# EXPLORING THE RELATIONSHIP BETWEEN *HELICOBACTER PYLORI* AND NATIVE AMERICANS LIVING IN NORTHERN ARIZONA

By Enoque P. Costa S. Júnior

A Thesis

Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Science

in Biology

Northern Arizona University

August 2021

Approved:

Fernando Monroy, Ph.D., Chair

Heidi Brown, Ph.D., M.P.H.

Jason Ladner, Ph.D.

# ABSTRACT

## EXPLORING THE RELATIONSHIP BETWEEN *HELICOBACTER PYLORI* AND NATIVE AMERICANS LIVING IN NORTHERN ARIZONA

ENOQUE P. COSTA S. JÚNIOR

Incidence and mortality rate of gastric cancer among Native Americans is significantly higher than for Non-Hispanic White people. Many factors can contribute to this discrepancy such as diet, lifestyle, adequate sanitary living conditions and genetic factors. One known cause of stomach cancer is infection with the bacterium *Helicobacter pylori*. Some bacterial genetic factors give an advantage to the bacterium, the presence of the *cagA* gene and the genotype of the *vacA* gene ("s" and "m" regions). In this study, we collected and screened 210 biopsies from Native Americans patients living in Northern Arizona. DNA was isolated from these specimens and tested for the presence of *H. pylori.* If positive, a PCR was conducted to determine the status of *cagA* and *vacA* genes. Sequencing allowed us to determine the genotypes of these two genes and epidemiological data from patients was used to determine associations between clinical findings and *H. pylori cagA* and *vacA* genotypes. Our research found a positive association between the *cagA* genotype and esophageal findings. Severe cases of gastric disease were linked with presence of the *vacA* "m1" allele. The *vacA* s1m1genotype was the most common and was associated with *cagA* presence. The *cagA* ABC EPIYA motif was the most common genotype, and isolates expressing *cagA* ABC EPIYA motifs and *vacA* s1m1 alleles are known to be virulent. These results show that more virulent *H.*

*pylori* strains are associated with more severe outcomes among Native Americans living in Northern Arizona.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# Introduction

The cancer burden in the world is high but varies depending on the region, population, and cancer site. Stomach cancer is the fifth most diagnosed cancer type and the fourth leading cause of death among cancer cases (Sung et al., 2021). Amid Alaska Natives (AN), cancer is the leading cause of death (Wiggins et al., 2008) and when it comes to gastric cancer, the incidence rates are higher for AN and American Indians than for Non-Hispanic White (NHW) persons (Wiggins et al., 2008). Other studies found an increased risk for gastric adenocarcinoma in non-white people (Schlansky & Sonnenberg, 2011) and high incidence rates of gastric cancer among people in low socioeconomic status (Uthman et al., 2013).

This is also in accordance with what it's found among in the Navajo Nation. An extensive report (NNDOH, 2013) found that between 2005 and 2013, stomach cancer was the 6$^{th}$ most diagnosed cancer among the Navajo. The incidence rate was three times higher than NHW and it's the second leading cause of death among cancers, killing Navajos four times more than NHW. The same report suggests that this disproportionate burden could be explained by environmental (heavy metals), behavioral (diet, lifestyle) and/or genetic factors.

In Northern Arizona, part of Native Americans live among Coconino, Navajo and Apache counties with approximately 80% affiliated with the Navajo tribe (NNDOH, 2013). In our study, we wanted to investigate the association between gastric cancer in this population with the presence of a bacterium, *Helicobacter pylori.*

The human stomach is one of the most hostile places in the body. With a pH varying from 1 to 2, it is capable of dissolving and destroying most living materials, proteins and lipids (Kong, F., & Singh et al., 2008). Just as a reference, the acid found in car batteries, a sulfuric acid solution, has a pH of <1, which is close in scale and known for being highly corrosive (Lu, 2002).

Before the 1980's, it was thought that nothing could survive inside the stomach and any microorganism would be killed. The theory that the stomach was aseptic went unquestioned for many centuries. But two scientists later refuted that belief (Kusters et al., 2006). Barry Marshal and Robin Warren in 1982 were able to isolate and culture for the first time a species of bacteria able to survive in the stomach (Marshall, 2002). *Helicobacter pylori* is an organism with the following characteristics: rod to S-shape, gram-negative, lophotrichous flagella, microaerophile, fastidious, catalase, oxidase and urease positive (Ley et al., 2015).

These traits all help *H. pylori* to thrive in the harsh stomach environment and to call it home. Once in the stomach, the bacterium needs first to avoid being destroyed by the hydrochloric acid. To achieve this, it produces urease to break down the urea, releasing in the process ammonia and carbon dioxide in the process and elevating the pH. Constantly creating this buffer around it, *H. pylori* then swims toward to the gastric mucus membrane. Its helical shape and flagella provide the mobility necessary to achieve that (Ansari & Yamaoka, 2019).

The gastric mucus membrane protects the gastric epithelial cells from the gastric acid and arises the pH to 6-7. Once reaching the gastric mucosa, *H. pylori* will be in

contact with these cells and dive inside the gastric pits successfully colonizing the stomach using several adhesins (Ansari & Yamaoka, 2017).

The mode of transmission for *H. pylori* is still unclear. However, current evidence suggests that it's likely from an oral-oral or fecal-oral route (Ansari & Yamaoka, 2017; Bui et al., 2016; Kusters et al., 2006). Usually, infection occurs during childhood from a household member and is associated with low socioeconomic status and poor living conditions (Marshall, 2002). Most likely some bacteria are eliminated with the feces and if sanitation and proper waste disposal are lacking, *H. pylori* can then be ingested. Studies have found the presence of bacterial DNA in dental plaque, further supporting oral-oral transmission (Mendoza-Cantú et al., 2017).

For *H. pylori* to colonize the stomach, it also needs to overcome many of the host immunological barriers. Genetic advantages give some strains a better chance to achieve that. In this study, we focused on two genes that can greatly benefit the pathogen in detriment to the host. They are the cytotoxin-associated gene A (*cagA*) and the vacuolating cytotoxin A (*vacA*).

The *cagA* can be present or absent in a *H. pylori* strain and the last gene in the pathogenicity island (cagPAI). This a 40 kb segment with 31 genes primarily encodes a type 4 secretion system (T4SS), which is a syringe-like structure that originates in the bacterium and penetrates the gastric epithelial cells aiding in the translocation of the cagA protein and peptidoglycan (Jones et al., 2010).

Once the resulting cagA protein (120-145 kDa) is inside the host cell, it can act in an unphosphorylated form or in a phosphorylated-dependent form. When

phosphorylated, this is mediated by Src family kinase or Abl kinase (Tammer et al., 2007). This happens in specific sites with a sequence of glutamic acid, proline, isoleucine, tyrosine and alanine (Glu-Pro-Iso-Tyr-Ala) located in the carboxy terminus region. Phosphorylated tyrosine subsequently interacts with Src homology 2 phosphatase (SHP 2) via the Ras-ERK MAP kinase signaling pathway interfering with IL-8 production, cell polarity, proliferation and differentiation, loosens tight junctions, elongates cells giving them the “hummingbird phenotype” and exacerbates inflammation (Argent et al., 2016; Chang et al., 2016; Nguyen et al., 2008).

The Glu-Pro-Iso-Tyr-Ala sequence is known as the EPIYA-motif and depending on the sequence surrounding an EPIYA they can be denominated as EPIYA-A, EPIYA-B, EPIYA-C and EPIYA-D (Hatakeyama, 2014). Most *cagA* positive *H. pylori* strains have one -A and -B EPIYA motifs. There are regional differences; in East Asian strains, the EPIYA-D motif is present while in Western countries the EPIYA-C motif is present, the latter varying in number from one to three repeats. Isolates with a higher number of EPIYA-C motifs present more phosphorylation sites and are more virulent than those with fewer repeats (Ferreira et al., 2012; Miura et al., 2009).

Many studies have linked the presence of *cagA* with a risk for peptic ulcer disease and gastric cancer development, demonstrating the oncogenic potential of this protein (Matozaki et al., 2009; Murata-kamiya, 2011; Suzuki et al., 2015). The mere presence of *H. pylori cagA*+ puts an individual at a higher risk for upper gastrointestinal diseases. For this reason, we investigated the prevalence of various genotypes of this virulence factor among consenting Native American clients from an endoscopy clinic in the Navajo Nation.

The second virulence factor of interest is the *vacA* gene. This gene is transcribed and translated into a 140 kDa long precursor that will later be broken down to an 88 kDa toxin. Subsequently, it undergos proteolytic cleavage into two amino-terminal portions, p33 containing the signal region and p55 containing the middle region (Jones et al., 2010).

The *vacA* gene is present in almost all *H. pylori* strains and its name comes from its primary effect, the ability to induce vacuolation in the host cell (Shiota et al., 2014). With recent research, many other *vacA* toxin activities have been unveiled, including, but not limited to, anion channel development with the outflow of ions such as bicarbonate and urea, apoptosis, mitochondrial distress, inhibition of immune cells, disturbance of phagosome maturation and cellular pathways (Foegeding et al., n.d.; Haas, 2016).

With this toxin present and produced by most *H. pylori*, it would be expected that most cases would lead to acute severe symptoms and probably not be a chronic infection, but that's not what's seen. Aside from environmental, host and other *H. pylori* factors, this can be explained by the heterogeneity of the *vacA* gene. Two regions are of interest for this study for their heterogeneity and toxicity, they are the signal and middle region.

The region located in the amino-terminal signal peptide encompasses the family s1 and s2. The s2 form cannot produce an anion channel nor cause vacuolation in gastric epithelial cells, while the s1 can. Due to its hydrophobic nature, the s1 can be more easily inserted in the cell than s2 allowing it to act on many cellular pathways, proving once more to be more virulent (Clain et al., 2001; Letley et al., 1999).

The middle region of the *vacA* toxin can have the m1 or m2 variation. The "m" region, in general, impacts host cell tropism and m1 is toxic to a broader scope of host

cells. Some studies have associated the *vacA* m1 with the development of peptic ulcers in many populations (Figueiredo et al., 1998; Sahara et al., 2012).

The combination of the "s" and "m" regions can lead to different *vacA* genotypes, with the s1m1 having the highest toxicity while s2m2 is not toxic. The s1m2 form is of intermediate toxicity and s2m1 is not found very frequently (McClain et al., 2017; Miehlke et al., 2000). The varying degrees of the *vacA* genotype and its toxic consequences to the host led us to investigate what are the common genotypes in *H. pylori* infecting Native Americans living in Northern Arizona, since this population is at a higher risk for gastric cancer.

The association between the *cagA* and *vacA* is of interest in our study since they can synergically act to alter cell shape, the immune cell reaction to infection and triggering oncogenic pathways. But they can also antagonize each other in apoptotic processes on host cells and in immune modulation (Yokoyama et al., 2005). Therefore, this study aimed to investigate the *cagA* and *vacA* genotype of *Helicobacter pylori* strains infecting Native Americans living in Northern Arizona. We hypothesized that *H. pylori*-positive samples would have the *cagA* gene present and carry the more virulent version of the *vacA* gene (s1m1).

# Materials and Methods

## Sample collection

A total of 210 Native American participants undergoing routine esophagogastroduodenoscopy (EGD) were recruited to participate in this study. This project was approved by the University of Arizona Institutional Review Board (IRB) under protocol number 1610912650 and the Navajo Nation IRB with protocol number NNR-16-263T. The EGD took place at the Winslow Indian Health Care Clinic, a 638 organization, in Winslow (35°01'56.5"N 110°42'46.9"W) in Arizona, USA.

Before the procedure, the study was described, and participants asked to participate by signing the study consent form and were asked eligibility questions: tribal affiliation and age. Additional medical information, including, height and weight, tissue heath of the esophagus, stomach and duodenum, and a medical history were abstracted from the medical record per the study protocol. During the EGD, two additional pin-sized gastric biopsies were collected from consented patient for this study. One from the fundus (F) and the other from the antrum (A) of the stomach. These biopsies were stored in RNAlater and stored at 4$^{o}$ C for no more than 4 days until transported to Flagstaff where they were frozen at -20$^{o}$ C until they were processed.

