Clinical and environmental risk factors of *Helicobacter pylori,* gastric cancer, and gastric microbiome signature in a South African cohort

By:

Innocent Ekparolaguaziba Francis

FRNINN001

Submitted to the University of Cape Town as fulfillment for the award of a PhD in Medicine

Faculty of Health Sciences, University of Cape Town

Supervisor: Professor Mashiko Setshedi

Co-supervisor: Professor Adrian Brink

DECLARATION

I, *Dr. Innocent Ekparolaguaziba Francis*, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: .. Signed by candidate

Date: 01 October 2025

DEDICATIONS

I dedicate this work to God Almighty for the grace and strength He granted me during my studies, and to my wife Umavie Ruth Francis, you are simply the best and the backbone of my success. Words cannot express the role you have played in my life. I love you dearly. Countless times, I slept on my laptop instead of staying close to you, but you loved me the same. My children, Eluan, Enola, and Enapu, thank you for bearing the pain of daddy always being on the computer and not willing to go for outings because of research and studies. I love you dearly. My siblings, Dr Daminola, Nator, Laguo, and Ibhosobo thank you for your prayers and encouragement.

ACKNOWLEDGEMENTS

I wish to acknowledge the support and encouragement given to me by my supervisors, Prof Mashiko Setshedi and Prof Adrian Brink. Thank you for guiding and mentoring me through the years. It was a memorable and impactful journey that left an indelible mark on me for the rest of my life. Undergoing gastroenterology training concurrently with a PhD pursuit was a huge task to embark upon, but you stood by me, encouraged me, and mentored me throughout my journey. Thanks so much, Prof Setshedi.

I also want to thank my teacher, A/Prof Gillian Watermeyer, for teaching me the rudiments of IBD, you are a giant among many. Drs. Dion Levin, Hlatshwayo, and Prof. Sandie Thomson, thank you for teaching me endoscopic skills and standing with me throughout my gastroenterology training. Thanks to A/Prof Galya Chinnery and Dr Matthias Scriba for your support and contribution to my research work and for equally building me in upper GI interventional endoscopy. To my hepatobiliary surgeons, Prof Jonas, Dr Burmeister, and A/Prof Marc Bernon, thank you for teaching me and for being part of my journey.

Thanks to a mother, mentor, and teacher Prof Wendy Spearman, thank you for tutoring and mentoring me in the management of Liver diseases. Thanks to Prof Mark Sonderup and Dr Gogela for all the teachings and encouragement.

I also wish to acknowledge the enormous contributions of Dr Fadheela Patel for all the wet laboratory work as well as proofreading this work and making valuable contributions. You ensure that we were able to successfully culture *Helicobacter pylori*. Thank you, Dr Ephifania Geza, you are indeed a superb bioinformatician. I admire all the magic manipulations on the computer, the countless times I have had to run back to both of you, and you were always keen to assist. Your effort played a big role in the completion of this project.

To all my colleagues during my training, Drs Isneen Hilmy, Dominic Mutura, Elizabeth Gatley, Nasief van der Schyff, Muaz Moosa, Sandy Mosenye, Simon Veenstra, Cecilia Coccia, and Thania Kahn; thank you for all your support and for being great colleagues to work with. Thanks to my friends Drs Brian and Joy for all the advice and support.

I cannot exhaust this list without mentioning my nursing colleagues, Sister Dias, Sister Alexander, Sister Johnson, and the whole team, thanks for supporting this work.

ABSTRACT

Background: *Helicobacter pylori (H. pylori*) is a Type 1 carcinogen associated with gastric cancer (GCA) and other gastroduodenal diseases. Despite high *H. pylori* prevalence in Africa, reported GCA incidence remains paradoxically low, a phenomenon known as the 'African enigma'. This study aimed to investigate the validity of this paradox through a systematic review, retrospective cohort analysis, and prospective study of *H. pylori*-infected individuals.

Methods: The study comprised three components: (1) a systematic review evaluating the epidemiology of GCA in Africa, (2) a retrospective cohort study analyzing GCA incidence and associated risk factors at Groote Schuur Hospital, and (3) a cross-sectional study examining the molecular signature of *H. pylori* strains and gastric microbiome diversity in affected patients.

Results: the systematic review highlighted significant limitations in current data, stressing the need for standardized national registries and comprehensive epidemiological studies. The retrospective study confirmed that GCA mainly affects males aged 60 and above, with non-cardia cancers being the most common subtype. The protective effect of proton pump inhibitors against antral cancer was also observed. The cross-sectional study revealed that the *H. pylori* strains among Africans had fewer virulence factors (*cagPAI* variants, *virB/D* genes) compared to Southeast Asian strains, potentially explaining lower GCA incidence. Additionally, *H. pylori* infection tends to alter the gastric microbiome diversity, with distinct microbial compositions across racial groups.

Conclusions: Findings support the 'African enigma', suggesting that specific *H. pylori* strain characteristics and microbiome alterations may contribute to the lower observed GCA incidence in Africa. Future research should integrate multi-omic analyses to further elucidate host-pathogen interactions and disease mechanisms.

TABLE OF CONTENTS

Chapter 5: Cross-sectional study: Epidemiology of *Helicobacter pylori* at Groote Schuur Hospital

Chapter 6: The molecular signature of *Helicobacter pylori* at Groote Schuur Hospital

LIST OF FIGURES

LIST OF SUPPLEMENTARY FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY TABLES

LIST OF ABBREVIATIONS

AMR: Antimicrobial resistance

CAG: Chronic atrophic gastritis

GCA: Gastric cancer

GSH: Groote Schuur Hospital

MLST: Multi-locus sequence typing

NCGC: Non-cardia GCA

NSAIDs: Non-steroidal anti-inflammatory drugs

PPI: Proton pump inhibitor

RGI: Resistance gene identifier

RUT: Rapid urease test

SCFA: Short chain fatty acid

SSA: SSA

VF: Virulent factors

WGS: Whole genome sequencing

CHAPTER 1

BRIEF LITERATURE REVIEW

1.1 BACKGROUND

Helicobacter pylori (H. pylori) is a gram-negative spiral-shaped bacterium that affects up to 40% of the population worldwide, with a higher prevalence in developing countries [1-3]. *H. pylori* causes chronic atrophic gastritis, peptic ulcers, gastric lymphoma, and gastric cancer [4]. There are various extra-intestinal manifestations that are associated with *H. pylori* infection such as iron deficiency anaemia and immune thrombocytopenia [5]. *H. pylori* infection is typically acquired in early childhood and persists without treatment [6]. Transmission of *H. pylori* can occur via the faecal-oral, gastric-oral, oral-oral, or sexual routes [7]. A major risk factor for a higher prevalence of infection is lower socioeconomic status [1], hence the prevalence is higher in Africa for instance.

H. pylori has an affinity for the stomach and causes disease by its urease activity which enables it to counter the acidic environment of the stomach. Other mechanisms include the flagella that help *H. pylori* attach to the gastric mucosa and bacterial adhesins interacting with the host cell receptors, leading to successful colonization and persistent infection. Finally, known effector proteins/toxins namely cytotoxin-associated gene A (Cag A) and vacuolating cytotoxin A (VacA) released by *H. pylori* result in inflammation and tissue damage [8]. *H. pylori* universally causes gastric inflammation even in asymptomatic patients [9, 10], with infiltration of neutrophils, eosinophils, mast cells, and dendritic cells [11]. Symptoms, if present, are usually of gastritis or peptic ulcer disease such as abdominal pain, nausea, vomiting or dyspepsia, or the complications thereof.

H. pylori is diagnosed by both invasive and non-invasive methods. Non-invasive tests include the detection of *H. pylori* stool antigens (SAG), detection of antibodies against *H. pylori* in serum, and a urea breath test (UBT). The SAG and UBT have high sensitivity and specificity similar to the invasive methods [12]. Serological assays detecting the presence of IgG antibodies have poor sensitivity, as they cannot differentiate current from past infection. They are, therefore, not reliable for use in the clinical setting [13]. The exception is in high prevalence regions where they can be used as a screening tool [8]. Invasive tests require gastric tissue for detecting *H. pylori* and include a RUT, histopathology, culture, polymerase chain reaction (PCR), and fluorescent in situ hybridization (FISH). Culture is the only method with 100% specificity, and a positive culture

confirms a diagnosis of *H. pylori* infection. It is recommended that at least 4 weeks after treatment, eradication should be confirmed [14].

Treatment of *H. pylori* comprises of a combination of an antisecretory drug and antibiotics. The choice of antisecretory drugs is either PPIs or potassium-competitive acid blockers (P-CABs). At least two antibiotics are generally chosen in the various regimens. The choice of regimen depends on antimicrobial resistance patterns in the region, previous regimen used, patient factors such as drug allergies and availability and access to medications [15-17]. The treatment goal for *H. pylori* infection is an eradication rate of at least 90% [18].

1.2 REFERENCES

1. Chen Y-C, Malfertheiner P, Yu H-T, Kuo C-L, Chang Y-Y, Meng F-T, et al. Global prevalence of *Helicobacter pylori* infection and incidence of GCA between 1980 and 2022. Gastroenterology. 2024;166(4):605-19

2. Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, Malfertheiner P, Graham DY, Wong VWS, Wu JCY, Chan FKL, Sung JJY, Kaplan GG, Ng SC. Global Prevalence of *Helicobacter pylori* Infection: Systematic Review and Meta-Analysis. Gastroenterology. 2017 Aug;153(2):420-429.

3. Graham DY, Adam E, Reddy GT, Agarwal JP, Agarwal R, Evans DJ, Malaty HM, Evans DG. Seroepidemiology of *Helicobacter pylori* infection in India. Comparison of developing and developed countries. Dig Dis Sci. 1991 Aug;36(8):1084-8.

4. Iannone A, Giorgio F, Russo F, Riezzo G, Girardi B, Pricci M, Palmer SC, Barone M, Principi M, Strippoli GF, Di Leo A, Ierardi E. New fecal test for non-invasive *Helicobacter pylori* detection: A diagnostic accuracy study. World J Gastroenterol. 2018 Jul 21;24(27):3021-3029.

5. Wong F, Rayner-Hartley E & Byrne MF Extraintestinal manifestations of Helicobacter pylori: a concise review. World J. Gastroenterol 20, 11950–11961 (2014).

6. Jones NL, Sherman PM. *Helicobacter pylori* infection in children. Curr Opin Pediatr. 1998 Feb;10(1):19-23.

7. Zamani M, Vahedi A, Maghdouri Z, Shokri-Shirvani J. Role of food in environmental transmission of *Helicobacter pylori*. Caspian J Intern Med. 2017 Summer;8(3):146-152.

8. Malfertheiner P, Camargo MC, El-Omar E, Liou JM, Peek R, Schulz C, Smith SI, Suerbaum S. *Helicobacter pylori infection*. Nat Rev Dis Primers. 2023 Apr 20;9(1):19. doi: 10.1038/s41572-023-00431-8.

9. Denic M, Touati E, De Reuse H. Pathogenesis of Helicobacter pylori infection. Helicobacter. 2020 Sep;25:e12736.

10. Sugano, K., et al., Kyoto global consensus report on *Helicobacter pylori* gastritis. Gut, 2015. 64(9): p. 1353-67.

11. Diaconu S, Predescu A, Moldoveanu A, Pop CS, Fierbințeanu-Braticevici C. *Helicobacter pylori infection*: old and new. J Med Life. 2017 Apr-Jun;10(2):112-117.

12. Fischbach W, Malfertheiner P. *Helicobacter Pylori* Infection. Dtsch Arztebl Int. 2018 Jun 22;115(25):429-436.

13. Rajindrajith S, Devanarayana NM, de Silva HJ. *Helicobacter pylori* infection in children. Saudi J Gastroenterol. 2009 Apr;15(2):86-94.

14. Malfertheiner P et al. Management of *Helicobacter pylori* infection — the Maastricht V/Florence consensus report. Gut 66, 6–30 (2017).

15. Chey WD, Leontiadis GI, Howden CW, Moss SF. ACG Clinical Guideline: Treatment Of *Helicobacter pylori* Infection. Am. J. Gastroenterol. 2017; 112: 212–39.

16. Mahachai V, Vilaichone RK, Pittayanon R et al. *Helicobacter pylori* management in ASEAN: The Bangkok consensus report. J. Gastroenterol. Hepatol. 2018; 33: 37–56.

17. Malfertheiner P, et al. Management of *Helicobacter pylori* infection: the Maastricht VI/Florence consensus report. Gut. 2022; 71: 1724–62.

18. Graham DY, Lu H, Yamaoka Y. A report card to grade *Helicobacter pylori* therapy. *Helicobacter*. 2007; 12: 275–8.)

CHAPTER 2

OVERVIEW OF THE STUDY

2.1 BACKGROUND

Helicobacter pylori (*H. pylori*) is a common global infection, infecting over 40% of the world population [1], with the highest prevalence reported in Africa. Initial prevalence data reported rates ranging from 70-80% for Africa. More recent data, however, has reported a decline in the prevalence of *H. pylori* with the African pool prevalence falling to 53% [2], attributable to better antibiotic access. Unfortunately, treating *H. pylori* requires multiple antibiotic regimes, thereby threatening the development and spread of antibiotic resistance and posing a significant threat to the eradication of *H. pylori. H. pylori* is reported to be a major contributor to the aetiopathogenesis of GCA globally. This understanding followed the discovery of *H. pylori* in 1982 by Barry Marshall, and Robin Warren [2] and subsequently, the knowledge of its role in the pathogenesis of GCA. Therefore, in 1994, the International Agency for Research on Cancer (IARC) designated *H. pylori* as a type1 carcinogen and was later reaffirmed in 2009 [3].

H. pylori in Africa is associated with gastritis, especially antral gastritis, and less commonly with duodenitis and duodenal ulcers. Even less common is the association of *H. pylori* and gastric ulcers and GCA. Studies have reported that while the incidence of gastritis can be as high as 90% among patients presenting with dyspeptic symptoms in an African cohort, the incidence of duodenal ulcers was 111/1 000 in the same population [4,5], while in a UK similar cohort with dyspeptic symptoms, the incidence of duodenal ulcers ranges between 175-305/1 000 [4]. Gastric ulcers are even rarer, reported as 6-30 times less common than duodenal ulcers among Africans despite the high prevalence of gastritis [4].

Holcombe *et al*. also reported the disproportionately low prevalence of GCA despite the relatively high prevalence of *H. pylori* reported at 70-80% [1,2]. According to the WHO, the incidence of GCA in Africa ranges between 3.3-4.4/100 000 [6]. These findings contradict what is reported in East Asian countries and other developed countries where the burden of GCA correlates well with the prevalence of *H. pylori*. In China, Japan, South Korea, and Mongolia, the incidence of GCA was reported at 20.6/100 000, 31.6/100 000, 27.9/100 000, and 32.5/100 000 respectively [6]. Also, a study in the UK reported an odd ratio of 2.8 in GCA cases among patients with *H. pylori* positive.

The so-called 'African enigma' has been a very contentious topic, with various explanations suggested to the findings reported by Holcombe et al. Among them are: (a) the possibility that the

enigma is not true and rather due to poor diagnostic capabilities because of poor health infrastructure (b) Africa has a different virulent phenotype of *H. pylori* (c) the high prevalence of helminths and the propagation of Th2 mediated cytokine pathway which opposes the *H. pylori* mediated Th1 pathway is possibly protective against the development of GCA among Africans.

The mechanisms of *H. pylori* chronic gastritis-dysplasia-metaplasia-cancer sequence are not well known but are thought to be related to *H. pylori* virulence factors (e.g., *cagA*, *vacA*, and others) [7], the host inflammatory response as well as environmental factors, [8] and importantly, diets and smoking and alcohol had been positively associated with GCA [9,10]. More recently, the role of microbiomes in *H. pylori*-related pathogenesis has increasingly been recognized. Patients with *H. pylori* infection have a fecal microbiome with increased diversity [11]. Additionally, those with higher *H. pylori* antigen stool tests showed a harmful microbial phenotype, which may predispose them to the development of GCA [11]. In another study, dysbiotic microbiota in *H. pylori-positive* gastric biopsies was associated with chronic atrophic gastritis and intestinal metaplasia/dysplasia [12], lending credence to the theory that microbiota may play a role in *H. pylori* oncogenesis.

Previous studies in South Africa have evaluated *H. pylori* virulence factors, but to our knowledge, no studies locally have evaluated the gastric microbiome. To further understand the oncogenesis of *H. pylor*i and predict which *H. pylori-infected* patients are at risk of progressing to GCA, longitudinal and further mechanistic studies are required. In this study, we aim to evaluate the possible existence of the so-called 'African enigma' and evaluate the epidemiological and microbiological factors influencing this enigma, if it exists.

We hypothesize that in Africa:

1. Patients infected with *H. pylori* may be protected against more aggressive disease phenotypes i.e. the 'African enigma' is true.
2. Clinical risk factors and histological activity in *H. pylori*-infected individuals will provide clues to risk factors for developing GCA.
3. The gastric microbiome of patients infected with *H. pylori* exhibits dysbiosis with less diversity similar to other gastric microbiome studies.

2.2 OBJECTIVES

The objectives of the study are as follows:

1. To investigate the possibility of the existence of the 'African enigma' by analyzing the incidence of GCA and associated risk factors. This was achieved by performing a systematic review of the literature on the incidence and epidemiology of GCA in Africa.
2. To determine the burden of GCA in Groote Schuur Hospital (GSH), Cape Town, South Africa, and its associated predisposing factors and characterizing the clinical presentation (symptoms, risk factors, endoscopic findings, and histology), to determine the contributions of *H. pylori* in the development of GCA. This will be achieved through a retrospective study design.
3. To investigate a local cohort of patients with *H. pylori* infection compared to non-infected control by characterizing:
    a. The clinical presentation (symptoms, risk factors, endoscopy findings, and histology)
    b. The gastric microbiome
    c. The molecular structure and genetic composition of *H. pylori* in the local cohort, and to compare these findings to regions with higher and similar prevalence of GCA.

2.3 MATERIALS AND METHODS

2.3.1 Study design and population

The study was conducted in the Division of Gastroenterology, GSH, Cape Town, South Africa.
A flow of the study is shown in Figure 2, the following was performed:

1. A systematic review on the burden and epidemiological profile of patients with GCA in Africa. All publications for Africa with biopsy-proven GCA for adults of 18 years and above published in the last 11 years (June 2010 and June 2021), were included in the systematic review. A literature search was conducted on PubMed/Medline, Embase, and the African database (Chapter 3).
2. A retrospective study, assessing the epidemiological profile of patients with biopsy-proven GCA including demographic data, clinical risk factors, endoscopic and histologic pathology, that was collated to assess the factors associated with the development of GCA. The

retrospective study cohort included adults of 18 years and above with confirmed diagnoses of GCA at GSH (chapter 4).

3. A cross-sectional study investigating a local cohort of patients with confirmed *H. pylori* infection compared to non-infected controls by characterizing: the clinical presentation (symptoms, risk factors, endoscopy findings, and histology), and molecular and microbiome signatures using gastric samples, which was investigated in two studies to further understand the role of *H. pylori* in the aetiopathogenesis of GCA (chapters 5-7).
    a. Whole genome sequencing (WGS) for *H. pylori*, including the virulence factors, mutations, and antibiotic resistance pattern, comparing results to countries with higher and similar prevalence of GCA (China, South Korea, Japan, Mongolia, the United States of America, the United Kingdom, and Europe).
    b. The gastric microbiome targeting comparing a cohort of *H. pylori-positive* to the control *H. pylori-negative* group.

The cross-sectional study included a South African cohort of patients who are 18 years and above, attending the Gastroenterology clinic at GSH, requiring a gastroscopy for appropriate clinical indications and who consented to participate in the study.

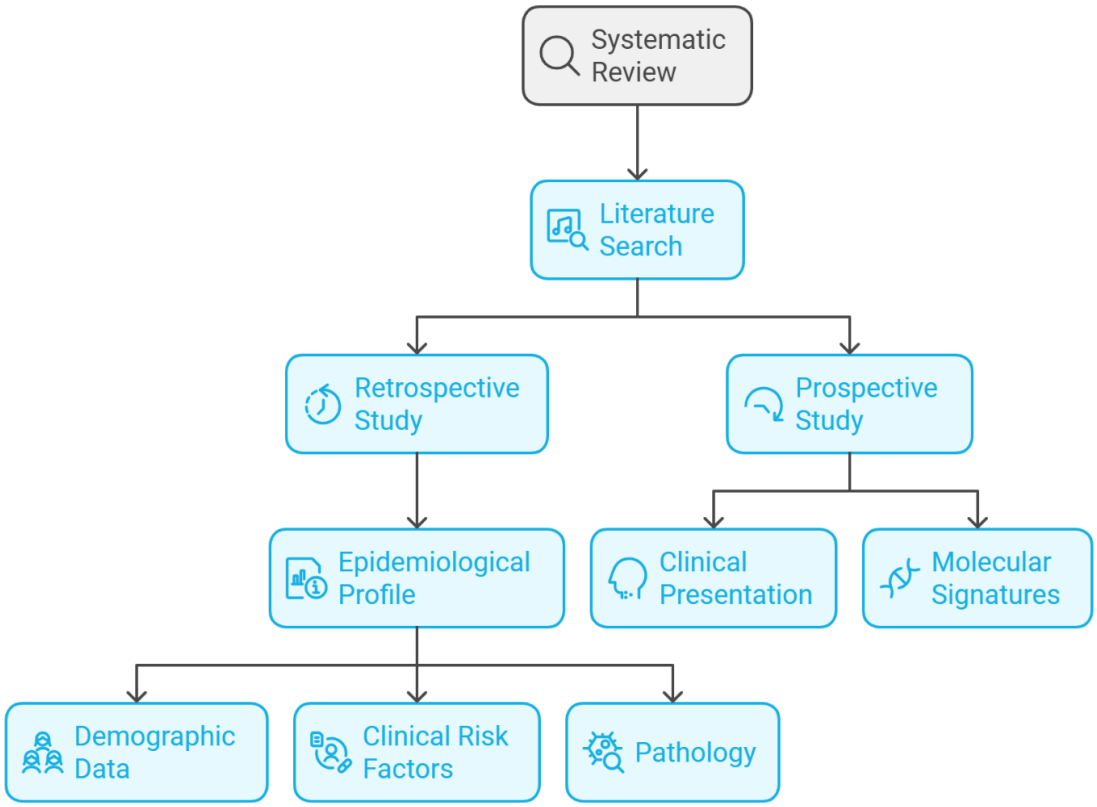


Figure 2: The schematic diagram showing the flow of the study

2.4 ETHICS

Ethical approval for this study was obtained from the University of Cape Town Human Research Ethics Committee (HREC), reference number 552/2021. Written consents from patients were obtained based on the Helsinki guidelines. Approval of the protocol and consent forms from participants was obtained before enrolment. Confidentiality was maintained; therefore, the names and addresses of patients are not published, the computer was kept in a locked office and data was secured with a password. Any amendments to the protocol or informed consent required review and approval from the HREC, before any change was implemented.

2.5 SIGNIFICANCE OF THE STUDY

This study aims to understand the aetiopathogenesis of *H. pylori* through the lens of the so-called 'African enigma', to determine if indeed it exists, and to determine the factors at play in *H. pylori* infection in a region with high prevalence that may or may not increase the risk for GCA. These can in turn be exploited to advance preventative and therapeutic strategies to better manage *H. pylori* in our region and potentially reduce rates of GCA in areas where the incidence is high.

2.6 REFERENCES

1. Chen Y-C, Malfertheiner P, Yu H-T, Kuo C-L, Chang Y-Y, Meng F-T, et al. Global prevalence of *Helicobacter pylori* infection and incidence of GCA between 1980 and 2022. Gastroenterology. 2024;166(4):605-19.

2. Malfertheiner P, Camargo MC, El-Omar E, Liou J-M, Peek R, Schulz C, et al. *Helicobacter pylori* infection. Nature Reviews Disease Primers. 2023;9(1):19.

3. Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, et al. A review of human carcinogens—Part B: biological agents. The Lancet oncology. 2009;10(4):321-2.

4. Holcombe C. *Helicobacter pylori*: the African enigma. Gut. 1992;33(4):429.

5. Holcombe C, Umar H, Lucas SB, Kaluba J. Low incidence of clinically significant gastroduodenal pathology despite a high incidence of *Helicobacter pylori* infection. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1994;88(5):569-71.

6. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2021;71(3):209-49.

7. Kim JY, Kim N, Nam RH, Suh JH, Chang H, Lee JW, et al. Association of polymorphisms in virulence factor of *Helicobacter pylori* and gastroduodenal diseases in South Korea. J Gastroenterol Hepatol. 2014;29(5):984-91.

8. Wroblewski LE, Peek Jr RM, Wilson KT. *Helicobacter pylori* and GCA: factors that modulate disease risk. Clinical microbiology reviews. 2010;23(4):713-39.

9. Nishino Y, Inoue M, Tsuji I, Wakai K, Nagata C, Mizoue T, et al. Tobacco smoking and GCA risk: an evaluation based on a systematic review of epidemiologic evidence among the Japanese population. Japanese journal of clinical oncology. 2006;36(12):800-7.

10. Jedrychowski W, Wahrendorf J, Popiela T, Rachtan J. A case-control study of dietary factors and stomach cancer risk in Poland. International journal of cancer. 1986;37(6):837-42.

11. Frost F, Kacprowski T, Rühlemann M, Bang C, Franke A, Zimmermann K, et al. *Helicobacter pylori* infection associated with fecal microbiota composition and diversity. Scientific reports. 2019;9(1):1-10.

12. Guo Y, Zhang Y, et al. Effect *of Helicobacter pylori* on gastrointestinal microbiota: a population-based study in Linqu, a high-risk area of GCA.Gut.2020;69(9):1598-607.

CHAPTER 3

SYSTEMATIC REVIEW ON THE EPIDEMOLOGICAL PROFILE OF GASTRIC CANCER IN AFRICA

Abstract

Background: The incidence of GCA is reported to be low in Africa, despite a high prevalence of *H. pylori*, the so-called African enigma. This study aimed to determine the pooled prevalence of GCA and incidence in Africa.

Methods: PRISMA guidelines were followed. PubMed/Medline, African Journal online, and EMBASE searches were performed for cohorts from January 2011 to June 2022. Case-control and cohort studies among adults ≥18 years were included. The study protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) (registration no: CRD42019130348).

Results: Forty studies were included with the majority from West and North Africa. 70% and 30% were cross-sectional and cohort studies respectively. The median age at diagnosis of GCA was 54 (range 45-61). The diagnosis was made endoscopically in 79% and histologically in 95%.

The median sample size of included studies was 455 (range 26-8565) and the median number of GCAs was 114 (range 13-876), however, this included some studies where all patients in the sample had GCA. 52% of cancers were in the antrum and 26% gastric body. Six studies reported *H. pylori* on histology and in these only 5% were positive. The majority (90%) of GCAs were adenocarcinoma.

Conclusions: This systematic review revealed deficiencies in the reported data. In many studies background sample/population sizes were not reported, making determinations of prevalence/incidence inaccurate, as a result the intended meta-analysis was not possible. Antral cancers occurred in 52%, suggesting *H. pylori* as a cause of GCA in Africa, despite limited data. This systematic review revealed a paucity of quality data on GCA in Africa which cannot support or contradict the argument for the 'African enigma'. Reliable country registry data is urgently needed.

3.1 BACKGROUND

3.1.1 Epidemiology of GCA

Morbidity and mortality from all cancer has remained high globally, and up until the 1980s, GCA was the most common cause of cancer death before it was overtaken by lung cancer [1]. Although there has been a steady decline in the incidence of GCA in developed countries, mortality and morbidity remain high [1]. The survival rate at 5 years is reported as 20% in most countries. GCA disproportionately affects men more than women with a ratio of 2-3:1 [2, 3]. The reduction in the incidence of GCA globally is related to a better understanding of the associated risk factors and better diagnostic modalities. For example, the discovery of *H. pylori* in 1982 by Barry Marshall and Robin Warren [4] and subsequently understanding its role in the pathogenesis of GCA, coupled with insight into the role of diet and other associated factors, was pivotal to this global reduction. According to the GLOBOCAN report for 2020, an estimated 19.3 million new cases of cancer were diagnosed, with 10 million cancer deaths for the year 2020. GCA alone accounts for over 1 million newly diagnosed cancers with 769,000 deaths (equating to 1 in 13 deaths), making GCA the 5th most common cancer and 4th most common cause of cancer-related mortality globally [5]. The incidence is highest in the East Asian countries (Japan, China, and South Korea) and lowest in North America and the African continent.

3.1.2 Risk factors for the development of GCA

H. pylori is reported to be a major contributor to the aetiopathogenesis of GCA globally. In 1994, the International Agency for Research on Cancer (IARC) designated *H. pylori* as a type1 carcinogen; this was reaffirmed in 2009[1,4]. *H. pylori* has the highest prevalence in SSA with a prevalence rate of 61-100% depending on the country [6]. Ironically, GCA is reported with a low prevalence rate in Africa compared to the Asian continent with a low prevalence of *H. pylori* infection 20-40% [6], but a higher incidence of GCA. Other risk factors include smoking, alcohol, a diet high in salt and low in vegetables, Epstein Bar Virus, genetic factors, Menetrier's disease, and autoimmune gastritis [7].

3.1.3 Epidemiology of GCA in Africa

The African continent is reported to have the lowest prevalence of GCA despite the high prevalence of *H. pylori*. There is also a significant variability in the reported incidence of GCA

even in the same country. Kidd et al reported that in Nigeria, Sudan, and Zimbabwe the prevalence of GCA is 2-3% of all malignancies, while in Southern Africa, it's suggested that the prevalence is as low as 0.8/100 000 to 2.4-20/100 000 [6]. The reason for this low incidence rate of GCA remains largely unknown. Some studies have suggested environmental factors such as diet and smoking [8] as key players in driving this variability, others have suggested that the strain of *H. pylori* in Africa might be of a low virulent sub-type [9], yet some have considered this paradoxical low incidence rate to be due to be possibly parasitic infections which induce a protective mechanism against the virulence factors expressed by the *H. pylori* for the promulgation of GCA[10]. Other possible factors could be due to the poor diagnostic ability that has plagued the continent over the years. For instance, in Kenya, McFarlane et al conducted a retrospective study in the Eastern part of the country on the prevalence of GCA between 1991-1993 compared to between 1965-1970 and found a 10-fold increase in GCA cases; of note is that endoscopic services were introduced in Kenya in the 1980s [11]. Similarly, in Uganda, East Africa, the most recent data shows a seven-fold increase incidence of GCA from 0.8/100 000 in the 1960s to 5.6/100 000 [12]. In the recent past, there has been a significant improvement in the diagnostic capabilities in most African countries which is expected to reflect on the current epidemiological profile of GCA in Africa.

3.2 AIM

This review is aimed at investigating the current epidemiological profile (including prevalence and incidence) of GCA in Africa.

3.3 METHODS

We searched PubMed/Medline, African Journal Online, and EMBASE for the African cohorts. All articles dealing with the subject matter in adults 18 years and above were included, from January 2011 and June 2021 without language restrictions. The systematic review and meta-analysis used the Meta-analysis of Observational Studies in Epidemiology guideline (MOOSE) and reported based on the Preferred Items for Systematic Review and Meta-Analysis for Protocol (PRISMA-P). Finally, the study protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) (registration no: CRD42019130348).

3.3.1 Search Strategy

The above-listed search engines were used to conduct a search involving the key terms of incidence, prevalence, gastric/stomach, cancer, neoplasia, malignancy, and non-benign lesion. This was combined search strategy for Africa; the African sub-regions, and member countries were used as part of the search strategy. We used the following search terms: Africa OR Africa OR Algeria OR Angola OR Benin OR Botswana OR "Burkina Faso" OR Burundi OR Cameroon OR "Canary Islands" OR "Cape Verde" OR "Central African Republic" OR Chad OR Comoros OR Congo OR "Democratic Republic of Congo" OR Djibouti OR Egypt OR "Equatorial Guinea" OR Eritrea OR Ethiopia OR Gabon OR Gambia OR Ghana OR Guinea OR "Guinea Bissau" OR "Ivory Coast" OR " Cote d`Ivoire" OR Jamahiriya OR Kenya OR Lesotho OR Liberia OR Libya OR Libia OR Madagascar OR Malawi OR Mali OR Mauritania OR Mauritius OR Morocco OR Mozambique OR Mocambique OR Namibia OR Niger OR Nigeria OR Principe OR Reunion OR Rwanda OR "Sao Tome" OR Senegal OR Seychelles OR "Sierra Leone" OR Somalia OR "South Africa" OR "St Helena" OR Sudan OR Swaziland OR Tanzania OR Togo OR Tunisia OR Uganda OR "Western Sahara" OR Zaire OR Zambia OR Zimbabwe OR "Central Africa" OR "Central African" OR "West Africa" OR "West African" OR "Western Africa" OR "Western African" OR "East Africa" OR "Eastern African" OR "North Africa" OR "North African" OR "Northern Africa" OR "Northern African" OR "South African" OR "Southern Africa" OR "Southern African" OR "sub Saharan Africa" OR "sub Saharan African" OR "subSaharan Africa" OR "subSaharan African"

3.3.2 Inclusion criteria

Case-control, cohort, and cross-sectional studies for GCA among adults of 18 years and above, conducted in Africa, were included.

3.3.3 Exclusion criteria

Studies published outside of the window period were excluded. In addition, studies focused on patients younger than 18, and articles not written in English for which we are unable to obtain a translation were excluded.

3.3.4 Data extraction

Data was extracted according to the first author's surname, year of publication, country of study, study design, sample size, mean or median age, gender, specific characteristics of the study population (diet, smoking history, alcohol history), histological type of cancer and location, presence of *H. pylori* and mode of cancer treatment.

3.3.5 Selection of articles

Two reviewers, IEF and MS independently assessed the searched articles for inclusion and quality of study based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) classification. Duplicate articles were eliminated, and grey areas from the selection process were resolved after deliberation.

3.3.6 Data analysis

The planned meta-analysis was not performed due to the heterogeneity of included studies, and the lack of key data such as the background population at risk. Descriptive statistics (median and interquartile ranges) were used to characterize the sample. Frequencies and percentages, bar charts and pie charts were used for categorical variables. Data exploration and analysis was done in Stata (Version 13.1; Stata Corp, College Station, Texas, USA) ®.

3.4 RESULTS

Following the search, a total of 139 articles were initially sampled between 2001-2021 and screened by titles, abstracts, and full text, following which, 40 studies met the inclusion criteria. The flow chart is shown in Figure 3.1. The majority of the published data emanates from the West and North African sub-region as detailed in Tables 3.1 and 3.2.

