DNA Subway-Cyverse data workspace for trimming and consensus identification. 5a) *Forward (21p002-HF183) and reverse (21p002-HFDrev) were paired and initial system-generated trimming performed.* 5b) *After the Cyverse system generated the reverse complement of HFDrev, the two sequences were aligned and the consensus, sequence of complete identity chosen.* 5c) *The best match alignment generated by the NCBI nucleotide BLAST system (blast.ncbi.nlm.nih.gov/Blast.cgi).*

with the reported specificity of the DG3 primer set (Green, White, et al., 2014). The other sequenced canine product 21D4 from a reaction with the human 1 primer set resulted in the closest match

to uncultured microorganisms containing the 16S ribosomal RNA gene. However, exact sequence matches were also found for a number of *Faecalibacterium prausnitzii* strains, which also contain the 16S rRNA gene. The human 1 primer set is intended to target the 16S rRNA gene (Haugland et al. 2010]), so these results are not surprising.

The 21p002 sequence resulted in a 73 base pair, high quality consensus that was used in a BLAST search and matched the *Bacteroides* 16s rRNA gene, as expected. The 21p003 product (human 2 primers) was of poor quality, but a 25 base pair sequence was put through a BLAST search and was a match to the 16s rRNA gene in *Bacteroides*. The results of this search showed matches to the same gene in multiple species, specifically *Bacteroides dorei* and *Bacteroides vulgatus*, whereas the human 2 primer set resulted in matches only to *Bacteroides dorei*.

The sample of p003 product with the *esp* primers was sent for sequencing even though it had two bands present in the agarose gel. The sequences were of low quality with high background which suggested that there may be multiple DNA species present, and was expected based on the agarose gel results. A portion of the consensus was used in a BLAST search and matched the *esp* gene in *Enterococcus faecium* and *Enterococcus faecalis*, which are both known to inhabit the human gastrointestinal tract.

Determination of standard curve for dog-specific primers

A standard curve was created using the dog primers and a dilution series of BDCanine sample 21D6. This sample had the lowest average Cq value, at about 15 cycles with a low SD across replicates and assay dates and was thus chosen for the dilution series (Figure 6). Undiluted and six ten-fold dilutions of the sample were used in the assay. Dilutions greater than 10^4 did not result in amplification. The average Cq values for the other samples, undiluted through 10^4 were approximately 14, 18, 22, 27, and 29 cycles, respectively.

Examining the average Cq values as a function of the sample dilution displays a linear relationship, Cq=3.85x+14.4 (r^2 =0.996), consistent with an expected slope of a bit greater than 3.3 which reflects the exponential nature of the amplification process.

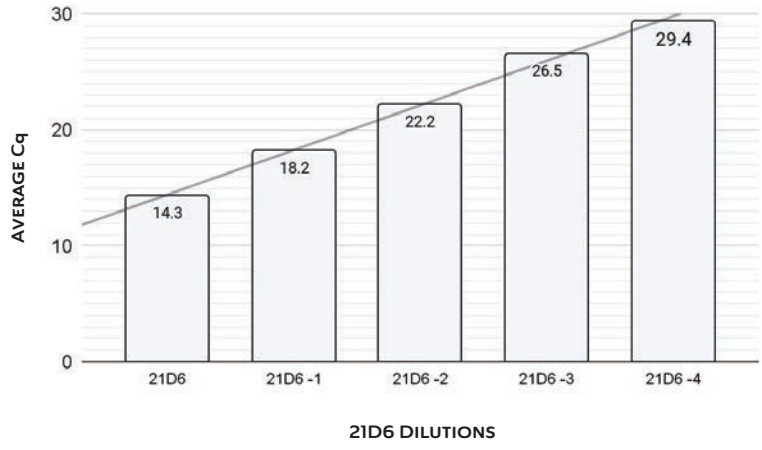


FIGURE 6

Average Cq values (n=3) for each dilution of sample 21D6 with dog primers.

DISCUSSION

Based on the assays conducted and analyses performed, the dog primer set shows promise in differentiating between BDCanine and BDHuman samples. The dog primer set produced significant differences in Cq value for known positive and known negative samples, indicating that it was selectively amplifying BDCanine target sequences. The dog primer set also produced consistent Cq results, with a maximum SD of 1.9 cycles for canine samples across assays and SDs of less than 0.6 cycles for all assays using the dog primer set. These together indicate that the dog primer set is a good candidate for further testing and eventual use in the monitoring of environmental waters. Further testing will involve the use of specific positive controls, such as synthetic plasmids of targeted DNA sequences rather than BDCanine, and the use of alternate probes, such as the Taqman probes used in previous

studies (Green, White, et al., 2014), in order to further improve the specificity of the primer set.

Results from assays with BDHuman and human-specific primers suggest that this procedure for identifying BDHuman samples is promising. The average Cq values of all results from each primer showed that the two human specific primer sets, human 1 (n=36) and human 2 (n=18), resulted in near-identical Cq values of 23.0 cycles. Both primer sets amplified the bacterial DNA with reasonable values across multiple different assays with different DNA samples. The specificity of the two primer sets were compared by looking at the average SDs for the same total number of assays (n=18). The human 1 primers were more precise, with a SD of 0.9 compared with the human 2 primers, SD of 9.5. In light of the results in this study, which did not agree with a previous report (Green, Haugland et al. 2014), the human 1 primer set was utilized in most assays. In addition, the Green study discussed the increased likelihood of very small primer-dimer formation when using the human 1 primer set. Our study did not show significant primer-dimer formation, which would result in bands smaller than the smallest band in the 100 bp standard (Figure 4). The study by Green et al. utilized samples from a wider range of species, including chickens, cattle, cats and deer, a slightly larger human sample population (n=6), and a large number of samples from wastewater facilities from across the United States (n=54) (Green, Haugland et al. 2014), which likely reflected more overall sequence variation. Future studies in our laboratory will include a wider range of species and wastewater samples, although our focus will be on methods with the highest specificity for strains in Anne Arundel County.

While this study shows great promise for use of qPCR for molecular MST, there are several areas that will benefit from method and technique optimization. In multiple experiments containing primers specific for one species and template from

the other species amplification products were detected, although those reactions were intended as negative controls and predicted not to amplify any product. Sequencing revealed that the expected target gene was amplified (figures 2 and 3 and BLAST results). These results suggest that either amplification conditions need to be altered to increase specificity of the primers, or that the primer sequences need to be optimized to decrease non-specific annealing. Although in all cases such reactions resulted in significantly higher Cq values than reactions containing the matched primers and template, confidently distinguishing between contamination sources will require optimization to more clearly differentiate between source species. The goals in this type of assay are low Cq values for assays containing target species' DNA templates, and the absence of amplification, so no Cq value in assays containing non-target species' DNA.

In rare cases, water only controls, containing no template DNA, resulted in amplification. These positive results were typically only found in one of the sample triplicates and not reflected in the other two triplicates of the sample primer and sample combination, indicating the probability of intermittent cross-contamination over non-specificity of the primer set. As cross-contamination can easily invalidate the results of an assay, the techniques used in future assays must be improved to remove all potential for cross-contamination. Additional improvements include a more objective and explicit definition of what constitutes a positive result, rather than a relative comparison between assay samples, and the use of more specific positive controls such as synthetic plasmids of targeted DNA sequences, to ensure that primers are amplifying targeted DNA sequences rather than unintended sequences present in less specific samples.

Other method improvements will include increased use of bleach to clean the area and instruments when working with the different samples of DNA. Currently, the bacterial samples from

human and dog fecal samples are handled in separate laminar flow hoods with separate instruments. As much as space allows, samples from different species will be handled in different rooms and at different times. The added control would prevent contamination between the different samples from the same sources.

Other primer sets of interest from other studies conducted by our laboratory (unpublished data) were employed in this study. The presence of the *esp* sequence was assessed in human fecal bacterial samples with the *esp* primer set. The *esp* primers have been reported to be specific for strains of *Enterococcus* isolated from hospitalized patients (Ahmed et al. 2008). The human 1 primer set is specific for sequences in *Bacteroides*, but since DNA was isolated from total bacteria collected from fecal samples, *Enterococcus* would also be expected in the sample. However, because the human fecal samples analyzed in this study were not from clinical settings, the lack of a strong positive result, a low Cq value, is not unexpected. A good positive control and samples from clinical settings are necessary to pursue further use of the *esp* primer set.

Moving forward the next steps in this project include obtaining synthetic positive control plasmids that will be diluted to known copy number to produce a standard curve precise enough for determination of sequence copy number in samples (USEPA, 2015), adding Taqman probes to our analysis for comparison, and obtaining complex samples such as influent from wastewater reclamation facilities and environmental samples from local rivers both after rain events, when bacterial concentrations are high and during dry periods when concentrations are low. Taqman chemistry utilizes target specific, internal probes that may be more specific and less prone to amplification of non-target sequences, although some studies have shown that careful optimization can make SYBR technology equally specific and accurate (Tajadini et al. 201).

With the goal of this work to provide communities and local

governments information about the source of fecal contamination, the stakes are high. Decisions about funding for programs to decrease sources of contamination, the possibility of costs to homeowners that might upgrade septic systems, the cost to public agencies trying to locate broken or leaky sewage pipes, and the confidence in community members in the safety of their beaches may be made based on methods developed in this study. Consequently, every effort will be made to optimize and validate results and every possible quality control mechanism will be added to the final protocol.

ACKNOWLEDGMENTS

The authors would like to thank Dean Bowen for obtaining the funding for equipment necessary for this work to be conducted. In addition, thanks to Jason Burkholder for the initial studies that provided the basis for this work, thanks to the support of communities around Anne Arundel County that support the Operation Clearwater monitoring program and provided funding to develop molecular methods for quantifying and identifying the source of contamination in local rivers. Finally, thanks to the Biology laboratory technical staff that has supported our efforts by providing invaluable assistance in finding reagents, setting up equipment and troubleshooting issues.

REFERENCES

Ahmed W, Stewart J, Gardner T, Powell D. 2008. A real-time polymerase chain reaction assay for quantitative detection of the human-specific enterococci surface protein marker in sewage and environmental waters. Environ Microbiol. 10(12):3255-3264. DOI: 10.1111/j.1462-2920.2008.01715.x

Anne Arundel County total maximum daily load restoration plan for bacteria. 2017. Anne Arundel County: Anne Arundel County Public Works; [accessed 2022 Jan 22]. https://www.aacounty.org/departments/public-works/wprp/bacterial-tmdl-plan/3_Draft_Bacteria_TMDL_Restoration_Plan_February_2016.pdf

Green HC, Haugland RA, Varma M, Millen HT, Sorchardt MA, Field KG, Walters WA, Knight R, Sivanganesen M, Kelty CA, et al. 2014. Improved HF183 quantitative

real-time PCR assay for characterization of human fecal pollution in ambient surface water samples. Appl Environ Microbiol. 80(10):3086-3094. DOI: 10.1128/AEM.04137-13

Green HC, White KM, Kelty CA, Shanks OC. 2014. Development of rapid canine fecal source identification PCR-based assays. Environ Sci Technol. 48(19):11453-11461. DOI: 10.1021/es502637b

Harwood VJ, Staley C, Badgely BD, Borges K, Korajkic A. 2013. Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. FEMS Microbiol Rev. 38(1):1-40. DOI: 10.1111/1574-6976.12031

Haugland RA, Varma M, Sivaganesan M, Kelty C, Peed L, Shanks OC. 2010. Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected Bacteroidales species and human fecal waste by qPCR. Syst Appl Microbiol. 33(6):348-57. DOI: 10.1016/j.syapm.2010.06.001

Layton A, McKay L, Williams D, Garrett V, Gentry R, Sayler G. 2006. Development of Bacteroides 16s rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. Appl Environ Microbiol. 72(6):4214-4224. DOI: 10.1128/aem.01036-05

Simpson JM, Santo Domingo JW, Reasoner DJ. 2002. Microbial source tracking: state of the science. Environ Sci Technol. 36(24):5279–5288. DOI: 10.1021/es026000b

Tajadini M, Panjehpour M, Javanmard SH. 2014. Comparison of SYBR Green and TaqMan methods in quantitative real-time polymerase chain reaction analysis of four adenosine receptor subtypes. Adv Biomed Res. 3:85-90. DOI: 10.4103/2277-9175.127998

[US EPA] US Environmental Protection Agency. 2005. Microbial source tracking guide document. Cincinnati (OH): US Environmental Protection Agency. Report No.: EPA/600-R-05-064. [accessed 2022 Jan 22] https://cfpub.epa.gov/si/si_public_record_Report.cfm?Lab=NRMRL&dirEntryID=133523

[US EPA] US Environmental Protection Agency. 2012. Recreational water quality criteria. Washington (DC): US Environmental Protection Agency. Report No.: 820-F-12-058

[US EPA] US Environmental Protection Agency. 2015. Method 1609.1: Enterococci in water by TaqMan® quantitative polymerase chain reaction (qPCR) with internal amplification control (IAC) assay. Washington (DC): US Environmental Protection Agency. Report No.: EPA-820-R-15-099

[US EPA] US Environmental Protection Agency. 2019. Method 1696: characterization of human fecal pollution in water by HF183/BacR287 TaqMan® quantitative polymerase chain reaction (qPCR) assay. Report No.: EPA-821-R-19-002

ASHLEY DYJACK

An Examination of Effort-Based Grading Effectiveness

ABSTRACT

The traditional standards-based or competency-based grading systems, preferred by American schools and colleges, do not incorporate non-academic student achievement factors, such as effort, attendance, or attitude when evaluating student performance. This paper reviews current literature on the effectiveness of an alternative grading system, called effort-based grading, that includes criteria for educators to represent non-academic achievement factors in student evaluation. The research will show that effort-based grading is effective in motivating student achievement up until students are able to exert the least amount of effort for the maximum achievement, where then standards-based or competency-based grading systems become more effective.

KEY WORDS

effort-based grading

mindset

motivation

standards or competency-based grading

alternative institutions

FACULTY MENTOR

Jackie Gambone, Ph.D.
Professor, Teach Institute

INTRODUCTION

Despite American schools and colleges using mostly a standards-based education system, many instructors tend to incorporate non-academic achievement factors into their grading systems, such as student effort, attendance, and attitudes (McMillan, 2018). This addition to the grading scale may be due to instructors wanting to represent the work the students are putting forth that may not be represented solely in their performance-based grades. However, according to McMillan, "Most assessment experts agree that nonacademic indicators should have little or no bearing on the academic performance grade" (McMillan, 2018, p. 438). This paper will look at the effectiveness of incorporating student effort

in grading, whereas effectiveness is defined as the ratio of effort to performance. The following sections include a literature review, research gap identification, and suggested areas for future study.

Literature Review

Over the years, researchers have looked theoretically and empirically at the correlation between student effort and student grades. They looked within the parameters of student performance in paradigms that included student effort as part of the grading criteria (Swinton, 2010), absolute and relative grading systems (Paredes, 2017), and instructor influence(s) on the gap between student effort and grades (Highfill & Marcum, 2019). In each case, the data shows that effort-based grading is effective up until a certain level of student ability or achievement, where then absolute or standards-based grading becomes more effective (Swinton, 2010; Paredes, 2017, Highfill & Marcum, 2019).

Swinton (2010) examines the effectiveness of Benedict College's Success Equals Effort (SE2) policy for freshmen and sophomore level courses where the student's grade is calculated using weighted categories for knowledge and effort. The model includes 40% knowledge and 60% effort for freshman courses and the reverse for sophomore courses (Swinton, 2010). *See Appendix A for the grading matrices.* Since the model used by Benedict College includes grades for both effort and knowledge, the matrices indicate how each grade impacts the final overall grade. In the freshmen model, where student effort is weighted at 60% of the final grade, a student's final grade will reflect more on their effort than knowledge; whereas in the sophomore model, students must display their knowledge rather than rely on their effort to receive their desired grade.

Benedict College implemented the policy to increase the market value of its graduates to future employers and its graduation rates (Swinton, 2010). A follow-up study by Swinton (2014), showed

that the policy did not significantly increase graduation rate, but did contribute to a reduced amount of time for degree completion. Swinton (2010) argues that the policy's driving question is "how do you induce all students to give effort without lowering the amount of knowledge gained by the students or weakening or minimizing the signal that is sent to future employers" (p. 1178). The results showed that there is a positive correlation between a student's effort grade and a student's knowledge grade up until a certain point where a student's academic ability allowed them to achieve the maximum desired grade with the minimum amount of effort (Swinton, 2010). This is evidence shown in the gap between students who have natural academic ability and those who must consistently strive to match their peers and get the desired or expected grades. This view can also be impacted by the percentage of effort verses true knowledge. Employers often want to know how much a candidate already knows verses the effort they would put in to learn what they need to know for the required position.

