IJHSR International Journal of High School Research

July 2020 | Volume 2 | Issue 2
ijhighschoolresearch.org
ISSN (Print version) 2642-1046
ISSN (Online version) 2642-1054

GENIUS OLYMPIAD

"Let's build a better future together"

www.geniusolympiad.org

International Environment Project Fair For Grades 9-12

RIT Rochester Institute of Technology

June 7-12 | Rochester, New York

Application Deadline: March 1, 2021

@GeniusOlympiad

terra science and education

Table of Contents

July 2020 | Volume 2 | Issue 2

Editorial Board

International Journal of High School Research

Bioremediation of Wastewater – Effect of Algae in Bioremediation of Nitrate and Phosphate Content in Wastewater

Hrishika Roychoudhury

Ardrey Kell High School, 10220 Ardrey Kell Rd, Charlotte, NC, 28277, USA
hrishikarc@gmail.com

ABSTRACT: The progressive adoption of an urban, industrial lifestyle has deteriorated the quality of freshwater reservoirs and caused a global challenge for satisfying the demand and maintaining water quality. Industrial, agricultural and domestic processes contaminate water with chemical and biological pollutants that cannot be released into the environment until treated. Traditional wastewater treatment plants are effective in remediating suspended solids and many harmful elements, but they are not always effective in nitrate and phosphate removal causing them to be discharged into the environment. This study aims to determine the effectiveness of algae to bioremediate nitrate and phosphate content from wastewater. Four algae species were tested with wastewater in a bioreactor setup for ten days. The nitrate and phosphate contents were measured every day using a HACH-DR-890 colorimeter to see how effectively the specimen remediated the nitrate and phosphate contents. The analysis shows all four species were able to effectively remediate 93-96% of nitrate and 73-86% of phosphate. *Chlamydomonas reinhardtii* was the most effective in nitrate remediation with a 95.66% remediation and *Scenedesmus quadricauda* was the least effective with 93.97% remediation. For phosphates, *S. quadricauda* (85.22%) was most the effective with 85.22% remediation and *A. platensis* was the least effective with a 73.36% remediation. This study concludes that algae bioremediation is viable in treating nitrates and phosphates in wastewater in a natural, sustainable way compared to conventional treatment processes.

KEYWORDS: Environment pollution; Water treatment; Wastewater; Bioremediation; Bioreactor; Algae; Nitrate; Phosphate.

INTRODUCTION

Maintaining the quality of water is a global challenge. The growing population along with the progressive adoption of an urban, industrial lifestyle has deteriorated the quality of freshwater reservoirs around the world. Remediation of wastewater generated from domestic sewage, industrial, and agricultural discharges has become a large concern for both developed and developing countries. Wastewater contains physical, chemical, and biological pollutants including harmful substances that cannot be released into the environment until the wastewater is treated[1] per the EPA guidelines.

Traditional wastewater treatment plants are effective in the remediation of suspended solids but toxic heavy metals, like arsenic and mercury, and nitrate and phosphate content removal processes are not always very effective and sustainable, causing these toxic elements to be discharged into ground water. Bioremediation is an effective and eco-friendly method in removing those elements. It uses naturally occurring microorganisms to break down hazardous substances into less toxic or nontoxic substances.[2] This study's objective is to determine the effectivity of algae bioremediation on the nitrate and phosphate content in wastewater. Algae absorbs nitrates and phosphates and can clean the water by separating the nutrients as biomass. This algal biomass can also be used later as an energy source and turned into biofuel, which is renewable and has a small carbon footprint.

RESULTS AND DISCUSSION

Based on the analysis, it is observed that all four types of algae were able to reduce the nitrate concentration from wastewater. The initial concentration of nitrate in the wastewater sample ranged from 178.25 - 188.50 mg/L, an average of 183.28 mg/L. The four algae types were able to reduce it to the following: *Chlamydomonas reinhardtii* (average: 7.96 mg/L), *Arthrospira platensis* (average: 9.96 mg/L), *Scenedesmus quadricauda* (average: 11.06 mg/L) and *Chlorella vulgaris* (average: 9.20 mg/L) as shown in Table 1.

Table 1. Nitrate concentration over 10 days (Average of 5 trials).

Days	Control Solution	Sp-1 (C. reinhardtii)	Sp-2 (A. platensis)	Sp-3 (S. quadricauda)	Sp-4 (C. vulgaris)
0	183.28	183.28	183.28	183.28	183.28
1	177.79	143.62	153.90	160.96	166.48
2	170.20	101.87	127.38	130.30	141.55
3	164.33	76.84	88.97	103.83	98.41
4	159.48	47.14	62.13	79.77	72.56
5	155.05	26.83	35.45	53.79	36.99
6	152.53	19.32	21.96	34.74	20.91
7	149.61	13.49	16.94	22.77	16.69
8	146.13	11.19	13.27	16.21	12.82
9	142.90	10.03	11.93	13.51	10.49
10	140.05	7.96	9.96	11.06	9.20

Based on this data, the nitrate remediation efficiency of the algae specimens was calculated. The average percentage nitrate

DOI: 10.36838/v2i2.1

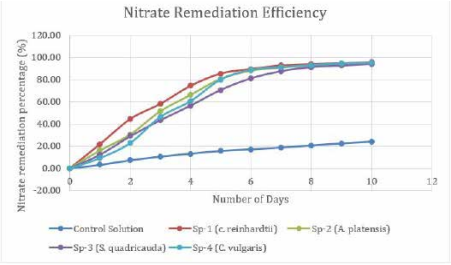


Figure 1. Nitrate remediation efficiency (%) of the algae specimens.

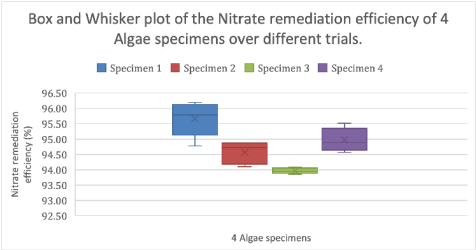


Figure 2. Phosphate remediation efficiency (%) of the algae specimens.

removal achieved by the four specimens were *C. reinhardtii* (95.66%), *A. platensis* (94.57%), *S. quadricauda* (93.97%) and *C. vulgaris* (94.98%) as shown in Fig. 1. Statistical analysis was performed to determine the different remediation efficiencies over the trials. The box and whisker plot shown in Fig. 2 shows the nitrate remediation efficiencies of the four algae species. It can be concluded *C. reinhardtii* is able to achieve the highest nitrate removal with 95.66%. All the specimens were able to remove 93-96% of wastewater nitrates. The EPA's nitrate concentration limit in water is 10 mg/L. All but one algal specimen was able to bring the nitrate concentration to that level over the ten-day study.

Table 2. Phosphate concentration over 10 days (Average of 5 trials)

Days	Control Solution	Sp-1 (c. reinhardtii)	Sp-2 (A. platensis)	Sp-3 (S. quadricauda)	Sp-4 (C. vulgaris)
0	24.77	24.77	24.77	24.77	24.77
1	24.08	20.55	22.32	21.72	20.74
2	23.29	18.44	19.80	19.11	17.31
3	22.39	16.31	16.77	17.28	14.70
4	21.83	12.08	14.71	14.29	12.05
5	21.32	10.55	12.48	11.63	10.73
6	20.23	9.02	10.46	9.86	9.10
7	19.48	7.87	8.46	7.90	7.17
8	18.93	6.86	7.71	6.47	5.94
9	18.23	6.18	7.03	5.21	4.95
10	17.14	5.42	6.60	3.66	4.34

For the phosphate remediation, all four algal species efficiently removed phosphorus from the wastewater. The phosphate concentration remediation by each algae specimen is presented in Table 2. Prior to algae treatment the initial phosphorus content was, on average, 24.77 mg/L. After ten days the phosphorus concentration was reduced to no more than 3.66 mg/L.

Based on this data, the phosphate remediation efficiency by *S. quadricauda* was the greatest during the experimental period. 85.22% of the initial phosphate content was consumed by *S. quadricauda*. It was then followed by *C. vulgaris* (82.46%)

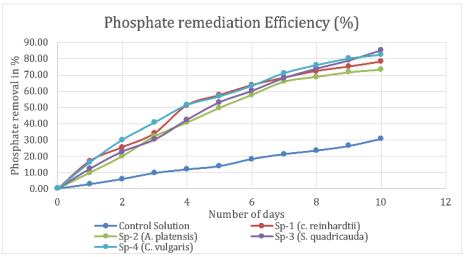


Figure 3. Phosphate remediation efficiency (%) of the algae specimens.

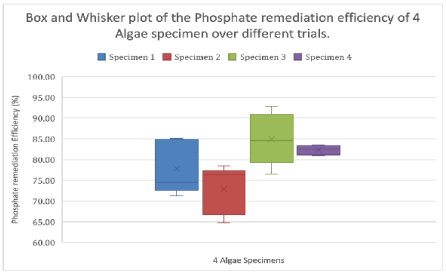


Figure 4. Statistical analysis of Phosphate remediation efficiency (%) data.

C. reinhardtii (78.12%) and *A. platensis* (73.36%) as shown in Fig. 3. The statistical analysis in Fig. 4 illustrates the phosphate remediation efficiencies of the algae species.

It was concluded that by the end of the ten-day period, regardless of the algae species, a range of 73-85% of the initial wastewater phosphate content was removed in each of the reactors.

CONCLUSION

Based on the observations and test results, I conclude that all the algae specimens are able to remediate nitrate and phosphate content from wastewater. *Chlamydomonas reinhardtii* was shown to be most effective for nitrate removal, remediating up to 96% of the concentration and making the water safe for the environment. While all four specimens were effective, *Scenedesmus quadricauda* was shown to be the most effective for phosphate remediation of wastewater, remediating up to 85.22% of the phosphate and making the water able to be discharged into the environment. This experiment concludes that algae bioremediation of wastewater provides an effective and environmentally acceptable option for wastewater treatment.

This study was focused on nitrate and phosphate removal aspect of wastewater bioremediation. There are many other characteristics of wastewater treatment such as BOD and COD remediation and removal of toxic heavy metals, which can be studied further as future goals of this project.

MATERIALS AND METHODS

Collection of Wastewater and Andalgae Specimen: Due to regulations on handling wastewater, this study's wastewater was simulated by mixing fertilizer ammonium nitrate and diammonium phosphate with distilled water and filtering using Whatman No.1 filter paper to remove suspended solids. There are many different species of algae available, and for this study, species that are readily available and easy to grow

DOI: 10.36838/v2i2.2

were collected. These species do not have adverse effects on environment or humans. They are also known to be pollution tolerant and have a high absorption capacity. Such species are considered to be effective in remediation of wastewater and so *Chlamydomonas reinhardtii, Arthrospira platensis (Spirulina), Scenedesmus quadricauda,* and *Chlorella vulgaris* were chosen for this study.

***Experiment Setup*:** Algae need a healthy environment to grow and bioremediate, so a bioreactor system was designed using 500 ml Erlenmeyer flasks connected with an air pump to provide a source of carbon dioxide in the flasks. A fluorescent light was kept on for fast photosynthesis of green algae. The bioreactor was used as an experimental prototype of the Algae-based Wastewater Treatment System (AWTS). The filtered, untreated wastewater was used as the control solution. The algae solutions were made by adding 4 ml of the algae and 200 ml of filtered wastewater in a bioreactor flask. The test was conducted under controlled conditions (temperature 72 ± 2°F) for a total duration of ten days, repeated in five different trials. Using a portable Hach DR-890 colorimeter, measurements of the nitrate and phosphate concentration were taken every day for ten days.[3]

The nitrate concentration of the solution was measured by cadmium reduction method (Hach Method 8039 [4]). In this method, the cadmium metal reduces nitrates present in the sample to nitrite. This nitrite forms an amber-colored product by reacting in an acidic medium with sulfanilic acid. This amber color indicates the presence of nitrate and the intensity is converted to ($N\text{-}NO_3$) mg/L or ppm.

Similarly, the phosphate concentration of the solution was measured by the molybdate-ascorbic acid method (Hach Method 8048 [4]). The orthophosphate present in the sample reacts with molybdate in the acid medium to produce a phospho-molybdate complex. Ascorbic acid reduces this complex and forms a blue color. This blue color indicates the presence of phosphate and the intensity is converted to ($P\text{-}PO_4$) mg/L or ppm.

Calculations of Remediation Efficiency: The nitrate and phosphate remediation efficiency of the algal specimen was calculated by the formula $Ei=\{(Co-Ci)/Co\} \times 100$ where C_0 and Ci were the concentrations of contaminant at the start of experiment (day 0) and day i, respectively.

ACKNOWLEDGEMENTS

I am thankful to my mentor and guide for the project Mr. Matthew G. Welch, a chemistry teacher at Ardrey Kell High School, and Ms. Alisa Wickliff the associate director of the Center for STEM Education at the University of North Carolina for providing support for my project.

REFERENCES

1. Environmental Protection Agency. Wastewater treatment manuals-Primary, Secondary & Tertiary Treatment 1997.
2. Das Surajit (2014). Microbial Biodegradation and Bioremediation.
3. Tredici MR, Margheri MC, Zittelli GC, Biagiolini S, Capolino E and Natali M. 1992. Nitrogen and phosphorus reclamation from municipal wastewater through an artificial food-chain system. Bio resource Technology.
4. HACH DR-890 Colorimeter procedures manual, Hach Company, 1997-2009, 2013.

AUTHORS

Hrishika Roychoudhury is a sophomore from Ardrey Kell High School in Charlotte, NC. She is the founder of non-profit StemmEdHelp that helps students in leaning and spreading awareness on STEM education. She is deeply passionate about science research and won many awards in science competitions. The World Science Foundation selected her as one of 30 World Science Scholars from all over the world in 2019. She is also ISEF 2019 finalist and National TSA Conference 2019 finalist. Miss Roychoudhury is also an Indian classical dancer and is among the top 75 female chess players of her age in USA.

DOI: 10.36838/v2i2.1

Straw Biochar: The Eco-Environment Protector

Chili Wang

Shanghai Youth Science Society, Shanghai 200020, China

ABSTRACT: In this study, a series of experiments was used to improve methods of removing river pollutants using biochar. The experimental results show the amount of CO_2 released from the soil after straw biochar is applied to soil is lower than that of the control and straw directly being applied. This indicates that straw biochar is not easily decomposed into CO_2 by soil microorganisms and thus has the potential for reducing farmland CO_2 emissions. Returning straw biochar to the soil also significantly increases soil water content of about 25-48% compared to the control so it can save agricultural irrigation water. The biochar bricks made of straw biochar and silt have significantly higher hygroscopicity and permeability than silt-only bricks so it can be used for the construction of sponge cities to alleviate urban flooding during rainstorms. Finally, biochar and silt are used to prepare a rigid biochar strip which can be suspended at any position in the water. The biochar strip is effective at removing pollutants. In short, straw biochar has many eco-environmental functions that can aid in environmental conservation of farmland.

KEYWORDS: Biochar brick, biochar strip, soil respiration; soil water content, hygroscopicity and permeability.

INTRODUCTION

China produces 1.4 billion tons of agricultural and forestry straw every year.[1] A large amount of straw is not used effectively and is burned. Straw incineration not only causes haze, but also produces greenhouse gases like CO_2 which seriously damage the environment and endanger people's health.[2] Additionally, untreated straw pollutes the water environment and releases CO_2 as it decomposes.

Previous studies show that straw biochar amendments to soil can significantly improve soil texture and promote crop growth. The Chinese government advocates for straw application to agricultural land.[3] However, farmers are not active in returning straw directly to farmlands because straw is incompatible with the soil. Too many straw surface applications increase the difficulty of crop planting for the next season.[4] The more effective, deeper applications are difficult because they require more manpower and money.

In the past 20 years, the popularity of biochar has increased. Straw biochar is a fine-grained, carbon-containing, porous material obtained by pyrolysis of straw in the absence of oxygen at high temperatures.[5] Its make-up is similar to soil and has very stable properties. Biochar is also rich in ash elements, like potassium, calcium and magnesium, that are nutrients for plants. It has been reported that the application of biochar can increase soil fertility and crop yields by 30-50%.[5] Additionally, due to its porous nature, biochar has a strong adsorption capacity for pollutants and can be usd for water pollution control. Therefore, producing biochar is an effective way to utilize straw.

Straw biochar is a highly stable organic matter. It is porous and very light. When used as a water pollutant treatment, it adsorbs pollutants. However, the biochar floats on the surface and/or is washed away making it difficult to effectively recover the pollutant-containing biochar.

This project aims to answer the following questions through experiments:

(1) Can straw biochar be retained in soil long-term?

(2) Does soil release less CO2 after biochar is applied to the field?

(3) Does straw biochar application increase soil moisture, thus decreasing agricultural irrigation water use?

(4) Do the bricks made from straw biochar and silt have better hygroscopicity and permeability than silt bricks?

(5) How can one enhance the ability of straw biochar to remove pollutants in flowing water?

RESULTS AND DISCUSSION

Effects of Straw Biochar Returning to Field on Soil CO2 Release: Figure 1 shows that the amount of biochar applied to the soil is directly proportional to the content of organic carbon in the soil. Biochar is rich in organic carbon so the soil's

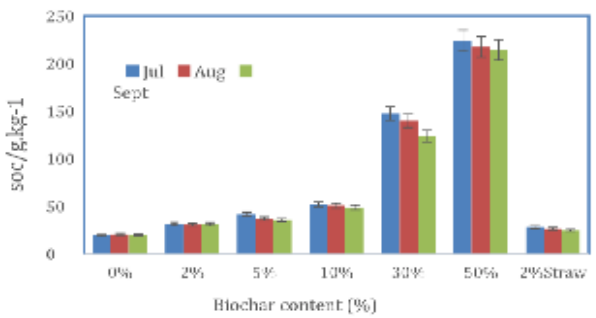


Figure 1. Soil organic carbon content with different proportions of biochar.

DOI: 10.36838/v2i2.2

carbon content in is rapidly supplemented.Soil absorbs oxygen and release CO_2 through soil respiration.[6] Microorganisms in the soil absorb oxygen and decompose the soil's organic compounds into CO_2. Therefore, the higher the soil respiration, the more released CO_2. If organic compounds returned to the soil cannot be rapidly decomposed by the microorganisms, the compounds can remain in the soil for an extended period. Using biochar maintains soil nutrients and decreases CO_2 content, helping combat global warming.

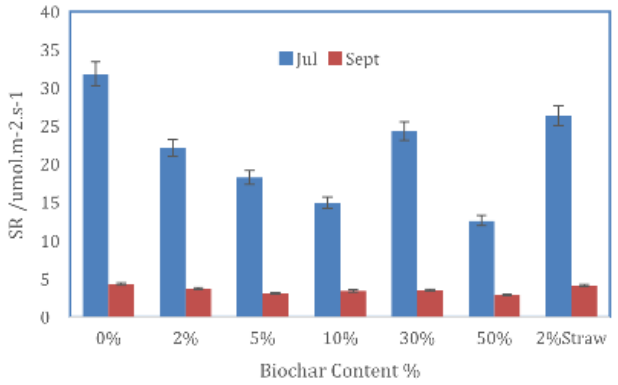


Figure 2. Respiration intensity of the soil with different proportions of biochar.

Figure 2 shows that the soil respiration intensity (the amount of CO_2 released per unit area of soil in a unit time) decreases with the increase of straw biochar application (except 30% biochar). Both straw direct application and the control group had higher respiration intensities than biochar. These results indicate that, compared to the organic carbon in soil and straw, the organic carbon in biochar is harder for soil microorganism to decompose. Therefore, biochar can be retained in soil for a longer time which reduces CO_2 released by farmland (soil).

The above results show that biochar straw release less CO_2 into the atmosphere which can alleviate global warming caused by excessive greenhouse gas emissions. Moreover, the biochar is compatible with soil so application to fields can significantly promote plant growth. Plants fix CO_2 from the atmosphere which further reduces the atmospheric CO_2 concentration

Soil Moisture after Straw Biochar Application to Soil: Soil water content was measured in June, July, August and September 2016. The results show the soil water content with biochar-containing soils was higher than that of the blank control each month as shown in Figure 3. Soil with 30% and 50% biochar was significantly higher than the blank control and the straw application control. The water content of 50% biochar soil was consistently the highest and reached 35.49% in July, a 48% increase compared to the blank control (24.0%). The results show that water absorption and retention were greatly improved after straw biochar application. There was a direct relationship between amount of biochar applied and the water absorption and retention. Therefore, applying straw biochar to soil can decrease agricultural irrigation.

Hygroscopicity and Permeability of Straw Biochar Bricks: The hygroscopicity and permeability of straw biochar brick were preliminarily analyzed. The overall water absorption capacity of bricks containing 5% biochar was nearly 60% higher than that of bricks made of pure silt. Furthermore, the biochar bricks' water absorption in the first minute was twice as high as that of silt bricks, shown in Figure 4 and Table 1.

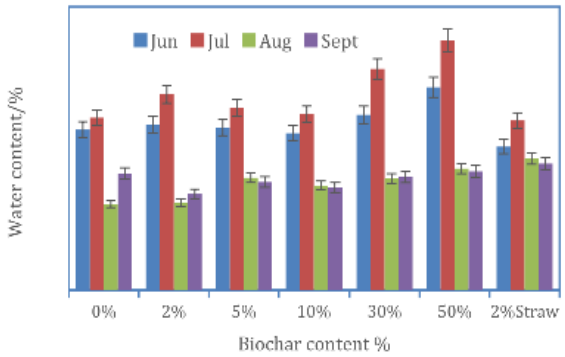


Figure 3. Water content of the soils with different proportions of biochar.

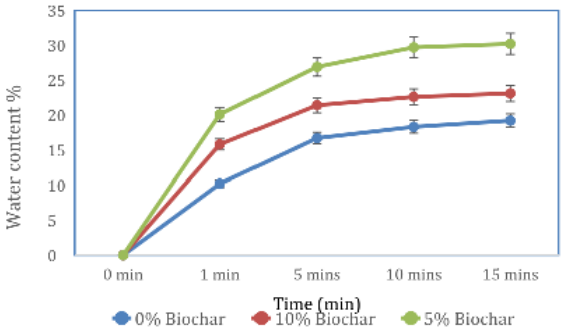


Figure 4. Water absorption capacity of bricks with different proportions of

Table 1. Water absorption and permeability of bricks with different proportions of biochar.

Biochar proportions	0%	5%	10%
Water absorption capacity in 1 min (g/g)	0.101	0.201	0.169
Final water absorption capacity (g/g)	0.189	0.301	0.224

The results show the hygroscopicity and permeability of biochar bricks is better than silt bricks so it can be used in road and sidewalk pavement to alleviate urban flooding caused by rainstorms.

Strangely, the 5% straw biochar bricks have a better hygroscopicity and permeability than the 10% straw biochar bricks. It may be because too much biochar (the mass ratio of 10% is almost equal to the volume ratio of 50%) cannot form a stable, porous structure with clay during hydrothermal reactions. Further research must be done to test this hypothesis.

The results indicate the biochar strips suspended in water can effectively remove eutrophic substances in flowing water. Moreover, unlike straw biochar powder, once the efficiency of the biochar strip fails, the strips can be easily recovered and can be regenerated for reuse.

DOI: 10.36838/v2i2.2

Purification Effect of Biochar Strips on Pollutants in Flowing Water: As shown in Figure 5a, biochar and silt were combined into biochar balls containing 5-10% mass ratio of biochar. These balls were made with a specified strength through a hydrothermal reaction.

To ensure the biochar ball can be suspended in water without floating away, a nylon strip mesh bag was constructed as shown in Figure 5b. The biochar balls are placed inside to make biochar strips that can be suspended in different positions in the water to improve contact efficiency between biochar balls and the waters pollutants.

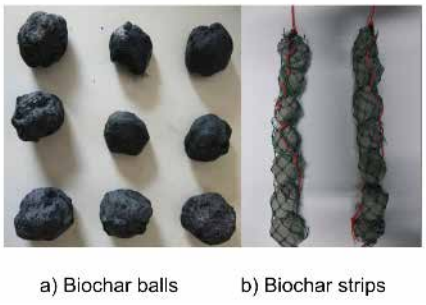


Figure 5. The Pictures of the biochar balls and biochar strips.

Figure 6. Experimental device for verifying purification efficiency of biochar strips.

The biochar strips were placed in simulated flowing eutrophic water system as seen in Figure 6. During the experiment, ammoniacal nitrogen (NH_4^+-N), total nitrogen (TN), and total organic carbon (TOC) concentrations in the water were measured and compared with those in the control water without biochar strips. The results in Figure 7 show that after one day with the biochar strip the NH_4^+-N, TN, and TOC concentrations were reduced by about 60%, 40%, and 15% respectively.

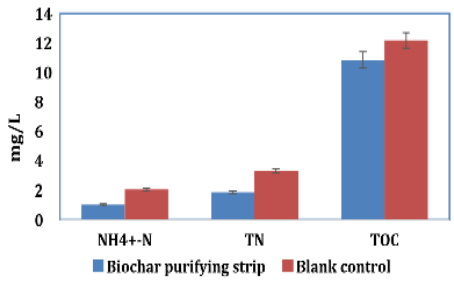


Figure 7. Removal efficiency of biochar strips of NH_4^+-N, TN and TOC in flowing water of fish jar.

CONCLUSION

Through this project, several conclusions were made. Applying straw biochar to soil can reduce CO_2 emissions from farmland and increase soil water content. This helps alleviate global warming and lessens the water required for agricultural irrigation. The water absorption and permeability of straw biochar bricks is better than those of silt bricks while maintaining the same strength so they can be used for construction of sponge cities. Straw biochar balls within nylon strip mesh bags can be suspended in flowing water, enhancing pollutant removal and ease of recovery. The biochar strip can also be regenerated after use.

METHODS

Preparation of Biochar: Wheat straw was collected in early summer to be sun dried and smashed. The project commissioned the Shanghai Jinghua Company to make the straw biochar using a small pyrolysis device at 500°C, shown in Figure 8.

Figure 8. A small pyrolysis device.

Effects of Biochar Application on Soil Respiration and Soil Water Content: In this experiment, five treatments were set up with differing amounts of carbonized straw accounting for 2%, 5%, 10%, 30%, and 50% of the soil weight respectively. The biochar was completely mixed with soil and placed in an uncovered 50cm x 40cm x 40cm plastic box. A group of blank soil controls was set up, recorded as 0%. A control group was set up with 2% straw applied. Triplicates were set for each treatment. The soil was collected from Chongming Island, Shanghai. The appearance of the mixed soil samples is shown in Figure 9.

Figure 9. The Picture of Biochar Returning Experiment.

In June 2016, eight soybean seedlings were planted in each box. During plant growth, the plants were watered equally. I observed the plants, recorded their growth and tested the soil's organic carbon content, soil respiration, and soil water content each month.