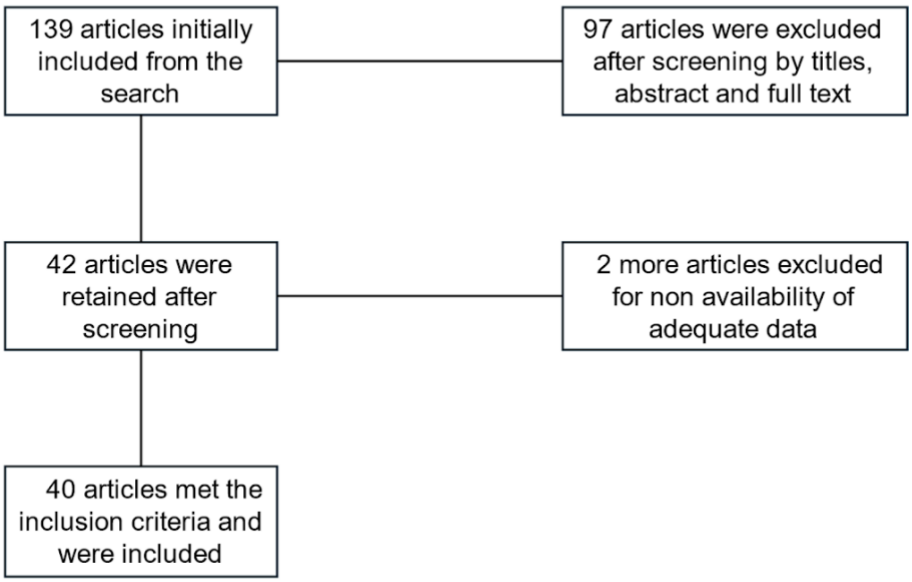


Figure 3.1: Prisma flow diagram

Table 3.1: Research articles: topic, author, year of publication, and country of study

Serial numbers	Study title	Author	Year of publication	Country of study
1	GCA in SSA	Segal et al.	2001[13]	Systematic review
2	GCA in Africa: current management and outcome	Asombang AW et al.	2014 [12]	Systematic review
3	GCA in Africa: what do we know about incidence and risk factors?	Akwi W. Asombang, et al.	2011[14]	Systematic review
4	GCA: what responsibility is borne by *Helicobacter pylori*? Should it be combated in the African context?	A R P Walker, et al.	2000 [15]	A review
5	Upper Gastrointestinal Disease in Nairobi and Nakuru Counties, Kenya	Makanga W et al.	2014 [16]	Kenya
6	A minimum estimate for the incidence of GCA in Eastern Kenya	G McFarlane et al.	2011 [17]	Eastern Kenya
7	GCA in Zambian adults, a prospective case-control study that assessed dietary intake and antioxidant status by using urinary isoprostane excretion	Akwi W Asombang et al.	2022 [18]	Zambia
8	Gastric Malignancy Survival in Zambia, Southern Africa: A Two-Year Follow-up Study	Akwi A Asombang et al.	2014 [19]	Zambia
9	Epidemiological, clinical, pathological, and therapeutic aspects of GCA in Morocco	Sanna Elmajjaoui et al.	2013 [20]	Morocco

10	Incidence of GCA in Marrakech and Casablanca, Morocco	Brittney L Smith et al.	2015 [21]	Morocco
11	Epidemiological profile of GCA in the northwestern region of Algeria: about 116 cases	Safia Fehim et al.	2017 [22]	Algeria
12	Retrospective epidemiological study on stomach cancer in a region of western Algeria: about 394 cases between 2011 and 2015	Dalale Behar et al.	2020 [23]	Algeria
13	Epidemiology of GCA in Jos University Teaching Hospital Jos: A 20-year review of cases	*Mandong et al.	2010 [24]	Nigeria
14	Management and outcome of gastric carcinoma in Zaria, Nigeria	*Ahmed A et al.	2011 [25]	Nigeria
15	A review of the current profile of GCA presentation in the University College Hospital Ibadan, a tertiary health care institution in the tropics	Oludolapo O et al.	2012 [26]	Nigeria
16	Gastric malignancies and associated pre-malignant lesions in a teaching hospital in South-West Nigeria	Komolafe et al.	2008 [27]	Nigeria
17	Pattern of GCA in North-eastern Nigeria: A clinico-pathological study	Bakari A A et al.	2010 [28]	Nigeria
18	Histopathological analysis of GCAs in the University of Ilorin Teaching Hospital: A 20-year review	Suleiman KA et al.	2018 [29]	Nigeria
19	Gastric carcinoma - A big challenge in a poor economy	O. Clement et al.	2010 [30]	Nigeria
20	Etiologic and clinicopathological correlates of gastric carcinoma in the Egyptian Delta	Mohamed Farouk Akl et al.	2018 [31]	Egypt

21	Gastric carcinoma at Tanta Cancer Center: a comparative retrospective clinicopathological study of the elderly versus the non-elderly	Zeeneldin et al.	2014 [32]	Egypt
22	The epidemiology of cancer in Angola—results from the cancer registry of the National Oncology Centre of Luanda, Angola	Armando et al.	2015 [33]	Angola
23	Gastrointestinal malignancies at five regional referral hospitals in Uganda	Siraji et al	2015 [34]	Uganda
24	Gastric adenocarcinomas in Central Tunisia: Evolution specificities through two decades and relation with *Helicobacter pylori*	Mohamed Amine Elghali et al.	2018 [35]	Tunisia
25	GCA at Gabriel-Touré Teaching Hospital (Bamako, Mali)	Togo et al.	2010 [36]	Mali
26	GCA and *Helicobacter pylori* infection in eastern Libya: a descriptive epidemiological study	Abdel-Naser Elzouki et al.	2012 [37]	Libya
27	GCAs at Kibogora Hospital	G. Ntakiyiruta et al.	2009 [38]	Rwanda
28	Management of digestive cancers in Ivory Coast: Experience of the Gastroenterology Unit of the Yopougon Teaching Hospital	H.Y. Kissi Anzouan-Kacou et al.	2016 [39]	Cote d' Ivoire
29	Clinical epidemiology and mortality risk factors of GCA in a SSA setting: a retrospective analysis of 120 cases in Yaoundé (Cameroon)	Guy Aristide Bang et al.	2020 [40]	Cameroon
30	GCA at a university teaching hospital in northwestern Tanzania: a retrospective review of 232 cases	Joseph B Mabula et al.	2012 [41]	Tanzania

31	The spectrum of GCA as seen in a large quaternary hospital in KwaZulu-Natal, South Africa	Benamro et al.	2017 [42]	South Africa
32	Aspects cliniques et histologiques des cancers de l'Estomac au Centre Hospitalier et Universitaire de Brazzaville	Fortuné et al.	2022 [43]	Congo
33	Malignant gastric tumors in Sudan: a report from a single pathology center	El Hassan et al.	2008 [44]	Sudan
34	Clinical epidemiology and mortality risk factors of gastric cancer in a sub-Saharan African setting: a retrospective analysis of 120 cases in Yaoundé (Cameroon)	Bang GA et al	2020	Cameroon
35	Analysis of the proportion of university teaching hospital gastric cancer data included in the Zambia national cancer registry	Shumba S et al	2023	Zambia
36	Study of epidemiological, clinical, and pathological characteristics of gastric adenocarcinoma in a Moroccan population	Amreni Joutei et al	2020	Morroco
37	Gastric Cancer: Epidemiological Aspects in Morocco	Fadlouallah et al	2014	Morroco
38	Clinical presentation and histopathology of gastric cancer at University Teaching Hospital Brazaville	Bolenga Liboko et al	2022	Congo Brazeville
39	Epidemiological Study and the Stomach, abbott 55 cases	Karamoko Diallo	2008	Mali

40	Management of Cancer of the stomach in Malungo Hospital, Kampala	CBR Ibingira	2001	Uganda

Table 3.2: Distribution of published articles based on regions and countries

Country of study	Frequency
North Africa 35%	
Algeria	2
Egypt	3
Morocco	4
Tunisia	3
Sudan	1
Libya	1
West Africa 40%	
Nigeria	7
Senegal	1
Mali	3
Cameroon	3
Cote d Ivoire	1
Togo	1
Southern Africa 2.5%	
South Africa	1
Eastern Africa 15%	
Kenya	1
Tanzania	1
Rwanda	1
Burundi	1
Uganda	2

Central Africa 7.5%	
Congo Brazzaville	2
Angola	1

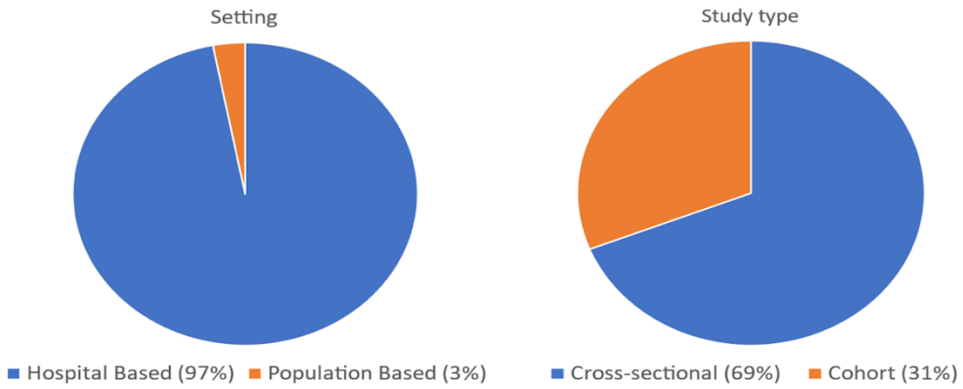


Figure 3.2: Study setting and study type

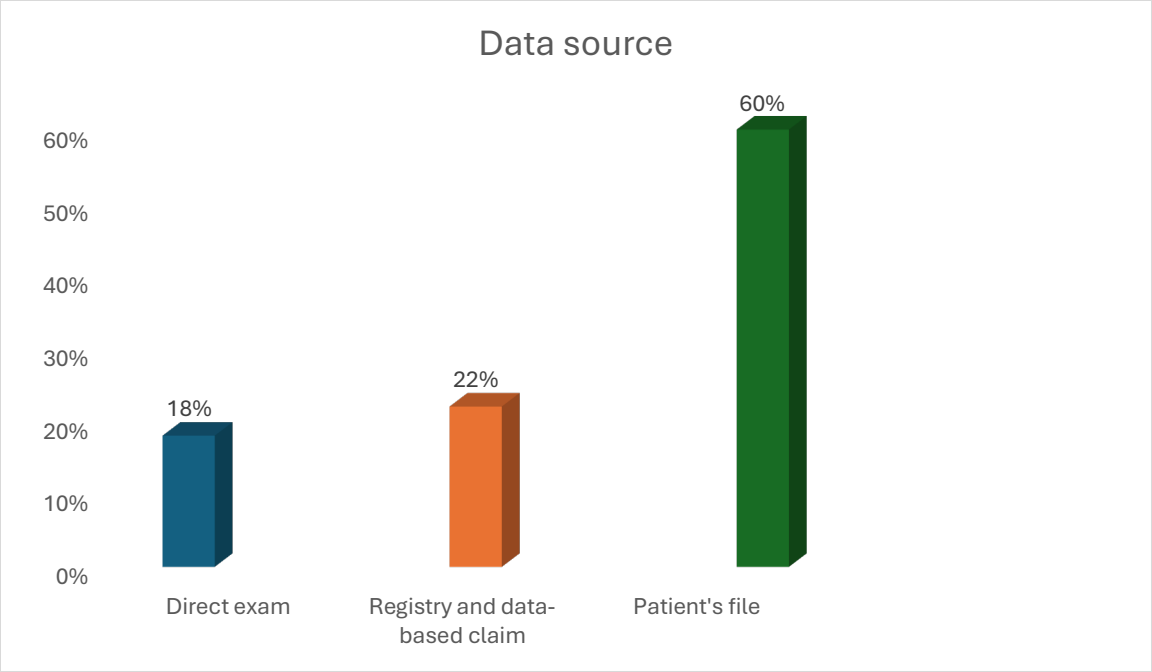


Figure 3.3: Bar charts showing the sources of the study data expressed in percentages

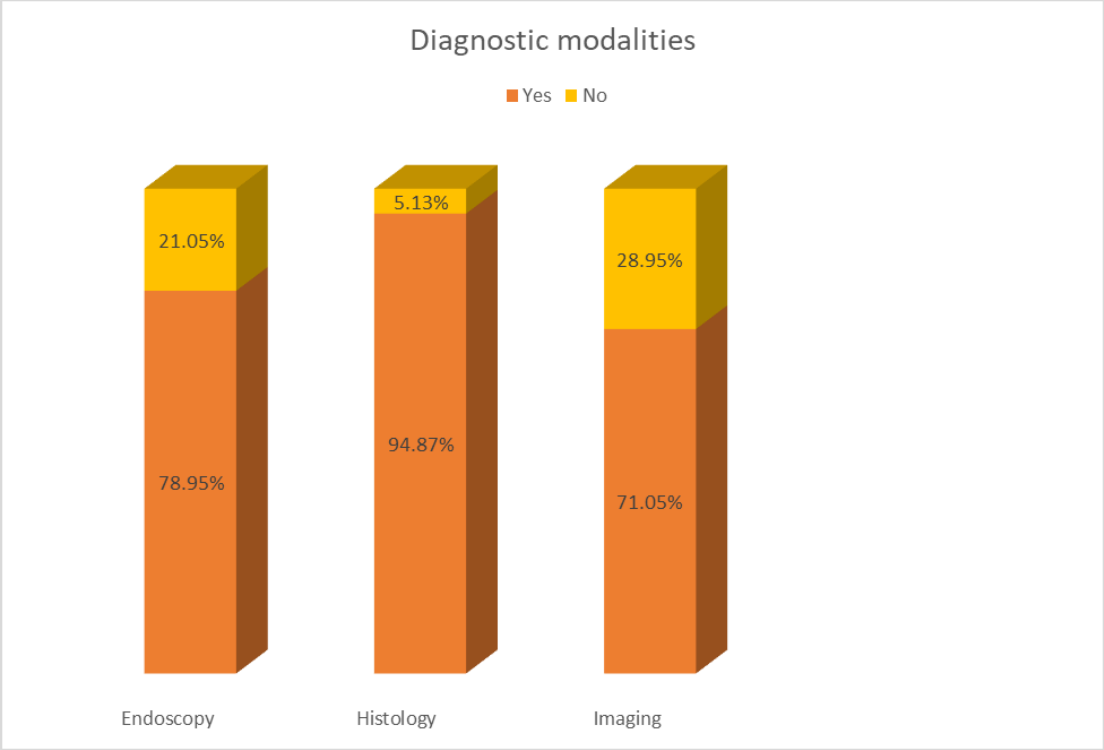


Figure 3.4: Bar charts of diagnostic modalities in the data expressed in percentages

All the studies included in this review were retrospective, with 69% being cross-sectional and 97% of the data sampled was from the hospital setting (Figure 3.2). 60% of the collated data are directly from patient files, while 22% was obtained from the registry and data-based claims (Figure 3.3). Regarding the mode of diagnosis, 79% of the reported cases had access to upper endoscopy, with 95% of the diagnoses having histological confirmation. Access to imaging modalities was limited to 28% of the published data (Figure 3.4).

The median sample size of included studies was 455 (26-8565), with a median 114 (13-876) GCA cases in the total cohort. The median age at diagnosis of GCA was 54 (45-61) years. 52% of the reported GCA cases were in the antrum, 26% located in the gastric body, and 11% in the cardia with only 6 (15%) studies reporting the presence of *H. pylori* in the histological sample. In these studies, only 33% were positive for *H. pylori*. The majority (90%) of the diagnosed GCAs were adenocarcinoma.

3.5 DISCUSSION

In this study, we reviewed the existing data to understand the epidemiology of GCA in Africa. The main findings were that: a) despite a high prevalence of *H. pylori* in Africa, a few studies were done (n=139) and even fewer eligible for inclusion (n=40), b) there was a bias towards hospital-based patients who are likely symptomatic with end-stage disease, and c) the majority of studies were cross-sectional in design, with inherent biases. Arguably, the most important finding was that key elements of data were not reported, which impaired our ability to conduct a meta-analysis. These factors combined speak to the challenges of doing research on the continent, which limits our ability to contribute to knowledge. Notwithstanding, this data provides evidence for the need in Africa for health systems development, collaboration, and an investment in data collection and management, to facilitate our understanding of diseases.

Another factor to consider is the role of availability and access to modern endoscopic procedures on the epidemiology of GCA in Africa. This review shows that patients have limited access to imaging modalities as part of diagnostic and management workup for GCA in African countries, with only 79% having access to upper endoscopy. This is indeed a major limiting factor in assessing the true prevalence of GCA as many patients with GCA will be missed. The findings of Mc Farlene et al. on the sequential rise in the GCA rate both in Kenya and Uganda which correlate with the advent of the introduction of endoscopic services attest to this fact [(10)]. It is therefore important to tease out the confounding factors including the paucity and/or poor quality of available data among African nations as depicted in this review when considering the 'African enigma'.

GCA occurs more commonly among the elderly, which is a phenomenon that is seen globally. In our study, the median age at diagnosis was 54 years. In a retrospective study conducted in Cameroon, Bang et al. reported in a total of 120 GCA cases, the mean age at diagnosis was 53.4 ± 13.7 [(45)] In another retrospective study conducted in Zambia, Samson Shumba et al. reviewed 94 patient records with GCA; they noted a similar mean age of 59 ± 14.9 years [(46)]. These findings may be explained by increasing DNA damage as cells age.

The global burden of GCA remains high with a reported wide variability in prevalence rate globally. The highest prevalence of GCA is reported in East Asian countries [(5)], with *H. pylori* implicated as the main carcinogen associated with GCA, especially the non-cardia types [(4)]. Our review has shown that the dominant form of GCA in the African continent is the non-cardia type

with isolated antral cancers accounting for 52% of reported GCAs; this location correlates with the high prevalence of *H pylori* in Africa.

Also, in this review, only 6 (15%) studies report the presence of *H. pylori* in the histological sample of which 33% were positive for *H. pylori*. In a Peruvian study, Carlos et al. found that, in a total of 288 GCA samples, *H. pylori* detection using a Giemsa stain was 58.2%, with immunohistochemistry the yield was 48.6% and with time PCR the yield was 60.7% [47]. They also reported that haematoxylin and eosin (H&E) stain had good concordance with immunohistochemistry, but poor concordance is noted with PCR [47]. In another study, Gabbo et al. evaluated *H. pylori in* 80 gastric biopsy samples of which 18 had GCA. They noted that using H&E stain, *H. pylori* was detected in 22.2% while the yield was 77.8% using real time PCR [48]. In our review, we noted that though only 6 (15%) studies reported that *H. pylori* status, the positivity was low at 33%, but still within the same range as other studies. A possible explanation to this low incidence of *H. pylori* in gastric cancers is the hit and run theory postulated by Hatakeyama. He explained that *cagA*-induced gastric carcinogenesis progresses through a hit-and-run mechanism in which pro-oncogenic actions of *cagA* are propagated by a series of genetic and/or epigenetic alterations compiled in cancer-predisposing cells during long-standing infection with *cagA*-positive *H. pylori.*[49] Therefore following the initial alterations caused by *cagA* in *H. pylori* infected gastric cells, its oncogenic impact could persist post eradication due to the in situ changes it triggered such as altered pH and microbiome changes, which eventually leads to the disappearance of the bacterium while the oncogenic properties are sustained by internal mechanisms. Other plausible explanations are the possible effect of prior eradication treatment for *H. pylori*, or unfavourable environment due to the gastric cancer.

3.5.1 The strengths and weakness of this review

There are limited studies on the epidemiology of GCA in Africa, hence the significance of this review. The findings from this review also reveal the gaps in the study of GCA in Africa, suggesting the need for further research. An additional strength is that where endoscopy was performed, a significant proportion of GCAs were diagnosed on histology, thus reducing the risk of ascertainment bias. This systematic review revealed weaknesses in the reported data with missing data, and in many studies background sample/population sizes were not reported, making determinations of prevalence and incidence including a meta-analysis not possible. Moreso, there

was a high heterogeneity in the data and because of the few publications, quality assessment was not done.

3.6 CONCLUSIONS

The paucity of quality data on GCA in Africa, as revealed in this study, provides no sufficient evidence to support or contradict the argument for the 'African enigma'. Given this, reliable country registry data is urgently needed, which should include accurate data collection during endoscopy and histology, and hospital/population data. There is also the need to conduct well-designed case-control studies that accurately assess exposures (*H. pylori,* others) in a bid to establish the contributions of these risk factors in the development of GCA in the African cohort. To maximize these studies, collaboration (unified protocols) among African countries will be important.

3.7 REFERENCES

1. Thrift AP, El-Serag HB. Burden of GCA. Clinical Gastroenterology and Hepatology. 2020;18(3):534-42.

2. Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. GCA: Descriptive Epidemiology, Risk Factors, Screening, and Prevention GCA. Cancer epidemiology, biomarkers & prevention. 2014;23(5):700-13.

3. Hoed CM, Kuipers EJ. GCA: how can we reduce the incidence of this disease? Current gastroenterology reports. 2016;18:1-8.

4. Malfertheiner P, Camargo MC, El-Omar E, Liou J-M, Peek R, Schulz C, et al. *Helicobacter pylori* infection. Nature Reviews Disease Primers. 2023;9(1):19.

5. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2021;71(3):209-49.

6. Kidd M, Louw JA, Marks IN. *Helicobacter pylori* in Africa: observations on an 'enigma within an enigma'. Journal of gastroenterology and hepatology. 1999;14(9):851-8.

7. Grantham T, Ramachandran R, Parvataneni S, Budh D, Gollapalli S, Gaduputi V. Epidemiology of Gastric Cancer: Global Trends, Risk Factors and Premalignant Conditions. Journal of Community Hospital Internal Medicine Perspectives. 2023;13(6):100.

8. Ladeiras-Lopes R, Pereira AK, Nogueira A, Pinheiro-Torres T, Pinto I, Santos-Pereira R, et al. Smoking and GCA: systematic review and meta-analysis of cohort studies. Cancer causes & control. 2008;19(7):689-701.

9. Kidd M, Peek R, Lastovica A, Israel D, Kummer A, Louw J. Analysis of iceA genotypes in South African *Helicobacter pylori* strains and relationship to clinically significant disease. Gut. 2001;49(5):629-35.

10. Fox JG, Beck P, Dangler CA, Whary MT, Wang TC, Shi HN, et al. Concurrent enteric helminth infection modulates inflammation and gastric immune responses and reduces helicobacter-induced gastric atrophy. Nature medicine. 2000;6(5):536-42.

11. McFarlane G, Forman D, Sitas F, Lachlan G. A minimum estimate for the incidence of GCA in Eastern Kenya. British Journal of Cancer. 2001;85(9):1322-5.

12. Asombang AW, Rahman R, Ibdah JA. GCA in Africa: current management and outcomes. World Journal of Gastroenterology: WJG. 2014;20(14):3875.

13. Segal I, Ally R, Mitchell H. Gastric cancer in sub-Saharan Africa. European Journal of Cancer Prevention. 2001;10(6):479-82.
14. Asombang AW, Kelly P. Gastric cancer in Africa: what do we know about incidence and risk factors? Trans R Soc Trop Med Hyg. 2012;106(2):69-74.
15. Walker AR, Segal I, Adam F. Gastric cancer: what responsibility is borne by *Helicobacter pylori*? Should it be combated in the African context? Eur J Cancer Prev. 2000;9(1):1-4.
16. Makanga W, Nyaoncha A. Upper gastrointestinal disease in Nairobi and Nakuru counties, Kenya; a two-year comparative endoscopy study. Annals of African Surgery. 2014;11(2):35-9.
17. McFarlane G, Forman D, Sitas F, Lachlan G. A minimum estimate for the incidence of gastric cancer in Eastern Kenya. Br J Cancer. 2001;85(9):1322-5.
18. Asombang AW, Kayamba V, Mwanza-Lisulo M, Colditz G, Mudenda V, Yarasheski K, et al. Gastric cancer in Zambian adults: a prospective case-control study that assessed dietary intake and antioxidant status by using urinary isoprostane excretion. Am J Clin Nutr. 2013;97(5):1029-35.
19. Asombang A, Turner-Moss E, Banda L, Trinkaus K, Kayamba V, Colditz G, et al. Gastric Malignancy Survival Outcomes in Zambia, Southern Africa. American Journal of Gastroenterology. 2012;107:S598-S.
20. Elmajjaoui S, Ismaili N, Zaidi H, Elkacemi H, Hassouni K, Kebdani T, et al. Epidemiological, clinical, pathological, and therapeutic aspects of gastric cancer in Morocco. Clinical Cancer Investigation Journal. 2014;3(1):3-8.
21. Smith BL, Khouchani M, Karkouri M, Lazenby AJ, Watkins K, Tahri A, et al. Incidence of Gastric Cancer in Marrakech and Casablanca, Morocco. J Cancer Epidemiol. 2015;2015:704569.
22. Fehim S, Bouhaous R, Diaf M, Drici AM, Khaled MB. Epidemiological profile of gastric cancer in the northwestern region of Algeria: about 116 cases. J Gastrointest Oncol. 2017;8(4):659-64.
23. Behar D, Boublenza L, Chabni N, Hassaine H, Dahmani B, Masdoua N, et al. Retrospective epidemiological study on stomach cancer in a region of western Algeria: about 394 cases between 2011 and 2015. J Gastrointest Cancer. 2021;52(2):706-10.
24. Mandong BM, Manasseh AN, Tanko MN, Echejoh GO, Madaki AJ. Epidemiology of gastric cancer in Jos University Teaching Hospital Jos a 20-year review of cases. Niger J Med. 2010;19(4):451-4.

25. Ahmed A, Ukwenya AY, Makama JG, Mohammad I. Management and outcome of gastric carcinoma in Zaria, Nigeria. Afr Health Sci. 2011;11(3):353-61.
26. Afuwape OO, Irabor DO, Ladipo JK, Ayandipo B. A review of the current profile of gastric cancer presentation in the university college hospital Ibadan, a tertiary health care institution in the tropics. J Gastrointest Cancer. 2012;43(2):177-80.
27. Komolafe AO, Ojo OS, Olasode BJ. Gastric malignancies and associated pre-malignant lesions in a teaching hospital in South-West Nigeria. African Journal of Biotechnology. 2008;7(13):2104-11.
28. Bakari AA, Ibrahim AG, Gali BM, Dogo D, Nggada HA. Pattern of gastric cancer in North-eastern Nigeria: a clinicopathological study.
29. Suleiman KA, Afolayan EA, Ibrahim OO, Ahmed A, Abubakar-Akanbi SK. Histopathological analysis of gastric cancers in university of Ilorin teaching hospital: a 20-year review. Journal of BioMedical Research and Clinical Practice. 2021 Apr 30;4(1):85-95.
30. Clement Osime O, Momoh MI, Irowa OO, Obumse A. Gastric carcinoma - A big challenge in a poor economy. Journal of Gastrointestinal Cancer. 2010;41(2):101-6.
31. Akl MF, Ibrahem MA, Khater A, El-zahaf E, Farag K, Abdallah H. Etiologic and Clinicopathological Correlates of Gastric Carcinoma in the Egyptian Delta. Indian Journal of Surgical Oncology. 2018;9(4):472-6.
32. Zeeneldin AA, Ramadan H, El Gammal MM, Saber MM, Elgamal D, Sherisher MA. Gastric carcinoma at Tanta Cancer Center: a comparative retrospective clinico-pathological study of the elderly versus the non-elderly. J Egypt Natl Canc Inst. 2014;26(3):127-37.
33. Armando A, Bozzetti MC, de Medeiros Zelmanowicz A, Miguel F. The epidemiology of cancer in Angola—results from the cancer registry of the national oncology centre of Luanda, Angola. ecancermedicalscience. 2015;9.
34. Obayo S, Lukwago L, Orem J, Faulx AL, Probert CS. Gastrointestinal malignancies at five regional referral hospitals in Uganda. Afr Health Sci. 2017;17(4):1051-8.
35. Eighali MA, Gouader A, Bouriga R, Mahjoub M, Jarrar MS, Ziadi S, et al. Gastric Adenocarcinomas in Central Tunisia: Evolution Specificities through Two Decades and Relation with *Helicobacter pylori*. Oncology. 2018;95(2):121-8.
36. Togo A, Diakité I, Togo B, Coulibaly Y, Kanté L, Dembélé BT, et al. Gastric cancer at Gabriel-Touré teaching hospital (Bamako, Mali). Journal Africain du Cancer. 2011;3(4):227-31.

37. Elzouki AN, Buhjab SI, Alkialani A, Habel S, Sasco AJ. Gastric cancer and *Helicobacter pylori* infection in the eastern Libya: a descriptive epidemiological study. Arab J Gastroenterol. 2012;13(2):85-8.
38. Ntakiyiruta G. Gastric cancers at Kibogora hospital. East and Central African Journal of Surgery. 2009;14(1):130-4.
39. Kissi Anzouan-Kacou HY, Doffou SA, Bangoura AD, Kouamé DH, Fanou CD, Bathaix YF, et al. Management of digestive cancers in Ivory Coast: Experience of the Gastroenterology unit of the Yopougon Teaching Hospital. Journal Africain d'Hepato-Gastroenterologie. 2017;11(1):13-8.
40. Bang GA, Savom EP, Oumarou BN, Ngamy CKM, Moto GB, Boukar YME, et al. Clinical epidemiology and mortality risk factors of gastric cancer in a sub-Saharan African setting: a retrospective analysis of 120 cases in Yaoundé (Cameroon). Pan Afr Med J. 2020;37:104.
41. Mabula JB, McHembe MD, Koy M, Chalya PL, Massaga F, Rambau PF, et al. Gastric cancer at a university teaching hospital in northwestern Tanzania: a retrospective review of 232 cases. World J Surg Oncol. 2012;10:257.
42. Benamro F, Sartorius B, Clarke DL, Anderson F, Loots E, Olinger L. The spectrum of gastric cancer as seen in a large quaternary hospital in KwaZulu-Natal, South Africa. S Afr Med J. 2017;107(2):130-3.
43. Bolenga Libiko, A.F., Dagnagnéwendé, D., Ndingossoka, R.J., Rissia, F. and Ndounga, E. (2022) Aspects cliniques et histologiques des cancers de l'estomac au Centre Hospitalier et Universitaire de Brazzaville. *Health Sciences and Disease*, 23, 63-68.
44. El Hassan A, El Hassan L, Mudawi H, Gasim B, Own A, Elamin E, et al. Malignant gastric tumors in Sudan: a report from a single pathology center. Hematol Oncol Stem Cell Ther. 2008;1(2):130-2.
45. Bang GA, Savom EP, Oumarou BN, Ngamy CK, Binyom PR, Essomba A, Sosso MA. Clinical epidemiology and mortality risk factors of gastric cancer in a sub-Saharan African setting: a retrospective analysis of 120 cases in Yaoundé (Cameroon). Pan African Medical Journal. 2020 Sep 30;37(1).
46. Shumba S, Kayamba V. Analysis of the proportion of university teaching hospital gastric cancer data included in the Zambia national cancer registry. Afr Health Sci. 2023 Mar;23(1):438-443. doi: 10.4314/ahs.v23i1.46. PMID: 37545961; PMCID: PMC10398468.

47. Castaneda CA, Castillo M, Sanchez J, Casavilca S, Sanchez J, Bernabe LA, Suarez N, Chavez I, Ruiz E, Tello K, Villa MR. Detection of Helicobacter pylori in gastric cancer tissue through histopathology, immunohistochemistry and real-time reverse transcription-PCR. Future Microbiol. 2020 Aug;15:1131-7.

48. Gobo A, Maluf P, Marques S, Rahal P, Silva A. Comparison of histological and molecular diagnosis of *Helicobacter pylori* in benign lesions and gastric adenocarcinoma. Brazilian J. Microbiol. 36(1), 12–16 (2005).

49. Hatakeyama M. Helicobacter pylori CagA and gastric cancer: a paradigm for hit-and-run carcinogenesis. Cell host & microbe. 2014 Mar 12;15(3):306-16.

CHAPTER 4
RETROSPECTIVE COHORT: EPIDEMIOLOGICAL PROFILE AND INCIDENCE OF GASTRIC CANCER IN GROOTE SCHUUR HOSPITAL

ABSTRACT

Background: Gastric cancer (GCA) is the 5th most common cancer and 3rd leading cause of cancer-related death. *H. pylori* is the chief risk factor. The pooled prevalence of *H. pylori* in Africa is 53%, despite this, GCA appears to be low, the so-called 'African enigma'.

Aim: To document the epidemiological profile and risk factors of GCA in a South African cohort.

Methods: A retrospective folder review of patients with GCA from 2018-2022.

Results: 443 patients were included; median age 63 (range 29-86), 65% male (61% ≥ 60 years old). Hypertension (38.2%) and diabetes (17.6%) were the main comorbidities. 46% and 8.4% had a history of smoking and alcohol use respectively. NSAID and PPI use was reported in 8.8% and 7% respectively. The commonest symptom was weight loss (38.6%) and epigastric pain (29%). 68% had a Hb<12g/dl with a median Hb=10 (range 3.8-17). Endoscopically, gastritis duodenitis or ulcers were recorded in 63% of cases. The commonest site of GCA was non-cardia (82.2%); antral and corpus cancers in 40.9% and 53.2% respectively. Histologically, *H. pylori* was reported in 11.5%, while chronic atrophic gastritis (CAG) in 15.6%. *H. pylori* was not associated with either antral or corpus cancer, or CAG, however PPIs were associated with a reduced odds of antral cancer, OR=0.42, CI=0.18-0.97, p=0.04. None of the other demographic or clinical factors were associated with the location of GCA.

Conclusions: GCA is higher in males and those ≥ 60 years. Loss of weight and anaemia are key red flags. Non-cardia GCAs are the most common in this cohort. The low prevalence of *H. pylori* suggests previous eradication, although CAG was high, thus maintaining the increased risk of non-cardia CGA. Interestingly, PPI use was significantly protective against developing antral cancer. This highlights the need to treat *H. pylori* before CAG occurs.

4.1 BACKGROUND

4.1.1 Global burden of GCA

According to GLOBOCAN 2020, GCA remains the 5th commonest cancer and the 3rd commonest cause of cancer death for the 2020 statistics, with over 1 million new cases per year [(1)]. Though

there had been a steady decline in the incidence and mortality rate for gastric cancer globally over the years, it remained one of the cancers with a higher burden and poorer prognosis. While a total of 1, 033 701 new cases of GCA (representing 5.7% of all cancer cases diagnosed) and 782 685 deaths related to GCA were reported according to the 2018 estimates from the International Agency for Research on Cancer (IARC) GLOBOCAN project [2], there seems to be no significant difference from the 2020 statistics.

According to the GLOBOCAN 2020, a total 1, 089 103 new cases of GCA were reported, with the African continent contributing 32 403 (3%) of these cases [1]. There is a wide variability in the reported incidence of GCA across the various countries and regions, Africa continent inclusive. The incidence of GCA is highest in Eastern Asia, with an incidence of 32.5/100 000 reported in Mongolia, 31.6/100 000 in Japan, 27.9/100 000 in Korea and 20.6/100 000 in China [1]. In Northern America and Northern Europe however, incidence rates are generally low and equivalent to those seen across the African regions [1].

In South Africa it was reported that GCA is the 16th most common cancer with an annual incidence of 3.5/100 000, in Botswana, GCA ranked 16th with an incidence rate of 1.8/100 000 and in Zimbabwe 9.4/100 000 [1]. In Nigeria, GCA ranked number 10 with a total of 2 621 new cases, the incidence rate was 2.8/100 000, with a mortality rate of 2.5/100 000. The incidence of GCA in Africa is highest in the eastern region of the continent with Mali and Kenya reporting rates of 12.8/100 000 and 7.6/100 000 respectively [1]. Among the Northern countries, the incidence of GCA for the same year reports the following incidence rates in Morocco 5.7/100 000, Algeria 5.7/100 000 and Egypt 4.1/100 000 [1]. Despite, this regional variability, the incidence of GCA remains low compared to the East Asian countries and many developed countries (Figure 4.1) [1].

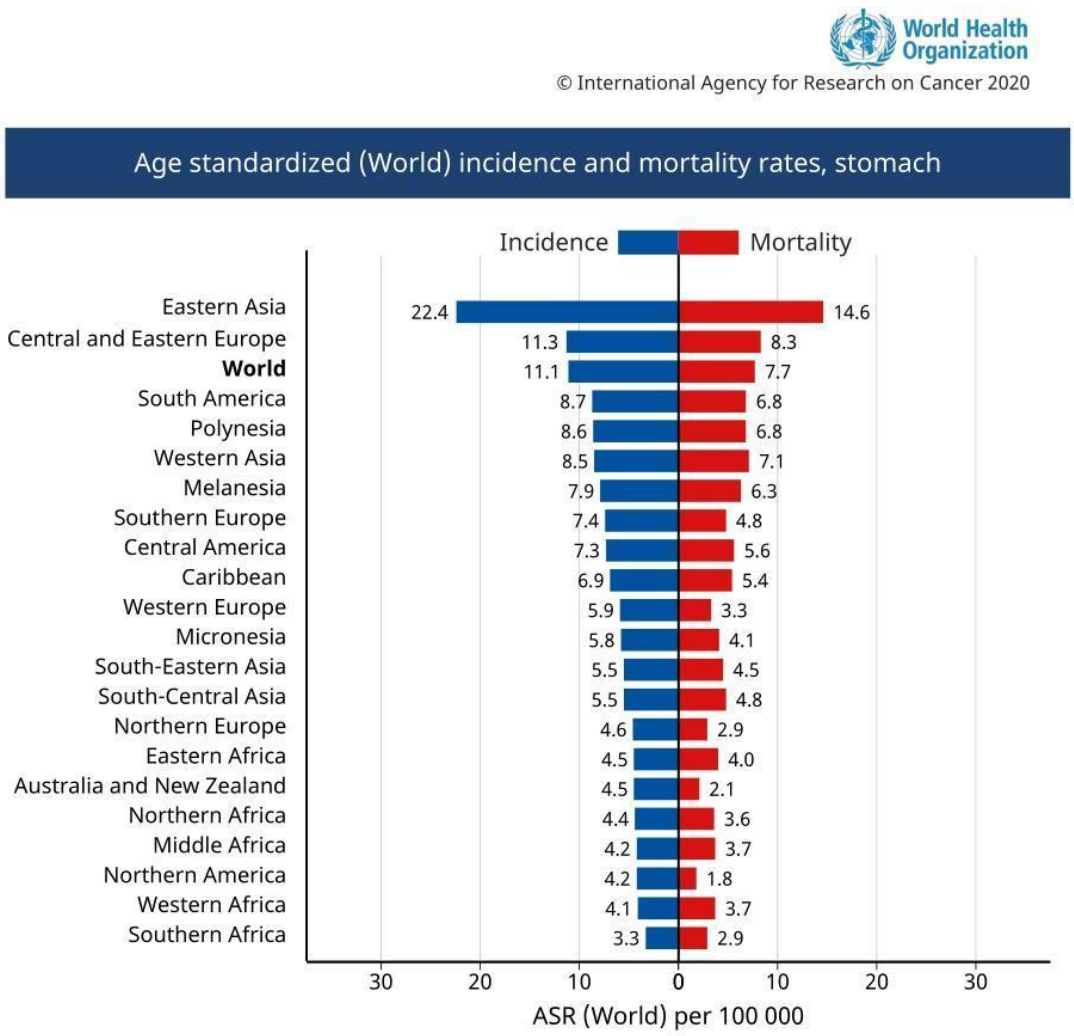


Figure 4.1 Incidence and mortality of GCA across various regions

4.1.2 *Helicobacter pylori* as a risk factor for GCA and the 'African enigma'

Chronic *H. pylori* infection is reported to be a major risk factor in the development of GCA globally, especially distal GCAs. It has a causal relationship with various diseases such as gastric ulcers, duodenal ulcers, non-ulcer dyspepsia, mucosal-associated gastric lymphoma and GCA, [3] and is reported to be responsible for about 66% of the global GCA [4], hence it was designated as a type 1 carcinogen by the IARC [5]. Ironically, despite its highest prevalence in SSA with a prevalence rate of 61-100% depending on the country, [6] GCA in Africa is reported with a low incidence rate when compared to the Asian continent with a low prevalence *of H. pylori* infection

of 20-40% [6], and higher incidence rate of GCA. Other risk factors for GCA include smoking, alcohol, high salt diet and a diet low in vegetables, Epstein Bar Virus, genetic factors, Menetrier's disease and autoimmune gastritis.

This disproportionate low prevalence of GCA in Africa despite the ubiquitous nature of *H. pylori* is referred to as the 'African Enigma' [7,8] and is yet to be understood. The possibility that the enigma exists, could be explained by inadequate diagnostic facilities in Africa, or cytokines and/or molecules that may confer protection against GCA among Africans or because of a less virulent strain of the *H. pylori*.

A retrospective study by McFarlane *et al* done in the Eastern part of Kenya between 1991-1993 compared to data from the same area obtained between 1965-1970 showed a 10-fold increase in the incidence rate of GCA [9,10]. An important note was that endoscopy services were established in the Eastern part of Kenya in the 1980s, which could have impacted on diagnostic capabilities and the increased recognition of GCA [9]. Similar findings of increased GCA incidence over decades in Uganda have been documented but attributed to increased access to healthcare and endoscopy availability, with most recent data showing a seven-fold increase in the incidence of GCA from 0.8/100 000 in the 1960s to 5.6/100 000 [10,11]. However, despite this increase in the incidence of GCA in these African regions, the reported incidence is relatively lower compared to East Asian countries. There is also the issue of paucity of data and poor registry data from the African subcontinent regarding the epidemiology of GCA.

This study is therefore designed to study the prevalence of GCA in a South African cohort and its associated risk factors. The data obtained will be useful for designing prospective studies that will improve our understanding of *H. pylori* in our context.

4.2 METHODS

4.2.1 Study design

We retrospectively assessed the epidemiological profile of patients with biopsy-proven GCA.

4.2.2 Population of study

We included all adult patients (≥ 18 years of age) with histologically proven GCAs diagnosed between March 2018 and September 2022. Patients with no histological evidence were excluded.

4.2.3 Setting

This study is a single-centre study conducted in GSH, a tertiary/quaternary hospital in Cape Town in the Western Cape region which serves an area with a population of about 3.2 million people.

4.2.4 Procedures

Patients diagnosed with GCA were identified from an upper endoscopy. The REDCap electronic database was used, and folder numbers of identified patients were subsequently retrieved. Thereafter, we assessed the hospital folders and the National Health Laboratory Service electronic database for histological reports, blood results, documented demographic profiles, and clinical risk factors.

4.2.5 Data collection

Data collected included demographics such as age at diagnosis, sex, comorbidities, alcohol use, smoking history, use of NSAIDs or proton pump inhibitors, presenting symptoms, haemoglobin level at presentation, and indications for gastroscopy. We also collated endoscopic findings such as the presence of ulcers or inflammation, the location of cancer, and histological findings including the presence of *H. pylori.*

4.2.6 Ethics

This study received approval from the University of Cape Town Human Research Ethics Committee (HREC: Ref 552/2021) and approval from the hospital.

4.2.7 Data analysis

Data exploration and analysis was done in Stata (Version 13.1; Stata Corp, College Station, Texas, USA) ®. Descriptive statistics of mean, median, standard deviation, quartile, and interquartile ranges were used to characterize the sample in terms of history, demographics, biodata, clinical presentation, and endoscopic and histologic findings. Frequencies and percentages were used for categorical variables. Noncontinuous variables are represented using bar charts, pie charts, and frequency distribution, while for continuous variables, means and (± standard deviations) were used for normally distributed continuous variables, and- medians and (interquartile ranges) for skewed data. Where appropriate, the means are compared using a two-sample t-test or its non-

parametric equivalent. Categorical variables were compared using nonparametric tests. For all, a test of significance and the p-values were calculated. P-values of <0.05 were considered statistically significant.

4.3 RESULTS

A total of 443 patients were included in this cohort, spanning over a four-and-a-half-year period (March 2018-September 2022). The annual incidence was 2.98/100 000. The mean age of this population was 61.7±12. 61.2% of these patients were ≥ 60 years of age. The majority (63.2%) were males (Table 4.1) whilst 38.2% were hypertensive, and 17.6% were diabetic. At least 8.4% and 8.8% had a history of smoking and alcohol use respectively. In addition, while 8.8% of this cohort reported NSAID use, PPI use was reported in 7.0% of patients (Figure 4.2).