Swinton (2010) highlights that the impact often comes from the instructor's view of the learning process, which falls under three categories: maximum grades, effort, or knowledge. At the university level, instructors need to consider potential employers will evaluate a student's grades and what that information will tell the employer about the student's ability. For example, if an instructor just gives all students the maximum grade, the employer will have no knowledge of the student's ability and the student would not be motivated to apply effort to the learning process. Whereas, if the instructor decides to maximize student effort, students will apply the effort needed to achieve their desired grade but does not tell employers anything about student knowledge or ability. Lastly, if the instructor maximizes knowledge, employers will be able to determine student ability, however lower ability students may not put effort into the learning process.

Like Swinton's (2010) conclusions that implemented a partial

effort-based system, Paredes, (2017) shows that the same conclusion applies to a full implementation of an effort-based grading system. Paredes theoretically and empirically explores the relationship between a student's ability and the amount of effort the student puts forth within a relative (performance-based) and absolute (standards-based) grading environment. The model developed for this study, "shows that the grading system can influence both the total amount of effort in a class and the level of individual effort throughout the ability distribution" (Paredes, 2017, p. 114). Assuming low ability students did not give up, the model predicted that these students would flourish under a relative grading system, but struggle within an absolute system where the standards are higher, and the cost of the extra effort would not be worth the results (Paredes, 2017). The reverse is true for high ability students as in a relative system, they may not be inclined to exert as much effort due to lower standards (Paredes, 2017). Using a unique data set from the University of Chile (where the grading system changed from absolute to relative and then back to absolute), the author showed that the model predicted the data trends correctly. Paredes (2017) concludes that student effort has a positive effect on a relative grading environment until the student's ability increases to the point where the effort no longer is needed to influence the grade and that the choice of grading system will depend on the instructor or school's target student audience.

The conclusions drawn by Swinton (2010) and Parades (2017), that students want to achieve the maximum grade for the least amount of effort, supports the reflection of Highfill and Marcum (2019), showing that students may "discuss strategies to 'game' the system to ultimately achieve a desired score without strenuous student effort by the student" (p. 61). Highfill and Marcum reflect on how instructor choices may affect the gap between a student's effort and grades. To create the gap between student effort and grades, the adapted model used by the authors introduces a

random component to reflect that there is not a perfect correlation between the amount of effort a student puts forth and the grade the student earns (Highfill & Marcum, 2019). Instructors may see the gap as the result of students misjudging their own ability when attempting to calculate how much effort they need to receive their desired grade or by their own choices due to the subjective nature of grading (partial credit, extra credit, student bias, etc.) or the setup of the grading environment (letter grades, pass/fail, plus/minus), assuming the instructor has leeway, where they must designate cut offs for each achievement (Highfill & Marcum, 2019). Overall, the efficiency of effort-based grading will depend on your student audience as well as paradigms that exist in the instructor's grading structure.

Research Gap Identification

Given that student ability and achievement tends to limit the effectiveness of effort-based grading, potential research gaps include evaluating whether the school environment (i.e. an alternative institution) could potentially increase the effectiveness of effort-based grading, developing a grading system to bridge the gap between the effort-based and the standards-based grading systems, and determining how a student's intrinsic motivation and mindset to learn can affect how much effort a student is willing to exert when it comes to effort-based learning.

In the United States, "within public education, alternative schools exist as one form of dropout prevention and youth re-engagement in school, with approximately 3% of United States (U.S.) high school students attending alternative high schools" (Tierney, 2020, p. 242). An alternative institution often provides "at risk" students, who may not have been successful in a traditional comprehensive school based on behavior or academic performance, opportunities to learn in a non-traditional school environment. At-risk students are defined as those who require interventions

for continued academic success. One advantage an alternative institution offers is that student success is frequently redefined to include non-academic factors such as student social and academic engagement, student ownership of learning goals, and education (including graduation progression) and assisting students in identity development within an academic community (Tierney, 2020). This can be reflected upon by higher education institutions as well.

Since alternative institutions may already include non-academic factors in their definition of student success, the addition of effort-based grading may assist them in reaching their goals since it would allow for student effort to be accounted for and visible towards student academic achievement. However, there may be a risk to adding an effort-based grading system to an alternative institution since students could potentially find a way to exert the minimal amount of effort for the maximum grade, like students in comprehensive institutions, except in this case, these students most likely would not be ready to transition to a standards-based grading system. If this occurred, a stop gap or bridge system would need to be researched or developed, to assist the students in the alternative institution to continue to progress towards their goals.

Carol Dweck (2006) discusses the fixed and growth mindsets. Fixed mindsets are characterized as needing to be perfect, failure being the result, negative self-talk, that your qualities are permanent, etc. Whereas with a growth mindset, failure is opportunity, seeking progress not perfection, embracing constructive criticism to grow and improve, etc. *See Appendix B.* If a student has a fixed mindset, the less likely they will be to strive for a desired grade because they would have already decided it is not possible. With a growth mindset, effort becomes critical and the tool a student can use to achieve their desired goal.

The student population's learning motivation and mindset should be determined and potentially improved before introducing an effort-based grading system to an alternative institution

environment to assist with getting the maximum benefit for the students. This is due to the majority of these "at risk" students having negative experiences with school and thus, even with the addition of an effort-based grading system, may not attempt work if they believe they will just fail in the end. Again, higher education institutions should consider the impact as students transition to them from the K-12 environment.

Alfie Kohn (1994) defines two types of motivation: intrinsic motivation, which is 'an interest in the task for its own sake' and extrinsic motivation, in which 'the completion of the task is seen chiefly as a prerequisite for obtaining something else.' Kohn's research shows that students who are extrinsically motivated are more likely to lose interest in the task they are working on since the end goal is to get the promised reward. Applying this to a classroom environment may mean that students who are motivated by obtaining passing grades will only put forth the minimum effort to receive their desired grade, thus curbing a student's desire to learn and their creativity. On the other hand, if students are intrinsically motivated, an effort-based grading system may allow them to explore their desire to learn about a topic, while rewarding them for the effort they put into the task.

Lastly, in another article, Kohn (2015) describes the two mindsets set forth by Carol Dweck (2006), the fixed mindset, where a person's intelligence and talent is set and unable to be changed and the growth mindset, that says that every person is capable of learning something with enough effort. Students who have a growth mindset and believe they can learn the subject or topic at hand, are more likely to put forth the effort needed to persevere through a task, as mentioned previously. In an alternative institution environment, aligning an effort-based grading system with encouraging the development of growth mindsets in students, may assist students in having a positive school experience, which in turn may increase the student's intrinsic motivation to learn and

persevere through challenging tasks.

The identified research gaps are not suggesting that one gap is more important than another, however, changing the institutional environment may increase the benefit of the implementation of an effort-based grading system. Due to the specialized nature, staff at alternative institutions may receive more training or professional development opportunities in mindset and motivation theories.

Areas for Future Study

The literature review and research gap identification sections have highlighted several areas that would benefit from additional research. These areas include: the conditions in which an effort-based grading system may be effectively implemented, the effects on intrinsic motivation and growth mindsets on student effort, and the development of a grading system that bridges the gap between an effort-based and standards-based grading system.

One area of future study should focus on the effectiveness of effort-based grading in an alternative institution environment, where student achievement already includes non-academic factors. This study or studies should include how to improve student intrinsic motivation, changing student's mindset from fixed to growth, and the effectiveness of effort-based grading for "at-risk" students. The researcher would like to implement a structured grading system that includes student effort as compared to the current system. A randomized field experiment could prove useful where a control group maintains the current grading system and a treatment group works within the confines of a partial effort-based system. The end-goal of the researcher's alternative institution program is to prepare students to return to their comprehensive school.

Other future areas of study include how being evaluated under an effort-based grading system may affect employer's outlook of students, how effort-based grading may inflate student grades

in an alternative environment and the effects of returning to a comprehensive school and standards-based grading system, how effort-based grading could help students who have a low ability in a specific subject to avoid failing and start the conversation on how the instructor can help the student overcome the low ability in the subject, and the effect an effort-based grading system could have on low ability students on having a positive interaction with a school environment overall.

Conclusion

This paper reviewed current literature on the effectiveness of effort-based grading, research gap identification, and areas of future study. The conclusion drawn from the literature review is that effort-based grading is effective up until students can perform the minimum amount of effort for the maximum grade. Lastly, the research gap identified multiple areas of future study that may help determine if effort-based grading may be more effective in an alternative education environment.

References

Dweck, C. (2006). *Mindset*. New York, NY: Random-house Publishing.

Highfill, J., & Marcum, T. M. (2019). Modeling undergraduate student effort: Exploring the gap between effort and grade. *Journal of Higher Education Theory and Practice, 19(1)*, 56-66. https://doi.org/10.33423/jhetp.v19i1.668

Kohn, A. (1994). *The Risks of Rewards* (ED3769990). ERIC. https://files.eric.ed.gov/fulltext/ED376990.pdf

Kohn, A. (2015). *The "mindset" mindset: What We Miss By Focusing on Kids' Attitudes*. https://www.alfiekohn.org/article/mindset/

McMillan, J. H. (2018). *Classroom assessment: Principles and practice that enhance student learning and motivation*. (7th ed.). Pearson.

Paredes, V. (2017). *Grading system and student effort. Education Finance and Policy 2017; 12(1)*: 107–128. https://doi.org/10.1162/EDFP_a_00195

Swinton, O. H. (2015). *An A for effort. American Economic Review, 105(5)*, 616–620. https://doi.org/10.1257/aer.p20151116

Swinton, O. H. (2010). *The effect of effort grading on learning. Economics of Education Review, 29(6)*, 1176–1182. http://doi.org/10.1016/j.econedurev.2010.06.014

Tierney, G. (2020). *Ideational resources and alternative definitions of success: Reorienting to Education and Developing Identities in an Alternative High School. High School Journal, 103(4)*, 241–261. https://doi.org/10.1353/hsj.2020.0015

APPENDIX A

Benedict College Success Equals Effort (SE2) Grading Matrices

TABLE 1
Freshman level grade matrix.

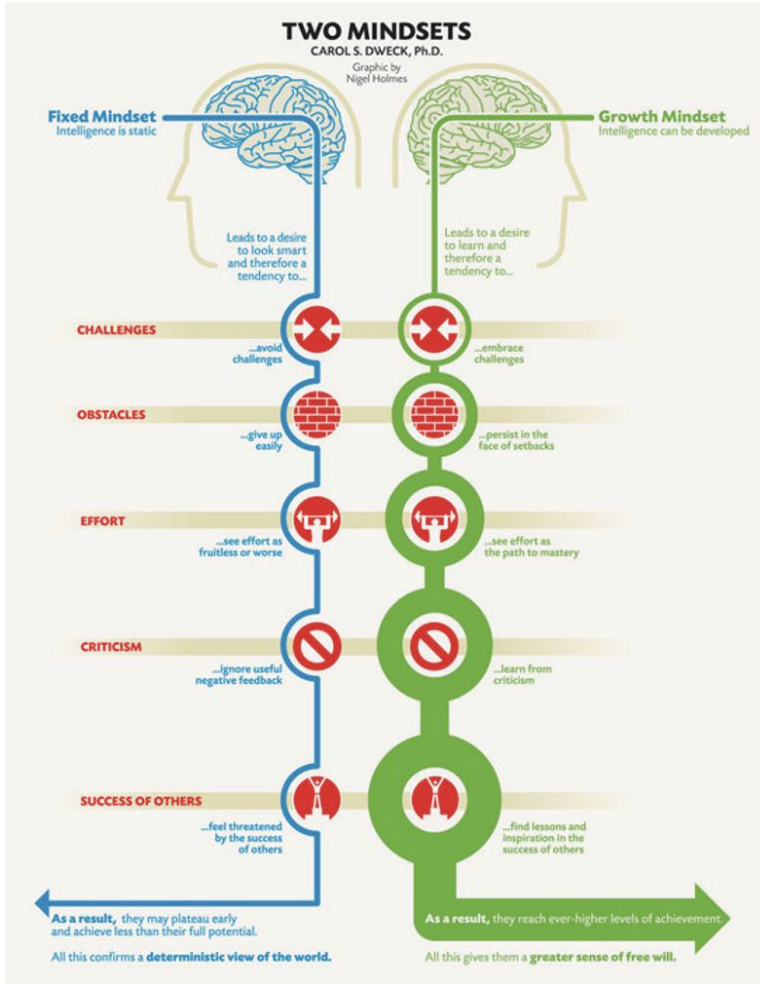
		Knowledge grade				
		A	B	C	D	F
Effort Grade	A	A	A	B	C	C
	B	B	B	B	C	D
	C	B	C	C	C	D
	D	C	C	D	D	F
	F	C	D	D	F	F

TABLE 2
Sophomore level grade matrix.

		Knowledge grade				
		A	B	C	D	F
Effort Grade	A	A	B	B	C	D
	B	A	B	C	C	D
	C	B	B	C	D	D
	D	B	C	C	D	F
	F	C	C	D	D	F

(Source: Swinton, 2010, p. 1177)

APPENDIX B



(Dweck, 2006)

ACKNOWLEDGEMENTS

The researcher would like to thank her mentor Dr. Jaclyn Gambone for her tremendous support on this project. The researcher would also like to acknowledge the students at the alternative institution whom she teaches; they inspired this research.

Thao-Nhi Luu and Maria Nicos Alain Pasaylo

Exploring the Hill Cipher through Linear Algebra and Python

Abstract

For most of human history, security of data communication has been essential. The relevance of encryption in times of war and (more recently) in the information age is difficult to overestimate. During the 20th century, advances in mathematics and technology prompted the proliferation of many new methods of encryption. Among these methods, the Hill cipher, a polygraphic substitution cipher introduced in 1929, pioneered the use of modular arithmetic and linear algebra in an encryption algorithm. In this paper, we explore the Hill cipher. This expository article includes a discussion of the mathematical framework and implementation of the cipher, as well as examples, a method of plaintext attack, and Python code for the Hill cipher.

1. Foundations of Cryptography

Oftentimes, we interchangeably use the words cryptography, cryptology, and cryptoanalysis. Nevertheless, these terms have different meanings. Cryptography deals with the techniques essential for data protection over communication systems; cryptology is the general term given to the study of communication over unprotected channels; cryptoanalysis is the process of breaking secure communication systems (for example, frequency analysis). See [9] for further details.

To introduce some standard concepts and terminology, let

us assume that we have two people, Lin and Al, who are sharing messages using an encryption method. In our story, Lin needs to securely send a message to Al. The message Lin is going to send to Al is called the plaintext. To securely send this message, Lin will use a key to encode, or encrypt, the plaintext. This key is a piece of information, sometimes a number, a string of characters, or a matrix, which encodes the plaintext using an encryption algorithm. The encoded message is called the ciphertext. Lin then sends this ciphertext over public or unsecured channels. Al receives this ciphertext and then uses a key to privately decode, or decrypt, the ciphertext back into the original plaintext message using a decryption algorithm.

To illustrate these ideas, we briefly describe one of the simplest – and perhaps one of the earliest – encryption algorithms known: the Caesar cipher. Developed around 100 BC, the Caesar cipher was used by Julius Caesar to send secret messages to his generals in the field. With this method, Lin and Al associate to every letter of the alphabet the corresponding number: *a* corresponds to 0, *b* corresponds to 1, and so on, all the way to *z*, which corresponds to 25. The Caesar cipher key, a whole number between 0 and 25 (inclusive), is privately agreed upon between Lin and Al before any encoding. Lin chooses this number to be 5, her favorite season of *The Simpsons*, and secretly shares this choice with Al. Then Lin and Al are separated to opposite ends of the battlefield. The following morning, Lin takes her plaintext message, which reads *attack*, and replaces every letter with the corresponding number, yielding the array of numbers [0, 19, 19, 0, 2, 10]. Then Lin encrypts the message by adding the preselected key, the number 5, to each number in the array to yield [5, 24, 24, 5, 7, 15]. This is translated into the ciphertext *FYYFHP*. This ciphertext is carried by a brave soldier across the battlefield. Although the ciphertext may be easily intercepted, any prying eyes do not have access to the key, and so they are unable to read it. When Al receives the

message, they convert the ciphertext into numbers, apply the key to these numbers by subtracting 5, and convert the shifted numbers back into the original message, *attack*.

The Caesar cipher is an example of private key cryptography, where the key used to encrypt and decrypt the message is known only to the sender and receiver. The Caesar cipher is also an example of symmetric key cryptography because both the sender and receiver use the same private key. Public key cryptography, in contrast, uses a public and a private key. In this case, a public key is known to everyone, and it is used to convert plaintext to ciphertext. The receiver then uses a distinct private key to decode the ciphertext. The most prominent and widely used public key cryptography system today is RSA named after its inventors Rivest, Shamir, and Adleman. RSA uses very large prime numbers to create public keys and leverages the computational difficulty of factoring large numbers. Further discussion of public key cryptography and this particular algorithm is beyond the scope of this paper. See [9] for further details.