Organic carbon was measured by a TOC analyzer and soil respiration was measured by a LICOR-8100A portable soil respirator. Soil water content was determined by wet weight of soil minus weight after drying at 100°C for one hour. Figure 10 shows the soil respiration test instrument and breathing ring

Figure 10. Soil respiration test instrument and test process: a) left is the soil respiration test instrument, b) right is breathing ring

Verification of Hygroscopicity and Permeability of Biochar Bricks: Bricks containing silt plus 5% and 10% straw biochar respectively were made via a hydrothermal reaction at 180 °C, shown in Figure 10. The hygroscopicity of the biochar brick was measured by weighing at different times after soaking bricks in water. Water absorption is directly related to hygroscopicity. The water absorption in one minute represents the brick's water permeability.

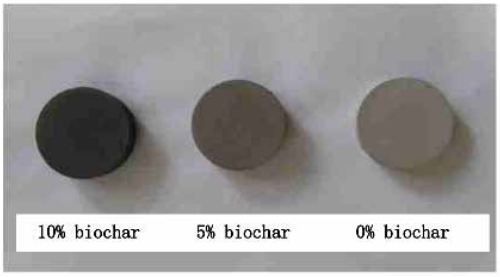


Figure 11. Picture of Biochar Bricks.

Preparation of Suspendible Biochar Strip and Its Purification Effect on Pollutants in Flowing Water: When the formed biochar ball is put into the water it sinks to the bottom and the contact efficiency with pollutants is poor. So, a strip mesh bag was made to hold the biochar ball suspended in the water. The eutrophic water from a fishpond is placed in the fish jar and circulated by a pump to make the water flow. After running for two days, the TN and TOC in the water were measured and compared with the water without biochar strips. The content of ammoniacal nitrogen in water was determined by Nessler reagent spectrophotometry, [7] the total nitrogen content was determined by potassium persulfate digestion, [8] and the total organic carbon content was determined by TOC analyzer.

ACKNOWLEDGEMENTS

I would like to thank the Shanghai Youth Science Society for its support and help on this topic and the Genius Olympiad Organizing Committee for the opportunity to participate in the Genius Olympiad competition and awarding me the bronze prize.in the Genius Olympiad competition and awarding me the bronze prize.

REFERENCES

1. Yuyun, B., Chunyu, G., Yajing, W., et al. (2009). Estimation of straw re sources in China. *Transactions of the CSAE, 25(12), 211-217.*
2. XB, L., YY, Y., Y, F., et al. (2014). Characterization and identification methods of ambient air quality influence by straw burning. *Environmental Monitoring Management and Technology, 26(4), 16-20.*
3. Wang, H., Wang, L., Zhang, Y., Hu, Y., Wu, J., Fu, X., & Le, Y. (2017). The variability and causes of organic carbon retention ability for different agricultural straw types returned to soil. Environmental Technology, 38(5), 538-548. DOI: 10.1080/09593330.2016.1201545.
4. Warnock, D. D., Mummey, D. L., McBride, B., Major, J., Lehmann, J., Rillig, & M. C. (2010) Influences of non-herbaceous biochar on arbuscular mycorrhizal fungal abundances in roots and soils: Results from growth-chamber and field experiments. *Applied Soil Ecology, 46(3), 450-456. DOI: 10.1016/j.apsoil.2010.09.002.*
5. Ahmed, A., Kurian, J., & Raghavan, V. (2016) Biochar influences on agri cultural soils, crop production, and the environment: A review. *Environ mental Reviews, 24(4). DOI: 10.1139/er-2016-0008.*
6. Yan, J., Li, H., Li., J, Xue, Y., Ding, G., & Shao, H. (2013) Response of soil respiration to temperature and soil moisture: Effects of different vegetation types on a small scale in the eastern Loess Plateau of China. *Plant Biosystems, 147(4), 1191-1200.* DOI: 10.1080/11263504.2013.859182.
7. Ministry of Environmental Protection. HJ 535-2009:Determination of Ammonia Nitrogen in Water- Nessler Reagent Spectrophotometry. Bei jing: China Environmental Science Press, 2009.
8. Ministry of Environmental Protection. HJ 636-2012: Determination of total nitrogen in water--alkaline potassium persulfate digestion ultraviolet spectrophotometry. Beijing: China Environmental Science Press, 2012.

AUTHORS

Chili Wang is a grade eleven student at the international division of WeiYu High School in Shanghai, China.

DOI: 10.36838/v2i2.2

Comparative Study of the Antifungal Potential of the Lichen Extracts of Stereocaulon sp. and Cladonia sp.

Roberto O. Vasquez[1], Ricardo E. Medina[1]

Complejo Educacional Chimbarongo, Blanco Encalada 151, Chimbarongo, Chile
obertoorlandov@gmail.com

ABSTRACT: Lichens produce compounds with inhibitory effects on some pathogenic microorganisms. In this survey, we evaluate the inhibitory effects of seven different lichen extracts of the genera *Stereocaulon* and *Cladonia* on the growth of yeast *Saccharomyces cerevisiae* and green mold *Penicillium digitatum*. The extracts that produced the greatest inhibition were those of ethanol, isopropanol, chloroform, ethyl acetate, and acetone while the aqueous and hexane extracts were less effective. Extracts of Stereocaulon sp. had greater inhibitory effect on fungal growth compared to Cladonia sp. Both species of lichens present secondary metabolites with antifungal activity.

KEYWORDS: Antimycotic activity; Growth inhibition; Lichen; Mold; Secondary metabolites.

INTRODUCTION

Microorganisms such as fungi and bacteria develop antimicrobial resistance.[1,2] This creates the need to search new compounds to control the pathogens of animals and plants. Additionally, the use of synthetic pesticides in agriculture is being restricted due to their harmful effects on the environment and human health.[1,2,3,4] New compounds have been created from natural sources such as plants, algae, fungi and lichens; the latter are especially interesting from a biochemical viewpoint due to their particular adaptations to hostile environments.[5,6]

Lichens are fungi living in close symbiosis with a photosynthetic organism, such as microalgae or cyanobacterium. The fungus component of this association is the mycobiont and the photosynthetic organism is the photobiont.[7,8] This symbiosis has allowed lichens to colonize all over the planet, even extreme habitats like Antarctica, deserts, and high mountains.[9,10]

The lichen body is the talus. There three main forms are as follows: crustose lichens form crusts on the substrate, foliose lichens resemble leaves, and fruticose lichens have erect or decumbent growth.[11-13] Most lichens form an internally stratified talus consisting of several ordered layers including the upper cortex, the layer of the photobiont, the medulla, and the lower cortex. The photobiont is protected by a fungal layer. Most of their secondary metabolites, collectively known as lichen compounds, accumulate in the medulla.[13] Lichens, along with bryophytes such as mosses and liverworts, play an important role in ecosystems as they capture nitrogen and carbon and create microhabitats favorable to organisms like amphibians and insects.[14] Lichens are excellent indicators of an ecosystem's health due to their high sensitivity to air pollution and habitat destruction.[11,15,16]

Lichens produce compounds that protect them from adverse physical and biological factors such as herbivorous animals and pathogens.[12,13] Studies have demonstrated these metabolites have antiviral properties and contain antibiotics, enzyme inhibitors, sunscreens, and growth inhibitors of plants and microorganisms.[17-22] Some of these substances correspond to the lichenic acids, the organic acids that allow lichens to degrade rock and make perforations for adherence while contributing to soil formation and the colonization of new areas. Some of these acids, like usnic acid and rhizocarpic acid, also act as photoprotective compounds.[23,24] Usnic acid is present in species of genera such as *Usnea* and *Cladonia* and has recognized antibacterial and antifungal properties.[25,26,27]

To date, 1,383 species of lichen entailing 304 genera have been identified in Chile[28,29] but it is still necessary to increase our taxonomic knowledge of these lichen and to further characterize their biochemical properties to aid researchers obtain novel medicines or pesticides. The central-southern zone of Chile is especially diverse, but its lichen flora is only partially known and even less is known about their potential biochemical properties. In Chile, there are no bibliographic data on the use of lichen extracts in traditional medicine and the existing information in relation to metabolites with biological activity is extremely limited.[30]

The species of two edaphic lichen genera of the order *Lecanorales, Stereocaulon* and *Cladonia* have been studied for the biological properties of their secondary metabolites. The genus *Stereocaulon* has a cosmopolitan distribution and includes approximately 125 species[31] and *Cladonia* includes over 450 species worldwide.[32,33] In Chile, the diversity of both genera is not fully known; only 7 species of *Stereocaulon* and 38 species of *Cladonia* have been documented in Aysén[34] and only two species of *Cladonia* have been reported in La Campana National Park.[35] The lichens of these genera have been used in folk medicine to treat fever, diarrhea, pains and wounds in other parts of the world.[36] Therefore, it is of interest to study the biological properties of these species against pathogenic microorganisms. The antimicrobial effect of the extracts of several species of both *Stereocaulon* and *Cladonia* have been tested

DOI: 10.36838/v2i2.3

on bacteria and fungi.[36,37,38,39,40,41,42,43,44,45,46,47,48] In Chile, the effect of *Cladonia* aff. *rappii* on yeasts has been evaluated.[49]

In this study, the biological activity of different lichen extracts of Stereocaulon sp. (Figure 1) and Cladonia sp. (Figure 2) were tested on the growth of baker's yeast (*Saccharomyces cerevisiae*) and green mold (*Penicillium digitatum*).

Figure 1. Stereocaulon sp. collected in Manquemapu locality.

Figure 2. Cladonia sp. collected in El Sauce locality.

RESULTS AND DISCUSSION

In the graphs, the bars followed by the same letters do not show significant statistical difference ($p = 0.05$).

In the trials testing Stereocaulon sp. against *S. cerevisiae*, as shown in Figure 3, 50 µL ethanol extract (ET 50) is shown to be the best inhibitor since its growth inhibition halo has the largest diameter. ET 50 is followed by chloroform (CL 50), acetone (AC 50), and isopropanol (IS 50). In the experiment with Cladonia sp. extracts, as shown in Figure 4, the best performance is using chloroform (CL 50) and ethyl acetate (AE 50).

For *P. digitatum* treated with Stereocaulon sp., the best performance was obtained with the ethanol extract (ET 50), followed by chloroform (CL 50) and isopropanol (IS 50). The 20 µL ethanol (ET 20) treatment was also very effective as it had a 9.0 mm halo diameter. This shows that *P. digitatum* is especially sensitive to lichen compounds dissolved in ethanol.

In the experiment with Cladonia sp. extracts, the best performance was observed with the isopropanol extract (IS 50), followed by ethanol (ET 50), chloroform (CL 50) and ethyl acetate (AE 50), as shown in Figure 5.

Aqueous extracts (AD 20 and 50) and hexanes (HE 20 and 50) had low performances in all experiments..

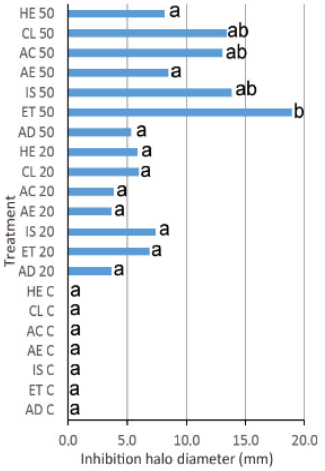


Figure 3. Inhibitory effect of Stereocaulon sp. extracts on S. cerevisiae. Solvents: distilled water (AD), absolute ethyl alcohol (ET), isopropyl alcohol (IS), ethyl acetate (AE), acetone (AC), chloroform (CL), and hexane (HE).

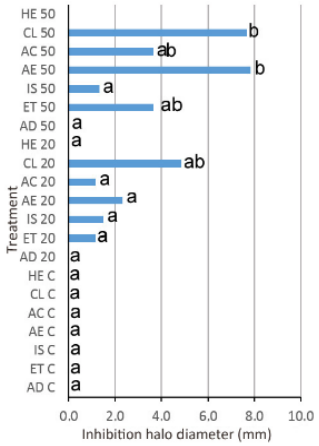


Figure 4. Inhibitory effect of Cladonia sp. extracts on S. cerevisiae.

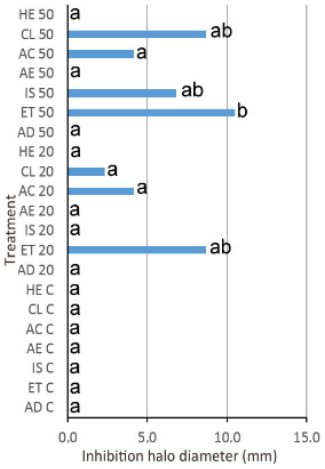


Figure 5. Inhibitory effect of Stereocaulon sp. extracts on P. digitatum.

DOI:10.36838/v2i2.3

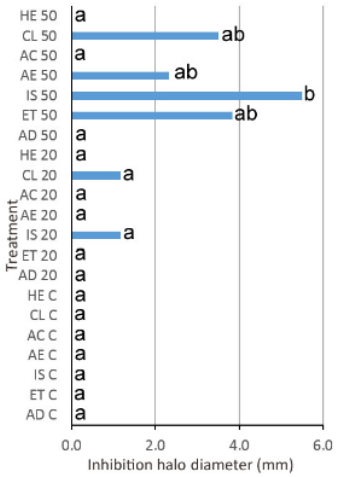


Figure 6. Inhibitory effect of Cladonia sp. extracts on P. digitatum.

CONCLUSION

The lichens of the genera *Stereocaulon* and *Cladonia* used in this project present substances that inhibiting the growth of *S. cerevisiae* and *P. digitatum*.

Extracts of Stereocaulon sp. had a better performance than the Cladonia sp. Extracts because they generated larger-diameter halos under the same cultivation conditions for both fungi.

Greater inhibition was observed in the treatments using a higher extract dose (50 µL) which shows the inhibitory effect is dependent on the extract concentration.

The solvents used have no inhibitory effect on fungal growth as demonstrated in the control treatments.

In all the experiments that utilized solid culture medium, the extracts with the greatest inhibition of fungal growth were those of ethanol, chloroform, and isopropanol while the hexane and aqueous extracts had the lowest performance. The ethyl acetate and acetone extracts had varied effects.

The lichen extracts of Stereocaulon sp. and Cladonia sp. have the potential to formulate phytosanitary products or drugs for controlling pathogenic fungi. For this, it will be necessary to determine the components of each extract.

Further research must be conducted to test these extracts on other fungal and bacteria species of agricultural interest as well as to determine the minimum dose necessary to inhibit microorganism growth and study the biochemical profile of each extract.

METHODS

The research was carried from March to August 2018 in the teaching laboratory at the Complejo Educacional Chimbarongo in Chimbarongo, Chile.

Stereocaulon sp. was collected in the Manquemapu locality of Purranque, Chile and the Cladonia sp. was collected in a sclerophyllous forest located in El Sauce. The lichen material was dried for two weeks then crushed in a food processor.

The solvents used to make the extracts were distilled water, absolute ethyl alcohol, isopropyl alcohol, ethyl acetate, acetone, chloroform, and hexane. To make the extracts, 80 ml of each solvent was mixed with 20 g of lichen material and macerated for a week at 6 °C in hermetic glass jars then sieved and filtered.

The fungi were cultivated on Potato-Dextrose-Agar medium in 90 mm plastic Petri discs. Inoculation was carried out by applying 1 mL of physiological saline solution containing approximately 5,000 colony-forming units (CFU) of *Saccharomyces cerevisiae* and 3,800 CFU of *P. digitatum* onto the agar medium.

Mycological susceptibility tests were performed using the disk diffusion method in solid media using 6 mm diameter filter paper discs. In each Petri dish, 3 discs with 20 µL and 3 discs with 50 µL were placed. The Petri dishes were placed in a digital incubator at 25 °C for 72 hours.

Four experiments were carried out with 28 treatments. Doses of 0 (C = control), 20 and 50 µL of each of the seven types of extract were evaluated. The control treatment consisted of disks with 50 µL of each solvent. The growth inhibition halos were measured with a ruler for the data analysis.

ACKNOWLEDGEMENTS

We thank Professor Alejandro Vargas for his constant support. We also thank Williams Vásquez for his collaboration in the collection of the samples and Diego Urbina for reviewing the translation.

REFERENCES

1. Saga, T., Yamaguchi, K. (2009). History of antimicrobial agents and resis tant bacteria. *Japan Medical Association Journal, 52(2), 103-108.*
2. Chakrabarti, A. (2011). Drug resistance in fungi – an emerging problem. *Regional Health Forum, 15(1), 97-103.*
3. Rechcigl, J.E., Rechcigl, N.A. (2000). Chapter 3: Ecologically based use of insecticides. In Edwards, C.A. (Ed.), *Insect pest management: Techniques for environmental protection, p. 370,* New York, NY: Lewis Publishers.
4. Harris, C.A., Renfrew M.J., Woolridge, M.W. (2001). Assess ing the risk of pesticide residues to consumers: Recent and future devel opments. *Food Additives and Contaminants, 18(12), 1124-1129.* https://doi.org/10.1080/02652030110050122
5. Koul, O., Cupreus, G.W. (2007). Ecologically based integrated pest man agement: *Present concept and new solutions. United Kingdom: CABI.*
6. Ware G. (2000). The pesticide book (5th ed.). Fresno, CA: Thomson Publications.
7. Vandeputte, P., Ferrari, S., Coste, A. (2012). Antifungal resistance and new strategies to control fungal infections. *International Journal of Microbiolo gy, 2012(3), 713687.* https://doi.org/10.1155/2012/713687
8. Yoon, M., Cha, B., Kim, J. (2013). Recent trends in studies on botani cal fungicides in agriculture. *Plant Pathology Journal, 29(1), 1-9.* https://doi.org/10.5423/PPJ.RW.05.2012.0072
9. Lindsay, D.C. (1978) The role of lichens in Antarctic ecosystems. The Bryologist, 81(2), 268-276. https://doi.org/10.2307/3242188
10. Robinson, S., Wasley, J., Tobin, A. (2003). Living on the edge – plants and global change in continental and maritime Antarctica. *Global Change Biology, 9, 1681–1717.* https://doi.org/10.1046/j/1365-2486.2003.00693.x
11. Pereira, I. (2007). Lichenized fungi of Parque Katalapi, X Región, Chile. *Gayana Botánica,* 64(2), 192-200. http://dx.doi.org/10.4067/S0717-66432007000200004
12. Toledo, F., García, A., León, F., Bermejo, J. (2004). Ecología química en hongos y líquenes. *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, 28(109), 509-528.
13. Nash, T. (Ed.). (2008) Lichen biology. Cambridge, England: Cambridge University Press.
14. Ippi, S., González, N. (2004) Arte y ciencia en los bosques en miniatura del Cabo de Hornos. *Revista Feria Cientifica Escolar* (UMAG), 4, 17-18.
15. Gries, C. (2008). Lichens as indicators of air pollution. In Nash, T. (Ed.), Lichen biology (pp. 240-254). Cambridge, England: Cambridge University Press
16. Barreno, E., Pérez-Ortega, S. (2003). Líquenes de la Reserva Natural de Muniellos, Asturias, Asturias, Spain: KRK Edicione.

17. Lauterwein, M., Oethinger, M., Belsner, K., Peters, T., Marre, R. (1995). In vitro activities of the lichen secondary metabolites vulpinic acid, (+) -us nic acid, and (-) -usnic acid against aerobic and anaerobic microorganisms. *Antimicrobial Agents and Chemotherapy, 39(11), 2541-2543.* https://doi.org/10.1128/aac.39.11.2541
18. Amaro, J. (2007). *Evaluación de la actividad antibacteriana y antifúngica de cuatro especies de líquenes de la Región del Maule, Chile* [Unpublished thesis]. Universidad de Talca, Chile
19. B alaji, P., Harijaran, G. (2007). In vitro antimicrobial activity of *Parmo trema praesorediosum* thallus extract. *Research Journal of Botany*, 2(1), 54-59. https://doi.org/10.3923/rjb.2007.54.59
20. Román, C. (2007). *Uso potencial de extractos de líquenes nativos como t ratamiento alternativo en la preservación de madera de Pinus radiata.* [Un published thesis]. Universidad de Talca, Chile
21. Rankovic, B., Misic, M., Sukdolak, S. (2008). The antimicrobial ac tivity of substances derived from the lichens *Physcia aipolia, Umbilicaria polyphylla, Parmelia caperata* and *Hypogymnia physodes*. World Journal of Microbiology and Biotechnology, 24, 1239–1242. https://doi.org/10.1007/s11274-007-9580-7
22. Akpinar, A., Ozturk, S., Sinirtas, M. (2009) Effects of some terricolous lichens [*Cladonia rangiformis Hoffm*., *Peltigera neckerii Hepp* ex Müll. Arg., *Peltigera rufescens* (Weiss) Humb.] on soil bacteria in natural condi tions. *Plant Soil Environmental,* 55(4), 154-158. https://doi.org/10.17221/1616-PSE
23. Rubio, C., Fernández, E., Hidalgo, M., Quilhot, W. (2002). Effects of solar UV-B radiation in the accumulation of rhizocarpic acid in a lichen species from alpine zones of Chile. *Boletín de la Sociedad Chilena de Química,* 47(1), 91-100
24. O tt, S. (2004). Early stages of development in *Usnea antarctica* Du Rietz in the South Shetland Islands, northern maritime Antarctica. The Lichenolo gist, 36(6), 413-423. https://doi.org/10.1017/S0024282904014380
25. Bustinza, F. (1951). Contribución al estudio de las propiedades antibacter anas y antifungicas del ácido úsnico y de algunos de sus derivados. *Anales del Jardín Botánico de Madrid*, 10(1), 151-175
26. Morales, P., Robles, C., Abram, A., Gallagher, K., Valdez, C. (1993). Estu dio fitoquímico de Usnea sp. *Revista de Química*, 7(1), 13-19
27. Vaillant, D. I. (2014). Los líquenes, una alternativa para el control de fito patógenos. *Fitosanidad*, 18(1), 51- 57.
28. Quilhot, W., Pereira, I., Guzmán, G., Rodríguez, R., Serey, I., Barrera, E. (1998) Categorías de conservación de líquenes nativos de Chile. *Boletín del Museo Nacional de Historia Natural, Chile, 47, 9-22.*
29. Guzmán, G. (2008). Líquenes,. In Comisión Nacional del Medio Am biente (Eds.), Biodiversidad de Chile: Patrimonio y desafíos (2nd ed.) (pp 376-377). Santiago, Chile: Ocho Libros Editores.
30. Piovano, M., Garbarino, J. A., Giannini, F. A., Correche, E. R., Feresin, G., Tapia, A., Zacchino, S., Enriz, R. D. (2002). Evaluation of antifungal and antibacterial activities of aromatic metabolites from lichens. Boletín de la Sociedad Chilena de Química, 47(3), 235-240. http://dx.doi.org/10.4067/S0366-16442002000300006
31. Park, J. S., Park, S. H., Park, S. Y., Jayalald, U., Hura, J. S. (2018). Revision of the lichen genus *Stereocaulon* (Stereocaulaceae, Ascomycota) in South Korea. *Mycobiology*, 46(2), 101-113. https://doi.org/10.1080/12298093.2018.1461314
32. Stenroos, S., Hyvonen, J., Myllys, L., Thell, A., Ahti, T. (2002). Phylogeny of the genus Cladonia s.lat. (Cladoniaceae, Ascomycetes) inferred from mo lecular, morphological, and chemical data. Cladistics, 18(3), 237–278. https://doi.org/10.1006/clad.2002.0202
33. Pérez-Vargas, I., González-Montelongo, C., Hernández-Padrón, C., Pérez de Paz, P. L. (2015). Contribution to the knowledge of the genus *Cladoni in Macaronesia. Botanica Complutense*, 39, 31-35. https://doi.org/10.5209/rev_BOCM.2015.v39.49131
34. Quilhot, W., Cuellar, M., Díaz, R., Riquelme, F., Rubio, C. (2012) Líquenes de Aisén, sur de Chile. *Gayana Botánica*, 69(1), 57-87. http://dx.doi.org/10.4067/S0717-66432-12000100007
35. Redón, J., Walkowiak, A. (1978). Estudio preliminar de la flora liquén ca del Parque Nacional "La Campana": I. Resultados sistemáticos. *Anales del Museo de Historia Natural de Valparaíso*, 11, 19-36
36. Açikgöz, B., Karalti, İ., Ersöz, M., Coşkun, Z. M., Cobanoğlu, G., Ses al, C. (2013). Screening of antimicrobial activity and cytotoxic effects of two Cladonia Species. *Zeitschrift für Naturforschung C*, 68(5-6), 191-197.
37. González, A. G., Rodríguez, E. M., Hernández, C. E., Bermejo, J. (1992). Chemical constituents of the lichen *Stereocaulon azoreum*. *Zeitschriftfür Naturforschung C*, 47(7-8), 503-507. https://doi.org/10.1515/znc-1992-7-802
38. Yilmaz, M., Türk, A. O., Tay T., Kivanç, M. (2004). The antimicrobial activity of extracts of the lichen *Cladonia foliacea* and its (-) -usnic acid, atranorin, and fumarprotocetraric acid constituents. *Zeitschrift für Naturforschung C*, 59(3-4), 249-54. https://doi.org/10.1515/znc-2004-3-423
39. tark, S., Kytöviita, M. M., Neumann, A. B. (2007). The phenolic com pounds in *Cladonia* Lichens are not antimicrobial in soils. *Oecologia*, 152(2), 299-306. https://doi.org/10.1007/s00442-006-0644-4
40. Neeraj, V., Behera, B. C., Parizadeh, H., Sharma, B. O. (2011). Bactericidal activity of some lichen secondary compounds of *Cladonia ochrochlora*, *Parmotrema nilgherrensis* & *Parmotrema sancti-angelii*. International Journal of Drug Development & Research. 3(3), 222-232
41. Dieu, A., Millot, M., Champavier, Y., Chulia, J. A., Vergnaud, J., Bressollier, P., Sol, V., Gloaguen, V. Antimicrobial activity of Cladonia incrassata acetone extract. Planta Medica, 78(11), 931-935. https://10.1055/s-0032-1321148
42. Dzomba, P., Togarepi, E., Musekiwa, C. (2012). Phytochemicals, antioxidant and antibacterial properties of a lichen species *Cladonia digitata*. *African Journal of Biotechnology,* 11(31), 7995-7999. https://doi.org/10.5897/AJB12.4130
43. Bhattarai, H. D., Kim, T., Oh, H., Yim, J. H. (2013). A new pseudodep sidone from the antarctic lichen *Stereocaulon alpinum* and its antioxidant, antibacterial activity. *The Journal of Antibiotics,* 66(9), 559-561. https://doi.org/10.1038/ja.2013.41
44. Kosanić, M., Ranković, B., Stanojković, T., Rančićc, A., Manojlović, N. (2014). *Cladonia lichens* and their major metabolites as possible natural antioxidant, antimicrobial and anticancer agents. *LWT – Food Science and Technology*, 59(1), 518-525. https://doi/org/10.1016/j.lwt.2014.04.047
45. Ramos, D. de B. M., Gomes, F. S., Napoleão, T. H., Paiva, P. M. G., Correia da Silva M. D., Coelho, L. C. B. B. (2014). Antimicrobial activity of *Cladonia verticillaris* lichen preparations on bacteria and fungi of medical importance. Chinese Journal of Biology, 2014(219392), 1-7. https://doi.org/10.1155/2014/219392
46. Rankovic, B., Kosanic, M., Stanojkovic, T. (2014). *Stereocaulon paschale* as antioxidant, antimicrobial and anticancer agent. *Farmacia*, 62(2), 306-317.
47. Studzińska-Sroka, E., Hołderna-Kędzia, E., Galanty, A., Bylka, W., Kacprzak, K.; Ćwiklińska, K. (2015). In vitro antimicrobial activity of extracts and compounds isolated from *Cladonia uncialis*. *Natural Product Research,* 29(24), 2302-7. https://doi.org/10.1080/14786419.2015.1005616
48. Moura, J. B., Vargas, A. C., Gouveia, G. V., Gouveia, J. J. S., Ramos-Júnior, J. C., Botton, S. de A., Pereira, E. C., Costa, M. M. (2014). In vitro antimicrobial activity of the organic extract of Cladonia substellata Vainio and usnic acid against Staphylococcus spp. obtained from cats and dogs. *Pesquisa Veterinária Brasileira*, 37(4), 368-378 https://doi.org/10.1590/s0100-736x2017000400011
49. Plaza, C. M., Pérez de Salazar, C., Vizcaya, M. Rodríguez-Castilo, C. G., Medina, G. E., Plaza, R. E. (2017) Potential antifungal activity of *Cladonia aff. rappii* A. Evans. *Journal of Pharmacy & Pharmacognosy Research*, 5(5), 301-309.