The commonest symptoms and indications for endoscopy in this cohort are weight loss reported in 38.6% and epigastric pain in 29.1%. Anaemia was a common finding with a mean haemoglobin of 10.21 ± 3.2, and 67.9% had haemoglobin ≤12g/dl (Table 4.2).

Endoscopically, gastritis, duodenitis or ulcers were recorded in 63.0% of cases. The commonest site of GCA is non-cardia, encountered in 82.2% of the cohort. Gastric antral cancers occurred at 40.86% while cancers involving corpus were reported at 53.3% (Table 4.3 and Figure 4.3).

Histologically, *H. pylori* was reported in 11.5%, while chronic atrophic gastritis (CAG) was reported in 15.6%. 90% of the reported cancers were adenocarcinomas (Table 4.4). *H. pylori* was not associated with either antral, corpus, or CAG; however, PPIs were associated with reduced odds of antral cancer, OR=0.42, CI=0.18-0.97, P=0.04. None of the other demographic or clinical factors were associated with the location of GCA.

Table 4.1: Biodata and social status of the patients

	Frequency	**Percentages**
Age ≥ 60 @ diagnosis		
Yes	272	61.4
No	171	38.6
Sex distribution		
Female	152	34.4
Male	288	65.2
NR	2	0.5
Smoking history		
Yes	203	45.8
No	189	42.7
NR	51	11.5
Alcohol history		
Yes	37	8.4
No	320	72.2
NR	86	19.4

*NR = not recorded

Table 4.2: Distribution of symptoms and signs

	Frequency	**Percentages**
Epigastric pain		
Yes	129	29.1
No	269	60.7
NR	45	10.2
Loss of weight		
Yes	171	38.6
No	229	51.7
NR	43	9.71
Haemoglobin ≤ 12g/dL		
Yes	249	67.9
No	118	32.1

*NR = not recorded

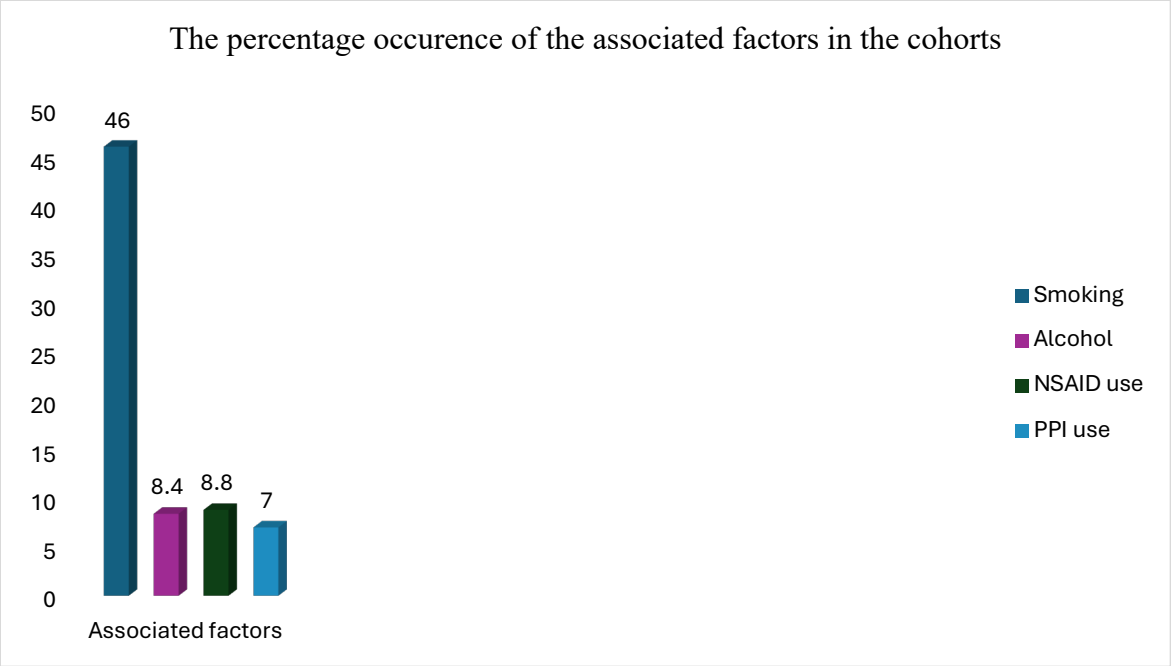


Figure 4.2: Bar charts showing the frequencies of lifestyle factors and drugs associated with GCA in our cohort

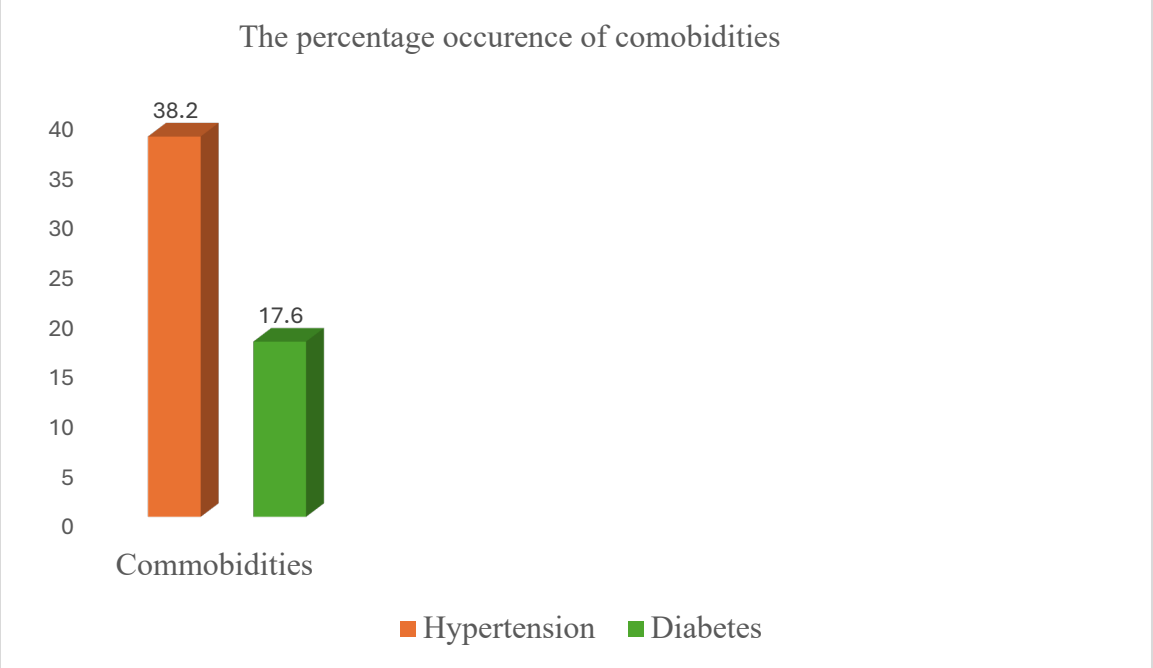


Figure 4.3: Bar charts showing the frequencies of comorbidities associated with GCA in our cohort

Table 4.3: Reported endoscopic findings during gastroscopy

	Frequency	Percentages
Ulcer, gastritis /duodenitis		
Yes	279	63.0
No	26	5.9
NR	138	31.1
Non-cardia cancer		
Yes	364	82.2
No	75	17.0
NR	4	0.9
Cancer involving the corpus		
Yes	236	53.3
No	203	45.8
NR	4	0.9
Antral cancer		
Yes	181	40.9
No	258	58.2
NR	4	0.9

*NR = not recorded

Table 4.4: Histological findings showing the occurrence of atrophy and *H. pylori*

	Frequency	**Percentages**
Chronic atrophic gastritis on biopsy		
Yes	69	15.5
No	51	11.5
NR	323	72.9
***H. pylori* on biopsy**		
Yes	51	11.5
No	242	54.6
NR	150	33.9

*NR = not recorded

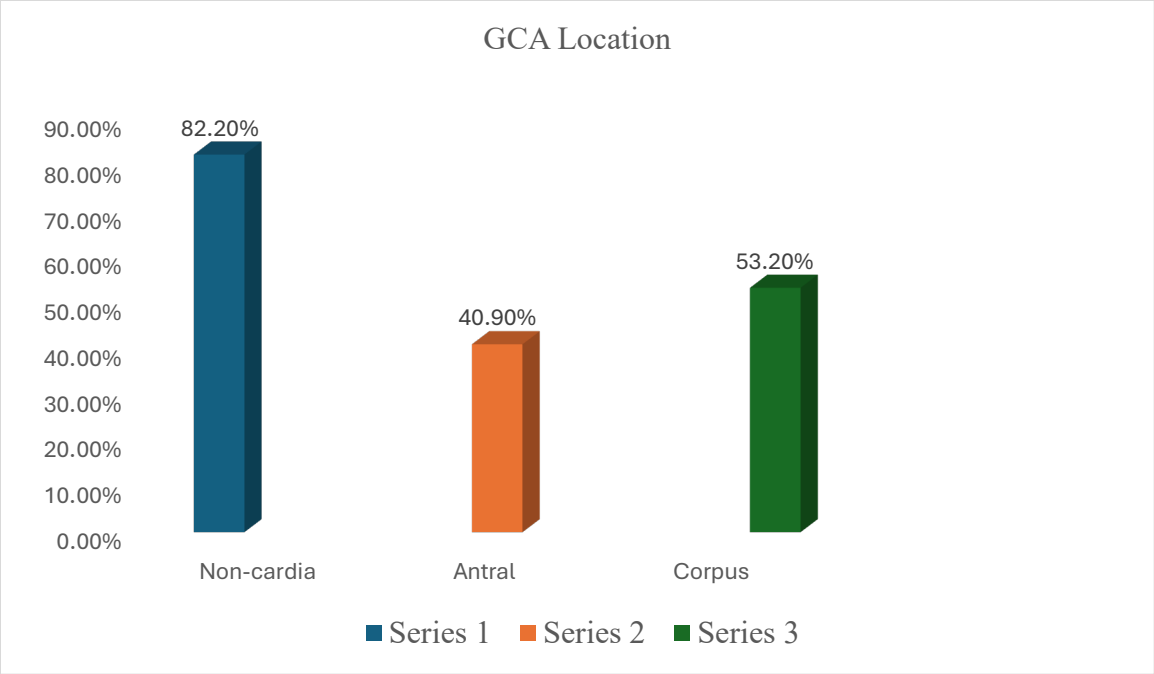


Figure 4.4: Bar charts showing the distribution of GCA based on anatomical locations

4.4 DISCUSSION

This study which sought to understand the incidence and associated risk factors of GCA in South Africa, had a total of 443 patients with histologically confirmed GCA included. Important findings in this study include the fact that the annual GCA incidence in our study was 2.98/100 000. This estimate is based on the premise that the Groote Schuur Hospital covers a drainage area with a population of 3.3 million. Also, a record of 443 gastric cancers over 4.5 years equates to 98.4 gastric cancers per year, and therefore the calculated annual incidence. We also noted that GCA in our population was commoner among males, with 61.2 % of this cohort being older than or equal to 60 years of age. Also, non-cardia cancers were the most common, seen in 82.2% of the cases. Adenocarcinoma was the predominant histological type of GCA noted in 90% of the reported cases.

Histologically, *H. pylori* was reported in 11.5%, while chronic atrophic gastritis (CAG) in 15.6%. *H. pylori* was not associated with either antral or corpus cancer, or CAG, however, PPIs were

associated with a reduced odd of antral cancer, OR=0.42, CI=0.18-0.97, p=0.04. None of the other demographic or clinical factors were associated with the location of GCA.

According to the International Agency for Research on Cancer published by WHO 2020, the annual incidence of GCA in Africa ranges from 3.3/100 000 to 4.4/100 000 with the lowest incidence reported in Southern Africa. Globocan 2020 for South Africa also reported GCA as the 16th most common cancer with an annual incidence of 3.5/100 000 [1]. In Botswana, the incidence was reported at 1.8/100 000 and in Zimbabwe, the reported annual incidence was 9.4/100 000. The incidence of GCA is also reported to be low in developed countries like North America 4.2/100 000 and Australia/New Zealand 4.5/100 000. On the contrary, the East Asian countries have a significantly higher annual incidence of 22.4/100 000 according to the IARC 2020 [1]. In our study, we noted a yearly incidence of 2.98/100 000. These findings are similar to what was reported by Globocan 2020 [1]. Holcombe in his leading publication had reported that while gastritis is common among Africans infected with *H. pylori,* the incidence of gastric ulcers and GCAs are lowest despite the high prevalence of *H. pylori* [7]. This had been named the 'African enigma' and remained an area of research interest. Several debates emanated from this study, with plausible arguments that the enigma is not true, but rather, Africa is laden with inadequate diagnostic capabilities. There is also the consideration of less virulent strains and mutants and possibly counter-inflammatory pathways due to helminthiasis or other parasitic infections. Considering that this study was conducted in a tertiary/quaternary hospital with easy access to endoscopy and imaging, one would think that possibly the 'African enigma' is true and hence warrants further study to unravel this enigma.

Although the annual incidence of GCA globally is 15.6-18.1/100 000 for men and 6.7-7.8/100 000 for women, non-cardia GCAs account for 82% of all GCAs [2,5]. There is variable distribution for cardiac and non-cardiac GCAs geographically and for both sexes. Particularly, male sex had been associated with increased risk in the development of GCAs and NCGCs [12]. The exact reason for this is not well known however there are speculations that the female sex hormones may have some protective benefit against GCA.

In 2012, 260 000 cardia GCA with age-standardized ratio (ASR) of 3.3/100 000 and 691 000 non-cardia gastric (NCGC) cases with ASR of 8.8/100 000 were reported globally, with the East /Southeast Asian countries recording the highest rates with the mean ratio for male to female of 2:1 [12]. In another study by Qiang Yao et al. a total of 18 997 new cases of cardia cancer (14 614

males and 4 383 females) and 38 537 non-cardia GCA (21 134 males and 17 403 females), given the male-to-female ratios in the age-standardized incidence rate of 4.2 for cardia cancer and 1.6 for non-cardia GCA [13]. Overall, incidence rates for cardia GCA are lower than those for non-cardia GCA (in 2012, reported cases were 3.3 per 100 000 persons compared with 8.8 per 100 000 persons). Among men, the ratio of non-cardia GCA to cardia GCA is as much as 40:1 in SSA. However, it is almost 1:1 in Northern America and Oceania. In the U.K., the rates of cardia GCA are 1.5-fold higher than those for non-cardia GCA (in 2012, 3.9 per 100 000 men compared with 2.6 per 100 000 men). For women in Northern America and Oceania, incidence rates for non-cardia GCA remain 2-fold higher than those for cardia GCA. Rates for cardia and non-cardia GCAs among women in the U.K. are comparable (1.5 and 1.7 per 100 000 women, respectively) [12].

In our cohort, 65.2% of the GCAs reported are males with a male-to-female ratio of 1:1.9. According to the South Africa cancer registry for 2017, the male-to-female ratio was reported as 1:1.75. We also noted in our study that non-cardia GCAs were the predominant type of cancer, accounting for 82.2% of the cases, with the male to female ratio of 1:1.6 which is similar to what is obtainable in East Asian countries.

In our study, we noted that the incidence of *H. pylori* on histology was 11.5% while CAG was 15.6%. *H. pylori* is the major risk factor for non-cardia GCA and B-cell lymphoma [14], but the exact magnitude of the association between *H. pylori* and GCA has always been a dilemma because the *H. pylori* infection and circulating antibody response can be lost following the development of GCA. It has been estimated that *H. pylori* is responsible for about 90% of non-cardia GCAs globally [15,16].

In a combined analysis of 12 case-control studies nested within prospective cohorts, the relative risk for NCGC was reported as 3.0 (95% Cl 2.3-3.8) and was stronger when blood samples for *H. pylori* were collected at least 10 years before the GCA diagnosis, with relative risk of 5.9 at 95% CI of 3.4-10.3 given attributable factor for *H. pylori* at 74.7%, with an average prevalence of *H. pylori* infection 90% in these cases. These assays were based on ELISA test. In a second review, Martyn Plummer et al. estimated the attributable factor of *H. pylori* to NCGC using immunoblot (Western blot, which is more sensitive compared to the ELISA) from prospective studies, they noted that the fraction of all cancers and NCGC attributable to *H. pylori* was higher at 6.2 and 89% respectively [16].

Although the reported incidence of *H. pylori* in our cohort was 11.5%, these were based on histological findings only, more so with 15.6% of CAG noted in this study, this will imply the significant role of *H. pylori* in the development of GCA in this cohort. One can also presuppose that the reported incidence of *H. pylori* would be higher if a serological test was performed.

4.4.1 Strengths and limitations

The strength of this study stems from the fact that the reported GCA cases are all histologically proven, with minimal risk of ascertainment bias, thus reflecting an accurate capture of the cases of CGA. When paired with the fact that we have the population at risk (that is the population GSH serves), it was possible to calculate an accurate incidence rate, which is new data. There are, however, some limitations to this study. Firstly, it is a retrospective study, with some missing data, also because the study was conducted in a referral centre, there might be a referral bias. This may not be representative of the disease in the community. Additionally, some of the GCA cases would have been managed in the private sector and will therefore not be accounted for, which might counterbalance the possibilities of referral bias.

4.5 CONCLUSIONS

This study showed that GCA is higher in males and those ≥ 60 years, as expected. Loss of weight and anaemia were key red flags and non-cardia GCAs were the most common in this cohort. The low prevalence of *H. pylori* seen may suggest previous eradication, although CAG was high, thus maintaining the increased risk of non-cardia GCA. The annual incidence of GCA in this study is 2.98/100 000, this figure is synonymous with what had been reported initially and tends to support the ‘African enigma’. Novel findings were that PPI use was significantly protective against developing antral cancer. This highlights the need to treat *H. pylori.* There is, however, a need for a robust prospective study to understand the so-called ‘African enigma’ including the possibilities of host protective molecules and cytokines, the strains of *H. pylori* in the African community. In addition, an analysis of the role of the gastric microbiome in modulating disease expression in the African community is key to unlocking the possibility of the existence of the ‘African enigma’.

4.6 REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2021;71(3):209-49.

2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018;68(6):394-424.

3. Suerbaum S, Michetti P. *Helicobacter pylori* infection. New England Journal of Medicine. 2002;347(15):1175-86.

4. Parkin DM. The global health burden of infection-associated cancers in the year 2002. International journal of cancer. 2006;118(12):3030-44.

5. Thrift AP, El-Serag HB. Burden of GCA. Clinical Gastroenterology and Hepatology. 2020;18(3):534-42.

6. Kidd M, Louw JA, Marks IN. *Helicobacter pylori* in Africa: observations on an 'enigma within an enigma'. Journal of gastroenterology and hepatology. 1999;14(9):851-8.

7. Holcombe C. *Helicobacter pylori*: the African enigma. Gut. 1992;33(4):429.

8. Ghoshal UC, Chaturvedi R, Correa P. The enigma of *Helicobacter pylor*i infection and GCA. Indian Journal of Gastroenterology. 2010;29:95-100.

9. McFarlane G, Forman D, Sitas F, Lachlan G. A minimum estimate for the incidence of GCA in Eastern Kenya. British Journal of Cancer. 2001;85(9):1322-5.

10. Asombang AW, Rahman R, Ibdah JA. GCA in Africa: current management and outcomes. World Journal of Gastroenterology: WJG. 2014;20(14):3875.

11. Galukande M, Luwaga A, Jombwe J, Fualal J, Kigula-Mugamba J, Kanyike A, et al. GCA diagnosis and treatment guidelines 2008: Uganda Cancer Working Group. East and Central African Journal of Surgery. 2008;13(2):142-9.

12. Colquhoun A, Arnold M, Ferlay J, Goodman K, Forman D, Soerjomataram I. Global patterns of cardia and non-cardia GCA incidence in 2012. Gut. 2015;64(12):1881-8.

13. Yao Q, Qi X, Xie S-H. Sex difference in the incidence of cardia and non-cardia GCA in the United States, 1992–2014. BMC gastroenterology. 2020;20:1-7.

14. Yang L, Kartsonaki C, Yao P, de Martel C, Plummer M, Chapman D, et al. The relative and attributable risks of cardia and non-cardia GCA associated with *Helicobacter pylori* infection in China: a case-cohort study. The Lancet Public Health. 2021;6(12):e888-e96.
15. González C, Megraud F, Buissonniere A, Barroso LL, Agudo A, Duell E, et al. *Helicobacter pylori* infection assessed by ELISA and by immunoblot and noncardia GCA risk in a prospective study: the Eurgast-EPIC project. Annals of oncology. 2012;23(5):1320-4.
16. Plummer M, Franceschi S, Vignat J, Forman D, de Martel C. Global burden of GCA attributable to *Helicobacter pylori*. International journal of cancer. 2015;136(2):487-90.

CHAPTER 5

CROSS-SECTIONAL STUDY: EPIDEMIOLOGY OF *HELICOBACTER PYLORI* AT GROOTE SCHUUR HOSPITAL

ABSTRACT

Background: The highest incidence of GCA is seen in East Asia, South America, and Middle America. Several factors are associated with the development of GCA. These factors are classified into non-modifiable factors including age, sex, race/ethnicity, and genetics, and modifiable factors which include *H. pylori* infection, the gastric microbiome, obesity, dietary habits, and lifestyle behaviours such as smoking and alcohol intake, radiation and chemical exposure, gastroesophageal reflux disease, gastric ulcers, and previous gastric surgery.

Aim: To compare patients positive for *H. pylori* to an *H. pylori* negative control group, where gastric biopsies were collected for molecular and microbiome studies (chapter 6 and 7). Here we compare their demographic data, clinical risk factors, and endoscopic findings. This was aimed at understanding the baseline risk factors for *H. pylori* pathogenesis and its impact in the development of GCA in our population.

Methods: For both cohorts of patients, (*H. pylori* positive and an *H. pylori* negative control group), demographic data, clinical risk factors, endoscopic and histologic pathology were collated and compared.

Results: 79 patients were recruited, 23 (29%) were *H. pylori* positive and 56 (71%) *H. pylori* negative. Sixteen (69.15%) and 41 (73.21%) of patients were less than 45 years of age in the *H. pylori* positive and *H. pylori* negative cohorts, respectively. There was no statistically significant difference between both cohorts for age, gender, smoking, alcohol use, PPI use, hypertension, or diabetes. We noted that anaemia was significantly associated with *H. pylori* positivity (4.5-fold). Our study showed that the rapid urease test (RUT) used in our facility is as sensitive as biopsy culture in diagnosing the presence of *H. pylori*.

Conclusions: Our study showed a strong association between anaemia and the occurrence of *H. pylori* infection. Therefore, screening and eradication of *H. pylori* should be fundamental in the management of unexplained anaemia.

5.1 BACKGROUND

According to the International Agency for Research on Cancer *H. pylori* is a type 1 carcinogen in the development of GCA [(1)]. This was a sequel to the discovery of *H. pylori* by Barry Marshal and Robin Warren in 1982 and the subsequent understanding of the role of *H. pylori* in the aetiopathogenesis of GCA among other gastrointestinal disease phenotypes. Studies have shown that the incidence of GCA is low in developed countries like North America 4.2/100 000 and Australia/New Zealand 4.5/100 000 [(2)], which is attributed to the lower incidence of *H. pylori*. On the contrary, East Asian countries have a significantly higher annual incidence of GCAs at 22.4/100 000 according to the IARC 2020 [(2)], with the prevalence of *H. pylori* seen at 40%. Therefore, eradicating *H. pylori* has become a cornerstone in reducing the incidence of GCA. A study done in China and published in Nature Medicine 2024 by Kai-Feng Pen et al revealed that in a total of 180 284 eligible participants from 980 villages who were enrolled and had over 11.8 years of follow-up, a total of 1 035 cases of incident GCAs were documented [(3)]. Individuals who received a 10-day course of *H. pylori* treatment showed a modest reduction in GCA incidence in intention-to-treat analyses (hazard ratio = 0.86, 95% confidence interval 0.74–0.99), with a more significant cancer reduction effect observed for those having successful *H. pylori* eradication (hazard ratio 0.81, 95% confidence interval 0.69–0.96) than for those who failed treatment [(3)]. In another study by Lingjun Yan et al, 1 630 patients were randomized to the *H. pylori triple* therapy eradication group (817) versus the placebo arm (813) with the primary endpoint being the incidence of GCA. After a follow-up period of 26.6 years, they noted that 21 participants (2.57%) in the treatment arm and 35 (4.31%) in the placebo arm were diagnosed with GCA [(4)]. There was a lower incidence of GCA among patients who received treatment (hazard ratio = 0.57; 95% CI, 0.33–0.98). A more significant risk reduction of 63% was observed among those without premalignant gastric lesions (hazard ratio = 0.37; 95% CI, 0.15–0.95) [(4)]. In Africa, available studies have noted a lower prevalence of GCA despite the high burden of *H. pylori*. In our recent retrospective study (unpublished) performed in a quaternary hospital GSH, Cape Town, a center with sophisticated facilities and easy access to endoscopy procedures, the yearly average incidence of GCA was noted to be 2.98/100 000, which is similar to previously reported data. From these findings, possibilities exist that the incidence of GCA might truly be low as earlier reported. Hence, understanding the genetic variability in the *H. pylori* strain among Africans and the gastric microbiome could explain the 'African enigma'.

5.1.1 Epidemiology and risk factors for GCA

90-95% of GCAs are adenocarcinomas and develop from the epithelium of the gastric mucosa [7,8]. Other types of GCAs include gastric lymphoma which arises from the immune system, sarcomas from the connective tissues, carcinoid tumours from the neuroendocrine cells and stromal tumours from the interstitial cells of Cajal [8,9].

Adenocarcinomas are further classified into two types based on Laureen's classification viz, intestinal type which developed following inflammation, gastritis/gastric atrophy, metaplasia, dysplasia, and GCA [7,9,10]. The development of this cancer type is associated with environmental factors and commonly metastasizes via the bloodstream and to the liver [11]. The patients are usually older than the diffuse type with a male preponderance. The diffuse type is more associated with genetic factors, patients are younger, generally less than 50 years of age, have late presentations and have unfavourable prognoses [7,9,10].

The highest incidence of GCA is seen in East Asia, South America, and Middle America [7,8]. Several factors are associated with the development of GCA. These factors are classified into non-modifiable factors including age, sex, race/ethnicity, and genetics, and modifiable factors which include *H. pylori* infection, the gastric microbiome, obesity, dietary habits, and lifestyle behaviours such as smoking and alcohol intake, radiation and chemical exposure, gastroesophageal reflux disease, gastric ulcers, and previous gastric surgery [12-18].

5.1.2 Age, ethnicity and sex

The incidence of GCA increases with age, peaking between the ages of 50 and 70 years of age [11,19]. According to the epidemiological profile of GCA in the USA, 1.8% of GCA occurs in those less than 34 years, 36% occurs in those between the ages of 35-64 years, and 59.6% occurs in the elderly who are more than 65 years [20]. There is male preponderance with a male-to-female ratio of 2:1 [20], with some studies suggesting a protective effect from the oestrogen [21]. There is a wide variability in the geographical distribution of GCA. Even in the same region, the distribution of GCA based on ethnicity varies [22]. Consequently, a study in the USA showed that GCA occurred less among Black Americans (1.8/100 000) compared to Whites with an incidence of 2.9/100 000. Also, in New Zealand, studies have shown a higher incidence of GCA among the non-Maoris with an incidence of 2.9/100 000 compared to the Maoris with an incidence of 1.5/100 000 [22].

5.1.3 Genetics and family history

The risk of GCA is 2 times higher among those with a family history of GCA [7,10].

5.1.4 Diet

Studies have also reported an increase in the occurrence of GCA in people ingesting diets rich in high salts, smoked fish, and picked food, with an increase in nitrates and nitrites. On the contrary, diets rich in fresh fruits and vegetables reduce GCA, suggesting the possible role of vitamins C, A, E, and beta-carotene in reducing GCA [7,8,19]. Increased salt intake is associated with mucosal injury, DNA destruction, and cell proliferation [23]. In a study done in China, it was shown that 5g/day of salt intake increases the risk of GCA by 12% [24].

5.1.5 Smoking and alcohol

Tobacco smoking is associated with an increased risk of GCA. It increases the risk of GCA in males by 50% and females by 20% [25,26,27]. Tobacco smoking causes chronic gastrointestinal inflammation and is characterized by immune dysfunction, and mucosal cell proliferation, thereby enhancing carcinogenesis [28,29]. Kaster et al showed that smoking is a dose-dependent risk factor for GCA, with a synergistic effect on the risk for GCA if higher exposure is combined with alcohol [26]. He showed that more than 20/day with alcohol of more than 5 occasions/14 days has a hazard ratio of 4.9 [26]. While smoking alone is associated with a hazard ratio of 2 [26]. Their study also revealed that the risk of GCA among cigarette smokers is associated with the combination of 2 or more tobaccos, and earlier age at initiation of daily smoking [26].

5.1.6 *H. pylori* infection and gastric dysbiosis

H. pylori is a key driver of gastric microbiome dysbiosis [30,31]. Its colonization of the stomach causes chronic inflammation, and it's associated with decreased diversity of the gastric microbiome, culminating in increased risk for more complicated gastric disorders [5]. Studies have shown that the severity *of H. pylori* infection, as reflected in the extent of gastritis, is proportional to the degree of microbiome dysbiosis [32]. Also, successfully eradicating *H. pylori* has been shown to partially or completely restore the microbiome toward a healthier state, although complete recovery is not guaranteed [31]. The specific changes in microbiome composition vary depending on factors like the strain of *H. pylori*, the host's genetic background, and the duration of infection.

Studies have shown that the abundance of certain bacterial genera, such as *Streptococcus* and *Lactobacillus*, is reduced in the presence of *H. pylori*, while other genera, such as *Campylobacte*r and *Helicobacter*, become more prevalent [5,32].

5.2 AIM

This study aimed to prospectively compare a cohort of patients positive for *H. pylori* and an *H. pylori-negative* control group, and compare demographic data, clinical risk factors, endoscopic and histologic pathology.

5.3 METHODS

5.3.1 Data and sample collection

Clinical data for both cohorts (patients referred for upper endoscopy) who were *H. pylori* positive or *H. pylori* negative were enrolled. The data collected included patients' history of smoking, alcohol use, diet, PPI use, demographics (age, sex), biodata including weight, height, endoscopic findings (gastritis, gastric ulcers, mass lesions/location, duodenitis, duodenal ulcer or any abnormal looking area) and histological findings (density of *H. pylori*, degree of inflammatory activity, atrophy, and metaplasia).

Gastric biopsies for RUT and histology (normal clinical care) and additional biopsies for microbiome analysis and *H. pylori* culture were collected. Three biopsies were taken from 3 different sites (antrum, corpus, and the fundus for each patient) and placed in one Sarstedt tube containing Prime store MTM media [12]. If lesions were observed, one biopsy was placed in a separate Sarstedt tube containing Prime store media. These specimens were transported to the molecular laboratory of the Division Medical Microbiology, University of Cape Town, where samples for the microbiome were stored at – 80 C in a designated freezer. In addition, biopsies for culture were placed into a sterilized tube with 1ml of sterile NaCl (0.9%) and processed as below by the Division Medical Microbiology, University of Cape Town. Biopsy samples for histology were stored in formalin solution and sent to the NHLS laboratory at GSH.

Inclusion criteria: All adults who are 18 years and above referred to the gastroenterology unit at Groote Schuur Hospital for gastroscopy for gastrointestinal symptoms.

5.3.2 Data analysis

All data exploration and analysis were in Stata (Version 13.1; Stata Corp, College Station, Texas, USA) ®. Descriptive statistics of mean, median, standard deviation, and interquartile ranges were used to characterize the sample. Frequencies and percentages were used for categorical variables. The means were compared using a two-sample t-test or its non-parametric equivalent where appropriate. For all variables, a test of significance was calculated and a $p = <0.05$ was considered statistically significant.

5.4 RESULTS

Seventy-nine (79) gastric biopsied specimens were collected, of which 56 (71%) samples were negative for *H. pylori* and served as controls. Twenty-three (29%) samples that were positive for *H. pylori* were used as test samples. The mean age of the population was 52 ± 17.25 with 57 (72 .15 %) of the population under study being at least 45 years of age. Females had a higher proportion at 46 (58.23%). PPI use within 2 weeks of the study was noted in 56 (70.89%) of the population. Moreover, 23 (29.11%) were positive for *H. pylori* with 7 (8.86%) reported with gastric ulcers on endoscopy. The distribution of the clinical profile, and endoscopic findings are shown in Table 5.1.

Table 5.1: The clinical profile of patients in the study (*H. pylori* negative and *H. pylori* positive patients)

Serial number	Clinical data	Absolute value	Percentages (%)
1	**Age in years**		
	Mean (52.2 ± 17.2)		
	<45	22	27.8
	≥45	57	72.1
2	**Gender**		
	Female	46	58.2
	Male	33	41.8
3	**Race**		
	Black	14	17.7
	Coloured	62	78.5
	White	2	2.5
	Other	1	1.3
4	**Alcohol use**		
	No	64	81.0
	Yes	15	19.0
5	**PPI use**		
	Yes	56	70.9
	No	21	26.6
	NR	2	2.5
6	**Smoking**		
	Yes	37	46.8
	No	42	53.2
7	**NSAID use**		
	Yes	27	34.2
	No	52	65.8
8	**Hypertension**		

	Yes	31	39.2
	No	48	60.8
9	**Diabetes**		
	Yes	15	19
	No	64	81
10	**Gastric ulcer**		
	Yes	7	8.8
	No	44	55.7
	NR	28	35.4

*NR = not recorded

Table 5.2: Comparison of the clinical data between the *H. pylori* positive and the *H. pylori* negative cohort

Serial number	Variables	Hp positive (n=23)	Hp negative (n=56)	p-value
1	Age < 45			
	Yes	16 (69.6%)	41(73.2%)	0.7424
	No	7 (30.4%)	15 (26.8%)	
2	Anaemia (corrected for sex)			
	Yes			
	No	16 (69.6%)	22 (39.3%	0.0078
		4 (17.4%)	27(48.2%)	
3	Gender			
	Male	8 (34.8%)	25 (44.64%)	0.4195
	Female	15 (65.2%)	31(55.36%)	
4	Vegetarian			
	Yes	0	6 (10.7%)	
	No	23 (100%)	49 (87.5%)	Unable to compute
5	NSAID use			
	Yes	9 (39.1%)	18 (32.1%)	0.5519
	No	14 (60.9%)	38 (67.9%)	
6	Smoking			
	Yes	10 (43.5%)	27 (48.2%)	0.7015
	No	13 (56.5%)	29 (51.8%)	
7	Alcohol use			
	Yes	6 (26.1%)	9 (16.1%)	0.3025
	No	17 (73.9%)	47 (83.9%)	
8	PPI use			
	Yes	18 (78.3%)	38 (67.9%)	0.4767
	No	5 (21.7%)	16 (28.6%)	

9	Hypertension			
	Yes	10 (43.5%)	21(37.5%)	0.6211
	No	13 (56.5%)	35 (62.5%)	
10	Diabetes			
	Yes	5 (21.7%)	10 (17.9%)	0.6894
	No	18 (78.3%)	46 (82.1%)	
11	Gastric ulcers			Unable to compute
	Yes	0	5 (8.93%)	
	No	13 (56.52%)	31(55.36%)	
12	RUT/Biopsy culture			
	Yes	11(47.8%)	1 (1.8%)	0.0000
	No	12 (52.2%)	55 (98.2%)	

*NSAID-nonsteroidal anti-inflammatory drug, PPI- proton pump inhibitor, RUT- rapid urease test, Hp- *Helicobacter pylori*

Among the *H. pylori* positive cohort, 16 (69.15%) were less than 45 years of age. In the cohort of the *H. pylori* negative 41 (73.21%) were less than 45 years of age. Additionally, females were 31 (55.36%) and 15 (65.22%) for *H. pylori* negative and *H. pylori* positive respectively. There was no statistical significance between both cohorts for age, gender, alcohol use, PPI use, hypertension, or diabetes. We were, however, unable to determine the presence of any statistical difference for both cohorts for gastric ulcers, gastro-duodenitis and vegetarian diets. (Table 5.2).

We noted that anaemia was significantly associated with patients *H. pylori* positivity with a p-value of 0.0078. Our study showed that the RUT used in our facility was as sensitive as biopsy culture in diagnosing the presence of *H. pylori*, p-value (0.0000).

5.5 DISCUSSION

Reports suggest a correlation between *H. pylori* and other systemic and localized diseases outside of the stomach, including cardiovascular diseases such as hypertension, ischaemic heart disease, metabolic syndrome, metabolic dysfunction associated with steatotic liver disease (MASLD), hepatobiliary disease, and neurogenerative disease. The pathophysiology attributed to these

disorders is due to chronic inflammation, endothelial injury with arterial narrowing, and immune dysregulation [33-35]. However, the data supporting these findings is not robust, with others showing either inconclusive or conflicting results [33].

In our study, we noted no statistically significant differences for age, gender, alcohol use, PPI use, hypertension, or diabetes between *H. pylori* positive and *H. pylori* negative cohorts. It has been shown that male gender is associated with a higher risk for GCA. However, the exact mechanism by which the male gender contribute to the development of GCA is not fully understood. Studies that looked at the occurrence of *H. pylori* on the bases of gender have shown conflicting results. In a meta-analysis of 244 studies by Abrahim et al. noted that the occurrence of *H. pylori* was higher among the males compared to the females [36]. On the contrary, Y Qiao et al. reported females have a higher incidence of *H. pylori* than males [37]. These findings were corroborated by Zhao et al [38]. Yet, some studies find no significant difference between genders [39], just as we noted in our study.

Similarly, diabetes and hypertension have been reported according to some studies to be associated with *H. pylori* infection. In a review by S Smith et al. they reported that *H. pylori* infection is more common among patients with type 2 diabetes [40]. They attributed this to the increased production of the proinflammatory cytokines and differential hormonal imbalance imposed by the bacterium. Mansori et al. in 2020 however reported an inconclusive correlation between *H. pylori* and diabetes [41]. Similarly, while some studies had reported up to 13.4% increased risk in the incidence of hypertension among patients infected with *H. pylori* [42], many studies showed mixed results with some showing no association between *H. pylori* infection and the occurrence of hypertension [43-45]. Our study is unable to establish any correlation between *H. pylori* infection and hypertension or diabetes likely because of the many cofounding factors such as lifestyle and age among many others.

An important finding in our study is the odds of anaemia among patients with *H. pylori* infection with a p-value of 0.0078. Previous studies have reported a strong association between *H. pylori* infection and the increased risk of anaemia [46-48]. The pathophysiology of this is multifactorial. *H. pylori* cause chronic gastritis leading to achlorhydria. Also, the bacterium competes with the body for iron as it requires iron for growth [46]. There is also evidence suggesting the ability of the bacteria to trigger antibody production against intrinsic factors thereby leading to vitamin B12 deficiency. Other vitamin deficiencies such as vitamin C, A, B6, and folic acid are also associated

with *H. pylori* infection [49]. This will also imply the importance of screening for *H. pylori* as part of a work-up for unexplained anaemia.

Another important finding is the sensitivity of rapid urease test in the diagnosis of *H. pylori* which is comparable to the biopsy and culture with a p-value of 0.0000. Studies have reported sensitivity of 80-99% and specificity of 92-100% [50]. Factors that are capable of reducing the sensitivity of RUT include the bacterial load, bleeding, the use of PPI and site of sample collection [51,52]. The fact that this diagnostic technique has a rapid turnover time makes it ideal in the management of *H. pylori* in clinical settings.