We return to our discussion of (symmetric) private key cryptography (where the sender and receiver share the same secret key) and describe some particular types of encryption algorithms in this category.

The Caesar cipher is just one example of a substitution cipher, in which letters of the alphabet, represented by the numbers 0 through 25, are 'scrambled' according to a fixed permutation – or reordering – of those numbers. Although simple to implement, these ciphers are easily broken by using frequency counts of letters. There are more sophisticated methods for producing this 'scrambling'. This includes the affine cipher, which uses an affine linear transformation to scramble the letters of the alphabet. Another example is the well-known Enigma Machine. It was used by Nazi Germany in WWII and cracked by allied forces.

Another class of encryption methods relies on a block cipher

algorithm. Block ciphers are any encryption method that receives all blocks of characters of fixed block size and produces an output of encrypted blocks. Well-known block encryption methods include the Vigenère cipher, the Playfair cipher, the ADFGX cipher, and the Hill cipher. We focus on the Hill cipher for the remainder of this paper.

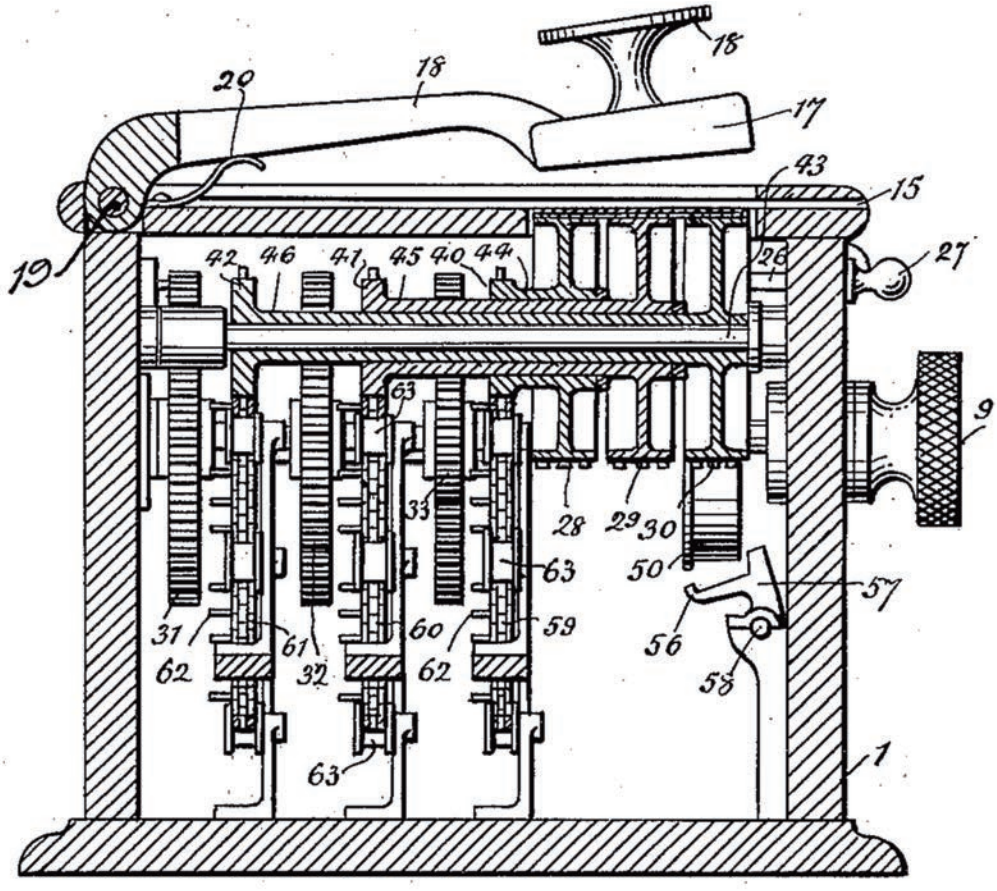


FIGURE 1
Hill cipher Machine

Lester Hill, an American mathematician and educator, invented the encryption method that now bears his name in 1929. See [3, 4] for further details. Hill's method uses elementary methods in modular arithmetic and linear algebra. The Hill cipher "… seems never to have been used much in practice. Its significance is that it was perhaps the first time that algebraic methods (linear algebra and modular arithmetic) were used in cryptography in an essential way." [9]

2. MATHEMATICAL FORMALISM

2.1 Modular Arithmetic

Careful readers may have noticed one possible flaw in our earlier description of the Caesar cipher. The algorithm instructs the parties involved, Lin and Al, to establish a secret key – a single number to shift the letters of the alphabet *forward* when encoding and *backward* when decoding. But, what happens if the shift pushes the numerical value of a letter over 25? For example, the letter x is replaced by 23. If we shift by Lin and Al's secret key of 5, we obtain 28 – and this number does *not* correspond to any letter of our alphabet. The simple solution is to take the remainder of 28 when divided by 25. The letter x encodes as 3. In this way, we can shift every letter of the alphabet. This process of taking remainders and carrying out numerical computations is formalized in mathematics with modular arithmetic. Modular arithmetic is essential in almost all cryptographic methods that convert plaintext to numbers and apply algebraic tools to encode this numerical data. We describe few basic definitions and properties of integer arithmetic here.

First, we define the greatest common divisor of two positive integers, m and n, denoted $\gcd(m, n)$, to be the largest integer that divides both m and n. For example, we have $\gcd(6, 8) = 2$ and $\gcd(3, 14) = 1$. Next, let us fix an integer $d > 1$. For any integer n, the Division Algorithm guarantees that we can always find unique integers q and r, where $0 \leq r < d$, such that

$$n = qd + r.$$

A proof of this result can be found in [7]. We call q the quotient and r the remainder when n is divided by d. This result guarantees that division of any integer by d results in a unique remainder between 0 and $d - 1$. For example, let $d = 7$ and divide a few numbers by d:

$$13 = 1 \cdot 7 + \mathbf{6}, \quad \text{and so 13 has a remainder of 6 when divided by 7.}$$
$$6 = 0 \cdot 7 + \mathbf{6}, \quad \text{and so 6 has a remainder of 6 when divided by 7.}$$
$$-78 = (-12) \cdot 7 + \mathbf{6}, \quad \text{and so -78 has a remainder of 6 when divided by 7.}$$

With a divisor d fixed, we say that two integers m and n are congruent mod d, written $m \equiv n \bmod d$, if they have the same remainder when divided by d. For the following examples, we again fix $d = 7$:

$$\mathbf{13} \equiv \mathbf{6} \bmod 7, \quad \text{because 13 and 6 both have a remainder of 6 when by divided by 7.}$$
$$-\mathbf{6} \equiv \mathbf{1} \bmod 7, \quad \text{because -6 and 1 both have a remainder of 1 when by divided by 7.}$$
$$\mathbf{12} \equiv \mathbf{96} \bmod 7, \quad \text{because 12 and 96 both have a remainder of 5 when divided by 7.}$$

We provide a few elementary but important results in modular arithmetic here; see [7] for further discussion and details.

Theorem 1. Let $d > 1$.

1) If $r \equiv s \bmod d$ and $u \equiv v \bmod d$, then $r + u \equiv s + v \bmod d$, and $ru \equiv sv \bmod d$.
2) Let $0 \leq r, s < d$. Then $r \equiv s \bmod d$ if and only if $r = s$.

Here is an example to illustrate part 1). Let's take $d = 7$. For brevity, we write $r \equiv s$ if $r \equiv s \bmod 7$. Now let us take the product of the two integers 13 and -6 and reduce mod 7: $13 \cdot (-6) = -78 \equiv 6$. Since $13 \equiv 6$ and $-6 \equiv 1$, we could also write $13 \cdot (-6) \equiv 6 \cdot 1 = 6 \equiv 6$. For the sum, $13 + (-6) \equiv 6 + 1 = 7 \equiv 7$. All of this just says that we can add and multiply integers mod d. Reducing our sums or products along the way at any step and the result will always be the same mod d. Part 2) of the Theorem says that any integer is equivalent to exactly *one* integer from the set $\{0, 1, \ldots, d - 1\}$, mod d. This number is just its remainder. For example, with a modulus of $d = 7$, 13 is congruent to one and only one integer from the set $\{0, 1, 2, 3, 4, 5, 6\}$. The fact that this can be done uniquely will be important in the next section. We will refer to replacing any integer by its equivalent unique remainder as 'reduction mod d'.

Standard arithmetic of real numbers has a well-known property: if r is a nonzero number, we can always multiply by $(1/r)$ to get 1. For example, $6 \cdot (1/6) = 1$, and $(-1/5) \cdot (-5) = 1$. We say that $1/r$ is the multiplicative inverse of r (and vice versa), and we write $1/r = r^{-1}$. So, 6 and $1/6$ are multiplicative inverses of each other, and so again for $-1/5$ and -5. We encounter multiplicative inverses in modular arithmetic as well. If the product ab of two integers a and b is congruent mod d, we say that b is the modular inverse of a (and vice versa). To illustrate, let's use the modulus $d = 7$ again. Since

$$3 \cdot 5 = 15 \equiv 1 \bmod 7,$$

we say that the modular inverse of 3 is 5 (and vice versa). However, unlike (regular) inverses of nonzero numbers, an integer does not necessarily have a modular inverse mod d. For example, if our modulus is $d = 4$, there is *no* integer k such that $2 \cdot k \equiv 1 \bmod 4$; that is, 2 has no modular inverse mod 4. This is because just $2 \cdot k$ is a multiple of 2, and multiples of 2 can only ever have remainders of 0 or 2 when divided by 4. Nevertheless, there is an easy way to determine if an integer n has a modular inverse mod d, given by the following result, which is obtained by an application of the Euclidean Algorithm:

Theorem 2. Let $d > 1$. An integer k has a modular inverse mod d if and only if $\gcd(d, k) = 1$.

See [7] for details. (Note: two integers d, k satisfying the property $\gcd(d, k) = 1$ are said to be relatively prime, and we will use this terminology). Modular inverses allow us to write some fractions as integers, mod d. If we take $d = 7$ for our modulus, then 5 is guaranteed to have an inverse mod 7, since $\gcd(5, 7) = 1$. By an exhaustive search through the remainders $\{0, 1, 2, 3, 4, 5, 6\}$ of 7, we can find that $5 \cdot 3 = 15 \equiv 1 \bmod 7$. Hence, the modular inverse of **5** is **3**. This allows us to write $\frac{1}{5} = 5^{-1} \equiv 3 \bmod 7$. We will use this property of modular inverses below, when we define the modular inverse of a *matrix*.

2.2 Matrices and Matrix Algebra

Let m and n be positive integers. An $m \times n$ matrix is a rectangular array

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

in which each entry, a_{ij}, of the matrix is a real number[1] (we also refer to a_{ij} as the (i, j) entry). An $m \times n$ matrix has m rows and n columns. Matrices are usually denoted by capital letters. For our applications, we are interested only in square matrices where $m = n$; i.e., where the number of rows and columns are the same. A 2×2 square matrix looks like

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

where a_{11}, a_{12}, a_{21}, and a_{22} are numbers. Matrices can be multiplied together. This method of multiplication needs to be described. First suppose A is a $1 \times n$ matrix and B is an $n \times 1$ matrix:

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}, \qquad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

[1] Matrices can also be constructed with other kinds of numbers for these entries, such as the complex numbers, the quaternions, and elements of finite fields, to name a few. However, these will not be needed here.

We call A a row matrix and B a column matrix. Then the product AB of these two matrices is obtained by multiplying all corresponding entries, and adding all of these products: $AB = a_1b_1 + a_2b_2 + \cdots + a_nb_n$. For example,

$$\text{if } A = \begin{bmatrix} 2 & 3 & 0 & 4 \end{bmatrix} \text{ and } B = \begin{bmatrix} -1 \\ 5 \\ -2 \\ 3 \end{bmatrix}, \quad AB = 2 \cdot (-1) + 3 \cdot 5 + 0 \cdot (-2) + 4 \cdot 3 = 25.$$

Now let A be an $m \times n$ matrix and B an $n \times q$ matrix (Note: we require the number of rows of B and columns of A to be equal). Then the matrix product AB is the $m \times q$ matrix whose (i, j) entry is the matrix product of the i^{th} row of A and the j^{th} column of B, as defined above. We illustrate with an example. Let $C = \begin{bmatrix} 1 & 3 \\ 2 & 0 \end{bmatrix}$ and $D = \begin{bmatrix} -1 & 4 \\ 2 & -3 \end{bmatrix}$. Then the product of C and D is the 2×2 matrix

$$CD = \begin{bmatrix} 1 \cdot (-1) + 3 \cdot 2 & 1 \cdot 4 + 3 \cdot (-3) \\ 2 \cdot (-1) + 0 \cdot 2 & 2 \cdot 4 + 0 \cdot (-3) \end{bmatrix} = \begin{bmatrix} 5 & -5 \\ -2 & 8 \end{bmatrix}.$$

While multiplication of real numbers is commutative, matrix multiplication is not commutative. In general, $AB \neq BA$ (although there are exceptions). The reader is encouraged to multiply the above 2×2 matrices C and D in the order DC to verify this.

Of special interest is the $\mathrm{n} \times \mathrm{n}$ matrix I_n, with entries of 1 down the main diagonal and 0 in all other entries. For example, we have

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We call I_3 the 3×3 identity matrix because if we multiply it by any other 3×3 matrix A, we have $AI_3 = I_3A = A$ (as the reader should check). In general, we call I_n the $n \times n$ identity matrix. We can think of I_n as acting like the number 1, at least when it comes to multiplication of square matrices. If we have two square matrices A and B such that $AB = I_n$, we say that B is the inverse of A (written $B = A^{-1}$), and that A is the inverse of B (written $A = B^{-1}$). For example, the reader can verify the following:

$$\text{for } E = \begin{bmatrix} 5 & 2 \\ -7 & -3 \end{bmatrix} \text{ and } F = \begin{bmatrix} 3 & 2 \\ -7 & -5 \end{bmatrix}, \text{ we have } EF = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = FE.$$

So, we can write $E = F^{-1}$ and $F = E^{-1}$.

While any nonzero real number has a multiplicative inverse (for example, $2 \cdot (1/2) = 1$), many nonzero square matrices *do not* have an inverse. In this case, we say that a matrix is *not invertible*. For example, we leave it to the reader to verify that any 2×2 matrix with a single row of all 0's *cannot* have an inverse. An *invertible* matrix, on the other hand, is a square matrix which does have an inverse.

For example, we can see that the matrices E and F in the example above are both invertible matrices. With these definitions and properties at hand, we are faced with two important questions: how do we know *which* square matrices are invertible, and, if a matrix is invertible, how do we find an inverse for it? We give a partial answer to both questions in the following section.

2.3 The Determinant of a Matrix

For any 2×2 matrix A, we will compute a number called its determinant, denoted $\det(A)$:

$$\text{if} \quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \det(A) := ad - bc.$$

That's easy enough to calculate, but what's the use in doing so? For our purposes, the most important property of the determinant is the following theorem.

Theorem 3. Let $A = \begin{bmatrix} q & r \\ s & t \end{bmatrix}$. Then

1) A is invertible if and only if $\det(A) \neq 0$, and
2) if A is invertible, $A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} t & -r \\ -s & q \end{bmatrix}$.

The theorem above provides a way to find the inverse of square matrices of size 2. What about square matrices of size > 2? These matrices also have a determinant, and part 1) of the above theorem remains true. Calculating the inverse of such a matrix becomes computationally lengthy as n gets larger and larger. The reader is referred to any standard reference on linear algebra, such as [5], for a comprehensive treatment of matrix inverses and determinants, including standard algorithms for finding the determinant of any $n \times n$ matrix.

2.4 The Modular Inverse of a Matrix

We combine the ideas of the preceding sections to define modular equivalence and the modular inverse of a matrix. From this point onward, we are interested mainly in matrices with integer entries,

and we will call such a matrix an **integer matrix**. For this section, we fix a modulus $d > 1$. We can carry out matrix multiplication and find matrix inverses, working entirely with the integers k ranging from 0 to $d - 1$. If A and B are two matrices, we will write $A \equiv B \bmod d$ if the corresponding entries of A and B are congruent mod d. For example, if $d = 7$, then

$$\begin{bmatrix} -1 & 5 & 9 \\ 0 & 7 & -21 \\ 8 & 22 & 50 \end{bmatrix} \equiv \begin{bmatrix} 6 & 5 & 2 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \bmod 7.$$

We say that an $n \times n$ integer matrix A has a *modular inverse mod d* if there is an integer matrix B such that $AB \equiv I_n \bmod d$. If so, we say that A is *invertible mod d*. It is possible for an $n \times n$ integer matrix to have an inverse, but fail to be invertible mod d – as this example shows: let $d = 4$ and $A = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}$. Then by Theorem 3, A is invertible, because $\det(A) = 2 \cdot 4 - 2 \cdot 2 \neq 0$. We cannot produce *any* integer matrix B satisfying $AB \equiv I_2 \bmod 4$, because the entries of any product AB will always be even numbers. So, the diagonal entries will never be congruent to $1 \bmod 4$. Fortunately, the criterion that tells us *when* a matrix is invertible mod d is known:

Theorem 4. Let A be an $n \times n$ matrix. Then A is invertible mod d if and only if $\gcd(d, \det(A)) = 1$.