AUTHORS

Roberto Vásquez is a 17-year-old student in 4th grade "D" at the Complejo Educacional Chimbarongo. He has participated in several national and international science fairs including Milset Expo-Sciences International América Latina (ESI AMLAT) in 2018, Genius Olympiad in 2019, and Milset Expo-Sciences International (ESI) in 2019.

Ricardo Medina is a 17-year-old student in 4th grade "D" at the Complejo Educacional Chimbarongo. He has participated in several national and international science fairs including Genius Olympiad in 2019 and Milset Expo-Sciences International (ESI) in 2019.

DOI: 10.36838/v2i2.3

Using the Santa Fe Ant Trail Benchmarking Problem to Optimize Genetic Algorithms

Nimisha Gupta

Yorktown High School, 2727 Crompond Road, Yorktown Heights, NY 10598, United States
Nimisha.gupta@yorktown.org

ABSTRACT: Genetic algorithms have recently gained popularity for solving complex problems through their evolutionary nature. They implement biological mechanisms such as crossovers and mutations, yet it is not known how each variable independently effects the algorithm's ability to work. If found, it would improve the optimization time in genetic algorithms by decreasing the amount of trials the program has to run. To test the benchmarking system, three trials were conducted in which the probability of mutating and mating were changed independently to understand their effects. As the base case, the probability of mating and mutating were set to default values from previous testing to efficiently complete the benchmarking device used (Santa Fe Ant Trail). In trial 2, the probability of mating was independently tested while trial 3 tested the probability of mutating. The increase in crossover rates had a superior efficiency in completing the task given while the increase in mutation rates decreased overall efficiency of the program. In the future, these results can be used in genetic algorithms to increase their efficiency in completing complex problem sets.

KEYWORDS: Evolutionary Computation; Optimization; Genetic Algorithm; Benchmarking; Fitness.

INTRODUCTION

As society evolves, certain complex problems arise that are virtually impossible for a human to solve. One example is creating a satellite antenna for broadcasting messages and data transmission. These antennas must be asymmetrical for maximum efficiency; it is both time and material consuming for a human to create because of the amount of trial and error required.[1] The field of evolutionary computation was created at the intersection of biology and computer science to solve such complex problems. Genetic algorithms are used in scenarios for optimizing behavior of artificial agents in achieving some goal. The algorithms use evolutionary tactics, such as mating and mutating, to "evolve" its programs/solutions until a solution is found. As the popularity of these algorithms grows the efficiency of genetic algorithms must be tested. Problems like the Santa Fe Ant Trail have been developed to decipher how factors liked mating and mutating effect the efficiency of the algorithm.[1] This knowledge will lead to the optimization of genetic algorithms.

Evolutionary Computation: Evolutionary computation techniques can produce highly precise and quick solutions to a wide range of problem settings.[1] It has applications in computer science and can be implemented in algorithms inspired by evolutionary biology because of its similarity to real-world problems.

Evolutionary computation encompasses genetic programs that implement mechanisms inspired by biological evolution such as reproduction and mutation. Each genetic program starts by generating a large amount of possible solutions, or evaluations. Through different strategies the less efficient solutions are removed. As a result, the solution population will gradually "evolve" to increase fitness, or efficiency.[2] This repeats until the most effective solution is outputted, shown in Figure 1. The greater the fitness, the greater the optimization of the solution. These algorithms have characteristics which eliminate the need for detail. They take random parameters from various subsets and combine them to create new solutions. For example, when simulating a race car, the focus is on aerodynamics and speed. Taking into account factors such as height and weight, the simulator combines several values of each respective value and tests them to provide the best outcome. Thus, there is no need for exact measurements. Despite this, there has been no research into how a change in mating/reproduction rate and mutation rate can affect an algorithm's efficiency.

Certain benchmarking problems are used to test the efficiency of a genetic algorithm. One benchmarking system, the Even-Parity problem, focuses on finding how many prime numbers less than a given integer have an even number of

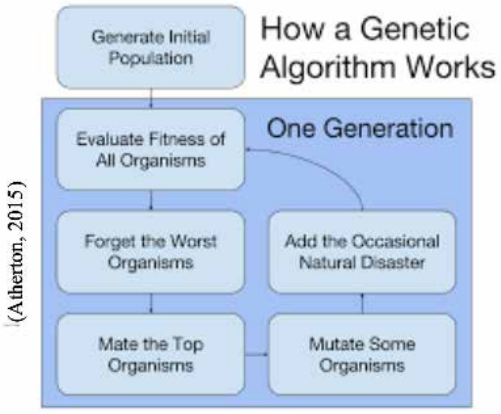


Figure 1. The genetic algorithm process to optimize solutions for difficult problem sets.

DOI: 10.36838/v2i2.4

prime factors.[3] Another benchmarking problem is called the Artificial Ant problem. In this, a program is supposed to follow a food path and over time holes start appearing on the trail and growing in size. Due to its correlation with ants and food trails the programs are called ants.

Santa Fe Ant Trail: A version of the Artificial Ant problem is the Santa Fe Ant Trail. It comes with a layout of 32x32 spatial elements and incorporates a predefined food trail the ant is supposed to follow. Figure 2 shows the food trail consisting of 144 cells with 89 cells containing food. It has the following irregularities which makes it difficult for the ant to follow the trail easily: single gaps, double gaps, single gaps at corners, double gaps at corners, and triple gaps at corners.[4] The ant starts at the top left corner facing east and continues down the trail. The ant can turn left, right, or move forward one step and sense whether there is food ahead of it. With each step the ant loses energy, regaining energy by ingesting food. The goal is for the ant to eat all of the food pellets. Therefore, the more food pellets the ant eats, the more efficient the genetic algorithm is.

Through various tests, researchers have called this problem highly deceptive. The Santa Fe Ant Trail is difficult to solve efficiently because of the large, hole-filled landscape and the

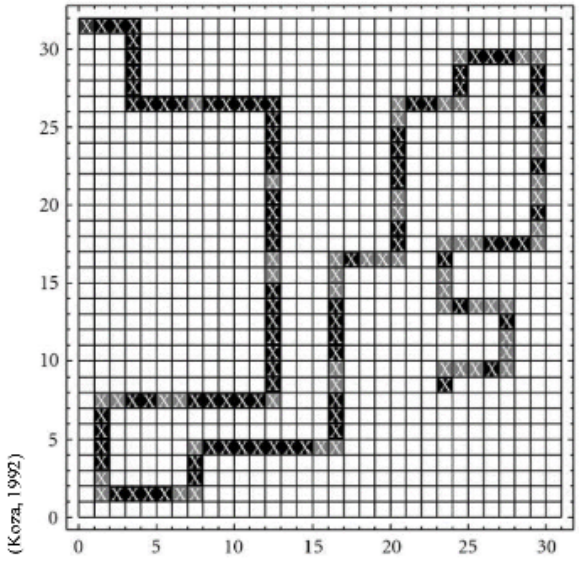


Figure 2. The Santa Fe Ant Trail. The dark squares contain food and the light squares are holes in the trail.

deceptive nature of the search space limited by a fixed amount of energy.[5,6] Additionally, the fitness space associated with the Santa Fe Trail has a great deal of randomness associated with it, creating difficulties.[7]

Problems: It is unknown how mating, also known as crossover, and/or mutating directly effects a genetic algorithm's output when benchmarked. By calculating this, one can increase the optimization of the efficiency of genetic algorithms.

The goals of this project are

1. To test the effect of mating on a previously created genetic algorithm and understand how it affects evaluation population, or the amount of possible solutions, and the maximum amount of food an "ant" could eat.

2. To test the effect of mutation on a genetic algorithm and understand how it affects the evaluation population and the maximum amount of food an "ant" could eat.

Hypotheses: It is hypothesized that

1. Increasing crossovers in the genetic algorithm will cause an increase in the amount of evaluations and the maximum amount of food eaten. There will be more evolution, causing "adaptable traits" or those that allow the "ant" to eat more food to become more common throughout

2. Increased mutations in the genetic algorithm will cause a decrease in evaluations and increase in food eaten. Positive mutations will lead to adaptive traits being developed quicker.

RESULTS

Three trials using the Santa Fe Ant Trail were run to test the effect of the mating and mutation rates in genetic algorithms and find an optimal solution to the trail problem.

During the first trial, the algorithm was set so the probability of mating and mutating between generations was 0.5 and 0.2, respectively. This trial was used as the control group. At the program's start, the number of evaluations (population) immediately decreased to 152. As shown in Figure 3, the number of evaluations remained around 150 for the remainder of the program. Throughout the 40 generations tested, the average amount of food eaten showed a positive correlation with the amount of generations, shown in Table 1. Therefore, the efficiency of the ant's capability to eat pellets on the trail increased overall. By generation 23, a configuration of the original algorithm had completed the trail (eating all 89 pellets) and provided a solution.

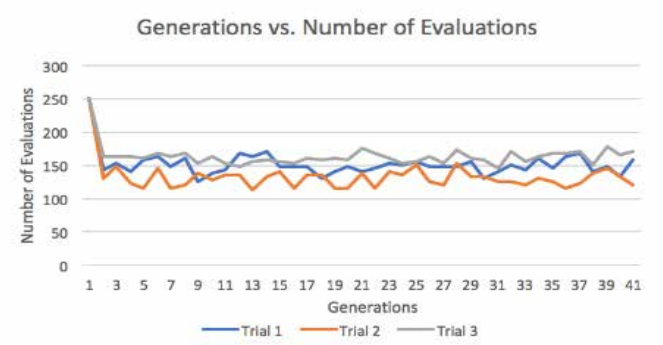


Figure 3. The number of evaluations for each generation in trials 1-3.

Table 1. The average and standard deviation of food pellets eaten for generations for trial 1.

Gens.	Average (Food Pellets)	St. Deviation (Food Pellets)
0	1.62	2.91266
5	14.7	11.1826
10	45.98	33.7082
15	61.32	37.2605
20	62.88	37.5869
25	64.608	37.1065
30	69.068	33.6293
35	71.24	31.4602
40	71.684	31.6462

DOI: 10.36838/v2i2.4

In trial 2, the probability of mating, or crossover, was set to 0.4 and the probability of mutating remained at 0.2. This was done to understand the independent effect of mating. Once again, the number of evaluations immediately dropped, this time to 129, shown in Figure 3. Continuing from generation 2, the system's variance stayed around the 130-evaluation mark. The average amount of food showed a positive correlation with the amount of generations, representing that the ant's efficiency in terms of eating food was increasing as the program evolved, shown in Table 2. However, the maximum amount of food eaten by generation 40 was only 60, demonstrating the set variables did not successfully create a solution.

Table 2. The average and standard deviation of food pellets eaten for generations for trial 2.

Gens.	Average (Food Pellets)	St. Deviation (Food Pellets)
0	1.62	2.91266
5	19.224	12.4889
10	35.872	20.3114
15	44.58	22.8081
20	44.348	21.8409
25	51.204	18.9311
30	51.86	18.6269
35	54.376	15.7072
40	54.508	15.0336

Table 3. The average and standard deviation of food pellets eaten for generations for trial 3.

Gens.	Average (Food Pellets)	St. Deviation (Food Pellets)
0	1.62	2.91266
5	16.236	12.6426
10	28.04	18.8895
15	35.436	25.6487
20	30.44	25.9983
25	36.844	25.8617
30	43.98	24.7406
35	40.524	25.673
40	49.172	22.2007

In trial 3, the crossover probability returned to the control probability of 0.5 and the = mutation probability was set to 0.3. The number of evaluations dropped from 250 to 163 in between generation 1 and 2. From there, the average amount of evaluations was 161.2. As the generation number grew, there was an increase in the average amount of food eaten, shown in Table 3. However, the maximum amount of food eaten by generation 40 was 62, demonstrating the problem setting did not create a solution. generation 40 was 62, demonstrating the problem setting did not create a solution.

DISCUSSION

Effect of Mating: Three trials were run to understand the effect of crossover and mutation in genetic algorithms. Through the results, the independent effects of mating and mutating are clear. For trials 1 and 2, they both began at 250 evaluations but drastically decreased in generation 2. However, trial 2, which tested crossover/mating, experienced a more drastic decrease in number of evaluations. This could be due to the fact that more programs were combining and evolving, which created an increase in the population counter. This outcome supports the results from previous research.[4] Along with the population count, the maximum amount of food eaten decreased between trial 1 and 2 as there was a 39-pellet-decrease from trial 1. This could also be caused by the increase in mating. Because more programs were crossing over, they evolved faster. Data from trials 1 and 2 support the first hypothesis that an increase in the probability of crossover will increase the amount of evaluations and the maximum amount of food pellets eaten.

Effect of Mutating: Comparing trials 1 and 3, the probability of mutating was changed from 0.2 to 0.3. Both trials began at 250 evaluations but experienced a significant drop. The decrease in trial 1 was greater than trial 3, with populations of 150 and 160 respectively. This drop could be due to the greater mutation rate slightly increasing the chances of crossover occurring. Because more programs were being mutated, more mating occurred and took on the new programs' features. Continuing, with the increase in mutating probability there was a decline in the maximum amount of food eaten. A possible cause for this is there were more negative mutations than positive mutations, or more traits that caused the ant to fall for the holes in the trail were being generated. Therefore, the mutations set the programs back on their goal to venture the trail. This falsifies the second hypothesis because as the mutation rate increased, the number of evaluations increased and the maximum amount of food eaten decreased. This shows that although it increased the algorithm's efficiency, higher mutations caused fewer possible solutions to be created.

Trends Found in Evaluations and Average Amount of Food Eaten: By the 5th generation, most data points regarding the number of evaluations and the maximum amount of food eaten had linearized. In trial 1, it linearized around 150, trial 2 linearized around 130, and trial 3 linearized around 160. This indicates that, over time, the effectiveness of evaluation diminishes. Similarly, the average food eaten constantly grew throughout the generations and trials, slowing down around generation 20. The standard deviation grew until about generation 15 and then began to decline slowly. This could be because most of the programs reached their peak by generation 20 and did not reproduce to create more efficient program.

Application: The results presented above can be used to optimize genetic algorithms by decreasing the amount of trial-and-error necessary. By using semi-random techniques to generate a population of programs and narrow its search to the solution, the efficiency of the algorithms is increased and allows the users to get their solutions faster.

Genetic algorithms are mainly applied for designing parts with specific tasks. For example, genetic algorithms are used to create race car parts to increase speed and aerodynamics. Another application is gene expression profiling. This method measures the activity of thousands of genes at once to determine a pattern the genes express. Genetic algorithms make this analysis more efficient.

If the results from this experiment are implemented, it may further increase the speed of such programs, making it possible for researchers to focus on individual patients' unique gene expression profiles.

Future Research: The code from trial 1 that was able to obtain all 89 food pellets in the trail at the end of generation 40,

can be compared to Koza's solution.[9] This may demonstrate this experiment's algorithm is more effective than the previous solution. Through comparison, a more efficient algorithm can be found and/or created. This could then be applied to genetic benchmark problems outside of the Santa Fe Ant Trail. Along with comparing the generated program with the previous solution, the code can be applied to an Arduino bot to understand if algorithms can be successfully transitioned from a virtual world to the real world. This could help decrease the amount of time it takes to test certain programs because testing in the real world is less efficient than in the virtual world.

CONCLUSION

This project aimed to find the effect of crossovers and mutations on a genetic algorithm using the Santa Fe Ant Trail. Crossover refers to when two programs in the population combine their traits or features to create a new program. Mutation refers to when a random feature is generated and introduced to the population of solutions. When the probability of mating increased the amount of the evaluation population increased and the maximum amount of food pellets increased. When the probability of mutating decreased the number of evaluations increased but the maximum amount of food eaten decreased. This data can optimize the process of creating efficient genetic algorithms and change the way engineers implement algorithms into the real world.

METHODS

Mentor Role: Throughout the research period of 18 weeks encompassing the summer and fall of 2018, my mentor and I collaborated on various aspects of the experiment. My mentor provided advice and background on genetic algorithms in evolutionary computation, their implementation, and how to benchmark them. My mentor also provided a pseudo-code to give a visual representation of any inputted genetic algorithm. Following the pseudo-code and my mentor's guidance, I created a genetic algorithm benchmarking test that mirrored the Santa Fe Ant Trail and implemented it using several different settings.

Virtual World: A virtual world, or computer-based stimulated environment, was created and implemented to display graphics of the implemented genetic algorithm. It was designed to display the Santa Fe Ant Trail and the program carrying out its path, shown in Figure 4. The programming language Python was used to create the virtual world because of its incorporated wide range of modules that allow for code to be created in a short and efficient manner.[11] The compiler/tester Eclipse was used to implement Python. The cloud/interpreter Anaconda, which comes preinstalled with popular Python packages and environments, was used to import packages, or groups of modules, for the program including NumPy and DEAP (Figure 5). NumPy allows the user to use high-level mathematical functions and DEAP allows the user to use genetic algorithm functions.

Santa Fe Ant Trai: The Santa Fe Ant Trail was implemented to provide a benchmark for the genetic algorithm. One necessary component to the Trail problem is the ant. The ant was designed to implement a genetic algorithm and eat the food. the algorithm would be tested based on the efficiency of the algorithm's ability to maneuver through the complex trail.

Algorithms: The algorithms tested in the Santa Fe Ant Trail included a random case to provide a base, a previously tested solution to provide a benchmark, and genetic programs to understand different methods to approach the problem. The solution algorithm was taken from Koza.[9] As shown in Figure 6, the old solution uses a combination of "if" statements and programs.

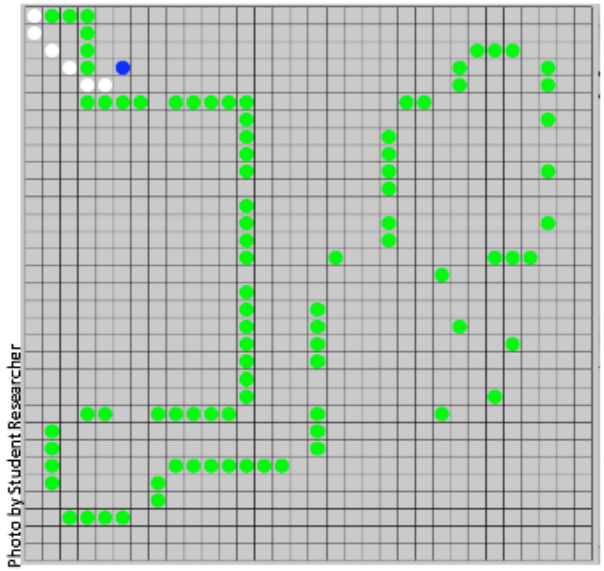


Figure 4. An example of how the virtual world appears when the program is run. The blue dot is the ant, the green dots are the food, and the white dots are the ant's trail

Figure 5. Examples of packages in the Anaconda cloud. The package name is on the left and a brief summary of its functions is on the right.

The code was created to implement new genetic algorithms in every trial. There were four inputs for the code including

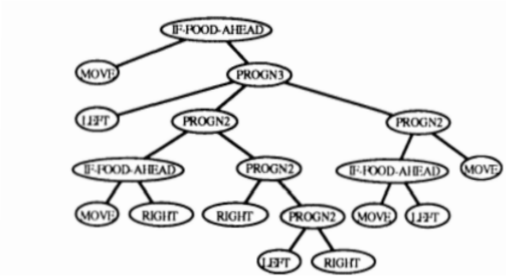


Figure 6. The pseudo code for Koza's 1992 solution reading from top to bottom and left to right.[9] "PROGN" means to carry out all of its following branches

DOI:10.36838/v2i2.4

the initial population, the probability of mating, the probability of mutating, and the amount of generations. Each trial maintained the initial population (250) and the number of generations (40). At the end of each trial, the most efficient set of directions is outputted and recorded. Along with this, the amount of evolutions, average amount of food eaten, and maximum amount of food eaten is outputted for each generation. Once the effects of each variable are calculated, the variables will be manipulated to output the best solution to the Santa Fe Ant Trail.

■ ACKNOWLEDGEMENTS

I would like to thank my parents, Ajeet and Sanjita Gupta, and my sister, Amani Gupta. I would also like to thank my teachers, Michael Blueglass and Rachel Koenigstein. Lastly, I would like to thank my mentor, Theodore van Kessel, for aiding me throughout the project.

■ REFERENCES

1. Dyer, D. W. (2010). *Evolutionary computation in Java*: A practical guide to the Watch maker Framework. Uncommons.org. https://watchmaker.uncommons.org/manual/

2. Mallawaarachchi, V. (2017). *Introduction to genetic algorithms – including example code.* Medium. https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3

3. *Even-parity problem.* (n.d.). DEAP Project. https://deap.readthedocs.io/en/master/examples/gp_parity.html

4. Wilson, D., & Kaur, D. (2014). How Santa Fe ants evolve. *Neural and Evolutionary Computing.* https://arxiv.org/pdf/1312.1858v2.pdf

5. Langdon, W., & Poli, R. (1998). *Why ants are hard* (CSRP-98-04). Birmingham, England: University of Birmingham. http://www0.cs.ucl.ac.uk/staff/W.Langdon/antspace_csrp-98-04/

6. Langdon, W., & Poli, R. (1998). *Better trained ants for genetic programming* (CSRP-98-12). Birmingham, England: University of Birmingham.

7. Miller, J. F., & Thomson, P. (2000). Cartesian genetic programming. In Poli, R.,Banzhaf, W., Langdon, W. B., Miller, J., Nordin, P., & Fogarty, T. C. (Eds.), *Genetic programming: European Conference, EuroGP 2000, Edinburgh, Scotland, UK, April 2000 proceedings* (pp. 121-132). Lecture notes in computer science (Vol. 1802).Berlin, Germany: Springer. https://doi.org/10.1007/978-3-540-46239-2_9

8. Chellapilla, K. (1997). Evolutionary programming with tree mutations: Evolving computer programs without crossover. In Koza, J. R., Deb, K., Dorigo, M., Fogel, D. B., Garzon, M., Iba, H., & Riolo, R.L. (Eds.), *Genetic programming 1997*: Proceedings of the second annual conference (pp.431-438). Stanford, CA: Stanford University.

9. Koza, J. R. (1992). Genetic programming: On the programming of computers by means of natural selection. Cambridge, MA: MIT Press.

10. Jefferson, D., Collins, R., Cooper, C., Dyer, M., Flowers, M., Korf, R., Taylor, C., & Wang, A. (1991). Evolution as a theme in artificial life: The Genesys/Tracker system. In Langton C. G., Taylor, C., Farmer, J. D., & Rasmussen, S. (Eds.), *Artificial Life II: Proceedings of the workshop on artificial life held February 1990 in Santa Fe, New Mexico* (Vol. X) (pp. 549-578). Redwood City, CA: Addison-Wesley. https://www.gwern.net/docs/ai/1992-langton-artificiallife-2.pdf

11. Mindfire Solutions. (2017). *Python: 7 important reasons why you should use Python. Medium.* https://medium.com/@mindfiresolutions.usa/_python-7-important-reasons-why-you-should-use-python-5801a98a0d0b

12. Galván-López, E., McDermott, J., O'Neill, M., & Brabazon, A. (2010). Towards an understanding of locality in genetic programming. *In GECCO' 10 : Proceedings of the 12th annual conference on Genetic and evolutionary computation* (pp. 901-908). https://doi.org/10.1145/1830483.1830646

13. McDermott, J., Galván-Lopéz, E., & O'Neill, M. (2010) A fine grained view of GP locality with binary decision diagrams as ant pheno types. In Schaefer, R., Coota, C., Kolodziej, J., & Rudolph, \G. (Eds.), *Parallel problem solving from nature – PPSN XI: 11th international conference, Krakov, Poland, September 2010 proceedings, part 1* (pp. 164-173). Lecture notes in computer science (Vol. 6238). Berlin,Germany: Springer. https://doi.org/10.1007/978-3-642-15844-5_17

14. O'Neill, M., & and Ryan, C. (1999) Evolving multi-line compliable C programs. In Poli, R., Nordin, P., Langdon, W. B., Fogarty, T. C. (Eds.), *Genetic programming: Second European workshop, EuroGP'99, Göteberg, Sweden, May 1999 proceedings* (pp. 83-92). Lecture notes in computer science (Vol. 1598). Berlin, Germany: Springer. https://doi.org/10.1007/3-540-48885-5_7

15. Oplatková, Z., & Zelinka, I. (2009). Investigation on evolutionary synthesis of movement commands. *Modelling and Simulation in Engineering,* 2009(845080), 1-12. https://doi.org/10.1155/2009/845080

■ AUTHOR

Nimisha Gupta is a rising senior at Yorktown High School. She loves computer science and is fascinated by algorithms and AI. She plans to double major in computer science and some form of engineering in college.

Designing PCR Primers for Detecting Clinically Actionable Single Nucleotide Variation for Non-Small Cell Lung Cancer

Jamie H. Kwon

Chadwick International, Art center-daero 97 beon-gil, Songdo-dong, Yeonsu-gu, Incheon, Republic of Korea
j2kwon2022@chadwickschool.org

ABSTRACT: Liquid biopsy has opened up a new era of noninvasive cancer detection, diagnosis, and prognosis using circulating tumor DNAs (ctDNAs). In this process, ctDNAs are amplified due to their small proportion in blood and are sequenced to collect information about cancer. This study focused on designing PCR primers that can detect clinically actionable single nucleotide variations (SNVs) in EGFR and BRAF genes specific to non-small cell lung cancer (NSCLC). Identifying the presence of genetic aberrations can help when choosing the type of treatments that can work best for the patient with a specific SNV. We evaluated three PCR primer sets for DNA amplification which target EGFR T790M, EGFR L858R, and BRAF V600E, which are SNVs that are clinically actionable for NSCLC. The primers that were designed using this method were eventually used in PCR and sequencing and were verified for their efficacy from the sequencing results. All PCR primer sets that were designed showed high accuracy for amplifying targeted regions when PCR was performed using the primers. This study suggests options for effective PCR primer sets and takes a step forward for clinical use of liquid biopsy since the usage of accurate PCR primers increases the probability of successful sequencing which in turn increases the reliability of liquid biopsy.