## DNA extraction

DNA was extracted from the biopsies using the FastDNA® Spin Kit manufactured by MP Biomedicals, LLC. The biopsies were transferred to a 1.5 mL lysing matrix A tube

containing 1 mL of CLS-TC buffer and homogenized at 5.0 m/s for 40 seconds for 4 cycles with 180 seconds of rest. Then centrifuged at 14,000 x g for 10 min, so that the supernatant could be transferred to a 2.0 ml microcentrifuge tube with an equal amount of binding matrix. The tube was incubated under gentle agitation for 10 minutes at room temperature. The mixture was transferred to a SPIN filter and centrifuged at 14,000 x g for 1 minute. Following that a SEWS-M buffer was added to the SPIN filter to resuspend the pellet and centrifuged under the same conditions. The DNA was eluted by the addition of 100 $\mu$L of DES onto the Spin filter and incubated at 55°C for 5 min. A final centrifugation at 14,000 x g for 1 min was done to collect the pure DNA that was then stored at -20°C.

## Polymerase chain reaction (PCR)

### Screening for *Helicobacter pylori* presence

All samples were initially screened for the presence of *Helicobacter pylori* using a quantitative polymerase chain reaction (PCR) method. The primers used were able to detect the 16S ribosomal DNA (Table 1) of *H. pylori*. The samples were submitted to the following protocol: 94° C for 3 minutes then 35 cycles of 94° C for 40 seconds, 61° C for 40s, and 72° C for 40s, and finally a final step at 72° C for 3 min. The melting temperature was also recorded following an increment of 0.5° C every 0.5s from 65° C to 95° C. The samples were compared to a positive control included in every run and checked for contamination with a negative control using just distilled water. Samples were considered positive with amplification within the 35 cycles and Relative Fluorescence Units (RFU) above 1500 and with melting temperature equal to the positive control at 83° C ± 1°C.

Positivity was reevaluated by running the qPCR product a 1% Agarose gel for 30 minutes at 62V. Positive bands were observed within 300bp (Figure 1).

## Screening for the presence of *cagA*, *vacA* s- and m- region

Samples positive for *H. pylori* were then assessed for the cytotoxin-associated gene A DNA. The *cagA* primers used are listed in Table 1. The quantitative PCR protocol used included a first step of 94°C for 3 min, then 35 cycles of 94°C for 40 seconds, followed by 52°C for 40s, then 72°C for 40s. A last step of 72°C for 3 min was then included. The melting temperature was assessed with increments of 0.5°C for 0.5s from 65°C to 95°C. Samples that amplified within the 35 cycles and had a Relative Fluorescence Units (RFU) above 1000 and with melting temperature similar to the positive control (78.5°C ± 1°C) were considered positive. The *H. pylori* strain 26695 (ATCC 700392) is *cagA*+ and was used as a reference strain/positive control.

Positivity for *H. pylori* led to the investigation of the signal region of the vacuolating cytotoxin A. The primers (Table 1) do not distinguish between the s1 and s2 alleles. The quantitative PCR had a first step of 94°C for 4 min, with 35 cycles of 94°C for 40 s, 50°C for 40s, and 72°C for 45s. The last step was 4 min at 72°C. The melting temperature was assessed with increments of 0.5°C for 0.5s from 65°C to 95°C. Amplifications with an RFU above 1000 and melting temperature of 82°C ± 1°C were considered positive. The reference isolate was *H. pylori* strain 26695 (*vacA* s1).

Samples positive for *H. pylori* were also evaluated for the presence of the mid-region of the vacuolating cytotoxin. The primers were designed (Table 1) to identify the

mid-region, but not to distinguish between the alleles m1 or m2. A quantitative PCR was then performed initiated by 5 min step at 94°C, followed by 35 cycles of 94°C for 40s, 50°C for 40s, and 72°C for 50s. The final step was 6 min at 72°C. The melting temperature was evaluated using increments of 0.5°C for 0.5s from 65°C to 95°C. If a sample had an RFU above 1000 and with a melting temperature of 77.5°C ± 1°C and 80.5°C ± 1°C, it was then considered positive. *H. pylori* strain 26695 expresses *vacA* m1 allele and was included as a reference strain.

## PCR product purification

Samples with amplification for any of the virulence factors tested were then processed to obtain a purified PCR product for subsequent sequencing. The purification process utilized the DNA Clean & Concentrator™-5 manufactured by Zymo Research. After the PCR run, the 20 $\mu$L in the well was mixed with 100 $\mu$L of the DNA Binding. Later, the content was transferred to a SPIN filter and centrifuged at 13,500 x g for 30 seconds. 200 $\mu$L of the wash buffer was added and centrifuged under the same conditions previously mentioned. This step was repeated before the addition of 12.5 $\mu$L of the DNA elution. The purified DNA was collected after centrifugation at 13,500 x g for 1 minute and stored at -20°C. The purified nucleic acid was quantified using the Nanodrop™ 2000c Spectrophotometer (Thermo Scientific™).

## Sequencing and data analysis

The purified qPCR product was sent for Sanger sequencing. All of the samples were adjusted to a concentration of 20 $\mu$g/$\mu$L and 5 $\mu$L of the corresponding primer was added to the well in a 96 well plate. The plate was shipped to Genewiz®.

All sequenced samples were evaluated for quality based on the confidence of each nucleotide peak in the electropherogram. Peaks needed to be long, not overlap, evenly spaced, and minimal background noise. The program FinchTV© version 1.5.0 was used to visualize and copy the sequences. For the *cagA* analysis these sequences were retrieved and subsequently translated into the corresponding protein using the "The sequence manipulation suite" (Stothard, 2000). Translated sequences were then aligned using the multiple sequence comparison by log-expectation (MUSCLE) algorithm from within the MEGA X program (Tamura et al., n.d.) and specific alleles were identified. As for the s- and m- alleles of the *vacA* gene the sequences were aligned using the nucleotides.

## *cagA* analysis

Once all sequences were aligned with MUSCLE protein alignment along with reference sequences, it was possible to determine the number of phosphorylation sites. These sites are a sequence of the following amino acids: glutamic acid, proline, isoleucine, tyrosine, and alanine. This sequence (Glu-Pro-Ile-Tyr-Ala) is called the EPIYA motif. These motifs can have different patterns potentially exhibiting EPIYA-A, EPIYA-B, and multiple EPIYA-Cs. That's determined using the following pattern for the EPIYA-A:

EPIYA(K/Q)VNKKK(A/T)GQ. For EPIYA-B, the sequence is EPIY(A/T)QVAKKV(N/T)(A/Q)KI. And finally, EPIYATIDDLGG(P/S) is EPIYA-C; repetitions of C-motif are common and expected.

The number of EPIYA motifs and their patterns were recorded for each sample. The reference sequences used were *H. pylori* 26695 (EPIYA-ABC), FJ 915841.1 (ABC), OK 111 (ABCC), NCTC 11637 (ABCCC). We also included strains with duplications or deletion of EPIYA-A such as FJ 915912.1 (AABC) and FJ 915878.1 (BC). Lastly, the Eastern Asian strain F32 (ABD).

### *vacA* m-region analysis

The vacuolating cytotoxin A may display two different alleles for the mid-region. These alleles are called m1 or m2. The gene sequence was aligned using MUSCLE alignment with known reference strains and manually compared, then the allele for each sample was determined. The lack of a mid-region is possible and samples that did not fall within an m1 or m2 region were excluded. For reference strains with the m1 allele, we used *H. pylori* 26695 and J99. For the m2 allele, we included Tx30a.

### *vacA* s-region analysis

The signal region of the vacuolating cytotoxin A gene can be expressed into alleles s1 or s2. To determine its allelic expression, reference strains were aligned using MUSCLE with our samples. Bacteria with no definitive pattern for s1 or s2 allele were not

accounted for. As reference sequences we used the Western 26696 (s1), J99 (s1) and Tx30a (s2).

## Phylogenetic analysis

A phylogenetic tree for the *cagA*, *vacA* m- and s-region was constructed using the maximum-likelihood method with the assist of the MEGA X program. The reference strains for each of the virulence factors used in the alignment were incorporated into the trees to add confidence to the analysis.

## Epidemiological findings

During the esophagogastroduodenoscopy (EGD), physicians evaluated and categorized the findings. For the esophagus, the categories were normal mucosa, inflammation (esophagitis), and Barrett's esophagus. The stomach findings were grouped into five categories: 1) normal stomach mucosa, 2) inflammation (gastritis), 3) gastric atrophy and fundal gastritis, 4) gastritis with intestinal metaplasia without dysplasia, and 5) erosive and reactive gastropathy. The duodenum was the last observed during EGD. Its findings were classified into normal duodenum and duodenitis.

## Statistical analysis

All statistical analyses were done using the open-source software R and RStudio version 1.2.1335. Chi-square and Fisher's exact tests were used for analysis between the

virulence factors and gender and the epidemiological findings in the esophagus, stomach, and duodenum.

# Results

## Determining the prevalence of *Helicobacter pylori* infection

A total of 65 patients, out of the 210, tested positive for the presence of *H. pylori* using a qPCR method. The proportion of the population testing positive was 30.95%.

## Prevalence of the *cagA* gene and its genotypes

The *cagA* gene can be present or absent in *H. pylori*. Of the 65 positive samples, 29 (44.62%) the *cagA* gene was present and was absent in 36 (55.38%).

The *cagA* gene can have different genotypes depending on the number of EPIYA motifs present. We sequenced the *cagA* gene of the 29 positive samples and aligned using the multiple sequence comparison by log-expectation (MUSCLE) algorithm on MEGA X software.

Most of the samples expressed the EPIYA-A motif. We found the following EPIYA motifs among the *cagA* positive samples: ABC, ABCC and AAABC. There was evidence of co-infection with multiple ABC/ABCC and ABC/BC motifs. There were no ABCCC and we did not find the eastern *cagA* ABD type. The frequency of each genotype is listed in Table 2.

## *vacA* s- and m- allele and *vacA* genotype

We ran a qPCR for the signal region of the *vacA* gene in all the positive samples. One of them did not amplify. All of the others amplified successfully and were sequenced.

The alleles s1 and s2 were determined by DNA alignment using reference strains. We found cases of multiple infections (n = 5) with individuals infected with both s1 and s2 strains (Table 3). 62.5% of the samples were only s1 and 29.6% were only s2.

The middle region of the *vacA* gene can present the m1 or m2 allele. However, we found 12 *H. pylori*-positive samples for which *vacA* m- did not amplify using a qPCR method. The remaining were sequenced accordingly and 3 cases of co-infection with m1 and m2 were identified (Table 4). We observed 66% of the samples were only m1 and 28.3% were only m2.

We genotyped the *vacA* gene in terms of s- and m-region, obtaining the following combinations: s1m1, s1m2, s2m1 and s2m2. The most common among these was s1m1 with 48% of the cases. The samples with multiple infection were categorized as "Co-infection" (Table 5).

## *Helicobacter pylori* genotype

After assessing each gene individually, we put together the genotype for *H. pylori* based on the *cagA* and *vacA* presence. The most common genotype was the ABC/s1m1 (32%). Similar to what we did with cases of mixed infection, these were included in the "Co-infection" category with a frequency of almost 20%. The samples with incomplete *vacA* gene were excluded (Table 6).

## Epidemiological findings

During EGD, the surgeons and nurses gathered information on the esophagus, stomach and duodenum of each patient. When evaluating the esophagus, they found that 75% of the cases had a normal mucosa with only 25% of them with clinical relevance.

As for the stomach, 75% of the patients were diagnosed with gastritis. There were also individual cases of many of the other lesions like gastric atrophy, intestinal metaplasia, erosive and reactive gastropathy. These findings are in general more serious than just gastritis and can potentially lead to the development of cancerous tissue (Correa et al., 1975), therefore, they were all grouped as “Pre-cancerous” findings. None of the patients were diagnosed with adenocarcinoma or MALT lymphoma.

The overwhelming majority of patients (92%) had a normal duodenal mucosa without any significant clinical findings.

## Statistical analysis

At the end of the data collection, each *H.* pylori-positive patient had the bacterium genotyped for the *cagA* and *vacA* (s_m_) gene and we included the clinical observations from the EGDs. A $x^2$-test was performed to assess any correlation.