Proof. We give a partial proof for the case $n = 2$. Let $A = \begin{bmatrix} q & r \\ s & t \end{bmatrix}$. For brevity we will write $x \equiv y$ if $x \equiv y \bmod d$. Assume $\gcd(d, \det(A)) = 1$. By Theorem 2, $\det(A)$ is invertible mod d. Then we let

$$e \equiv \det(A)^{-1} \cdot t, \quad f \equiv \det(A)^{-1} \cdot (-r), \quad g \equiv \det(A)^{-1} \cdot (-s), \quad \text{and} \quad h \equiv \det(A)^{-1} \cdot q.$$

Now if we let $C = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$, a verification (using Theorem 1) shows that $AC \equiv I_2 \equiv CA \bmod d$. The other direction, proving that $\gcd(d, \det(A)) = 1$ if A has a modular inverse, is left to the reader. □

Here is an example: let $d = 5$, and $A = \begin{bmatrix} 3 & 1 \\ 3 & 4 \end{bmatrix}$. Then $\det(A) = 3 \cdot 4 - 1 \cdot 3 = 9$. Since $\det(A) \neq 0$, A has an inverse, given by

$$A^{-1} = \frac{1}{9} \begin{bmatrix} 4 & -1 \\ -3 & 3 \end{bmatrix} = 9^{-1} \begin{bmatrix} 4 & -1 \\ -3 & 3 \end{bmatrix} = \begin{bmatrix} 4 \cdot 9^{-1} & (-1) \cdot 9^{-1} \\ (-3) \cdot 9^{-1} & 3 \cdot 9^{-1} \end{bmatrix}.$$

Since 5 and 9 are relatively prime, A *also* has a modular inverse, which can be obtained by reducing

all entries of $A^{-1} \bmod 5$. To do this, we use the fact that 4 is the modular inverse of $9 \bmod 5$, because $9 \cdot 4 = 36 \equiv 1 \bmod 5$. Therefore, we can write $9^{-1} \equiv 4 \bmod 5$, and we have

$$A^{-1} = \begin{bmatrix} 4 \cdot 9^{-1} & (-1) \cdot 9^{-1} \\ (-3) \cdot 9^{-1} & 3 \cdot 9^{-1} \end{bmatrix} \equiv \begin{bmatrix} 4 \cdot 4 & 4 \cdot 4 \\ 2 \cdot 4 & 3 \cdot 4 \end{bmatrix} = \begin{bmatrix} 16 & 16 \\ 8 & 12 \end{bmatrix} \equiv \begin{bmatrix} 1 & 1 \\ 3 & 2 \end{bmatrix} \bmod 5.$$

Hence, a modular inverse for A is $C = \begin{bmatrix} 1 & 1 \\ 3 & 2 \end{bmatrix}$. One can check that $\begin{bmatrix} 3 & 1 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 3 & 2 \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \bmod 5$.

For our applications to the Hill cipher, here is the important property of matrix A with a modular inverse. The proof follows from the definition of the modular inverse.

Theorem 5. Let A be an $n \times n$ matrix with $\gcd(d, \det(A)) = 1$, and let B be its modular inverse $\bmod d$. Let $\mathbf{u}$ be an $n \times 1$ column matrix. Then $\mathbf{u} \equiv BA\mathbf{u} \bmod d$.

3. HILL CIPHER IMPLEMENTATION WITH MODULAR ARITHMETIC AND MATRIX ALGEBRA

With the mathematical foundations of modular arithmetic and matrix algebra in place, we are ready to describe the implementation of the Hill cipher. We will see that the Hill cipher is a block cipher: after a plaintext message is converted into a sequence of integers (from 0 to 25), this sequence is partitioned into blocks of predetermined length and these blocks are encrypted one at a time.

We begin the process of encryption by choosing a block size n and then a *key*. Recall that the key for the Caesar cipher consists of a single integer which we used to 'shift' our message. For the Hill cipher, our key will be an $n \times n$ matrix which is invertible $\bmod 26$. From now on, we fix our modulus to be 26 because this is the number of letters in the English alphabet. We require the condition of invertibility so that our encrypted message can be decrypted later. We will soon see why and how this works.

3.1 Setting Up the Plaintext

The next step is to convert the letters of our plaintext message to their corresponding values as shown in Table 1 below, where a corresponds to 0, b to 1, and so on, up to z, which corresponds to 25. Then this sequence of integers is blocked into column matrices with a row size that matches the size of the encryption matrix key. For example, if we choose a block size of 4, our plaintext message

a	b	c	d	e	f	g	h	i	j	k	l	m
0	1	2	3	4	5	6	7	8	9	10	11	12

n	o	p	q	r	s	t	u	v	w	x	y	z
13	14	15	16	17	18	19	20	21	22	23	24	25

TABLE 1
English alphabet assigned to numeric values.

will be converted into a sequence of integers, and these integers will be partitioned into blocks of length 4. These length-4 blocks will then be used as the entries in a sequence of 4×1 column matrices. To illustrate, if our sequence of integers is $\{b_1, b_2, \ldots, b_{16}\}$, then we obtain the four column vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$:

$$\mathbf{v}_1 \downarrow \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \qquad \mathbf{v}_2 \downarrow \begin{bmatrix} b_5 \\ b_6 \\ b_7 \\ b_8 \end{bmatrix} \qquad \mathbf{v}_3 \downarrow \begin{bmatrix} b_9 \\ b_{10} \\ b_{11} \\ b_{12} \end{bmatrix} \qquad \mathbf{v}_4 \downarrow \begin{bmatrix} b_{13} \\ b_{14} \\ b_{15} \\ b_{16} \end{bmatrix}$$

In this illustration, it looks like there was a fortunate coincidence – the length of our text string of 16 characters was divisible by our choice of block size of 4. In general, this will not be the case. The standard practice is to add on a single repeated character so that our numerical sequence is divisible by the block size. For example, if our integer sequence had only 13 entries $\{b_1, \ldots, b_{13}\}$, we could end it with 3 repetitions of 25, to yield $\{b_1, \ldots, b_{13}, 25, 25, 25\}$. This 'padding' will not provide any practical obstruction to decoding because the recipient can determine that a long string of z's at the end of the message should be stripped away. With this convention and a block size of n, we can assume that the integer sequence representing our plaintext is always divisible by n – so it will look like $\{b_1, b_2, \ldots, b_{kn}\}$ for some integer $k > 0$. Here is the partition of our integer sequence into k blocks of length n, and their corresponding block column matrices:

$$\{\underbrace{b_1, b_2, \ldots, b_n}_{1^{\text{st}}\text{ block}}, \underbrace{b_{n+1}, b_{n+2}, \ldots, b_{2n}}_{2^{\text{nd}}\text{ block}}, \ \ldots, \ \underbrace{b_{n(k-1)+1}, b_{n(k-1)+2}, \ldots, b_{nk}}_{k^{\text{th}}\text{ block}}\}$$

$$\mathbf{v}_1 = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \qquad \mathbf{v}_2 = \begin{bmatrix} b_{n+1} \\ \vdots \\ b_{2n} \end{bmatrix} \qquad \cdots \qquad \mathbf{v}_k = \begin{bmatrix} b_{n(k-1)+1} \\ \vdots \\ b_{kn} \end{bmatrix}$$

3.2 Encryption

Now we have a block size (n), and a sequence of blocks: these are our $n \times 1$ column matrices $\mathbf{v}_1, \ldots, \mathbf{v}_k$, each with n components, representing our plaintext message. The next step is to choose a key, and to describe how this key will encrypt our blocks. For the Hill cipher method, our key will be an $n \times n$ matrix A which is invertible mod 26. To encrypt the first block $\mathbf{v}_1$, we will matrix

multiply $\mathbf{v}_1$ on the left by our chosen encryption key matrix A. Then we reduce the entries of the resulting column matrix mod 26:

$$\begin{bmatrix} \text{encryption} \\ \text{key matrix} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \xrightarrow{\text{reduce mod 26}} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

$$A \quad \mathbf{v}_1 \quad = \quad \mathbf{e}_1 \quad \xrightarrow{\text{reduce mod 26}} \quad \mathbf{u}_1$$

This provides an encrypted column matrix $\mathbf{u}_1$. We repeat this process for remaining block column matrices $\mathbf{v}_2$, $\mathbf{v}_3$, . . . , $\mathbf{v}_k$. The resulting entries of the encrypted block column matrices $\mathbf{u}_1$, . . . , $\mathbf{u}_k$ can be put into one long sequence of integers:

$$\mathbf{u}_1 = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \qquad \mathbf{u}_2 = \begin{bmatrix} u_{n+1} \\ \vdots \\ u_{2n} \end{bmatrix} \qquad \cdots \qquad \mathbf{u}_k = \begin{bmatrix} u_{n(k-1)+1} \\ \vdots \\ u_{kn} \end{bmatrix}$$

$$\Big\{ \underbrace{u_1, u_2, \ldots, u_n}_{1^{\text{st}} \text{ block}}, \quad \underbrace{u_{n+1}, u_{n+2}, \ldots, u_{2n}}_{2^{\text{nd}} \text{ block}}, \quad \ldots, \quad \underbrace{u_{n(k-1)+1}, u_{n(k-1)+2}, \ldots, u_{nk}}_{k^{\text{th}} \text{ block}} \Big\}$$

Since each matrix $\mathbf{u}_i$'s entries are reduced mod 26, these integers are between 0 and 25. We 'backwards-replace' each of these integers with an alphabet letter: $0 \mapsto a$, $1 \mapsto b$, and so on. Here's the full encryption process:

attackthecastleatdawn

↓ replace each letter with its corresponding integer, padding if necessary

$\{b_1, b_2, \ldots, b_{nk}\}$

↓ block these nk integers, n at a time, into k column matrices

$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$

↓ multiply each column matrix by the encryption matrix A

$\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k\}$

↓ reduce entries of all matrices $\mathbf{e}_i$ mod 26

$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$

↓ 'unblock': use column matrix entries to form a sequence of integers

$\{u_1, u_2, \ldots, u_{nk}\}$

↓ substitute letters for integers

MPWZLDKVNGOQPWSJCNVGZ

The full encrypted message now consists of this encrypted sequence of letters, which we call the ciphertext. If an eavesdropper were to intercept this letter sequence, the original message would remain hidden. Our final encrypted message is the last sequence of letters[2], MPWZLDKVNGOQPWSJCNVGZ, and it is ready to be shared with the world! We can send it in an unsecured email, write it on a postcard, or shout it from a rooftop for all to hear. Without the cipher key A, anyone eavesdropping will be unable to 'reconstruct' the original block column matrices $\mathbf{v}_1, \ldots, \mathbf{v}_k$ (unless they are clever enough to 'crack' the code, but more on this later).

3.3 Decryption

From a distant rooftop, our intended message recipient receives our encrypted sequence of letters on their laptop. Armed with the encryption key matrix A, they are ready to decode the message. Since they have the encryption key matrix A, which has been specifically chosen to satisfy $\gcd(d, \det(A)) = 1$, they are guaranteed to find a modular inverse for A: let us call this modular inverse C. The 'key' – no pun intended – that allows them to decrypt is the following result.

Corollary 1. Let C be the modular inverse of A, and let the column matrices $\mathbf{v}_i$, $\mathbf{e}_i$, and $\mathbf{u}_i$ be given as in Section 3.2 above. Let $\mathbf{w}_i$ be the reduction $\bmod d$ of $C\mathbf{u}_i$ (so all entries of the column matrix $\mathbf{w}_i$ are integers w_j with $0 \leq w_j < d$). Then $\mathbf{w}_i = \mathbf{v}_i$.

Proof. We have

$$\begin{aligned} C\mathbf{u}_i &\equiv C\mathbf{e}_i \\ &\equiv C(A\mathbf{v}_i) && \text{(Since } \mathbf{e}_i = A\mathbf{v}_i\text{)} \\ &\equiv \mathbf{v}_i && \text{(By Theorem 5).} \end{aligned}$$

Since $\mathbf{w}_i \equiv C\mathbf{u}_i$, this shows us that $\mathbf{w}_i \equiv \mathbf{v}_i$ for all i. Consider $\mathbf{v}_1$ and $\mathbf{w}_1$, and write

$$\mathbf{v}_1 = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}, \quad \text{and} \quad \mathbf{w}_1 = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}.$$

By construction, the entries of $\mathbf{v}_1$ are all integers between 0 and $d - 1$. Also, the entries of $\mathbf{w}_1$ are all integers between 0 and $d - 1$ because these entries have been reduced $\bmod d$. So, we have $b_1 \equiv w_1, \ldots, b_n \equiv w_n \bmod d$. Since $0 \leq b_i, w_j < d$, by Theorem 1 we can conclude

[2] This encrypted sequence is entirely random, and is only inserted here for illustrative purposes.

that $b_i = w_i$ for all $1 \leq i \leq n$. Hence, $\mathbf{v}_1 = \mathbf{w}_1$. The same argument provides $\mathbf{v}_i = \mathbf{w}_i$ for all such column matrices. □

Now the recipient calculates the modular inverse C of A and takes the following steps (which are essentially reversals of the encryption steps above):

MPWZLDKVNGOQPWSJCNVGZ

↓ replace each letter with its corresponding integer

$\{u_1, u_2, \ldots, u_{nk}\}$

↓ block these nk integers, n at a time, into k column matrices

$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$

↓ multiply each column matrix by decryption key matrix C

$\{C\mathbf{u}_1, C\mathbf{u}_2, \ldots, C\mathbf{u}_k\}$

↓ reduce entries of all matrices $C\mathbf{u}_i$ mod 26

$\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\} \overset{*}{=} \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$

↓ 'unblock': use column matrix entries to form a sequence of integers

$\{b_1, b_2, \ldots, b_{nk}\}$

↓ substitute letters for integers

attackthecastleatdawn

where we have used Corollary 1 for the $\overset{*}{=}$ equality.

4. EXAMPLES OF ENCRYPTION AND DECRYPTION USING THE HILL CIPHER

Now let's work out an explicit example. We will choose a plaintext message, fix a block size, build an encryption key matrix, encrypt, and decrypt. We return to our collaborators, Lin and Al, from Section 1. Lin would like to try to encrypt and send Al a simple message: *hello*. However, the world cannot know this! They choose a block size of 2. Their next step is to find an encryption key matrix.

4.1 Determine an Encryption Key Matrix

Lin needs to find a 2×2 integer matrix satisfying the property that $\gcd(26, \det(A)) = 1$; i.e., whose determinant is relatively prime to 26. Since 17 is relatively prime to 26, they can look for an integer matrix $A = \begin{bmatrix} q & r \\ s & t \end{bmatrix}$ such that $\det(A) = 17$. Finding the entries of A that provide them with this condition is done by solving the equation

$$qt - rs = \det(A) = 17.$$

This equation has infinitely many possible solutions, some of which can be found by writing

$$\begin{aligned} qt - rs &= 17 \\ &= 20 - 3 \\ &= 5 \cdot 4 - 3 \cdot 1. \end{aligned}$$

Hence, the encryption key matrix can be $A = \begin{bmatrix} 5 & 3 \\ 1 & 4 \end{bmatrix}$. Now that the encryption key matrix has been determined, they need to set up the plaintext before performing matrix multiplication.

4.2 Convert Plaintext into Matrices of Plain-numbers

The next steps are to convert the message to a sequence of integers and partition this sequence into blocks of size 2. The length of the message, 5, has a remainder of 1 when divided by the block size 2. So, they add a '*z*' at the end of the plaintext to ensure that the plaintext can be partitioned into blocks of length 2. The 'padded' plaintext message is now *helloz*.