5.5.1 Strengths and weaknesses

The strength of this study stems from the fact that it is a prospective study. It was aimed that the data gleaned may be hypothesis-generating and inform future studies which may help unlock the factors responsible for the 'African enigma'. The sample size, however, may be a weakness. We were also unable to control for confounding factors such as diet, drugs, the use of PPI etc. prior to enrolment, however these factors were not different in the cases and controls.

5.6 CONCLUSIONS

Although *H. pylori* is a type 1 carcinogen for GCA, with the male gender, advance age, smoking having been established among others as independent risk factors in the development of GCA, we found no association between the occurrence of these factors and the risk of *H. pylori* infection. Importantly we noted that there appear to be an association between anemia and the occurrence of *H. pylori* infection. Therefore, screening and eradication of *H. pylori* should be considered in the management of anaemia.

We planned this is a pilot study, however, with small data size and certain confounding factors such as diet and socioeconomic status, other comorbidities are not accounted for. Therefore, a more robust prospective study will be needed to ascertain these findings.

5.7 REFERENCES

1. Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, et al. A review of human carcinogens—Part B: biological agents. The lancet oncology. 2009;10(4):321-2.

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2021;71(3):209-49.

3. Pan K-F, Li W-Q, Zhang L, Liu W-D, Ma J-L, Zhang Y, et al. GCA prevention by community eradication of *Helicobacter pylori*: a cluster-randomized controlled trial. Nature Medicine. 2024:1-11.

4. Yan L, Chen Y, Chen F, Tao T, Hu Z, Wang J, et al. Effect of *Helicobacter pylori* eradication on GCA prevention: updated report from a randomized controlled trial with 26.5 years of follow-up. Gastroenterology. 2022;163(1):154-62. e3.

5. Yalda, Khosravi. "The Gastric Microbiome of Four Malaysian Gastroduodenal Disease Patients." None (2017). 10.36959/998/781.

6. Precious Mathebela, B.Damane, T. Mulaudzi, Zilungile Lynette Mkhize-Khwitshana, G. Gaudji, Zodwa Dlamini. "Influence of the Microbiome Metagenomics and Epigenomics on GCA." International Journal of Molecular Sciences (2022). 10.3390/ijms232213750.

7. J.R. Kelley, J.M. Duggan, GCA epidemiology and risk factors, J. Clin.Epidemiol. 56 (2003) 1–9.

8. B.J. Dicken, D.L. Bigam, C. Cass, J. Mackey, A.A. Joy, S.M. Hamilton, Gastric adenocarcinoma: review and considerations for future directions, Ann. Surg.241 (2005) 27–39.

9. GCA: introduction. [Internet] available from:http://ww.hopkinsmedicine.org/gastroenterology_hepatology/_pdfs/esophagus_stomach/gastric_cancer.pdf.

10. P. Correa, Pathology and molecular pathogenesis of GCA [Internet] Available from: http://www.uptodate.com/contents/pathology-andmolecular-pathogenesis-of-gastric-cancer.

11. M. Malinowska, A. Nasierowska-Guttmejer, Epidemiologia i patogeneza raka_zoła˛dka, Pol. J. Pathol. 64 (2013) 17–26.

12. Rawla P, Barsouk A. Epidemiology of GCA: global trends, risk factors and prevention. *Prz Gastroenterol* 2019;14: 26-38 [PMID: 30944675 DOI: 10.5114/pg.2018.80001]

13. Sun DQ, Yang F, Li H, Cao MM, Yan XX, He SY, Zhang SL, Xia CF, Chen WQ. [Regional disparities in trends of global GCA incidence and mortality from 1990 to 2019]. *Zhonghua Zhong Liu Za Zhi* 2022; 44: 950-954 [PMID: 36164696 DOI: 10.3760/cma.j.cn112152-20220120-00049]
14. National Health Commission of the People's Republic of China. National guidelines for diagnosis and treatment of GCA 2022 in China. *Chin J Cancer Res* 2022; 34: 207-237 [DOI: 10.21147/j.issn.1000-9604.2022.03.04]
15. Clinton SK, Giovannucci EL, Hursting SD. The World Cancer Research Fund/American Institute for Cancer Research Third Expert Report on Diet, Nutrition, Physical Activity, and Cancer: Impact and Future Directions. *J Nutr* 2020; 150: 663-671 [PMID: 31758189 DOI: 10.1093/jn/nxz268]
16. Ladeiras-Lopes R, Pereira AK, Nogueira A, Pinheiro-Torres T, Pinto I, Santos-Pereira R, Lunet N. Smoking and GCA: systematic review and meta-analysis of cohort studies. *Cancer Causes Control* 2008; 19: 689-701 [PMID: 18293090 DOI: 10.1007/s10552-008-9132-y]
17. Fortunato L, Rushton L. Stomach cancer and occupational exposure to asbestos: a meta-analysis of occupational cohort studies. *Br J Cancer* 2015; 112: 1805-1815 [PMID: 25928706 DOI: 10.1038/bjc.2014.599]
18. Singh S, Jha HC. Status of Epstein-Barr Virus Coinfection with *Helicobacter pylori* in GCA. *J Oncol* 2017; 2017: 3456264 [PMID: 28421114 DOI: 10.1155/2017/3456264
19. J., Ferlay, I., Soerjomataram, M., Ervik, R., Dikshit, S., Eser, C. Mathers, et al., Cancer incidence and mortality worldwide: IARC CancerBase No 11. International Agency for Research on Cancer [Internet]. Available from: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx. (2013).
20. Surveillance E, and End Results Program. Cancer Stat Facts: Stomach Cancer. [cited 26 December 2022]. In: National Cancer Institute. Available from: https://seer.cancer.gov/statfacts/html/stomach.html
21. Wesołowska M, Pawlik P, Jagodziński PP. The clinicopathologic significance of estrogen receptors in human gastriccarcinoma. *Biomed Pharmacother* 2016; 83: 314-322 [PMID: 27399808 DOI: 10.1016/j.biopha.2016.06.048]
22. Parkin DM, Whelan SL, Ferlay J, et al. Cancer incidence in five continents, vol VII. Lyon: International Agency for Research on Cancer; 1997. p. 822–3.

23. Bouras E, Tsilidis KK, Triggi M, Siargkas A, Chourdakis M, Haidich AB. Diet and Risk of GCA: An Umbrella Review. *Nutrients* 2022; 14 [PMID: 35565732 DOI: 10.3390/nu14091764]
24. Fang X, Wei J, He X, An P, Wang H, Jiang L, Shao D, Liang H, Li Y, Wang F, Min J. Landscape of dietary factors associated with risk of GCA: A systematic review and dose-response meta-analysis of prospective cohort studies.*Eur J Cancer* 2015; 51: 2820-2832 [PMID: 26589974 DOI: 10.1016/j.ejca.2015.09.010]
25. Ladeiras-Lopes R, Pereira AK, Nogueira A, Pinheiro-Torres T, Pinto I, Santos-Pereira R, Lunet N. Smoking and GCA: systematic review and meta-analysis of cohort studies. *Cancer Causes Control* 2008; 19: 689-701 [PMID:18293090 DOI: 10.1007/s10552-008-9132-y]
26. Sjödahl K, Lu Y, Nilsen TI, Ye W, Hveem K, Vatten L, Lagergren J. Smoking and alcohol drinking in relation to risk of GCA: a population-based, prospective cohort study. *Int J Cancer* 2007; 120: 128-132 [PMID: 17036324 DOI:10.1002/ijc.22157]
27. Nomura AM, Wilkens LR, Henderson BE, Epplein M, Kolonel LN. The association of cigarette smoking with GCA: the multiethnic cohort study. *Cancer Causes Control* 2012; 23: 51-58 [PMID: 22037905 DOI:10.1007/s10552-011-9854-0
28. Li LF, Chan RL, Lu L, Shen J, Zhang L, Wu WK, Wang L, Hu T, Li MX, Cho CH. Cigarette smoking and gastrointestinal diseases: the causal relationship and underlying molecular mechanisms (review). *Int J Mol Med* 2014; 34: 372-380 [PMID: 24859303 DOI: 10.3892/ijmm.2014.1786]
29. Rota M, Pelucchi C, Bertuccio P, Matsuo K, Zhang ZF, Ito H, Hu J, Johnson KC, Palli D, Ferraroni M, Yu GP, Muscat J,Lunet N, Peleteiro B, Ye W, Song H, Zaridze D, Maximovitch D, Guevara M, Fernández-Villa T, Vioque J, Navarrete-Muñoz EM, Wolk A, Orsini N, Bellavia A, Håkansson N, Mu L, Persiani R, Kurtz RC, Lagiou A, Lagiou P, Galeone C,Bonzi R, Boffetta P, Boccia S, Negri E, La Vecchia C. Alcohol consumption and GCA risk-A pooled analysis within the StoP project consortium. *Int J Cancer* 2017; 141: 1950-1962 [PMID: 28718913 DOI: 10.1002/ijc.30891]
30. Kountouras J, Papaefthymiou A, Polyzos SA, Kazakos E, Tzilves D, Lazaraki G, Zavos C, Vardaka E, Arapoglou S, Kyrailidi F, Mouratidou MC. Potential impact of Helicobacter pylori infection on hepatic steatosis and metabolic risk factors among patients with chronic hepatitis B. Journal of viral hepatitis. 2023 Aug;30(8):710-1.

31. Chunfeng Zhang, Chang-hui Geng, Zhihong Yang, Yan-Lin Li, Li-Quan Tong, P. Gao, Yue-Qiu Gao. "Changes in patients" symptoms and gastric emptying after *Helicobacter pylori* treatment." World Journal of Gastroenterology (2016). 10.3748/wjg.v22.i18.4585.

32. H. Erdal, Y. S. Sakin, M. etin, B. elik, Attila nmez, A. Uygun, M. Glen. "The Relationship Between Hyperplastic Gastric Polyps and *Helicobacter pylori."* Konuralp Tip Dergisi (2021). 10.18521/KTD.834666.

33. de Korwin JD, Ianiro G, Gibiino G, Gasbarrini A. *Helicobacter pylori* infection and extragastric diseases in 2017. Helicobacter. 2017 Sep;22:e12411.

34. Zuin M, Rigatelli G, Del Favero G, et al. coronary artery disease and *Helicobacter pylori* infection: should we consider eradication therapy as cardiovascular prevention strategy? Int J Cardiol. 2016;223: 711-712

35. Pellicano R, Ianiro G, Fagoonee S, Settanni CR, Gasbarrini A. Extragastric diseases and *Helicobacter pylori*. Helicobacter. 2020 Sep;25:e12741.

36. Ibrahim A, Morais S, Ferro A, Lunet N, Peleteiro B. Sex-differences in the prevalence of *Helicobacter pylori* infection in pediatric and adult populations: systematic review and meta-analysis of 244 studies. Digestive and Liver Disease. 2017 Jul 1;49(7):742-9.

37. Qiao Y, Zhou Y, Zhao L, Yang S, Zhang X, Liu S. Sex differences in *Helicobacter pylori* infection and recurrence rate among 81,754 Chinese adults: a cross-sectional study. BMC gastroenterology. 2024 Sep 11;24(1):305.

38. Zhao XX, Liu MH, Wang RL, Tian T. Effect of gender and age on the correlation between *Helicobacter pylori* and colorectal adenomatous polyps in a Chinese urban population: a single center study. Gastroenterology Research and Practice. 2020;2020(1):8596038.

39. Ibrahim A, Morais S, Ferro A, Lunet N, Peleteiro B. Sex-differences in the prevalence of *Helicobacter pylori* infection in pediatric and adult populations: systematic review and meta-analysis of 244 studies. Digestive and Liver Disease. 2017 Jul 1;49(7):742-9.

40. Smith S, Fowora M, Pellicano R. Infections with *Helicobacter pylori* and challenges encountered in Africa. World journal of gastroenterology. 2019 Jul 7;25(25):3183.

41. Mansori K, Moradi Y, Naderpour S, Rashti R, Moghaddam AB, Saed L, Mohammadi H. *Helicobacter pylori* infection as a risk factor for diabetes: a meta-analysis of case-control studies. BMC gastroenterology. 2020 Dec;20:1-4.

42. Fang Y, Xie H, Fan C. Association of hypertension with *helicobacter pylori*: A systematic review and metaanalysis. PLoS One. 2022 May 19;17(5):e0268686.
43. Zuin M, Rigatelli G, Del Favero G, et al. coronary artery disease and *Helicobacter pylori* infection: should we consider eradication therapy as cardiovascular prevention strategy? Int J Cardiol. 2016;223: 711-712
44. Wawro N, Amann U, Butt J, et al. *Helicobacter pylori* seropositivity: prevalence, associations, and the impact on incident metabolic diseases/risk factors in the population-based KORA Study. Front Public Health. 2019;7:96.
45. Pellicano R, Ianiro G, Fagoonee S, Settanni CR, Gasbarrini A. Extragastric diseases and *Helicobacter pylori*. Helicobacter. 2020 Sep;25:e12741.
46. Bini EJ. *Helicobacter pylori* and iron deficiency anemia: guilty as charged?. The American journal of medicine. 2001 Oct 15;111(6):495-7.
47. Monzón H, Forné M, Esteve M, Rosinach M, Loras C, Espinos JC, Viver JM, Salas A, Fernández-Bañares F. *Helicobacter pylori* infection as a cause of iron deficiency anaemia of unknown origin. World Journal of Gastroenterology: WJG. 2013 Jul 7;19(26):4166.
48. Haile K, Yemane T, Tesfaye G, Wolde D, Timerga A, Haile A. Anemia and its association with *Helicobacter pylori* infection among adult dyspeptic patients attending Wachemo University Nigist Eleni Mohammad Memorial Referral Hospital, Southwest Ethiopia: A cross-sectional study. PLoS One. 2021 Jan 14;16(1):e0245168.
49. Franceschi F, Annalisa T, Teresa DR, Giovanna D, Ianiro G, Franco S, Viviana G, Valentina T, Riccardo LL, Antonio G. Role of *Helicobacter pylori* infection on nutrition and metabolism. World J Gastroenterol. 2014 Sep 28;20(36):12809-17. doi: 10.3748/wjg.v20.i36.12809. PMID: 25278679; PMCID: PMC4177464.
50.Lim LL, Ho KY, Ho B, Salto-Tellez M. Effect of biopsies on sensitivity and specificity of ultra-rapid urease test for detection of *Helicobacter pylori* infection: a prospective evaluation. World Journal of Gastroenterology: WJG. 2004 Jul 7;10(13):1907.
51. Wong BC, Wong WM, Wang WH, Tang VS, Young J, Lai KC, Yuen ST, Leung SY, Hu WH, Chan CK, Hui WM. An evaluation of invasive and non-invasive tests for the diagnosis of *Helicobacter pylori* infection in Chinese. Alimentary Pharmacology & Therapeutics. 2001 Apr;15(4):505-11.

52. Chen YK, Godil A, Wat PJ. Comparison of two rapid urease tests for detection of *Helicobacter pylori* infection. Digestive Diseases and Sciences. 1998 Aug;43:1636-40

CHAPTER 6

THE MOLECULAR SIGNATURE OF *HELICOBACTER PYLORI* AT GROOTE SCHUUR HOSPITAL

ABSTRACT

Background: *Helicobacter pylori (H. pylori)* is a type 1 carcinogen and widely occurs in Africa, yet the incidence of GCA in Africa is among the lowest globally, a phenomenon generally referred to as the 'African enigma'. In Southeast Asia, countries like Japan and South Korea report relatively lower *H. pylori* prevalence than many African nations. Despite this, Southeast Asia experiences some of the highest GCA incidence rates globally. Several factors may contribute to these regional disparities including *H. pylori* strain variations.

Methods: We investigated the genetic composition of the African sub-type of *H. pylori* (n=23) to determine the degree of virulence, the potential origin, or proximity to other countries with either similar or varying GCA distribution in a bid to interrogate the so-called 'African enigma' using *H. pylori* DNA whole genome sequencing (WGS). We also included antimicrobial resistance (AMR) gene analysis.

Results: The occurrence of v*irB4/D* which is a type IV secreting system is said to potentiate the activities of *cagA* and *vacA*, which is low in the African cohort. Similarly, the occurrence of the CagPAI morphologic variants which are associated with the development of GCA is lower in the African cohort compared to the Southeast Asian populations. Also, using phylogenetic tree analysis, the *H. pylori* strains isolated from the South African cohort are most similar to American strains and less related to the Southeast Asian type**.** Unique sequence types for the African cohort were detected.

Conclusions: This study shows molecular evidence for the low pathogenicity of *H. pylori* in Africa; the African *H. pylori* strains appear to be less virulent compared to the Southeast Asian type. This data, therefore, suggests a biological argument for the existence of the 'African enigma'.

6.1 BACKGROUND

The World Health Organization (WHO)/International Agency for Research Cancer (IARC) 1994 declared *H. pylori* a type 1 carcinogen [(1)]. Anecdotally, the prevalence of GCA in the African sub-

continent should be among the highest, considering the high prevalence of *H. pylori* in Africa. However, the prevalence of GCA in Africa is among the lowest globally, leading to the conundrum of the 'African enigma' [2-4]. There is also a wide regional and global variation in the pattern of GCA distribution. It is therefore imperative that we investigate the molecular composition of the African sub-type of *H. pylori* to determine the degree of virulence, the potential origin, or proximity to other countries with either similar or varying GCA distribution in a bid to interrogate the so-called 'African enigma'. Other factors such as genetic susceptibility and environmental and lifestyle factors also need to be considered but, in this study, we focused on differences based on whole genome sequencing (WGS).

6.1.1 The role of *H. pylori* virulent genes in the pathogenesis of GCA

H. pylori virulent factors in association with environmental factors have been implicated in the pathogenesis of gastrointestinal disorders such as GCA, gastric ulcers, duodenal ulcers, and gastritis. The *Cytotoxic-associated gene A* (*cagA*)*, Vacuolating cytotoxin gene A* (*vacA*)*, Duodenal ulcer-promoting gene A* (*dupA*)*, Induced by contact with epithelium A* (*IceA*)*, Outer inflammatory protein)* (*oipA*)*,* and *Blood group antigen binding adhesion gene* (*babA*) are commonly studied virulence factors implicated in gastrointestinal disorders [5].

The *cagA* is implicated in the development of GCAs. It is located in the *cag* Pathogenicity Island (*cagPAI*), which is a protein arising from the insertion of DNA into the gene that encodes *cagA* and other T4SS. *cagA* is transported into the host cell via the T4SS, which is then phosphorylated and thereafter, initiates the oncogenic process [5].

Another crucial factor to consider is *cagPAI* Polymorphism. It has been shown that the presence of c*agA* does not invariably predict the occurrences of GCA, more so, several individuals are being infected with CagA-positive *H. pylori*, without the development of GCA. However, studies have shown that the micro-variants polymorphism in the *cagPAI* seems to influence the development of GCA. Federico et al reported that polymorphism in the *cagA, cagL*, and *cagI* were associated with the development of GCA [6]. There are, however, several other micro-variant polymorphs whose roles remain to be elucidated.

vacA gene is another virulence factor that has been widely studied. It is said to activate cell proliferative pathways and, hence promote cancer development pathways. It also causes vacuolation with subsequent release of the cytochrome C leading to the activation of the

proinflammatory pathway. There are multiple variants of the *vacA* viz v*acA* s1, s2, m1, and m2. These strains can occur in any combination, however, strains with either VacA s1 or m1 are more virulent [5].

Notably, the *virB* genes are cluster genes that form one of the types of Type 4 Secretion System (T4SS) in the *cagPAI* and are homologous to the duodenal ulcer pathogenic gene (*dupA* genes). Studies have also shown that the *dupA* gene in association with the *virB* homologous cluster genes is associated with the development of duodenal ulcers rather than the existence of *dupA* gene alone [7].

6.1.2 Eradication of *H. pylori* in the prevention of gastrointestinal disorders including GCA

To prevent the development of GCA and other gastrointestinal disorders *H. pylori* eradication is paramount. The treatment regime for the eradication of *H. pylori*, either as first-line treatment, second-line, or salvage therapies required a combination of more than one antibiotic with PPI and bismuth. The commonly recommended antibiotics in various combinations include clarithromycin, metronidazole, amoxicillin, levofloxacin, tetracycline, and rifabutin [8].

In recent years, the use of multiple antibiotics in the eradication of *H. pylori* has been marred by multiple antimicrobial resistances. The Maastricht VI guideline on the management of *H. pylori* recommends against the use of clarithromycin as first line if the resistance rate is unknown or more than 15% [8]. According to Argueta et al, metronidazole has the highest resistance rate globally, with a concomitant rise in the resistance rate for other antimicrobial agents (Table 6.1) [9]. Therefore, understanding the regional antimicrobial resistance profile is key to the eradication of *H. pylori*.

Although the advent of WGS has enabled us to identify a plethora of *H. pylori* resistance genes, there appears not to be an established direct correlation between the presence of resistance genes and antimicrobial susceptibility profiles. Understanding expression and the regional distribution of resistance genes should be investigated further. The AMR regional profile, the choice of empiric antibiotics, and thus eradication rates of *H. pylori*, may better inform the regional distribution of *H. pylori*-associated gastrointestinal disorders, including GCAs.

In this sub-study, we therefore focused on studying the genetic composition of *H. pylori*, including the virulence factors, mutations, and antibiotic resistance pattern utilizing whole-genome sequencing (WGS), and comparing molecular results to countries with higher prevalence of GCA

(China, South Korea, Japan, Mongolia), and with countries having similar prevalence of GCA to the African population, the United States of America and the United Kingdom.

6.2 METHODS

Patients who presented for routine clinic consultation at the GSH gastroenterology unit with the indications for gastroscopy and who had willingly consented to the study were included. Seventy-nine (79) gastric biopsies were collected, of which 23 samples were *H. pylori positive,* and therefore selected to perform WGS. The detection of *H. pylori* in these samples was confirmed by the RUT at GSH and thereafter, three gastric biopsies were placed in saline and transported on ice immediately to Medical Microbiology, University of Cape Town for further testing.

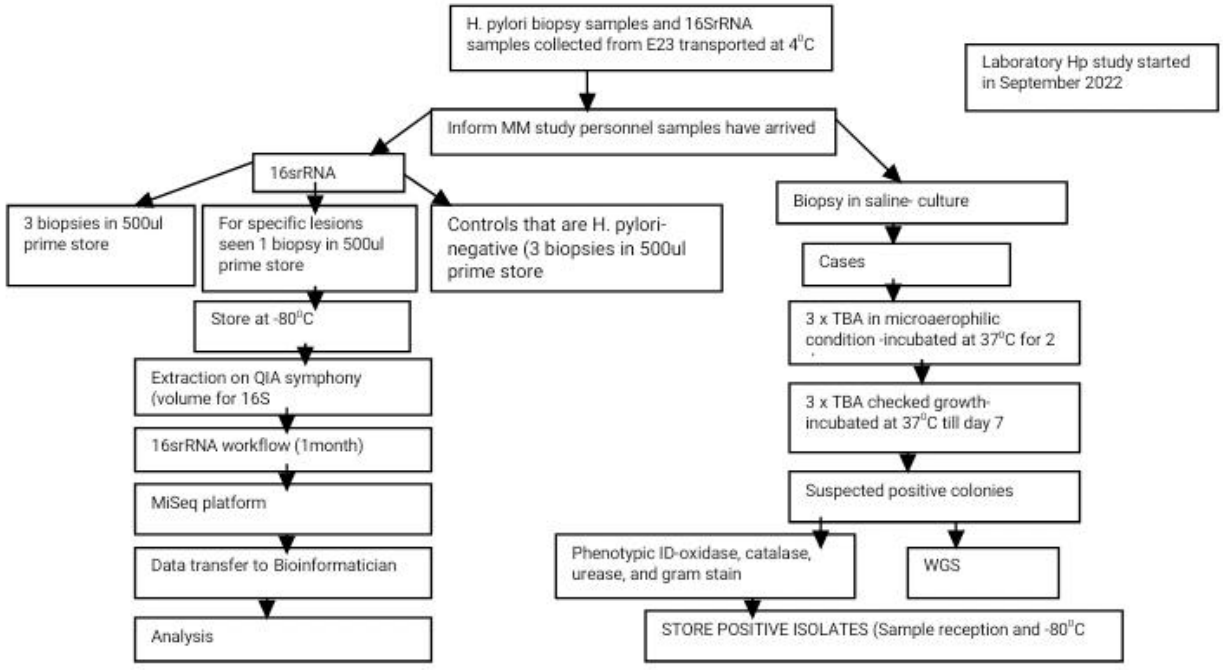


Figure 6.1: *H. pylori* study flow chart

6.2.1 *H. pylori* culture

H. pylori was isolated through a culture of gastric biopsy specimens sampled from the gastric antrum, corpus, and fundus of patients to maximize yield. These gastric biopsies were cultured using three Tryptose Blood Agar (TBA) plates under micro-aerophilic (10% CO2, 5% O2, and 85% N2) conditions typically at 35-37°C for 48 hours to 7 days. If suspected colonies were observed, a gram stain was performed for preliminary identification as well as a biochemical urease test and re-cultured on a fresh TBA plate to obtain pure colonies. These 23 positive cultured samples underwent a manual DNA extraction method and were subsequently used for WGS. Figure 6.1 shows the flow diagram.

The *H. pylori*-like colonies with translucent, convex morphology which grew on the plates, were identified based on biochemical properties (catalase, oxidase, and urease reactions) and microscopic morphology following a Gram staining (Gram-negative bacilli). They were then sub-cultured for 48 hours. Results were recorded on a data form (supplementary Figure S6.1). *H. pylori* DNA was extracted from cultured positive samples for whole genome sequencing (WGS) and then stored at −80 °C in a Brucella broth medium containing glycerol.

6.2.2 DNA extraction and quality assessment

Genomic DNA was extracted from cultured *H. pylori* isolates using the DNA Fungal/Bacterial Miniprep Kit, Zymo D6005, USA, following the manufacturer's protocol. The concentration and purity of the extracted DNA were assessed using a Nanodrop spectrophotometer (Thermo Fisher Scientific), and the integrity was evaluated by agarose gel electrophoresis. The DNA samples exhibited adequate A260/A280 ratios of 1.8 and 2.0. Samples with an A260/A230 ratio below the expected range of 2.0–2.2, indicating possible contaminants, were subjected to an additional cleanup step using Agencourt AMPure XP beads. The bead clean-up ratio of 2.5X ratio of beads to sample was included. The cleaned DNA extracts with intact high molecular weight bands were used for subsequent library preparation.

6.2.3 Library preparation

Whole genome sequencing libraries were prepared using the DNA Prep Kit (Illumina) [14] according to the manufacturer's instructions. The initial DNA was diluted to the desired input

concentration (note that while the acceptable range is 1–500ng, input that is less than 100ng will not be normalized by the bead-linked transposome). Briefly, 30 ng of genomic DNA was enzymatically fragmented using Bead-Linked Transposomes (BLT). During this process, the DNA was simultaneously fragmented and tagged with partial adapter sequences.

Following purification, the fragmented product was then subjected to a PCR where unique barcodes were assigned to each library to allow for downstream multiplexing (also known as indexing PCR) and to complete the adapters, to include the Illumina P5 and P7 flow cell attachment sequences. The PCR conditions were as follows: Preheated lid set to 100°C, 68°C for 3 minutes, 98°C for 3 minutes 8 cycles of: 98°C for 45 seconds, 62°C for 30 seconds, 68°C for 2 minutes, 68°C for 1 minute and the hold at 10°C. The amplified libraries were purified and quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific).

6.2.4 Library quality control

The quality and size distribution of the prepared libraries were evaluated using an Agilent D2100 Bioanalyzer, Tape Station (Agilent Technologies) with the High Sensitivity DNA Kit. The libraries had a peak size distribution of approximately 300-600 bp. The pooled library was determined using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific).

6.2.5 Sequencing

The prepared libraries were sequenced on the Illumina MiSeq platform using the MiSeq Reagent Kit v2 (300 cycles). The cartridge was removed from the -20°C freezer and allowed to thaw overnight in the fridge. Libraries were pooled in equimolar concentrations. The pooled libraries were diluted to 4 nM before denaturation and diluted to 20 pM. The PhiX (stock concentration of 10 nM) also underwent the same denaturation and dilution and was ultimately spiked into the library at the same final loading concentration as the library (3% PhiX spike-in). PhiX served as a sequencing control and to increase sequencing diversity for low-diversity library preparations. The denatured libraries (with Phix) were diluted to a final loading concentration of 10.5 pM and loaded onto the MiSeq flow cell. Sequencing was performed with a paired-end 150 bp read configuration.

6.2.6 Bioinformatics data analysis

Raw sequencing data were processed using the MiSeq Reporter software (Illumina) for initial base calling and demultiplexing. The quality of the sequencing data was checked by FastQC v0.12.1 [11] and MultiQC v1.17 [12], and adapter trimming and low-quality trimming were done by Trimmomatic v0.39 [13] (parameters: ILLUMINACLIP: Truseq3.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:30 MINLEN:36) to remove all adapters, while allowing a maximum of 2 mismatches, a palindrome clip threshold of 30 and a simple clip of 10, leading and trailing bases with quality below 3, using a sliding window of size 4 and trimming when the average quality within the window fell below 30, and finally discarding all reads that were shorter than 36 bases after trimming. The trimmed reads were assembled by SpaDes v3.15.5 [18], and the resulting assemblies were assessed by QUAST v4.5 [14,15]. DDBJ Fast Annotation and Submission Tool (DFAST) [16] was used to analyze and predict the features of the bacterial genome sequences. Antimicrobial Resistance (AMR) genes were determined using the Resistance Gene Identifier (RGI) v6.0.3 [17] (parameters: --clean --include_loose --local) to clean input sequences before analysis and include the less confident (loose) matches in the output using a local Comprehensive Resistance Database (CARD) v 3.2.9 [18]. This RGI version uses BLAST v2.14.0. Multi-locus sequence typing (MLST) was done by mlst v2.23.0 [19, 20].

6.2.7 Comparison WGS *H. pylori* (Genbank)

We extracted 224 of the available *H. pylori* whole genome sequences from the GenBank database of the National Centre for Biotechnology (NCBI) for China, South Korea, Japan, Mongolia, the United States of America, the United Kingdom, and Africa (South Africa 3: (GCF_000590775.1, GCF_001653415.1, GCF_000448525.1; Nigeria 1: GCF_900638475.1; Gambia 1: GCF_000185205.1). We then investigated the similarities and differences between the genomic compositions of *H. pylori* strains isolated in our study cohort in South Africa, (n = 23) and the available whole genome sequences retrieved from the GenBank database of the NCBI. We therefore considered 247 samples in total.

6.2.8 Virulence factors

We determined the virulent (VFs) factors including *vacA* and *cagA* that contribute to the ability of *H. pylori* to cause disease by ABRicate v0.7 [21] using the virulence factor database (VFDB). These

VFs enable the bacteria to colonize a host, evade or suppress the host's immune response, and obtain nutrients from the host. We screened for the VFs of the 23 samples generated in our lab and that of 224 publicly accessible *H. pylori* genomes (119 from Asia, 3 from South Africa, 1 from Nigeria and 1 from Gambia, 90 from the USA, and 10 from the United Kingdom).

6.2.9 Comparative genomics phylogenetic tree construction and generation

Lab samples were aligned to the reference genome (strain MT5135: GCF_017821535.1, NCBI Ref Seq assembly) using bowtie2 v2.2.8. SNP calling was performed with freebayes v1.2.0-4-gd15209e [(22)], and the resulting SNPs were integrated into the reference genome using bcfools v1.3.1, with htslib 1.3.1. For phylogenetic analysis, kSNP3 [(23)] was used to generate a pairwise distance matrix based on the number of SNP differences across genomes and to construct the phylogenetic tree. The SNP differences were identified by scanning k-mers (subsequences of length k) across genome sequences to pinpoint variable positions between genomes. The genetic distance between the two genomes was calculated as the proportion of differing SNPs. kSNP3 was run using the following parameters: -k optimal_k -ML -NJ -core -min_frac 0.75, where optimal_k=31 was determined using the Kchooser script bundled with kSNP3. The -ML option used the maximum likelihood method for tree construction, while -NJ enabled the neighbor-joining algorithm. The -core option ensured that only core genome SNPs were used, and -min_frac 0.75 set the minimum fraction of genomes required to call a core SNP. The resulting tree was visualized with iTOL [(24)] for final presentation and interpretation.

6.3 RESULTS

A total of 23 phenotypically identified as *H. pylori* isolates were cultured from our cohort. The microscopic view of *H. pylori* gram staining and colonies cultured from our original biopsy samples are depicted in Supplementary Figures S6.2 and Figure S6.3, respectively. All the *Helicobacter pylori* isolates showed high sequencing quality, with Q30 scores exceeding 96% for forward reads and 92% for reverse reads and read counts ranging from approximately 652,000 to 1.55 million before trimming. After trimming, read counts ranged from around 586,000 to 1.41 million. Taxonomic classification revealed that all but one isolate had over 80% of reads and over 90% of weighted scaffolds assigned to *H. pylori*, with FastANI coverage above 90%, confirming species-level identity. One outlier isolate showed signs of mixed taxonomy, with less than 70% of

reads and scaffolds assigned to *H. pylori* and *S. salivarius*, respectively and FastANI coverage below 90%. Genome assemblies ranged from 1.57 to 1.71 Mb in size, with N50 values between 66 kb and 199 kb, L50 values between 4 and 8, and contig counts between 25 and 117. GC content was consistently around 39%, except for one isolate at 42%, indicating overall high-quality draft genomes suitable for downstream analysis.

6.3.1 Antimicrobial genes

A heatmap showing the occurrence of AMR genes and a bar plot showing the frequency and degree of hits for the AMR genes in the African strains are shown in Supplemental Figure 6.4 and Supplemental Figure S6.5, respectively (for the other countries, see supplementary figure S7.15). The distribution of regional antimicrobial (AMR) genes is depicted in Table 6.1 whilst Table 6.3 summarizes the AMR gene function and antibiotic affected. Similar distribution of the AMR genes was evident except for *TEM-116* which was identified in American strains. Whilst regionally the *vanT* gene was prevalent in all regions, the vanY gene was present only in the African strains. The distribution and regional occurrence rates of respective AMR genes are depicted in Supplemental Table 1. It is important to mention that, based on the ABRicate v0.7 and the plasmidfinder database, no plasmids or replicons were detected in the strains.

6.3.2 Virulence factors

The comparison of the distribution of virulence genes is shown in Table 6.4. Supplementary Figure S6.6 showed the heatmaps of the virulence genes in the African cohort, and supplementary Table S6.2 summarizes the VFs associated with *H. pylori* and their functions. The number and range of VFs in *H. pylori* differed between regional strains. The lowest number of VFs was found among strains in the African cohort (n=71: range 71—113) compared to Japan (n=87—111), China (n=90—112), Mongolia (88—114), Thailand (n=111—117), South Korea (n=111—117), USA (n=85—112) and UK (n=91—121).

The *BabA/hopS* and *SabB/hopO* gene was absent among all countries in the Southeast Asian region while the *BabB/hopT* gene was absent in the strains from China. We also noted that the rate of occurrence of the *cagA* gene was highest among the USA/UK cohorts while the *VacA* gene was most common among the cohorts from Japan and Korea. Most importantly, the occurrence of *cagPAI* polymorphic variant genes was most prevalent in the cohorts from the East Asian countries

and lower among the African strains. Also, *Rfaj* was predominant in the Southeast Asian present in 94—100% of strains but occurred only in 50% of *H. pylori* in Africa.

6.3.3 Multilocus sequence typing (MLST)

We determined the multi-locus sequence types (ST) of the studied *H. pylori* strains and out of 247 we managed to characterize 40. Overall, Africa (n=6), South Korea (n=1), UK (n=5), and USA (n=28) were available for analysis. Supplementary Table S6.3 shows the sequence types obtained from our analysis. None of the African ST (1320, 3051, 3583, 3582, 1288, and 340) were found in the other regional cohorts. The US cohort was dominated by ST 3020. It is important to mention that in addition to determining sequence types using the MLST command-line tool, the genomes of the 23 lab generated *Helicobacter pylori* isolates were uploaded to the Centre for Genomic Epidemiology (CGE) to identify their nearest sequence types (STs). While most isolates matched clearly to a single ST, several showed proximity to multiple sequence types, suggesting either high genetic similarity across STs or potential recombination events.

6.3.4 Phylogenetic relatedness

Based on the maximum likelihood (ML) phylogenetic tree of *H. pylori* (Figure 6.2), the African genotypes are more closely related to the US and UK, but most distant from Chinese strains, followed by the Japanese, Mongolian, and Thai strains. The African *H. pylori* genotype is sandwiched between the US strains. The ectopic occurrence of foreign strains interspersed in other countries may denote migration.

6.4 DISCUSSION

H. pylori is known to cause several gastric pathologies including gastric and duodenal ulcers and GCAs. However, studies have shown a wide regional and geographical variability in the distribution of the various pathologies caused by the bacteria, warranting further interrogation into the molecular structure of *H. pylori* as it relates to these various regions. This section of the study centered on the molecular signatures of African *H. pylori* strains. Important findings include the following.

6.4.1 Antimicrobial resistance

We noted in our study that resistance genes that are not associated with drugs used in the treatment of *H. pylori* were found during the *H. pylori* sequencing. These genes include the *vanY gene* in *vanM* cluster, *vanY gene in vanG* cluster, and *vanY gene in vanB* cluster identified only in a South African strain, whilst the *Van T* gene was more widespread across the various regions. These genes are said to confer resistance to vancomycin. Similarly, the *TEM-116* resistance gene which is a B-lactamase resistance gene was restricted to 2.2% (2/90) population of the American cohort only.

We also noted that *gyrA*, a fluoroquinolone resistance gene was more common in the Asian and American strains, occurring at 25.2% (30/119) and 25.6% (23/90) respectively versus the African and UK cohorts, where 7.1% (2/28) and 0%, contained the gene. Levofloxacin which is a flouro-quinoline was recommended in the American Gastroenterology Association guideline in 2017 as a component of the salvage therapy for *H. pylori* eradication following the increased resistance noted with clarithromycin-based therapy [25]. However, antimicrobial susceptibility testing reports noted a rise in levofloxacin resistance with an incidence of 28-33% in Southeast Asian countries, 23-56% in the USA, and a lower resistance rate in African countries [9]. We also noted that the *rpoB* gene (rifampicin resistance) is least common among the East Asian population, as opposed to Africa, America, and the UK at 75%, 88.9%, and 80% respectively. Although rifampicin is not part of the regimen for the eradication of *H. pylori*, rifabutin shares the same *rpoB* resistance gene with rifampicin. Both drugs are important in the treatment of tuberculosis and leprosy which are endemic in Africa. Though global data regarding rifabutin resistance is sparse, data from Nigeria showed a high resistance rate for rifabutin [9].

In addition, our analysis noted that *frxA/rdxA* metronidazole resistance genes were found to be commoner in the American/UK and African cohorts, the Southeast Asian countries have rates comparatively lower than the Americans/Africans. These findings corroborate available data which showed a higher resistance rate for metronidazole in the African and American populations. Also of note, the African cohort has the least resistance gene for clarithromycin occurring at 17.9% as opposed to the UK at 100%, America at 98.9% and East Asia at 98.3%.