### *cagA* status and genotype

The first analysis consisted of checking if the presence of *cagA* positive strains would lead to a more severe case of disease. For the esophagus, there was no significant difference in disease manifestation between positive and negative *cagA* status with a p-value of 0.2778. As for the stomach, there was no significance as well and the p-value

was 0.0976. We did not find any correlation when checking the duodenum clinical reports either. The p-value was >0.9999 (Figure 2)

The second analysis consisted of checking the number of EPIYA motifs of the *cagA*-positive samples. We ran a similar analysis as in the previous one. Only the esophagus evaluation showed a significant difference with a p-value of 0.0202. The stomach and duodenum assessments had a p-value of 0.9170 and 0.4496, respectively (Figure 3).

### *vacA* s- and m- region and *vacA* genotype

For the *vacA* gene, the first analysis was with the s-region. No significant difference was observed among the three sites and s-allele. The p-values for the esophagus, stomach and duodenum were 0.2755, 0.1351 and 0.3537, respectively (Figure 4).

As for the m-region of the *vacA* gene, only in the stomach we could detect a difference between the clinical observations and the alleles present (p-value = 0.0181). Neither esophagus (p-value = 0.8302) nor duodenum (p-value = 0.8502) had significant variations (Figure 5).

In possession of the s- and m-regions allele configuration, we put together the *vacA* genotype for these regions and correlated with the EGD findings. With a p-value of 0.2465 for the esophagus and 0.7567 for the duodenum, we found no association. For the stomach, however, all the pre-cancerous lesions were found only with the s1m1 genotype (p-value = 0.0072, Figure 6). The cases with normal gastric mucosa were found with the s1m2 and s2m2 genotypes being of intermediate toxicity and non-toxic, respectively (McClain et al., 2017) .

### *Helicobacter pylori* genotype

Finally we combined the *cagA* genotype with the *vacA* genotype to obtain a single *H. pylori* genotype for every sample and to associate that with the EGD results for each site examined. The esophagus did not show any statistical implication (p-value = 0.5842). The stomach findings with serious lesions were distributed mainly with the *vacA* s1m1 genotype and showed a statistical difference with a p-value of 0.0007. The duodenum did not have any association (p-value = 0.2858) (Figure 7).

## Phylogenetic analysis

Three trees were generated. One with the different EPIYA-motifs for the *cagA* gene that resulted in four groups. Reference strains with EPIYA-ABC motifs were in groups A, Eastern Asian strain F32 (ABD) in Group B, most samples in Group C had EPIYA-ABC motifs and lastly group D had reference strains ABCCC and incomplete FJ915878.1 EPIYA-BC (Figure 8).

Two other cladograms were generated with the *vacA* gene, one for the s-region (s1 or s2) (Figure 9) and another for the m-region (m1 or m2) (Figure 10).

# Discussion

In this study, two gastric biopsies were collected from 210 Native American patients and 30.95% of them had a detectable amount of *Helicobacter pylori* DNA via qPCR. Worldwide, about 50% of the population is infected by this bacterium (Pucułek et al., 2018). However, in developing countries, this is about 85%-95%, while in developed countries this ranges from 30% to 50% (Baj et al., 2020).

At first, the prevalence found here seems to agree with the current data around the globe. But it's important to mention other essential findings. Native Alaskans had a near 70% prevalence rate using urea breath test (UBT) and serology (Miernyk et al., 2018) which contrasts with the rate commonly found in the rest of the country. The general population of the US had a seroprevalence of 27% (Cardenas et al., 2006).

While our results seem to be closer to the general US population, a previous study in Navajo Nation found 56.4% of the households among three communities had at least one individual testing positive for the presence of *H. pylori* in a urea breath test (Harris et al., in prep). One reason for this difference is that many patients that were undergoing EGD had previously been treated for *H. pylori* eradication. Therefore, our test via qPCR could be used by the surgeons as a confirmation that the treatment was successful.

Diseases that affect the upper gastrointestinal tract may have many etiologic factors. Diet, family history, smoking and alcohol consumption, stress and *H. pylori* infection are some of the reasons this can happen (Deng, 2000). When this bacterium colonizes a human's stomach a battle between this organism and the host starts.

Certain *H. pylori* strains can have genetic advantages to overcome the host's immune response and to successfully colonize the stomach. Two of them were explored here because of their proven effects in aiding the bacterium (Ali et al., 2000). They are the cytotoxin-associated gene A (*cagA*) and the vacuolating cytotoxin A gene (*vacA*).

Our intent in this study was to look for an association between the *H. pylori* genotype with disease progression among Native Americans living in Northern Arizona. With the findings from the EGDs, we were able to check if a more virulent strain could lead to more severe findings in the esophagus, stomach and duodenum of this population.

The presence of the *cagA* gene effectively benefits *H. pylori* infection and the resulting protein is oncogenic (Hatakeyama, 2014). Therefore, our first step was to determine the prevalence of the *cagA* gene among the positive samples. 44.62% of them had this virulence factor, while 55.38% did not. We hypothesized that the *cagA* positive samples were more likely to be from patients with a more compromising diagnosis than those that were negative. When we checked the esophageal diseases, we found no association, but only the *cagA*+ samples had a case of Barrett's esophagus; a condition where the lining of the esophagus has been damaged by the gastric acid for a long period, suggesting a chronic infection. There was also no correlation among the stomach findings, however, all the cases with no significant clinical findings were *cagA*-. Finally, no positive association between the *cagA* gene and duodenal diseases was found.

The C-terminal region of *cagA* gene has sites for tyrosine phosphorylation. These sites are the EPIYA motifs that can be denominated EPIYA-A, -B, -C and -D. The East Asian strains contain EPIYA-A, EPIYA-B and EPIYA-D, while the Western strains have

EPIYA-A, EPIYA-B and one to three EPIYA-C. In the West, the more EPIYA-C motifs a strain has the more severe are the clinical manifestations. They tend to progress from gastritis and ulcers to gastric atrophy and gastric carcinoma (Ferreira et al., 2012; Matozaki et al., 2009).

We wanted to know the number and types of EPIYA motifs found among the 29 *cagA* positive samples found. Not surprisingly, there was no EPIYA-D motif among these samples. The most common genotype was the EPIYA ABC (22/29), and there was no ABCCC. Similar to the previous analysis, we explored the association between the number of EPIYA motifs and the EGD diagnosis for each organ evaluated.

The association between esophageal diseases and the number of EPIYA-motifs was statistically significant ($p = 0.0202$). All patients with esophagitis that were infected by *cagA* positive strains had *H. pylori* EPIYA-ABC. Interestingly, among the *H.* pylori-positive samples, all with Barrett's esophagus were in patients with multiple infections (Figure 3). These cases presented a varying number of EPIYA-C motifs. When co-infection happens, recombination between different strains can occur, which can prolong the infection, giving the bacterium the chance for further recombination that could lead to an aggravated disease (Cao et al., 2015). This is accordance with Barrett's esophagus diagnosis since it requires a sustained infection to develop.

The analysis of the stomach findings with the *H. pylori cagA* genotype was not statistically significant. However, the only case with the unusual multiple EPIYA-A motifs (AAABC) was diagnosed with gastritis and intestinal metaplasia without dysplasia, a more advanced stage in gastric disease development. Similar diagnoses were only found with

ABC motifs. Most of the *cagA*+ cases presented a normal duodenal mucosa, and not unexpectedly, there was not any significant association with the number of EPIYA motifs.

We proceeded, then, with the next virulence factor of interest, the *vacA* gene. Almost all the *H. pylori* strains have this gene, however, there are two alleles (*s* and *m*) that play a role in the resulting protein's ability to vacuolate infected cells (McClain et al., 2017). We wanted to know if individual alleles were associated with an increased risk for disease severity.

The signal region of the *vacA* gene has two families, s1 and s2. The s1 has an increased capability of forming vacuoles than s2 and is associated with cytotoxin production (Atherton et al., 1995). Knowing the s1 allele is more virulent than the s2, we checked for any correlation with disease progression.

In the same manner, as our earlier analysis, we tested the three organs of interest. Although none of them showed any statistical significance we observed a trend with the gastric analysis. Most pre-cancerous conditions found in the stomach were among those patients with strains from the s1 family, with a few among the co-infection cases, and none among the s2 allele. The huge number of gastritis cases could be a factor that interfered with the statistical test here, but the trend is clear (Figure 4).

The mid-region of the *vacA* gene has the alleles m1 and m2. The m1 family has an increased cytotoxin production and bigger vacuolation power than m2 (McClain et al., 2017). Our analysis found significance only in the gastric findings ($p = 0.0181$). It's clear as shown in Figure 5 that all the pre-cancerous cases were typed as m1. In addition to the clear case where m1 seems to be more virulent than its counterpart, all of the cases

where there was a normal stomach mucosa were typed m2. Interestingly, all the patients with multiple infections were diagnosed with gastritis.

Now that we evaluated the individual contribution of each allele, we checked the *vacA* gene genotype including both regions (s,m). This will most likely be the final form of the toxin once it's been transcribed and translated. There are four possible combinations. s1m1 genotype is known to be more virulent with a high vacuolating activity while s2m2 is considered non-toxic. The s1m2 strain is of intermediate toxicity and the s2m1 type is not commonly found (Basso et al., 2008).

The most common genotype found among the biopsies screened was s1m1 and while s2m1 is infrequent and alleged to be negatively selected (Letley et al., 1999), it was the second most frequent genotype among our samples. There was no association between esophageal and duodenal findings in terms of the *vacA* genotype even though s1m1 form was the most abundant in both cases.

The stomach findings were statistically significant ($p = 0.0072$) when correlated with the *vacA* genotype. This is clear and consistent with the literature since all severe cases of gastric diseases were typed s1m1 (Figure 6). Unsurprisingly, most of the normal stomach results were among patients with s2m2 genotypes, the non-toxic version of the resulting vacA toxin.

Although the s2m1 arrangement has been reported before as potentially disadvantageous to the bacterium since it's less frequent, Keikha et al.(2020) found a 15.7% abundance of this genotype within an Iranian population. Furthermore, they

directly linked the s2m1 genotype to a higher risk for developing peptic ulcer disease in Iranian people.

Virulence factors do not usually lead to an effector response just by themselves, rather they work synergistically. Having genotyped each case for *cagA* and *vacA* genes, we could assess if these virulence factors were aiding each other in colonizing the stomach, establishing an infection and with time furthering any upper gastrointestinal disease.

The most common genotype was *cagA* ABC together with *vacA* s1m1. The second most frequent were the cases with multiple strain infections. Even though we did not find any association between these genotypes and esophageal diseases, the number of patients diagnosed with Barrett's esophagus had multiple strains infecting them. Similar to findings with the *cagA* alone, where this diagnosis was only found in co-infections, we believe they play a role in extending the infection and symptoms of gastric reflux that can damage the esophagus mucosa. Since we could not find any significance when checking just the *vacA* gene association, this is more likely to be an effect with the varying numbers of EPIYA-C motifs alone.

The positive association ($p = 0.0007$) found between gastric findings and the *H. pylori* genotype is most likely explained by the cases where the patient is infected with a *vacA* s1m1 genotype (Figure 7). All gastric pre-cancerous conditions had this *vacA* genotype, regardless of the *cagA* gene since they were found both in *cagA*+ and *cagA*- patients. The potential of the s1m1 genotype in exacerbating the infection is further demonstrated in the case with the *cagA* AAABC genotype. The EPIYA-A motif is

phosphorylated by the Abl kinase and can also bind the Csk kinase. This one will phosphorylate and inhibit Src kinases that usually act on EPIYA-C motifs (Hatakeyama, 2004). Therefore, Csk exerts negative feedback, diminishing *H. pylori*'s successful establishment in the gastric mucosa. With many EPIYA-A repeats, this process of restraining the infection is most likely aggravated, so the case with the AAABC genotype seen here is probably harmful to the bacterium and yet this patient has a pre-cancerous condition. This led us to assume that the seriousness of the infection can be explained by the more toxigenic *vacA* s1m1 genotype found here. As mentioned before no patients with normal gastric results were infected with the *H. pylori cagA*+ and they also had the m2 allele to downplay the infection.

We found no association between *cagA*/*vacA* genotype and the duodenal findings. This was expected since most individuals did not have any duodenal alterations during the EGD exam.

Multiple infections are defined by the presence of different bacterial strains infecting the same individual. Based on the *cagA* and *vacA* genotypes, we found a prevalence of nearly 20% (10/52) of co-infection with multiple *H. pylori* strains. Ben Mansour et al., 2016 linked a higher incidence of these cases to developing countries in contrast to developed ones, using participants from France and Tunisia, and found that 48% of Tunisians and 5% of French patients were simultaneously infected with multiple strains of *H. pylori*.

