

## **Examining the Accessible Chromatin Regions Between Duplicated Gene Copies in the Tetraploid *Leucaena trichandra***

Katie Banga, New Mexico State University

Dr. Donovan Bailey, New Mexico State University

Chromosomal and genetic duplications are key sources of genetic novelty in evolution. When duplication events occur, two copies of the genes are usually not necessary, which results in one being lost through mutational processes. However, duplicate copies can also be retained, going on to gain new functions or subfunctions between the daughter genes. For these reasons, the duplicated regions are often expressed differently than the original copy. The plant genus *Leucaena*, which contains 19 species with tetraploid genomes, provides exceptional opportunities to investigate the evolutionary changes brought forth by chromosomal and genic duplications. Species in the group are good models because they grow relatively quickly for trees and they have many uses as multipurpose crops in the tropics. Here we use the genome of the tetraploid, *L. trichandra*, because it is a parent of other octoploid species within the genus. Chromatin accessibility studies, such as the one conducted here, include methods used to identify heterochromatic DNA regions open to transcription factors and therefore expression. We used an Assay for Transposase-Accessible Chromatin (ATAC-seq) to determine the changes to the accessible chromatin regions between different copies of the chromosomes in the *L. trichandra* genome. This is possible because a Tn5 transposon is inserted into open chromatin regions during the library preparation. The results of the ATAC-seq analysis will provide an important annotation for future research and allow us to make conclusions about the accessible regions that have experienced gene expression differences since the duplication of the chromosomes. The data obtained from ATAC-seq will identify regions of accessible and inaccessible chromatin, providing key information regarding how chromatin accessibility varies among the different copies and the relative fate of duplicated gene copies.

**Poster #8**

Luis Carranza, New Mexico State University  
Dr. Sijo Augustine, New Mexico State University

The idea of this project was to innovate a Three-Level Boost Converter which will be implemented into a Photovoltaic system. Although there have already been many different designs of DC Boost Converters, we have taken it upon ourselves to create a design that will be more effective to avoid the replacement of an entire Printed Circuit Board due to any shorts in the circuit or other malfunction in components such as voltage measuring devices or the Gate Driver component. Although I am sure there have been designs implemented where they have separated the voltage sensors and Gate driver components, we believe that there can be a more efficient way, this can help reduce error as well as cost during the manufacturing process as well as in the long term use of the DC Boost Converter. During this research we took an existing design of a DC Boost Converter and decided to innovate by creating a PCB that is strictly the Power Circuit, we also proceeded to create another board for auxiliary power, a board for the gate driver and controller, and a series of different boards that can be used as voltage sensor (measurement devices). Overall the project is still in process but we have discovered that we will have a more efficient design that will increase the efficiency of the converter as well as elongate the life of the components.

**Poster #10**

Alexa Davis, University of New Mexico-Valencia Campus  
Marissa Vigil, University of New Mexico-Valencia Campus  
Benjamin Flicker, University of New Mexico-Valencia Campus

The Rio Grande river stretches across the lower southwest region of the United States. It goes near the farmlands in Alamosa Colorado, right through New Mexico, parts of Texas, and ending in the Gulf of Mexico. The Rio Grande has 15 dams, we chose to study the bacterial communities found around the Isleta Division Dam, New Mexico (34.90602 N, 106.68634 W) (Figures 1 & 2). The Isleta Diversion Dam diverts water from the river into irrigation canals that run through the Isleta pueblo. In the Rio Grande, there can be harmful bacteria that grow and get transferred into our crops. Our project is centered around using 16s sequencing which is used to find specific genetic codes for bacteria with over 89% accuracy (Illumina, 2022). We sampled around the dam upstream in the morning and evening and downstream in the morning and evening. We took 8 samples altogether, we took 4 samples at 6 am and 4 samples at 6 pm: Sample U1AM & U1PM, upstream was set 5-10 feet upstream from the dam. Sample D1A & D2P, Downstream was set around 50 feet downstream from the dam. In our results, we found the 7 most common phyla of bacteria both upstream and downstream included Proteobacteria, Bacteroidetes, Actinobacteria, Verrucomicrobia, Firmicutes, Cyanobacteria, and Chlorflexi, with Proteobacteria, have the highest number of reads.