The management of *H. pylori* requires a combination of more than one antibiotic. Though the eradication of *H. pylori* is key to the prevention of GCA and treatment of ulcers, the global epidemic of drug resistance poses a risk to the achievement of these goals, and understanding the local resistance profile is paramount in the management of *H. pylori*-associated diseases. A review

by Argueta et al has shown the growing trend of antibiotic resistance against *H. pylori* across regions [9]. Except for levofloxacin where the rate of resistance gene occurrence matches with the real-world resistance profile from the published data, the occurrence in the rest of the antibiotic resistance genes seen in our study does not show a corresponding increase in the real-life antibiotic resistance profile based on available data. The underlying reason for this remains unclear.

6.4.2 Virulence factors

The ability of *H. pylori* to cause diseases is associated with the existence of virulence factors, especially *VacA* and *CagA*. Importantly, *CagA* is reported to be carcinogenic. We, therefore, compared the occurrence of important virulence factors in the African cohort, where it is reported to have a lower incidence of GCA despite the higher prevalence of *H. pylori* compared to the Asian population with a higher incidence of GCA, as well as to the American and UK cohorts with a similar incidence of GCA to the African population.

In our study, we noted that the rate of occurrences of the *CagA* gene was the highest among the USA/UK cohorts but rare in the Japan and Chinese cohorts. We also noted that *CagPAI* morphologic variants were comparatively lower in the African cohort (46.6%) compared with the Southeast Asian countries (86-100%). In the context of the relatively higher incidence of GCA in Southeast Asian countries, despite the lower occurrences of the *cagA* gene which is an established *H. pylori* oncogenic gene, the question about the true contribution of the *cagA* gene as an independent factor in the pathogenesis of GCA and the possibilities of *cagPAI* morphologic variants genes playing a more important role in the pathogenesis of GCA becomes paramount. Especially, since the rate of occurrences of these *cagA* genes does not correspond to the incidence of GCA in those regions, rather, the rate of the occurrences of cag*PAI* variant genes appears to be related well with the incidence of GCA. This has been reported in earlier studies where they noted that the polymorphism in the *cagA, cagL*, and *cagI* were associated with the development of GCA [6]. The incidence of these *cagPAI* genes in the American cohorts was relatively higher than in the African cohorts. Yet, the incidence of GCA is similar to what is reported in the African sub-region. However, the lower incidence of GCA in the American population could be attributed to the lower prevalence of *H. pylori.* Conversely, among the African population, parasite co-infection could contribute to the low GCA in Africa.

We also noted the disproportionate increase in the *VirB/D* type 3 T4SS (secreting system) in the Southeast Asian countries compared to the African cohorts and Japan, which has the highest occurrence of the *VacA* gene. Andrzej Szkaradkiewicz et al reported that strong expression of the *VirB/D* channeling systems in the presence of moderate activity of either *CagA* or *VacA* is associated with chronic gastritis. Also, the progression in the oncogenic process is associated with the strong expression of *CagA* and *VirB/D* systems [26,27]. Therefore, the lower occurrence of the *VirB/D* genes in the African cohort may account for the lower incidence of GCA in the African populations.

6.4.3 Phylogenetic tree

From our study the African *H. pylori* is genotypically related to the US cohort, and least likely to the East Asian cohorts. Considering that the incidence of GCA in these regions is at par, it is possible that the *H. pylori* from the African and American cohorts share common genotypic features, and hence their virulence, therefore lending further credence to the 'African enigma'. This fact is strongly supported by the fact that from an earlier retrospective study, we conducted in GSH (chapter 4), a quaternary facility with open access to endoscopy, the average annual incidence of GCA was 2.98/100 000 (unpublished).

6.5 CONCLUSIONS

The findings from the *H. pylori* molecular signatures may indicate that the so-called 'African enigma' may be true. The following supports this; firstly, the African cohort has a lower incidence of the *CagPAI* morphologic variants which have been associated with the development of GCA. Secondly, *VirB/D*, a type IV secreting system has been shown to potentiate the oncogenic activity of *CagA/VacA*. These genes are, however, lower in the African cohort compared to the Southeast Asian countries. Finally, phylogenetically, the African strain appears to be similar to the American type and may share less virulence, hence both regions have a similar lower incidence of GCAs compared to the Southeast Asian countries.

However, a limitation of our study is that other risk factors such as genetic susceptibility, environmental and lifestyle factors were not studied and thus focused only on molecular differences based on WGS. Furthermore, the fact that the STs identified in the African strains were unique and not found elsewhere globally (i.e. ST1320, 3051, 3583, 3582, 1288, and 340) needs

further investigation and evolutionary analysis. This is required to understand their origins, genetic diversity, and potential role in disease dynamics. These unique STs may indicate region-specific evolutionary pressures, distinct ecological niches, or host-pathogen interactions unique to the African context.

6.6 REFERENCES

1. Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, et al. A review of human carcinogens—Part B: biological agents. The lancet oncology. 2009;10(4):321-2.

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2021;71(3):209-49.

3. Holcombe C, Umar H, Lucas SB, Kaluba J. Low incidence of clinically significant gastroduodenal pathology despite a high incidence of *Helicobacter pylori* infection. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1994;88(5):569-71.

4. Holcombe C. *Helicobacter pylori*: the African enigma. Gut. 1992;33(4):429.

5. Shiota S, Suzuki R, Yamaoka Y. The significance of virulence factors in *Helicobacter pylori*. Journal of digestive diseases. 2013 Jul;14(7):341-9.

6. Canzian F, Rizzato C, Obazee O, Stein A, Flores-Luna L, Camorlinga-Ponce M, Mendez-Tenorio A, Vivas J, Trujillo E, Jang H, Chen W. Genetic polymorphisms in the cag pathogenicity island of *Helicobacter pylori* and risk of stomach cancer and high-grade premalignant gastric lesions. International journal of cancer. 2020 Nov 1;147(9):2437-45

7. Alam J, Sarkar A, Karmakar BC, Ganguly M, Paul S, Mukhopadhyay AK. Novel virulence factor dupA of *Helicobacter pylori* as an important risk determinant for disease manifestation: An overview. World Journal of Gastroenterology. 2020 Aug 8;26(32):4739.

8. Malfertheiner P, Megraud F, Rokkas T, Gisbert JP, Liou JM, Schulz C, Gasbarrini A, Hunt RH, Leja M, O'Morain C, Rugge M. Management of *Helicobacter pylori* infection: the Maastricht VI/Florence consensus report. Gut. 2022 Sep 1;71(9):1724-62.

9. Argueta EA, Ho JJ, Elfanagely Y, D'Agata E, Moss SF. Clinical implication of drug resistance for *H. pylori* management. Antibiotics. 2022 Nov 23;11(12):1684.

10. Meijers E, Verhees FB, Heemskerk D, Wessels E, Claas ECJ, Boers SA. Automating the Illumina DNA library preparation kit for whole genome sequencing applications on the flowbot ONE liquid handler robot. Scientific Reports. 2024;14(1):8159.

11. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010.

12. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-8.

13. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.

14. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology. 2012;19(5):455-77.

15. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072-5.

16. Tanizawa Y, Fujisawa T, Nakamura Y. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics. 2018;34(6):1037-9.

17. Alcock, B., Huynh, W., Chalil, R., Smith, K., Raphenya, A., Wlodarski, M.,et al. (2020). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database (Version 6.0.3) [Computer software]. https://doi.org/10.1093/nar/gkac920

18. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic acids research. 2023;51(D1):D690-D9.

19. MLST. https://github.com/tseemann/mlst. 2.23.0

20. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018 Sep 24;3:124. doi: 10.12688/wellcomeopenres.14826.1. PMID: 30345391; PMCID: PMC6192448.

21. Seemann T, *Abricate*, **Github** https://github.com/tseemann/abricate

22. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907 [q-bio.GN]* 2012 v1.2.0-4-gd15209e

23. Shea N Gardner, Tom Slezak, Barry G. Hall, kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome, *Bioinformatics*, Volume 31, Issue 17, September 2015, Pages 2877–2878, https://doi.org/10.1093/bioinformatics/btv271

24. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024 Jul 5;52(W1):W78-W82. doi: 10.1093/nar/gkae268. PMID: 38613393; PMCID: PMC11223838.

25. Chey WD, Leontiadis GI, Howden CW, Moss SF. ACG clinical guideline: treatment of *Helicobacter pylori* infection. Official journal of the American College of Gastroenterology| ACG. 2017 Feb 1;112(2):212-39.

26. Szkaradkiewicz A, Karpiński TM, Linke K, Majewski P, Rożkiewicz D, Goślińska-Kuźniarek O. Expression of cagA, virB/D complex and/or vacA genes in *Helicobacter pylori* strains originating from patients with gastric diseases. PloS one. 2016 Feb 11;11(2):e0148936.

27. Rizzato C, Torres J, Obazee O, Camorlinga-Ponce M, Trujillo E, Stein A, Mendez-Tenorio A, Bravo MM, Canzian F, Kato I. Variations in cag pathogenicity island genes of *Helicobacter pylori* from Latin American groups may influence neoplastic progression to GCA. Scientific reports. 2020 Apr 16;10(1):6570.

Table 6.1: *H. pylori* regional antibiotic resistance rates [9]

Antibiotic	**Thailand**	**China**	**Japan**	**South Korea**	**USA**	**UK**	**South Africa**	**Nigeria**	**Morocco**
Clarithromycin	19.0%	37.0%	28.0%	18.0%	14.6-70.4%	36%	18.4%	36%	27.4%
Amoxicillin	1.0%	1.0%	NA	4.0%	0.0-6.7%	2.0%	21.5%	67.7%	0.0%
Metronidazole	44.0%	77.0%	NA	40%	12.0-82.4%	57%	91.3%	94.2%	0.0%
Tetracycline	0.0%	2.0%	NA	4.0%	0.0-1.7%	2.0%	8.7%	58%	0.0%
Levofloxacin	31.0%	33.0%	NA	28.0%	23.6-52%	11%	10.3%	0.0%	0.0%
Rifabutin	NA	NA	NA	NA	NA-0.5%	NA	NA	96.8%	NA

[*]**NA = not available**

Table 6.2: Distribution of antibiotic resistance genes in *H pylori* for the various countries under study

AMR Genes	Africa	USA/UK	China	Japan	South Korea/Thailand	Mongolia
Hp_pbp1_AMX,	✓	✓	✓	✓	✓	✓
Hp_pbp2_AMX,	✓	✓	✓	✓	✓	✓
Hp_pbp3_AMX	✓	✓	✓	✓	✓	✓
Hp1181	✓	✓	✓	✓	✓	✓
Hp_rdxA_MET	✓	✓	✓	✓	✓	✓
Hp_23S_rRNA_	✓	✓	✓	✓	✓	✓
Hp_frxA_MET	✓	✓	✓	✓	✓	✓
Hp_gyrA_FXNs	✓	✓	✓	✓	✓	✓
Hp_rpoB_RIF	✓	✓	✓	✓	✓	✓
vanT_vanLcluster	✓	✓	✓	✓	✓	✓
vanT_vanG cluster	✓	✓	✓	✓	✓	✓
vanY_vanB cluster	✓	✗	✗	✗	✗	✗
vanY_vanG cluster	✓	✗	✗	✗	✗	✗
vanY_vanM cluster	✓	✗	✗	✗	✗	✗
TEM-116	✗	✓	✗	✗	✗	✗
Total	14	12	11	11	11	11

NB: ✓ means resistance is present and ✗ means the resistance gene is absent

Table 6.3: Function of AMR genes

Resistance genes	**Antibiotic affected**
gyrA	Fluoroquinolones
pbp1, pbp2 and pbp3	Amoxicillin
hp1181	Membrane multidrug efflux protein
rdxA	Metronidazole
23S Rrna	Clarithromycin
FrxA	Metronidazole
rpoB	Rifampicin
vanT gene in the vanL cluster *vanT* gene in the vanG cluster *vanY* gene in vanM cluster *vanY* gene in vanG cluster *vanY* gene in vanB cluster	Vancomycin
TEM -116	Beta-lactams

Table 6.4: Comparison of the distribution and prevalence of virulent genes

Virulent gene present	Africa % n=28	America % n=90	UK % n=10	Japan % n=28	South Korea % n=12	China % n=37	Mongols % n=38	Thailand% n=4
HP0256	82.1	98.9	100	100	100	100	100	100
HP_RS02435	53.6	88.8	100	0	50	81.1	60.5	83.3
HP_RS03030	100	100	100	100	100	97.3	100	100
HP_RS03480	100	100	90	100	100	97.3	100	100
HP_RS04690	100	100	100	100	100	97.3	100	100
HP_RS07005	85.7	98.9	100	100	100	97.3	100	100
HP_RS07240	100	100	100	100	100	100	100	100
babA.hopS	36	10.1	50	0	0	0	0	0
babB.hopT	25	16.9	30	14.3	50	0	7.9	66.7
cag1	46.4	89.9	100	96.4	100	97.3	89.5	100
cag2	3.6	2.2	50	0	50	2.7	42.1	41.7
cag3	46.4	91	90	96.4	100	100	89.5	100
cagA	32.1	77.5	90	0	25	0	15.8	41.7
cagD	46.4	85.4	90	96.4	100	97.3	86.8	100
cagF	46.4	86.5	90	96.4	100	97.3	86.8	100
cagG	46.4	86.5	90	96.4	100	97.3	86.8	100
cagH	46.4	86.5	90	96.4	100	97.3	86.8	100
cagI	46.4	86.5	90	96.4	100	97.3	86.8	100
cagM	46.4	86.5	90	96.4	100	97.3	86.8	100
cagN	46.4	86.5	90	96.4	100	97.3	86.8	100
cagP	46.4	86.5	90	96.4	100	91.9	86.8	100
cagQ	46.4	80.9	90	14.3	25	13.5	21	100
cagS	46.4	87.6	90	96.4	100	94.6	86.8	100
cagU	46.4	86.5	90	96.4	100	97.3	86.8	100
cagZ	46.4	88.8	90	96.4	100	97.3	86.8	100

cds6	100	98.9	100	100	100	100	100	100
cheA	100	85.4	100	100	100	100	100	100
cheV1	96.4	100	100	100	100	100	100	100
cheV2	100	98.9	100	100	100	100	100	100
cheV3	100	100	100	100	100	97.3	100	100
cheW	89.3	100	100	100	100	100	100	100
cheY	100	100	100	100	100	100	100	100
flaA	100	100	100	100	100	97.3	100	100
Flab	100	100	100	100	100	100	100	100
Flag	89.3	100	100	100	100	100	100	100
flgA	89.3	100	90	100	100	100	100	100
flgB	96.4	100	100	100	100	100	100	100
flgC	100	100	100	100	100	100	100	100
flgD	100	100	100	100	100	97.3	100	100
flgE	100	100	100	100	100	100	100	100
flgE_1	100	100	100	100	100	97.3	100	100
flgG	100	100	100	100	100	97.3	100	100
flgG_2	100	100	100	100	100	100	100	100
flgH	100	100	100	100	100	100	100	100
flgI	100	100	100	100	100	100	100	100
flgK	100	100	100	100	100	100	100	100
flgL	100	100	100	100	100	100	100	100
flgM	92.9	100	100	100	100	100	100	100
flgR	100	100	100	100	100	100	100	100
flgS	100	100	100	100	100	100	100	100
flhA	100	100	100	100	100	97.3	100	100
flhB	96.4	100	100	100	100	100	100	100
flhB2	100	100	100	100	100	100	100	100
flhF	100	100	100	100	100	97.3	100	100
fliA	96.4	98.9	100	100	100	100	100	100
fliD	100	100	100	100	100	97.3	100	100
fliE	96.4	100	100	100	100	100	100	100
fliF	100	98.9	100	100	75	97.3	100	100

fliG	100	98.9	100	100	75	100	100	100
fliH	100	100	100	100	75	100	100	100
fliI	100	100	100	100	100	100	100	100
fliL	100	100	100	100	100	100	100	100
fliM	100	98.9	100	100	100	100	100	100
fliN	100	100	100	100	100	100	100	100
fliP	100	100	100	100	100	97.3	100	100
fliQ	100	100	100	100	100	100	100	100
fliR	100	100	100	100	100	97.3	100	100
fliS	100	100	100	100	100	100	100	100
fliY	89.3	100	100	100	100	100	100	100
futA	0	30.3	20	0	0	0	0	0
futB	0	30.3	10	0	0	0	0	8.3
futC1	100	100	100	100	100	100	100	100
futC2	82.1	79.8	90	71.4	100	97.3	94.7	100
gluE	100	97.8	100	100	100	97.3	100	100
gluP	100	100	100	100	100	97.3	100	100
hopZ	32.1	48.3	90	0	25	2.7	18.4	41.7
hpaA2	100	95.5	90	57.1	75	8.1	55.3	83.3
kdtB	100	100	100	100	100	100	97.4	100
lpxB	100	100	100	100	100	100	100	100
motA	100	100	100	100	100	100	100	100
motB	100	100	100	100	100	100	100	100
napA	100	100	100	100	100	97.3	100	100
oipA.hopH	92.9	97.8	90	100	100	75.7	100	100
pdxA	100	98.9	100	100	100	100	100	100
pdxJ	100	100	100	100	100	100	100	100
pflA	100	100	100	100	100	100	100	100
pseB	96.4	98.8	100	100	100	100	100	100
pseC	89.3	98.8	100	100	100	100	100	100
pseFG	92.9	100	100	100	100	100	100	100
pseH.flmH.flaG1	100	100	100	100	100	100	100	100

pseI	100	100	100	100	100	100	100	100
rfaC	92.9	100	100	100	100	100	100	100
rfaJ	50	47.2	90	100	100	97.3	94.7	100
rfbD	96.4	100	100	100	100	100	100	100
rfbM	100	100	100	100	100	100	100	100
sabA.hopP	25	60.7	60	39.3	50	73	47.4	100
sabB.hopO	7.1	2.2	20	0	0	0	0	0
tlpA	100	98.8	100	100	100	94.6	97.4	100
tlpB	82.1	98.8	100	100	100	97.3	100	100
tlpC	89.3	98.8	100	100	100	97.3	97.4	100
ureA	100	100	100	100	100	100	97.4	100
ureB	100	100	100	100	100	100	100	100
ureE	100	100	100	100	100	100	100	100
ureF	85.7	100	100	100	100	100	100	100
ureG	100	100	100	100	100	100	100	100
ureH	92.9	100	100	100	100	100	100	100
ureI	100	100	100	100	100	100	100	100
vacA	46.4	56.2	40	92.9	75	62.1	34.2	41.7
virB1.cag4	46.4	87.6	90	21.4	100	45.9	68.4	83.3
virB10.cagY	0	3.4	20	0	0	32.4	2.6	33.3
virB11	46.4	88.8	90	96.4	100	100	86.8	100
virB2.cagC	46.4	86.5	90	96.4	100	97.3	86.8	100
virB4.cagE	46.4	86.5	90	96.4	100	97.3	86.8	100
virB5.cagL	46.4	86.5	90	96.4	100	97.3	86.8	100
virB6.cagW	46.4	86.5	90	96.4	100	97.3	86.8	100
virB7.cagT	46.4	86.5	90	96.4	100	97.3	86.8	100
virB8.cagV	46.4	86.5	90	96.4	100	97.3	86.8	100
virB9.cagX	46.4	86.5	90	96.4	100	97.3	86.8	100
virD4.cag5	46.4	89.9	90	96.4	100	100	86.8	100
wbcJ	100	100	100	100	100	100	100	100
wbpB	96.4	100	100	100	100	100	100	100
ylxH	100	100	100	100	100	100	100	100

*The virulence gene segments highlighted in red are marked by remarkable differences in the countries under study.

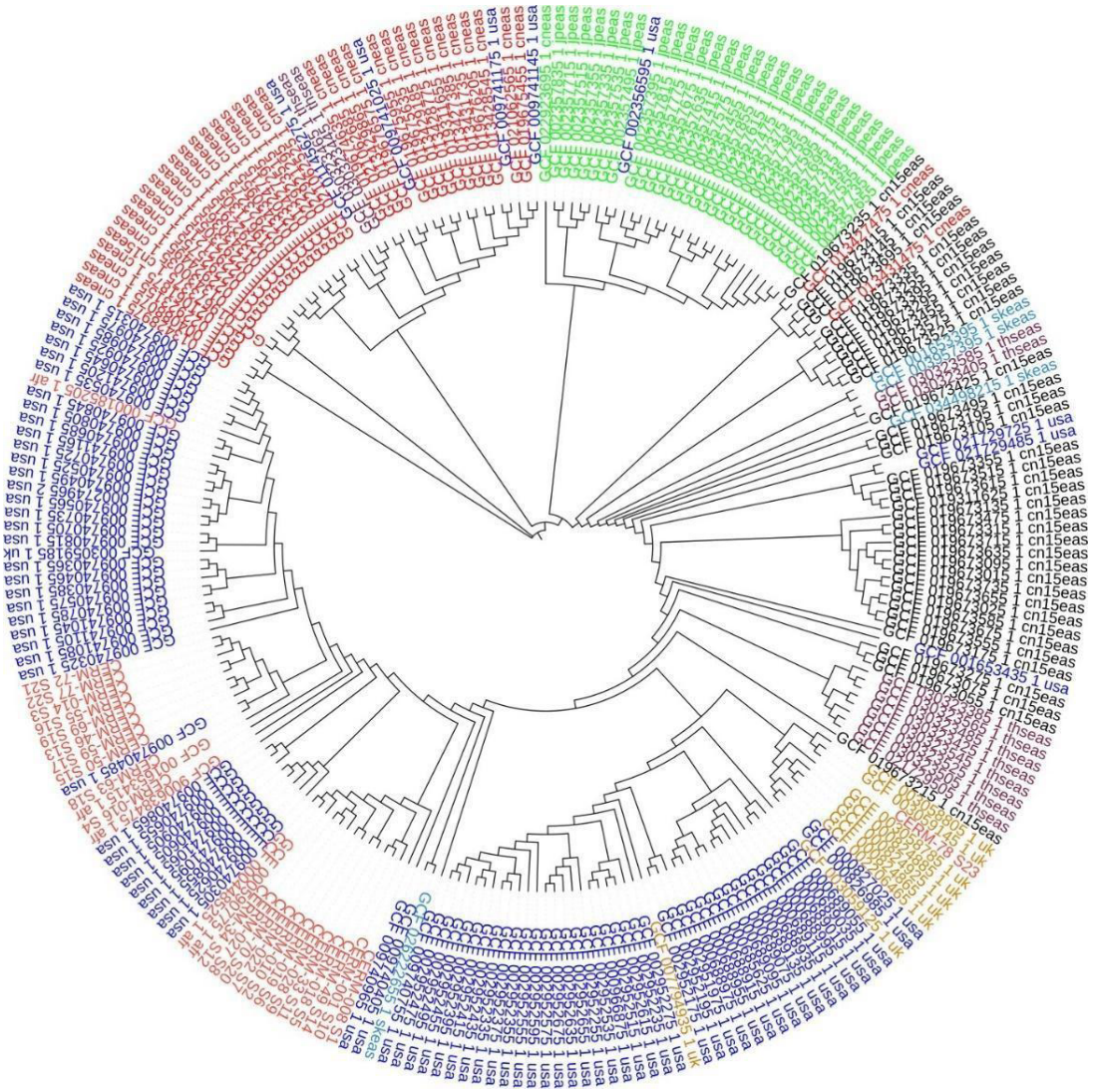


Figure 6.2: The maximum likelihood (ML) phylogenetic tree of *H. pylori* (n=247)

*Blue ending with *usa* represents USA, Yellow ending with *uk* represents UK, Green ending with *jpeas* - Japan, Brown ending with *afr* denotes South African, Nigerian and Gambian samples except our laboratory samples, Brown CERMS with numbers are our laboratory study samples (South Africa), Magenta colour ending with *cneas* denotes China, *cn15eas* stands for Mongols, *thseas* stands for Thailand, *skeas* stands for South Korea.

CHAPTER 7

THE GASTRIC MICROBIOME: COMPARING A COHORT WITH HELICOBACTER PYLORI POSITIVE TO A HELICOBACTER PYLORI NEGATIVE COHORT":

ABSTRACT

Background: *Helicobacter pylori* is common, affecting more than 40% of the world's population, with the highest incidence in developing countries. The role of *H. pylori* in the pathogenesis of GCA has been widely studied. It is said to cause dysbiosis. There is evidence suggesting a synergism *of H. pylori* with the gastric microbiome in the pathogenesis of GCA.

Methods: DNA of the gastric biopsy samples for both the control (*H. pylori* negative, n=56) and the test (*H. pylori* positive, n=23) cohorts were extracted and subsequently 16S rRNA gene amplicon sequencing on Illumina MiSeq platform was conducted. Using bioinformatics**,** taxonomic classification was performed using DADA2 and the Silva v138 database. All downstream analysis was done in R v4.3.0 (2023-04-21 ucrt). Alpha diversity was determined after rarefication at a depth of 10 000 reads per sample using Chao1, Simpson, Shannon, and observed species indices. While the Bray-Curtis dissimilarity was used for β-diversity, and measuring compositional dissimilarity based on abundance data. The β-diversity distances were visualized using the principal coordinate analysis (PCoA). For statistical testing whether microbial community composition significantly differs between groups and to compare ranked dissimilarities among groups, we utilized PERMANOVA (Permutational Multivariate Analysis of Variance) and ANOSIM (Analysis of Similarities). We also compared gastric microbiome diversity in our cohorts with regard to epidemiological and clinical determinants and published data from Asia, Europe, and the United States.

Results: In the test group, though there was some degree of dysbiosis with less diversity in the *H. pylori* positive cohort, no statistically significant difference in the microbial richness between the cases and control were observed, despite the controls having higher microbial diversity compared to the cases. The comparison of α−diversity using the Shannon index between controls and cases revealed a statistically significant difference (p-values of 0.0154) as did the analysis of microbial diversity using the Simpson index (p-value < 0.05). The most abundant phylum in the control and

test group was *Campylobacterota* and *Firmicutes*, respectively. No statistically significant differences in gastric microbiome composition in our cohorts were found for sex, race, diet, peptic ulcer, and urease positivity nor with published data from other countries.

Conclusions: *H. pylori* infection causes gastric microbiome dysbiosis, but no significant differences between test and control groups were observed in microbial richness nor were there significant differences in the gastric microbiome diversity in the South African cohort and compared to Southeast Asian and European countries or the United States.

7.1 BACKGROUND

According to Hoo et al 2017, *H. pylori* is common, affecting more than 50% of the world's population with the highest seen in developing countries [(1)]. The role of *H. pylori* in the pathogenesis of GCA has been widely studied, with evidence suggesting the possibility of synergistic effects with the gastric microbiome. In the past, gastric microbiota isolates were mainly culture-dependent methods with *Veillonella, Lactobacillus,* and *Clostridium spp* being the frequently isolated genera [(2)]. However, with the advent of PCR such as the short gun sequencing and new-generation sequencing, a greater number of bacteria taxa can be detected in the stomach despite the high acidic gastric internal milieu.

Based on these newer techniques, the predominant phyla have been *Proteobacteria, Firmicutes, Bacteroidetes, Actinobacteria,* and *Fusobacteria* in normal conditions [(3)]. Bacteria such as *Veillonella, Lactobacillus,* and *Clostridium* are present in the oral cavity and duodenum and can transiently colonize the duodenum [(3)]. It is also important to note that the gastric microbiome can be influenced by the mode of childbirth in infants, age, sex, ethnicity, diet, lifestyle, geography, use of antibiotics, PPI, and the presence of *H. pylori* [(4-9)]. However genetic variations have been shown to have no effects on the gastric microbiome as seen in a study using the twin model [(10)]. In a study by Noto JM et al, they noted a very diverse and complex microbial biome in a *H. pylori* negative cohort compared to a *H. pylori* positive group, with the predominant phyla being *Proteobacteria* in the *H. pylori* positive cohort, and *Firmicutes* and *Actinobacteria* predominant in the *H. pylori* negative cohort [(11)] (Supplemental Figure S7.1).

7.1.1 Microbiome and GCA

In the recent past, the role of the gastric microbiome in the development of GCA has become a topic of great interest. In an animal model study done by Lofgren JL et al, they demonstrated that INS-GAS mice harboring a complex microbiota developed GCA within 7 months following *H. pylori* infection, whereas the development of GCA was prolonged in germ-free mice that were colonized by *H. pylori* only; lending credence to the contributions of the gastric microbiome in the pathogenesis of GCA [12]. The gastric mucosa is regularly bathed with a highly concentrated acid, impairing the growth and survival of many bacteria. Chronic infection with *H. pylori* is associated with an increase in the pH of the internal milieu of the stomach, hence the changes in the gastric microbiome. Studies have shown that there is significant microbiome diversity among patients who are *H. pylori* negative with *Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Bacteroidetes,* and *Fusobacteria* being the commonest phyla, while among those who are *H. pylori* positive, it became the predominant organism with 72-97% occurrence [13-15]. In another study, it was also noted that *H. pylori* was present at relatively low abundance in patients with advanced premalignant lesions and that the microbiota of patients with GCA were dominated by species of *Lactobacillus*, *Streptococcus*, *Veillonella*, and *Prevotella* [11,15]. Also, a study by Dicksved J *et al*. on the gastric carcinogenesis pathway demonstrated a steady decrease in bacterial diversity of the gastric microbiota, with an increasing abundance of *Lactobacillus* and *Lachnospiraceae* in the late phase of this process [11-16], with similar bacteria noted in gastric biopsied samples in other studies [9, 11, 17]. These findings corroborate the fact that among other factors, the development of GCA is possibly an interplay between *H. pylori* and other gastric microbiomes. The intricate interplay between the host immune response, the tumor microenvironment, genetic factors, and the gastric microbiome is thought to impact the pathogenesis of GCA and the response of cancer to chemotherapy and immunotherapy [18,19], necessitating the possibilities of microbiome-based markers as surrogates for determining and individualizing cancer treatment options. Specific bacterial taxa and metabolic pathways associated with GCA prognosis have been identified in studies, paving the way for better treatment options in the future [20].

To ascertain differences in gastric microbiome composition in South African patients at GSH gastric biopsies were submitted for 16SrRNA sequencing by the Elizayo biome unit, Division of Medical Microbiology, Faculty of Health Sciences, University of Cape Town.

7.1.2 The interplay between dietary, lifestyle factors, and the gastric microbiome

Diets both in type and quantity consumed directly influence the gastric microbiome [21,22].

A diet rich in fiber and prebiotics, which act as substrates for beneficial bacteria, can promote their growth, leading to a more diverse and balanced microbiome [21,23]. Conversely, a diet high in processed foods, saturated fats, and simple sugars can lead to dysbiosis, characterized by a decrease in beneficial bacteria and an increase in potentially harmful species [23].

Lifestyle factors, including smoking, alcohol consumption, and stress, can indirectly influence the gastric microbiome [24,25]. Smoking and alcohol abuse are well-established risk factors for various gastric diseases, and they can significantly alter the composition and function of the gastric microbiome [24]. These effects are likely mediated through various mechanisms, including direct toxicity to bacterial cells, altered gastric pH, and impaired mucosal immunity. Chronic stress indirectly impacts the gastric microbiome through its effects on the gut-brain axis [25]. Stress can alter gut motility, inflammation, and immune function, which can have downstream consequences for the gastric microbiome.

7.1.3 Drugs and the gastric microbiome

Drugs such as antibiotics, proton pump inhibitors (PPIs), and other medications can profoundly affect the gastric microbiome [24,26]. This is either by directly killing some bacteria and giving room for opportunistic bacteria to grow as in the case of antibiotics or by changing the microenvironment by altering the pH as seen with PPIs, which then favours certain bacteria over others leading to dysbiosis [26,27].

7.1.4 Genetic factors

Host genetics can influence several aspects of the gastric microbiome, including susceptibility to infections, the intensity and duration of inflammatory responses, and the overall composition of the microbial community [28].

7.2 METHODS

Seventy-nine (79) gastric biopsied specimens were collected, of which 56 samples were negative for *H. pylori* and served as controls. 23 samples that were positive for *H. pylori* were used as test samples.

7.2.1 Preparation of 16S rRNA sequencing libraries using the JCVI protocol and MiSeq v3 600 Cycle Kit

7.2.1.1 DNA Extraction and quality assessment

Genomic DNA was extracted from gastric biopsies using the MN NucleoSpin Microbial DNA kit per the manufacturer's instructions. The concentration and purity of the extracted DNA were assessed using a Nanodrop spectrophotometer (Thermo Fisher Scientific), and the integrity was evaluated by agarose gel electrophoresis. High-quality DNA samples, showing an A260/A280 ratio between 1.8 and 2.0 and intact high molecular weight bands, were selected for subsequent 16S rRNA library preparation. The initial step involved aliquoting 18 µL of the extracted DNA into a 96-well PCR semi-skirted plate, followed by appropriate labeling. The controls were also included as per the plate manifest for the run. This was followed by aliquoting 7 µL.

7.2.1.2 qPCR

The bacterial load was determined from the extracted DNA and sequencing controls that target the 16S rRNA gene which was previously described by Bogaert and colleagues [29].

7.2.1.3 Amplification of 16S rRNA gene

The 16S rRNA gene was amplified using a two-step PCR approach. The 16S rRNA gene was amplified using primers targeting the V4 region of the 16S rRNA gene, as specified in a previously published manuscript [13]. The initial PCR termed the short PCR, consisted of a 7µl template of the DNA and controls. Following the short PCR, a long PCR was performed. The primer sequences used were 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3'). The PCR cycling conditions were as follows: initial denaturation at 95°C for 3 minutes, followed by 25 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds, and extension at 72°C for 30 seconds, with a final extension at 72°C for 5 minutes. Thereafter, the long PCR, which is the second PCR, used the same set of primers (515F and 806R), adapters, barcodes, and nucleotides were added to 5 µl of the PCR product from the first PCR [29, 30].

7.2.1.4 Purification and quantification of the amplicon products

The amplicons were purified using the Agencourt® AMPure® XP beads (Beckman Coulter) following the manufacturer's protocol. The bead-to-amplicon ratio was 0.65: 1 (bead: amplicon) to clean the products [13]. The purified PCR products were quantified using the QuantiFluor® dsDNA System (Promega, Madison, WI, USA) following the manufacturer's protocol. Fluorescence was measured on the GloMax® Discover Microplate reader (Promega, Madison, WI, USA) to determine the amplicon concentration of each sample. As a quality control step, an agarose gel electrophoresis was performed at 110 volts for 90 minutes to ensure the amplification of extracted DNA. The prepared 16S rRNA libraries were pooled in equimolar concentrations of 70 ng as established from the reading and calculations obtained from the GloMax®.

7.2.1.5 Purification of the pooled library and excised agarose gel with Qubit reading

Thereafter, the pooled amplicons were purified using Agencourt® AMPure® XP PCR purification solution at a 1:1 ratio. A total of 3500 ng of the purified pooled library was loaded onto a 1.6 % agarose gel and subjected to electrophoresis under the following conditions: 35 volts for 30 minutes, 40 volts for 45 minutes, 70 volts for 180 minutes, and 50 volts for 60 minutes. After electrophoresis, the excised 16S library was purified using the QIAquick Gel Extraction Kit (QIAGEN, MA, USA) and quantified using the Qubit dsDNA HS Assay Kit. The quality and size distribution of the pooled libraries were verified using an Agilent 2100 Bioanalyzer (Agilent Technologies) with the D1000 Kit (Agilent Technologies, CA, USA) to verify a peak size distribution of approximately 300-500 bp for sequencing. The library was quantified using the Takara Library Quantification Kit (catalogue no. 638324, Takara, USA).

7.2.1.6 Sequencing

The final pooled library was diluted to 9 pM spiked with 20 % PhiX as an internal control. The library-spiked PhiX pool was loaded onto the Illumina MiSeq platform and sequenced using the MiSeq Reagent Kit v3 (600 cycles) Illumina, San Diego, CA, USA with a paired-end 300 bp read configuration.

7.3 DATA ANALYSIS

Raw sequencing data were processed using the MiSeq Reporter software (Illumina) for initial base calling and demultiplexing. Raw read quality checking was done using FastQC v0.12.1 [31] and MultiQC v1.17 [32]. Adapters and low-quality reads were trimmed or filtered using bbduk.sh from the BBMap v38.79 https://github.com/BioInfoTools/BBMap (with parameters ktrim=r qtrim=rl trimq=20 maq=20 minilength=60 maxns=0 k=23 tbo tpe). The trimmed or filtered reads were then aligned to the human genome ftp://ftp.ensembl.org/pub/release100/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz (downloaded on the 20th of September 2024) using bowtie2 v2.5.4. [33] The reads that did not map to the human genome were extracted and further processed by nf-core/ampliseq version 2.9.0 [34] of the nf-core collection of workflows [35] utilizing reproducible software environments from the Bioconda [36] and Biocontainers [37] projects. Sequences were processed sample-wise (independent) with DADA2 [37] to eliminate PhiX contamination, trim reads (before median quality drops below 25), discard reads with > 2 expected errors, correct errors, merge read pairs and remove polymerase chain reaction (PCR) chimeras. Taxonomic classification was performed using DADA2 and the Silva v138 [38] database. ASV sequences, abundance, and DADA2 taxonomic assignments were loaded into R v4.3.0 (2023-04-21 ucrt) [39]. All downstream analysis, that is taxonomy and sample preprocessing (removal of samples with less than 10 reads after quality filtering and ASVs with less than 5% prevalence), rarefication at a depth of the number of reads in the sample with the lowest read count, alpha (Chao1, Observed, Simpson, and Shannon) and beta diversity, taxonomy visualization, and multivariate association testing was done in R with phyloseq [40], vegan [41], ggplot2 [42] microbial [43] and MaAsLin2 [44].

7.3.1 Taxa visualization

7.3.1.1 Bar plots

Bar plots for the relative abundance of bacteria were plotted, using the microbial R-package, where possible, we plotted the top 10 taxonomy.

7.3.1.2 Heatmap visualizations

Although RNA sequence analysis methods (DESeq2 and edgeR) has been useful in normalizing microbiota data, Mcknight et al have shown that proportions and rarefying yield better

comparisons among communities [45]. As a result, to normalize plotting heatmaps, we used rarefying.

7.3.2 Diversity analysis and visualization

7.3.2.1 *α*-diversity

In microbiota analysis, $\alpha-$diversity is a measure of species diversity within a sample. These indices indicate the richness (number of species in the sample, e.g. Observed and Chao1), evenness (the extent of closeness in numbers of species in an environment), dominance (or rarity), and diversity of the species within a sample. Depending on the data and the study objective, it is advisable to make decisions based on measures that consider more than one class of indicators (richness, evenness, rarity, diversity).

Although it is not recommended to use rarefaction curves to estimate the total richness of a sample, or to extrapolate anything from them, they provide some insight into the studied samples depending on the data. For example, they are useful for determining if the sequencing effort is sufficient and if the total diversity within the sample has been captured.

7.3.2.2 Richness and diversity estimate plots

Chao1 accounts for the likeliness of having more undiscovered species in the sample. It is important to note that richness does not consider the abundance of the species types. A sample with an even distribution of species is more diverse than a sample with the same number of species yet one of the species dominates. Both Shannon and InvSimpson indices account for species' abundance, richness, and evenness. Unlike Shannon, Simpson is less sensitive to richness than it is to evenness, hence the use of InvSimpson. We estimated these measures using the microbial package with counts normalized by the *rarefy* function to account for differences in library sizes.

7.3.2.3 *β*-diversity

β-diversity is the measure of diversity between samples. Metrics used include UniFrac which compares samples based on phylogeny information. When the distance between samples has been determined, different ways to visualize them include graphs, networks, and ordination methods (such as principal coordinate analysis: PCoA). Other non-phylogeny $\beta-$diversity measures exist including Bray–Curtis dissimilarity, Jaccard index, and Euclidean distance. Due to the sparsity of

microbial data, the Euclidean distance is not recommended for microbial analysis. Hence, we used Bray-Curtis.