KEYWORDS: Liquid Biopsy; Circulating Tumor DNA; Bioengineering; Polymerase Chain Reaction; Primer Design.

INTRODUCTION

Each year there are around 18.1 million patients who get diagnosed with new cases of cancer, a disease that leads to a mortality rate of around 27%.[1] Non-small cell lung cancer (NSCLC) is a type of lung cancer in which more than 85% of all lung cancer patients are diagnosed as having one of three subtypes of NSCLC.[2] Cancer is very difficult to detect in early stages and often times, patients start treatment for cancer too late.[3] The golden standard for diagnosing and detecting cancer has been invasive tissue biopsies, which are done by collecting samples from the actual tumor mass through painful, time consuming, and costly surgical procedures.[4] However, because of the invasiveness of tissue biopsy, it is difficult to periodically monitor the status of cancer patients and track treatment responses.

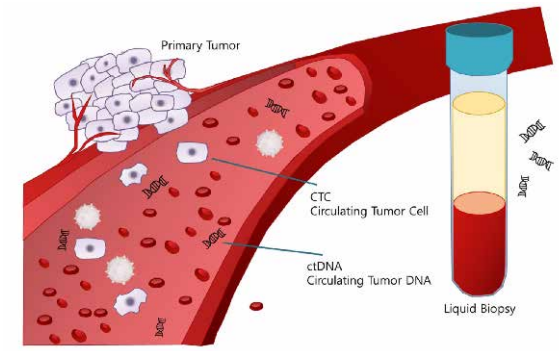


Figure 1. Tumor cells and DNA shed from the primary tumor and circulate around the whole body. Liquid biopsy collects and sequences the ctDNA and gains information about tumor heterogeneity, metastasis, and progression. Unlike conventional tissue biopsies, liquid biopsy is noninvasive, only requiring a vial of blood and providing more accurate data that better reflects the patient's current status.

Liquid biopsy is an emerging noninvasive alternative to detect tumor burdens, identify subclonal sites or transferred metastatic tumors, and track the treatment response through just 5~10ml of blood without actually having to dissect the tumor[5] (Figure 1). Liquid biopsy is done by collecting and sequencing the tumor DNAs (ctDNAs) from the circulating blood that is known to be shed from the tumor cells of the tumor mass. Cell free DNA, which are degraded DNA fragments that are generated during apoptosis and necrosis, are released to the blood plasma and circulate the blood stream. When a growing tumor mass intrudes the blood vessels, the cells and cell-free DNAs (cfDNA) from the tumors are disseminated into the blood vessel and circulate.[6] Because the ctDNA contains the information of the primary and metastatic tumors of a patient, it can be used as a clinical biomarker to monitor, diagnose, and prognose cancer.[5] Not only it is possible to gain information about the originating tumor mass, but it is also possible to determine the metastatic progression of cancer by looking at the amount of ctDNA in the blood.[6] After obtaining the initial values using real-time quantitative PCR, the changing concentration of ctDNA in blood can be significant in clinical settings. According to Corcoran's study, the quantity of ctDNA directly reflects the progressing stages of cancer – which can be used when monitoring residual cancer after rounds of treatments.[5]

ctDNA is an accessible and accurate clinical biomarker due to its short lifespan and thus reflects the patient's status in real-time and its noninvasiveness, which makes it possible

DOI: 10.36838/v2i2.5

to periodically keep track of cancer patients. The half-life of ctDNA is less than 1.5 hours, which allows for data collected from the ctDNA most accurately reflects the current status of the patient.[6] Accurate reflection of the patient status allows the development of precision medicine and personalized treatment that best fits the patient and avoiding unnecessary treatments that can be costly and toxic to the body. However, it is difficult to accurately detect clinically actionable biomarkers due to the heterogeneity of the tumor.

Different tumor cells display distinct features such as cellular morphology and gene mutations.[7] Since cancer displays tumoral genetic heterogeneity, different types and stages of cancers each have distinct genetic aberrations that can be used as clinically actionable biomarkers. Single nucleotide variation (SNV) is a type of genetic aberrations that occurs when there is a mismatched or missing nucleotide in DNA which leads to codon mismatch, resulting in abnormal activities in cells such as the production of abnormal proteins and severe clinical conditions. For example, for NSCLC, epidermal growth factor receptor (EGFR) T790M, EGFR L858R are common single nucleotide mutations that occur in the EGFR gene which develops resistance towards tyrosine kinase inhibitors, a crucial pharmaceutical drug used to treat cancer. B-raf encoding gene (BRAF) V600E is a single nucleotide variation in the BRAF gene which results in oncogenic mutations that cause tumors to develop. Due to the advancement of sequencing technology, even a single mutated nucleotide can be detected through sequencing in high throughput. In order to detect the targeted variation, sequencing technologies are used to analyze the captured ctDNAs. However, according to Corcoran's study, ctDNA is less than 0.1% of total cell-free DNAs.[5] Due to the low population of ctDNAs among other cfDNAs, precise and accurate amplification of ctDNA is crucial. For accurate amplification of ctDNAs, using polymerase chain reaction (PCR), target-specific primers are required to amplify the mutated region.

In this study, I designed and verified PCR primer sets that amplify the single nucleotide variations clinically actionable for NSCLC patients (Figure 2). After executing PCR on samples using the designed primer sets, the reliability of the primers was verified by finding the mutated region in each sequencing result.

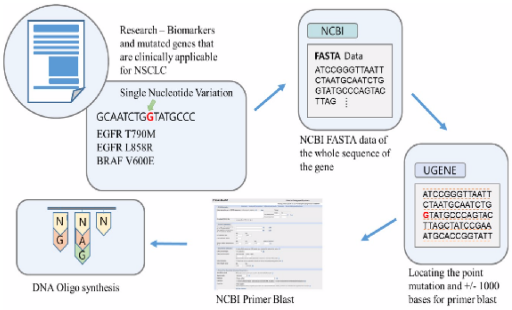


Figure 2. The primer designing and verification process is shown above. Primers are designed through PrimerBlast, an online platform published by NCBI. After choosing the primer sets, primer sequences are sent for oligo synthesis. Using the designed primer, DNA samples, Taq master mix, four samples were amplified. After PCR products were verified for successful amplification through gel electrophoresis the PCR products are purified and sent for Sanger sequencing service. From the sequencing results, the presence of genetic aberrations is verified by comparing the sequence with the original reference sequence of the gene.

RESULTS AND DISCUSSION

***Designing the Primer Sets*:** For the three single nucleotide variations that are clinically actionable for NSCLC, three PCR primer sets (forward/reverse) that target each SNV were designed. In the table below, the melting temperature, GC%, product length, annealing temperature, and the actual sequence is shown (Table 1). All the primer sets are in the appropriate range of Tm and GC% for successful PCR. The measured annealing temperature that was found through PCR and gel electrophoresis is shown in ranges, which is close to the calculated annealing temperature.

Table 1. Three different primer sets used in this study.

Product Name	Tm(c)	GC%	Expected Product Length	Calculated Annealing Temperature(°C)	Measured Annealing Temperature (°C)	Sequence (5' → 3')
EGFR T790M	F: 60.5 R: 60.5	F: 55 R: 55	355bp	68	68~72	F: GCGTAAACGTCCCTGTGCTA R: CCTTTGCGATCTGCACACAC
EGFR L858R	F: 60.5 R: 62.5	F: 50 R: 60	432bp	68	68~72	F: CTCAGAGCCTGGCATGAACA R: GCTCTGGCTCACACTACCAG
BRAF V600E	F: 60.5 R: 62.5	F: 55 R: 55	433bp	69	68~72	F: AAGAGCCTTTACTGCTCGCC R: CTGATGGGACCCACTCCATC

PCR Validation: After designing the PCR primer sets that target EGFR T790M, EGFR L858R, and BRAF V600E, the validity and accuracy of the primer sets were tested (Table 2). The suitability of primers for PCR was verified by looking at features of the primers such as GC%, melting temperature (Tm), product length, annealing temperature and more. Table 2 shows the DNA concentration of the amplified product before sequencing and the presence of SNV found from the sequencing result. The initial value of the DNA samples were 1 ng per µl, from which 1 µl was used for all . Samples with known concentration was diluted to 1 ng per µl. From the DNA concentration that was measured after amplification, it was evident that all samples were successfully amplified. However, for all samples, none of the SNVs were present, which was an expected result since all four samples weren't from an NSCLC patient.

GC percentage is a percentile that shows how much guanine-cytosine base pairs are contained in the total primer sequence. GC percentage plays a crucial role in PCR because guanine-cytosine pairs have stronger bonds than adenine-thymine pairs, resulting in unsuccessful amplification when the GC percentage is out of appropriate range of 40~60%.[8] When there are too much GC bases in the primer, the DNA will coil due to the strong bonds of GC pairs while the primers won't prime well to the DNA if there are not enough GC bases. All the primer sets that were designed for this study were in the appropriate GC percentage of 40%~60%.

Tm, the melting temperature of the primers is also a major factor when validating the effectiveness of the designed primers. Too low Tm will result in loss of specificity while too high Tm will increase the chance of mispriming. Mispriming indicates primers priming to unintended regions, primers not priming to the DNA strand, and degradation of primers

Table 2. PCR using the designed primer sets was tested on four samples: SKBR3, HL60, Sample 1, and Sample 2.

Sample name	Primer name	DNA concentration (ng/μl)	Presence of Single Nucleotide Variation
SKBR3	EGFR T790M	11.7	c2369c
	EGFR L858R	7.2	t2573t
	BRAF V600E	7.0	t1799t
HL60	EGFR T790M	5.5	c2369c
	EGFR L858R	7.2	t2573t
	BRAF V600E	4.7	t1799t
Sample 1	EGFR T790M	7.7	c2369c
	EGFR L858R	7.5	t2573t
	BRAF V600E	8.9	t1799t
Sample 2	EGFR T790M	6.2	c2369c
	EGFR L858R	8.1	t2573t
	BRAF V600E	7.3	t1799t

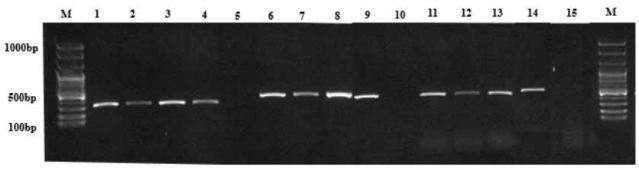


Figure 3. The accuracy of PCR was verified by gel electrophoresis. The lanes of the figure above are the results of amplified products that used primer set for EGFR T790M, EGFRL858R, and BRAF V600E with genomic DNAs from SKBR3, HL60, Sample 1, Sample 2, and nuclease-free water, consecutively.

Figure 4. Presence of single nucleotide variation was verified in the Sanger sequencing result. (a) The reference SNP was found in NCBI database in order to locate the single nucleotide variation point in the gene sequence. A part in front of the SNV point (marked with blue box) was used to locate the SNV point in the Sanger sequencing result. (b) This is an example of Sanger sequencing result of the amplified product of Sample 1 amplified using EGFR L858R primer set. Using a part of the sequence that was found in the reference data, the SNV point (marked with red box) was found in the Sanger sequencing result. Since the base was equal to the reference gene, there was no EGFR L858R SNV present in Sample.1

-which at last reduces the specificity and accuracy of PCR. The Tm of the designed primers were all in the appropriate range of 56~62 degrees.[8]

Verifying the Presence of SNV: After the samples were amplified using the designed primers, the accuracy of PCR was verified by gel electrophoresis (Figure 3). Using 10μl of samples and 100bp ladder, gel electrophoresis was performed at 160V for 35 minutes. Looking at the gel and comparing with the expected product size of each primer, it was possible to verify whether or not the amplification was successful. 1,2,3,4 and 5 lanes of the figure above are the results of amplified products that used primer set for EGFR T790M with genomic DNAs from SKBR3, HL60, Sample 1, Sample 2, and nuclease-free water, consecutively. Since all four marks are between the 300 and 400 bp mark of the ladders, and the expected product size of EGFR T790M primer was 355bp, it could be concluded that the amplification was successful. 6,7,8,9, and 10 lanes of the figure above are the results of amplified products that used primer for EGFR L858R with SKBR3, HL60, Sample 1, Sample 2, and nuclease-free water. Since the expected product size was 432bp and the band is between the 400 and 500 ladder, it can be verified that the amplification was successful. Lastly, lane 11,12,13,14 were the results of amplified products that used primer for BRAF V600E with SKBR3, HL60, Sample 1, Sample 2, and nuclease-free water. Since the expected product size was 433 bp, and the bands are between 400 and 500bp ladders, it verifies that the amplification was successful. Lane 5, 10, and 15 were the negatives that used nuclease free water instead of samples with amplified DNAs. Since no band was shown, the negative was accurate.

The purified PCR products were then sent for Sanger sequencing for SNV validation (Figure 4). The presence of single nucleotide variation was verified by searching a part of reference sequence in the Sanger sequencing result. Verifications were done by searching the reference SNP in the NCBI data base, which, for example, was rs121434569 for EGFR T790M. Comparing the reference genome and finding the location of the SNV point (marked with blue box) in the Sanger sequencing result, I was able to determine if the genomic DNA had the clinically actionable genetic aberration or not. Since the SNV point (marked with red box) is the same as the reference sequence, it was verified that there is no genetic aberration present.

Likewise, the accuracy of the primers when detecting the targeted SNVs were verified using UGENE (Unipro UGENE) by locating both forward and reverse primers in the amplified sequence and determining if the single nucleotide variation point is in the range between both primers. PCR was performed on four DNA samples from SKBR3, HL60, Sample 1, and Sample 2 using the designed primers. Sample 1 and Sample 2 were DNAs extracted from healthy volunteers; however, for privacy, names and details weren't provided. All primer sets successfully amplified the targeted region including the single nucleotide variation point nucleotide. The presence of the genetic aberration was found by searching a part of the gene sequence that is in front of the single nucleotide variation from

DOI: 10.36838/v2i2.5

the original sequence at the sequencing result of the amplified PCR products. The sequencing result of the amplified product showed that no sample had the single nucleotide variation, however; the position of the SNV was present in all products. It was an expected result since the samples weren't from actual NSCLC patients.

CONCLUSION

In this study, we designed PCR primer sets for detecting clinically actionable single nucleotide variation in NSCLC and demonstrated their application to four samples. From the results, it was verified that the designed PCR primers were successful in accurate amplification of targeted sections of DNA. Therefore, the three PCR primer sets that were designed in this research can be used in actual clinical settings, allowing for clinical detection of single nucleotide variations for NSCLC patients. By designing the effective primer sets and successfully amplifying the ctDNA, it will decrease the errors in PCR and sequencing. Reducing amplification and sequencing errors are highly crucial in order for liquid biopsy to replace conventional invasive biopsies in clinical settings because it will prevent extra costs from unsuccessful trial. Higher chances of successful diagnosis will then lead to more conventional use of liquid biopsy. However, there are many improvements in technology needed as well as integration with other platforms and technologies.

In further research, if the amplified products using the designed primers are sequenced through next generation sequencing (NGS), a more sensitive technology can be developed due to the higher sensitivity of NGS than Sanger sequencing. Furthermore, in this study none of the samples were cancerous, therefore no genetic aberrations were detected. By using actual ctDNA from NSCLC patients, the effectiveness and accuracy of the designed primers will be better verified. It will also allow us to research what kind of primers better amplify the SNVs with fewer errors. Also, different genetic aberrations can be detected and amplified with the use of multiplexed PCR where multiple targeted regions of biomarkers can be amplified using multiple PCR primer sets simultaneously. In order to multiply multiple SNVs at the same time through high-throughput detection, the melting temperature of the primer sets should be in equal range, as the specificity of primers depend on the melting temperature.

An example of integrating liquid biopsy with other technology is barcode-free NGS error validation technology. This combined technology can decrease the sequencing errors that often occur when searching for ultra-rare variants such as SNVs that are detected in ctDNAs.[9] According to Yeom *et al.*, this technology allows us to validate if a detection of ultra-rare variant is a sequencing error without extra consumption of resource by validating the variant by amplifying the used sample from the NGS substrate.[9] Since among all variants that are detected, only 1 in 90 detections is a true variant, validating the raw NGS data through the PCR platform can more accurately distinguish true genetic aberrations. Also, for more accurate amplification, I could use Pfu and other types of polymerase instead of Taq polymerase that creates more errors. In this process of PCR validation, the designed primers can be used to amplify the targeted region. Increased accuracy will decrease costs of sequencing processing, thus increasing the accessibility of liquid biopsy in clinical settings.

Another way of integrating this study with other technology is to use the designed primers to amplify circulating tumor cells (CTCs). Although sequencing results of ctDNA provide valuable information about tumor, analysis of CTCs cells can provide different genomic profile of cancer. According to Kim et.al, single CTCs can be captured from whole blood using an optomechanically-transferrable chip and laser-induced isolation of microstructure on optomechanically-transferrable-chip sequencing (LIMO seq), which allows each cell to be sequenced in single cell level.[10] In Kim *et. al*'s study the captured single CTCs are whole genome amplified and go through NGS which then provides a detailed analysis of heterogeneous genome profile of cancer. However, whole genome amplification has a limitation in lower accuracy due to amplification artifacts, which can result in errors when looking for ultra-rare variants such as SNV. By replacing whole genome amplification with PCR, the designed primers can be used for more accurate amplification of DNAs or single CTCs where only clinically valuable targeted regions will be amplified. In addition, by performing multiplex PCR with multiple designed primer sets, it will be possible to obtain more clinical information in higher accuracy than the sequencing result of whole genome amplified CTCs.

With integration with other technologies, liquid biopsy has high potential of becoming the next golden standard of accurate and noninvasive cancer diagnosis, prognosis, and monitoring. Designing clinically actionable primer sets, as described herein, will be the first step in high-performance liquid biopsy.

METHODS

The primers were designed through an online platform, Primer Blast (NCBI). By entering 1000 base pairs upstream and downstream of the single nucleotide variation point from the original sequence and setting the product size between 400~800 base pairs, primer blast provides 5~10 options of primer sets. By looking at the position of the primer sets, I chose the set that contains the single nucleotide variation point.

The template DNAs were extracted from human leukemia cell line (HL60, NCBI) and human breast cancer cell line (SKBR3, NCBI) that were bought from the American Type Culture Collection. Using the QiaGen DNeasy Blood & Tissue kit, DNA was extracted using the DNeasy spin column which includes the DNeasy membrane. By mixing buffers and repeatedly centrifuging the spin columns, the DNAs are extracted. The extraction procedure was done following the manufacturer's instructions.

The Annealing temperature of the designed primers were calculated using the formula below:

$$T_a \text{ Opt} = 0.3 \times (T_m \text{ of primer}) + 0.7 \times (T_m \text{ of product}) - 14.9$$

The PCR cycling condition of the solution (Taq Mix 10µl + Water 6µl + Tem 2µl + Primer F+R 2µl) using the Bio-Rad Thermal cycler was 95˚C for 5:00 min, 95˚C for 0:30 min, 69˚C for 0:30 min, 72˚C for 1:00 min, Goto 2, 32 times, 72˚C for 1:00 min 4˚C forever. PCR products were purified using the GeneAll (Seoul, Republic of Korea) PCR Purification kit, following the manufacturer's instructions.

The amplified PCR products were verified using gel electrophoresis. The PCR products that were used for gel electrophoresis were purified before sequencing using MinElute PCR purification Kit by QiaGen Inc (Hilden, Germany). The purification was done using the min elute spin columns, following the manufacturer's instructions.

Successfully amplified PCR products went through Gel Electrophoresis using SolGent (Daejeon, Republic of Korea) PS600C machine set at 160V for 35 minutes. The PCR products were loaded in an agar gel (Biofactory, Singapore) with 1% dilution.

Before sequencing, the samples were checked if they were amplified through measuring the DNA concentration of the amplified PCR product was through UV absorption method, using NanoDrop by Thermos. (Chicago, USA)

The purified PCR products were sent to Macrogen Inc. (Seoul, Republic of Korea) for standard Sanger sequencing. The Sanger sequencing file was in ABIF format, which was analyzed through UGENE program.

■ ACKNOWLEDGEMENTS

I would like to sincerely thank my mentor Amos Chungwon Lee from Seoul National University for helping me design the experiments and supervising the execution of the experiments.

■ REFERENCES

1. Bray, Freddie, et al. "Global Cancer Statistics 2018: GLOBOCAN Esti mates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries." *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, American Cancer Society, Nov. 2018, pp. 394–424, doi:10.3322/caac.21492.

2. García-Campelo, R., et al. "SEOM Clinical Guidelines for the Treatment of Non-Small Cell Lung Cancer (NSCLC) 2015." *Clinical and Translational Oncology,* vol. 17, no. 12, Springer-Verlag Italia s.r.l., Dec. 2015, pp. 1020–29, doi:10.1007/s12094-015-1455-z.

3. Etzioni, Ruth, et al. "The Case for Early Detection." *Nature Reviews Cancer*, vol. 3, no. 4, Apr. 2003, pp. 243–52, doi:10.1038/nrc1041.

4. Brock, Graham, et al. "Liquid Biopsy for Cancer Screening, Patient Stratification and Monitoring." *Translational Cancer Research*, vol. 4, no. 3, 2015, pp. 280–90, doi:10.21037/4546.

5. Corcoran, Ryan. "Circulating Tumor DNA: Clinical Monitoring and Early Detection." *Annu. Rev. Cancer Biol*, vol. 3, 2019, pp. 187–201, doi:10.1146 /annurev-cancerbio-030518.

6. Bettegowda, Chetan, et al. "Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies." *Science Translational Medicine*, vol. 6, no. 224, NIH Public Access, Feb. 2014, p. 224ra24, doi:10.1126/scitranslmed.3007094.

7. Marusyk, Andriy, and Kornelia Polyak. "Tumor Heterogeneity: Causes and Consequences." *Biochimica et Biophysica Acta - Reviews on Cancer,* vol. 1805, no. 1, Jan. 2010, pp. 105–17, doi:10.1016/j.bbcan.2009.11.002.

8. Dieffenbach, C. W., et al. General Concepts for PCR *Primer Design*.

9. Yeom, Huiran, et al. "Barcode-Free next-Generation Sequencing Error Validation for Ultra-Rare Variant Detection." *Nature Communications*, vol. 10, no. 1, Nature Publishing Group, Dec. 2019, doi:10.1038/ s41467-019-08941-4.

10. Kim, Okju, et al. "Whole Genome Sequencing of Single Circulating Tumor Cells Isolated by Applying a Pulsed Laser to Cell-Capturing Microstructures." *Small*, vol. 15, no. 37, Sept. 2019, p. 1902607, doi:10.1002/smll.201902607.

■ AUTHOR

Jamie Kwon is a current sophomore in Chadwick International, located in South Korea. She is interested in bioengineering, diagnosis technologies, as well as in marine biology and shark-related studies. She hopes to major electrical engineering and bioengineering in university. This research qualified for Korea Scholars Conference for Youth.

DOI: 10.36838/v2i2.5

Choosing the Best Candidate: An Analysis of Voting Systems using a Monte Carlo Method

Ashton Keith

Franklin Academy High School, 42 Husky Lane, Malone, NY 12953, United States
xxphignewtonxx@gmail.com

ABSTRACT: Democracy is one of the fundamental principles underpinning the United States. Recently, the plurality voting system used in our presidential elections has come under scrutiny. Problems such as the spoiler effect and the two-party system have been shown to be directly caused by plurality. Discussions over an alternative cannot determine which alternate system should be used. Every system is susceptible to unwanted paradoxes and potential for tactical voting.[1] Warren Smith used computer simulations under a Monte Carlo method with voters' feelings toward candidates based on stances on a number of issues to determine the system with maximum voter satisfaction. He found that range voting worked best in every trial.[2]

In this study, I will extend Smith's research by using more voters and basing feelings towards candidates on the distance between leanings on a political spectrum using six different voting systems. 10 million elections with 3, 4, 5, 7, 8, 10, 12, 15, 20 and 25 candidates were tested. This study determined that Copeland's Method worked best in every situation, matching a theory made by Davis, DeGroot, and Hinich,[3] followed by approval voting under manipulation and Borda under honesty. Additionally, both approval and plurality voting performed better under tactical voting while the Borda count chose the worst candidate most of the time under the same manipulation.

KEYWORDS: Statistics; Monte Carlo Study; Social Choice; Utility; Voting System; Strategic Voting; Moving Average Strategy.

INTRODUCTION

Citizens believe they understand how voting works: go into a voting booth, select our favorite candidate, and the candidate that receives the most votes wins. Most will be surprised to learn that this is not the only way votes could be cast and counted and that, in fact, it's one of the more mediocre ways of choosing a candidate that satisfies the voters, wishes. Due to the controversial elections of 2000 and 2016, national attention has been focused on how exactly votes are cast. Along with the Electoral College, not addressed in this paper, the plurality voting system has come under considerable scrutiny. Some might say that the system the U.S. uses has been good enough to work for over 200 years and doesn't require a replacement. However, supposing this is true, countries that have only recently adopted democracy are at serious risk of losing it again. If voters who have never experienced a democratic election before are disappointed by the results, they may be willing to abandon democracy all together. We've seen democracies in Latin America, Africa, and the Middle East fall apart after implementing a plurality system. By finding the right voting system to use, a newly established democracy may be able to keep their voters satisfied long enough to maintain stability. Additionally, the effects of different voting systems are relevant in other fields such as computer science and machine learning, corporate decisions, and the stock market.

Alternatives such as range voting, Borda count, instant runoff, and the Condorcet methods have advantages over plurality, but each have their own problems. Which system selects the best candidate is hard to determine and is subject to philosophical discussion. It has been shown that no voting system is perfect: the Gibbard-Satterthwaite Theorem states that all systems discourage voting honestly in some situations.[1] Arrow's theorem states that the only ranked voting system that does not restrict how one votes, does not pick one candidate when voters unanimously prefer another, and is unaffected by adding in losing candidates, is a dictatorship.[4] Furthermore, some systems that seem reasonable are, in reality, so bad that it is better to choose a winner randomly, like bullet voting.[2]

Research Goals: Historically, the effectiveness of voting systems has been determined through mathematical models. This study plans to run simulations on how well each system determines the best candidate and how consistently they do so. This was partially inspired by the work of Smith. However, I will base how well each voter likes a candidate on how close their political leanings are instead of stances on specific issues. This study assumes a utilitarian model of quality, or that the best voting system is that which maximizes the expected utility of society.

I compared four of the most favored voting systems in the literature: range voting,[2] Borda count,[5] Condorcet methods (here represented by Copeland's method),[6] and Hare STV,[7] along with plurality and the lottery system as references.