The relatively higher prevalence of multiple infections among Native Americans suggests that cases of chronic infection and reinfection happen. This can be explained

by the sanitary conditions most of these participants live in, which are comparable to underdeveloped countries. Experiences range from homelessness to overcrowding, plumbing problems, limited access to clean water and improper sanitary structure (NDWR, 2011; Riley, 2021).

One possible outcome of these multiple infections is an increase in bacterial recombination sustaining a genetic variation among these strains which could confer an advantage to the bacterium (Krebes et al., 2014). In sum, the *H. pylori*-associated diseases in this population could be more severe if the infection is left untreated.

To strengthen this work, future studies such as an antibiotic resistance assay to assess the presence of resistant subclones and strains where bacterial recombination and the presence of mixed infections among Native Americans can be further evaluated.

In this study, we were able to observe that Native Americans living in Northern Arizona were infected with most of the possible genotypes for two virulence factors of the *H. pylori* bacterium. None of the participants were diagnosed with gastric cancer, however, other clinical conditions could potentially lead to cancer were observed. Among these samples, the more virulent genotype of *cagA* and *vacA* genotypes were associated with a more severe diagnosis. However, other reasons could lead to this outcome like other virulence factors that *H. pylori* can present and that it would be of relevance to accurately affirm the part this bacterium plays. Further screening of the intermediate region of the *vacA* gene and virulence factors such as *OipA*, *DupA* and *BabA* are strongly suggested. Another recommendation is acquiring more data on the population studied such as diet habits, age and weight.

# References:


Ali, A., Ashour, R., Gusmão, V. R. De, Magalhães, P. P., Collares, G. B., Mendes, E. N., Maria, D., Queiroz, D. M., Sales, A., & Carvalho, T. (2000). Associação entre cagA e alelos do vacA de Helicobacter pylori e úlcera duodenal em crianças no Brasil VacA alleles , cagA , and duodenal ulcer in children in Brazil. *Jornal Brasileiro de Parologia e Medicina Laboratorial*, *38*(2), 79–85.

Ansari, S., & Yamaoka, Y. (2017). *Survival of Helicobacter pylori in gastric acidic territory*. 1–13. https://doi.org/10.1111/hel.12386

Ansari, S., & Yamaoka, Y. (2019). Helicobacter pylori virulence factors exploiting gastric colonization and its pathogenicity. *Toxins*, *11*(11), 1–26. https://doi.org/10.3390/toxins11110677

Argent, R. H., Thomas, R. J., Letley, D. P., Rittig, M. G., Hardie, K. R., & Atherton, J. C. (2016). *Functional association between the Helicobacter pylori virulence factors VacA and CagA*. *May*, 145–150. https://doi.org/10.1099/jmm.0.47465-0

Atherton, J. C., Cao, P., Peek, R. M., Tummuru, M. K. R., Blaser, M. J., & Cover, T. L. (1995). Mosaicism in vacuolating cytotoxin alleles of helicobacter pylori. Association of specific vacA types with cytotoxin production and peptic ulceration. *Journal of Biological Chemistry*, *270*(30), 17771–17777. https://doi.org/10.1074/jbc.270.30.17771

Baj, J., Forma, A., Sitarz, M., Portincasa, P., Garruti, G., Krasowska, D., & Maciejewski, R. (2020). Helicobacter pylori Virulence Factors-Mechanisms of Bacterial Pathogenicity in the Gastric Microenvironment. *Cells*, *10*(1), 1–37.

https://doi.org/10.3390/cells10010027

Basso, D., Zambon, C. F., Letley, D. P., Stranges, A., Marchet, A., Rhead, J. L., Schiavon, S., Guariso, G., Ceroti, M., Nitti, D., Rugge, M., Plebani, M., & Atherton, J. C. (2008). Clinical Relevance of Helicobacter pylori cagA and vacA Gene Polymorphisms. *Gastroenterology*, *135*(1), 91–99. https://doi.org/10.1053/j.gastro.2008.03.041

Ben Mansour, K., Fendri, C., Battikh, H., Garnier, M., Zribi, M., Jlizi, A., & Burucoa, C. (2016). Multiple and mixed Helicobacter pylori infections: Comparison of two epidemiological situations in Tunisia and France. *Infection, Genetics and Evolution*, *37*, 43–48. https://doi.org/10.1016/j.meegid.2015.10.028

Bui, D., Brown, H. E., Harris, R. B., & Oren, E. (2016). *Serologic Evidence for Fecal – Oral Transmission of Helicobacter pylori*. *94*(1), 82–88. https://doi.org/10.4269/ajtmh.15-0297

Cao, Q., Didelot, X., Wu, Z., Li, Z., He, L., Li, Y., Ni, M., You, Y., Lin, X., Li, Z., Gong, Y., Zheng, M., Zhang, M., Liu, J., Wang, W., Bo, X., Falush, D., Wang, S., & Zhang, J. (2015). Progressive genomic convergence of two Helicobacter pylori strains during mixed infection of a patient with chronic gastritis. *Gut*, *64*(4), 554–561. https://doi.org/10.1136/gutjnl-2014-307345

Cardenas, V. M., Mulla, Z. D., Ortiz, M., & Graham, D. Y. (2006). Iron deficiency and Helicobacter pylori infection in the United States. *American Journal of Epidemiology*, *163*(2), 127–134. https://doi.org/10.1093/aje/kwj018

Chang, C., Kuo, W., Chen, Y., & Perng, C. (2016). *Fragmentation of CagA Reduces*

*Hummingbird Phenotype Induction by Helicobactor pylori*. 1–16. https://doi.org/10.1371/journal.pone.0150061

Clain, M. S. M. C., Cao, P., Iwamoto, H., Vinion-dubiel, A. D., Szabo, G., Shao, Z., & Cover, T. L. (2001). *A 12-Amino-Acid Segment , Present in Type s2 but Not Type s1 Helicobacter pylori VacA Proteins , Abolishes Cytotoxin Activity and Alters Membrane Channel Formation*. *183*(22), 6499–6508. https://doi.org/10.1128/JB.183.22.6499

Correa, P., Haenszel, W., Cuello, C., Tannenbaum, S., & Archer, M. (1975). A MODEL FOR GASTRIC CANCER EPIDEMIOLOGY. *The Lancet*, *306*(7924), 58–60. https://doi.org/10.1016/S0140-6736(75)90498-5

Deng, D. J. (2000). Progress of gastric cancer etiology: N-nitrosamides in the 1990s. *World Journal of Gastroenterology*, *6*(4), 613–618.

Ferreira, R. M., Machado, J. C., Leite, M., Carneiro, F., & Figueiredo, C. (2012). The number of Helicobacter pylori CagA EPIYA C tyrosine phosphorylation motifs influences the pattern of gastritis and the development of gastric carcinoma. *Histopathology*, *60*(6), 992–998. https://doi.org/10.1111/j.1365-2559.2012.04190.x

Figueiredo, U., Sanna, R., Plaisier, A., & Doorn, L. J. A. N. V. A. N. (1998). *Clinical Relevance of the*. 58–66.

Foegeding, N. J., Caston, R. R., Mcclain, M. S., Ohi, M. D., & Cover, T. L. (n.d.). *An Overview of Helicobacter pylori VacA Toxin Biology*. 1–21. https://doi.org/10.3390/toxins8060173

Haas, C. U. and R. (2016). *VacA's Induction of VacA-Containing Vacuoles (VCVs) and*

*Their Immunomodulatory Activities on Human T Cells*. https://doi.org/10.3390/toxins8060190

Hatakeyama, M. (2004). *ONCOGENIC MECHANISMS OF THE HELICOBACTER PYLORI CagA PROTEIN*. *4*(September), 688–695. https://doi.org/10.1038/nrc1433

Hatakeyama, M. (2014). Helicobacter pylori CagA and gastric cancer: A paradigm for hit-and-run carcinogenesis. *Cell Host and Microbe*, *15*(3), 306–316. https://doi.org/10.1016/j.chom.2014.02.008

Jones, K. R., Whitmire, J. M., & Merrell, D. S. (2010). A tale of two toxins: Helicobacter pylori CagA and VacA modulate host pathways that impact disease. *Frontiers in Microbiology*, *1*(NOV), 1–17. https://doi.org/10.3389/fmicb.2010.00115

Keikha, M., Ali-Hassanzadeh, M., Karbalaei, M., & Karbalaei, M. (2020). Association of Helicobacter pylori vacA genotypes and peptic ulcer in Iranian population: A systematic review and meta-analysis. *BMC Gastroenterology*, *20*(1), 1–11. https://doi.org/10.1186/s12876-020-01406-9

Kong, F., & Singh, R. P. (, Reviews, C., & Science, F. (2008). *Disintegration of Solid Foods in Human Stomach*. *73*(5), 67–80. https://doi.org/10.1111/j.1750-3841.2008.00766.x

Krebes, J., Didelot, X., Kennemann, L., & Suerbaum, S. (2014). International Journal of Medical Microbiology Bidirectional genomic exchange between Helicobacter pylori strains from a family in Coventry , United Kingdom. *International Journal of Medical Microbiology*, *304*(8), 1135–1146. https://doi.org/10.1016/j.ijmm.2014.08.007

Kusters, J. G., Van Vliet, A. H. M., & Kuipers, E. J. (2006). Pathogenesis of Helicobacter

pylori infection. *Clinical Microbiology Reviews*, *19*(3), 449–490. https://doi.org/10.1128/CMR.00054-05

Letley, D. P., Lastovica, A., Louw, J. A., Hawkey, C. J., & Atherton, J. C. (1999). Allelic diversity of the Helicobacter priori vacuolating cytotoxin gene in South Africa: Rarity of the vacA s1a genotype and natural occurrence of an s2/m1 allele. *Journal of Clinical Microbiology*, *37*(4), 1203–1205. https://doi.org/10.1128/jcm.37.4.1203-1205.1999

Ley, D., Paster, B. J., Genetics, M., Fenway, T., Fox, J. G., Medicine, C., Ma, C., Vandamme, P., Gent, U., Wetenschappen, F., & Genetica, M. (2015). *Helicobacter*. https://doi.org/10.1002/9781118960608.gbm01073.

Lu, G. (2002). *Corrosion resistance of ternary Ni Ã P based alloys in sulfuric acid solutions*. *47*, 2969–2979.

Marshall, B. (2002). *Helicobacter pylori : 20 years on*. *2*(2), 147–152.

Matozaki, T., Murata, Y., Saito, Y., Okazawa, H., & Ohnishi, H. (2009). Protein tyrosine phosphatase SHP-2: A proto-oncogene product that promotes Ras activation. *Cancer Science*, *100*(10), 1786–1793. https://doi.org/10.1111/j.1349-7006.2009.01257.x

McClain, M. S., Beckett, A. C., & Cover, T. L. (2017). Helicobacter pylori vacuolating toxin and gastric cancer. *Toxins*, *9*(10), 23–25. https://doi.org/10.3390/toxins9100316

Mendoza-Cantú, A., Urrutia-Baca, V. H., Urbina-Ríos, C. S., De La Garza-Ramos, M. A., García-Martínez, M. E., & Torre-Martínez, H. H. H. (2017). Prevalence of

Helicobacter pylori vacA Genotypes and cagA Gene in Dental Plaque of Asymptomatic Mexican Children. *BioMed Research International*, *2017*. https://doi.org/10.1155/2017/4923640

Miehlke, S., Kirsch, C., Agha-Amiri, K., Günther, T., Lehn, N., Malfertheiner, P., Stolte, M., Ehninger, G., & Bayerdörffer, E. (2000). The Helicobacter pylori vacA sl, ml genotype and cagA is associated with gastric carcinoma in Germany. *International Journal of Cancer*, *87*(3), 322–327. https://doi.org/10.1002/1097-0215(20000801)87:3<322::AID-IJC3>3.0.CO;2-M

Miernyk, K. M., Bulkow, L. R., Gold, B. D., Bruce, M. G., Hurlburt, D. H., Griffin, P. M., Swerdlow, D. L., Cook, K., Hennessy, T. W., & Parkinson, A. J. (2018). Prevalence of Helicobacter pylori among Alaskans: Factors associated with infection and comparison of urea breath test and anti-Helicobacter pylori IgG antibodies. *Helicobacter*, *23*(3), 1–8. https://doi.org/10.1111/hel.12482

Miura, M., Ohnishi, N., Tanaka, S., Yanagiya, K., & Hatakeyama, M. (2009). *Differential oncogenic potential of geographically distinct Helicobacter pylori CagA isoforms in mice*. *2504*(June), 2497–2504. https://doi.org/10.1002/ijc.24740

Murata-kamiya, N. (2011). Pathophysiological functions of the CagA oncoprotein during infection by Helicobacter pylori. *Microbes and Infection*, *13*(10), 799–807. https://doi.org/10.1016/j.micinf.2011.03.011

NDWR, N. (2011). *DRAFT WATER RESOURCE DEVELOPMENT STRATEGY FOR THE NAVAJO NATION*. *July*.