To test if the within-condition variability is greater than the between-condition variability, we use the ANalysis Of SIMilarity (ANOSIM) test (Clarke 1993, Warton *et al*.) [46,47].

7.3.4 Analysis of similarity (ANOSIM)

ANOSIM is a non-parametric test that tests the similarity given there are categorical variables. The results of this test are based on the p-value significance or R2/R values. $R = 0$ means groups/clusters are similar, $R = 1$ means they are very different and $R < 0$ means the between-group differences are smaller than within-group differences.

7.3.5 Permutational multivariate analysis of variance (PERMANOVA)

Here we tested if condition groups were different concerning centroid and dispersion. H0: the different conditions have no significant effect on the species composition ("the centroids of the groups, as defined in the space of the chosen resemblance measure, are equivalent for all groups") vs H1.

7.3.6 Differential expression

We determined features that are up or down-regulated in the cases and control using the DESEQ2 package. In other words, we evaluated whether the abundances of microbial taxa differ significantly between groups of samples based on the grouping factor, which is diseased vs *H. pylori* negative controls in this case.

7.3.7 Biomarker selection

In addition, we implemented classification using a random forest classifier and LEfSe method to find the important differentially expressed bacteria/taxa in the microbial community. We used the Mean Decrease in Accuracy to measure the importance of each bacteria/taxon. The biomarker function does the random forest classification and returns the significant table including the important values. We drew the top 30 important markers, where possible. We performed Linear Discriminant Analysis (LDA) Effect Size effects (LEfSe) which statistically identifies microbial taxa that are differentially abundant across the cases (orange) and controls (green). We used an

LDA threshold for significant biomarkers of 5, with an adjusted p-value (padj) of 0.05 for significant results. It combines standard statistical tests with LDA to emphasize taxa that are not only significantly different between groups but also have consistent biological relevance or effect size.

7.3.8 The relationship between epidemiological and clinical parameters and the gastric microbiome at the genus level

We analyzed the gastric microbiome in relation to the patient's clinical data such as sex differences, race, diet, the presence or absence of gastric ulcers, and the status of the RUT. Using the Inverse Simpson (InvSimpson) and Shannon, we tested for the significance of the microbiome diversity based on clinical data. We also compared diversity to published gastric biome data from other countries.

7.4 RESULTS

Supplemental Figure S7.2 depicts the FastQC per Sequence Quality Score

7.4.1 Exploratory analysis

Data was analyzed according to the following taxonomy ranks: Kingdom, Phylum, Class, Order, Family, Genus, Species. Our domain consists of 1 unique kingdom which is Bacteria. We identified phyla comprising *Campylobacterota, Firmicutes, Actinobacteriota, Proteobacteria, Bacteroidota*, and *Fusobacteriota* and 9 classes (including *Campylobacteria, Bacilli, Actinobacteria, Negativicutes, Clostridia*, *Gammaproteobacteria*). 16 orders were identified (incl. *Campylobacterales, Lactobacillales, Actinomycetales, Veillonellales-Selenomonadales, Micrococcales, Clostridiales*), 25 families (including *Helicobacteraceae, Streptococcaceae, Actinomycetaceae, Veillonellaceae, Micrococcaceae, Clostridiaceae*); 35 genera (including *Helicobacter, Streptococcus, Actinomyces, Veillonella, Rothia, Sarcina);* and 34 species (including *H. pylori, Streptococcus oralis, Streptococcus salivarius, Entamoeba dispar, Rothia mucilaginosa*) were established. No missing values existed at the kingdom, Phylum, Class, and Order levels. On the other hand, there was 1 missing taxon at the family level, 2 at the Genus level, and 36 at the species level.

7.4.2 Taxa

Figure 7.1 depicts the bar plot of relative abundance of bacteria at the genus level. Supplemental Figures S7.3- 7.7 show the heatmap for each sample at phylum, class, order, family, and species levels. Table 7.1 summarizes the most common phyla occurring in both the *H. pylori* positive and *H. pylori* negative cohorts arranged in the order of relative abundance at the phylum level. Supplemental Figure S7.8 depicts the bar relative abundance at the Phylum level in both the *H. pylori* negative and *H. pylori* positive groups.

Firmicutes were the most abundant phylum in the control group while *Campylobacterota* was the most common in the cases. *Bacteriota* and *Fusobacteria*, though occurred sparingly, were relatively common in the *H. pylori* negative group. At the genus level, *Streptococcus* was more abundant in controls than in the cases such as *Rothia* and *Sarcina.*

7.4.3 Diversity analysis and visualization

7.4.3.1 α-diversity

7.4.3.1.1 Richness and diversity estimate plots

Figure 7.2 and Figure 7.3 show the distribution of Observed and Chao1, Shannon, and InvSimpson indices, respectively. Generally, diversity was low; the lowest diversity observed in cases. The plots show that the within-condition variation in sexennial burning samples is smaller than the within-condition variation given another condition (no burning, quadrennial, annual, and biennial plots). On the other hand, within-condition, variation in quadrennial samples was greater than it is in other conditions.

For all conditions, all samples have richness in the range of 1 and 8 ASVs; Chao1 has an ASV range of 1 and 16. Thus, Chao1 was slightly higher in its richness which may suggest that the sequence depth was not sufficient to catch all diversity that is present in the samples. For Shannon, the ASV range is 0, 2.0253262; InvSimpson, ASV range is 1, 7.1428571. Cases have a wider interquartile range (IQR) than controls (ctl), indicating greater variability in the middle 50% of its data. However, cases have a smaller overall range than controls, suggesting that while the central portion of cases is more widespread, the overall extent from minimum to maximum value is more compact. As shown in Figure 7.2, the species richness using both the Observed and Chao1 indices resulted in adjusted p-values (0.57 for Chaol and 0.1 for the Observed), which are both greater

than 0.05. This indicates no statistically significant differences in microbial richness between the cases and controls. As such, we explored other indices, that is, Shannon or InvSimpson.

The comparison of $\alpha-$diversity using the Shannon index between controls (ctl) and cases revealed a statistically significant difference (adjusted p-values of 0.0154) (Figure 7.3.) This suggests that controls have a higher microbial diversity reflecting both a greater number of species and a more even distribution of those species compared to cases.
The analysis of microbial diversity using the InvSimpson index revealed a statistically significant difference (adjusted p-value < 0.05) between controls and cases (Figure 7.3). This suggests that the microbial community in controls is more diverse, characterized by a higher evenness of species distribution compared to cases.

The means of the Observed and Chao1and Shannon and InvSimpson were presented on a bar graph to determine if there were trends in the cases or controls, and we observed no trends. Using pairwise scatterplots and the correlation matrix, we also determined if the α-diversity indices are correlated. From the scatter plot of measures and the correlation matrix, we observed that the Observed measure is highly correlated to Chao1, and Shannon is correlated to all three (InvSimpson, Observed, and Chao1). This confirms that Observed and Chao1 give similar information about the samples. Subsequently, we tested the statistically significant effect of conditions on the $\beta-$ diversity measures in the next section.

7.4.3.1.2 *β*-diversity

Figure 7.4 represents every sample as a dot, which is colored according to their sampling condition (case in red and control in blue). Firstly, this two-dimensional PCOA plot shows $\approx 47\%$ of the total variance between the samples. There are two clusters in our data: that are made of cases and a few controls, which are more diverse than the cases. It was difficult to differentiate the two condition samples in the PCoA using the Bray-Curtis method. Generally, the within-conditioning is greater than the between-condition variability.

7.4.3.1.3 Analysis of similarity (ANOSIM)

From the ANOSIM test results, the p-value = 0.001, suggests that the observed differences between

the groups (cases and controls) are statistically significant. The ANOSIM value 0 < R = $ 0.2951906 < 1 indicates some separation between the cases and controls, though the difference is moderate, not extreme. The upper quantiles of permutations show the distribution of R values from the permutations (under the null hypothesis where there's no real difference between groups). The observed R = 0.2951906 is higher than the 99th percentile of these permutations (0.1285), further reinforcing that the result is significant given the dissimilarity ranks between and within classes: the higher the median rank (50% quantile), the more dissimilar the communities within that group or between groups. The between-group dissimilarity ranks are higher than within-group ranks, supporting the conclusion that the groups differ Supplemental Figure S7.9. In short, based on the Bray-Curtis dissimilarity metric, the community composition of cases was statistically significantly different from the composition of controls.

7.4.3.1.4 Permutational multivariate analysis of variance (PERMANOVA)

Based on the PERMANOVA test, conditions explain 20.2821014 % of the variability and p-value = 0.001, therefore we reject H0 and conclude that the groups have a significant effect on species composition. This confirms the ANOSIM results.

7.4.3.1.5 Differential expression

Based on the adjusted p-value (padj), no taxa show significant differences in their abundances between the cases and controls (that is, 0).

7.4.3.1.6 Biomarker selection

The important differentially expressed bacteria/taxa to the microbial community at the Genus level is shown in Supplemental Figure S7.10. The Linear Discriminant Analysis (LDA) Effect Size effects (LEfSe) is depicted in Figure 7.5 and Supplemental Figure S7.11.

7.4.3.1.7 The relationship between epidemiological and clinical parameters and the gastric microbiome at the genus level

The relative abundance of the gastric microbiome at the Genus level based on epidemiological and clinical determinants in the South African cohorts is shown in Table 7.2. Supplemental Figures S7.12-14 depict bar plots correlations between racial groups, sex gastric ulcer, and the gastric

microbiome in the South African cohort. There was no significant difference in the microbiome diversity between male and female genders with the predominant genera in their order of relative abundance for both sexes being *Streptococcus, Helicobacter, Rothia, Lachnoanaerobaculum,* and *Neisseria.* Concerning race, though there was also no significant difference in the diversity of the microbiome, we noted that while *Helicobacter* was the predominant genus in both Black and Coloured patients, *Streptococcus* was the most common genera in the White population. Additionally, among patients with gastric ulcers, *Sarcina* was the most common organism as compared to *Streptococcus* in patients without ulcers. No significant differences between vegetarian and non-vegetarian gastric microbiomes were noted whilst *Veillonella* and *Neisseria* were more abundant in vegetarians. Table 7.3 provides a comparison of inter-country gastric microbiome composition.

7.5 DISCUSSION

In the context of gastric carcinogenesis, *H. pylori* infection is a critical trigger in the development of GCA. Chronic infection with *H. pylori* can lead to hypochlorhydria (reduced gastric acid secretion) and this in turn allows for the colonization of non-*H. pylor*i bacteria that are typically suppressed by the acidic milieu. The shift in the gastric microbiome composition may result in the proliferation of bacteria that produce carcinogenic byproducts such as nitrosamines. These compounds, alongside chronic inflammation induced by *H. pylori*, exacerbate mucosal damage and genomic instability. Detailed analysis of the gastric microbiota has confirmed that patients with gastric carcinoma exhibit a dysbiotic microbial community with genotoxic potential, which is distinct from that of patients with chronic gastritis. [37]. Gastric microbial community profiling reveals dysbiotic cancer-associated microbiota [37]. So far, only a very small number of studies characterized the human gastric microbiota in health and disease. Major findings suggest that *H. pylori* negative subjects contain a diverse microbiota in their stomach, whereas in *H. pylori* positive patients the gastric mucosa is dominated by certain species [48].

This was confirmed in our South African cohort. In our study *H. pylori* infection was associated with gastric microbiome dysbiosis, with significantly reduced diversity but no significant differences between test and control groups in microbial richness. Thus, microbial communities

appear to share a similar number of species, suggesting that environmental factors did not markedly affect species richness in the respective samples. Although statistical significance was not achieved, it does not imply that the microbial communities are identical. Instead, it suggests that they are relatively similar in terms of richness, which can still have biological implications depending on the ecological context.

7.5.1 α-diversity

The Shannon index summarizes the diversity in the population while assuming all species are represented in a sample and randomly sampled. The Shannon index increases as both the richness and evenness of the community increase. The Simpson index is another indicator, which could also be used to estimate microbial diversity. In this study, the comparison of α−diversity using the Shannon index between controls and cases revealed a statistically significant difference as did the analysis of microbial diversity using the Simpson index. Such differences in diversity may be related to underlying biological or environmental factors or in this case the impact and influence of *H. pylori*. Thus, controls could have a healthier or more stable ecosystem since high diversity is often linked to resilience in biological communities. On the other hand, the lower Shannon index in cases could mean dominance of a few species, which could reflect environmental stress or disease states or more likely, prior receipt of antibiotics and the indirect effect of *H. pylori* on the gastric microbiome composition. Similarly, a higher Simpson index indicates not only a greater number of species but also a more equitable distribution of those species. This can suggest a more stable or resilient community structure, which is often linked to better ecosystem functioning. The lower value of Simpson indicates that the diseased samples community may be dominated by a few species, potentially signaling stress or instability in the ecosystem caused by *H. pylori.*

7.5.2 β-diversity

β-diversity analysis quantifies the similarity or distance between microbiome pairs; on the basis of beta-diversity analysis. Therefore, β- diversity in a gastric microbiome measures the dissimilarity between microbial communities in different environments. It is a key tool in microbiome studies, where it is used to identify differences between groups, such as treatment and control groups. Beta diversity is also known as sample dissimilarity. It is calculated using indices like Bray-Curtis,

which we used and found the community composition of cases to be statistically significantly different from the composition of controls. We confirmed this with the ANOSIM and PERMANOVA statistical tests. Together, these findings confirm gastric microbiome dysbiosis associated with *H. pylori* in infected South African patients.

Using pyrosequencing analysis Jo et al. established that *Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes,* and *Fusobacteria* are the major phyla in both *H. pylori* positive and negative patients [49]. Also, Maldonado-Contreras A et al showed that *Proteobacteria, Spirochetes,* and *Acidoacteria* are more associated with *H. pylori-positive* patients while *Actinobacteria, Bacteroidetes,* and *Firmicutes* were detected at low levels among this cohort [50]. The distribution of phyla differed in our study. In our study, the most abundant phyla included *Firmicutes, Campylobacterota, Actinobacteria, Bacteriota, Proteobacteria*, and *Fusobacteria*. *Firmicutes* were the most abundant phylum in the control group while *Campylobacterota* was the most common in the cases. The reasons and implications of these differences are unknown.

The most prevalent gastric mucosa-associated genera frequently described in scientific literature are *Neisseria, Prevotell, Haemophilus, Fusobacterium, Streptococcus, and Veillonella,* though the abundance of each varies among subjects [51-54]. *Neisseria, Haemophilus, Fusobacterium,* and some species of *Streptococcu*s and *Prevotella* are typical oral bacteria [52,55,56]. It is noteworthy that these prevailing genera within the stomach are rarely detected in the lower gastrointestinal tract, which is primarily characterized by microbial communities dominated by *Ruminococcus, Faecalibacterium*, and *Bacteroides* [52,57], hence the significance of compartmental gastrointestinal microbiome studies. In our study at genus and species levels, *H. pylori* was the predominant organism in the cases, while *Streptococcus* spp was the most abundant in the control group. This is per previous studies that reported *H. pylori,* the predominant gastric bacterium in *H. pylor*i-infected patients [58]. Notably, with regards to the differences in epidemiological parameters in our cohort [i.e sex, race, diet (vegetarian or not), and presence or absence of gastric ulcers], we observed, that while the dominant genus in both the-Black and Coloured patients was *Helicobacte*r, *Streptococcus* was predominant in the White population. We also noted that among patients diagnosed with a gastric ulcer in endoscopy, *Sarcina* which is a member of the *Clostridiacae*, was the most predominant genus. Once again, the implications or significance of these differences are unknown.

While we did not investigate the gastric microbiome of GCA patients, according to Yang et al, GCA-specific microbiota is characterized by a paradoxical decrease in the abundance of *H. pylori* and an increase in other non-*H. pylori* bacteria taxa, pointing to the fact that the pathogenesis of GCA is a complicated interplay between *H pylori* with determinants in an altered gastric microbiome [59]. In that study, they also noted strong exclusion interactions in the gastric microbiota between *H. pylori* and *Fusobacterium, Neisseria, Provetella, Veillonella,* and *Rhothia* which are found only in patients with advanced gastric lesions such as chronic atrophic gastritis, intestinal metaplasia, and dysplasia, but absent in patients with normal and superficial gastritis [59]. Yang et al, also reported that *H. pylori* eradication has a beneficial effect in the restoration of gastric dysbiosis, with the effect more pronounced amongst patients who had successful eradication compared to the cohort with partial eradication, thereby pointing to the role of *H. pylori* in gastric dysbiosis [59]. They also noted a strong co-excluding interaction in nine genera in both CAG, IM, and dysplasia groups, and ~~included~~ *Alloprevotella, Fusobacterium, Neisseria, Porphyromonas, Prevotella, Rothia* and *Veillonella* [60].

Notably, in our study, strong co-excluding interaction was also observed. Whilst at Phylum level *Campylobacterota* dominated amongst *H. pylori* cases, a consequential reduction in the abundance in *Firmicutues, Actinobacteriota, Bactriodetes*, *Proteobacteria*, and *Fusobacteria* compared to the *H. pylori* negative cohort, was documented. The same findings were noted in the studies from America, Japan, and the UK. Similarly, at the genus level, among our cohort of patients who were *H. pylori* positive, the relative abundance of *Lachnonaerobaculum, Actinomyces*, *Veillonela, Provetella, Haemophilus, Neisseria, Lactobacillus* among others, were significantly reduced and, in some cases, rarely seen, compared to the *H. pylori* negative cohort.

7.6 CONCLUSIONS

Our results suggest a slight alteration in the microbial diversity in the presence of *H. pylori*, although there is no statistical significance in taxa differentiation between the *H. pylori* negative and positive cohorts. While the predominant phyla in the *H. pylori*-negative cohort were *Firmicutes*, *Campylobacterota* was the dominant phyla in the *H. pylori* positive cohort. There was also no significant difference in the taxa distribution between the South African cohorts and those

from the Southeast Asian and American populations. Furthermore, whilst no significant differences were observed relating to males vs females, vegetarian vs non-vegetarians, the dominant genus in both the Black and Coloured patients and White patients in our study was *Helicobacter* and *Streptococcus*, respectively.

The reasons and implications of these differences are unknown and therefore future studies should include multi-omic analysis beyond the microbiome specifically transcriptomics (RNA expression levels), proteomics (the entire set of proteins expressed), metabolomics (small molecules and metabolites), and epigenomics (epigenetic modifications). An integrative approach that combines data from multiple biological "omics" layers may provide a comprehensive understanding of biological systems, disease mechanisms, or treatment responses related to *H. pylori* infections as pivotal enablers for GCA. By leveraging various omics datasets, the complex interactions among genes, proteins, metabolites, and other biomolecules can be explored to uncover deeper insights into *H. pylori* infections.

7.7 REFERENCES

1. Hooi JK, Lai WY, Ng WK, Suen MM, Underwood FE, Tanyingoh D, et al. Global prevalence of *Helicobacter pylori* infection: systematic review and meta-analysis. Gastroenterology. 2017;153(2):420-9.

2. Zilberstein B, Quintanilha AG, Santos MA, Pajecki D, Moura EG, Alves PRA, et al. Digestive tract microbiota in healthy volunteers. Clinics. 2007;62(1):47-56.

3. Sung J, Kim N, Kim J, Jo HJ, Park JH, Nam RH, et al. Comparison of gastric microbiota between gastric juice and mucosa by next generation sequencing method. Journal of Cancer Prevention. 2016;21(1):60.

4. Tsuda A, Suda W, Morita H, Takanashi K, Takagi A, Koga Y, et al. Influence of proton-pump inhibitors on the luminal microbiota in the gastrointestinal tract. Clinical and translational gastroenterology. 2015;6(6):e89.

5. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. Science translational medicine. 2016;8(343):343ra82-ra82.

6. Haro C, Rangel-Zúñiga OA, Alcalá-Díaz JF, Gómez-Delgado F, Pérez-Martínez P, Delgado-Lista J, et al. Intestinal microbiota is influenced by gender and body mass index. PloS one. 2016;11(5):e0154090.

7. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. Genome medicine. 2016;8:1-11.

8. Yang I, Woltemate S, Piazuelo MB, Bravo LE, Yepez MC, Romero-Gallo J, et al. Different gastric microbiota compositions in two human populations with high and low GCA risk in Colombia. Scientific reports. 2016;6(1):18594.

9. Nardone G, Compare D, Rocco A. A microbiota-centric view of diseases of the upper gastrointestinal tract. The lancet Gastroenterology & hepatology. 2017;2(4):298-312.

10. Dong Q, Xin Y, Wang L, Meng X, Yu X, Lu L, et al. Characterization of gastric microbiota in twins. Current microbiology. 2017;74:224-9.

11. Noto JM, Peek Jr RM. The gastric microbiome, its interaction with *Helicobacter pylori*, and its potential role in the progression to stomach cancer. PLoS pathogens. 2017;13(10):e1006573.

12. Lofgren JL, Whary MT, Ge Z, Muthupalani S, Taylor NS, Mobley M, et al. Lack of commensal flora in *Helicobacter pylori*–infected INS-GAS mice reduces gastritis and delays intraepithelial neoplasia. Gastroent erology. 2011;140(1):210-20. e4.
13. Bik EM, Eckburg PB, Gill SR, Nelson KE, Purdom EA, Francois F, et al. Molecular analysis of the bacterial microbiota in the human stomach. Proceedings of the National Academy of Sciences. 2006;103(3):732-7.
14. Andersson AF, Lindberg M, Jakobsson H, Bäckhed F, Nyrén P, Engstrand L. Comparative analysis of human gut microbiota by barcoded pyrosequencing. PloS one. 2008;3(7):e2836.
15. Dicksved J, Lindberg M, Rosenquist M, Enroth H, Jansson JK, Engstrand L. Molecular characterization of the stomach microbiota in patients with GCA and in controls. Journal of medical microbiology. 2009;58(4):509-16.
16. Riley DR, Sieber KB, Robinson KM, White JR, Ganesan A, Nourbakhsh S, et al. Bacteria-human somatic cell lateral gene transfer is enriched in cancer samples. PLoS computational biology. 2013;9(6):e1003107.
17. Eun CS, Kim BK, Han DS, Kim SY, Kim KM, Choi BY, et al. Differences in gastric mucosal microbiota profiling in patients with chronic gastritis, intestinal metaplasia, and GCA using pyrosequencing methods. Helicobacter. 2014;19(6):407-16.
18. Precious Mathebela, B.Damane, T. Mulaudzi, Zilungile Lynette Mkhize-Khwitshana, G. Gaudji, Zodwa Dlamini. "Influence of the Microbiome Metagenomics and Epigenomics on Gastric Cancer." International Journal of Molecular Sciences (2022). 10.3390/ijms232213750.
19. A. Almeda, S. Grimes, Hojoon Lee, S. Greer, G. Shin, M. McNamara, A. C. Hooker, M. M. Arce, M. Kubit, Marie C Schauer, P. Van Hummelen, Cindy Ma, M. Mills, Robert J. Huang, Joo Ha Hwang, M. Amieva, Summer S. Han, J. Ford, Hanlee P. Ji. "The Gastric Cancer Registry: A Genomic Translational Resource for Multidisciplinary Research in Gastric Cancer." Cancer Epidemiology, Biomarkers and Prevention (2022). 10.1158/1055-9965.EPI-22-0308.
20. Y. Sunakawa, R. Matoba, T. Takayama, N. Okumura, N. Sugimoto, K. Miura, S. Yuki, Y. Shindo, H. Kuramochi, Sho Sato, T. Ogura, S. Kinami, N. Nagao, S. Ueda, E. Inoue, H. Kawakami, K. Muro, T. Nakajima, W. Ichikawa, M. Fujii. "Host-related biomarkers including gut microbiome to predict toxicities of nivolumab in advanced gastric cancer: DELIVER trial (JACCRO GC-08).." Journal of Clinical Oncology (2022). 10.1200/jco.2022.40.4_suppl.245.

21. Audrey Rivire, Marija Selak, David Lantin, Frdric Leroy, Luc De Vuyst. "Bifidobacteria and Butyrate-Producing Colon Bacteria: Importance and Strategies for Their Stimulation in the Human Gut." Frontiers Media (2016). https://doi.org/10.3389/fmicb.2016.00979.
22. Sarah M Trujillo, E. McKenney, G. Hilderbrand, Lindsey S. Mangipane, M. Rogers, K. Joly, D. Gustine, Joy A. Erlenbach, B. Mangipane, Diana J. R. Lafferty. "Intrinsic and extrinsic factors influence on an omnivores gut microbiome." PLoS ONE (2022). 10.1371/journal.pone.0266698.
23. M. Bartsch, Andreas Hahn, S. Berkemeyer. "Bridging the Gap from Enterotypes to Personalized Dietary Recommendations: A Metabolomics Perspective on Microbiome Research." Metabolites (2023). 10.3390/metabo13121182.
24. Lesley A. Houghton, Margaret Heitkemper, Michael D. Crowell, Anton Emmanuel, Albena Halpert, James A. McRoberts, Brenda B. Toner. "Age, Gender, and Womens Health and the Patient." Elsevier BV (2016). https://doi.org/10.1053/j.gastro.2016.02.017.
25. K. Maki, G. Wallen, T. Bastiaanssen, LiYueh Hsu, Michael E. Valencia, Vijay A Ramchandani,Melanie L Schwandt, N. Diazgranados, J. Cryan, R. Momenan, Jennifer J Barb. "The gut-brain axis in individuals with alcohol use disorder: Anexploratory study of associations among clinical symptoms, brain morphometry, and the gut microbiome." None (2024). 10.1111/acer.15346.
26. B. Bicknell, Anita Saltmarche, Orla Hares, Geoffrey Herkes, A. Liebert. "Parkinsons disease and the Interaction of Photobiomodulation, the Microbiome, and Antibiotics: A Case Series A Case Series." Medical Research Archives. 10.18103/mra.v12i1.4929.
27. Walsh J, Griffin BT, Clarke G, Hyland NP. Drug–gut microbiota interactions: implications for neuropharmacology. British journal of pharmacology. 2018 Dec;175(24):4415-29.
28. A. Szymczak, Katarzyna Gembara, J. Majewska, P. Miernikiewicz, Stanisaw Ferenc, J. Gnus, M. Kak, Natalia Jdruchniewicz, Wojciech Witkiewicz, K. Dbrowska. "Triadic Dynamics of Gastric Bacterial Microbiome, Phageome, and Host Genotype, with Implications for Disease Associations." bioRxiv (2024). 10.1101/2024.06.16.598898
29. Bogaert D, Keijser B, Huse S, Rossen J, Veenhoven R, van Gils E, et al. Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. PLoS One. 2011;6(2):e17035.
29. Claassen-Weitz S, Gardner-Lubbe S, Nicol P, Botha G, Mounaud S, Shankar J, et al. HIV-exposure, early life feeding practices and delivery mode impacts on faecal bacterial profiles in a

South African birth cohort. Sci Rep. 2018;8(1):5078.
30. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A. 2011;108 Suppl 1(Suppl 1):4516-22.
31. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010.
32. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-8.
33. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012 Apr;9(4):357-9.
34. Straub D, Blackwell N, Langarica-Fuentes A, Peltzer A, Nahnsen S, Kleindienst S. Interpretations of environmental microbial community studies are biased by the selected 16S rRNA (gene) amplicon sequencing pipeline. Frontiers in Microbiology. 2020;11:550420.
35. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nature Biotechnology. 2020;38(3):276-8.
36. Grüning B, Dale R, Sjödin A, Chapman BA, Rowe J, Tomkins-Tinch CH, et al. Bioconda: sustainable and comprehensive software distribution for the life sciences. Nat Methods. 2018;15(7):475-6.
37. da Veiga Leprevost F, Grüning BA, Alves Aflitos S, Röst HL, Uszkoreit J, Barsnes H, et al. BioContainers: an open-source and community-driven framework for software standardization. Bioinformatics. 2017;33(16):2580-2.
38. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013 Jan;41(Database issue):D590-6. doi: 10.1093/nar/gks1219. Epub 2012 Nov 28. PMID: 23193283; PMCID: PMC3531112.
39. https://cran.r-project.org/src/base/R-4/R-4.3.0.tar.gz
40. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. Paul J. McMurdie and Susan Holmes (2013) PLoS ONE 8(4):e61217
41. Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter

Braak C, Weedon J (2024). _vegan: Community Ecology Package_. R package version 2.6-8, https://CRAN.R-project.org/package=vegan

42. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

43. Guo K, Gao P (2024). microbial: Do 16s Data Analysis and Generate Figures_. R package version 0.0.21, https://CRAN.R-project.org/package=microbial.

44. Mallick H et al. (2021). Multivariable Association in Population-scale Meta-omics Studies, http://huttenhower.sph.harvard.edu/maaslin2. To cite the MaAsLin 2 software, please use: Mallick H, Rahnavard A, McIver LJ (2020). MaAsLin 2: Multivariable Association in Population-scaleMeta-omics Studies. R/Bioconductor package, http://huttenhower.sph.harvard.edu/maaslin2

45. McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, Zenger KR. Methods for normalizing microbiome data: an ecological perspective. Methods in Ecology and Evolution. 2019;10(3):389-400

46. Clarke, K. R. (1993). Non-parametric multivariate analysis of changes in community structure. Australian Journal of Ecology 18, 117–143.

47. Warton, D.I., Wright, T.W., Wang, Y. 2012. Distance-based multivariate analyses confound location and dispersion effects. Methods in Ecology and Evolution, 3, 89–101.

48. Ferreira RM, Pereira-Marques J, Pinto-Ribeiro I, Costa JL, Carneiro F, Machado JC, Figueiredo C. Gastric microbial community profiling reveals a dysbiotic cancer-associated microbiota. Gut. 2018 Feb 1;67(2):226-36.

49. Jo HJ, Kim J, Kim N, Park JH, Nam RH, Seok YJ, et al. Analysis of gastric microbiota by pyrosequencing: minor role of bacteria other than *Helicobacter pylori* in the gastric carcinogenesis. Helicobacter. 2016;21(5):364-74.

50. Maldonado-Contreras A, Goldfarb KC, Godoy-Vitorino F, Karaoz U, Contreras M, Blaser MJ, Brodie EL, Dominguez-Bello MG. Structure of the human gastric bacterial community in relation to *Helicobacter pylori* status. The ISME journal. 2011 Apr;5(4):574-9.

51. Li X-X, Wong GL-H, To K-F, Wong VW-S, Lai LH, Chow DK-L, et al. Bacterial microbiota profiling in gastritis without *Helicobacter pylori* infection or non-steroidal anti-inflammatory drug use. PloS one. 2009;4(11):e7985.

52. Kashiwagi S, Naito Y, Inoue R, Takagi T, Nakano T, Inada Y, et al. Mucosa-associated microbiota in the gastrointestinal tract of healthy Japanese subjects. Digestion. 2020;101(2):107-20.

53. Ndegwa N, Ploner A, Andersson AF, Zagai U, Andreasson A, Vieth M, et al. Gastric microbiota in a low–*Helicobacter pylori* prevalence general population and their associations with gastric lesions. Clinical and Translational Gastroenterology. 2020;11(7):e00191.
54. Monstein H-J, Tiveljung A, Kraft C, Borch K, Jonasson J. Profiling of bacterial flora in gastric biopsies from patients with *Helicobacter pylori*-associated gastritis and histologically normal control individuals by temperature gradient gel electrophoresis and 16S rDNA sequence analysis. Journal of medical microbiology. 2000;49(9):817-22.
55. Castaño-Rodríguez N, Goh K-L, Fock KM, Mitchell HM, Kaakoush NO. Dysbiosis of the microbiome in gastric carcinogenesis. Scientific reports. 2017;7(1):15957.
56. Sedghi L, DiMassa V, Harrington A, Lynch SV, Kapila YL. The oral microbiome: Role of key organisms and complex networks in oral health and disease. Periodontology 2000. 2021;87(1):107-31.
57. Vasapolli R, Schütte K, Schulz C, Vital M, Schomburg D, Pieper DH, et al. Analysis of transcriptionally active bacteria throughout the gastrointestinal tract of healthy individuals. Gastroenterology. 2019;157(4):1081-92. e3.
58. Yu G, Torres J, Hu N, Medrano-Guzman R, Herrera-Goepfert R, Humphrys MS, et al. Molecular Characterization of the Human Stomach Microbiota in GCA Patients. Frontiers in Cellular and Infection Microbiology. 2017;7.
59. Yang J, Zhou X, Liu X, Ling Z, Ji F. Role of the gastric microbiome in GCA: from carcinogenesis to treatment. Frontiers in microbiology. 2021;12:641322.
60. Bik EM, Eckburg PB, Gill SR, Nelson KE, Purdom EA, Francois F, et al. Molecular analysis of the bacterial microbiota in the human stomach. Proceedings of the National Academy of Sciences. 2006;103(3):732-7.

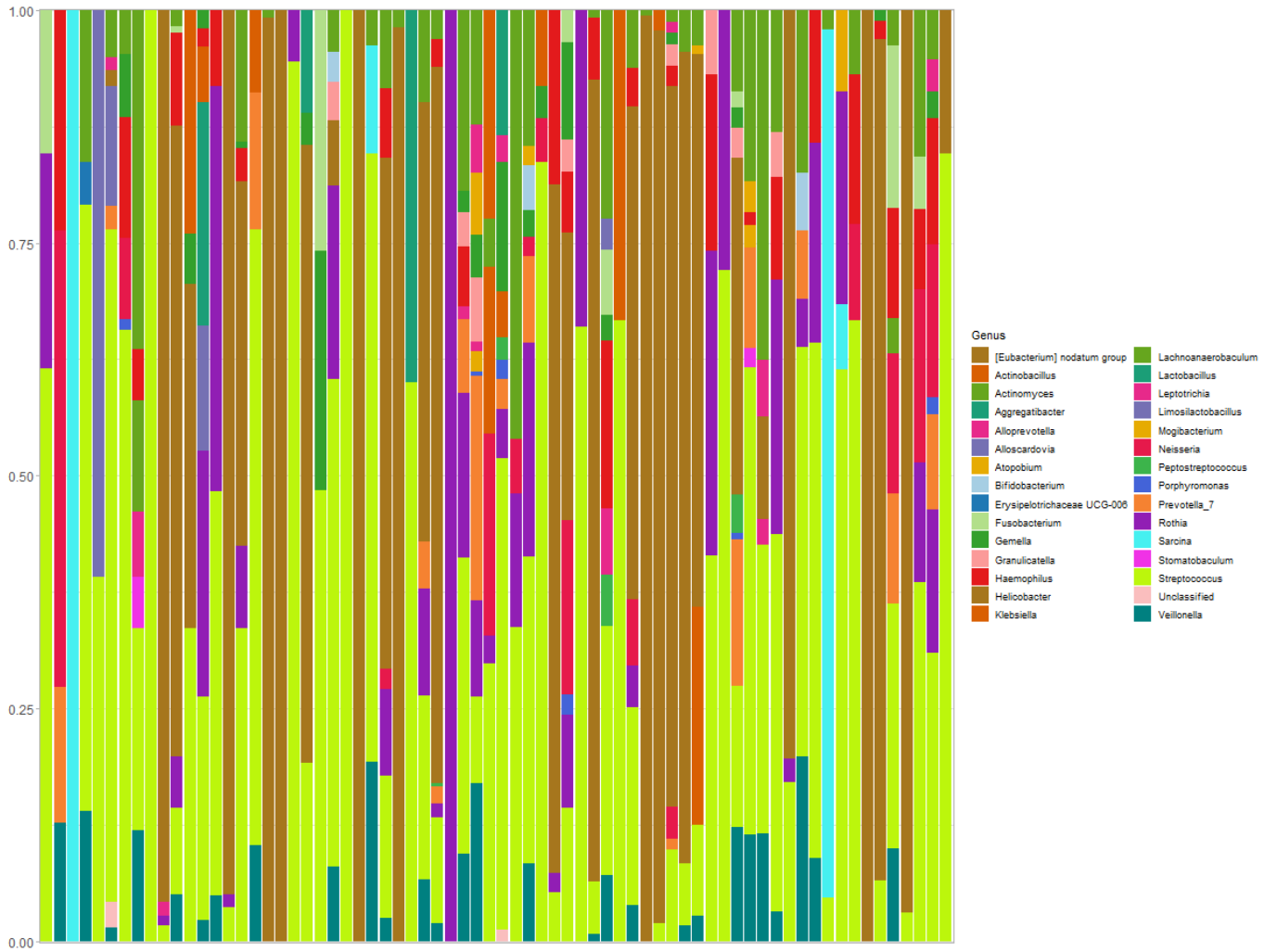


Figure 7.1: Bar plot of relative abundance of bacteria at the genus level

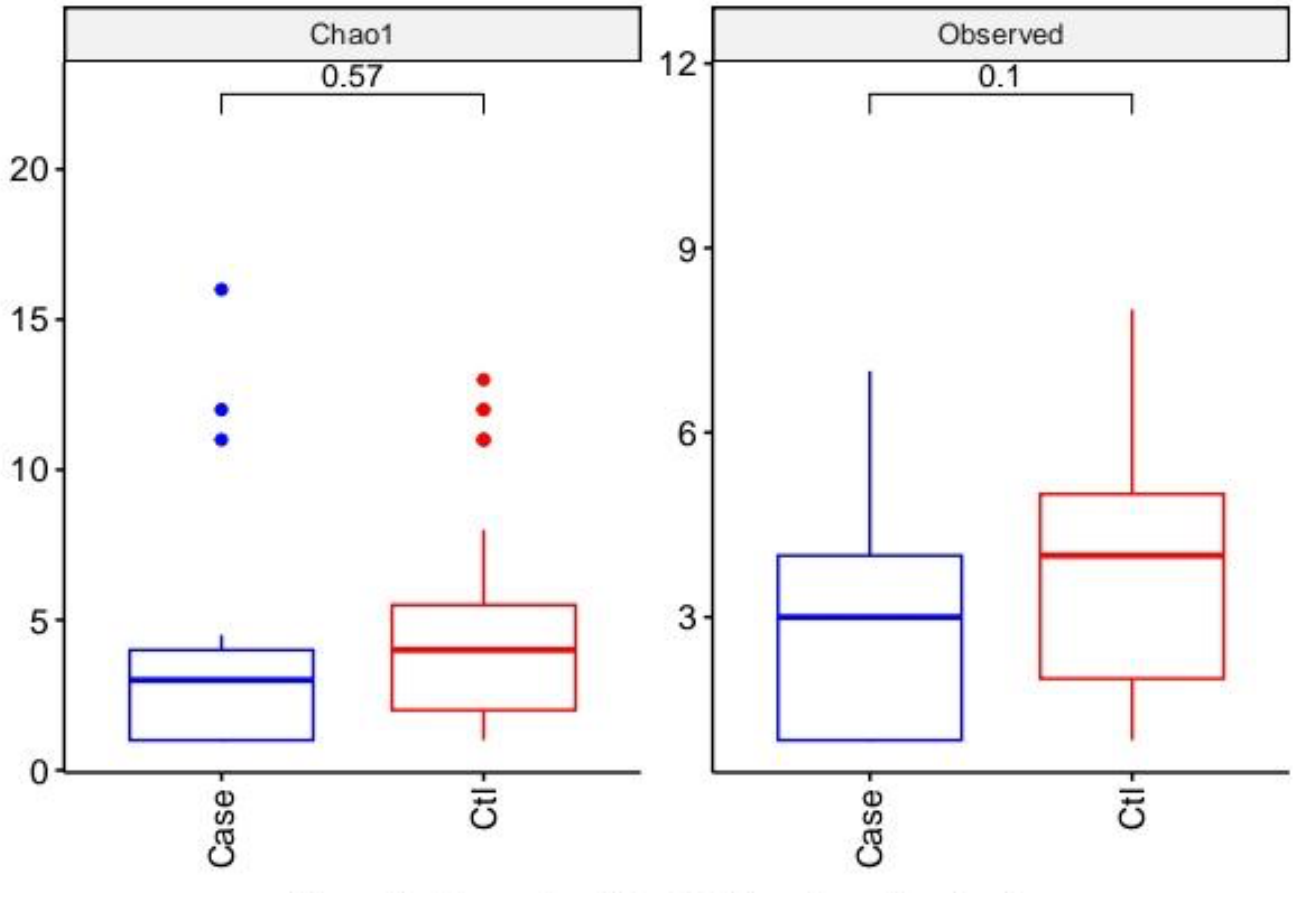


Figure 7.2: Observed and Chao1 indices at the condition level, between cases and controls