Hypothesis: It has been shown that, under a distribution of voters symmetric across some point and a utility function that decreases with distance, Condorcet methods will have zero Bayesian regret as the number of voters goes to infinity.[3,8] Therefore, I expect the Copeland's method to have near zero expected Bayesian regret for each situation simulated.

After Copeland's method, I predict that range voting will perform the best in deciding the most desired candidate due to

DOI: 10.36838/v2i2.6

the study done by Smith.[2] My research also shows that range voting has received the least criticism and those made typically deal with the potential for voters to reduce the ballot to either approval voting or plurality.

RESULTS

For the sake of brevity while showing trends with the number of candidates, only results for 3, 15, and 25 candidates are shown. All errors represent the 99% confidence interval for the recorded averages. The highlighted rows in Table 1 are the voting systems with the least Bayesian regret among elections with more candidates and are graphed in Figure 1.

In every scenario, Copeland's method resulted in the highest voter satisfaction. The range/approval and plurality voting systems performed significantly better when voters decide to vote tactically based on the polls, while the ranked voting systems all suffered under tactical voting. The plurality, Hare STV, and honest approval voting systems consistently performed worse in every scenario. The Borda count and range voting's performance improved relative to the other systems as the number of candidates grew as long as the voters were honest. The Borda count under manipulation suffered terribly, having an average regret of 0.28984 vs. 0.00031 under 25 candidates. The lottery system had one-fifth the regret with 25 candidates. I determined that, due to how the strategy works and that voters tend to like those higher in the polls, the very worst candidate ends up placing just above halfway in nearly every ballot, enough for them to win almost every time.

Table 1. 1000x average Bayesian regret of voters under each voting system in 3, 15, and 25 candidate elections.

Voting System	3 Candidates	15 Candidates	25 Candidates
Honest Plurality	2.11808±0.00567	25.3281±0.01941	34.8251±0.02219
MA Plurality	0.38688±0.00195	13.3740±0.01269	20.0535±0.01536
Honest Range	1.40410±0.00452	0.94987±0.00210	0.73014±0.00152
Honest Approval	2.33777±0.00726	3.21130±0.00483	2.51405±0.00349
MA Approval	0.03051±0.00027	0.23412±0.00107	0.24150±0.00097
Honest Borda	0.70203±0.00320	0.45114±0.00130	0.30524±0.00084
MA Borda	0.09226±0.00258	231.883±0.10201	289.837±0.09411
Honest Copeland	0.03050±0.00027	0.04243±0.00021	0.04338±0.00019
MA Copeland	0.09070±0.00255	1.35911±0.01584	2.19883±0.02155
Honest Hare	0.38687±0.00195	8.24258±0.00754	11.1631±0.00813
Lottery	25.9036±0.03878	53.5472±0.04567	59.5996±0.04728

DISCUSSION

Honest Condorcet, moving average approval, and the honest Borda methods all performed better than range voting. What benefitted each can be traced back to the model used. The Condorcet methods benefit due to the distribution of voters was perfectly symmetrical across 0; therefore, it was able to have near zero regret according to the study by Davis, DeGroot, and Hinich.[3] Honest Borda was able to benefit from the fact that it suffers from the reverse of the spoiler effect: candidates do better when a losing candidate of similar views is also running.[9] Because the views nearer the mean of the voters' will be more crowded by candidates, more preferred candidates will have an extra boost in the election. The distribution of views in Smith's study was uniform, preventing this from occurring. As for the moving average approval vote, the candidates at the top of the polls are considered first weeds out competition for the most preferred candidate who is likely to end up second or third in the ranking for honest range voting. The moving average strategy usually corrects this, with the candidate receiving a majority approval and the nature of the polls causing all other candidates to receive far less than the majority. The Borda count works in the same way, but approval voting avoids choosing the worst candidate by not being a ranked system; there is no half-way up the ballot for the worst candidate to be forced on.

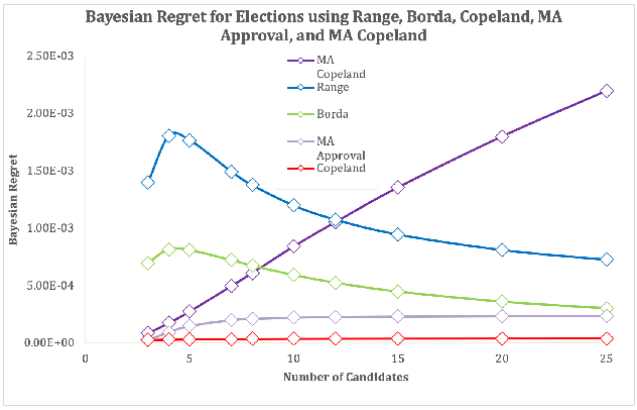


Figure 1. The average Bayesian regrets of Approval Voting with Moving Average Strategy, Honest Borda Count, Honest Range Voting, and Copeland's Method both Honest and with Moving Average Strategy.

Prior research has shown that systems similar to range, plurality, and approval voting are very susceptible to tactical voting[10],[11] but as this study shows, this is not necessarily a downside. In fact, according to a study by James Quinn, plurality, approval, and range voting were the most susceptible to tactical voting of all methods tested yet they benefitted from this manipulation. Smith's study showed manipulation performs significantly worse under every voting system,2 but this study disproved that. Borda count's quirk of almost always choosing the worst candidate appears in Quinn's study but is not further elaborated.[10] My study demonstrates the effects of manipulation depend heavily on the model voters fit into. With the issues-based utilities used in Smith's study views on candidates were heavily divided and the effect of the worst candidate always ending up in the top half of the ballots does not occur.

The conclusions made in this study and Smith's study differ in several ways. The principal difference between our studies focuses on how voters decide how much they like each candidate, or the systems' utility functions. Previous research does not consider how the views of candidates may be similar between voters and instead focuses on worst-case scenarios. This study shows that utility function, whether it involves politicians' proposed policies or the personal economic impact of

DOI: 10.36838/v2i2.6

each possible alternative, is very important in determining a voting system's effectiveness.

There have been several attempts to implement the Hare STV system in mayoral and gubernatorial elections in the United States, but these have often been repealed not long after being implemented. The system was repealed by voters from Burlington's mayoral elections after only two elections and is currently facing legal trouble in Maine's gubernatorial elections.[12],[13] By determining what underlies their voters' views of each candidate, a voting district can decide which voting system results in the highest voter satisfaction.

CONCLUSION

By using a simulation and a utilitarian approach to investigate which voting system has the highest voter satisfaction, this study reveals new information about the nature of a variety of voting systems. First, optimal strategies affect voting systems very differently depending on the cause of voters' views of each candidate. Previous work has shown that some voting systems suffer terribly from strategic voting, including bullet voting, but these findings are true regardless of what the utility function is. Borda count suffers from the same problems, electing the very worst candidate the majority of the time only when voters' opinions of each candidate are highly correlated. Second, if voters vote tactically, the performance of a voting system does not necessarily perform any worse than if voters are honest. Both plurality and range/approval voting result in greater voter satisfaction when voters act in their own best interests. Third, approval voting under manipulation and Borda under honesty are the most utilitarian voting systems with the utility function used in this study. This is contrary to Smith's study in which range voting had the best performance overall.[2] Each of the three systems were able to exploit something in the model to their advantage, Condorcet exploited the symmetric distribution, Borda used the concentration of candidates around the mean view, and Approval exploited the strategy used. Hopefully, this study will help those deciding what voting system to use.

METHODS

Set Up Elections: I created three candidates who will be running in the elections, each assigned two numbers representing their political leanings. This was inspired by how The Political Compass™ assigns political persuasions, as seen in Figure 2. These values were assigned according to a Gaussian distribution with a standard deviation of 0. 2 and mean of 0.

One thousand voters were created to vote in the elections. The same process to determine political leanings was used as for the candidates.

Running the Elections: For each voter, the utility of each candidate was calculated using the following:

$$U_C = -||\vec{v} - \vec{C}||^2,$$

where U_C is the utility of the candidate for that voter and V and C are the vectors representing the views of the voter and candidate, respectively. A higher utility represents a greater desire that a candidate wins.

Using the utilities of each candidate for each voter, ballots

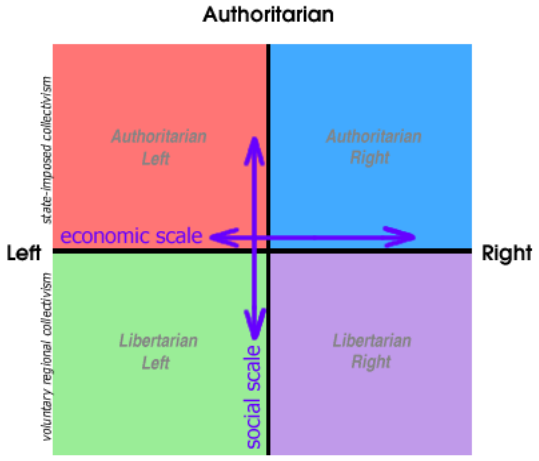


Figure 2: The 2D political spectrum that The Political Compass™ uses in assigning political leanings.[14]

were filled out according to the following voting strategies: honest range, honest plurality, honest ranked [Borda count, Copeland's method, and Hare STV], and honest approval. The winner was determined under each of the following voting systems with their corresponding ballots: plurality voting, range/approval voting, Borda count, Copeland's method, Hare STV, and lottery. Using the results of the plurality, range, Borda count, and Copeland's method elections, ballots for each voter were filled out under their moving average strategy. The winner was determined according to the corresponding voting system.

I calculate the Bayesian Regret, or the sum of the winner's utility to all of the voters minus the sum of the most preferred candidates, of each of the elections run. This can be found using the following equation:

$$D = ||\vec{W} - \vec{u}||^2 - ||\vec{B} - \vec{u}||^2,$$

where D is the Bayesian Regret of the election and W ,B and μ are the vectors representing the winning candidate's views, most preferred candidate's views, and the mean of the voters' views, respectively. The most preferred candidate has the views closest to the mean of the voters.

Repeat : For the elections with three candidates, a total of 10 million elections were run and their regrets analyzed for each of the voting systems and strategies. The most utilitarian voting system under the model used ends up producing the lowest average regret overall.

More elections were run using 4, 5, 7, 8, 10, 12, 15, 20, and 25 candidates using the same method used for 3 candidates. By doing this, I was able to determine whether a voting system's performance stays consistent in large and small elections.

The source code used in this study is available on GitHub™.[15]

Voting Systems A. Plurality voting is the system most people are familiar with. Each voter casts a vote by marking a single candidate from a list; the candidate with the most votes wins the election. This system is known to suffer from the spoiler effect, an inevitable trend towards only two candidate races and failing to meet the independence of irrelevant alternatives

which means that removing a losing candidate from the ballots might change who wins the election.

B. In range voting each voter gives a rating for every candidate on a ballot. The candidate with the highest average rating wins the election. This system is known to fail the later-no-harm criterion, which means that voters will be encouraged to give some of their non-favorite candidates the lowest possible ratings, as well as the Condorcet criterion where a candidate may win even though a majority of voters prefer a different candidate. Approval voting is equivalent to range voting except every rating is restricted to 1 or zero. The paradoxes of range voting also apply to approval voting.

C. In Borda count each voter ranks the candidates in order of most to least favorite. Each candidate gets a vote for every other candidate they beat on each ballot; ex: if there are five candidates, the top preference gets four votes, second preference gets three, etc. The candidate with the most votes wins. This system does not satisfy the Condorcet criterion nor the independence of irrelevant alternatives. Voters are sometimes encouraged not to rank their top preference first.

D. In Copeland's method each voter ranks their candidates in order of preference. After all of the ballots are cast, each candidate is compared pairwise with every other. A candidate receives a point for every other candidate they rank higher than in a majority of ballots. The candidate with the most points wins. This voting system is the only one listed here that satisfies the Condorcet criterion and was chosen for this purpose. It is not independent of irrelevant alternatives, fails later-no-harm, and sometimes encourages voters to not rank their favorite candidate first. In fact, sometimes voters are better off not voting at all than to vote honestly.

E. In Hare STV, each voter ranks their candidates in order of preference. Votes are assigned to the top preference of each ballot and the candidate with the least votes is eliminated. The votes are redistributed to the top preference still running on each ballot. This continues until there is one candidate left standing. This voting system fails to satisfy the Condorcet criterion, independence of irrelevant alternatives, and the consistency criterion. Voters are sometimes discouraged from ranking their favorite candidate first or even from voting at all, and a voter may actually cause their favorite candidate to lose by ranking them higher on their ballot. Despite all of this, this voting system is the most popular system for head-of-state elections, behind plurality.

F. In lottery, a voter is chosen at random and their top preference wins the election. This system is not taken seriously and is only included for theoretical reasons. Although this voting system fails several criteria, it has one very important property: voters are always encouraged to vote honestly. Therefore, we can predict exactly how every voter is going to fill out their ballots if we know their views of each candidate.

Voting honestly in range voting and approval voting is not well defined. Voters can put down the exact utility each candidate has for them in range voting, but voters often only know the utility of one candidate compared to another. For this to work, all voters need to agree on what to compare the candidates against, which is not feasible. For this study, I defined honest range voting as marking the utilities of each candidate given by the simulation, transformed linearly so that a voter's favorite candidate receives a 1 and their least favorite candidate a zero. In approval voting I defined the threshold for marking 1 as the average utility of all of the candidates running. If a voter thinks a candidate is better than average, a one is put on the ballot for them.

I will also be using the moving average strategy for plurality, approval, Borda, and Copeland. Each voter considers each candidate in order of their likelihood of winning, starting with the top two. The voter gives the most votes available to the candidate they prefer more and the least to the one they prefer less. Each successive candidate is compared with all of the ones considered before them. If said candidate is liked more than the considered candidates on average, then the maximum number of votes available is given, and the least number of votes otherwise. This strategy was proven to be the optimal strategy for voters under plurality, range/approval, and Borda by Smith.[2]

■ ACKNOWLEDGEMENTS

I thank my Biology teacher and Science National Honors Society advisor, Ms. Rogers, for repeatedly reviewing the research design and final paper and pushing for my project to get presented to a wider community.

I also thank Mary Eileen Wood and the Terra Northeast Fair Judges for their interest and feedback, for choosing my project as a Ying Scholars Semifinalist, and for suggesting that I submit to this journal when my application for the Ying Scholarship failed to go through.

■ REFERENCES

1. Gibbard, A. Manipulation of Voting Schemes: A General Result. *Econometrica* 1973, 41 (4), 587.
2. Smith, W. D. Range voting, 2000. RangeVoting.org - Center for Range Voting. https://www.rangevoting.org/WarrenSmithPages/homepage/rangevote.pdf (accessed Mar 30, 2019).
3. Davis, O. A.; DeGroot, M. H.; Hinich, M. J. Social Preference Orderings and Majority Rule. Econometrica 1972, 40 (1), 147-157.
4. Arrow, K. J. A Difficulty in the Concept of Social Welfare. Journal of Political Economy 1950, 58 (4), 328–346.
5. Saari, D. G. Geometry of Voting. Handbook of Social Choice and Welfare 2011, 897–945.
6. Voting Research - Voting Theory. https://www.princeton.edu/~cuff/voting/theory.html (accessed Mar 30, 2019).
7. Bartholdi, J.; Orlin, J. Single Transferable Vote Resists Strategic *Voting. Social Choice and Welfare* 1991, 8 (4).
8. Smith, W. D. Best Voting Systems in D-Dimensional Politics Models. RangeVoting.org - Center for Range Voting. https://www.rangevoting.org/BestVot2.html (accessed Nov 08, 2019).
9. Olson, B. Election Methods In Space! https://bolson.org/voting/sim_one_seat/www/ (accessed Dec 15, 2019).
10. Quinn, J. Voter Satisfaction (VSE) FAQ | James Quinn. https://electionscience.github.io/vse-sim/VSE/ (accessed Nov 10, 2019).
11. Balinski M.; Laraki R. Election by Majority Judgment: Experimental Evidence. In: *In Situ and Laboratory Experiments on Electoral Law Reform*; Dolez B., Grofman B., Laurent A., Eds.; Studies in Public Choice 25; Springer: New York, 2011; pp 13-54. https://www.fc.up.pt/dmat/engmat/2012/seminario/artigos2012/bruno_neto/ElectionByMajorityJudgment(ExperimentalEvidence)Final.pdf (accessed Dec 14, 2019).
12. .McCrea, L. Burlington Voters Repeal Instant Runoff Voting. https://vprarchive.vpr.net/vpr-news/burlington-voters-repeal-instant-runoff-voting/ (accessed Mar 30, 2019).

DOI: 10.36838/v2i2.6

13. Seelye, K. Q. Ranked-Choice Voting System Violates Maine's Constitution,Court Says. https://www.nytimes.com/2017/05/23/us/maine-ranked-choice-elections-voting.html?_r=0 (accessed Mar 30).
14. About the Political Compass. https://www.politicalcompass.org/analysis2 (accessed Mar 30, 2019).
15. https://github.com/AshtonKeith/Voting-System-Simulations

AUTHOR

Ashton Keith graduated from Franklin Academy High School in June 2019 and is currently attending SUNY Binghamton as a math and physics major. He has always been interested in math, statistics, and programming, and is grateful to have found a project that incorporates all three.

Identification of Escherichia Coli O157:H7 Outer Membrane Proteins Which Mediate Adherence to Bovine Endothelial Cells

Vishnu Iyer

University High School of Indiana, 2825 W. 116th Street, Carmel, Indiana-46032, USA
iyer.cvenk@gmail.com

ABSTRACT: *Escherichia coli (E. coli)* O157:H7 is a human pathogenic bacterium known to cause foodborne outbreaks of bloody diarrhea and hemolytic uremic syndrome. Cattle are a primary reservoir of *E. coli*, which are shed in the feces and transmitted to humans through contaminated dairy products and undercooked or raw meat. The purpose of this study is to identify bacterial outer membrane proteins (OMPs) involved in the colonization of bovine endothelial cells. It was hypothesized that similar *E. coli* O157:H7 OMPs would be involved in adhesion to both human and bovine endothelial cells. An optimized "pull-down" technique was developed to selectively anchor biotin-labeled bovine endothelial cell surface proteins (CSP) onto a streptavidin bead matrix followed by incubation with *E. coli* OMPs. After washing the beads, bound OMPs were eluted and subsequently analyzed by peptide sequencing. A total of 90 proteins were identified, including significant hits of OmpA, OmpX, OmpSlp and flagellin, in addition to several chaperone proteins which represented a group of previously well-characterized mediators of E. coli adhesion to human cells. This is the first study in the literature to demonstrate the role of these OMPs for attachment of *E. coli* to bovine endothelial cells..

KEYWORDS: Microbiology, *E. coli*; O15:H7; outer membrane proteins; bovine; OmpA.

INTRODUCTION

The bacterium *Escherichia coli* is a normal resident of humans and animal intestines. Most *E. coli* do not cause any harm, actually helping in food digestion and preventing colonization of harmful bacteria within the intestinal tract. O157:H7 serotype of *E. coli* belongs to the family of enterohemorrhagic zoonotic bacteria is one of the leading causes of food-borne disease outbreaks in the United States.[1]Infection is transmitted from livestock to humans via consumption of uncooked and contaminated dairy and meat products. *E. coli* O15:H7 strain produces Shiga toxin which results in severe systemic infections that can lead to symptoms such as bloody diarrhea and hemolytic uremic syndrome.[1]

E. coli O157:H7 attach to the human intestinal epithelial cells and flatten the finger-like projections of the epithelial cells described as effacement.[2] The attachment and effacement lesions on epithelial cells are a result of outer membrane proteins (OMPs) binding to host cells and releasing bacterial toxins and virulence factors. Proteins such as OmpA and OmpX, in addition to flagellins, may be critical for bacterial binding to human epithelial cells. Bacterial binding causes rearrangement of the host's cytoskeleton and forms pedestals under the surface of attached bacteria resulting in irreversible damage to the intestinal epithelial cells. *E. coli* O157:H7 inserts an intimin receptor into the epithelial cells which binds to intimin on bacteria resulting in tight attachment and efficient host colonization.[2] Next, Shiga toxins are released, causing death of the epithelial cells, breakdown of intestinal barrier, and hemolytic diarrhea. Tissue damage can stimulate the host's immune system which causes further damage to the intestines, including increased intestinal permeability and apoptosis.

E. coli O157:H7 is a concern for the food industry since the bacteria reside in bovine intestines without causing any pathogenic symptoms.[3] The bacteria are shed in the feces and pose a risk to people in contact with or consume contaminated food or drink. Additionally, *E. coli* can be transmitted between humans through occupational exposure. Water-borne outbreaks can occur in swimming pools contaminated with animal feces. Shedding of bacteria by beef cattle is higher in spring and summer than during winter months. Every outbreak of *E. coli* can be traced back to cattle and recalling contaminated products pose huge financial burdens to cow farmers. Several pre-harvest controls have been adopted in farms, including house cleaning, feed and water management, and vaccines against bacterial iron transport proteins such as siderophores.[1,2] Post-harvest control measures include thermal and chemical decontamination of infected carcasses.

In contrast to human colonization of *E. coli* O157:H7, binding of bacteria to cattle epithelium does not result in diarrhea. Cattle are known to lack Shiga toxin receptors and may be resistant to the toxic lesions in the gut, making them ideal reservoirs for the pathogen.[4] Earlier studies have shown bacterial binding to bovine epithelial cells but the exact bacterial OMPs involved in adherence and colonization has not been characterized. *E. coli* binding to human endothelial cells is mediated in part by OMPs and results in upregulation of several adhesion molecules, such as ICAM-1 and VCAM-1, leading to enhanced leukocyte binding to endothelial cells.[5] It is

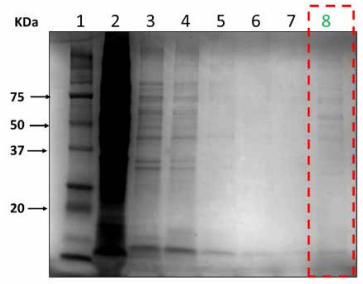
DOI: 10.36838/v2i2.7

unclear whether bacterial binding to bovine endothelial cells results in binding and activation of the immune system. As a first step, we sought to identify bacterial OMPs which are potentially utilized to bind the bovine endothelium.binding and activation of the immune system. As a first step, we sought to identify bacterial OMPs which are potentially utilized to bind the bovine endothelium.

The purpose of this study is to identify bacterial OMPs involved in the colonization of bovine endothelial cells. The hypothesis is that previously described *E. coli* O157:H7 OMPs which were involved in colonization of human cells are responsible in binding bovine endothelial cells. This would highlight common mechanisms in the colonization of human and bovine endothelial cells that lead to rapid transmission of infection from bovine to human hosts.

RESULTS AND DISCUSSION

Isolation of E. coli O15:H7 OMPs binding to bovine endothelial cells: Cell surface biotinylated bovine endothelial proteins were immobilized onto streptavidin beads. Unbound host proteins were washed away with mild detergent (0.1% Triton-X) followed by incubation with bacterial OMP preparation. Results are shown in Figure 1. The majority of the unbound host and bacterial proteins were detected in the flow-through fractions (Lanes 2 and 3). The efficiency of the washing procedure was monitored by a subsequent reduction of detectable proteins in the wash steps (Lanes 4-7). The acid



12% SDS-PAGE
Coomassie blue stain

Lane	Samples
1	Molecular weight standard
2	Flow-through - host proteins
3	Flow-through- E. coli OMP
4.	Wash 1
5	Wash 2
6	Wash 3
7	Wash 4
8	Elution output

Figure 1. Proteins from the flow-through of the host and bacterial proteins were run under reducing and denaturing conditions using 12% SDS-PAGE and visualized with Coomassie blue stain.

Table 1. Summary of bacterial OMPs and their proposed functions

Protein	Gene	Coverage	#peptides	# amino acids	MW (kDa)	Predicted protein function
Outer membrane protein A	OMP A	40.17	8	346	37.2	Adhesin/invasin; binds GlcNAc1,4-GlcNAc epitopes on glycoprotein. Target for immune system
Outer membrane protein X	OMP X	13.45	2	171	18.6	Adhesion and invasion of lung epithelial cells
Outer membrane protein slp	OMP slp	9.57	1	188	20.9	Initial adhesion. Binds to human polymeric immunoglobulin receptor
Flagellin	Flagellin	56.75	21	585	59.9	Motility. Target for immune attack

elution procedure resulted in the isolation of a small subset of proteins from the total fraction, observed in Lane 8. Stringent washing conditions with 5M NaCl did not yield any proteins.

Identification of Bacterial OMPs: Partial sequence analysis of peptides revealed about 90 proteins including three OMPs, OmpA, OmpX and OmpSlp. The data is summarized

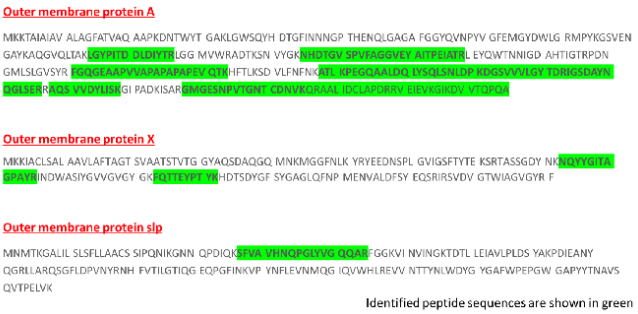


Figure 2. Full-length protein sequences of OMPs isolated from E. coli O157:H17 which bind bovine endothelial cell surface. Identified peptide sequences are highlighted.

in Table 1. A total of 8 peptides were identified for OMP A while fewer peptides were identified for OmpX and OmpSlp (Figure 2). Additionally, flagellin, a protein important for bacterial mobility, was identified as an interacting partner with bovine cell surface proteins. The putative roles of these proteins previously described for human cells are included in Table 1..

In addition to these four well-characterized *E. coli* OMPs, several chaperone proteins were identified. The exact connection of these proteins to mediate bacterial adhesion to bovine endothelial cells is unclear since these proteins may represent a pool of non-specific proteins commonly detected during preparation of bacterial membranes. Other potential OMPs such as fimbrial biogenesis outer membrane protein, Phosphoporin PhoE, and murein lipoprotein OS were identified with lower peptide coverage and need further confirmation.

We have successfully developed a pull-down technique to isolate bacterial OMPs which bind to bovine endothelial cells. LC/MS/MS results showed several significant hits including OmpA, OmpX, OmpSlp, and flagellin. These adhesins should be further studied to develop therapeutic interventions such as blocking antibodies or vaccine-based approaches.

OmpA is a highly expressed protein of *E. coli* and can aid in adhesion and invasion. The exposed surface of OmpA can trigger a host immune response.[6] OmpA has a Swiss army knife shape and forms a pore or barrel-like structure, although the exact function is unclear. OmpA and OmpX expression are tightly regulated by the bacteria under different growth conditions.[6]

Approaches to anchor host cell surface proteins to binding columns can potentially disrupt biologically relevant interactions between host cells and bacterial proteins. Alternative approaches, such as whole cell labeling with biotin followed by incubation of fixed bacteria and crosslinking prior to peptide sequencing, can reveal novel host-pathogen interactions which may have been missed during this study. Additional studies are needed to study the impact of bacterial adhesion to host cells using OMP-specific blocking antibodies to further understand the functional significance of our findings.