Nguyen, L. T., Uchida, T., Murakami, K., Fujioka, T., & Moriyama, M. (2008).

*Helicobacter pylori virulence and the diversity of gastric cancer in Asia*. 1445–1453. https://doi.org/10.1099/jmm.0.2008/003160-0

NNDOH. (2013). *Cancer among the Navajo 2005-2013*.

Pucułek, M., Machlowska, J., Wierzbicki, R., Baj, J., Maciejewski, R., & Sitarz, R. (2018). Helicobacter pylori associated factors in the development of gastric cancer with special reference to the early-onset subtype. *Oncotarget*, *9*(57), 31146–31162. https://doi.org/10.18632/oncotarget.25757

Riley, R., & Dissertation, E. (2021). *Prioritization of Potable Water Infrastructure Investments on the Navajo Nation In the Graduate College*.

Sahara, S., Sugimoto, M., Vilaichone, R., Mahachai, V., & Miyajima, H. (2012). *Role of Helicobacter pylori cagA EPIYA motif and vacA genotypes for the development of gastrointestinal diseases in Southeast Asian countries : a meta-analysis*. https://doi.org/10.1186/1471-2334-12-223

Schlansky, B., & Sonnenberg, A. (2011). Epidemiology of Noncardia Gastric Adenocarcinoma in the United States. *The American Journal of Gastroenterology*, *106*(11), 1978–1985. https://doi.org/10.1038/ajg.2011.213

Shiota, S., Romiko, S., & Yoshio, Y. (2014). The significance of virulence factors in Helicobacter pylori Seiji. *Bone*, *23*(1), 1–7. https://doi.org/10.1111/1751-2980.12054.The

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). *Global Cancer Statistics 2020 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. *71*(3), 209–

249. https://doi.org/10.3322/caac.21660

Suzuki, N., Murata-kamiya, N., Yanagiya, K., & Suda, W. (2015). Mutual reinforcement of inflammation and carcinogenesis by the Helicobacter pylori CagA oncoprotein. *Nature Publishing Group*, *May*, 1–14. https://doi.org/10.1038/srep10024

Tammer, I. N. A., Brandt, S., Hartig, R., König, W., & Backert, S. (2007). *Activation of Abl by Helicobacter pylori : A Novel Kinase for CagA and Crucial Mediator of Host Cell Scattering*. 1309–1319. https://doi.org/10.1053/j.gastro.2007.01.050

Tamura, K., Stecher, G., & Kumar, S. (n.d.). *MBE Brief Communication MEGA11: Molecular Evolutionary Genetics Analysis Version 11*. 1–13.

Uthman, O. A., Jadidi, E., Moradi, T., Uthman, O. A., Jadidi, E., & Moradi, T. (2013). *Socioeconomic position and incidence of gastric cancer : a systematic review and meta- analysis Stable URL : https://www.jstor.org/stable/43281627 Socioeconomic position and incidence of gastric cancer : a systematic review and meta-analysis*. *67*(10), 854–860.

Wiggins, C. L., Espey, D. K., Wingo, P. A., Kaur, J. S., Wilson, R. T., Swan, J., Miller, B. A., Jim, M. A., Kelly, J. J., & Lanier, A. P. (2008). *Cancer Among American Indians and Alaska*. *August*, 1999–2004. https://doi.org/10.1002/cncr.23734

Yokoyama, K., Higashi, H., Ishikawa, S., Fujii, Y., Kondo, S., Kato, H., Azuma, T., Wada, A., Hirayama, T., Aburatani, H., & Hatakeyama, M. (2005). *Functional antagonism between Helicobacter pylori CagA and vacuolating toxin VacA in control of the NFAT signaling pathway in gastric epithelial cells*. *102*(27).

# Tables

**Table 1. Primers used to identify *Helicobacter pylori*-positive samples and to amplify and sequence the genes for *cagA* and *vacA* s- and m-regions.**

| Gene | Primer | Sequence | Tm (°C) | Reference |
|---|---|---|---|---|
| 16S rRNA | | 5' GCTAAGAGATCAGCCTATGTCC 3' | 57 | Chang et al., 2006 |
| | | 5' CAATCAGCGTCAGTAATGTTC 3' | 62 | |
| *cagA* | *cagA*-2-F | 5' GGAACCCTAGTCGGTAATG 3' | 55 | Argent et al., 2005 |
| | *cagA*-4-R | 5' ATCTTTGAGCTTGTCTATCG 3' | 52 | |
| *vacA* m | *vacA*m1-m2-F | 5' CAATCTGTCCAATCAAGCGAG 3' | 54 | Yamoaka et al., 1998 |
| | *vacA*m1-m2-R | 5' GCGTCAAAATAATTCCAAGG 3' | 50 | |
| *vacA* s | VA1s-F | 5' ATGGAAATACAACAACACCAC 3' | 50 | Atherton et al., 1995 |
| | VA1s-R | 5' CTGCTTGAATGCGCCAAAC 3' | 56 | |

**Table 2. Frequency of *cagA Helicobacter pylori* genotypes.**

| ***cagA* genotype** | **Frequency** |
|---|---|
| **ABC** | 22 |
| **ABCC** | 2 |
| **AAABC** | 1 |
| **ABC/ABCC** | 3 |
| **ABC/BC** | 1 |
| ***cagA* negative** | 36 |

**Table 3. *vacA* signal-region allele frequency.**

| ***vacA* s-region allele** | **Frequency n (%)** |
|---|---|
| **s1** | 40 (62.5) |
| **s2** | 19 (29.6) |
| **Co-infection** | 5 (7.9) |

**Table 4. *vacA* mid-region allele frequency.**

| ***vacA* m-region allele** | **Frequency n (%)** |
|---|---|
| **m1** | 35 (66) |
| **m2** | 15 (28.3) |
| **Co-infection** | 3 (5.7) |

**Table 5. *vacA* genotype (s_m_) frequency.**

| ***vacA* genotype** | **Frequency n (%)** |
|---|---|
| **s1m1** | 25 (48) |
| **s1m2** | 7 (13.4) |
| **s2m1** | 8 (15.6) |
| **s2m2** | 6 (11.5) |
| **Co-infection** | 6 (11.5) |

**Table 6. *Helicobacter pylori* genotype (*cagA*/*vacA*) frequency.**

| *H. pylori* genotype | Frequency n (%) |
|---|---|
| ***cagA*-/s1m1** | 6 (11.5) |
| ***cagA*-/s1m2** | 3 (5.7) |
| ***cagA*-/s2m1** | 7 (13.4) |
| ***cagA*-/s2m2** | 6 (11.5) |
| **ABC/s1m1** | 17 (32.7) |
| **ABCC/s1m2** | 2 (3.8) |
| **AAABC/s1m1** | 1 (2.2) |
| **Co-infection** | 10 (19.2) |

**Table 7. Esophagogastroduodenoscopy results and scores in Native Americans living in Northern Arizona.**

| EGD finding | Frequency n (%) |
|---|---|
| **Normal esophagus** | 49 (75.3) |
| **Esophagitis** | 14 (21.5) |
| **Barrett's esophagus** | 2 (3.2) |
| | |
| **Normal stomach** | 4 (6.3) |
| **Gastritis** | 49 (75.3) |
| **Pre-cancerous stomach findings** | 12 (18.4) |
| | |
| **Normal duodenum** | 60 (92.3) |
| **Duodenitis** | 5 (7.7) |

**Table 8. Frequency of *vacA* alleles and genotype and *cagA* status and EPIYA motif determination in relation to esophageal findings in biopsies from a population of Native Americans living in Northern Arizona.**

| | Total | Normal esophagus | Esophagitis | Barrett's esophagus |
|---|---|---|---|---|
| ***vacA* alleles** | | | | |
| *s1* | 40 | 32 | 7 | 1 |
| *s2* | 19 | 14 | 4 | 1 |
| *s1/s2* | 5 | 2 | 3 | 0 |
| *m1* | 35 | 26 | 8 | 1 |
| *m2* | 15 | 11 | 3 | 1 |
| *m1/m2* | 3 | 3 | 0 | 0 |
| ***vacA* genotype** | | | | |
| *s1m1* | 25 | 20 | 5 | 0 |
| *s1m2* | 7 | 6 | 0 | 1 |
| *s2m1* | 8 | 6 | 1 | 1 |
| *s2m2* | 6 | 4 | 2 | 0 |
| *Co-infection* | 6 | 3 | 3 | 0 |
| ***CagA* status** | | | | |
| *Positive* | 29 | 21 | 6 | 2 |
| *Negative* | 36 | 28 | 8 | 0 |
| ***cagA* genotype** | | | | |
| *ABC* | 22 | 16 | 6 | 0 |
| *ABCC* | 2 | 2 | 0 | 0 |
| *AAABC* | 1 | 1 | 0 | 0 |
| *Co-infection* | 4 | 2 | 0 | 2 |
| ***cagA*/*vacA* genotype** | | | | |
| *cagA-/s1m1* | 6 | 5 | 1 | 0 |
| *cagA-/s1m2* | 3 | 3 | 0 | 0 |
| *cagA-/s2m1* | 7 | 6 | 1 | 0 |
| *cagA-/s2m2* | 6 | 4 | 2 | 0 |
| *ABC/s1m1* | 17 | 13 | 4 | 0 |
| *ABCC/s1m2* | 2 | 2 | 0 | 0 |
| *AAABC/s1m1* | 1 | 1 | 0 | 0 |
| *Co-infection* | 10 | 5 | 3 | 2 |

**Table 9. Frequency of *vacA* alleles and genotype and *cagA* status and EPIYA motif determination in relation to gastric findings in biopsies from a population of Native Americans living in Northern Arizona.**

| | Total | Normal stomach | Gastritis | Gastric atrophy & fundal gastritis | Gastritis & intestinal metaplasia without dysplasia | Erosive gastropathy | Reactive gastropathy |
|---|---|---|---|---|---|---|---|
| ***vacA* alleles** | | | | | | | |
| *s1* | 40 | 2 | 27 | 4 | 3 | 3 | 1 |
| *s2* | 19 | 2 | 17 | 0 | 0 | 0 | 0 |
| *s1/s2* | 5 | 0 | 4 | 0 | 1 | 0 | 0 |
| *m1* | 35 | 0 | 27 | 4 | 1 | 3 | 0 |
| *m2* | 15 | 3 | 12 | 0 | 0 | 0 | 0 |
| *m1/m2* | 3 | 0 | 3 | 0 | 0 | 0 | 0 |
| ***vacA* genotype** | | | | | | | |
| *s1m1* | 25 | 0 | 17 | 4 | 1 | 3 | 0 |
| *s1m2* | 7 | 1 | 6 | 0 | 0 | 0 | 0 |
| *s2m1* | 8 | 0 | 8 | 0 | 0 | 0 | 0 |
| s2m2 | 6 | 2 | 4 | 0 | 0 | 0 | 0 |
| **Co-infection** | 6 | 0 | 6 | 0 | 0 | 0 | 0 |
| ***CagA* status** | | | | | | | |
| *Positive* | 29 | 0 | 25 | 3 | 1 | 0 | 0 |
| *Negative* | 36 | 4 | 24 | 1 | 3 | 3 | 1 |
| ***cagA* genotype** | | | | | | | |
| *ABC* | 22 | 0 | 19 | 3 | 0 | 0 | 0 |
| *ABCC* | 2 | 0 | 2 | 0 | 0 | 0 | 0 |
| *AAABC* | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| *Co-infection* | 4 | 0 | 4 | 0 | 0 | 0 | 0 |
| ***cagA/vacA* genotype** | | | | | | | |
| *cagA-/s1m1* | 6 | 0 | 2 | 1 | 0 | 3 | 0 |
| *cagA-/s1m2* | 3 | 1 | 2 | 0 | 0 | 0 | 0 |
| *cagA-/s2m1* | 7 | 0 | 7 | 0 | 0 | 0 | 0 |
| *cagA-/s2m2* | 6 | 2 | 4 | 0 | 0 | 0 | 0 |
| *ABC/s1m1* | 17 | 0 | 14 | 3 | 0 | 0 | 0 |
| *ABCC/s1m2* | 2 | 0 | 2 | 0 | 0 | 0 | 0 |
| *AAABC/s1m1* | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| *Co-infection* | 10 | 0 | 10 | 0 | 0 | 0 | 0 |