The corresponding integers based on Table 1 would be 7, 4, 11, 11, 14, and 25. With blocks of size 2, they have the following block matrices representing the plaintext:

$$\mathbf{v}_1 = \begin{bmatrix} 7 \\ 4 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 11 \\ 11 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 14 \\ 25 \end{bmatrix}.$$

4.3 Multiplying the Block Matrices $\mathbf{v}_i$ *by the Encryption Key Matrix*

Matrix multiplication now gives Lin

$$\mathbf{e}_1 := A\mathbf{v}_1 = \begin{bmatrix} 5 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 7 \\ 4 \end{bmatrix} = \begin{bmatrix} (5)(7) + (3)(4) \\ (1)(7) + (4)(4) \end{bmatrix} = \begin{bmatrix} 47 \\ 23 \end{bmatrix},$$

$$\mathbf{e}_2 := A\mathbf{v}_2 = \begin{bmatrix} 5 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 11 \\ 11 \end{bmatrix} = \begin{bmatrix} (5)(11) + (3)(11) \\ (1)(11) + (4)(11) \end{bmatrix} = \begin{bmatrix} 88 \\ 55 \end{bmatrix},$$

$$\mathbf{e}_3 := A\mathbf{v}_3 = \begin{bmatrix} 5 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 14 \\ 25 \end{bmatrix} = \begin{bmatrix} (5)(14) + (3)(25) \\ (1)(14) + (4)(25) \end{bmatrix} = \begin{bmatrix} 145 \\ 114 \end{bmatrix}.$$

These column matrices $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ have been encrypted via multiplication by our encryption key matrix A, but they have not been converted into ciphertext.

4.4 Convert Matrices of Cipher-numbers into Ciphertext

Conversion of $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ into a ciphertext message takes three steps:

1) reduce each $\mathbf{e}_i \bmod 26$ to obtain three column matrices $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ with entries between 0 and 25,

2) 'unblock' these matrix entries to provide a string of integers, and
3) convert each such integer into a letter of the alphabet.

For example, for the vector $\mathbf{e}_1$ whose entries are 47 and 23, Lin needs to reduce each entry mod 26. The remainders of these numbers when divided by 26 are, respectively, 21 and 23. Using the notation of the previous section, Lin then has

$$\mathbf{u}_1 := \mathbf{e}_1 \bmod 26 = \begin{bmatrix} 47 \\ 23 \end{bmatrix} = \begin{bmatrix} 21 \\ 23 \end{bmatrix}, \quad \text{and similarly...}$$

$$\mathbf{u}_2 := \mathbf{e}_2 \bmod 26 = \begin{bmatrix} 88 \\ 55 \end{bmatrix} = \begin{bmatrix} 10 \\ 3 \end{bmatrix},$$

$$\mathbf{u}_3 := \mathbf{e}_3 \bmod 26 = \begin{bmatrix} 145 \\ 114 \end{bmatrix} = \begin{bmatrix} 15 \\ 10 \end{bmatrix}.$$

Unblocking and converting these integers back into letters yields the ciphertext *VXKDPK*. In a grand romantic gesture, Lin hires a skywriting airplane to write this message in the sky over the beaches of Annapolis.

Nursing a hot cocoa by the oceanside, Al sees the message *VXKDPK* appear in the sky above them – and they know what to do. A few days prior, Lin had shared with Al the secret encryption key matrix for their communication, unkown to all others – the matrix $A = \begin{bmatrix} 5 & 3 \\ 1 & 4 \end{bmatrix}$.

4.5 Finding the Decryption Key Matrix

Al proceeds by finding the modular inverse of the encryption key matrix A. They first calculate the determinant of this matrix: $\det(A) = 5 \cdot 4 - 1 \cdot 3 = 17$, and then set out to find the modular inverse of 17. Al can do this using the Euclidean Algorithm. Or, he can calculate the reduced value of $17 \cdot k \bmod 26$ for all integers k with $0 \leq k < 26$ to find k such that $17 \cdot k \equiv 1 \bmod 26$. After doing this, Al finds that $17 \cdot 23 = 391 \equiv 1 \bmod 26$, and so the modular inverse of $\det(A)$ is 23.

Next, using this modular inverse and the well-known formula for the inverse of a 2×2 matrix given in Theorem 3, Al can find the modular inverse of A:

$$A^{-1} = \frac{1}{17}\begin{bmatrix} 4 & -3 \\ -1 & 5 \end{bmatrix} = 17^{-1}\begin{bmatrix} 4 & -3 \\ -1 & 5 \end{bmatrix} = \begin{bmatrix} 23 \cdot 4 & 23 \cdot (-3) \\ 23 \cdot (-1) & 23 \cdot 5 \end{bmatrix} \equiv \begin{bmatrix} 92 & -69 \\ -23 & 115 \end{bmatrix} \equiv \begin{bmatrix} 14 & 9 \\ 3 & 11 \end{bmatrix} \bmod 26.$$

Therefore, Al's decryption matrix is the modular inverse $C = \begin{bmatrix} 14 & 9 \\ 3 & 11 \end{bmatrix}$.

4.6 Convert Ciphertext into Matrices of Cipher-numbers

With this decryption matrix C in hand, Al's next step is to take Lin's skywritten ciphertext *VXKDPK* and convert it into a string of integers. Using our Table 1, Al gets the integer sequence $\{21, 23, 10, 3, 15, 10\}$. Al knows that Lin used blocks of size 2 to encrypt the matrix, since that was the size of their encryption key matrix A. Al uses these integer entries to write down three column matrices:

$$\mathbf{u}_1 = \begin{bmatrix} 21 \\ 23 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 10 \\ 3 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 15 \\ 10 \end{bmatrix}.$$

4.7 Multiplying Block Matrices $\mathbf{u}_i$ by the Decryption Key Matrix

With this decryption matrix C in hand and carefully concealed from prying eyes, Al now multiplies each encoded matrix $\mathbf{u}_i$ by C to reveal the *deciphered* matrices, and then *reduces* the result of this multiplication mod 26, as we see here.

$$C\mathbf{u}_1 = \begin{bmatrix} 14 & 9 \\ 3 & 11 \end{bmatrix} \begin{bmatrix} 21 \\ 23 \end{bmatrix} = \begin{bmatrix} (14)(21) + (9)(23) \\ (3)(21) + (11)(23) \end{bmatrix} = \begin{bmatrix} 501 \\ 316 \end{bmatrix} \equiv \begin{bmatrix} 7 \\ 4 \end{bmatrix} \bmod 26,$$

$$C\mathbf{u}_2 = \begin{bmatrix} 14 & 9 \\ 3 & 11 \end{bmatrix} \begin{bmatrix} 10 \\ 3 \end{bmatrix} = \begin{bmatrix} (14)(10) + (9)(3) \\ (3)(10) + (11)(3) \end{bmatrix} = \begin{bmatrix} 167 \\ 63 \end{bmatrix} \equiv \begin{bmatrix} 11 \\ 11 \end{bmatrix} \bmod 26,$$

$$C\mathbf{u}_3 = \begin{bmatrix} 14 & 9 \\ 3 & 11 \end{bmatrix} \begin{bmatrix} 15 \\ 10 \end{bmatrix} = \begin{bmatrix} (14)(15) + (9)(10) \\ (3)(15) + (11)(10) \end{bmatrix} = \begin{bmatrix} 300 \\ 155 \end{bmatrix} \equiv \begin{bmatrix} 14 \\ 25 \end{bmatrix} \bmod 26.$$

Following the notation used in Section 3, our deciphered and mod 26-reduced column matrices are

$$\mathbf{w}_1 = \begin{bmatrix} 7 \\ 4 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 11 \\ 11 \end{bmatrix}, \mathbf{w}_3 = \begin{bmatrix} 14 \\ 25 \end{bmatrix}.$$

The reader is encouraged to check that Al's deciphered column matrices $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$ are precisely the column matrices $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ written down by Lin in Section 4.2, as guaranteed by Corollary 1. Al can take the integer entries from these matrices, and write them down in order to obtain the deciphered integer sequence $\{7, 4, 11, 11, 14, 25\}$.

4.8 Convert Deciphered Numbers to Deciphered Text

Al's final step is just to read off the letters corresponding to these integer, using Table 1. Doing so Al recovers the message *helloz*. The convention of padding the plaintext is well-known to Al, so he strips off the appended '*z*' to reveal the deciphered text, Lin's original message.

5. PLAINTEXT ATTACKS ON THE HILL CIPHER

As mentioned in Section 1, the Hill cipher is notable for its introduction of more sophisticated (at the time) linear algebraic methods of encryption. Nevertheless, the method had very limited practical use. At the time of its introduction, lack of computing technology made implementation of the algorithm impractical. Ironically, the development of computing power sufficient to implement the algorithm brought with it enough power to easily attack the cipher.

There are several ways to attack a Hill cipher. For example, if an eavesdropper intercepts a significant amount of ciphertext, a row-by-row reconstruction of the encryption key matrix is feasible using statistical methods paired with data on frequencies of n-grams in the English language (See [1] for more details). Another possibility is that an eavesdropper intercepts some ciphertext *and* knows (or suspects) the corresponding plaintext. An attack based upon this data is called a plaintext attack. We will focus on such a plaintext attack on the Hill cipher.

A plaintext attack assumes that the eavesdropper has access to part of the plaintext. In other words, the eavesdropper has intercepted the ciphertext and (knows or has suspicions on) the corresponding original plaintext. According to [8], "making a good guess as to how just two letter blocks should be decrypted, we can deduce the matrix that will decrypt the entire message."

Suppose Lin sends Al another encrypted message, which is intercepted by a nosy and clever eavesdropper named Trinity. She intercepts the ciphertext *VX KD SO KD GG FY RE*. Suppose Trinity also knows part of the original message: the plaintext for the first two blocks *VX* and *KD*, which are *he* and *ll* respectively. Trinity's goal is to reconstruct the 2×2 encryption matrix A used by Lin to encode the message. Since Trinity knows that *VX* decrypted is *he*, they can replace these letters with their corresponding integers to write

$$\begin{bmatrix} q & r \\ s & t \end{bmatrix} \begin{bmatrix} 21 \\ 23 \end{bmatrix} \equiv \begin{bmatrix} 7 \\ 4 \end{bmatrix} \bmod 26.$$

Similarly, Trinity knows that *KD* decrypted is *ll*, and so they can write

$$\begin{bmatrix} q & r \\ s & t \end{bmatrix} \begin{bmatrix} 10 \\ 3 \end{bmatrix} \equiv \begin{bmatrix} 11 \\ 11 \end{bmatrix} \bmod 26.$$

Performing the matrix multiplication between the decryption matrices and the cipher-numbers on the left hand side and setting its equivalence to the corresponding row on the right hand side, we get following equations:

$$\begin{aligned}(21)(q) &+ (23)(r) &&\equiv 7 \bmod 26\\ (21)(s) &+ (23)(t) &&\equiv 4 \bmod 26,\\ (10)(q) &+ (3)(r) &&\equiv 11 \bmod 26\\ (10)(s) &+ (3)(t) &&\equiv 11 \bmod 26.\end{aligned}$$

Combining the equations with the same variables, we have the following equations:

$$\begin{aligned}(21)(q) &+ (23)(r) &&\equiv 7 \bmod 26\\ (10)(q) &+ (3)(r) &&\equiv 11 \bmod 26\\ (21)(s) &+ (23)(t) &&\equiv 4 \bmod 26\\ (10)(s) &+ (3)(t) &&\equiv 11 \bmod 26.\end{aligned}$$

These equations can be rewritten as matrix equations. For the equations with variables q and r, Trinity obtains

$$\begin{bmatrix}21 & 23\\ 10 & 3\end{bmatrix}\begin{bmatrix}q\\ r\end{bmatrix} \equiv \begin{bmatrix}7\\ 11\end{bmatrix} \bmod 26.$$

For the equations with variables s and t,

$$\begin{bmatrix}21 & 23\\ 10 & 3\end{bmatrix}\begin{bmatrix}s\\ t\end{bmatrix} \equiv \begin{bmatrix}4\\ 11\end{bmatrix} \bmod 26.$$

To find the variables q, r, s, and t, Trinity only needs to find the modular inverse of the matrix $\begin{bmatrix}21 & 23\\ 10 & 3\end{bmatrix}$ (mod 26). This matrix has determinant $21 \cdot 3 - 23 \cdot 10 = -167$. The Euclidean Algorithm can be used to reveal that this determinant is relatively prime to 26, hence the existence of a modular inverse for this matrix is guaranteed. Note that $-167 \equiv 15 \bmod 26$, and the modular inverse of 15 is 7. Therefore, the method given in the proof of Theorem 5, and the following example, yields

$$\begin{bmatrix}21 & 23\\ 10 & 3\end{bmatrix}^{-1} = \frac{-1}{167}\begin{bmatrix}3 & -23\\ -10 & 21\end{bmatrix} = \begin{bmatrix}3\cdot 7 & -23\cdot 7\\ -10\cdot 7 & 21\cdot 7\end{bmatrix} \equiv \begin{bmatrix}21 & -161\\ -70 & 147\end{bmatrix} \equiv \begin{bmatrix}21 & 21\\ 8 & 17\end{bmatrix} \bmod 26.$$

Trinity can use this modular inverse $D = \begin{bmatrix} 21 & 21 \\ 8 & 17 \end{bmatrix}$ to find the entries q, r, s, and t of Lin's encryption matrix A:

$$\begin{bmatrix} q \\ r \end{bmatrix} \equiv \begin{bmatrix} 21 & 21 \\ 8 & 17 \end{bmatrix} \begin{bmatrix} 7 \\ 11 \end{bmatrix} \equiv \begin{bmatrix} 14 \\ 9 \end{bmatrix} \mod 26;$$

$$\begin{bmatrix} s \\ t \end{bmatrix} \equiv \begin{bmatrix} 21 & 21 \\ 8 & 17 \end{bmatrix} \begin{bmatrix} 4 \\ 11 \end{bmatrix} \equiv \begin{bmatrix} 3 \\ 11 \end{bmatrix} \mod 26.$$

With this calculation, Trinity has recovered the entries of the modular inverse of Lin's original encryption key matrix A: $q = 14$, $r = 9$, $s = 3$, and $t = 11$, hence

$$C \equiv \begin{bmatrix} 14 & 9 \\ 3 & 11 \end{bmatrix}.$$

Exercise for the reader: Decode the rest of Lin's message, using the encryption key cracked by Trinity!

6. CODING THE HILL CIPHER IN PYTHON

We provide this code as a resource for readers who might be interested in further exploring the Hill cipher and its uses in other encryption schemes. In this section, we will explain sections of the code and include code cells in the greyed out enclosures. This section will include all lines of code to run your own Hill cipher in the Python programming language. This code does not generate new encryption and decryption key matrices. These key matrices were given to the program. It currently only allows encryption and decryption of the 26 letters of the English alphabet. In each code cell, comments begin with a hashtag (#) or are enclosed in quotations marks (“”), and they are in blue.

```
#import numpy library
import numpy as np
```

6.1 Encryption

The determinant of the encryption key matrix has to be relatively prime to 26. The following encryption matrices' determinants, 17, 3, and 5, are all relatively prime to 26. The following matrices are our encryption keys. With them, our code can encrypt in blocks of size 2, 3, and 4.