CONCLUSION

Our study has revealed remarkable similarities between the adhesion molecules involved in binding *E. coli* to human and bovine endothelial cells. The lack of expression of immunoreactive receptors for Shiga toxin by bovine endothelial cells in

contrast to the expression in other tissues such as convoluted kidney tubules[4] may provide an explanation for the lack of hemorrhagic diarrhea in cattle in spite of conserved colonization mechanisms in human and bovine hosts. Further studies are needed to profile additional bacterial OMPs which may interact with bovine endothelial cells by culturing bacteria under different stress conditions including nutrient deprivation. Additionally, bacterial invasion results in the release of inflammatory mediators by host cells. Such inflammatory mediators can upregulate expression of additional cell surface host proteins which interact with *E. coli* OMPs. Thus, manipulation of both host and bacterial culture conditions prior to membrane preparations may reveal additional binding partner of bacterial OMPs. In summary, this study is the first to demonstrate the role of these OMPs for attachment of *E. coli* to bovine endothelial cells.

METHODS

Bacterial OMP Extraction: Bacterial cultures were grown in 1L RAS-BH1 broth for 12-14 hours. Cells were pelleted by centrifugation, resuspended in 20 ml of cold 0.75 M sucrose in 10 mM Tris, pH 7.8, and incubated on ice for 20 minutes. Following the addition of 1.5 mM EDTA at a constant rate, lysis was performed by ultrasonication in an ice-water bath. Cell debris was removed by centrifugation at 6500 g for 15 minutes at 4 °C. Supernatant was then centrifuged at 386,984 g for 2h at 2-4 °C. After discarding the supernatant, the pellet was resuspended in 10 ml of cold 0.25M sucrose-3.3 mM Tris buffer and 2 ml of Triton-X 100 (20 mg/ml). The volume was

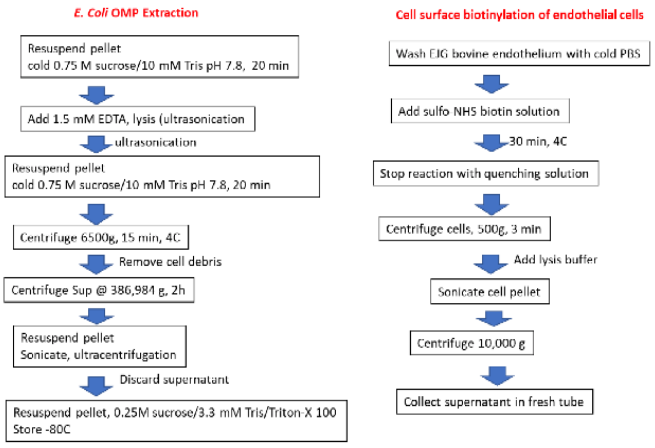


Figure 3. Stepwise protocol for extraction of bacterial OMPs (Left panel) and biotinylation and membrane preparation of bovine endothelial cells (Right panel) is described.

adjusted to the original sonicate suspension and the ultracentrifugation step was repeated. The pellet was resuspended in cold 0.25M sucrose-3.3 mM Tris buffer and stored in -80 °C until further use. A stepwise protocol is shown in Figure 3.

Extraction of cell surface proteins from bovine endothelial cells (EJG cells): Confluent EJG cells were quickly washed twice with ice-cold phosphate buffered saline (PBS) to prevent rounding and detachment. Cell surface biotinylation with sulfo-NHS-SS-biotin was performed according to the manufacturer's protocol (Pierce Biotechnology). Cells were incubated with 10 ml biotin solution in an orbital shaker for 30 minutes at 4 °C. The reaction was stopped by the addition of a quenching solution. Cells were gently scraped and centrifuged at 500 g for 3 minutes. Cell pellets were resuspended in lysis buffer and disrupted by sonication. Cell lysates were centrifuged at 10,000 g for 2 minutes and the clarified supernatant was collected in a fresh tube and stored at -20 °C. A stepwise protocol is shown in Figure 3.

Pull-down Assay and Peptide Analysis: All procedures were conducted using the Pierce cell surface protein isolation kit. Briefly, bovine EJG endothelial lysate was incubated with streptavidin beads for 60 minutes at room temperature in a rotating mixer. Unbound proteins were washed three times with 0.5% Triton-X Tris buffer followed by incubation with the bacterial OMP preparation for an additional 60 minutes at room temperature. Free OMPs were washed three times with 0.5% Triton-X Tris buffer followed by elution under acidic conditions.

Peptide Analysis: Eluted proteins were digested with trypsin, concentrated and desalted according to standard procedures by an external contract research organization (Poochon Scientific, Frederick, MD). Samples were analyzed using LC/MS/MS analysis and bacterial OMP proteins were identified based on peptide sequence analysis. Briefly, LC/MS/MS analysis was carried out using a Thermoscientific Q-Extractive hybrid mass spectrometer. Peptide mixtures were loaded on reverse phase PicoFrit column and trapped peptides were eluted using 3-36% linear gradient of acetonitrile in 0.1% formic acid. Eluted peptides were ionized and sprayed into the mass spectrometer. Raw data was searched against *E. coli* O15:H7 strain protein sequence database downloaded from UniportKB using the Proteome Discoverer 1.4 software. The maximum false peptide rate was specified as 0.01 (confidence scores were not provided by the external CRO).

ACKNOWLEDGEMENTS

I would like to thank Ms. Prabha Bista and Dr. Sanjeev Narayanan for their guidance to perform this research in the Department of Comparative Pathobiology at Purdue University. Ms. Prabha offered help to learn all the relevant protocols and trouble-shooting tips during the study.

REFERENCES

1. Pfizer Animal Health. (2011). A guide to E. coli O157 in cattle. Retrieved from zoetisus.com/_locale-assets/mcm-portal assets/services/documents/srpecoli/e_coli_tech_manual_final.pdf.
2. Saeedi, P., Yazdanparast, M., Behzadi, E., Salmanian, A.H., Mousavi, S.L., Nazarian, S., & Amani, J. (2017). A review on strategies for decreasing E. coli O157:H7 risk in animals. *Microbial Pathogenesis*, 103, 186-195. DOI: 10.1016/j.micpath.2017.01.001.
3. Lim, J.Y., Yoon, J., & Hovde, C.J. (2010). A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. J *Microbiol Biotechnol*, 20(1), 5-14. Retrieved from ncbi.nlm.nih.gov/pubmed/20134227.
4. Pruimboom-Brees, I.M., Morgan, T.W., Ackermann, M.R., Nystrom, E.D., Samuel, J.E., Cornick, N.A., & Moon, H.W. (2000) Cattle lack vascular receptors for *Escherichia coli* O15:H7 Shiga toxins. *Proc Natl Acad* Sci U.S.A, 97(19), 10325-10329. DOI: 10.1073/pnas.190329997..
5. Kim, J.H., Yoon, Y.J., Lee, J., Choi, E.J., Yi, N., Park, K.S., Park, J.,

DOI: 10.36838/v2i2.7

Lotvall, J., Kim, Y.K., & Gho, Y.S. (2013). Outer membrane vesicles derived from *Escherichia coli* upregulate expression of endothelial cell adhesion molecules in vitro and in vivo. *PLoS One*, 8(3), e59276. DOI: 10.1371/journal.pone.0059276

6. Smith, S.G., Mahon, V., Lambert, M.A., & Fagan, R.P. (2007). A molecular Swiss army knife: OmpA structure, function and expression. *FEMS Microbiol Lett*, 273(1), 1-11.DOI: 10.1111/j.1574-6968.2007.00778.x.

AUTHOR

Vishnu Iyer is a sophomore at University High School of Indiana. He is very interested in STEM and presented his research work at the Annual Student Research conference last year. He competed at the United States National Chemistry Olympiad and placed in the Top 150 in the nation. He aspires to be a surgeon and plans to spend part of his time volunteering at non-for-profit organizations to offer free medical care to the needy.

Analysis of the Activity of Natural Biopesticides on Different Plant Varieties

A. K. Akshay, Dayajanaki, M.Anoop*

Sree Gokulam Public School, Cheruvallimukku Mamomkshetram Rd, Attingal, Thiruvananthapuram, Kerala 695104, India
mailmeanp86@gmail.com

ABSTRACT: Agriculture has become high investing and low yielding due to uncontrolled use of synthetic chemicals that are harmful to ecosystems. To protect our biosphere, organic farming replaces harmful chemicals with bio-fertilizers and bio-pesticides. This process maintains soil fertility. Biopesticides can support the life of soil organisms and can generate high crop yields. This project formulated biopesticides from milk as well as a combination of chili, garlic, and decayed vegetables and fruits. These components were applied to spinach, brinjal, and tomato plants. The methods provided rapid plant growth and action against pests. The results are promising for future biopesticide development that should be further researched.

KEYWORDS: Biopesticides; Milk; Chili; Garlic; Tomato; Spinach; Brinjal.

INTRODUCTION

The earliest agricultural biocontrols were plant extracts. For example, nicotine was used to control plum beetles as early as the 17th century. Experiments involving biological controls for agricultural pests date back to 1835, when entomologist Agostino Bassi demonstrated that white-muscadine fungus (*Beauveria bassiana*) causes an infectious disease in silkworm. Experiments with mineral oils as plant protectants were also reported in the 19th century. The 20th century saw a rapid institutional expansion of agricultural research and, with it, an ever-growing number of studies and proposal for biocontrols.[1]

Biopesticides generally display an erratic performance under field conditions and hence represent a small percentage of the global crop protection market in the performance of microbial insecticides including Baculoviruses which is compromised due to its sensitivity to desiccation and ultraviolet light. Therefore, successful implementation of a pest control program involving Baculoviruses depends upon several factors, beginning with a detailed understanding of host–parasite biology and disease dynamics. A major factor in biocontrols is the selection of the most virulent isolate of the virus for an insect pest in a given geographical location. Isolates differ in their insecticidal phenotypes and a study has identified genes in SeMNPV that are likely involved in the pathogenicity and virulence of OBs.[1] Biopesticides generally possess several advantages over conventional pesticides. A serious health hazard for chemical pesticides is the presence of pesticide residues in food. Other risks include developed resistance in the targeted pest populations and a decrease in biodiversity. An advantage of microbial pesticides is that they replicate in their target hosts and persist in the environment due to horizontal and vertical transmission which may cause long-term suppression of pest populations even without repeated application.[2]

The increased social pressure to replace synthetic pesticides with other alternatives that are safe to humans and non-target organisms has led to increased development of compounds. Biopesticides include an array of microbial pesticides and biochemicals derived from microorganisms, photochemical and other natural sources, and processes involved in the genetic modification of plants to express genes encoding insecticidal toxins.3 Biological compounds are used to control pests, pathogens, and weeds by a variety of means. Microbial biocontrols may include a pathogen or parasite that infects the target.3 They often act as competitors or inducers of plant host resistance. Biochemical controls can also act through inhibiting the growth, feeding, development or reproduction of a pest or pathogen. Other biocontrols form a barrier on the host to act as a feeding or infection inhibitor through a multifunctional approach.3biochemicals derived from microorganisms, photochemical and other natural sources, and processes involved in the genetic modification of plants to express genes encoding insecticidal toxins.[3] Biological compounds are used to control pests, pathogens, and weeds by a variety of means. Microbial biocontrols may include a pathogen or parasite that infects the target.[3] They often act as competitors or inducers of plant host resistance. Biochemical controls can also act through inhibiting the growth, feeding, development or reproduction of a pest or pathogen. Other biocontrols form a barrier on the host to act as a feeding or infection inhibitor through a multifunctional approach.[3]

Biopesticides are usually applied similarly to chemical pesticides. With all pest management products, especially microbial agents, effective control requires appropriate formulation and application.

Biochemical biopesticides are naturally occurring compounds and are characterized by a nontoxic mode of action that may affect the growth and development of a pest, its ability to reproduce, or pest ecology. The growth and development of treated plants, including their post-harvest physiology, is deeply dependent on the activity of application of biopesticides.[2]

RESULTS AND DISCUSSION

The treatment using all of the aforementioned methods was found to be very effective. In the initial phase, the plants were

DOI: 10.36838/v2i2.8

Table 1. Analysis of the application of different fertilizer combinations in different stages of growth of Brinjal.

Trial	Crop Stage	Duration (days)	Fertilizer Type	Fertilizer Effect
1	Plant establishment	10	Chili & garlic	effective
	Vegetative	26	Chili & garlic	effective
	Flower insertion to first picking	35	Chili & garlic	effective
	Harvesting	69	Chili & garlic	effective
2	Plant establishment	10	From milk	effective
	Vegetative	30	From milk	effective
	Flower	30	From milk	effective
	Harvesting	75	From milk	effective
3	Plant establishment	10	Decayed vegetables and fruits	effective
	Vegetative	30	Decayed vegetables and fruits	effective
	Flower insertion to first pick up	30	Decayed vegetables and fruits	effective
	Harvesting	70	Decayed vegetables and fruits	effective

Table 2. Analysis of the application of different fertilizer combinations in different stages of growth of Tomato.

Trial	Crop Stage	Duration (days)	Fertilizer Type	Fertilizer Effect
1	Plant establishment	10	Chili & garlic	effective
	Vegetative	26	Chili & garlic	effective
	Flower insertion to first picking	40	Chili & garlic	effective
	Harvesting	50	Chili & garlic	effective
2	Plant establishment	10	From milk	effective
	Vegetative	30	From milk	effective
	Flower	40	From milk	effective
	Harvesting	46	From milk	effective
3	Plant establishment	10	Decayed vegetables and fruits	effective
	Vegetative	30	Decayed vegetables and fruits	effective
	Flower insertion to first pick up	38	Decayed vegetables and fruits	effective
	Harvesting	42	Decayed vegetables and fruits	effective

Table 3. Analysis of the application of different fertilizer combinations in different stages of growth of spinach.

Trial	Crop Stage	Duration (days)	Fertilizer Type	Fertilizer Effect
1	Plant establishment	3	Chili & garlic	effective
	Vegetative	5	Chili & garlic	effective
	Flower insertion to first picking	7	Chili & garlic	effective
	Harvesting	9	Chili & garlic	effective
2	Plant establishment	3	From milk	effective
	Vegetative	6	From milk	effective
	Flower	8	From milk	effective
	Harvesting	10	From milk	effective
3	Plant establishment	2	Decayed vegetables and fruits	effective
	Vegetative	4	Decayed vegetables and fruits	effective
	Flower insertion to first pick up	5	Decayed vegetables and fruits	effective
	Harvesting	8	Decayed vegetables and fruits	effective

categorized and marked, then divided into different groups. All groups were exposed to the different varieties of treatment.

The plants were all grown simultaneously. The experiments have remarkable results. Using chili and garlic with brinjal was found to be the most effective treatment out of all the crops. The treatment decayed vegetables and fruits protected tomatoes and spinach from pest attacks.

Brinjal was the first crop subjected to our field studies. The chili and garlic combination was found to be the most effective as it promoted the establishment of the plant within 10 days and achieved remarkable growth and harvest within 69 days. The Brinjal treated with milk-based fertilizer produces a similar effect. Its flower stage is within 30 days and harvest occurred at 75 days. The fertilizer made from decayed vegetables and fruits is a variation of the conventional practices and presented growth duration to the harvesting stage at 70 days, shown in Table 1.

Tomato was the second crop tested with the different fertilizers. The tomato plant took 50 days to reach the harvesting stage when treated with chili and garlic. Milk-based fertilizer and decayed fruits and vegetables treatments allowed harvest at 46 days and 42 days, respectively, as shown in Table 2.

Spinach was the third crop to be tested. Spinach was harvested within 8 days with the decayed fruits and vegetables treatment. When milk-based fertilizer was used, harvest occurred within 9 days occurred.

CONCLUSION

The results observed using the three biofertilizer treatments showed that the fertilizer made from chili and garlic was the most effective when used on brinjal. The normal time duration for brinjal crop to be harvested is between 90 – 110 days. This was significantly reduced to 69 days by the addition of chili and garlic biofertilizer. The decayed vegetables and fruits were found to be more effective for both spinach and tomatoes. For tomatoes, the normal time duration of 50 – 60 days was reduced to 42 days due to the biofertilizer made from leafy vegetables and fruits. Spinach, usually harvested at 10 – 12 days, was harvested within 8 days through the application our biofertilizer made from leafy vegetables and fruits.

Biopesticides and biofertilizers can help farmers transition from toxic conventional chemical pesticides to an era of truly sustainable agriculture. Questions remain regarding the safety of biopesticide products for both human and ecosystem health. Green chemistry is about continuous improvement aimed to reduce or eliminate health hazards. We must encourage pest management solutions and regulations to continuously evolve. It must also be ensured that multi-disciplinary teams including green chemists, environmental specialists, and other researchers approach these novel innovations more holistically.

METHODS

Selection of the Plant Varieties: Brinjal, tomato, and spinach seeds were collected from the seed purchasing center Krishi Bhavan from Alleppey, Kerala, India. The seeds were initially dried in sun and then used for further experiments.

Growth of the Plant Varieties: *Spinach seeds were sprinkled with a small amount of water and then tied in muslin cloth until germination. The seeds were then planted in soil. B*rinjal and tomato seeds were planted in sandy soil.

Fertilizer Using Chili and Garlic: To make the fertilizer, ½ cup of chopped hot pepper and ½ cup of chopped garlic were combined in a blender to make a thick paste. The paste was added to ½ liter of warm water. The solution was kept for 24 hours and then strained to remove the solid vegetable pieces. The collected liquid was used as biofertilizer.

Fertilizer from Milk. To make the fertilizer, ½ liter of milk was poured into a vessel. Juice from ½ a lemon was added and mixed well and left for one day. The product was sieved into another vessel. The liquid obtained is a source of calcium, minerals and potassium. Water was added to ensure an equal ratio of milk to water.

Fertilizer from Milk: To make the fertilizer, ½ liter of milk was poured into a vessel. Juice from ½ a lemon was added and mixed well and left for one day. The product was sieved into another vessel. The liquid obtained is a source of calcium, minerals and potassium. Water was added to ensure an equal ratio of milk to water.

Fertilizer from Leafy Vegetables and Fruits: Grated coconut along was tied onto a cotton cloth. 500mL fermented curd was added into a mud pot and coconut water was added. The vegetables were tied with edges of cotton cloth and kept dipped for five days in the mud pot. Water was added to the mixture.

■ ACKNOWLEDGEMENTS

We would like to express our heartfelt gratitude to our biology teacher, Dr. Anoop, for constantly looking out for opportunities and platforms for us to present this project. We also express our deepest gratitude to Mrs. Jayakumari. They have both been a source of inspiration and encouragement to us. We would also like to thank Dr. Meena Shukoor, our beloved principal for constant motivation and providing us the freedom to perform the field trials on campus as well as permitting us to use our school laboratory.

■ REFERENCES

1. Reddy, C. A. & Saravana, R. S. (2013) Polymicrobial multi-functional approach for enhancement of crop productivity. *Advances in Applied Microbiology*, 82, 53-113. DOI: 10.1016/B978-0-12-407679-2.00003-X.

2. Gillespie, R. G. & New, T. R. (1998) Compatibility of conservation and pest management strategies. *Sixth Australasian Applied Entomological Research Conference*,195-208. ISBN: 1864990511.

3. Serrano, A., Pijlman, G. P., Vlak, J. M., Munoz, D., Williams, T., & Caballero, P. (2015) Identification of Spodoptera exigua nucleopolyhedro virus genes involved in pathogenicity and virulence. *Journal of Invertebrate Pathology,* 126, 43-50. DOI: 10.1016/j.jip.2015.01.008.

■ AUTHOR

A. K. Ashkay is a studious student who enjoys discovering new things. He is hardworking and always striving to reach new heights.

Dayajanaki possesses creative ideas on various experimental approaches which helped her create such a beautiful idea of biopesticides. She intends to pursue her higher studies in the engineering fields.

DOI: 10.36838/v2i2.8

Does Downside Risk of Financial Institutions Predict Future Economic Downturns and Housing Market Crashes?

Kaan M. Bali

McLean High School, 1633 Davidson Rd McLean, VA 22101, USA
kaanbali@gmail.com

ABSTRACT: This study's purpose is to estimate the systemic risk based on the downside risk of large financial institutions and test whether the systemic risk in the financial sector predicts future economic downturns and housing market crashes. The project finds that the newly proposed measure of systemic risk elevates with the start of the Great Recession in 2007, peaks during the heart of the financial crisis in 2008, and levels back out in the ensuing recovery period. The empirical analyses indicate that systemic risk in the financial sector predicts economic downturns seven months into the future, giving financial regulators time to react to possible financial turmoil in the United States. It also predicts housing prices over two years into the future again giving financial institutions and regulators time to respond to potential housing market crashes. Using this type of predictive power provided by the downside risk of financial institutions is imperative in monitoring the financial system. Although there will never be a perfect solution to the financial system's many complexities, more stringent regulation surrounding the risks taken by financial institutions can reduce the probability of another Great Recession.

KEYWORDS: Economics; Finance; Great Recession; Housing Market Crashes; Downside Risk.

INTRODUCTION

The global financial crisis, or the Great Recession, occurred in 2007-08 and is often considered the worst financial crisis since the Great Depression in the 1930s. The Great Recession was rooted in the bursting of the U.S. housing market.[1] The Great Recession was caused, in part, by greed and negligence of financial institutions in controlling risk. Prior to the credit boom in 2007 and the subprime mortgage industry meltdown that followed, financial institutions were bullish on lending and did not assess the magnitude risk faulty loans would have on their portfolios. When loans began to default at accelerated rates, the liquidity of banks was put into question and the financial sector plummeted, leading to the national and global recession.[2] Given the deep integration of investment banks within the global economy, it can be logically inferred that a strong correlation exists between economic distress and the systematic risk of the financial sector.

This study aims to determine whether systemic risk in the financial sector, as measured by average Value at Risk (VaR) of large financial institutions, can successfully predict future downturns in the economy and the housing market. Using large banks' VaR to predict recessions aids in determining policy and regulation to prevent future crises. For example, the financial crisis of 2008 could have been anticipated ahead of time by examining the banks' downside risk. Instead, the market was shocked by the fall of the investment bank and brokerage firm Bear Stearns in March 2008. Given the interconnectedness of large financial institutions, the fall of Bear Stearns sent waves of panic throughout the financial services industry. By September 2008, Lehman Brothers collapsed. While there were other warning signs, the fall of Bear Stearns and Lehman Brothers is often considered as sparking the financial crisis.

Leading up to the crisis, banking sector regulation was minimal. Banks were incredibly leveraged and had taken on a significant amount of risk. At the time, the idea that asset prices, such as real estate, would continue to rise for the indefinite future was widely accepted. Sub-prime candidates were granted risky mortgages using their homes as collateral. These mortgages were then bundled and sold in the market. When asset prices began falling, borrowers began defaulting on their loans. Many investment banks had large amounts of mortgage-backed securities in their portfolios resulting in devastating losses across the industry. Additionally, because there were very few capital and liquidity requirements financial institutions had difficulty funding their large losses. The overall economy is so tied to large investment banks that their failure can lead global markets into a recession.

As a result of the 2008 crisis, financial regulators enhanced the oversight of financial institutions, introduced new capital, leverage, and liquidity requirements, and affected significant changes in the derivatives and securitization market.[3] VaR was used to limit the amount of risk banks can take relative to the capital on hand. Therefore, it can be inferred that VaR of large financial institutions is a significant indicator of the economy's overall stability. This project will attempt to show the industry's average VaR can be used to forecast future downturns.

METHODS

The purpose of this research is to determine whether the systemic risk of the financial sector is a reliable measure to predict future economic and housing market conditions in the United States. The first step is to estimate the 1% Value at Risk (VaR) of the leading financial institutions in the United States. The 1% VaR is the 1st percentile of the daily return

DOI: 10.36838/v2i2.9

distribution, or approximately the average of the second and third lowest daily return observations out of 250 daily returns, calculated for each month using the past year's daily data. This estimation is calculated through Excel's percentile function. The VaR's are computed over an almost 20-year span for the prominent financial firms and averaged to determine the average worst-case downfall for the financial industry as a whole. The average 1% VaR is compared across financial institutions using the Chicago Fed National Activity Index (CFNAI) to determine the correlation between this project's proxy for systemic risk and the future economic states of the country. The final step is to investigate the performance of this project's systemic risk proxy in predicting housing market crashes. Overall, this research aims to prove that excessive risk-taking behavior of financial institutions during the Great Recession led to housing market crashes and macroeconomic downturns.

Data and Variables: *Largest Financial Institutions in the U.S.*: Several factors are considered in determining which financial institutions to use as the key ingredients of the systemic risk proxy, particularly the assets under management and the market capitalization of each financial institution. The assets under management and the market cap are valuable consideration points because the financial firms need to have a significant financial impact on the economy to develop an accurate proxy for the systemic risk of the entire financial system. Eleven U.S. financial institutions fell within the top 13 for both assets under management and market capitalization: JP Morgan Chase (JPM), Bank of America (BAC), Citigroup (C), Well Fargo (WFC), Goldman Sachs (GS), Morgan Stanley (MS), U.S. Bancorp (USB), PNC Financial Services (PNC), Capital One Financial (COF), the Bank of New York Mellon (BK), and Charles Schwab Corporation (SCHW). In using these 11 firms, it is assumed that the firms' market value and the size of the assets they hold and manage are influential enough to be used in the estimation of systemic risk of the financial sector. This assumption is tested later in the research.

Table 1 lists the top 15 U.S. bank holding companies ranked by total assets as of March 2019. Their market capitalizations are also shown in Table 1. The full list of bank holding companies with total assets greater than $10 billion are reported by the Federal Financial Institutions Examination Council (FFIEC).[4] To understand the weight of the chosen 11 financial firms in the entire financial sector, the ratio of the total assets of those financial firms to the total assets of all financial firms with assets greater than $10 billion was calculated. It was found that the top 11 firms form 65% of the major financial system. Similarly, the ratio of the total market capitalization of the top 11 financial firms to the total market capitalization of all financial firms was calculated and found that the 68% of total market capitalization for all financial firms comes from these 11 financial firms. Thus, one of the project's hypotheses is that the market value and asset size of these 11 financial firms are large enough to be qualified in the estimation of systemic risk of the financial sector. TD Bank was not used in the project's estimation of the financial industry's systematic risk because its data starts in 2015 and therefore does not have enough time-series observations required for the main tests. TIAA was also not used because the FFIEC does not report its market capitalization.

Table 1. List of the 15 Largest Banks in the United States.