## **Signals Features for Classification of Power System Disturbances using PMU Data**

Daniel Diaz, New Mexico State University  
Alexa Templeton, University of New Mexico  
Patrick Jojola, New Mexico State University  
Jeremy Nichols, University of New Mexico

Disturbance classification in our power systems is vital in ensuring that energy, which powers our communities and homes, stays protected and under control. With the usage of PMU data collected from the Phasor Data Concentrator (PDC), we are able to implement machine learning techniques to accurately classify the status of relays or disturbances. This poster shows that S3 feature extraction is the better method overall by capturing the trend of sequence by calculating the changes of sequence values. This in turn helps classify signal disturbances.

## **A Machine Learning Approach for Optimizing Antenna Design.**

Matthew Kube, Eastern New Mexico University

Sarbagya Ratna Shakya, Eastern New Mexico University

With an increasing demand for smarter antenna design in advanced technology applications like 5G, IoT (Internet of Things), and WLAN applications, the development of Antenna design has been a complicated process, which requires multiple iterations and extensive testing to produce a final product. With recent development in Machine Learning (ML) algorithms and the availability of data for antenna design, we investigated different Machine learning algorithms, for optimizing the output strength of an antenna by analyzing different dimensions of the patch antenna. We utilized Regression based ML models to learn the behaviors and efficiency of the antenna and predict the output strength (S11) for a range of frequencies. The preliminary comparative study of different algorithms shows that Random Forrest (RF) has higher accuracy as compared to other ML algorithms. The application of this study will help to analyze the design using ML algorithms and will provide a quicker and smarter antenna design as compared to traditional design methods.

## **Growth of Long Emission Wavelength (1.3 – 2 $\mu\text{m}$ ) InAs Quantum Dots using Metamorphic Buffers**

Carter Heinrich, University of New Mexico  
F. Ince, University of New Mexico,  
M. Moreno, University of New Mexico,  
G. Balakrishnan, University of New Mexico

InAs Stranski-Krastanov (SK) quantum dots have been used for single photon emitters. The SK growth mode involves the growth of a planar wetting layer of InAs on a substrate like GaAs, and subsequently, the high mismatch drives the growth three dimensional leading to an ensemble of nanoscale islands or quantum dots. This growth phenomenon has been observed in a few material systems; however, InAs on GaAs has been the most widely used embodiment of this growth resulting in lasers, solar cells, and detectors. The coverage of InAs that is used for quantum dots ranges from 1 ML which is a pure wetting layer to up to 2.7 MLs. Beyond this, the dots experience Oswald ripening and result in photonically inactive quantum dots. This translates to a wavelength range of  $\sim 1000$  nm to 1250 nm. While these quantum dots are typically very high in density at  $> 10^{10}$  dots/cm<sup>2</sup>, through growth methods the density can be reduced and using advanced microscopic techniques a region with a few quantum dots can be probed. In this paper, we attempt to increase the emission wavelength of InAs quantum dots to telecommunication wavelengths of 1.33  $\mu\text{m}$  and subsequently to 1.55  $\mu\text{m}$ . To do this, we must be able to grow larger quantum dots and that will require reduced mismatch between the quantum dot and the underlying substrate. We, therefore, employ a metamorphic buffer on GaAs comprised of step-graded AlGaAsSb. The results presented will include XRD and TEM-based characterization of the buffer along with AFM and photoluminescence-based analysis of the quantum dots.

## **Concerns of the 2016 WHO Diagnosis Criterion for Polycythemia Vera**

Sophia Lim, Albuquerque Academy

Amalia S. Parra, University of New Mexico

Christopher A. Johnston, University of New Mexico

Polycythemia Vera is a Myeloproliferative disorder involving an increase in red blood cells and white blood cells. Increased red blood cells result in hyperviscosity of the blood causing it to clot and interfere with blood flow through small blood vessels. In 2016, the WHO altered its guidelines regarding the diagnosis of PV from 2008 due to the emergence of "masked PV" for which patients fared worse outcomes, which mimicked Essential Thrombocythemia, increasing the risk of thrombosis due to delayed diagnosis. Modern diagnosis status for PV requires the mutation status of JAK2V617F. 3 While more accurate the 2008 guidelines, there are still issues due to the limitations in JAK2 testing caused by a lack of a broadly-used method. These can include qPCR, PCR, melting curve analysis, pyrosequencing, RFLP, and sanger sequencing as well as the intellectual property rights of Ipsogen which influence the choice of method. In this poster, I will highlight the issues of the multiple methods and the prospect of endogenous erythroid colony formation, which is highly specific for JAK2 mutations and could be an effective addition to diagnosis criteria.