**Table 10. Frequency of *vacA* alleles and genotype and *cagA* status and EPIYA motif determination in relation to duodenal findings in biopsies from a population of Native Americans living in Northern Arizona.**

| | Total | Normal duodenum | Duodenitis |
|---|---|---|---|
| ***vacA* alleles** | | | |
| *s1* | 40 | 36 | 4 |
| *s2* | 19 | 18 | 1 |
| *s1/s2* | 5 | 5 | 0 |
| *m1* | 35 | 33 | 2 |
| *m2* | 15 | 14 | 1 |
| *m1/m2* | 3 | 3 | 0 |
| ***vacA* genotype** | | | |
| *s1m1* | 25 | 24 | 1 |
| *s1m2* | 7 | 6 | 1 |
| *s2m1* | 8 | 7 | 1 |
| *s2m2* | 6 | 6 | 0 |
| *Co-infection* | 6 | 6 | 0 |
| ***CagA* status** | | | |
| *Positive* | 29 | 27 | 2 |
| *Negative* | 36 | 33 | 3 |
| ***cagA* genotype** | | | |
| *ABC* | 22 | 20 | 2 |
| *ABCC* | 2 | 2 | 0 |
| *AAABC* | 1 | 1 | 0 |
| *Co-infection* | 4 | 4 | 0 |
| ***cagA*/*vacA* genotype** | | | |
| *cagA-/s1m1* | 6 | 6 | 0 |
| *cagA-/s1m2* | 3 | 2 | 1 |
| *cagA-/s2m1* | 7 | 6 | 1 |
| *cagA-/s2m2* | 6 | 6 | 0 |
| *ABC/s1m1* | 17 | 16 | 1 |
| *ABCC/s1m2* | 2 | 2 | 0 |
| *AAABC/s1m1* | 1 | 1 | 0 |
| *Co-infection* | 10 | 10 | 0 |

# Figures

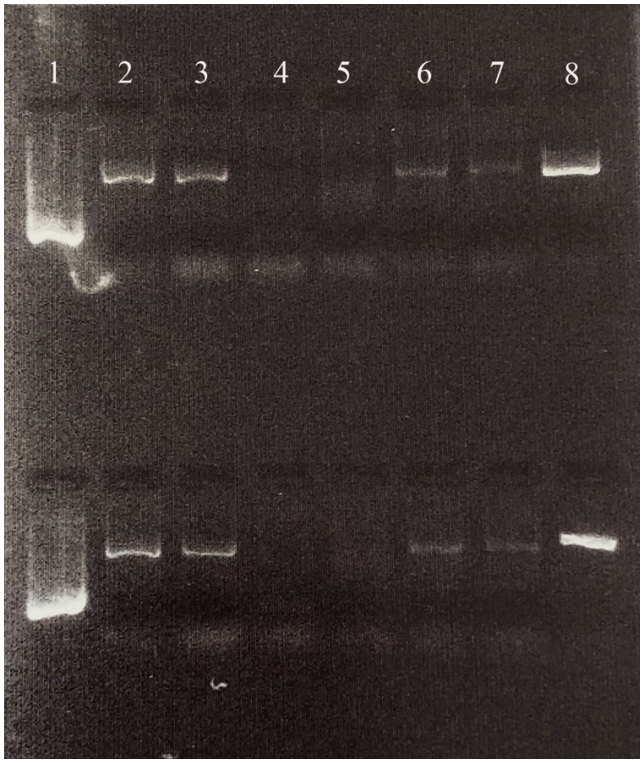



**Figure 1.** Preliminary screening in 1% agarose gels for *Helicobacter pylori*. Each Sample was loaded in duplicate wells. Lane 1 = Marker. Lane 2 = 115A. Lane 3 = 115F. Lane 4 = 120A. Lane 5 = 120F. Lane 6 = 119F. Lane 7 = 131A. Lane 8 = Positive control.

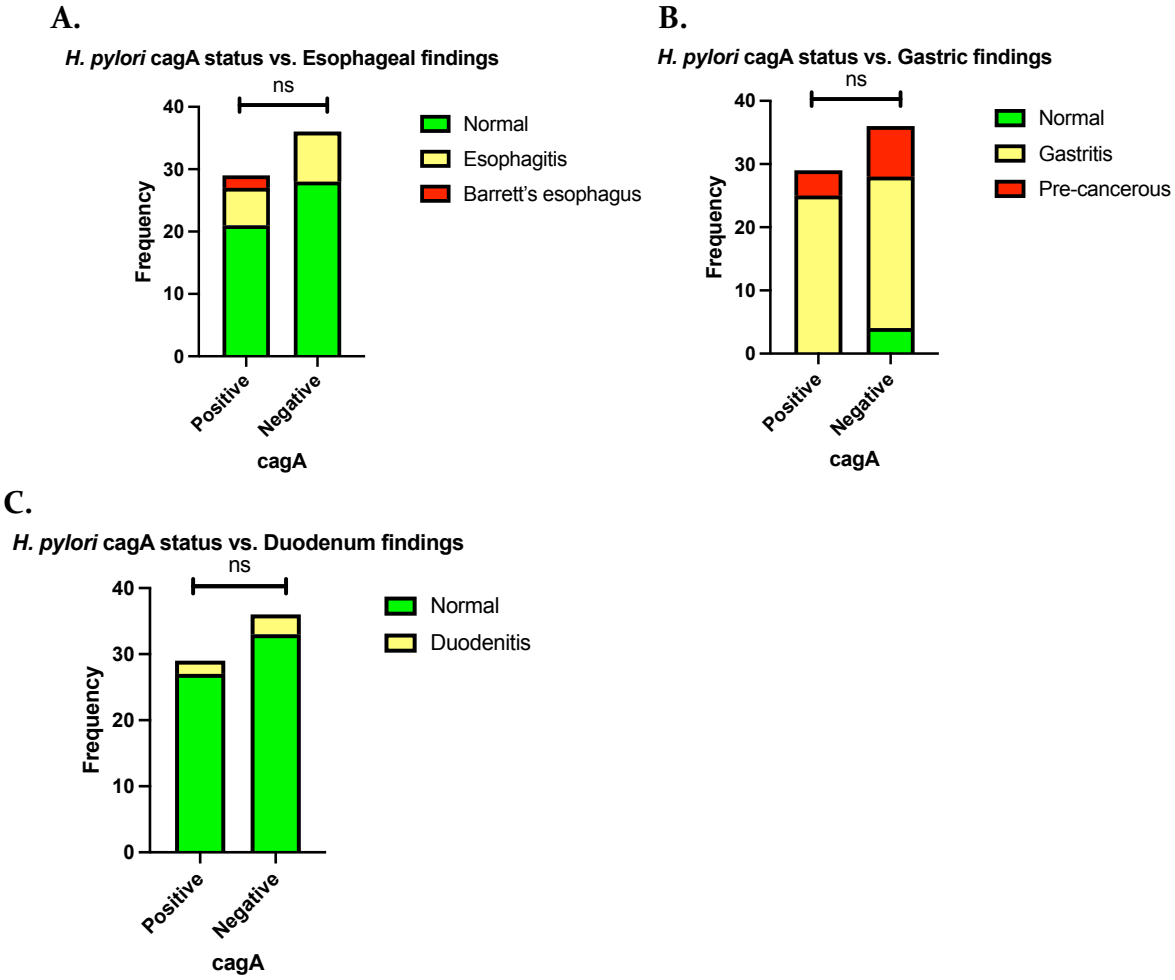
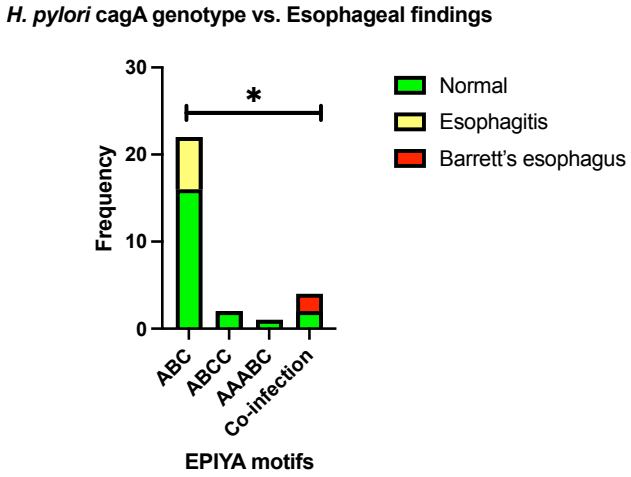
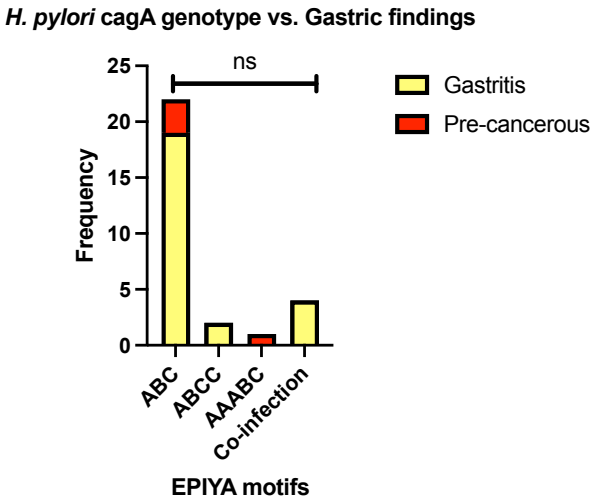



**Figure 2.** *Helicobacter pylori cagA* status and EGD findings in the esophageal (A), gastric (B), and duodenal (C) regions. ns = not significant, $p \leq 0.05$.

A.




B.




C.

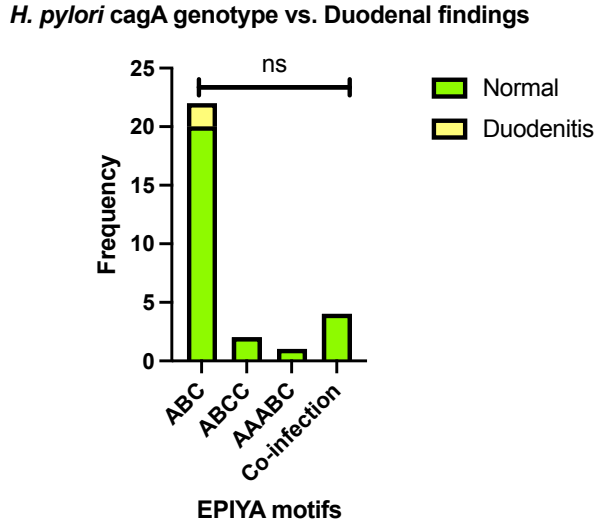
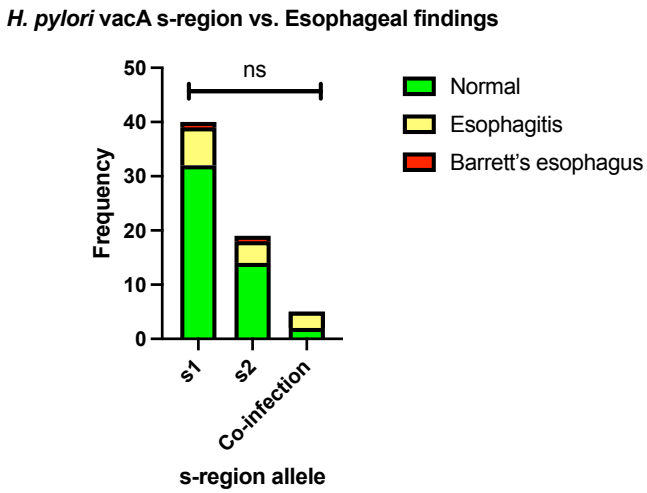
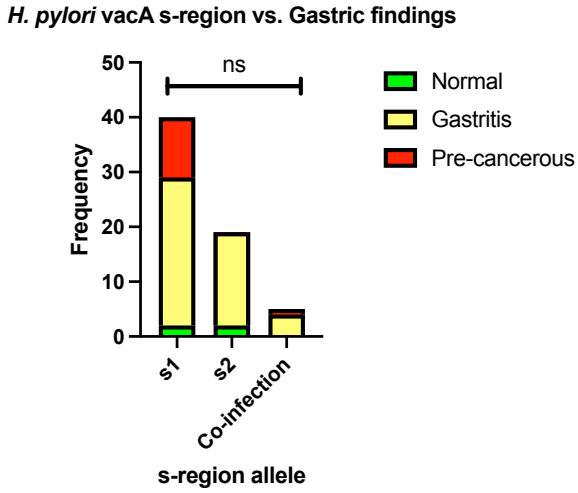



**Figure 3.** *Helicobacter pylori cagA* genotype (EPIYA-motif) frequency and EGD findings in the esophageal (A), gastric (B), and duodenal (C) regions. *, $p \leq 0.05$.