Rank	Name	Headquarters	Total Assets (in billions USD)	Market capitalization (in billions USD)
1	JPMorgan Chase	New York, NY	2,737	327
2	Bank of America	Charlotte, NC	2,737	302
3	Citigroup	New York, NY	1,958	174
4	Wells Fargo	San Francisco, CA	1,887	273
5	Goldman Sachs	New York, NY	925	87
6	Morgan Stanley	New York, NY	875	72
7	U.S. Bancorp	Minneapolis, MN	475	83
8	T.D. Bank	Cherry Hill, NJ	384	108
9	PNC Financial Services	Pittsburgh, PA	392	67
10	Capital One Financial	McLean, VA	373	47
11	The Bank of New York Mellon	New York, NY	346	53
12	TIAA	New York, NY	307	N/A
13	Charles Schwab Corp.	San Francisco, CA	282	39
14	HSBC Bank USA	New York, NY	279	N/A
15	State Street Corp.	Boston, MA	228	31

Measuring Economic Downturns: Following the original work of the Federal Reserve Bank of Chicago, this project uses the Chicago Fed National Activity Index (CFNAI) to measure economic downturns. The CFNAI is the weighted average of 85 monthly indicators of national economic activity and is used to predict U.S. economic growth.[5] It is computed monthly and is, historically, very accurate in predicting business cycle fluctuations. Its economic indicators have been drawn from four major categories of data:

- Production and Income (23 indicators)
- Employment, Unemployment, and Hours (24 indicators)
- Personal Consumption and Housing (15 indicators)
- Sales, Orders, and Inventories (23 indicators

The CFNAI is intended to provide a real-time statistic of economic stability in the U·S· market· An index of zero is interpreted as the economy growing at a typical expectancy. A positive value represents above-average economic growth and a negative value indicates a below-average economic growth· If its results are –0.7 or below, it correctly predicts a recession with 86% accuracy.[6]

The CFNAI's proves its importance through its performance in identifying economic downturns compared to the National Bureau of Economic Research (NBER). Figure 1 displays the monthly time-series plot of the CFNAI for April 2000 – June 2019, the sample period used in this project's main empirical analyses. The index is negative and larger in absolute magnitude during bad states of the economy. As shown in Figure 1,

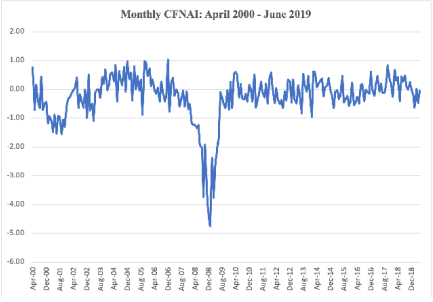


Figure 1. Chicago FED National Activity Index (CFNAI) from April 2000 – June 2019.

DOI: 10.36838/v2i2.9

the CFNAI level reached historical lows during the Great Recession – 4.38 in December 2008 and -4.75 in January 2009.

Over the past 35 years, the CFNAI level has explained almost 40% of the variation in the next quarter's economic output (measured by gross domestic product).[6] Figure 2 compares the end-of-quarter CFNAI levels with the quarterly real GDP growth rate over the past 51 years. The correlation between the end-of-quarter CFNAI and the quarterly real GDP growth rate is positive, 72%, and highly significant for the period 1967-2018. In Figure 2, the blue line representing the end-of-quarter CFNAI closely mirrors the red line representing the quarterly real GDP growth. Additionally, the CFNAI moves before the real GDP does. With this information, it can be concluded that the CNFAI is a strong indicator of future economic output because CFNAI and GDP encompass many of the same factors.

Figure 2. Comparing CFNAI with Real GDP Growth. Presents the time-series plots of the end-of-quarter CFNAI and the quarterly real GPD growth for the sample period of 1967:Q1 – 2018:Q4

Housing Market Index: For this project, the S&P/Case-Shiller U.S. National Home Price Index is used to track the value of single-family housing within the United States. The Case-Shiller index is a widely used indicator of the U.S. housing market and the broader economy.[8] Figure 3 shows that the Case-Shiller U.S. National Home Price Index value was 102.54 in April 2000 and peaked at 184.62 in July 2006, indicating that the house prices increased by 80% over the six-year period. The index almost monotonically declined to 133.99 in February 2012, showing a 27% decline during those six years. It took until February 2017 for the index to regain its earlier peak.

Value-at-Risk of Financial Institutions: A primary tool for financial risk assessment is the Value at Risk (VaR), which is defined as the maximum loss expected on a portfolio of assets over a certain holding period at a given confidence level (probability).[9,10] For this project, the 1% VaR of financial institutions is estimated using daily equity returns of the largest 11 financial firms for the longest common sample period from May 5, 1999 – June 28, 2019. The one year of daily returns of each institution from May 5, 1999 – April 28, 2000 was used (total of 250 daily returns) to generate the first 1% VaR estimate. This estimate corresponds to the 1st percentile of the daily return distribution, or approximately the average of the second and third lowest daily return observations from May 5, 1999 – April 28, 2000. This is used as a measure of April 2000's 1% VaR. Then, the sample is rolled one month forward and each financial institution's 1% VaR was re-estimated using one year of daily returns from June 1, 1999 – May 31, 2000 to generate the second 1% VaR estimate. The monthly rolling VaR estimation procedure is repeated for each month until the sample is exhausted in June 28, 2019.

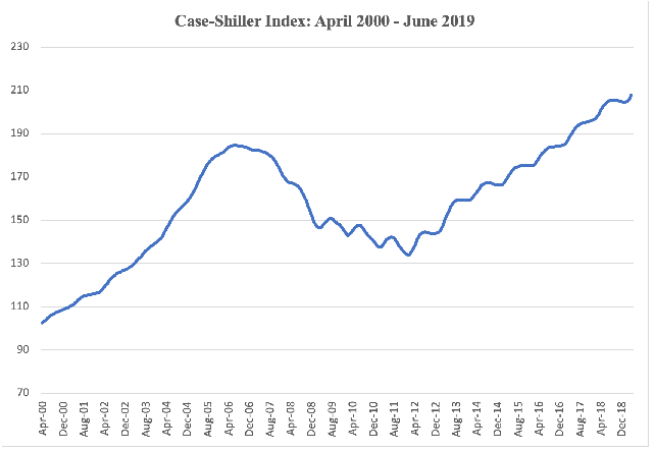


Figure 3. Monthly time-series plot of the Case-Shiller U.S. National Home Price Index for the sample period of April 2000 – June 2019

The systemic risk of the financial sector is defined as the average of the 1% VaR estimates of the largest 11 financial firms. Note that the original 1% VaR measures are negative since they are obtained from the left tail of the return distribution. The original systemic measure is multiplied by -1 for convenience of interpretation; larger positive values of the systemic risk measure indicate higher catastrophic risk in the financial sector.

Figure 4 displays the monthly time-series plot of the systemic risk of the financial sector (SYSRISK) for the sample period April 2000 – June 2019. As the graph shows, the systemic risk of financial institutions elevated with the start of the Great Recession and peaked during the heart of the financial crisis.

Figure 4. Monthly time-series plot of the systemic risk in the financial sector (SYSRISK) for the sample period of April 2000 – June 2019.

It leveled back out in the ensuing recovery period with a slight increase in 2011.

RESULTS AND DISCUSSION

The following monthly time-series regressions are run to test whether the SYSRISK predicts future economic turndowns and housing market crashes:

$$(1)\ CFNAI_{t+1} = \alpha + \beta \cdot SYSRISK_t + \varepsilon_{t+1}$$

$$(2)\ HOUSE_{t+1} = \phi + \lambda \cdot SYSRISK_t + e_{t+1}$$

where SYSRISK is the systemic risk of the financial sector in month t, CFNAI... is the Chicago FED National Activity Index in month t+1, and HOUSE... is the Case-Shiller U.S. National Home Price Index in month t+1. The slope coefficients (β, λ) in regression equations (1) and (2) will determine whether higher systemic risk in the financial sector predicts lower economic growth and lower home prices in the U.S. The project tests whether the slope coefficients (β, λ) are negative and statistically significant.

Regressing the one-month-ahead CFNAI on SYSRISK generated the following regression output:

$CFNAI_{t+1}$	*Coefficients*	*Std. Err.*	*t-stat*	*p-value*
Intercept	0.46	0.09	5.29	0.0000
$SYSRISK_t$	-13.91	1.41	-9.85	0.0000

As postulated, the slope coefficient (β) on SYSRISK is estimated to be negative, -13.91, and highly statistically significant with a *t-statistic of* -13.91 (or a p-value of zero). The significantly negative slope coefficient on SYSRISK indicates that an increase in the systemic risk of the financial sector predicts a significant decline in the CFNAI or predicts economic downturns. The economic magnitude of the slope coefficient can be interpreted as follows: Given that the standard deviation of SYSRISK is 3.29% in the sample, a 2-standard deviation increase in SYSRISK leads to a decline of 0.92 in the CFNAI, which is economically large given that the sample minimum, mean, and maximum values of the CFNAI are, respectively, -4.75, -0.26, and +1.02 over the estimation period of April 2000 – June 2019. This concludes that the systemic risk of the financial sector successfully predicts future economic downturns.

$HOUSE_{t+1}$	*Coefficients*	*Std. Err.*	*t-stat*	*p-value*
Intercept	175.04	2.98	58.74	0.0000
$SYSRISK_t$	-326.01	47.87	-6.81	0.0000

Regressing the one-month-ahead Case-Shiller index on SYSRISK generated the following regression output:

The slope coefficient (λ) on SYSRISK is estimated to be negative, -326.01, and highly statistically significant with a t-statistic of -6.81 (or a p-value of zero). The significantly negative slope coefficient on SYSRISK indicates that an increase in the systemic risk of the financial sector predicts a significant decline in U.S. house prices. The economic magnitude of the slope coefficient can be interpreted as follows: Given that the standard deviation of SYSRISK is 3.29% in the sample, a 2-standard deviation increase in SYSRISK leads to a decline of 21.45 in the Case-Shiller index. This is economically large given that the sample minimum, mean, and maximum values of the Case-Shiller index are respectively 102.54, 157.59, and 207.97 over the estimation period of April 2000 – June 2019. This concludes that the systemic risk of the financial sector accurately predicts housing market crashes.

This project also investigates if the systemic risk of the financial sector has long-term predictive power for economic downturns and housing market crashes or if the predictability evidence provided thus far is just a one-month affair. The following longer-term predictability regressions are run:

$$(3)\ CFNAI_{t+k} = \alpha + \beta \cdot SYSRISK_t + \varepsilon_{t+k}$$

$$(4)\ HOUSE_{t+k} = \phi + \lambda \cdot SYSRISK_t + e_{t+k}$$

where k denotes 2-month-ahead to 12-month-ahead Chicago FED National Activity Index in regression equation (3), and k denotes 12-month-ahead to 36-month-ahead Case-Shiller index in regression equation (4).

Table 2 presents the slope coefficients (β), standard errors, t-statistics, and p-values from the time-series regressions of the two-month-ahead (*CFNAI*t+1) to 12-month-ahead (*CFNAI*t+12) Chicago FED National Activity Index on SYSRISK. A notable point in Table 2 is that SYSRISK predicts the *CFNAI* seven months into the future because the slope coefficients are statistically significant at least at the 5% level (|t|>1.96) for predicting *CFNAI*t+2 to *CFNAI*t+7. The predictive power of SYSRISK diminishes starting *CFNAI*t+8 with an insignificant slope coefficient of -2.78 with a t-statistic of -1.64.

Table 2. Predictive Power of SYSRISK for Future Economic Downturns

	Coefficients	*Std. Err.*	*t-stat*	*p-value*
$CFNAI_{t+2}$	-12.19	1.48	-8.22	0.0000
$CFNAI_{t+3}$	-10.19	1.55	-6.57	0.0000
$CFNAI_{t+4}$	-8.20	1.61	-5.10	0.0000
$CFNAI_{t+5}$	-6.53	1.64	-3.97	0.0001
$CFNAI_{t+6}$	-5.19	1.67	-3.12	0.0021
$CFNAI_{t+7}$	-3.84	1.69	-2.28	0.0236
$CFNAI_{t+8}$	-2.78	1.70	-1.64	0.1028
$CFNAI_{t+9}$	-1.84	1.71	-1.08	0.2827
$CFNAI_{t+10}$	-1.06	1.71	-0.62	0.5360
$CFNAI_{t+11}$	-0.37	1.72	-0.22	0.8278
$CFNAI_{t+12}$	0.15	1.72	0.09	0.9288

Table 2 shows that after month t+7, the significance of SYSRISK completely disappears. This concludes that the predictive power of systemic risk in the financial sector is not just a one-month phenomenon. SYSRISK predicts economic downturns seven months into the future, giving financial regulators enough time to react to possible financial turmoil in the United States.

Table 3 presents the slope coefficients (λ), standard errors, t-statistics, and p-values from the time-series regressions of the 12-month-ahead (*HOUSE*t+12) to 36-month-ahead (*HOUSE*t+36) Case-Shiller index on SYSRISK. As shown in Table 3, SYSRISK predicts the housing market index 29 months into the future because the slope coefficients are statistically significant at least at the 5% level (|t|>1.96) for predicting (*HOUSE*t+2) to (*HOUSE*t+29). The predictive power of SYSRISK diminishes starting with (*HOUSE*t+30) with an in-

DOI: 10.36838/v2i2.9

significant slope coefficient of -78.32 with a *t-statistic* of -1.83. After month *t*+30, the significance of SYSRISK completely disappears. This concludes that the predictive power of systemic risk in the financial sector is works to predict housing market crashes almost two and a half years into the future, giving financial regulators long enough to respond to housing market crashes.

Table 3. Predictive Power of SYSRISK for Housing Market Crashes.

	Coefficients	*Std. Err.*	*t-stat*	*p-value*
$HOUSE_{t+12}$	-215.18	38.60	-5.57	0.0000
$HOUSE_{t+24}$	-136.56	41.49	-3.29	0.0012
$HOUSE_{t+25}$	-127.62	41.73	-3.06	0.0025
$HOUSE_{t+26}$	-119.08	41.92	-2.84	0.0050
$HOUSE_{t+27}$	-110.41	42.11	-2.62	0.0094
$HOUSE_{t+28}$	-100.65	42.30	-2.38	0.0183
$HOUSE_{t+29}$	-89.47	42.55	-2.10	0.0367
$HOUSE_{t+30}$	-78.32	42.71	-1.83	0.0682
$HOUSE_{t+31}$	-65.18	42.88	-1.52	0.1301
$HOUSE_{t+32}$	-51.33	43.02	-1.19	0.2343
$HOUSE_{t+33}$	-36.17	43.15	-0.84	0.4029
$HOUSE_{t+34}$	-20.38	43.25	-0.47	0.6380
$HOUSE_{t+35}$	-4.97	43.32	-0.11	0.9088
$HOUSE_{t+36}$	9.69	43.40	0.22	0.8235

CONCLUSION

This experiment tests the correlation between financial institutions' Value-at-Risk (VaR) and overall macroeconomic activity and the housing market. The sample studied included the largest 11 financial institutions that are quantitatively and qualitatively similar and were strongly impacted by the Great Recession. By evaluating the average downside risk of these financial institutions over time, the project shows how systemic risk in the financial sector increases during recessionary periods. Not surprisingly, the regression analyses yield a statistically significant relationship between average VaR of financial institutions and the CFNAI and the Case-Shiller index. Moving forward, VaR can be viewed as an effective way to predict recessionary downfall; in the hands of regulators and other government officials this information can be used to strategically counteract the incidence of future recessionary periods.

This paper effectively proves that excessive risk-taking behavior of financial institutions during the financial crisis of 2007-2008 led to economic downturns and housing market crashes. Given these findings, one can say that financial sector performance is an accurate indicator of macroeconomic health. It is important that financial regulators place greater focus on this systemic risk measurement in the future in order to hedge against future economic downturns. By using the relatively simple measure of systemic risk, institutions have the heightened ability to anticipate economic downturns, eliminating economic surprises and the risk of future financial crises. These findings indicate that average VaR of financial institutions can be used as a reliable proxy for systemic risk in the financial sector and can be used to predict future economic downturns and trends in the housing market.

ACKNOWLEDGEMENTS

I am grateful to my high school teachers, especially my AP Economics and AP Statistics teachers, Ryan Abrams and Emily Jaffa, for their insightful and constructive feedback.

REFERENCES

1. Davis, B. (2009, April 22). What's a global recession? T*he Wall Street Journal.* Retrieved from https://blogs.wsj.com/economics/2009/04/22/whats-a-global-recession/

2. International Monetary Fund. (2009, April 22). World economic outlook — *April 2009: Crisis and recovery*. Retrieved from https://www.imf.org/en/Publications/WEO/Issues/2016/12/31/World-Economic-Outlook-April-2009-Crisis-and-Recovery-22575

3. Morrison & Foerster (2010). *The Dodd-Frank act*: *A cheat sheet*. Retrieved from http://media.mofo.com/files/uploads/Images/SummaryDodFrankAct.pdf

4. List of largest banks in the United States. (2020, April 27). *In Wikipedia*. https://en.wikipedia.org/wiki/List_of_largest_banks_in_the_United_States

5. Brave, S. (2009, Nov). The Chicago fed national activity index and business cycles. Chicago Fed Letter, 268

6. Brave, S. (2020, April 20). *CFNAI historical (real-time) data*. Federal Reserve Bank of Chicago. https://www.chicagofed.org/research/data/cfnai/historical-data

7. Koesterich, R. (2015, April 29). *Why the Chicago fed national activity index matters.* Market Realist. Retrieved from https://marketrealist.com/2015/04/chicago-fed-national-activity-index-matters/

8. Shiller, R. (n.d.). Online data. http://www.econ.yale.edu/~shiller/data.htm

9. Jorion, P. (2006) Value-at-Risk: The new benchmark for controlling market risk. Chicago, IL: McGraw-Hill Education.

10. Ryback, W. (2011). *Case study on Bear Stearns*. Retrieved from http://siteresources.worldbank.org/FINANCIALSECTOR/Resources/02BearStearnsCaseStudy.pdf

AUTHOR

Kaan Bali is a junior at McLean High School in Virginia. He has taken AP Economics and AP Statistics. Exposure to popular media about business has led him to take special interest in economics and finance. This is the second scientific paper Bali has written.

Dysfunctions in Alzheimer's Dementia Hallmarks with Pyrethroids and Piperonyl Butoxide Pesticide Synergy

Marguerite Li

Jericho Senior High School,99 Cedar Swamp Rd., Jericho, NY, 11753, USA
marguerite.li@jerichoapps.org

ABSTRACT: There are over 5 million Americans with Alzheimer's dementia (AD). A positive correlation between chronic pesticide exposure and AD prevalence. This study's goal is to elucidate the toxicity of synergistic pesticides pyrethroid (PY) and piperonyl butoxide (PBO) on the induction of cytotoxicity, tau phosphorylation, and amyloid-ß (Aß). An MTT assay measured cell viability and cytotoxicity. A lactate dehydrogenase (LDH) assay determined extracellular LDH and an ELISA determined Aß and tau protein expression. Exposure to PY and PBO synergistically reduced cell viability greater than PY and PBO individually. Exposure to synergistic PY and PBO increased LDH cytotoxicity in a time-dependent manner. PY and PBO synergistically increased Aβ and tau protein expression, indicating low concentrations and long-term treatment were capable of inducing Aβ and tau protein expression. PY and PBO synergistically induced multiple hallmarks of AD, suggesting chronic low concentration exposure to pesticides is likely correlated with AD development. Future investigations should focus on the pathophysiological mechanisms of PY and PBO's induction on AD hallmark protein expression. Researching these effects will improve the safety and health of pesticide users.

KEYWORDS: Neurodegenerative Diseases; Alzheimer's Disease; Pyrethroid; Piperonyl Butoxide; Amyloid-ß.

INTRODUCTION

There are currently 5.8 million Americans with Alzheimer's dementia (AD). By 2050, over 14 million Americans will be affected by AD, according to the Alzheimer's Association.[1] It is estimated that 95% of Alzheimer's patients suffer from late-onset Alzheimer's dementia (LOAD). LOAD is characterized by developing AD at age 65 or older. While LOAD is the most common form of AD, there are no known causes.[2] Risk factors for LOAD include both genetic and environmental factors; the main genetic factor identified is the Apolipoprotein E gene.[3] This study aims to research pesticides as a main environmental factor in the development of AD.

Annually, 5.6 billion pounds of pesticides are used worldwide, with 1 billion pounds used in the United States alone.[4] One popular agricultural pesticide is pyrethroid (PY), a synthetic neurotoxic insecticide developed from pyrethrin found in chrysanthemums.[5,6] PYs are considered nontoxic to large mammals but recent studies have shown it is high toxic to mice and human cell lines.[7,8] Meta-analysis data has demonstrated a positive association between chronic pesticide exposure and AD prevalence.[9] PY pesticides are known to induce oxidative stress, acetylcholinesterase, apoptosis, and mitochondrial dysfunction, all which mirror hallmarks of AD. Thus, pyrethroid presence in neurons is questioned as a risk factor of AD.[10,11] PY disrupts proper voltage gated sodium channel (VGSC) kinetics and prolongs depolarization of the channel.[12] PY can also act upon a voltage gated calcium channel to change its kinetics channel and cause an influx of calcium ions.[13] This imbalance of calcium ions in mitochondria can cause reactive oxidative species and cell death associated with AD neurological dysfunctions. Since PY can be metabolized quickly by esterases and cytochrome p450, PY is applied with piperonyl butoxide (PBO), a chemical synergist. PBO inhibits cytochrome p450 function to increase PY toxicity.5 In this study, a combination of PY and PBO were used synergistically to mimic the common application of PY pesticides.[5,14]

The purpose of this study is to examine the cytotoxicity of PY and PBO on dysfunctions to AD hallmark proteins. Aß and tau phosphorylation were the two main hallmarks studied. Previous studies have shown both Aß and tau-phosphorylation to be associated with the glycogen synthase kinase pathway can be activated with PY.[15]

RESULTS AND DISCUSSION

In eight replicative experiments, SK-N-SH human neuroblastoma cells were treated with increasing concentrations of PY, PBO, or in combination at 0.1 µM, 1.0µM, and 10µM for

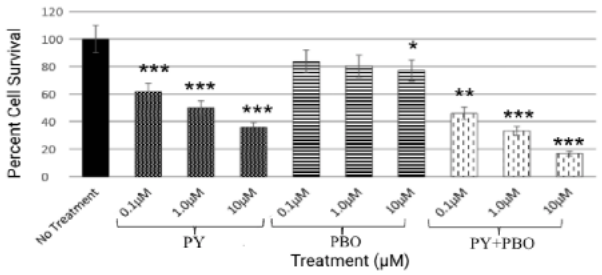


Figure 1. The effect of PY and PBO on human neuroblastoma cell survival. MTT assay was performed 72 hours after initial treatment of PY, PBO, or a combination of PY and PBO. * = p<0.05; ** = p<0.005; *** = p<0.001. Data is presented as mean +_ SEM from eight samples.

DOI: 10.36838/v2i2.10

for 72 hours prior to performing an MTT assay to determine cell viability.

The synergistic relationship between PY and PBO was present, indicating their synergistic ability functions in human neuroblastoma cells at low concentrations of 0.1μM treatment of PY and PBO. Cell survival percent at 0.1 μM of PY+PBO was 35.39%; individual treatments of PY and PBO had cell survival rates of 47.80% and 79.16%, respectively (Figure 1; $p<0.05$). Even at low concentrations of 0.1μM, PY+PBO still produced significant results, indicating its synergistic neurotoxicity could have implications in AD neurological hallmarks. All the concentrations applied to the cells demonstrated that PBO+PY reduced cell survival compared to individual treatments. The addition of PBO to PY increased the cytotoxic ability of the two chemicals thus causing less cell viability.

LDH Assay: In eight replicate experiments, SK-N-SH human neuroblastoma cells were treated with increasing concentrations of PY, PBO, or in combination at 0.1 μM, 1.0μM, and 10μM for 72 hours before an LDH assay was performed to determine cytotoxicity.

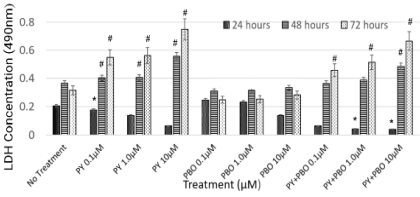


Figure 2. Cytotoxicity of PY and PBO in a Time Dependent Manner. Human neuroblastoma cells were treated with various concentrations of PY and PBO. Statistical analysis was performed in comparison to no treatment groups. * = $p<0.05$; # = $p<0.005$. Data is presented as mean ± SEM from eight samples

A time- and dose-dependent relationship was established between PY and PY+PBO. The synergistic ability of PY+PBO is not shown by the LDH assay, but rather in the PY which produced significantly greater cytotoxicity than PY+PBO. This suggests that PBO may have a neuroprotective role in mitochondrial related pathways. This is supported by the individual PBO treatment at 48 and 72 hours where the LDH concentrations released by the neuroblastoma cells was less than the overall LDH released by neuroblastoma cells receiving no treatment (Fig 2; $p<0.05$). The lack of PBO effect on LDH concentrations may be due to PBO's unique characteristic to act as both an inhibitor and inducer of cytochrome p450. The result of this is that PBO inhibits cytochrome p450 function thus maintaining the LDH concentration. Further experiments into PBO reactions is needed to determine the significant role of PBO in combination with PY. Nevertheless, individual PY produced significant cytotoxicity at low concentrations of 0.1μM, indicating its toxicity has great consequences in mitochondrial function associated with AD mitochondrial dysfunction.

Neuronal Death: A Trypan Blue Exclusion Assay was performed 72 post treatment of human neuroblastoma cells with 0.1μM, 1.0μM, and 10.0μM of PY and/or PBO to determine neuronal death count.

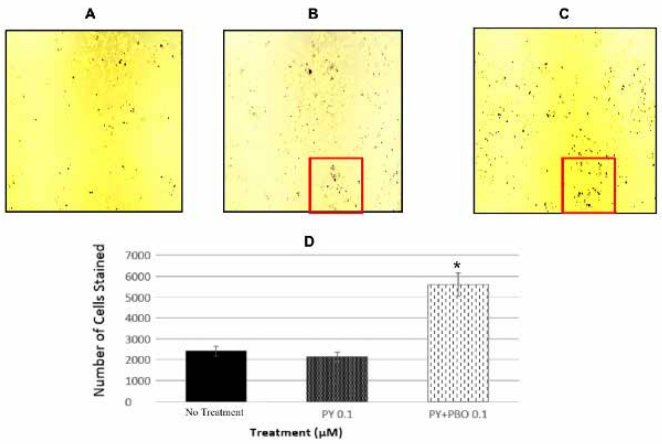


Figure 3. Neuronal death caused by long-term PY and PBO treatment. A) Control. B) PY; 0.1μM. C) PY and PBO; 0.1μM. D) Quantified human neuroblastoma cells 72 hours post Trypan Blue Exclusion Assay to determine cellular death via staining with Trypan Blue. * = $p<0.005$. Data is presented as mean ± SEM from twelve samples.