The next block of code enters these matrices into the program.

$$KEY\_A = \begin{bmatrix} 5 & 3 \\ 1 & 4 \end{bmatrix}, KEY\_B = \begin{bmatrix} 2 & 5 & 10 \\ 3 & 5 & 8 \\ 1 & 2 & 3 \end{bmatrix}, KEY\_C = \begin{bmatrix} 13 & 3 & 4 & 1 \\ 24 & 6 & 5 & 1 \\ 9 & 5 & 5 & 1 \\ 8 & 3 & 4 & 1 \end{bmatrix}$$

```
#Note:Each inner array is a row of the matrix
KEY_A= [[5,3],[1,4]]
KEY_B = [[2,5,10],[3,5,8],[1,2,3]]
KEY_C = [[13,3,4,1],[24,6,5,1],[9,5,5,1],[8,3,4,1]]
```

The following lines allow us to enter the encryption key size that we would like to use to encrypt the message. Based on the keys we have, we can type in 2, 3, or 4.

```
key = int(input()) #user input for encryption key size
print(key)
```

The user is prompted here to enter their plaintext message to be encrypted and decrypted.

```
message = input() #user input for plaintext message
print(message)
```

The function *plaintext* below converts letters of the plaintext to their corresponding values based on Table 1. The function removes the spaces, converts all of the characters in the string to lower case, and converts them to their corresponding integer values.

```
def plaintext(message):
    """Accepts a plaintext message, removes spaces, and converts to lower case.
        Then converts each character to numerical value. Returns an array of
        the corresponding integers. Note:This function does not include any
       special characters in the array of integers."""

    message_integers = []#make an array to store integers

    message_no_space = message.replace(" ","") #remove spaces from the string
    message_lower_case = message_no_space.lower() #convert to lower case
    message_char = list(message_lower_case)  #Split string into characters

    for a_letter in message_char :
        converted_letter = ord(a_letter) - 97 #Use ord() function to get ASCII
            value of character and subtract 97 to get a = 0
        if (converted_letter >= 0 and converted_letter<=25 ) : #add the letter
            to the message integers array, if it is a-z (0-25).
            message_integers = np.append(message_integers,converted_letter)
    return message_integers
```

These lines below save the array of integers to the variable named message_in_numbers.

```
message_in_numbers = plaintext(message) #store the array of integers
print(message_in_numbers)
```

The next step is to convert the array plain-numbers into blocks based on the size of the key. The *block* function below takes the encryption key matrix size chosen earlier and the array of integers to be blocked.

```
def block(block_size,to_block):
    """Accepts the block size and array of integers to be blocked. Returns
       reshaped array of integers."""
    blocked = [] #holds array of numbers that have been blocked

    #add the letter z as needed by looking at the remainders.
    remainder_of_block_sizes = len(to_block)%block_size #get the remainder of
       the size of to_block array divided by block_size
    #while loop to add to to_block array while remainder is not zero.
    while remainder_of_block_sizes != 0:
        to_block = np.append(to_block,25)#add 25 until remainder is zero.
        remainder_of_block_sizes = len(to_block)%block_size #update
           remainder_of_block_sizes with new remainder of to_block array
           divided by block_size

    #reshape the 1D array to 2D using reshape
    blocked = np.reshape(to_block,(-1,block_size))
    return blocked
```

To check the output, we will do an assignment and print statement of the message_in_numbers _blocked.

```
message_in_numbers_blocked = block(key,message_in_numbers)
print(message_in_numbers_blocked)
```

After setting up the plaintext, we can now encrypt using matrix multiplication and modular arithmetic. The function *encrypt* below accepts the array of arrays and multiplies it by one of the encryption key matrices, based on the chosen key size. This function then performs matrix multiplication for each plain-number (or plaintext of integers) matrix and the encryption key matrix. We also perform modular arithmetic mod 26 to ensure that we can later convert each number to an equivalent alphabet letter (using Table 1).

```
def encrypt(plaintext_integers):
    """Accepts plaintext integers array. Performs matrix multiplication based
    on the key chosen. Returns product mod 26. """
    encrypted_message = [] #holds the product of the matrix multiplication,
        the cipher-numbers.

    if key == 2:
        KEY = KEY_A
    if key == 3:
        KEY = KEY_B
    if key == 4:
        KEY = KEY_C

    for a_block_plaintext in plaintext_integers :
        encrypted_number = np.dot(KEY, a_block_plaintext)%26 #use np.dot
            function to perform matrix multiplication and reduce mod 26
        encrypted_message = np.append(encrypted_message,encrypted_number)
        #store the cipher-numbers
    return encrypted_message
```

Similarly, check that we have expected output from the *encrypt* function.

```
message_in_numbers_encrypted=encrypt(message_in_numbers_blocked)
print(message_in_numbers_encrypted)
```

Now that we have the cipher-numbers (which are encrypted plain-numbers), we can convert the cipher-numbers to their corresponding letters from Table 1 to get the ciphertext. The following function *ciphertext* takes in the array of arrays of cipher-numbers and converts it to letters. It returns the corresponding string in all upper case.

```
def ciphertext(encrypted_msg_num):
    """Accepts array of integers. Converts array into letters. Returns a
        string in upper case. Note: This function does not include any special
         characters."""
    encrypted_msg = []#make an array to store integers

    #for loop for integers in the encrypted_msg_int array
    for a_number in encrypted_msg_num :
        converted_number = chr(int(a_number) + 97)#convert a_number into an
            int using int() function,add 97 to match the ASCII values, and use
             chr() function to convert the integer into a character
        encrypted_msg = np.append(encrypted_msg,converted_number)

    encrypted_msg_str = ""
    encrypted_msg_str = encrypted_msg_str.join(encrypted_msg) #use join()
        function to convert array of characters into a string.

    return encrypted_msg_str.upper()
```

Now, we need to store the ciphertext to decrypt the same message. This will be a way to check that the calculations are correct.

```
message_in_letters_encrypted = ciphertext(message_in_numbers_encrypted)
print(message_in_letters_encrypted)
```

6.2 Decryption

All code below will be for the decryption process. This process is shorter in the program because we get to reuse some of the functions above from the encryption process. We are starting with the ciphertext when we decrypt using the inverse of the key.

The function *decrypt* accepts the ciphertext string from the encryption process above. Then, it uses the modular inverse of the key. It then uses the same *plaintext* function from encryption to convert the string into an array of integers, and the *block* function to block the array of integers into an array of arrays. Then we use matrix multiplication and modular arithmetic to get the deciphered array of integers.

```
def decrypt(ciphertext_str):
    """Accepts a string. Performs matrix multiplication using decryption key
       matrix. Returns product mod 26, an array of integers."""
    #Note: the following are modular inverses of the encryption matrix keys
    if key == 2:
        inv_KEY = [[14,9],[3,11]]
    if key == 3:
        inv_KEY = [[17,19,14],[17,16,22],[9,9,7]]
    if key == 4:
        inv_KEY = [[21,0,0,5],[23,1,25,3],[11,24,3,14],[5,5,17,0]]

    ciphertext_int = plaintext(ciphertext_str)#Converts str to int array
    ciphertext_int_blocked = block(key,ciphertext_int) #blocks ciphertext_int
    deciphered_message = [] #holds product of matrix multiplication of
       inverted key and ciphertext

    for a_block_ciphertext in ciphertext_int_blocked:
        deciphered_number = np.dot(inv_KEY, a_block_ciphertext)%26
        deciphered_message = np.append(deciphered_message,deciphered_number)
    return deciphered_message
```

We call our *decrypt* function to save the deciphered numbers as an array.

```
message_in_numbers_decrypted = decrypt(message_in_letters_encrypted)
print(message_in_numbers_decrypted)
```

Last, we need to convert deciphered-numbers into our final deciphered text string. This function is similar to the *plaintext* function in Section 6.1, but here we use a different formula to get back to the original message.

```
def decipheredtext(deciphered_msg_num):
    """Accepts the array of numbers. Converts it into letters. Returns the
       string in lowercase. """
    decrypted_msg = []
    #for loop for integers in deciphered_msg_num array
    for a_number in deciphered_msg_num :
        converted_number = chr(int(a_number) + 97)
        decrypted_msg = np.append(decrypted_msg,converted_number)

    deciphered_msg_str = ""
    deciphered_msg_str = deciphered_msg_str.join(decrypted_msg)

    return deciphered_msg_str
```

To check, we display the deciphered text. We should get the original message (with possible additional padding).

```
message_in_letters_decrypted = decipheredtext(message_in_numbers_decrypted)
print(message_in_letters_decrypted)
```

7. SUMMARY AND FUTURE RESEARCH DIRECTIONS

Although the computing power available today renders the Hill cipher obsolete as a stand-alone encryption method, the Hill cipher provides a valuable and accessible introduction to important methods in number theory and linear algebra that are used in a wide variety of encryption methods. In addition, the Hill cipher method is used as part of encryption of images along with 1D chaotic maps; see [2] for further details. Our research described a plaintext attack on the Hill cipher. Future coding projects could include the implementation of a software-based approach to breaking the Hill cipher using plaintext attack, ciphertext only, chosen plaintext or chosen ciphertext attacks, or the use of the Hill cipher as a diffusion method in combination with other encryption algorithms.

REFERENCES

[1] Craig Bauer and Katherine Millward, *Cracking matrix encryption row by row*, Cryptologia. 31 (2007), no. 1, 76–83, DOI 10.1080/01611190600947806.

[2] M. Essaid, I. Akharraz, A Saaidi, and A. Mouhib, *Image encryption scheme based on a new secure variant of Hill cipher and 1D chaotic maps*, Journal of Information Security and Applications. 47 (2019), 173–187, DOI 10.1016/j.jisa.2019.05.006.

[3] Lester S. Hill, *Concerning certain linear transformation apparatus of cryptography*, The American Mathematical Monthly. 38 (1929), no. 3, 135–54, DOI https://doi.org/10.2307/2300969.

[4] _____, *Cryptography in an algebraic alphabet*, The American Mathematical Monthly. 36 (1929), no. 6, 306–312, DOI https://doi.org/10.2307/2298294.

[5] Ron Larson, *Elementary linear algebra*, eighth ed., Cengage Learning, 2017.

[6] Weisner Louis and Lester Hill, *Message protector*, https://patents.google.com/patent/US1845947A/en, 1929.

[7] Kenneth Rosen, *Elementary number theory and its applications*, third ed., Pearson, 2010.

[8] The Department of Mathematics and Computer Science at Emory Oxford College, *The Hill Cipher*, http://mathcenter.oxford.emory.edu/site/math125/hillCipher/.

[9] Lawrence C. Washington, *Introduction to cryptography with coding theory*, second ed., Pearson, 2006.

LAUREN E. STREET

Alcohol Abuse: Causes, Effects, and Potential Solutions through a Biopsychosocial Lens

ABSTRACT

Substance use disorder is defined as the perpetual craving and repeated use of a drug despite its negative impact on the user and their overall well-being. One drug that is very common in substance use disorders is alcohol. Alcohol serves as a stimulant drug in small doses, but when large amounts are consumed, it acts as a depressant. There are a number of biological, psychological, and social causes and negative effects of alcohol use disorder. Despite often being taken for granted, alcohol contributes to a significant number of deaths in the United States every year. Studying and understanding alcohol use disorder through what is termed a "biopsychosocial lens" can help researchers and health officials continue to determine the causes for this disorder, as well as potential treatments for individuals living with it. Additionally, cultural differences must be considered when making any generalizations about alcohol use disorder, or the groups of people that it impacts. Future research will likely continue to build on what researchers already know and may eventually lead to a better understanding of the disorder, and even more effective methods of treatment.

KEY WORDS

alcohol use disorder

substance use disorder

alcoholism

abuse

treatment

FACULTY MENTOR

Rachell Tannenbaum, Ph.D.
Professor, Psychology Department

SUBSTANCE ABUSE AND ALCOHOLISM

Substances such as drugs and other chemicals have the ability to temporarily alter an individual's state of consciousness and bend their reality. There is a variety of reasons a person may choose to use drugs including to improve health or relieve pain, for religious purposes, or sometimes just for fun. Moderate use of prescribed or legalized recreational drugs can often give users their desired experience without any maladaptive consequences; however, many individuals use the drug so frequently that their moderate use becomes a substance abuse disorder (Myers & DeWall, 2018, 101). Substance abuse disorder refers to the perpetual craving and repeated use of a drug despite its negative impact on the user's physical health and overall life. Those who struggle with this disorder face its devastating impacts on their lives every day, and often struggle to recover from it.

One drug common with substance abuse disorders is alcohol. Alcohol is classified as a stimulant drug in small doses, but if the user consumes a large amount of the drug it functions as a depressant. Like any drug, alcohol alters the brain's regular functioning. When it enters the brain, its effects include (a) increasing the efficiency of the inhibitory neurotransmitter GABA, and (b) impeding the ability of glutamate, an excitatory neurotransmitter, to bind to receptor sites in the brain (Genetic Science Learning Center, 2013). This double inhibitory effect is what slows the brain's neural activity as well as the bodily functions, producing the depressant effect. For many Americans, drinking alcohol is a "cool" and highly sought-after way to relax or have fun. In many cases, it seems like a nearly harmless way to kick off a weekend for a wide range of ages. Unfortunately, this perception that drinking alcohol is a casual, completely harmless activity is far from accurate, as the drug contributes to roughly 95,000 deaths a year in the United States alone, making it the country's third most common preventable cause of death (National Institute on Alcohol Abuse and Alcoholism, 2021).

ALCOHOLISM: PROBLEM JUSTIFICATION

Because of the common misconception that alcohol is a casual and harmless drug, it is one that many individuals start using at a very young age. According to the National Institute on Alcohol Abuse and Alcoholism, approximately 4% of all alcohol consumed in the United States is consumed by individuals ages 12 to 20, making it the most used drug among the country's youth (National Institute on Alcohol Abuse and Alcoholism, 2021). While the reasons for consuming alcohol at any age are specific to the individual, the oftentimes tragic effects impact many.

An individual's drinking may physically harm only themselves directly, with slowed body functions, impaired memory, and liver damage, but it can indirectly harm others in numerous ways as well. The Motor Vehicle Crash Data Report released in 2021 by the National Highway Traffic Safety Administration (NHTSA) showed that in the year 2019, 28% of all traffic fatalities in the United States were alcohol related (pg. 1). In many cases, the impaired driver is not the only one injured in these car crashes, and innocent lives are lost. Aside from alcohol's effects behind the wheel, alcohol use contributes to around 700,000 assaults, including nearly 97,000 sexual assaults in the United States each year (Myers & DeWall, 2018). Like most abused drugs, alcohol affects many more people than just the user, with those additional people impacted typically being those closest to the user. When an individual has crossed the threshold to alcohol use disorder, family and friends begin to suffer the consequences of alcoholism. Every substance abuse disorder involves maladaptive effects on the user's daily life. Whether they are actively out drinking or away seeking treatment for their condition, an individual's reliance on alcohol may impede their ability to maintain relationships. In any case, the loss of these relationships may lead to the loss of support that an individual dealing with alcoholism may have in their life, possibly extending their battle with addiction by decreasing the

likelihood of them seeking treatment. In a broader spectrum, alcohol use disorder also negatively affects the nation overall, costing the United States more than $249 billion every year (Witkiewitz et al., 2019). While it is true that people who drink alcohol responsibly typically experience few negative effects, many do not do so responsibly. These devastating effects have the potential to worsen as an individual's drinking becomes more excessive, resulting in the development of alcohol use disorder.

Alcoholism: Causes and Effects

Approximately 14.5 million people were diagnosed with alcohol use disorder in 2019 (Substance Abuse and Mental Health Services Administration, 2019, pg. 35). Many individuals allow their cognitive bias to overrule logic and ignore the evidence, becoming overconfident in their ability to refrain from developing such disorders, and ultimately believing that they are an exception to the statistics. For years, researchers have studied possible causes as to why people develop alcohol use disorder, thoroughly studying the issue from biological, social-cultural, developmental, and physical/mental health perspectives.

Research suggests that individuals with a certain nucleotide polymorphism in their DNA may be more prone to alcohol dependence and abuse (Kareken et al., 2010). The altered DNA affects their brains' reward responses, which in experienced drinkers can lead to more positive experiences associated with alcohol, and therefore increased usage, which increases the odds of addiction. Regardless of age or gender, the more alcohol an individual consumes, the higher their tolerance becomes, and hence the more they must consume to achieve those reward responses in the brain. Biologically speaking, men are more likely than women to develop a dependence on the drug (National Institute on Alcohol Abuse and Alcoholism, 2021). This may be a result of differences in emotion processing or coping mechanisms for trauma or stress.

Furthermore, those diagnosed with mental illnesses such as bipolar disorder and depression are likely to use alcohol and other substances as a coping mechanism and worsen their illness as a result (Smith et al., 2021). Excessive consumption of alcohol contributes to the development of certain psychiatric disorders (U.S. Department of Health and Human Services, 2021). Though alcohol has temporary stimulating effects, and sometimes gives those who are struggling with stress an "escape," the drug provides no long-term positive effects.

From a social-cultural perspective, researchers have found that those who engage in drinking alcohol do so to fit in or keep up with what they think society's expectation is for them. One study found that drinking in adolescents is heavily influenced by their friendship statuses with their peers. The study concluded that in social groups where "friendship status" mattered (such as cooperative team sports), adolescents were more likely to drink with only "reciprocated friends" (in other words, those who mutually considered said peer as a friend). Conversely, in groups where friendships status was less important, such as school clubs and activities, adolescents drank with peers regardless of whether they were reciprocated friends (Fujimoto & Valente, 2013). In teams and cooperative groups, one may be more concerned with their peers' perception of them, only engaging in activities that they are sure will be accepted. In less cooperative groups, the adolescents subsequently paid less attention to what was acceptable in the group, perhaps in an attempt to act "cool" or rebellious. Regardless of what is socially acceptable or expected at a given time, people tend to act in ways based on what they believe would make them fit into a particular group. The more a person believes they should be drinking, the more they will, which can often lead to abuse, and is a large reason so many people ages 12-17 have alcohol use disorder (U.S. Department of Health and Human Services, 2021).

In terms of social-cultural effects of alcohol use disorder, an individual abusing alcohol tends to neglect many important social aspects of their life to make time for alcohol. This includes but is not limited to decreasing time spent with family and friends and declining motivation for work. This neglect results in a deterioration of the user's relationships and support systems.