## Using Bisulfite Sequencing to Characterize the Fate of Duplicate Gene Copies

Mike Lopez III, New Mexico State University

Donovan Bailey Ph.D., New Mexico State University

Katie Banga, New Mexico State University

The field of Epigenetics is a relatively new field of scientific study and research, with the focus being on the effect that nature has on an organism's development. However, there are still specific areas of epigenetics that are not fully understood. The current literature describes several mechanisms that play key roles in how an environment can cause changes in an organism's genetic structure over time, such as can be described as the fate of duplicate gene copies; Pseudogenization, conservation of gene function, subfunctionalization, or neofunctionalization are all ways the genome deals with duplicated genes. We are focused on the methylation patterns in the tetraploid *Leucaena trichandra* genome as an example of what can happen to genic methylation following genome duplication. These patterns are identified by a methyl group attached to the phosphate group of a nucleotide in the genome; methylation (presence of a methyl group on a nucleotide) is commonly found on cytosines, with some adenines being methylated as well. The goal of our project is to use bisulfite sequencing (a type of gene library that accounts for methylation) to detect the methylcytosine patterns in the *Leucaena trichandra* genome to identify gene copies that are likely to be available (unmethylated) or unavailable (methylated) for expression and function. The expected results of our experiment include a description of the *Leucaena* genome that shows three classes of methylation patterns for duplicated genes, including only one copy methylated (one copy retains function), neither copy methylated (both copies retain function), or both copies methylated (neither copy retain function).



## **Comparing Semi-Supervised Anomaly Detection Algorithms that use PMU Data to Detect Cyber Attacks in Smart Grids**

Crystal Cervering, University of New Mexico

Allison Kilpatrick, University of New Mexico

Larry Maes, Eastern New Mexico Univeristy

Damian Lovato, Central New Mexico Community College

There are a number of different machine learning algorithms capable of detecting cyber attacks. Previous methods have used supervised algorithms which use both normal and attack data to train a detection model. Supervised learning may have poor performance in detecting cyber-attacks because it is difficult to get various instances of attack data. Whereas semi-supervised anomaly detection algorithms rely only on normal event data to train a detection model, making it more capable of detecting unknown attack types. Data is gathered from the Phasor Measurement Unit (PMU), which is a sensing device that provides real-time measurements of the power system state. The data from the PMU is then used to train and improve the accuracy of learning algorithms. smart grid network Preprocessing

## **Metagenomic 16s Sequencing of the Rio Grande Dam in Isleta, New Mexico**

Marissa Vigil, University of New Mexico - Valencia Campus

Alexa Davis, University of New Mexico - Valencia Campus

Professor Benjamin Flicker, University of New Mexico - Valencia Campus

The Rio Grande river stretches across the lower southwest region of the United States. It goes near the farmlands in Alamosa Colorado, right through New Mexico, parts of Texas, and ending in the Gulf of Mexico. The Rio Grande has 15 dams, we chose to study the bacterial communities found around the Isleta Division Dam, New Mexico (34.90602 N, 106.68634 W) (Figures 1 & 2). The Isleta Diversion Dam diverts water from the river into irrigation canals that run through the Isleta pueblo. In the Rio Grande, there can be harmful bacteria that grow and get transferred into our crops. Our project is centered around using 16s sequencing which is used to find specific genetic codes for bacteria with over 89% accuracy (Illumina, 2022). We sampled around the dam upstream in the morning and evening and downstream in the morning and evening. We took 8 samples altogether, we took 4 samples at 6 am and 4 samples at 6 pm: Sample U1AM & U1PM, upstream was set 5-10 feet upstream from the dam. Sample D1A & D2P, Downstream was set around 50 feet downstream from the dam. In our results, we found the 7 most common phyla of bacteria both upstream and downstream included Proteobacteria, Bacteroidetes, Actinobacteria, Verrucomicrobia, Firmicutes, Cyanobacteria, and Chlorflexi, with Proteobacteria, have the highest number of reads.