A.




B.




C.

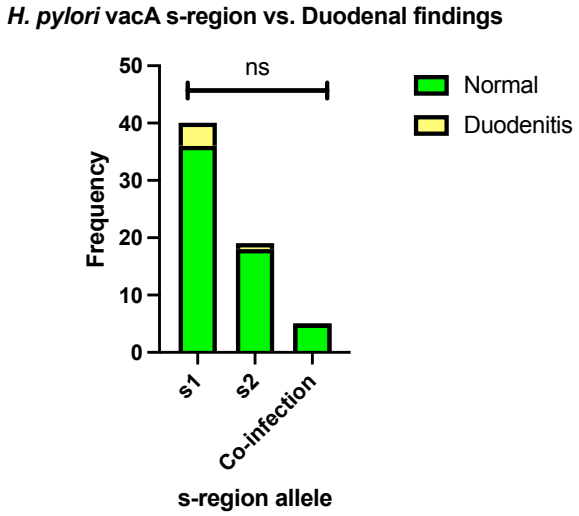
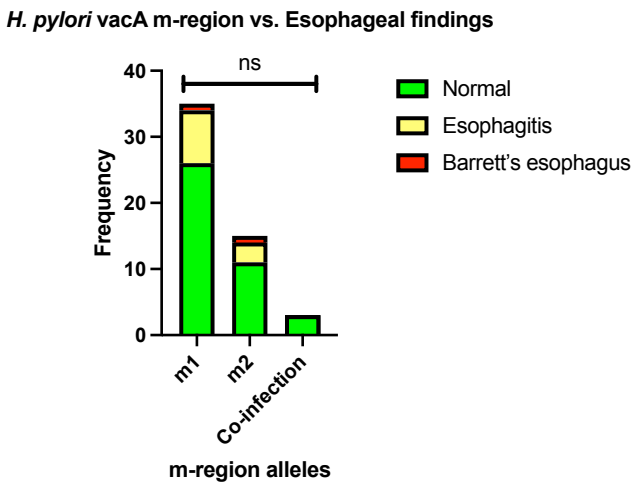
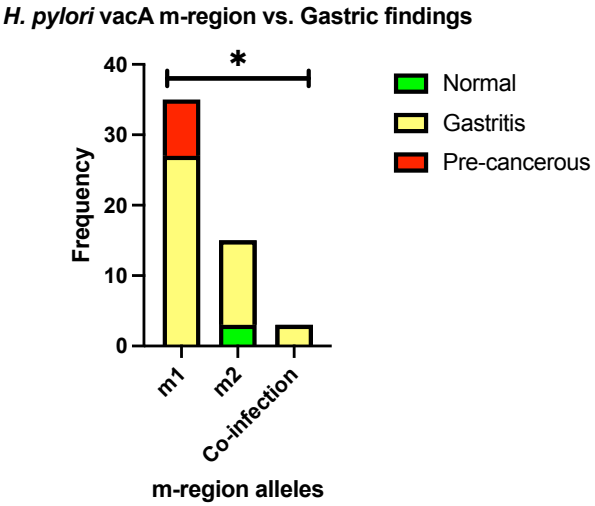



**Figure 4.** *Helicobacter pylori vacA* signal (s)-region frequency and EGD findings in the esophageal (A), gastric (B), and duodenal (C) regions. ns = not significant, $p \leq 0.05$.

A.




B.




C.

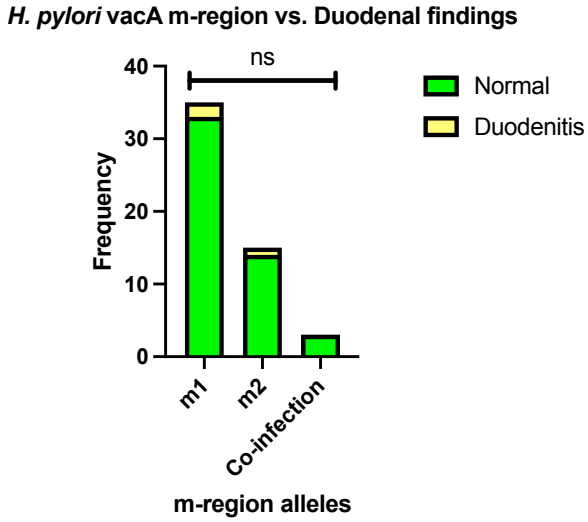
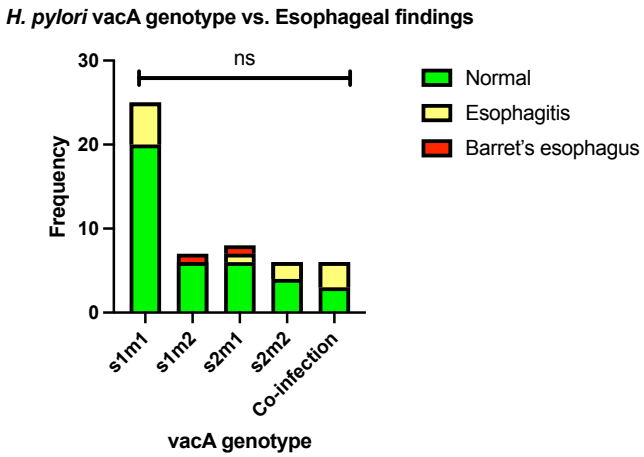
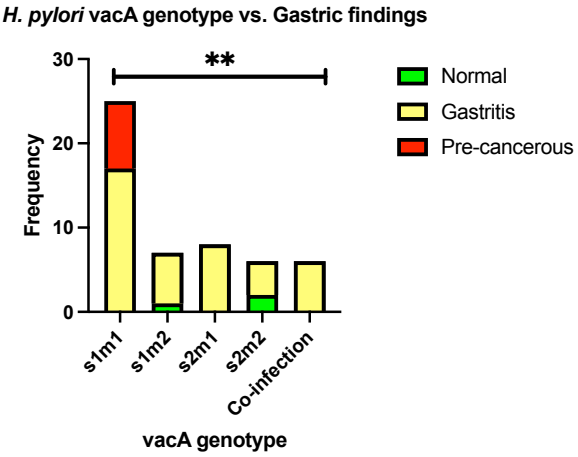



**Figure 5.** *Helicobacter pylori vacA* mid (m)-region frequency and EGD findings in the esophageal (A), gastric (B), and duodenal (C) regions. *, $p \leq 0.05$.

A.




B.




C.

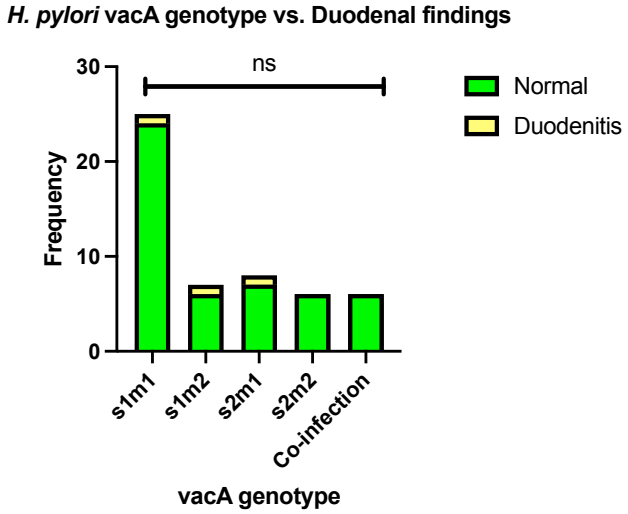



**Figure 6.** *Helicobacter pylori vacA* genotype frequency and EGD findings in the esophageal (A), gastric (B) and duodenal (C) regions. **, $p \leq 0.01$.

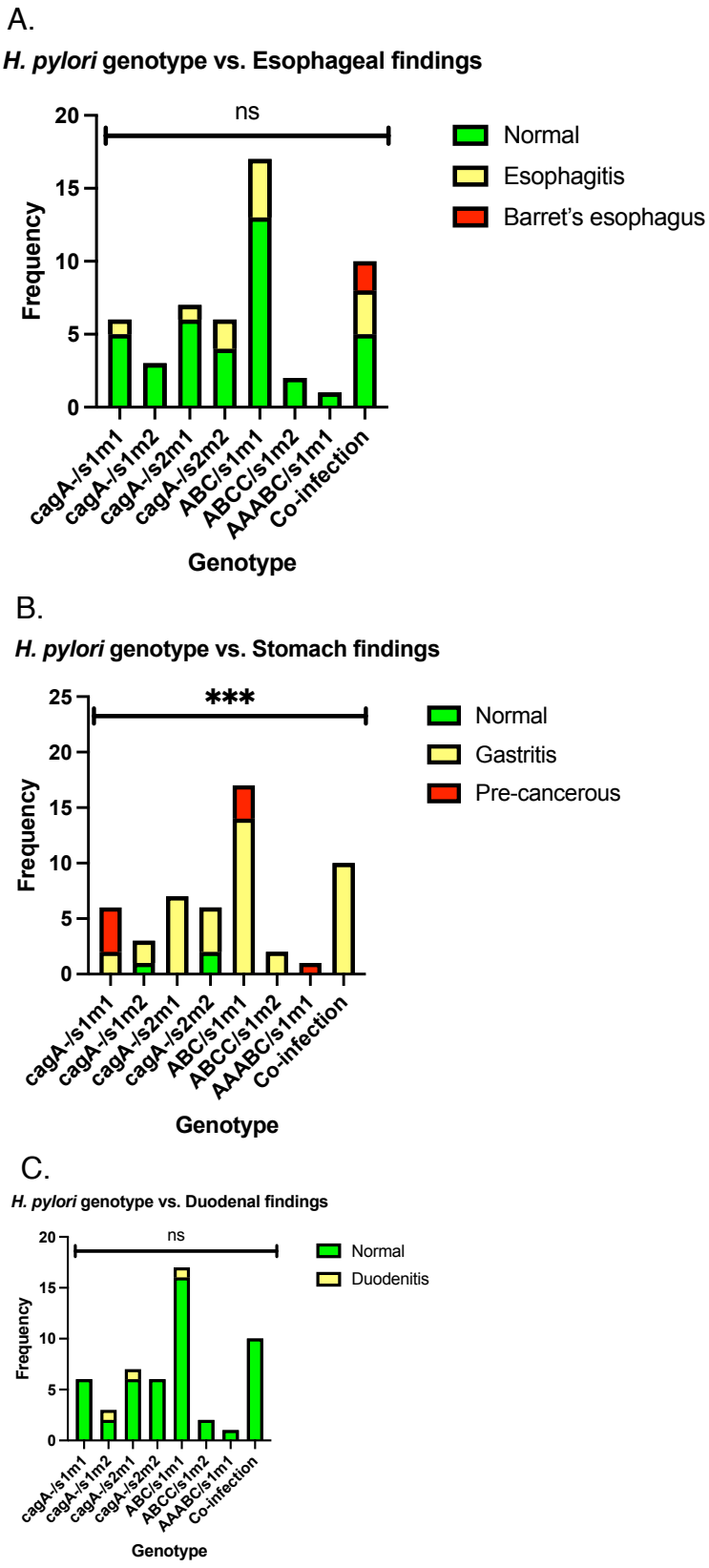



**Figure 7.** *Helicobacter pylori cagA*/*vacA* genotype frequency and EGD findings in the esophageal (A), gastric (B), and duodenal (C) regions. ***, $p \leq 0.001$.