ALCOHOLISM: POTENTIAL SOLUTIONS

As more knowledge is gained on the causes and effects of alcohol use disorder, researchers have investigated more effective methods of treatment and prevention. Currently, many individuals seek recovery through Alcoholics Anonymous (AA), a fellowship of "sobriety seekers" who meet with one another to share their stories, strength, and hope with one another as they commit themselves to accepting their wrongs, mending affected relationships, and actively work towards recovering themselves and their lives from their illness (Alcoholics Anonymous). According to one study, AA was 60% more successful than other methods of intervention or no intervention at all, by reducing the participants' consumption of alcohol and increasing the length of time they abstained from drinking alcohol (Erickson, 2020). Though the statistics support AA's effectiveness, some professionals skepticize that the lack of professional involvement in such treatment is cause for concern (Erickson, 2020). This led researchers to a newer method of treatment being studied, motivational interviewing and intervention.

Motivational interviewing and intervention is a communication style that, like AA, encourages the patient to establish their own meaning for their disorder and develop a genuine desire to change their behavior. Several studies, each utilizing different samples of people, have tested the effectiveness of motivational interviewing. In one, over-the-phone motivational interviewing followed by either feedback or psychoeducation (an approach that combines educating the participant about their disorder with

structure and feedback in a safe environment) was used to treat active members of the military dealing with untreated alcohol use disorder (Lukens, 2015). Though all participants decreased their alcohol consumption, those who received feedback reported fewer drinks per week than those who received psychoeducation (Walker et al., 2017). In a different study, individual motivational interventions were used to treat adolescents abusing alcohol. When combined with family checkups, the treatment was found to be even more effective than it was without family checkups at short-term follow-ups three, six, and twelve months into treatment (Spirito et al., 2011). While therapy has been used in the past, motivational interviewing and intervention has since proven to be a more effective method of treatment because it encourages the patient to actively take steps toward their recovery with professional guidance.

The effectiveness of these interviews will also depend on other factors such as cultural and societal expectations. As the country moves toward becoming more open about mental health awareness, and subsequently the rehabilitation of those suffering from said disorders, there will be more success in the researching and studying of ways to treat individuals with alcohol use disorder. However, if a struggling individual's culture and way of life inhibits them from reaching out for help, several previously stated methods will not benefit them. For example, in some cultural groups there is a stigma around seeking help for mental health problems or a preference for seeking help from spiritual or community leaders over health professionals; people may also legitimately be leery after prior experiences, either personal or observed with discrimination in treatment settings (Modir et al., 2022). In addition, although alcohol use disorder does not discriminate by wealth, those who lack financial resources may not be able to afford adequate treatment. Continuing to study this ever-growing issue and its causes may lead to a better understanding of substance abuse in general,

and eventually the development of more effective treatment options and methods of prevention. When discussing alcohol abuse and similar problems, it is important to understand not only the biological, psychological, and social causes and effects, but the interactions between them. Alcohol abuse does not have just one cause or one effect, and these causes are never only biological, one psychological, or only social. Any effective treatment needs to address multiple domains of thought and behavior.

References

Alcoholics Anonymous. (n.d.). *The twelve steps*. https://www.aa.org/the-twelve-steps

Erickson, M. (2020, March 11). *Alcoholics Anonymous most effective path to alcohol abstinence*. Stanford Medicine: News Center. https://med.stanford.edu/news/all-news/2020/03/alcoholics-anonymous-most-effective-path-to-alcohol-abstinence.html

Fujimoto, K., & Valente, T. W. (2013). Alcohol peer influence of participating in organized school activities: A network approach. *Health Psychology, 32*(10), 1084–1092. https://doi.org/10.1037/a0029466

Genetic Science Learning Center. (2013, August 30). *Mouse party*. https://learn.genetics.utah.edu/content/addiction/mouse/

Substance Abuse and Mental Health Services Administration. (2020, September). *Key substance use and mental health indicators in the United States: results from the 2019 national survey on drug use and health* (PEP20-07-01-001). U.S. Department of Health and Human Services. https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR1PDFW090120.pdf

Kareken, D. A., Liang, T., Wetherill, L., Dzemidzic, M., Bragulat, V., Cox, C., Talavage, T., O'Connor, S. J., & Foroud, T. (2010). A polymorphism in GABRA2 is associated with the medial frontal response to alcohol cues in an fMRI study. *Alcoholism: Clinical and Experimental Research, 34*(12), 2169–2178. https://doi.org/10.1111/j.1530-0277.2010.01293.x

Lukens, E. (2015). Psychoeducation. *Oxford Bibliographies Online Datasets*. https://doi.org/10.1093/obo/9780195389678-0224

Modir, S., Alfaro, B., Casados, A., & Ruiz, S. (2020, August 4). *Understanding the role of cultural stigma on seeking mental health services*. https://health.choc.org/understanding-the-role-of-cultural-stigma-on-seeking-mental-health-services/

Myers, D., & DeWall, N. (2018). *Exploring psychology* (11th ed.). Worth Publishers.

National Center for Statistics and Analysis. (2021, October). *State traffic data: 2019 data* (Traffic Safety Facts. Report No. DOT HS 813 183). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813183

National Institute on Alcohol Abuse and Alcoholism. (2021, June). *Alcohol facts and statistics*. U.S. Department of Health and Human Services. https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/alcohol-facts-and-statistics

Smith, M., Segal, J., & Robinson, L. (2021, October). *Dual diagnosis: Substance abuse and mental health*. HelpGuide.org. https://www.helpguide.org/articles/addictions/substance-abuse-and-mental-health.htm

Spirito, A., Sindelar-Manning, H., Colby, S. M., Barnett, N. P., Lewander, W., Rohsenow, D. J., & Monti, P. M. (2011). Individual and family motivational interventions for alcohol-positive adolescents treated in an emergency department. *Archives of Pediatrics & Adolescent Medicine, 165*(3), 269-–274. https://doi.org/10.1001/archpediatrics.2010.296

Walker, D. D., Walton, T. O., Neighbors, C., Kaysen, D., Mbilinyi, L., Darnell, J., Rodriguez, L., & Roffman, R. A. (2017). Randomized trial of motivational interviewing plus feedback for soldiers with untreated alcohol abuse. *Journal of Consulting and Clinical Psychology, 85*(2), 99–110. https://doi.org/10.1037/ccp0000148

Witkiewitz, K., Litten, R. Z., & Leggio, L. (2019). Advances in the science and treatment of alcohol use disorder. *Science Advances, 5*(9), Article No. eaax4043. https://doi.org/10.1126/sciadv.aax4043

AUTHOR NOTE

We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Rachelle Tannenbaum, Anne Arundel Community College, 101 College Parkway, Arnold, MD 21012. Email: retannenbaum@aacc.edu

ALEXANDER THOMPSON

Multispectral Analyses on Drone-Captured Images for Submerged Aquatic Vegetation (SAV) Monitoring

KEY WORDS

drones

SAV

remote sensing

multispectral analysis

FACULTY MENTOR

Tim Tumelty
Instructional Specialist,
Drone Center

ABSTRACT

The important ecological role of submerged aquatic vegetation (SAV) makes its year-to-year distribution of significant interest to environmental monitoring organizations. The use of drones to perform the task of SAV monitoring through multispectral analyses is a promising tool to achieve a methodology that is automatable, repeatable, time efficient, and accessible. A preliminary trial was conducted at Eagle Cove near Gibson Island on the Magothy River where a DJI Phantom 4 Drone with a Sentera special purpose camera captured multispectral digital images with five spectral bands. These were used to apply and compare four vegetation indices: Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Modified Normalized Difference Vegetation Index (mNDVI), and Normalized Difference Aquatic Vegetation Index (NDAVI). Analyses was done using the geographic information systems program known as ArcGIS Pro. The images generated by each index show some measure of successful identification of SAV, though there are many false-positives due to a variety of factors. The effectiveness of each index in our images was estimated by comparing the amount of pixels identified as SAV in the area of

observed SAV growth and outside of this area. The most effective index was indicated to be mNDVI. This methodology will continue to be developed at AACC, and future work will aim to improve upon this process and to make calculations of SAV acreage and density that can be compared to ground-truthed observations.

Introduction

The abundance of submerged aquatic vegetation (SAV) is a critical metric involved in the assessment and monitoring of the biological health of local waterways. SAV is defined as a rooted aquatic plant that grows completely underwater, and can be found throughout the Chesapeake Bay and its tributary rivers. SAV plays an important role in stabilizing water quality by providing oxygen to the water column, filtering sediment, absorbing excess nutrients, buffering pH, and neutralizing acidic conditions. It also protects shorelines from erosion, provides food and habitat for wildlife, and sequesters carbon dioxide ("Chesapeake Bay SAV Watchers"). Therefore, efforts to preserve and propagate SAV are critical, and the monitoring of SAV is of great interest to many organizations that work to conserve our environment.

The challenge of mapping SAV from year to year has been approached with a variety of methods, from traditional surveys using boats to remote sensing using airplanes and satellites. The Virginia Institute of Marine Science (VIMS) conducts annual surveys of the entire Chesapeake Bay, its tributaries, and the Delmarva Coastal Bays by using aerial photography to collect multispectral digital images taken from an aircraft at an altitude of approximately 13,200ft ("Monitoring Methods for SAV"). The aircraft-based methodology used by VIMS is useful because it captures a large amount of data over vast distances, but it is not always accurate and often needs to be verified through ground-truthing at the local level. The use of drones to map SAV may allow for more detailed and more reliable surveys to be performed on a

smaller scale. Drones provide a ground level local observation similar to kayak surveys combined with an overhead sensor similar to crewed aircraft. Drones provide an intermediate tool in terms of scale and detail (Figure 1), while allowing for a methodology that is automatable, easily repeatable, and time efficient. Through the use of computer software, drones are able to follow a programmed route over a study area. A key benefit of using an automated flight program and image capturing process is that it is easy to keep a consistent flight path when monitoring the same site from year to year. Another benefit of drone-based methodology is its accessibility to a wide variety of environmental organizations, many of which are interested in surveying only a particular river or reach in great detail.

Drones can capture detailed photographs in which SAV is visible under the surface of the water, though it is often not enough to rely simply on the viewing of images to accurately find SAV and distinguish it from its surroundings. Multispectral analysis offers a more rigorous methodology for processing the data that is captured

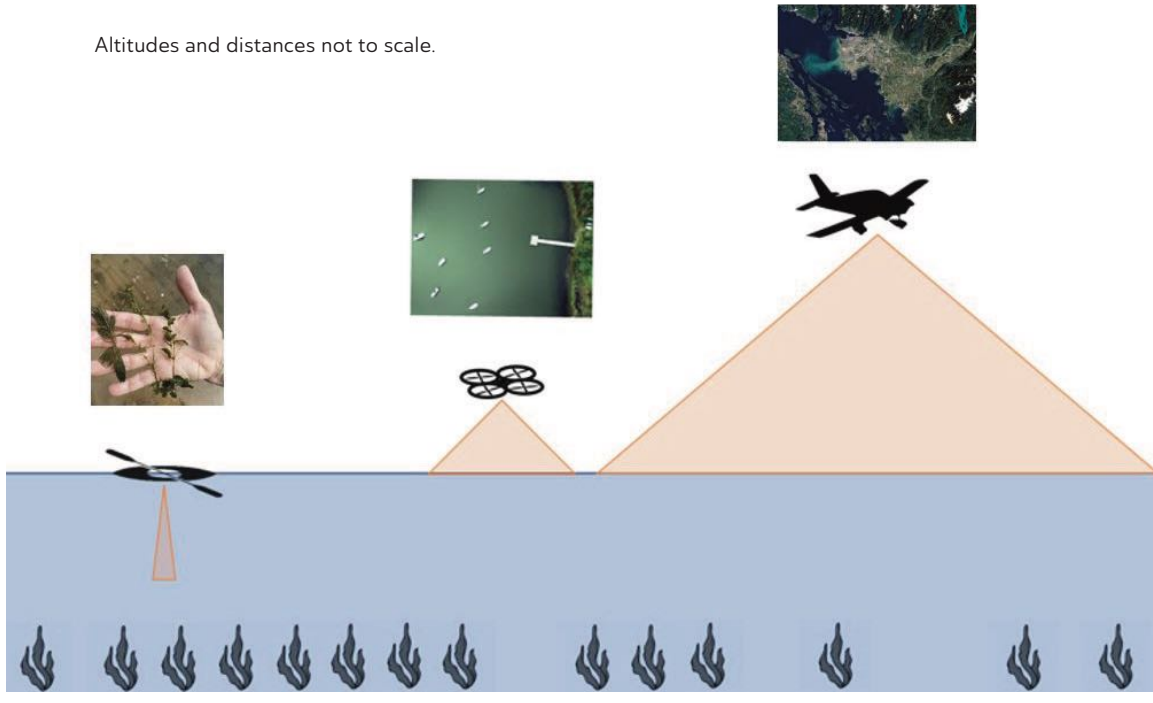


FIGURE 1

Illustration of the varying scales of coverage and the levels of detail observed from boats, drones, and aircraft. Coverage increases from left to right, while detail decreases.

by the drone. The Drone Center at Anne Arundel Community College (AACC), in coordination with the Environmental Center and the Geography department, has begun the development of a drone-based methodology to survey SAV using multispectral analysis. We collected data in the form of digital images captured from a drone on the Magothy River at Eagle Cove near Gibson Island, where SAV was seen to be present. Five bands of reflectance: red (R), green (G), blue (B), red edge (RE), and near-infrared (NIR) were captured and used to calculate vegetation indices, which are combinations of reflectance in two or more bands designed to highlight a particular property of vegetation.

The presence of chlorophyll and photosynthesis causes light absorption in the red region of the electromagnetic spectrum, and consequently vegetation has a very low red reflectance. Due to internal cellular structure, vegetation also has very high reflectance in the NIR region (Rowan and Kalacska). Vegetation indices such as the Normalized Difference Vegetation Index (NDVI) take advantage of this by using the difference between NIR and red reflectance to highlight vegetation. Another index, the Green Normalized Difference Vegetation Index (GNDVI) uses green reflectance rather than red to estimate photosynthetic activity. These indices are often used to gauge the health of crops and forests, but they can also be used to identify and map vegetation, including SAV. One study has shown that NDVI can be suitable for detecting SAV, with the condition that in deeper waters the depth is considered (Jung et al.). However, a drawback to using NDVI and GNDVI in under-water settings is that NIR frequencies have a high degree of absorption by the water column (Rowan and Kalacska). Because of this factor, we also considered other indices that are designed with the aquatic medium in mind. The Modified Normalized Difference Vegetation Index (mNDVI) addresses the issue of NIR attenuation by modifying NDVI to use the RE band instead of NIR (Brooks et al.). Another index that has been shown

to produce good results in under-water studies is the Normalized Difference Aquatic Vegetation Index (NDAVI) which uses blue reflectance (Rowan and Kalacska).

At Eagle Cove a series of 47 overlapping images was taken, which allows for the use of photogrammetric analysis methods. However, for the purposes of this preliminary trial we limited ourselves to selecting only one set of red, green, blue (RGB) and corresponding NIR/RE photos with which to work. The objective of this project was to begin the establishment of a process for image collection and analysis that can be used by AACC or other organizations in the future, and to identify a vegetation index that is effective at finding SAV. This project serves as the preliminary work for a future study that will be the basis of a 2022 Department of Natural Resources grant proposal.

Methodology

The drone images were captured on September 9, 2021 at Eagle Cove, located on the Magothy River at approximately 39° 05' 14" N, 76° 25'30" W. A DJI Phantom 4 drone was mounted with a Sentera 5-band multispectral double 4k camera. The peak wavelength and widths that define each band are given in Table 1. Images were captured using the software Pix4Dcapture to create a pre-programmed flight route over the study area, flying at a programed height of 120 meters. Pix4Dcapture creates a flight path within a user-defined area along which the drone automatically captures overlapping photos (Figure 2). Photos were taken looking straight down at an angle of 90° to the horizon, with a front overlap of 80% and side overlap of 80%. For purposes of documentation, water quality parameters were measured at Eagle Cove during the time of our flight. A Yellow Springs Instrument (YSI) was used to measure standard water quality parameters, in addition to water clarity being measured with a Secchi disk and turbidity with a field kit.

BAND	PEAK WAVELENGTH	WIDTH
Blue	446nm	60nm
Green	548nm	45nm
Red	650nm	70nm
Red Edge	720nm	40nm
Near-Infrared	840nm	20nm

TABLE 1

Spectral band specifications for the Sentera camera.

FIGURE 2

Flight path in Pix4D. Camera icons (black boxes) show the location of each photograph.