## Analyzing the Golden Age of Science Fiction: A Topic Modeling Approach

Stephen Villanueva, Eastern New Mexico University

Edgar Eduardo Ceh Varela, Eastern New Mexico University

Science Fiction is a popular literature genre, yet it lacks a clear definition. Even literature scholars have different interpretations of what Science Fiction is. Several Machine Learning (ML) techniques exist to perform literature analysis. Similarly, Natural Language Processing (NLP) comprises a set of tools and algorithms to analyze, manipulate, generate and understand human language. A typical analytical method for evaluating data is topic modeling. Topic modeling has been used as a research method in many different disciplines. These studies examine enormous volumes of text data to categorize document subjects objectively and examine the trends within each topic. In this study, we apply different ML and NLP techniques to a corpus of more than 2,000 Science Fiction novels from the so-called "Golden Age of Science Fiction" (i.e., the 1930s to 1960s). Our main goal is to find hidden topics in these documents to understand the Science Fiction genre better. In text analysis, we assume that documents from the same group must cover the same topic. Therefore, in this project, we use two different approaches to compare and contrast the topics extracted by each of them. In the first approach, we use a Latent Dirichlet Allocation (LDA) model for topic extraction. LDA is a generative probabilistic model considering the documents as random mixtures over latent topics. For the second, we use Doc2Vec. This model uses neural networks to represent each Science Fiction novel as an n-dimensional vector. Then, using a KNN algorithm, we use the generated vectors to find groups of similar documents. We obtained three important topics using LDA and five groups of documents using Doc2Vec with KNN. Some of the topics we found are related to "Horror and creatures," "Crimes," "Galaxy and space," and "Robots and Machines." Similarly, we analyze these topics' evolution during the "Golden Age" period. This analysis shows how different topics increased or decreased in popularity over this period. Applying NLP techniques to a corpus of literature helped us discover latent topics that give us insight into the most important themes written during this important period for the Science Fiction genre.

## Comparing the Growth of Xeric Trees in Soil vs. Aquaponics

Conner M. Wood, Santa Fe Community College

Stephen M. Gomez, Santa Fe Community College

R. Charlie Schultz, Santa Fe Community College

In my experiment, I was comparing the growth of xeric trees that grow in the southwest in Soil with traditional methods vs. Aquaponics. I felt as though this application could provide a potential solution to fighting climate change, saving massive amounts of water, as well as contribute to habitat restoration efforts. I also examine factors that are common in the field of aquaponics such as PH, EC, N03 concentration, and PAR. Aquaponics has been around for thousands of years however, we as a society today still have a lot of research into what answers to current problems aquaponics can give us. Today there is not a lot of knowledge about growing Xeric trees in aquaponics so far, this is a unique study. This is a study that took place over one year from the summer of 2021 to the summer of 2022, there were three trials over the course of the study. I originally started with seeds that were in plug trays and were ready to be transplanted. There are more than several varieties of trees in the study. Since I was comparing the growth rate of the trees in various methods, all of the seeds per each specie were sown in the trays at the same time. This offers a fair and sets a starting point for all of the trees to be fairly compared in terms of height, width, and root mass comparison. I will demonstrate how I transplanted them into the aquaponics system as well as how I transitioned them out and eventually hardened them off without harming them. I will also discuss methods for a successful transplanting of the trees from water-based media into soil media including specific media used to help the transition. Overall after all 3 trials, I answered my original questions which were, Can the growth rate of trees is execrated with aquaponics vs. growing them in soil with traditional methods? As well as, Can larger root masses be cultivated while growing trees in aquaponics? The answer was Yes, for both questions and all of the species in the three trials. I believe that my findings of achieving faster growth rates of trees with aquaponics have the potential to help produce trees faster which has many applications. This can be used to speed up Habitat Restoration and possibly help curb the effects of climate change by being able to speed up tree production which

## **Model and Power Flow Analysis of a Distribution Feeder using Data Recorded from Buildings, Facilities, and a Photovoltaic System**

Adrian Maez, Department of Electrical Engineering, New Mexico Institute of Mining and Technology

Casie Taylor, Department of Electrical Engineering, New Mexico Institute of Mining and Technology

Dawn Walaitis, Department of Electrical Engineering, New Mexico Institute of Mining and Technology

Kevin Wedeward, Department of Electrical Engineering, New Mexico Institute of Mining and Technology

This poster presents the model and power flow analysis of a distribution feeder developed for academic studies related to the optimization, control, and integration of distributed energy resources such as sources of renewable energy and energy storage systems. The 16-bus, three-phase model was created to provide a realistic test feeder the size of a small campus where radial underground cables connect the utility, the primary source of power to loads and a photovoltaic system. Measurements of power collected from actual buildings, facilities, and a photovoltaic system serve as the basis for the model and enable computer simulations to be performed with realistic profiles over time. A number of models of distribution test feeders are available from IEEE and others, but few are on the scale of a small campus, and fewer yet have associated time-series data. The proposed model consists of underground cables represented by total phase impedance and shunt admittance matrices, seven loads (five buildings, heat plant, and chiller plant) that have a peak total of approximately 2,000kW and have time-series values for average and reactive power, and a 346kW-rated photovoltaic system that has time-series values for average power. To provide options and confirm results, the model and associated power flow equations were solved using hourly data over multiple days in three software packages (Gridlab-D, OpenDSS, and MATLAB). The outcome of this effort is a new test distribution feeder with time-series solutions that may be used by researchers to develop more sustainable and resilient power distribution systems in the future.