(NB: Ctl- control)

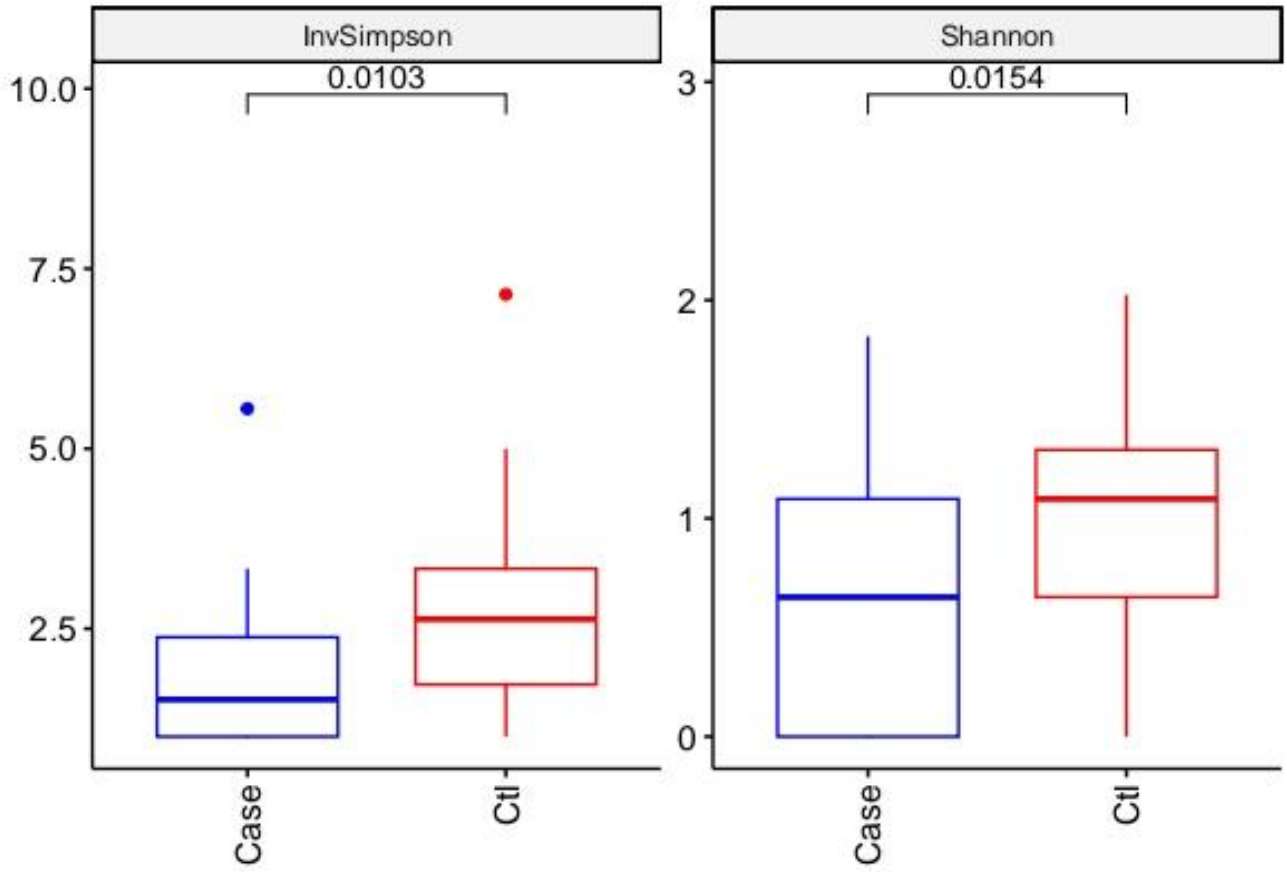


Figure 7.3: Shannon and Inverse Simpson indices at condition-level between cases and controls

(NB:Ctl-control)

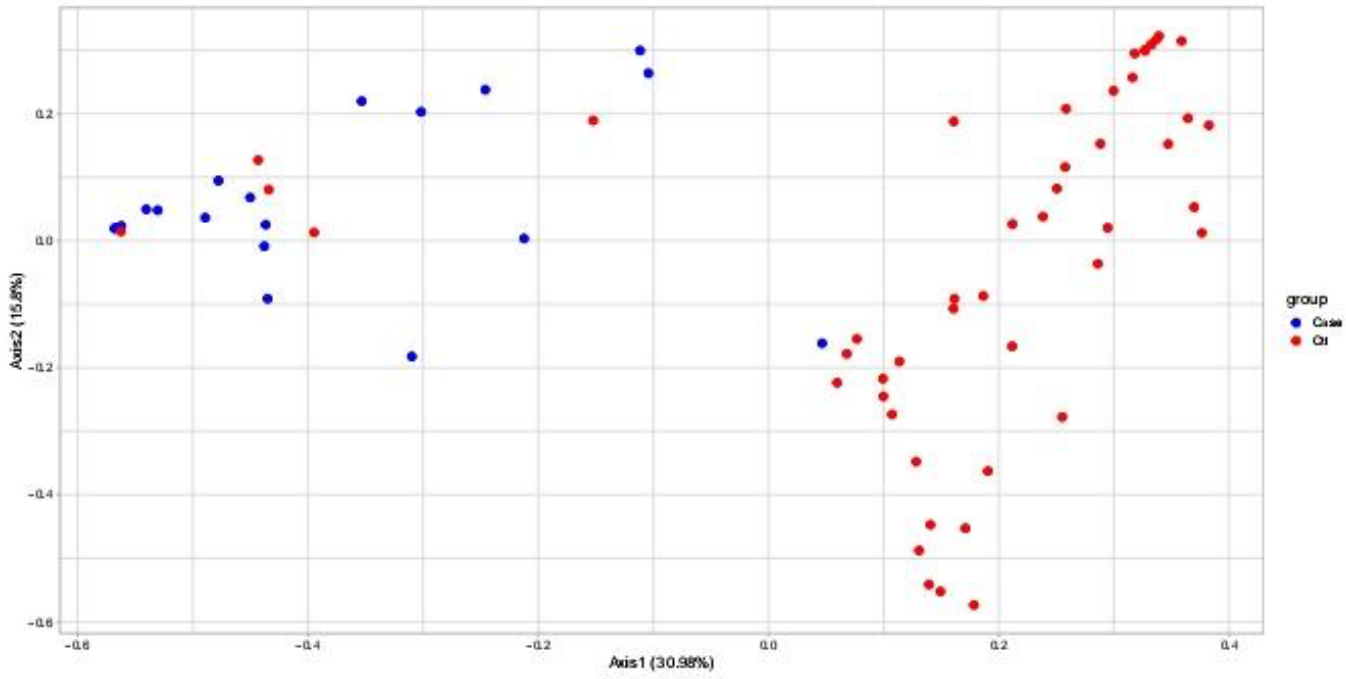


Figure 7.4: The two-dimensional Principal Coordinate Analysis (PCoA) using the Bray distance, showing approximately 47% of the total variance between the cases (blue) and control (red)- The controls are more diverse compared to the cases.

(NB: Ctl- control)

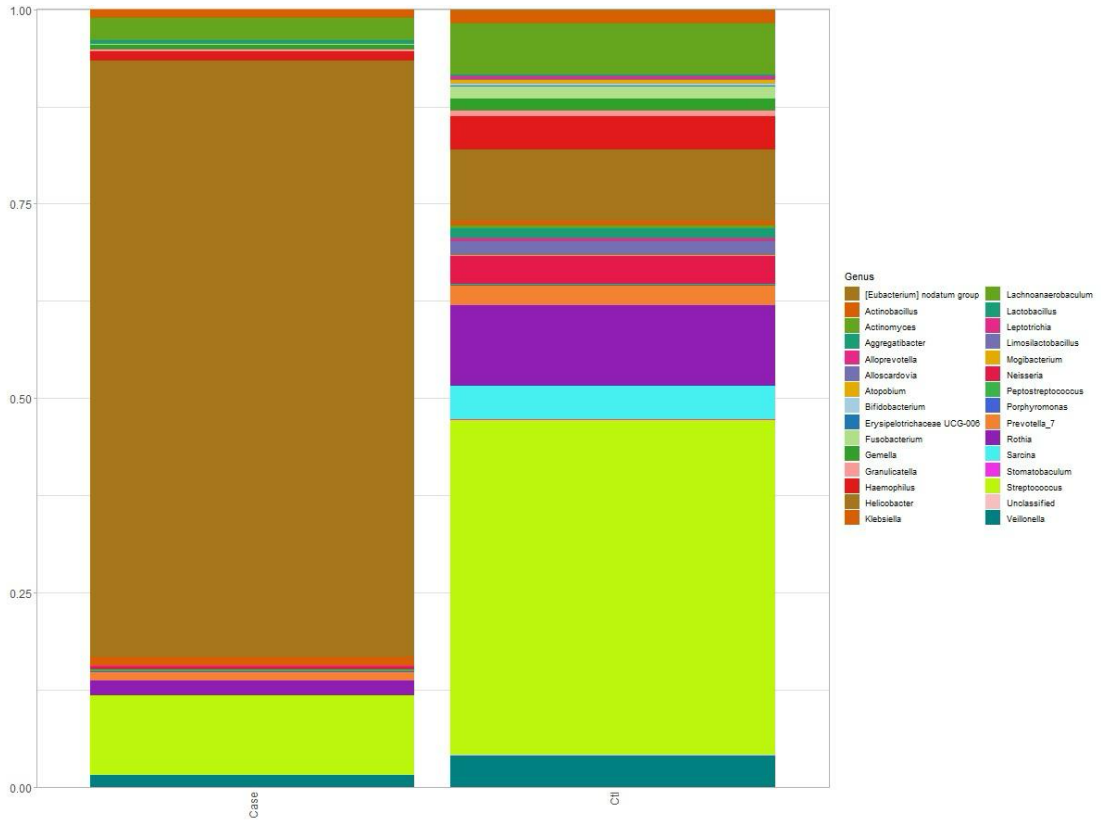


Figure 7.5: Taxa differential abundance between cases and controls based on the Linear Discriminant Analysis Effect Size effects (LEfSe) at the genus level

(NB: Ctl- control)

Table 7.1: The most common phyla occurring in both the *H. pylori* positive and *H. pylori* negative cohorts arranged in the order of relative abundance at the phylum level

***H. pylori* positive cohort**	***H. pylori* negative cohort level**
Campylobacterota (most common)	*Firmicutes (most common)*
Firmicutes	*Actinobacteriota*
Actinobacteriota	*Proteobacteria*
Bacteriodes	*Campylobacterota*
Proteobacteria	*Bacteriodata*
Fusobacteria	*Fusobacteria*

Table 7.2: Relative abundance of the gastric biome at the Genus level based on epidemiological and clinical determinants in the South African cohorts (both *H. pylori* positive and *H. pylori* negative)

No	Clinical parameter	The order of predominant genera	Shannon	Simpson	Statistical difference
1	**Sex**	**Male 33 (41.77%)**: *Streptococcus, Helicobacter, Rothia, Lachnoanaerobaculum, Neisseria* **Female 46 (58.23%)**: *Streptococcus, Helicobacter, Rothia, Lachnoanaerobaculum, Neisseria*	0.461	0.345	NS
2	**Race**	**Black**: **14 (17.72%)** *Helicobacter, Streptococcus, Lachnoanaerobaculum, Veillonella* **Coloured**: **62 (78.48)** *Helicobacter, Streptococcus, Rothia, Sarcina*	0.71	0.822	NS

		White: **2 (2.52)** *Streptococcus,Rothia, Lachnoanerobaculum, Veillonella*			
3	**Diet**	**Vegetarian**: **6(7.59)** *Streptococcus, Helicobacter, Rothia, Veillonella, Neisseria* **Non-vegetarian**: **72 (91.14)** *Streptococcus, Helicobacter,Rothia, Lachnoanerobaculum*,	0.783	0.818	NS
4	**Ulcer**	**Gastric Ulcer**: **7 (8.86)** *Sarcina, Helicobacter, Streptococcus, Veillonella, Neiserria* **No Gastric ulcer**: **44 (55.70)** *Streptococcus, Helicobacter, Rothia, Lachnoanaerobaculum, Veillonella*	0.485	0.283	NS
5	**Rapid Urease Test**	**Positive: 23 (29.11)** *Helicobacter, Streptococcus, Rothia, Lachnoanerobaculum, Sarcina* **Negative**: **56 (70.89)** *Streptococcus, Helicobacter, Lachnoanerobaculum, Rothia, Veillonella*	0.212	0.308	NS

NS: Not statistically significant

Table 7.3: Comparison of inter-country gastric microbiome composition

Country of Study	Gastric sample type	Subjects	Methods	Predominant taxonomic structure	
				Phyla	**Genus**
South Africa	Gastric biopsy	23 *H. pylori* positive adults (our study cohort)	16SrRNA sequencies	*Campylobacterota Firmicutes Proteobacteria Actinobacteria*	*Helicobacter Streptococcus Veilonela, Lachnoanaerobaculum*
South Africa	Gastric Biopsy	56 *H. pylori* negative control adults (our study cohort)	16SrRNA	Phyla (Control)-*Firmicutes, Bacteriodata, Actinobacteria, Proteobacteria, Fuso*bacteria	Genus: *Streptococcus, Veillonela, Rothia, Sarcina, Neisseria, Haemophilus, Lachnoaerobaculum, Helicobacter*
USA	Gastric mucosa	23 *H. pylori* negative control adults	16SrRNA sequences	*Proteobacteria, Firmicutes, Bacteroidetes, Actinobacteria, Fusobacteria*	*Genera: Streptococcus, Prevotella, Neisseria, Haemophilus, Porphyromonas* [50]
Hong Kong China	Gastric Mucosa	5 *H. pylori* negative control adults and 5 non-	16S rRNA sequencing	Phyla: *Proteobacteria, Bacteroidetes , Firmicutes, Actinobacteria, Fusobacteria* [42]	

		NSAID gastritis			
Japan	Gastric mucosa	7 *H. pylori* negative controls adults	16S rRNA sequencing	Phyla*: Proteobacteria, Bacteroidetes, Firmicutes,Fusobacteria, Actinobacteria*	*Genera:Neisseria, Prevotella,* f*; [paraprevotellaceae]-*g*;[Prevotella], Haemophilus, Fusobacterium, Streptococcus, Veillonella, Capnocytophage, Leptotrichia* [(42)]
Germany	Gastric mucosa	21 *H. pylori* negative controls	16S rRNA sequencing	Phyla: *Firmicutes, Proteobacteria, Bacteroidetes, Fusobacteria, Actinobacteria*	*Genera: Streptococcus, Prevotella, Pseudomonas, Fusobacterium, Gemella, Neisseria, Veillonell*[(47)]
Sweden	Gastric mucosa	*H. pylori* negative controls:171	16S rRNA	Phyla: *Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria, Fusobacteria*	Genera: *Streptococcus, Prevotella, Veillonella, Fusobacterium, Gemella, Neisseria, Haemophilus*[(43)]

		Non-atrophic *H. pylori* gastritis:33 AG:12 antral chemical gastritis:61			

CHAPTER 8
SUMMARY OF OVERALL STUDY RESULTS
8.1INTRODUCTION

Helicobacter pylori (*H. pylori*) is a type 1 carcinogen associated with the development of GCAs. It is also implicated in other gastroduodenal diseases such as gastric ulcers, duodenal ulcers, and gastric lymphoma. *H. pylori* has infected over 50% of the global population, with the highest burden observed in the African continent. However, reported data on GCA in Africa shows a paradoxical relationship with *H. pylori* incidence, a phenomenon known as the 'African enigma'. This study aimed to investigate the validity of this so-called 'African enigma'.
We hypothesize that in Africa:

1. Patients infected with *H. pylori* may be protected against more aggressive disease phenotypes i.e. The 'African enigma' is true.
2. Clinical risk factors and histological activity in *H. pylori*-infected individuals will provide clues to risk factors for developing GCA.
3. The gastric microbiome of patients infected with *H. pylori* exhibits dysbiosis with less diversity similar to the few other gastric microbiome studies reported before.

This study comprised three components: a systematic review, a retrospective cohort study, and a cross-sectional study. The summary of results for each sub-study is as follows:

8.2 Systematic review

We conducted a systematic review to explore the epidemiological profile of GCA in Africa, including incidence, prevalence, and associated risk factors. Notably, the review revealed significant weaknesses in the reported data, such as missing background sample/population sizes, making prevalence and incidence estimates unreliable. Despite these limitations, the review indicated that antral cancers accounted for 52% of cases, suggesting a link between *H. pylori* infection and GCA in Africa. The findings emphasize the urgent need for reliable national registry data, accurate data collection during endoscopy and histology, and robust hospital and population data. Future studies should include well-designed case-control studies to assess the

contributions of *H. pylori* and other risk factors to GCA in African cohorts. Collaboration among African countries, including unified protocols, is essential to maximize the utility of these studies.

8.3 Retrospective study

This study examined the epidemiological profile and incidence of GCA at Groote Schuur Hospital. The findings revealed that GCA was more common in males and individuals aged ≥60 years. Key diagnostic red flags included weight loss and anaemia. The annual incidence of GCA was 2.98 per 100 000, consistent with previous reports supporting the 'African enigma' may be true. Non-cardia GCAs were the most prevalent subtype. The low prevalence of *H. pylori* in histological samples suggested prior eradication, though chronic atrophic gastritis remained high, sustaining the risk of non-cardia GCA. Interestingly, proton pump inhibitors (PPIs) were found to be significantly protective against antral cancer, highlighting the importance of treating *H. pylori* before CAG develops.

8.4 Cross-sectional study

This study had two arms:

1. **Molecular signature of *H. pylori***

   Molecular analyses revealed findings that may explain the 'African enigma'. The African cohort showed a lower prevalence of *CagPAI* morphologic variants, previously associated with GCA development. Similarly, the oncogenic *VirB/D* genes, which potentiate the activity of *CagA/VacA*, were less common in the African cohort compared to Southeast Asian populations. Phylogenetic analysis suggested that African *H. pylori* strains are similar to American strains, both of which exhibit lower virulence features and correspondingly lower GCA incidence compared to Southeast Asia.

2. **Gastric microbiome analysis**

   This arm confirmed that *H. pylori* infection significantly reduced gastric microbiome biodiversity. While the predominant phylum in the *H. pylori* negative cohort was *Firmicutes, Campylobacterota* dominated the *H. pylori* positive cohort. Taxa distribution showed no significant differences between South African, Southeast

Asian, and American populations. Among Black and Coloured patients, *H. pylori* was the dominant genus, while *Streptococcus* was dominant in White patients. The reasons and implications for these differences remain unclear.

8.5 LIMITATIONS

Several limitations were identified in this project:

1. The systematic review suffered from poor-quality data in the published articles, which precluded a comprehensive meta-analysis, as originally intended.
2. Genetic susceptibility, environmental, and lifestyle factors were not evaluated, limiting the scope of the findings.
3. The retrospective study faced issues with missing data and potential referral bias.
4. Unique sequence types (STs) identified in African *H. pylori* strains (e.g., ST1320, 3051, 3583) require further investigation to understand their origins, genetic diversity, and potential role in disease dynamics. These unique STs may indicate region-specific evolutionary pressures, distinct ecological niches, or host-pathogen interactions unique to the African context.
5. Microbiome analysis did not account for confounding factors such as diet and drug use and was not powered for the confounders that can modify the microbiome.
6. The subtypes of vacA and cagA alleles were not analyzed in this current pilot study. These, however, were not the primary aims of the study.

8.6 FUTURE WORK

Future studies should expand beyond the microbiome to include multi-omic analyses, such as transcriptomics, proteomics, metabolomics, and epigenomics, to provide a more comprehensive understanding of *H. pylori*-related GCA mechanisms. An integrative approach that combines data from multiple biological "omics" layers may provide a comprehensive understanding of biological systems, disease mechanisms, or treatment responses related to *H. pylori* infections as pivotal enablers for GCA. By leveraging various omics data sets, complex interactions among genes, proteins, metabolites, and other biomolecules can be explored to uncover deeper insights into *H. pylori* infections.

In addition, the roles of pro-inflammatory and anti-inflammatory cytokines in *H. pylori* positive and *H. pylori* negative cohorts should be explored. Investigating genetic mutations and their contribution to GCA pathogenesis in the context of *H. pylori* infection is another crucial avenue.

SUPPLEMENTARY FIGURES AND TABLES

Biopsies - *Helicobacter pylori*

1.A. Culture x3 plates (2x middle and 1x streaked plate)

A.1. Date of collection of specimen (DD/MMM/YYYY) __ __ / __ __ __ / __ __ __ __

A.2. Date Performed of specimen (DD/MMM/YYYY) __ __ / __ __ __ / __ __ __ __

A.3. Growth: ☐ No Growth: ☐

ATCC growth:

Day 2

Day 6

Gram stain:

***Oxidase*:**

Negative Positive

***Urease*:**

Negative Positive

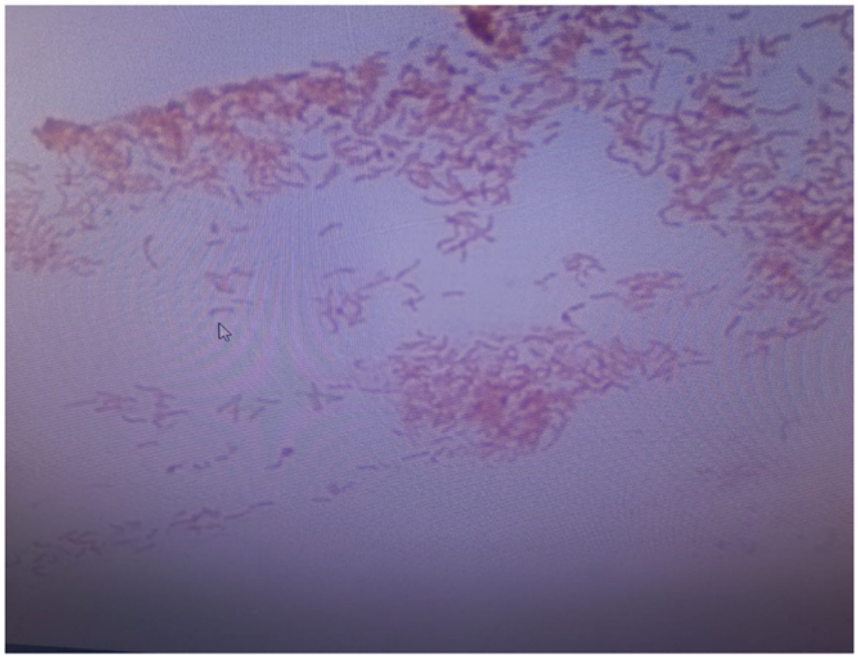
***Catalase*:**

Negative Positive

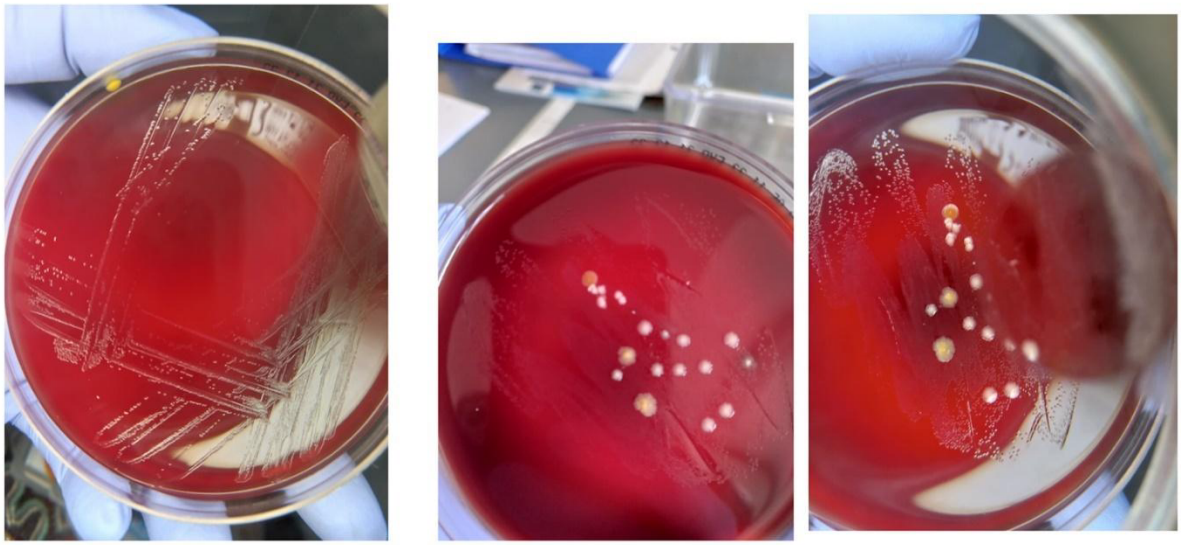
PCR Result:

Date: (__ __ / __ __ __ / __ __ __ __) QC: (__ __ / __ __

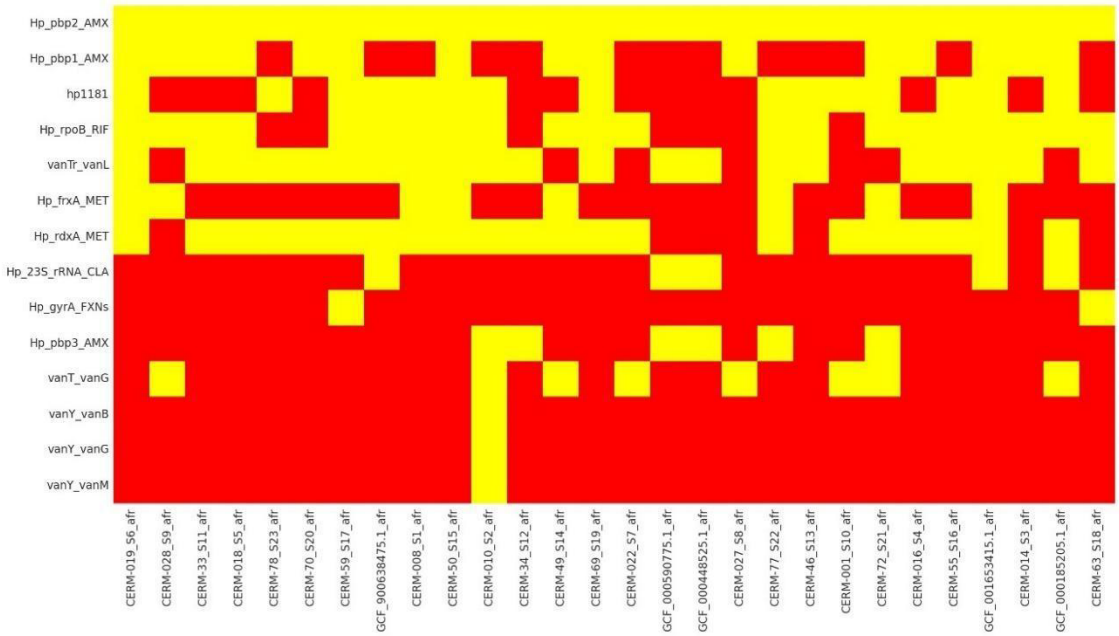
Supplementary Figure S6.1: The sample culture data form



Supplementary Figure S6.2: Microscopic view of *Helicobacter pylori*



Supplementary Figure S6.3: Colonies of *H. pylori* in the culture



Supplementary Figure S6.4: Heatmap showing the presence/absence of AMR genes in the African samples

(23 of our lab study samples, 3 South African samples, 1 Gambian and 1 Nigerian publicly available samples)

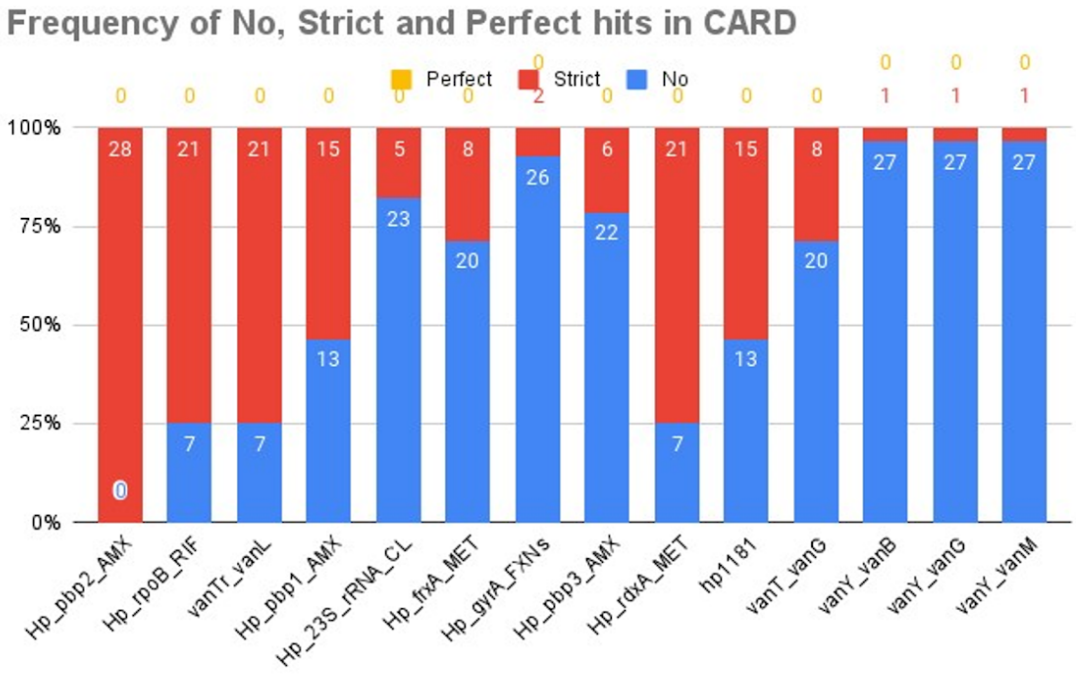
NB: In the legend Green: if seen would mean perfect hits to the CARD database

Red: Strict hits

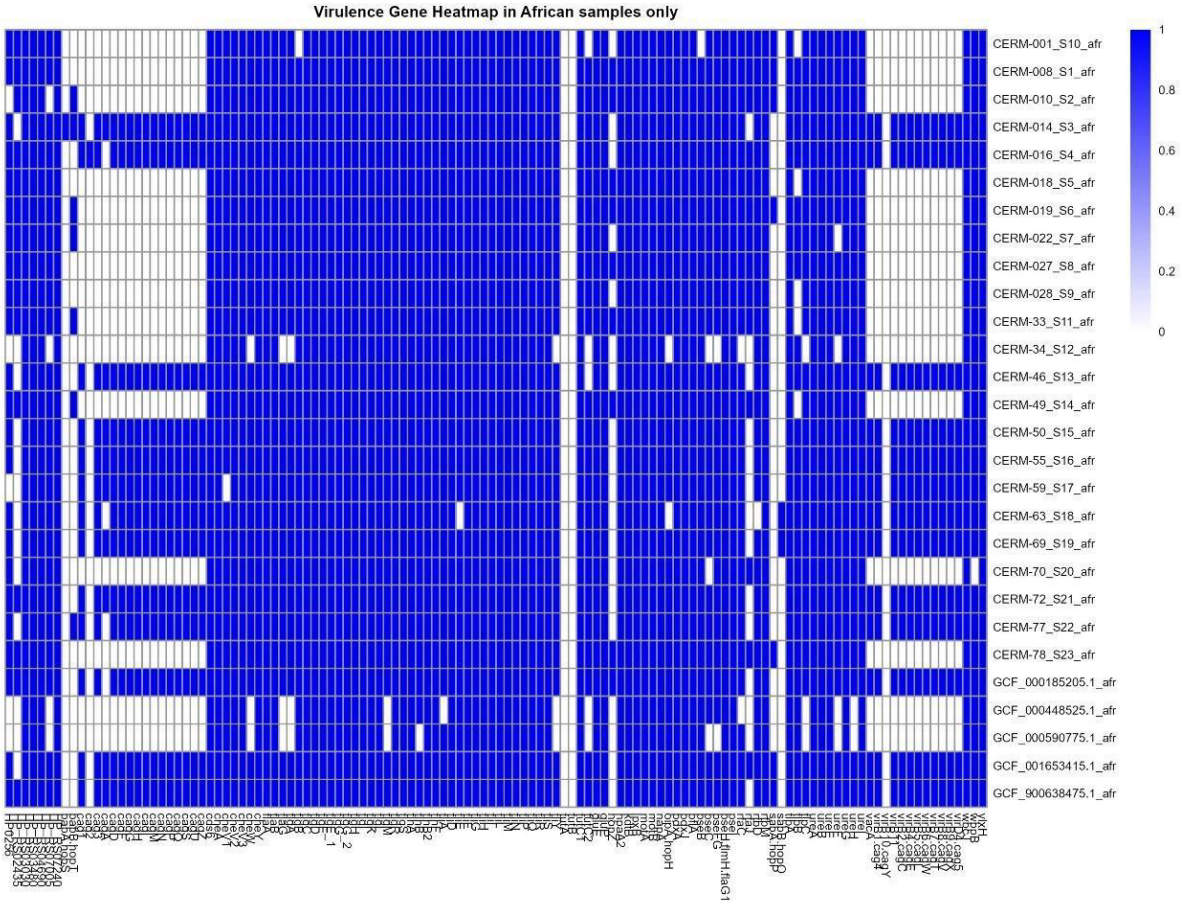
Yellow: No matching hits found in the CARD database

Gene	No	Strict	Perfect
Hp_pbp2_AMX	0	28	0
Hp_rpoB_RIF	7	21	0
VanTr_vanL	7	21	0
Hp_pbp1_AMX	13	15	0
Hp_23S_rRNA_CLA	23	5	0

Hp_frxA_MET	20	8	0
Hp_gyrA_FXNs	26	2	0
Hp_pbp3_AMX	22	6	0
Hp_rdxA_MET	7	21	0
hp1181	13	15	0
vanT_vanG	20	8	0
vanY_vanB	27	1	0
vanY_vanG	27	1	0
vanY_vanM	27	1	0



Supplementary Figure S6.5. A table and Bar plot showing the frequency and degree of hits for the AMR genes in the African strains (n-28). 2 represents perfect hit, 1 represents partial hit and 0 represents no hit



Supplementary Figure S6.6. Virulence gene heatmaps for African strains

Supplementary Table S6.1. Distribution and regional occurrence rates of respective AMR genes

S/No	Resistance Gene	Region	Total sample	Occurrences	Percentage
1	*Pbp1*	America/Europe	100	57	57%
		Africa	28	15	53.6%
		East Asia	119	116	97.5%
2	*pbp2*	Africa	28	28	100%
		America	90	89	98.9%
		European	10	10	100%
		Asia	119	118	99.2%
6	*pbp3*	Africa	28	6	21,4%
		America	90	1	1.1%
		European (UK)	10	0	0%
		Asia	119	27	22.7%
4	*Hp1181*	Africa	28	15	53.6%
		America	90	78	86.7%
		European (UK)	10	8	80%
		Asia	119	118	99.2%
5	*gyrA*	Africa	28	2	7.1%
		America	90	23	25.6%
		European (UK)	10	0	0%
		Asia	119	30	25.2%
6	*rdxA*	Africa	28	21	75%
		America/UK	100	83	83%
		Asia	119	114	25.2%
7	*frxA*	Africa	28	8	28.6%

		America European Asia	90 10 119	49 7 30	55% 70% 25.2%
8	23SrRNA	America Africa European Asia	90 28 10 119	89 5 10 117	98.9% 17.9% 100% 98.3%
9	*Van Tr gene* in Van L	Africa America/UK Asia	28 100 119	21 64 72	75% 64% 60.5%
10	*VanT* gene in Van G	Africa America European (UK) Asia	28 90 10 119	8 29 3 46	28.6% 32.2% 30% 38.7%
11	*TEM-116*	America Only	90	2	2.2%

Supplementary Table S6. 2. Virulence factors associated with *H. pylori* and their functions

GENE	PRODUCT
tlpC	(*tlpC*) membrane-bound chemoreceptor [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
futC1	(*futC1*) alpha-(12)-fucosyltransferase [Lewis antigen (VF0057) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
futC2	(*futC2*) alpha-(12)-fucosyltransferase [Lewis antigen (VF0057) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
tlpA	(*tlpA*) membrane-bound chemoreceptor sensing arginine and bicarbonate [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
tlpB	(*tlpB*) membrane-bound chemoreceptor sensing pH and autoinducer-2 [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flaB	(*flaB*) flagellin B FlaB [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
gluP	(*gluP*) glucose/galactose transporter [LPS (VF0056) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
flgE_1	(*flgE_1*) flagellar hook protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
lpxB	(*lpxB*) lipid A disaccharide synthetase [LPS (VF0056) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
pseB	(*pseB*) UDP-N-acetylglucosamine 46-dehydratase [Flagellar glycosylation/Pse biosynthetic pathway (VF0608) - Motility (VFC0204)] [Helicobacter pylori 26695]