The collected images were analyzed using ArcGIS pro. One single RGB and matching NIR/RE image were selected for analysis. Sentera cameras have two different lenses that are spaced

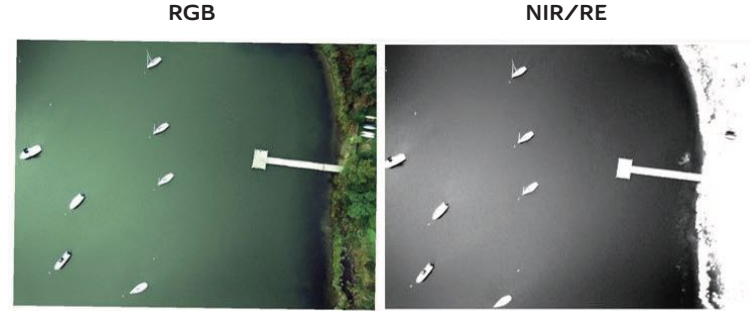
approximately one inch apart, one for RGB and one for NIR/RE. This created the need for the images to be aligned with each other using anchor points before generating a single composite image that includes all five spectral bands. This composite image was used to calculate a variety of different vegetation indices.

DATA

The drone-collected data is in the form of RGB and NIR/RE digital images (Figure 3) and the composite image that combines the spectral bands from both into one single image.

FIGURE 3

RGB and NIR/RE images taken from the drone at Eagle Cove.



The water clarity at Eagle Cove was measured to be 0.6m with a turbidity of 10.75 nephelometric turbidity units (NTUs). The parameters measured with the YSI at both surface and bottom depths are shown in Table 2.

TABLE 2

YSI data from Eagle Cove.

	SURFACE (0.2M)	BOTTOM (1.1M)
Temperature (°C)	23.2	22.5
Dissolved Oxygen (mg/L)	8.75	7.64
Salinity (ppt)	5.79	5.75
pH	8.45	8.64

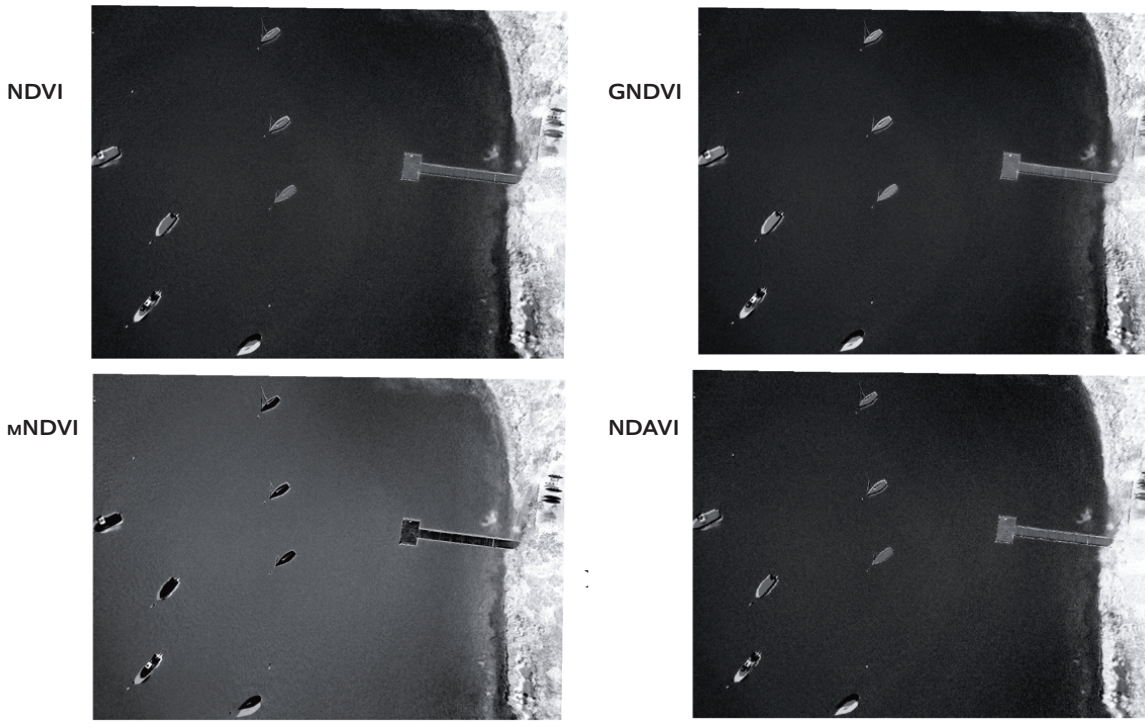


FIGURE 4

Resulting images from the application of vegetation indices.

RESULTS

Several vegetation indices were calculated from the composite image using ArcGIS Pro. Four that appeared to be potentially useful, NDVI, GNDVI, mNDVI, and NDAVI, were selected for further analysis. The images generated from the application of these indices are shown in Figure 4, and the formulas for each index are given in Table 3.

TABLE 3

Vegetation indices and corresponding formulas.

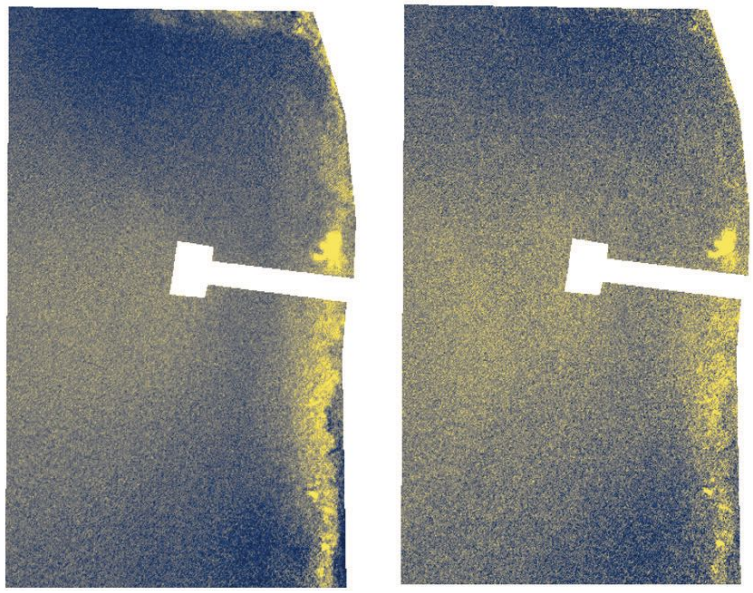
INDEX	FORMULA
NDVI	(NIR - R)/(NIR + R)
GNDVI	(NIR - G)/(NIR + G)
mNDVI	(RE - R)/(RE + R)
NDAVI	(NIR - B)/(NIR + B)

Each index converts the composite image into an image with one single band, shown with a black to white color gradient. Each pixel in the image is assigned a value between -1 and 1 based on the index formula, with higher values displaying brighter and lower values displaying darker. Since these indices are designed to highlight vegetation, the SAV as well as terrestrial plants are seen to be highlighted against the dark water surface. Some parts of manmade objects such as the boats are also highlighted. We would like to remove the noise of terrestrial plants and manmade objects to view only SAV against the water surface. For each image it is therefore necessary to clip out a section that shows only the area of the water where SAV might be found. This was done with the image masking tool in ArcGIS Pro which allows a polygon to be drawn and applied to multiple images to extract a section of each image as its own layer. The mask that was used to clip out our study area, referred to as the "full mask", is shown in Figure 5. The full mask was applied to each index-generated image. Each extracted image was given a new color gradient of blue to yellow, which makes the images easier to view. Two examples of the resulting final images are shown in Figure 6.

FIGURE 5

Masks that were used to analyze only a certain section of the photograph, shown overlain with the RGB image. The full mask was used to examine a broad area where one might look for SAV and that excludes the dock, boats, and land. The SAV mask was used later in the analysis to look at only the area where SAV was known to be present.

FULL MASK **SAV MASK**



MNDVI NDAVI

FIGURE 6

Extracted images from mNDVI and NDAVI, shown with a blue-yellow color gradient.

While these index-generated images succeed at highlighting SAV, a problem with how they turned out is that they also contain many highlighted pixels that are not actually representative of SAV. A comparison of each index's effectiveness in our images was carried out by computing the number of pixels at each .01 step in the index from 1.0 to -1.0 for both the full mask and the SAV mask. The mask that was used to analyze only the area of SAV growth is shown in Figure 5, referred to as "SAV mask". The count of pixels at each step value identified as SAV in the SAV mask was subtracted from that in the full mask to determine how many pixels are not inside of the area where SAV is expected. The assumption being that index returns close to positive 1 in the SAV mask were pixels containing SAV and the corresponding pixels in the count "outside SAV mask" were not SAV and would be considered a false positive. A limit of 85% success rate of bright pixels in the SAV mask was chosen as a way to compare indexes.

To complete Table 4, a summation of all pixel's values was made at each step from +1 to -1 at a .01 step interval. This summation gave a total of pixels in all bins below the index value.

A ratio of Pixels Count Inside SAV Mask/Pixel Count Outside SAV Mask was then determined with 85% being used as the acceptable success rate. This value was chosen because above 85% there was a sharp increase in the number of pixels outside the SAV mask for that index. The lower limit that yielded approximately 85% of identified pixels in the expected area corresponded to a pixel count inside the SAV mask area. The pixel counts inside SAV mask in Table 4 indicate the relative effectiveness of each index at 85% success rate.

TABLE 4

Counts of pixels identified as SAV and corresponding lower limits for each index.

INDEX	LOWER LIMIT FOR SAV PIXELS	PIXEL COUNT INSIDE SAV MASK	PIXEL COUNT OUTSIDE SAV MASK	PERCENTAGE OF PIXELS INSIDE SAV MASK
mNDVI	0.33	166,987	23,975	85.6%
GNDVI	-0.31	116,558	17,750	84.8%
NDVI	-0.17	89,342	13,989	84.3%
NADVI	-0.09	19,381	2,529	87.0%

CONCLUSIONS

The study area at Eagle Cove was observed to have SAV growing along the shoreline, as shown in Figure 7. The images generated using vegetation indices appear to indicate the presence of SAV along the shore, and in each case provide a more defined and visually identifiable picture of the precise area containing SAV when compared to the RGB photo. The use of multispectral analyses provides confirmation that what is seen as a dark area under the surface is in fact vegetation showing high NIR reflectance. One notable aspect of the study area was a patch of SAV broaching the water surface (Figure 7). This area appears very bright when any index is applied. These images may be used in future study to compare the spectral profile of SAV that is at the water surface

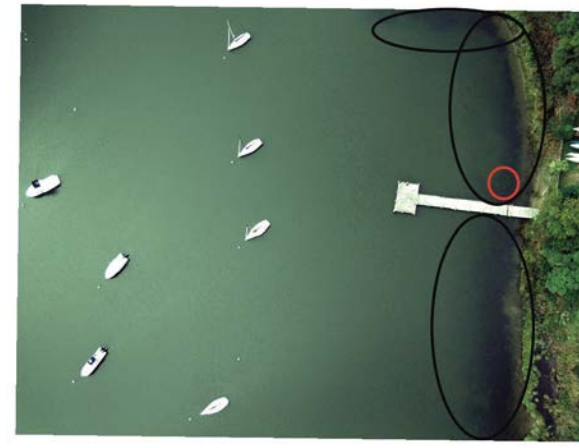


FIGURE 7

Areas of observed SAV growth at Eagle Cove are shown circled in black. A patch of SAV adjacent to the dock was seen to be broaching the water surface, circled in red.

with that of SAV that is below the surface.

A noticeable problem with all four final images is the area of high return in the deeper part of the water which appears as a sparse cloud of highlighted points (Figure 8). This is unlikely to be SAV. The exact source of this error is not known. It could be due to a variety of factors including reflectance problems, artifacts of image manipulation, or organic matter suspended in the water column. Based on knowledge of how SAV grows along the shoreline and how far from the shore it typically grows and at what depth, an educated guess could be made as to what highlighted areas are a result of this "noise" and what areas are SAV. An important challenge of future work will be to reduce the prominence of these errors. Another consideration of future work could be how to separate SAV from algae and other types of phytoplankton that are spectrally similar.

Among the four indices that were tested, mNDVI is shown to be the most effective based on the analysis method outlined in the results section. It identified the highest number of pixels in the SAV Mask within the success range used in the study. GNDVI was the second most effective index but yielded 30% less pixels in the SAV Mask than mNDVI. NDAVI stood out as having an extremely low count of pixels in the SAV mask, with 88% less than mNDVI. When visually comparing mNDVI and NDAVI (Figure

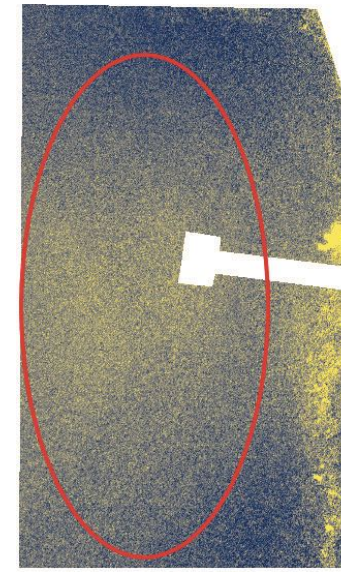


FIGURE 8

An area of higher return (circled in red) that is not SAV is present in all four indexes.

6), it is apparent that NDAVI has a higher prominence of noise that is harder to separate from what is really SAV. These statements of effectiveness apply only to these particular images, and not the viability of these indices in every case. There are many factors that can be adjusted in future trials to maximize the application of these indices, and thus produce more accurate images of SAV presence. These images, however, provide a useful preliminary trial for identifying what went well in this process, learning what needs to be improved on, and for providing a sense of what vegetation index may be the most worthwhile to pursue in future trials.

It is important to note that several environmental conditions determine the ideal time frame for the use of a drone to photograph SAV. The most important factor is the angle of sunlight at the time of flight. The sun being directly overhead allows for the least amount of light scattering by the water and allows light to penetrate deeper into the water column (Rowan and Kalascka). Another factor of importance is the tide. A lower tide will provide less of an obstacle to the camera when detecting submerged grasses. Flights should be conducted during the lowest tide possible, and should be avoided during the highest tides. Additionally, water turbidity, the clarity of the water as affected by suspended particles, in the study area is also a factor. Data collection during surges in turbidity, such as up to 48 hours after heavy rainfall, should be avoided. As it may not be possible to achieve ideal conditions in each of these categories on any particular day, drone operators should seek the best possible balance of all factors when flying for data collection. Recording water quality parameters for each flight may help to understand the distribution of SAV in the area, as well as how images from the drone are affected by water clarity. A noteworthy challenge of photogrammetric analysis over water surfaces is the lack of key points that can be matched between images due to the high uniformity and reflectiveness of water. Pix4D

recommends having at least 30% land area in each image when completing photogrammetric analysis over water, ("Is it Possible to Generate the Orthomosaic of Water Surfaces?").

Our future work will continue to determine a process by which drone-captured multispectral images can be reliably analyzed by ArcGIS Pro to yield an accurate count of SAV pixels and their corresponding acreage and density. An aim of future study should be to make acreage and density calculations and to compare the results to ground-based traditional surveys of the same site.

ACKNOWLEDGEMENTS

Thank you to Dr. Brad Austin for guidance and assistance using ArcGIS Pro, to Dr. Tammy Domanski for obtaining water quality data, and to the Magothy River Association for assistance with access to Eagle Cove and for all the work that they do to protect SAV.

WORKS CITED

Brooks, Colin N., et al. "Multiscale Collection and Analysis of Submerged Aquatic Vegetation Spectral Profiles for Eurasian Watermilfoil Detection." *Journal of Applied Remote Sensing*, vol. 13, no. 3, 27 Aug 2019. SPIE, https://doi.org/10.1117/1.JRS.13.037501.

"Chesapeake Bay SAV Watchers." *Chesapeake Monitoring Cooperative*, 2021, https://www.chesapeakemonitoringcoop.org/chesapeake-bay-sav-watchers/.

Cho, Hyun Jung, et al. "Test of Multi-spectral Vegetation Index for Floating and Canopy-forming Submerged Vegetation." *International Journal of Environmental Research and Public Health*, vol. 5, no. 5, 2008, pp. 477-483. MDPI, https://doi.org/10.3390/ijerph5050477.

"Is it Possible to Generate the Orthomosaic of Water Surfaces?" *Pix4D*, 2021, https://support.pix4d.com/hc/en-us/articles/202558999-Is-it-possible-to-generate-the-orthomosaic-of-water-surfaces.

"Monitoring Methods for SAV." *Virginia Institute of Marine Science*, 2022, https://www.vims.edu/research/units/programs/sav/methods/index.php.

Rowan, Gillian & Margaret Kalacska. "A Review of Remote Sensing of Submerged Aquatic Vegetation for Non-Specialists." *Remote Sensing*, vol. 13, no. 4, 9 Feb 2021. MDPI, https://doi.org/10.3390/rs13040623.


ANNE ARUNDEL
COMMUNITY COLLEGE