## **Survival Motor Neuron Uncovers a New Link Between Splicing and Defense Response in *Arabidopsis thaliana***

Jayden Montoya, Northern New Mexico College

Tristan Maestas, Northern New Mexico College

Regina Bedgood, Northern New Mexico College

Haidong Yan, University of Georgia

Xaun Zhang, University of Georgia

Michael Gonzales, University of New Mexico

Bob Schmitz, University of Georgia

Mario Izaguirre-Sierra, Colorado State University-Pueblo

Spinal Muscular Atrophy (SMA) is an autosomal recessive genetic disorder found in metazoans that are associated with the loss of motor neurons caused by mutations within the survival motor neuron (SMN) gene. A complete knockout of motor neurons and the progressive degeneration of muscle tissue. SMA is of the smn gene that results in embryonic lethality in humans as well as animals such as mice and flies. The SMN protein is a muscle-based protein, which is encoded in two genes in the human genome, plants contain only one smn gene although they lack muscular structures. Interestingly, smn mutant, plants are nonlethal. SMN protein is necessary for the assembly and transport of small nuclear ribonucleoproteins (snRNPs) which are required for the spliceosome in all eukaryotes. Our preliminary data showed that several defense response genes are over-expressed in our smn mutant background. Specifically, my goal is to use molecular genetics and bioinformatics to understand the role of smn and alternative splicing in defense response. Therefore, I am studying the chromatin behavior during and after a pathogen attack using Assay for Transposase-Accessible Chromatin with sequencing (ATAC-seq). Investigating the chromatin accessibility profiles in smn mutant plants during infection will allow us to understand the genomic response during pathogenesis.

## **Comparison of Bacterial Communities in the Middle of the Rio Grande**

Aida Nevarez, University of New Mexico - Valencia campus

Jarret Attell

Ben Flicker Ph.D., University of New Mexico - Valencia Campus

Victor French

While soil and aquatic bacteria are critical to ecosystem function, their abundance and diversity in river ecosystems, particularly the Rio Grande, have not previously been studied. The impact of wastewater treatment effluent discharge on these bacterial communities is also not well understood. In this study we used the 16s rRNA sequencing method to elucidate the diversity of bacteria in three different regions of the middle Rio Grande in New Mexico: Los Lunas upstream of the municipal effluent discharge; the site of the Los Lunas wastewater treatment effluent discharge into the river; and Belen, south of the Los Lunas municipal effluent discharge. Both water and soil samples were collected for whole genome extraction. We also compared the three sites in abiotic factors such as dissolved oxygen, conductivity, and temperature. We will present the bacteria found using these sequencing data and make comparisons of the species diversity between these three communities based on statistical analyses. We will also provide comparisons of the species diversity in collections from the fall of 2019 through the winter of 2022.

was previously thought would take much longer to produce. Aquaculture can be a sustainable form of food in areas that greatly limit food production capabilities. Growing trees indigenous to the specific region could offer a great product for habitat restoration, combining these two concepts with aquaponics could be a potential game changer for restoration and sustainability.



## **The Effects of Forrest Fires on Bacterial Communities**

Savanna Parada, University of New Mexico

Fatima Martinez, University of New Mexico

Julia Aragon, University of New Mexico

Benjamin Flicker, University of New Mexico - Valencia Campus

Wildfires can cause a significant impact on the surrounding ecosystem. It can be difficult to determine the effect these wildfires can have on microorganisms. For our experiment, we collected four total samples. Each pair was collected along the Rio Grande in Jarales, New Mexico (34.5518045° N, -106.5819789° W), which experienced a wildfire in June of 2021. Visibly burned soil was collected for samples A1 and A2 while visibly unaffected soil was collected for samples B1 and B2. We used 16s sequencing to determine the diversity of the classes of bacteria found. The results were consistent with the most abundant classes being Actinobacteria, Alphaproteobacteria, Gammaproteobacteria, Deltaproteobacteria, and Betaproteobacteria for all samples. These results do not offer any evidence that burned soil has less diversity in the bacteria found compared to regular soil.