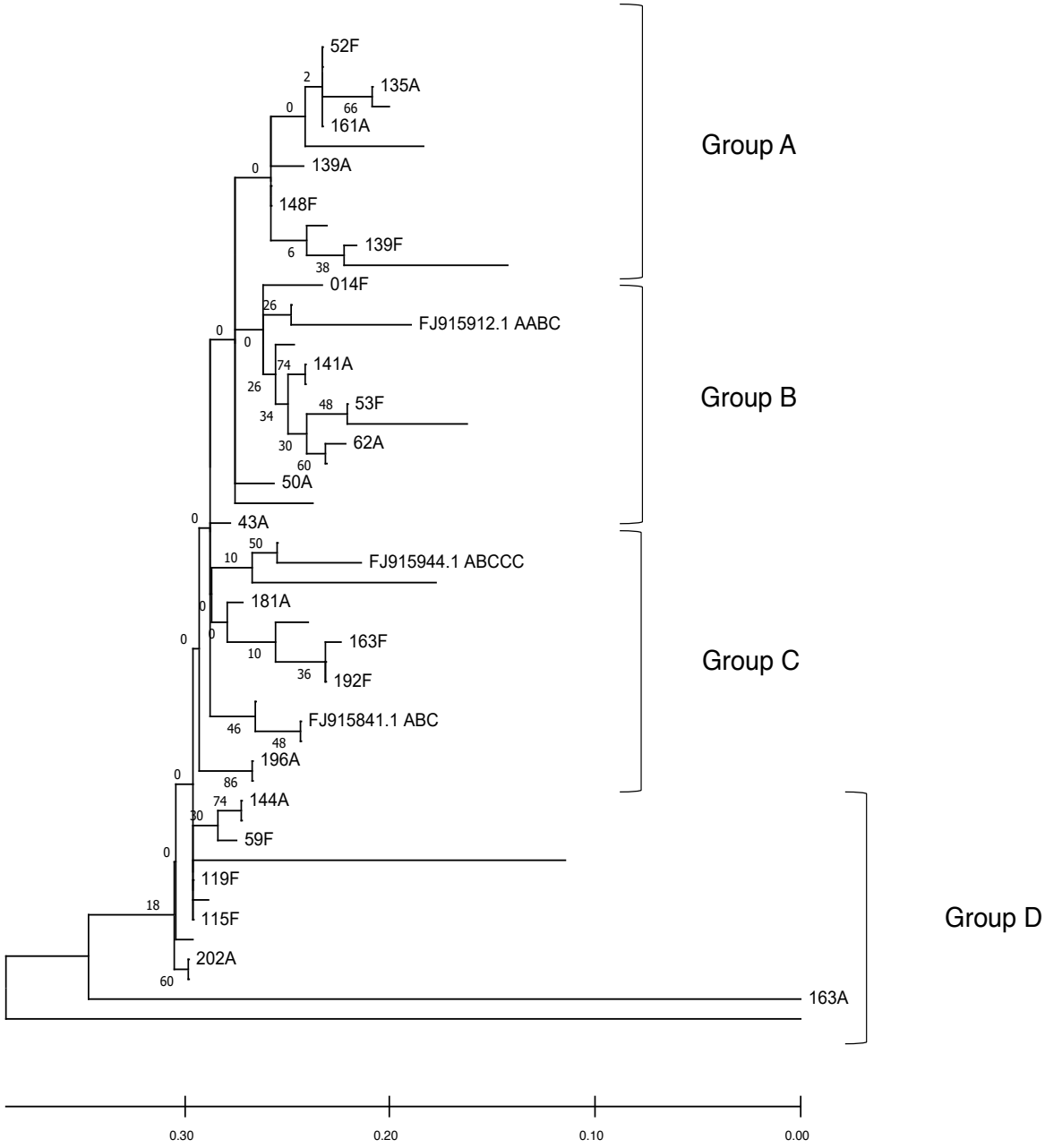



**Figure 8.** Phylogenetic tree of *cagA* gene in isolates from Native Americans living in Northern Arizona compared with the following reference strains: strain 26695, OK 111, FJ915912.1, FJ915891.1, FJ915944.1, FJ915841.1, FJ915878.1, NCTC 11637 and F32.

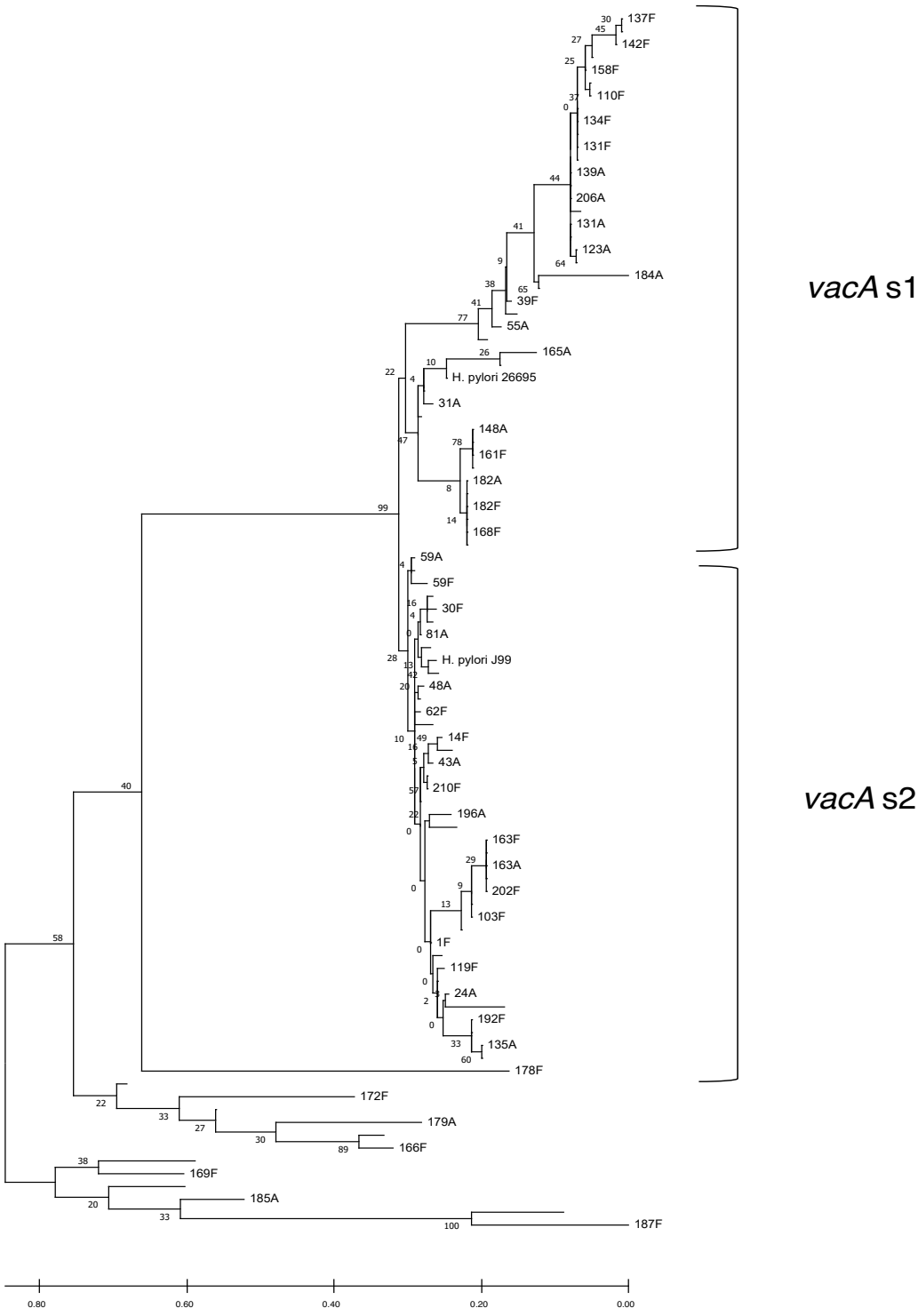



**Figure 9.** Phylogenetic tree of the *vacA* s-region in isolates from Native Americans living in Northern Arizona compared with the following reference strains: strain 26695, J99, and Tx30a.

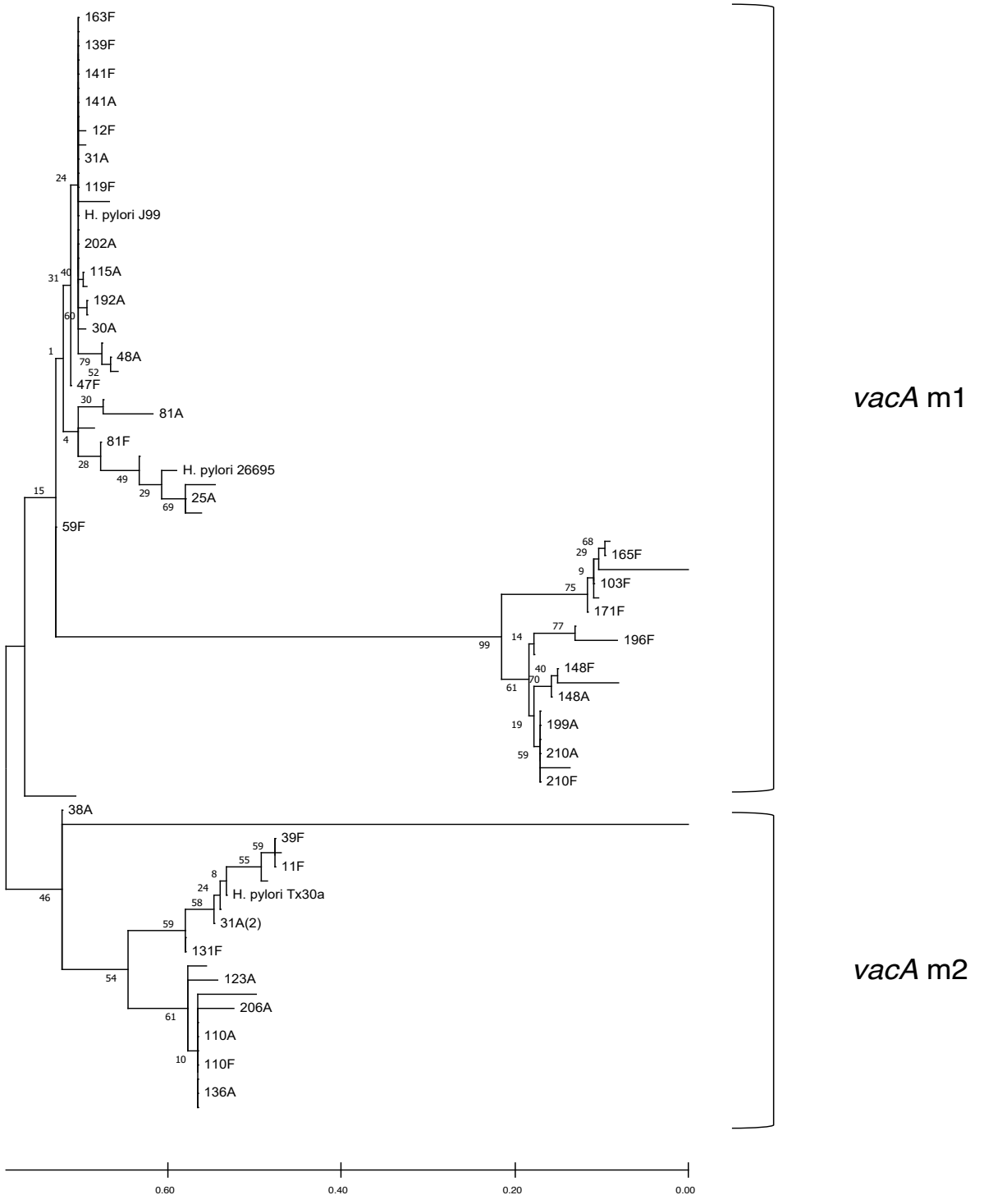



**Figure 10.** Phylogenetic tree of the *vacA* m-region in isolates from Native Americans living in Northern Arizona compared with the following reference strains: strain 26695, J99 and Tx30a.