Neuronal death was greater with PY+PBO than the individual treatment of PY. The synergy of PY+PBO is cytotoxic and most likely causes oxidative stress to produce cell death. This could indicate PY+PBO induces apoptosis via oxidative stress. The neuroblastoma cells without PY and/or PBO treatment had cell death that could be from the naturally occurring death of cells. Additionally, neuronal death (a less specific hallmark of AD) caused by PY+PBO could be associated with AD cellular hallmark development. PBO was not tested in this experiment because of previous results that determined PBO has no significant effects on cell viability at 0.1μM and 1.0 μM.

Aß ELISA: An ELISA was performed 72 hours after human neuroblastoma cell treatment with 0.1μM, 1.0μM, or 10μM of PY and/or PBO to determine dose-dependent Aß concentration with the synergy of PY+ PBO and Aß induction level on neuroblastoma cells.

Individual treatment with 0.1μM and 1.0μM PBO induced little to no Aß expression. However, PY individual treatment induced a 19.89% and 36.72% increase in Aß expression, respectively. The synergistic effect of PBO and PY significantly increased Aß expression to 30.47% and 42.31% at 0.1μM and 1μM, respectively, supporting the hypothesis that pesticides are able to induce Aß production (Fig. 4; $p<0.05$). Low concentration and chronic synergistic pyrethroid exposure are demonstrated as a contributing factor to the development of AD hallmarks. Furthermore, the increase in Aß expression could be caused by increased oxidative stress that PY and PBO induce, or PY+PBO stimulation of key enzymes in Aß production such as ß-secretase.

TAU ELISA: An ELISA was performed 72 hours after human neuroblastoma cell treatment with 0.1μM, 1.0μM, or 10μM of PY and/or PBO to determine dose-dependent tau protein concentration in treatments with a combination PY and PBO on neuroblastoma cells.

Individual treatment with 1.0μM PY and PBO resulted in an 18.62% and 10.47% respective increase compared to the no treatment group, indicating that long-term low concentration

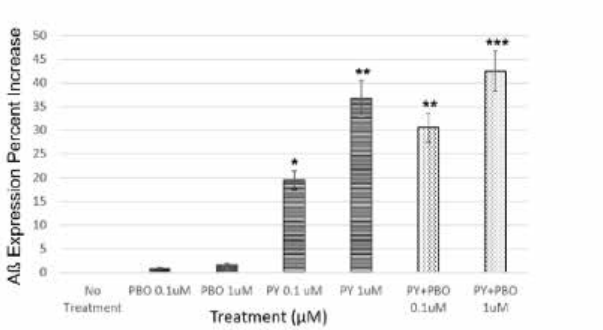


Figure 4. ELISA of PY and PBO on Aß protein concentration. * = p< 0.05; ** = p<.005; *** = p<0.001. Data is presented as mean +− SEM from four repeated experiments.

exposure to neuroblastoma cells elevates tau protein levels. PBO's influence on tau protein expression is significant as it suggests PBO cytotoxicity could cause tau expression. Since PBO does not significantly change cell viability but does significantly increase tau protein levels it is possible that, like tau protein in AD, PBO does not result in immediate neuron death but a gradual decline in neurons. The synergistic ability of PY and PBO caused significantly more tau protein expression compared to both the control group and the cells with treated individually with PY or PBO. At 1.0µM, tau protein expression increased by 27.81%, suggesting that at chronic low concentration exposure to PY+ PBO causes dysfunctions in the tau protein pathology (Figure 5; p<0.05)

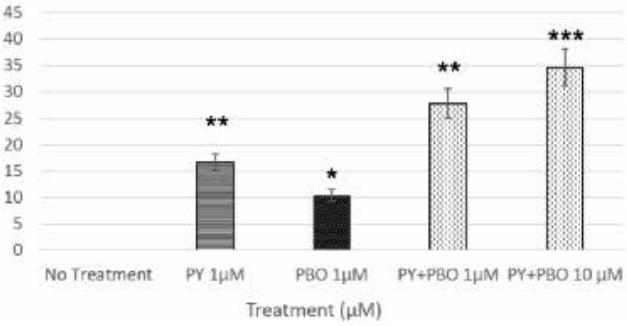


Figure 5. ELISA of PY and PBO on tau protein expression. Human neuroblastoma cells were treated with various concentration of PY and PBO for 72 hours before an ELISA was performed to determine tau phosphorylated protein concentration. * = p< 0.05; ** = p<.005; *** = p<0.001. Data is presented as mean ± SEM from eight samples.

CONCLUSION

PY and PBO both demonstrated consistent cytotoxic synergistic effects on human neuroblastoma cells. Their combines neurotoxic effects indicate their potential as contributing factors to AD development. Aß protein expression, tau protein expression, and cellular toxicity all increased after a 72-hour treatment of PY and PBO. This reveals PY and PBO's potential influence on AD neurological hallmark development as well as the exposure to low concentrations of 0.1µM of PY+PBO can cause cytotoxicity associated with neuronal death and decline. The increase in Aß and tau protein demonstrates that PBO and PY have the potential to contribute to the pathophysiological formation of AD. Aß increase can be induced by oxidative stress and increases in reactive oxidative species like superoxides or hydrogen peroxides.[16] Pyrethroid exposure to mitochondria reduces superoxide dismutase (SOD), a key antioxidant enzyme that reduces superoxides into oxygens. A reduction in SOD leads to a buildup of unmetabolized toxic superoxide molecules that increase oxidative stress.[17] Cytotoxicity and oxidative stress could result from the prolonged depolarization period caused by PY targeting VGSC. Cytotoxicity in low dosages of these pesticides may also lead to neuronal loss which can be contributed to other neurodegenerative diseases. Calculating the average amount of pesticides that an individual is exposed to is difficult, but the concentrations would be similar to the low concentrations used in the experiments.

METHODS

Chemical and Reagent Preparation

Pyrethroid extract, 2-(2-Butoxyethoxy) ethyl (6-propylpiperonyl) ether (PBO), and 3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide (MTT) were purchased from Sigma-Aldrich. Pyrethroid solutions were created using serial dilutions of 10µM, 1.0µM, and 0.1µM using ethanol concentrations. PBO solutions were created using serial dilutions of 10µM, 1.0µM, and 0.1µM using Human Tau ELISA kit and Amyloid-ß 42 Human ELISA Kit purchased from Invitrogen. All chemicals purchased were of analytical or technical grade

Cell Culture: SK-N-SH HTB-11 human neuroblastoma cells were obtained from the American Type Culture Collection. Neuroblastoma cells were cultured in a humidified 5% CO2 atmosphere at 37 ˚C in MEM media.

MTT Assay: An MTT assay was conducted to measure the induction of PY and PBO on SK-N-SH human neuroblastoma cell. The cell cultures were treated with 0.1mM, 1.0 µM, or 10.0 µM of either individual PY or PBO or synergistically treated with PY and PBO. Samples were incubated for 72 hours to mimic the chronic low concentration pesticide exposure in agricultural settings. After incubation, excess MTT was removed and 50µL of DMSO was pipetted into each well to solubilize the cells. Absorbance was measured at 490 nm using a BioRad iMark™ Microplate Absorbance Reader.

Lactate Dehydrogenase Assay: Lactate dehydrogenase (LDH) measured LDH concentration based on a coupled enzyme reaction. LDH assay kit was purchased from ThermoFisher Scientific. LDH catalyzes the reaction of lactate to pyruvate by reducing NAD+ to NADH. NADH is used by diaphorase to reduce tetrazolium salt to a red formazan product which can be measured spectrophotomically and is directly proportional to LDH concentration. Cells were treated with 0.1µM, 1.0µM, and 10.0µM, of PY, PBO, or PY+PBO for 72 hours before LDH assay was performed. Red formazan product was measured with a BioRad iMark™ Microplate Absorbance Reader at 490nm and 680nm.

Trypan Blue Exclusion Assay: Cell viability and cellular death were measured using 0.4% Trypan Blue Solution. Af-

DOI: 10.36838/v2i2.10

ter PY and/or PBO treatment for 72 hours, SK-N-SH human neuroblastoma cells were treated with Trypan Blue. Viable cells have membranes that are impermeable to the Trypan Blue. Images of each well were taken immediately after straining using a light microscope. ICTN plugin within ImageJ was used to calculate the number of viable cells.

Enzyme-Linked Immunosorbent Assay (ELISA): Aß protein expression was measured using ELISA conducted according to the manufacturer's manual. After ELISA was performed, the Bio-Rad iMark Microplate Absorbance Reader was set to 450 nm. Tau protein expression was also measured using an ELISA conducted according to the manual. All ELISA procedures were performed 72 hours after human neuron models were treated with 0.1μM, 1.0μM, and 10.0μM, of PY, PBO, or PY+PBO.

Data Analysis: One-way ANOVA was used to determine statistical significance where all experiments were repeated three times with $p<0.05$ as the statistical significance. The statistical significance of the data was in comparison with cells that received no treatment.

■ ACKNOWLEDGEMENTS

I want to thank my mentor, Dr. Wei Zhu, for his support and guidance during this research process, especially his continued encouragement to find opportunities to present my research. I would also like to thank my teacher, Dr. Serena McCalla, for being a constant source of inspiration and motivation.

■ REFERENCES

1. Alzheimer's Association. (n.d.). *Facts and Figures*. https://alz.org/alzheimers-dementia/facts-figures
2. Gauthier, E., Fortier, I., Courchesne, F., Pepin, P., Mortimer, J., & Gauvreau, D. (2001). Environmental pesticide exposure as a risk factor for Alzheimer's disease: *A case-control study. Environmental Research,* 86(1), 37-45. https://doi.org/10.1006/enrs.2001.4254
3. Sabarwal, A., Kumar, K., & Singh, R. P. (2018). Hazardous effects of chemical pesticides on human health – Cancer and other associated disorders. *Environmental Toxicology and Pharmacology,* 63, 103-114. https://doi.org/10.1016/j.etap.2018.08.018
4. Alavanja, M. C. R. (2009). Pesticides use and exposure extensive worldwide. *Reviews on Environmental Health,* 24(4), 303-309. https://doi.org/10.1515/reveh.2009.24.4.303
5. Tozzi, A. (1999). A brief history of the development of piperonyl butoxide as an insecticide synergist. In D. G. Jones (Ed.), In D. G. Jones (Ed.), *Piperonyl Butoxide* (pp. 1-5). Elsevier
6. El-Demerdash, F. M. (2011). Lipid peroxidation, oxidative stress and acetylcholinesterase in rat brain exposed to organophosphate and pyrethroid insecticides. *Food and Chemical Toxicology,* 49(6), 1346-1352. https://doi.org/10.1016/j.fct.2011.03.018
7. Vardavas, A. I., Stivaktakis, P. D., Tzatzarakis, M. N., Fragkiadaki, P., Vasilaki, F., Tzardi, M., Datseri, G., Tsiaoussis, J., Alegakis, A. K., Tsitsimpikou, C., Rakitskii, V. N., Carvalho, F., & Tsatsakis, A. M. (2016). Long-term exposure to cypermethrin and piperonyl butoxide cause liver and kidney inflammation and induce genotoxicity in New Zea land white male rabbits. *Food and Chemical Toxicology*, 94, 250–259. https://doi.org/10.1016/j.fct.2016.06.016
8. Chen, M., Du, Y., Zhu, G., Takamatsu, G., Ihara, M., Matsuda, K., Zhorov, B. S., & Dong, K. (2018). Action of six pyrethrins purified from the botanical insecticide pyrethrum on cockroach sodium channels expressed in Xenopus oocytes. *Pesticide Biochemistry and Physiology,* 151, 82-89. https://doi.org/j.pestbp.2018.05.002
9. Yan, D., Zhang, Y., Liu, L., & Yan, H. (2016). Pesticide exposure and risk of Alzheimer's disease: *A systematic review and meta-analysis. Scientific Reports*, 6, 32222. https://doi.org/10.1038/srep32222
10. García-Ayllón, M., Small, D. H., Avila, J., & Sáez-Valero, J. (2011). Revisiting the role of acetylcholinesterase in Alzheimer's disease: Cross-talk with P-tau and B-amyloid. *Frontiers in Molecular Neuroscience*, 4, 22. https://doi.org/10.3389/fnmol.2011.00022
11. Arora, S., Balotra, S., Pandey, G., & Kumar, A. (2016). Binary combinations of organophosphorus and synthetic pyrethroids are more potent acetylcholinesterase inhibitors than organophosphorus and carbamate mixtures: An in vitro assessment. *Toxicology Letters,* 268, 8-16. https://doi.org/10.1016/j.toxlet.2016.12.009
12. Soderlund, D. M. (2011). Molecular mechanisms of pyrethroid insecticide neurotoxicity: Recent advances. *Archives of Toxicology,* 86(2), 165-181. https://doi.org/10.1007/s00204-011-0726.x
13. Cao, Z., Shafer, T. J., & Murray, T. F. (2010). Mechanisms of pyrethroid insecticide-induced stimulation of calcium influx in neocortical neurons. *Journal of Pharmacology and Experimental Therapeutics*, 336(1), 197-205. https://doi.org/10.1124/jpet.110.171850
14. Haller, H. L., McGovran, E. R., Goodhue, L. D., & Sullivan, W. N. (1942). The synergistic action of sesamin with pyrethrum insecticides. *The Journal of Organic Chemistry,* 7(2), 183-184. https://doi.org/10.1021/jo01196a011
15. Maurya, S.K., Mishra, J., Abbas, S., & Bandyopadhyay, S. (2015). Cypermethrin stimulates GSK-3ß-dependent Aß and p-tau proteins and cognitive loss in young rats: Reduced HB-EGF signaling and down stream neuroinflammation as critical regulators. *Molecular Neurobiology*, 53(2), 968-982. https://doi.org/10.1007/s12035-014-9061-6
16. Kale, M., Rathore, N., John, S., & Bhatnagar, D. (1999). Lipid peroxidative damage on pyrethroid exposure and alterations in antioxidant status in rat erythrocytes: A possible involvement of reactive oxygen species. *Toxicology Letters*, 105(3), 197-205. https://doi.org/10.1016/s0378-4274(98)00399-3
17. Guven, C., Sevgiler, Y., & Taskin, E. (2018). Pyrethroid insecticides as the mitochondrial dysfunction inducers. In E. Taskin (Ed.), *Mitochondrial Diseases* (pp. 293-322). IntechOpen. https://doi.org/10.5772/intechopen.80283

■ AUTHOR

Marguerite Li is a senior at Jericho High School in Jericho, NY. She plans to pursue an education in neuroscience because of its complexity and mystery; majoring in biomedical sciences or biochemistry are also options she would like to pursue. When she is not busy reading journal articles for her research, Marguerite loves to solve puzzles and read a good thriller/mystery book.

Quality School Dining Experience Using Neural Network, Machine Learning and NLP Technologies

Taehoon Han

Northfield Mount Hermon, 1 Lamplighter Way, Gill, MA 01354, United States
thhan0130@gmail.com

ABSTRACT: Schools in the U.S. waste about $5 million worth of edible food every day, translating to $1.2 billion in losses per school year. [1] This is caused by not only the schools' inability to predict students' demand for their menus but also their failure to satisfy an ever-changing diverse range of student tastes. The purpose of this research is to evaluate the feasibility of harvesting a suite of artificial intelligence (AI) technologies to suggest solutions to this problem. Technologies explored in this research are neural network, machine learning, and Natural Language Processing (NLP). Starting from Northfield Mount Hermon (NMH), a private boarding school in Massachusetts, the scope of the research will expand to include other private boarding schools in the state and ultimately all private and public schools in the United States. The aforementioned technologies have been shown to be effective in a certain domain of the problem. This research highlights a need for government intervention financially and politically.

KEYWORDS: School dining; Artificial intelligence; NLP; Machine Learning; Moblie application.

INTRODUCTION

In the United States, nearly 7 billion meals are served in K-12 schools each year through the National School Lunch Program (NSLP) and National School Breakfast Program.[2-4] Despite growing interest in school dining, this has been no analysis on how students consume and react to the food offered. My research found that the most school food waste is generated due to a school administrations' inability to predict students' demand for a given menu and structural limitations to meet students' tastes. To provide feasible solutions applicable to any school settings (private and public) in the U.S., this study examined artificial intelligence including neural network, machine learning, and NLP (Natural Language Processing) and how technology can work as a guide for school systems to lessen food waste. Furthermore, the analysis is done of the technologies' effectiveness by comparing student satisfaction scores before and after the change. The research concludes with implications for the federal government's policy issues regarding the school dining system, whereby more streamlined regulations and sets of policies might contribute to mitigating food waste problems.

RESULTS AND DISCUSSION

Two types of inputs were gathered from the students: the three key in-app activity metrics and their responses (in "natural language") later processed according to NLP algorithms embedded in the application. Both results were cross-referenced with traditional surveys and in-person interviews.

Key in-app Activity Metrics: Key in-app activity metrics provide numeric measures of the level of student engagement in the study, thus serving as the foundational features of the project. Overall, the results on the three metrics were simplified according to the funnel that students were guided through while using the app, as well as the sex of the students as shown in Table 1.

Overall, the engagement level in the app activity was generally on par with other types of mobile applications, as can be seen from session length per login (5.2 for male students and 5.4 for female students). This demonstrates that the participating students showed moderate to strong level of interests in the activity itself. It is notable to see that both male and female students stayed for shorter periods of time during the second

Table 1. Results from the key in-app activity metrics.

1. Session length per login	Minutes (min)
OVERALL	
MALE STUDENTS	5.2
FEMALE STUDENTS	5.4
1ST HALF OF THE PERIOD	
MALE STUDENTS	6.1
FEMALE STUDENTS	5.8
2ND HALF OF THE PERIOD	
MALE STUDENTS	4.9
FEMALE STUDENTS	5.2
2. CTR; Click Through Rate	**Percentage (%)**
OVERALL	
MALE STUDENTS	73
FEMALE STUDENTS	60
1ST HALF OF THE PERIOD	
MALE STUDENTS	67
FEMALE STUDENTS	52
2ND HALF OF THE PERIOD	
MALE STUDENTS	79
FEMALE STUDENTS	66
3. Frequency of Response	**Percentage (%)**
OVERALL	
MALE STUDENTS	73
FEMALE STUDENTS	60
1ST HALF OF THE PERIOD	
MALE STUDENTS	67
FEMALE STUDENTS	52
2ND HALF OF THE PERIOD	
MALE STUDENTS	79
FEMALE STUDENTS	66

DOI: 10.36838/v2i2.11

half of the session. It was expected and confirmed that this was primarily because students became more apt at navigating throughout the whole process not because the level of engagement decreased. Click through rate (CTR) measured whether students clicked on the link provided. Male students showed significantly more willingness than their female counterparts did. Meanwhile, the rate for both groups increased during the second half of the session and it was confirmed that this was driven by the increased level of engagement after they were better informed about the possible positive effects of the application.

Finally, the rate of response (free response) reassured researchers that students enjoyed engaging in the activity as both groups showed higher levels of response rate than originally expected. In this category of metrics, female students showed slightly more willingness to provide their opinions in free form than their male counterparts although no one reason for this was confirmed. As with CTR, the rate of response increased significantly during the second half of the session after the students were provided more information on the context of the study.

It is important to consider that the results may have some innate biases due to the limited size of the sample test group, some of which will be further discussed in the later section of this paper. Additionally, characteristics of the sample test group might have influenced the results to some extent, due to its unique socioeconomic status as private school students.

Table 2. Results from the free response.

Metric	*Unit*
GENERAL FEEDBACK	**69%**
Very Positive	6
Somewhat Positive	35
Moderate or Mixed	6
Somewhat Negative	12
Negative	8
Very Negative	2
MENU SUGGESTIONS	**26%**
Calorie Related	15
Diversity Related	5
Others	6
MISCELLANEOUS	**5%**

In-depth Analysis on the Free Response through NLP Algorithm: In the free response section, comments were categorized as menu suggestions if they included menu related words. Approximately 15% of the free responses were not be able to be automatically categorized, so each comment was categorized by hand as a second step.

As a result of the NLP analysis, general feedback accounted for 69 percent of the total comments, followed by menu suggestions (26 percent). Within the general feedback comments, comments with 'somewhat positive' attitudes towards the dining or dining system as a whole had the most share while ones with 'very negative' accounted for the least share. With some due reservations, this result indicates that students at Northfield Mount Hermon School were generally satisfied with the current level of school dining.

Meanwhile, comments that were categorized as menu suggestions were further specified into three types: calorie-related, diversity-related, and miscellaneous. The majority of comments were calorie related as shown in Table 2. This could be due to the fact that NMH students receive three meals a day and the school dining is the only place where they get the food.

CONCLUSION

The most essential part of this research is how effectively student satisfaction level could be gauged regarding school dining and translate it into quantifiable measures. This project showed that the process was able to easily achieve the goal. However, there were areas that might have provided unfair advantages. First, the number of samples was quite limited. Second, the samples were limited only to participants from NMH. When tested in a more general setting (e.g. public school in California), the process might lose its validity in reflecting the student opinion about school dining in an accurate manner.

It is necessary to expand the scope of research to include different educational settings and socioeconomic backgrounds. Additionally, this research does not examine public policy perspectives on the school dining issue. Student satisfaction levels is related to how much money and public attention would be invested into the dining quality. Future research should include the possible effects that an increased budget and public attention might have on how students feel about their meals in the cafeteria.

METHODS

The data used in this study was obtained from students' in-app activity and surveys conducted in a controlled school dining environment at NMH. Data on the recommended daily consumption of vegetables, fruit, and fish for children aged 4-18 years in the U.S. was used as a reference benchmark. (2) The number of individual students that participated in the study was 120 (64 male students and 56 female students) and the study lasted for 10 days.

As a way to identify the level of engagement, the study complemented the key in-app metrics with free responses provided by the students. Students were given a choice between leaving comments and opting-out. The app automatically saved the response in the server where they are then processed through NLP algorithms for further analysis. We categorized the responses into three types: general feedback, menu suggestions, and miscellaneous.

A comment was categorized as general feedback if it contained general attitude or reactions to the food, dining system as a whole, service, and others. Using NLP algorithms, the comment was further categorized into one of six types of feedback: very positive, somewhat positive, moderate or mixed, somewhat negative, negative, and very negative. It is notable

that human intervention was an inevitable process in this step where the researchers had to set the standard of deciding a comment's degree of positive or negative. For example, a comment was categorized as very positive if it contained five or more words that expressed positive feelings whereas somewhat positive comments had two to four positive words. Implications of the human intervention are discussed in the final section of this paper.

Data Sources: The first key metric to measure the educational value of using NLP in improving dietary habits is the in-app activity. As students interact with a virtual assistant on a mobile app-based environment, the level of in-app activity including CTR, frequency of response, and total session time per login provides an important measure of student engagements with the proposed solution (Table 3). This also resonates with some of the key benefits of using a mobile application as an educational tool: student engagement can be measured in a clear and objective manner allowing data-driven approaches to further enhance interaction with students.Session length per login is measured as total minutes that a user stayed in the app during a single login. CTR indicates the level of a student's willingness to advance into the next step of the session. Frequency of response measures whether a student provides his or her response when given a question proposed by the app's virtual assistant. Each of these key metrics are later compared to industry averages of each corresponding metric to provide insight into the app's effectiveness for increasing student engagement levels.

Table 3. Key Metrics from In-App Activities

Metric	*Unit*
1. Session Length per Login	Minutes (min)
2. CTRS; Click through Rate	Percentage (%)
3. Frequency of Response	Percentage (%)

Another key measure of this study is based on surveys conducted in Northfield Mount Hermon (NMH), a private boarding school in Massachusetts. Survey questions are designed to highlight qualitative assessment of satisfaction levels towards a student's own dietary behavior.[3,4] Students are asked to fill out the identical survey forms before and after the app's introduction implementation for one school month.

Study Methods: One of the purposes of employing NLP in this study is to automate the feedback gathering process in the school dining environment. Traditional approaches require instructors on-site to distribute surveys or questionnaires to students. This method takes much time and effort and also does not guarantee the accuracy or confidence level of the responses. The researchers attempted to ameliorate potential drawbacks usually seen in the traditional approach by using NLP embedded in the mobile app.

The NLP used in the research follows typical NLP steps with some modifications and optimization conducted within the codes to reflect behavioral context of the students. The major steps of NLP used in this study are shown in Table 4.

Table 4. Major Steps in NLP Process in the Study

Step	*Tasks*
Pre-processing	- Clean Less usefull part of texts and annotate texts by mark-ups and tagging - Normalize texts by stemming, lemmatization, and other forms of standardization
Mopheme Analysis	- Apply transformation rules and eliminate false analysis results from candidates - Generate tagss and final results by reflecting dictionary information and variuos constraints
Semantic Analysis	- Identify requirements of the sentence through phonetic analysis - Decide what action to take based on neural netwrok or rule based on extracted sentence information

In this study, there were several modifications required for the semantic analysis step. Based on the extracted information (key clues) provided by the students, the virtual assistant is designed to guide students to several different scenarios. This step serves multiple purposes: 1) assessing the engagement level with the whole process, 2) acquiring additional key clues to proceed to the next step, and 3) identifying and delivering to the main server of the app about the student's response.

REFERENCES

1. Hart, Caroline Sarojini. "The School Food Plan and the Social Context of Food in Schools." Cambridge Journal of Education 46, no. 2 (2016): 211–31. https://doi.org/10.1080/0305764x.2016.1158783.
2. Anderson, Michael, Justin Gallagher, and Elizabeth Ramirez Ritchie. "School Lunch Quality and Academic Performance," 2017. https://doi. org/10.3386/w23218.
3. Weaver-Hightower, Marcus B. "Why Education Researchers Should Take School Food Seriously." Educational Researcher 40, no. 1 (2011):15–21. https://doi.org/10.3102/0013189x10397043.
4. Ball, Kylie. "Improving Eating Behaviours in Socioeconomically Disadvantaged Communities." http://Isrctn.org/>, May 2012. https://doi. org/10.1186/isrctn48771770.

AUTHOR

Taehoon Han has served as school dining leader at NMH, and he started an AI-based mobile application that automatically collects and analyze students' feedback on school dining and deliver them to dining administrators.

DOI: 10.36838/v2i2.11

ACASU

American Commission for Accreditation of Schools and Universities

STANDARDS

* Governance

* Education

* Culture

* Operations

* Quality

Lifetime Qualities

MORAL
Caring, Modest, Ethical, Respectful

THINKER
Knowledgeable, Inquirer, Curious

PRODUCTIVE
Achiever, Excellence Seeker

INNOVATIVE
Problem Solver, Explorer

VISIONARY
Global Citizen, Inclusive

ACASU BENEFITS

Enhances your institution's reputation and garners the esteem of Americans.

Assures students and their families that your institution meets American standards for high quality education.

Shows students and parents your institution's commitment to success and your ability to perform at high quality educational standards.

Shows your colleagues and the general public that your institution has a genuine interest in institutional improvement.

Sets your institution apart as a leader among others.

Provides you with a network of skilled professionals and resources to assist meeting your accreditation, enrollment, and college attainment goals.

Ensures that your standard operating procedures are compliant with U.S. international student regulations.

ACASU.ORG

The International Journal of High School Research is published
by Terra Science and Education