## Phytoremediation

Serena Prather, Santa Fe Community College

The Ambrosia Lake disposal site is the home of a nonoperational Uranium ore processing facility in New Mexico. The facility stopped operations in 1982. Algae polycultures that were isolated at the site, have been shown to remove radionuclides from the water. Our team intends to take the next step in this research project by turning the experiments into a realistic solution for remediating the water on the disposal site and developing an economic model to understand other impacts of this solution for the region. The current land steward for Ambrosia Lake is hauling in water from outside of the area for dust control and other uses at the Ambrosia Lake disposal site. The water available at the site has levels of uranium and radium that exceed the EPA standards for drinking water but not the NRC standards for radioactive material. The contamination is in the water in the mines and aquifers in the area. This problem is not isolated to Ambrosia lake, and the solution could be applied to many affected areas. For example, uranium plumes have been documented in the San Andreas, Glorieta, and Alluvial Aquifers at the Bluewater, New Mexico, disposal site. Two hundred and fifty uranium mining sites have operated across New Mexico alone, and over a hundred of these sites have no record of reclamation. The Navajo Nation contains five hundred and twenty-one abandoned uranium mines. The polycultures from the Ambrosia Lake site at Santa Fe Community College have shown successful results in remediating water samples collected from the disposal site. They quickly and effectively remove radionuclides from the contaminated groundwater. The algae are naturally occurring in the site's ecosystem and probably have an exopolysaccharide coating that adsorbs the uranium, but critical questions have yet to be answered. Our team proposes to create a system of bioremediation of heavy metals that can be implemented to help restore lakes and lands to usable levels. Scientific evidence has proven some algae to be incredibly productive in phytoremediation. Discovering practical implementation of algae's abilities in phytoremediation can undo the harm caused to our waterways and land by mining and contamination across the globe.

Algae's superpower of remediation has been overlooked for far too long. Scientific evidence has proven it to be incredibly productive in phytoremediation. It can help repair generations of thoughtless damage caused to the earth and help maintain healthy, thriving ecosystems for future generations to come. Moreover, we can help address environmental justice depravity that has existed for too long.

## Use of Machine Learning in Combination with Analytical Tools for Medical Screening

Luis E Tafoya, CURE Summer program for Cancer Research, University of New Mexico

Robert M Taylor, Department of Emergency Medicine, University of New Mexico

Jameel K Remtulla, Department of Emergency Medicine, University of New Mexico

Xander Augustson, Department of Emergency Medicine, University of New Mexico

Andrea Howard, Department of Emergency Medicine, University of New Mexico

Virginia Severns, Clinical Translational Sciences Center, University of New Mexico

Sarah Lavelle, Department of Emergency Medicine, University of New Mexico

Silas Bussmann, Department of Emergency Medicine, University of New Mexico

Justin T Baca, Department of Emergency Medicine, University of New Mexico

Gas Chromatography (GC), Mass Spectrometry (MS), or a combination of both, are used in many industries to identify substances and contaminants in complex mixtures. Analyzing GC and MS signatures using Artificial Intelligence (AI) allows for improved signal classification and predicting the composition of unknown mixtures. This research focuses on training Machine Learning (ML) algorithms capable of predicting SARS-CoV-2 infection status through GC analysis of exhaled breath. 150 human subjects were recruited for feasibility studies in an effort to develop a rapid, point-of-care test for SARS-CoV-2 infection based on GC-ML signal analysis. 100 had positive PCR tests for SARS-CoV-2 and 50 had negative PCR tests for SARS-CoV-2. Of the subjects with positive PCR results, 50 reported active symptoms while 50 were asymptomatic. After informed consent, subjects were provided 2.0 L of breath into a Tedlar collection bag secured via a mouthpiece connected to high-efficiency particulate air (HEPA) filters. Demographics were also collected. Retention time signatures were measured over a run time of <800 seconds. ML was performed using Scikit-learn libraries in python predicting positive, negative, and positive-asymptomatic individuals using GC intensity and retention time data. Models with metrics are generated and compared with the use of different algorithms as logistic regression. Based on preliminary data, we expect the breath analysis with GC, in combination with ML, will allow the prediction of SARS-CoV-2 cases with an accuracy >70%. This approach may be applicable to other screening strategies such as for diabetes or lung cancer.