HP_RS07005	(*HP_RS07005*) alpha-12/4 Glc transferase [LPS (VF0056) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
fliQ	(*fliQ*) flagellar biosynthesis protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
fliI	(*fliI*) flagellum-specific ATP synthase FliI [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flgK	(*flgK*) flagellar hook-associated protein 1 FlgK [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flgM	(*flgM*) negative regulator of flagellin synthesis [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
HP_RS04690	(*HP_RS04690*) zinc ribbon domain-containing protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
HP_RS03480	(*HP_RS03480*) RNA polymerase factor sigma-54 [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
HP_RS07240	(*HP_RS07240*) META domain-containing protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
kdtB	(*kdtB*) lipopolysaccharide core biosynthesis protein [LPS (VF0056) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
flgA	(*flgA*) flagellar basal body P-ring biosynthesis protein FlgA [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
pflA	(*pflA*) paralysed flagella protein (pflA) [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
babB/hopT	(*babB/hopT*) outer membrane protein adhesin [BabA (VF0053) - Adherence (VFC0001)] [Helicobacter pylori 26695]

fliY	(*fliY*) flagellar motor switch protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
fliM	(*fliM*) flagellar motor switch protein FliM [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
fliA	(*fliA*) flagellar biosynthesis sigma factor FliA [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
ylxH	(*ylxH*) ATP-binding protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flhF	(*flhF*) flagellar biosynthesis protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flhA	(*flhA*) flagellar biosynthesis protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
cheY	(*cheY*) chemotaxis response regulator CheY [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flgG	(*flgG*) flagellar basal-body rod protein FlgG [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
fliH	(*fliH*) flagellar assembly protein H [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
fliG	(*fliG*) flagellar motor switch protein G [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
fliF	(*fliF*) flagellar M-ring protein FliF [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]

pseH/flmH/flaG1	(*pseH/flmH/flaG1*) UDP-4-amino-46-dideoxy-N-acetyl-beta-L-altrosamine N-acetyltransferase; Pseudaminic acid biosynthesis protein H [Flagellar glycosylation/Pse biosynthetic pathway (VF0608) - Motility (VFC0204)] [Helicobacter pylori 26695]
pseFG	(*pseFG*) CMP-N-acetylneuraminic acid synthetase [Flagellar glycosylation/Pse biosynthetic pathway (VF0608) - Motility (VFC0204)] [Helicobacter pylori 26695]
flgH	(*flgH*) flagellar L-ring protein precursor FlgH [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flgL	(*flgL*) flagellar hook-associated protein 3 FlgL [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
rfaC	(*rfaC*) lipopolysaccharide heptosyltransferase-1 [LPS (VF0056) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
HP0256	(*HP0256*) involved in motility and cell envelope architecture [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flgI	(*flgI*) flagellar P-ring protein precursor FlgI [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flgS	(*flgS*) signal-transducing protein histidine kinase [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
napA	(*napA*) neutrophil activating protein NapA [HP-NAP (VF0052) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
ureA	(*ureA*) urease alpha subunit UreA [Urease (VF0050) - Stress survival (VFC0282)] [Helicobacter pylori 26695]
ureB	(*ureB*) urease beta subunit UreB urea amidohydrolase [Urease (VF0050) - Stress survival (VFC0282)] [Helicobacter pylori 26695]

ureI	(*ureI*) acid-activated urea channel protein [Urease (VF0050) - Stress survival (VFC0282)] [Helicobacter pylori 26695]
ureE	(*ureE*) urease accessory protein (ureE) metallochaperone [Urease (VF0050) - Stress survival (VFC0282)] [Helicobacter pylori 26695]
ureF	(*ureF*) urease accessory protein (ureF) [Urease (VF0050) - Stress survival (VFC0282)] [Helicobacter pylori 26695]
ureG	(*ureG*) urease accessory protein (ureG) [Urease (VF0050) - Stress survival (VFC0282)] [Helicobacter pylori 26695]
ureH	(*ureH*) urease accessory protein (ureH) [Urease (VF0050) - Stress survival (VFC0282)] [Helicobacter pylori 26695]
wbcJ	(*wbcJ*) GDP fucose synthase [LPS (VF0056) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
rfbD	(*rfbD*) GDP-D-mannose dehydratase [LPS (VF0056) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
rfbM	(*rfbM*) mannose-6-phosphate isomerase [LPS (VF0056) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
cheV1	(*cheV1*) chemotaxis coupling protein CheV1 [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flgG_2	(*flgG_2*) flagellar basal-body rod protein (flgG) [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
pdxA	(*pdxA*) 4-hydroxythreonine-4-phosphate dehydrogenase [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
pdxJ	(*pdxJ*) pyridoxine 5'-phosphate synthase [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]

flhB2	(*flhB2*) ABC transporter putative [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flgB	(*flgB*) flagellar basal body rod protein FlgB [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flgC	(*flgC*) flagellar basal-body rod protein FlgC [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
fliE	(*fliE*) flagellar hook-basal body complex protein FliE [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flgR	(*flgR*) response regulator [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
fliP	(*fliP*) flagellar biosynthetic protein FliP [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
wbpB	(*wbpB*) lipopolysaccharide biosynthesis protein [LPS (VF0056) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
oipA/hopH	(*oipA/hopH*) outer inflammatory protein A adhesin [OipA (VF0266) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
HP_RS03030	(*HP_RS03030*) chemotaxis protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flaA	(*flaA*) flagellin A FlaA [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
cheV3	(*cheV3*) chemotaxis coupling protein CheV3 [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
fliN	(*fliN*) flagellar motor switch protein FliN [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]

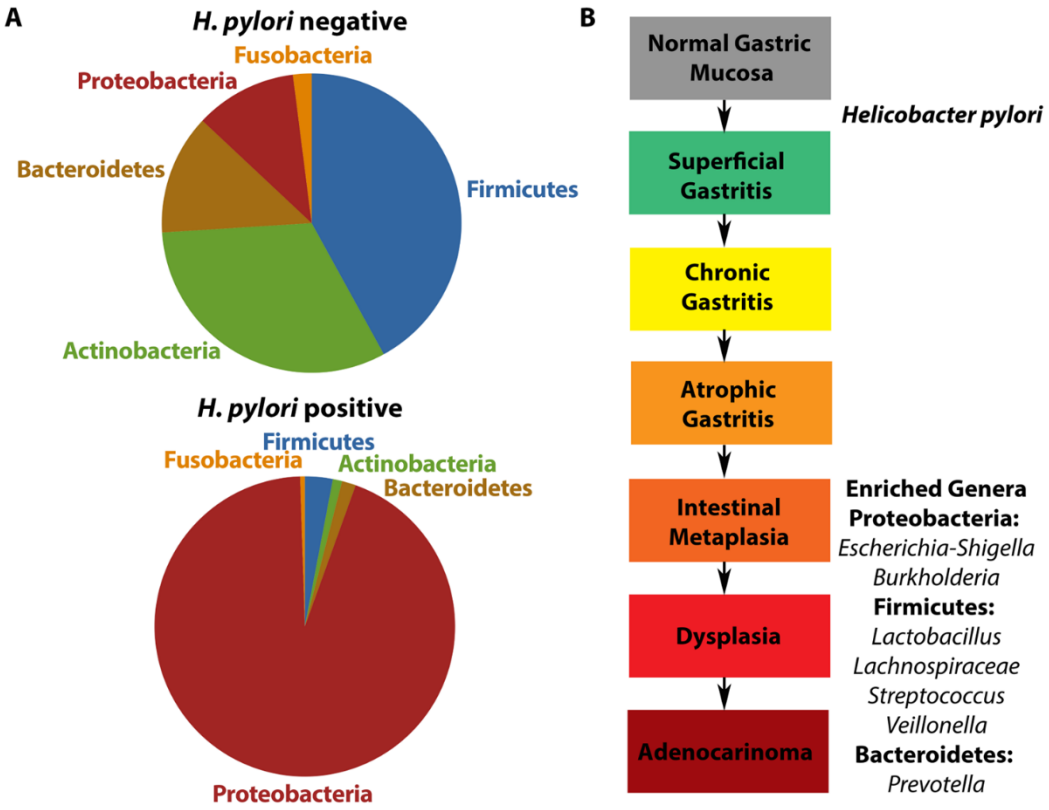
cagA	(*cagA*) cag T4SS translocated effector CagA [CagA (VF0059) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
virB2/cagC	(*virB2/cagC*) type IV secretion system protein Cag25/CagC VirB2 homolog [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cagD	(*cagD*) type IV secretion system protein Cag24/CagD [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
virB4/cagE	(*virB4/cagE*) type IV secretion system protein Cag23/CagE VirB4 homolog [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cagF	(*cagF*) type IV secretion system protein Cag22/CagF chaperone-like protein for CagA [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cagG	(*cagG*) type IV secretion system protein Cag21/CagG [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cagH	(*cagH*) type IV secretion system protein Cag20/CagH [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cagI	(*cagI*) type IV secretion system protein Cag19/CagI [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
virB5/cagL	(*virB5/cagL*) type IV secretion system protein Cag18/CagL VirB5 homolog [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cagN	(*cagN*) type IV secretion system protein Cag17/CagN membrane-associated protein [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]

cagM	(*cagM*) type IV secretion system protein Cag16/CagM transmembrane channel protein [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cagP	(*cagP*) type IV secretion system protein Cag15/CagP [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cagQ	(*cagQ*) type IV secretion system protein Cag14/CagQ [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cagS	(*cagS*) type IV secretion system protein Cag13/CagS [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
virB7/cagT	(*virB7/cagT*) type IV secretion system protein Cag12/CagT VirB7 homolog [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cagU	(*cagU*) type IV secretion system protein Cag11/CagU [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
virB8/cagV	(*virB8/cagV*) type IV secretion system protein Cag10/CagV VirB8 homolog [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
virB6/cagW	(*virB6/cagW*) type IV secretion system protein Cag9/CagW VirB6 homolog [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
virB9/cagX	(*virB9/cagX*) type IV secretion system protein Cag8/CagX VirB9 homolog [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cagZ	(*cagZ)* type IV secretion system protein Cag6/CagZ [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]

virB11	(*virB11*) type IV secretion system ATPase VirB11 homolog [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
virD4/cag5	(*virD4/cag5*) type IV secretion system protein Cag5 VirD4 homolog [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
virB1/cag4	(*virB1/cag4*) type IV secretion system protein Cag4 VirB1 homolog [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cag3	(*cag3*) type IV secretion system protein Cag3 component outer membrane subcomplex [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cag1	(*cag1*) type IV secretion system protein Cag1 [Type IV secretion system (VF0060) - Effector delivery system (VFC0086)] [Helicobacter pylori 26695]
cds6	(*cds6*) LD-carboxypeptidase [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flaG	(*flaG*) a negative regulator of flagellar assembly [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
fliD	(*fliD*) flagellar capping protein FliD [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
fliS	(*fliS*) flagellar protein FliS [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flhB	(*flhB*) flagellar biosynthesis protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
fliL	(*fliL*) flagellar basal body protein FliL [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]

motA	(*motA*) flagellar motor protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
motB	(*motB*) flagellar motor protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
gluE	(*gluE*) UDP-glucose 4-epimerase [LPS (VF0056) - Immune modulation (VFC0258)] [Helicobacter pylori 26695]
pseC	(*pseC)* flagellar modification protein aminotransferase PseC [Flagellar glycosylation/Pse biosynthetic pathway (VF0608) - Motility (VFC0204)] [Helicobacter pylori 26695]
cheW	(*cheW*) chemotaxis coupling protein CheW [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
cheA	(*cheA*) histidine kinase CheA [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
cheV2	(*cheV2*) chemotaxis coupling protein CheV2 [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
hpaA2	(*hpaA2*) ahesin protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
pseI	(*pseI*) pseudaminic acid synthase [Flagellar glycosylation/Pse biosynthetic pathway (VF0608) - Motility (VFC0204)] [Helicobacter pylori 26695]
fliR	(*fliR*) flagellar biosynthetic protein FliR [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flgE	(*flgE)* flagellar hook protein [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]
flgD	(*flgD*) flagellar basal-body rod modification protein FlgD [Flagella (VF0051) - Motility (VFC0204)] [Helicobacter pylori 26695]

babA/hopS	(*babA/hopS*) outer membrane protein adhesin [BabA (VF0053) - Adherence (VFC0001)] [Helicobacter pylori 26695]
vacA	(*vacA*) vacuolating cytotoxin [VacA (VF0058) - Exotoxin (VFC0235)] [Helicobacter pylori 26695]



Supplemental Figure S7.1. Alterations in the gastric microbiome following interactions with *H. pylori.* (Adapted from reference 11)

Supplementary Table S6.3 Multilocus sequence typing (MLST) of *H. pylori*

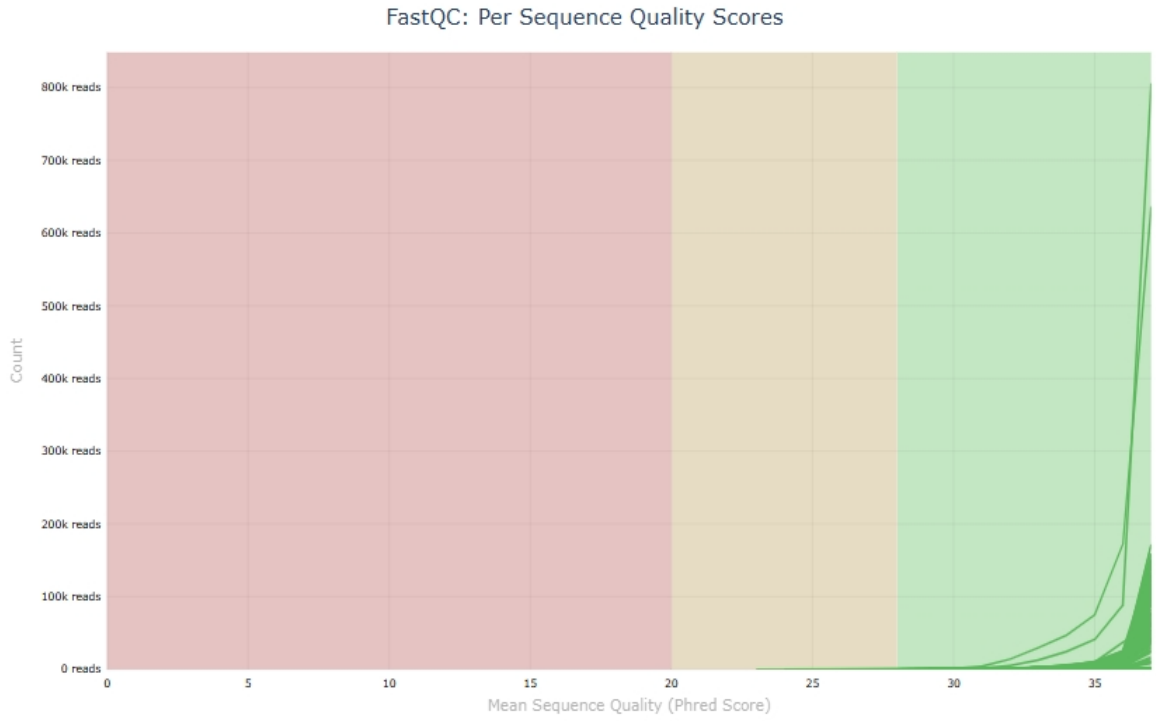
Sample ID	ST	*atpA*	*efp*	*mutY*	*Ppa*	*trpC*	*urel*	*yphC*
CERM-69_S19_afr	1320	atpA(966)	efp(938)	mutY(283)	ppa(952)	trpC(980)	ureI(985)	yphC(320)
GCF_000185205.1_afr	3051	atpA(1820)	efp(1723)	mutY(1774)	ppa(1783)	trpC(2665)	ureI(211)	yphC(2693)
GCF_000448525.1_afr	3583	atpA(2836)	efp(2652)	mutY(2881)	ppa(2713)	trpC(2938)	ureI(2938)	yphC(2975)
GCF_000590775.1_afr	3582	atpA(2835)	efp(2651)	mutY(2880)	ppa(2712)	trpC(2937)	ureI(2937)	yphC(2974)
GCF_000689015.1_usa	204	atpA(204)	efp(204)	mutY(204)	ppa(204)	trpC(204)	ureI(204)	yphC(204)
GCF_000689035.1_usa	204	atpA(204)	efp(204)	mutY(204)	ppa(204)	trpC(204)	ureI(204)	yphC(204)
GCF_000689055.1_usa	204	atpA(204)	efp(204)	mutY(204)	ppa(204)	trpC(204)	ureI(204)	yphC(204)
GCF_000824865.1_uk	594	atpA(433)	efp(434)	mutY(432)	ppa(427)	trpC(433)	ureI(438)	yphC(439)

GCF_000824885.1_uk	595	atpA(434)	efp(435)	mutY(433)	ppa(427)	trpC(434)	ureI(438)	yphC(440)
GCF_000824925.1_uk	594	atpA(433)	efp(434)	mutY(432)	ppa(427)	trpC(433)	ureI(438)	yphC(439)
GCF_000826985.1_usa	181	atpA(181)	efp(181)	mutY(181)	ppa(181)	trpC(181)	ureI(181)	yphC(181)
GCF_000827025.1_usa	181	atpA(181)	efp(181)	mutY(181)	ppa(181)	trpC(181)	ureI(181)	yphC(181)
GCF_000966875.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_001653395.1_skeas	45	atpA(45)	efp(45)	mutY(45)	ppa(45)	trpC(45)	ureI(45)	yphC(45)
GCF_001653415.1_afr	340	atpA(340)	efp(340)	mutY(340)	ppa(340)	trpC(340)	ureI(340)	yphC(340)
GCF_001653435.1_usa	1047	atpA(528)	efp(522)	mutY(522)	ppa(515)	trpC(526)	ureI(530)	yphC(532)
GCF_002206465.1_uk	181	atpA(181)	efp(181)	mutY(181)	ppa(181)	trpC(181)	ureI(181)	yphC(181)
GCF_002952235.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)

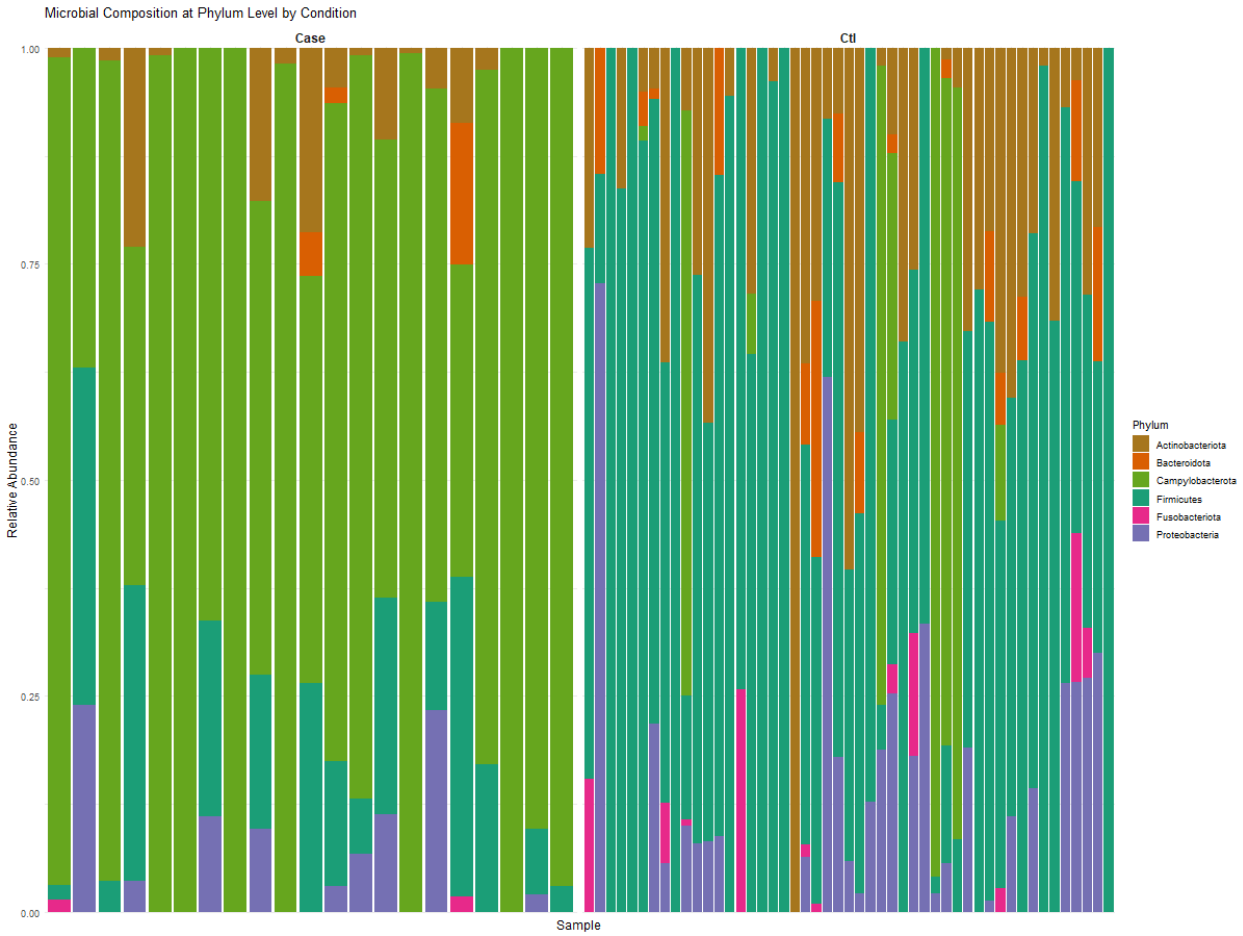
GCF_002952255.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952275.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952335.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952355.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952375.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952395.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952415.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952435.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952455.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952475.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)

GCF_002952495.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952515.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952535.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952555.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952575.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952595.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952615.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952635.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952655.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_002952675.1_usa	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)

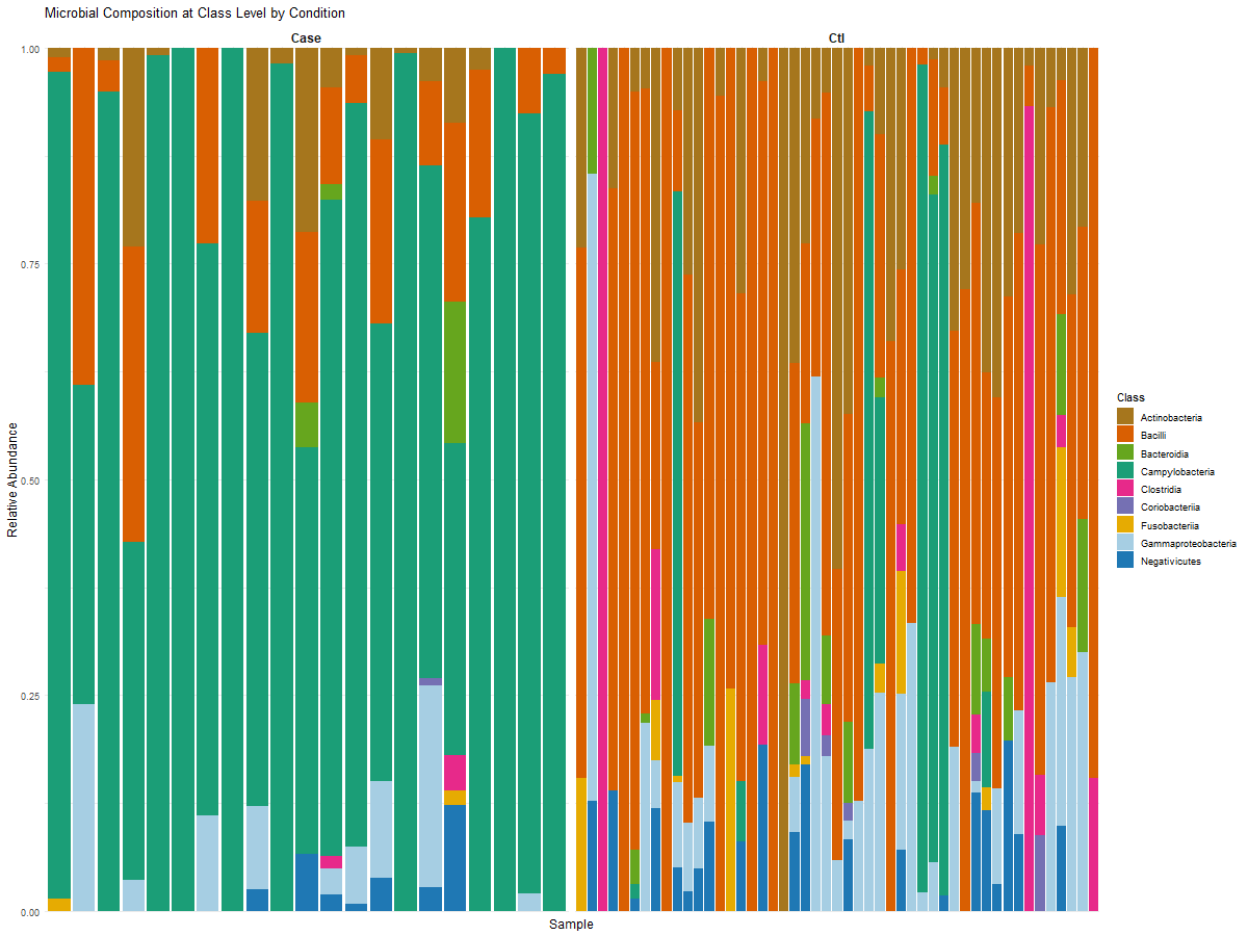
GCF_007794935.1_uk	3020	atpA(2558)	efp(626)	mutY(2567)	ppa(204)	trpC(2628)	ureI(2634)	yphC(2654)
GCF_900638475.1_afr	1288	atpA(1204)	efp(1151)	mutY(1200)	ppa(1174)	trpC(1213)	ureI(1228)	yphC(1223)



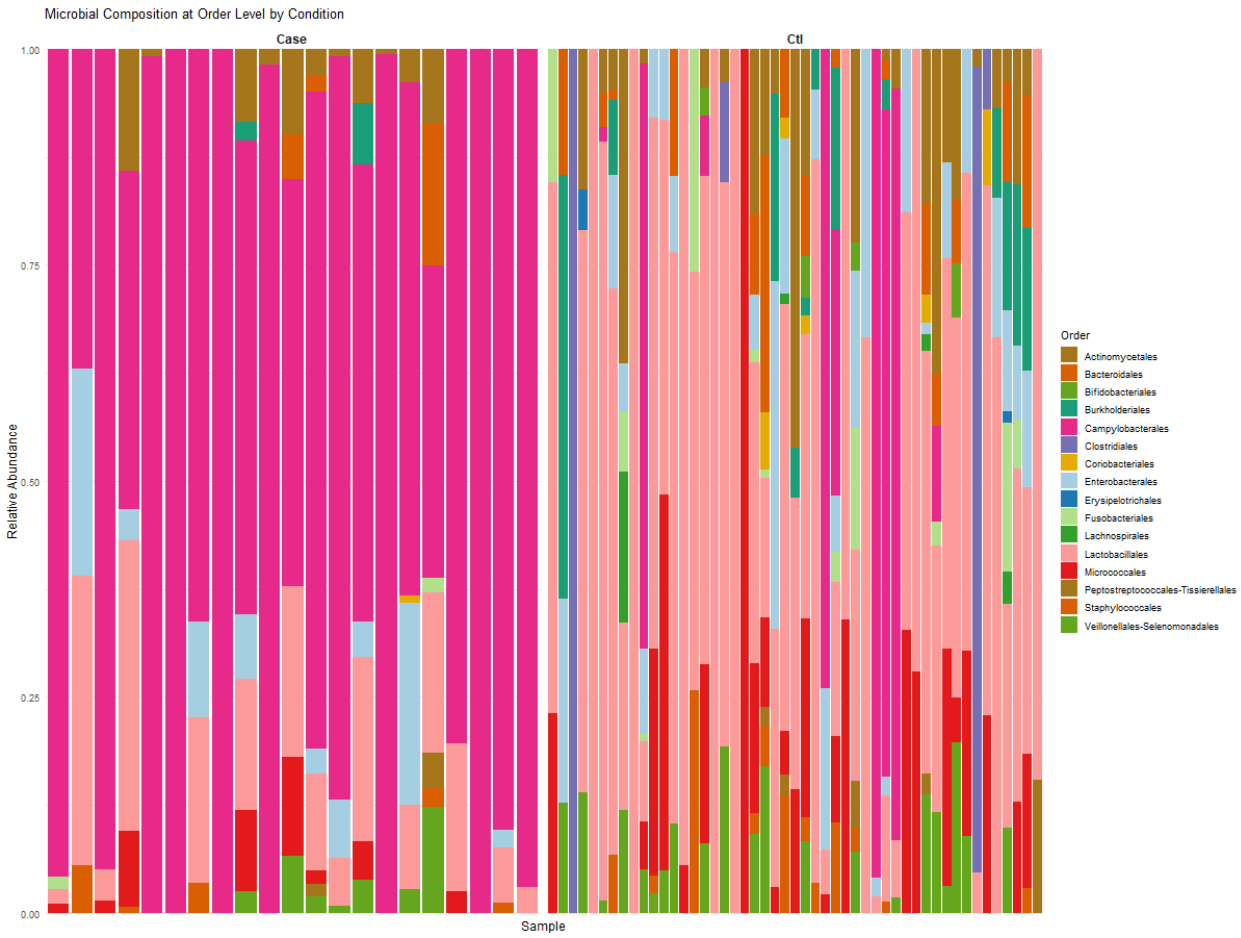
Supplementary Figure S7.2 shows FastQC per Sequence Quality Score



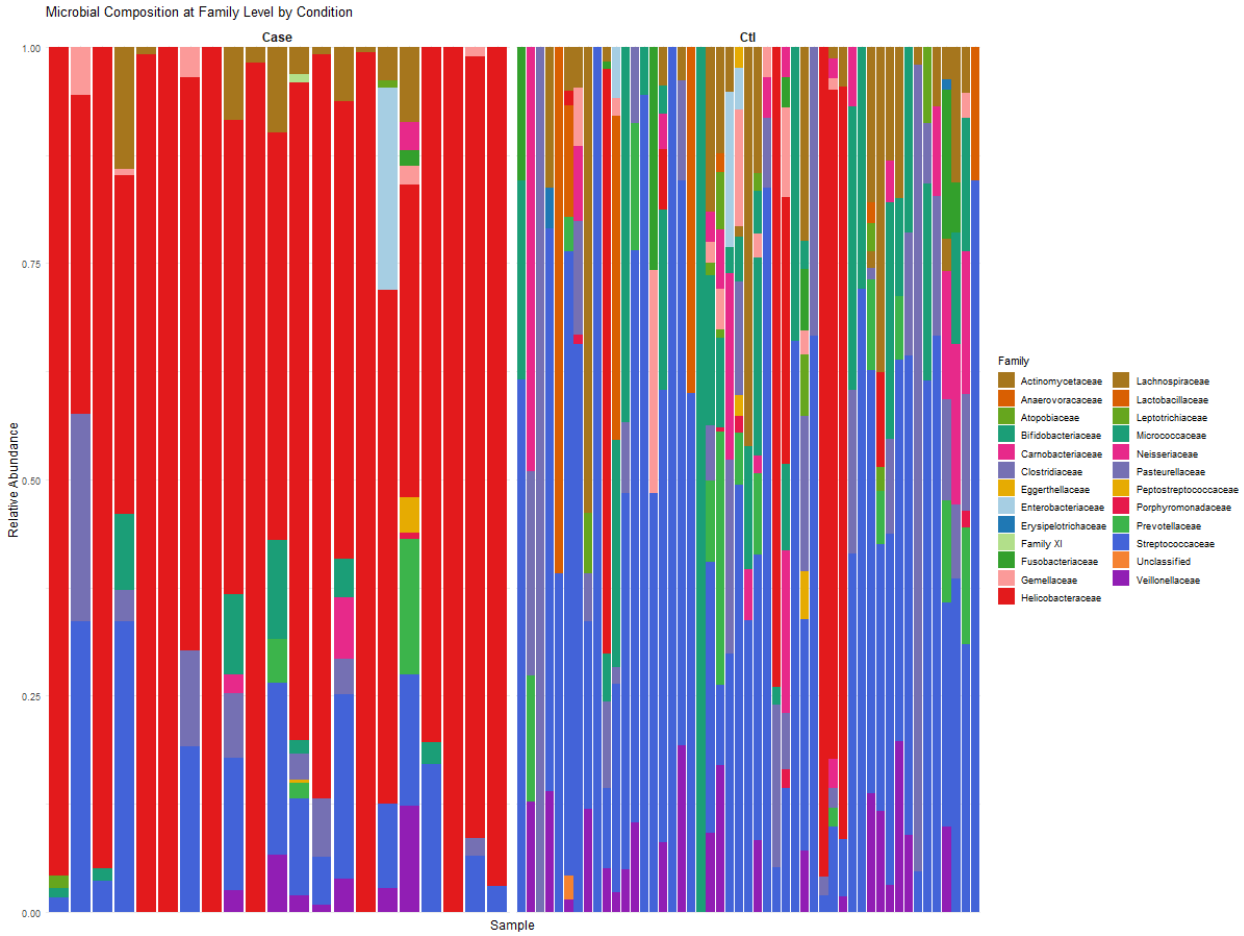
Supplementary Figure S7.3: Bar plot of relative abundance of bacteria at the Phylum level



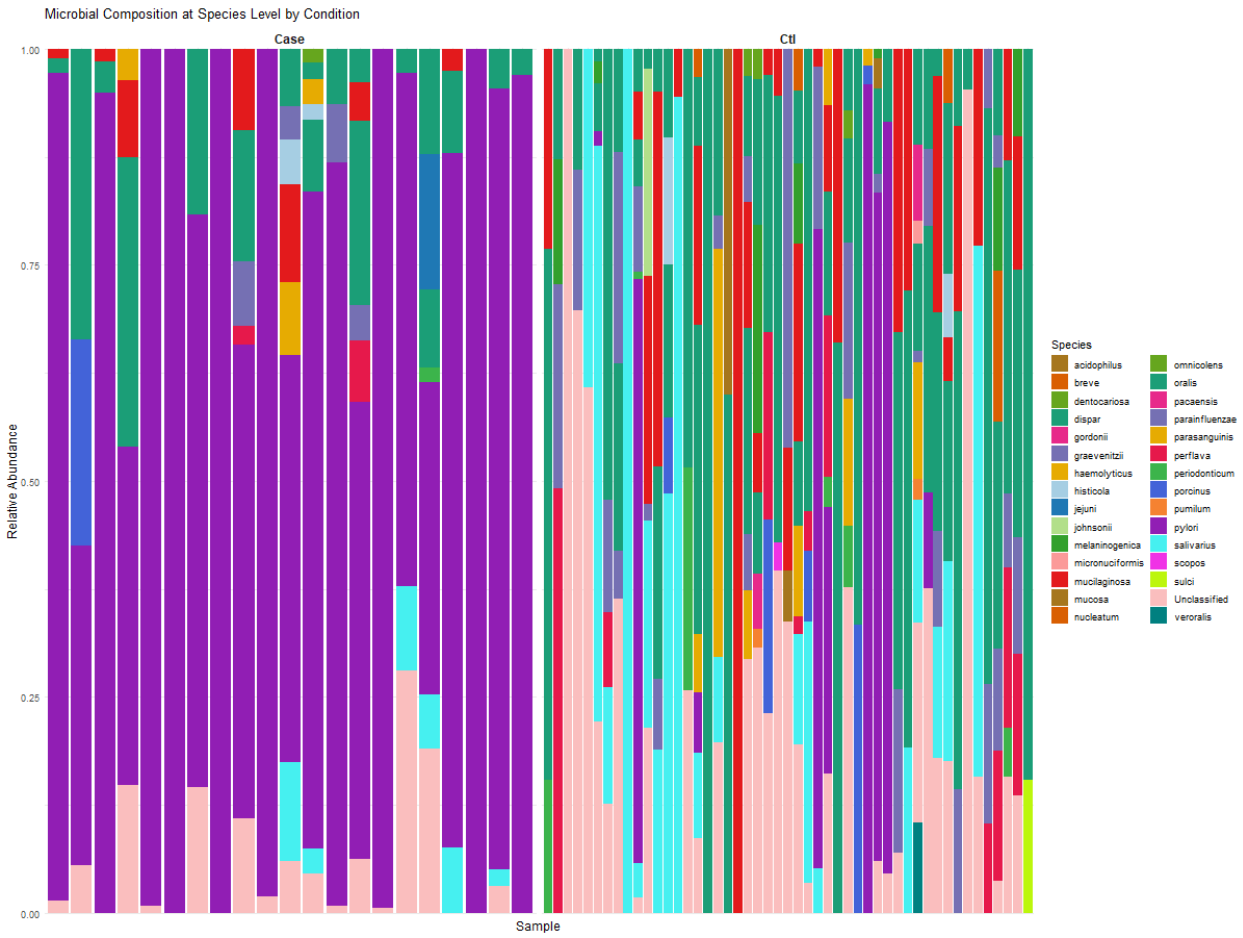
Supplementary Figure S7.4: Bar plot of relative abundance of bacteria at the class level



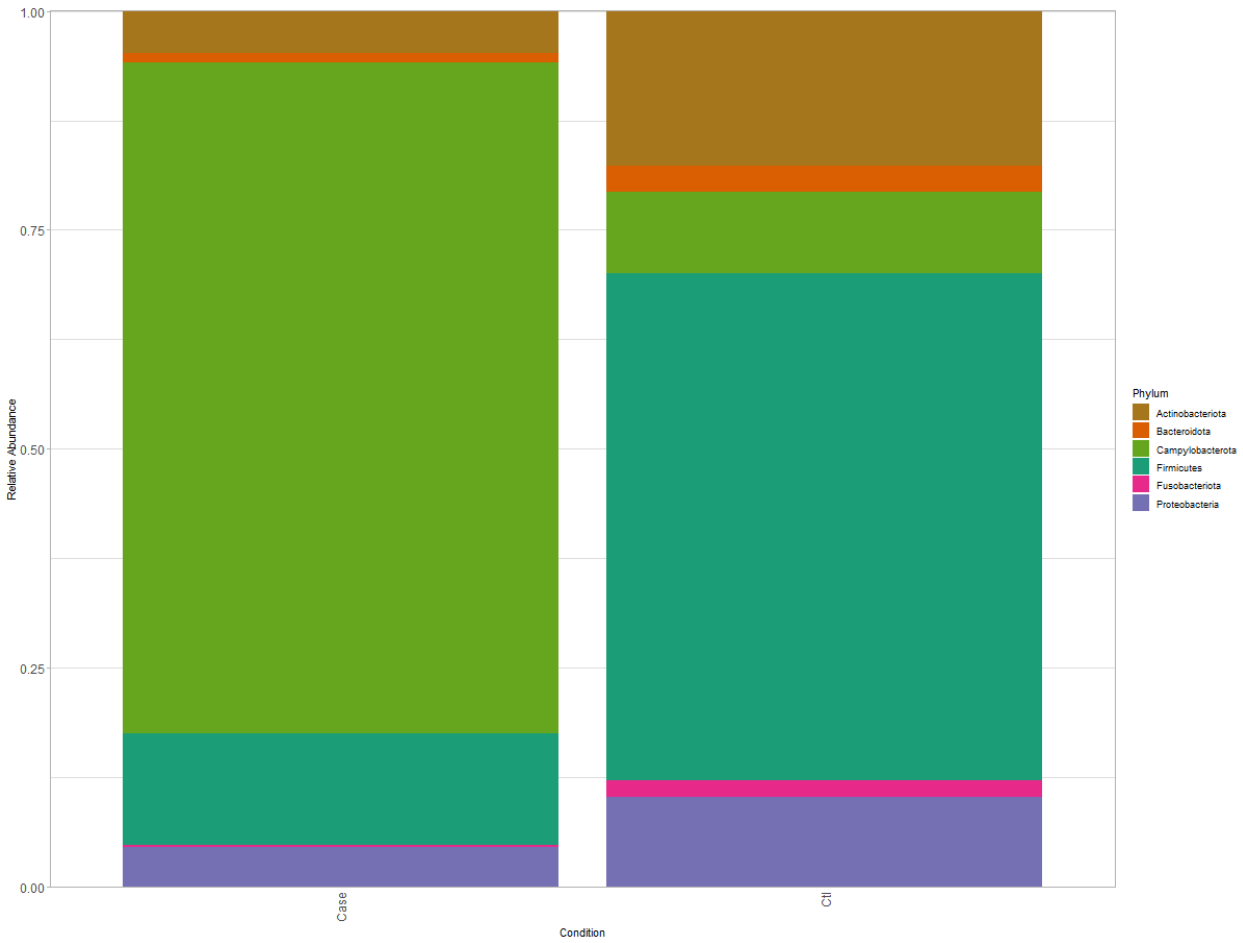
Supplementary Figure S7.5: Bar plot of relative abundance of bacteria at the Order level



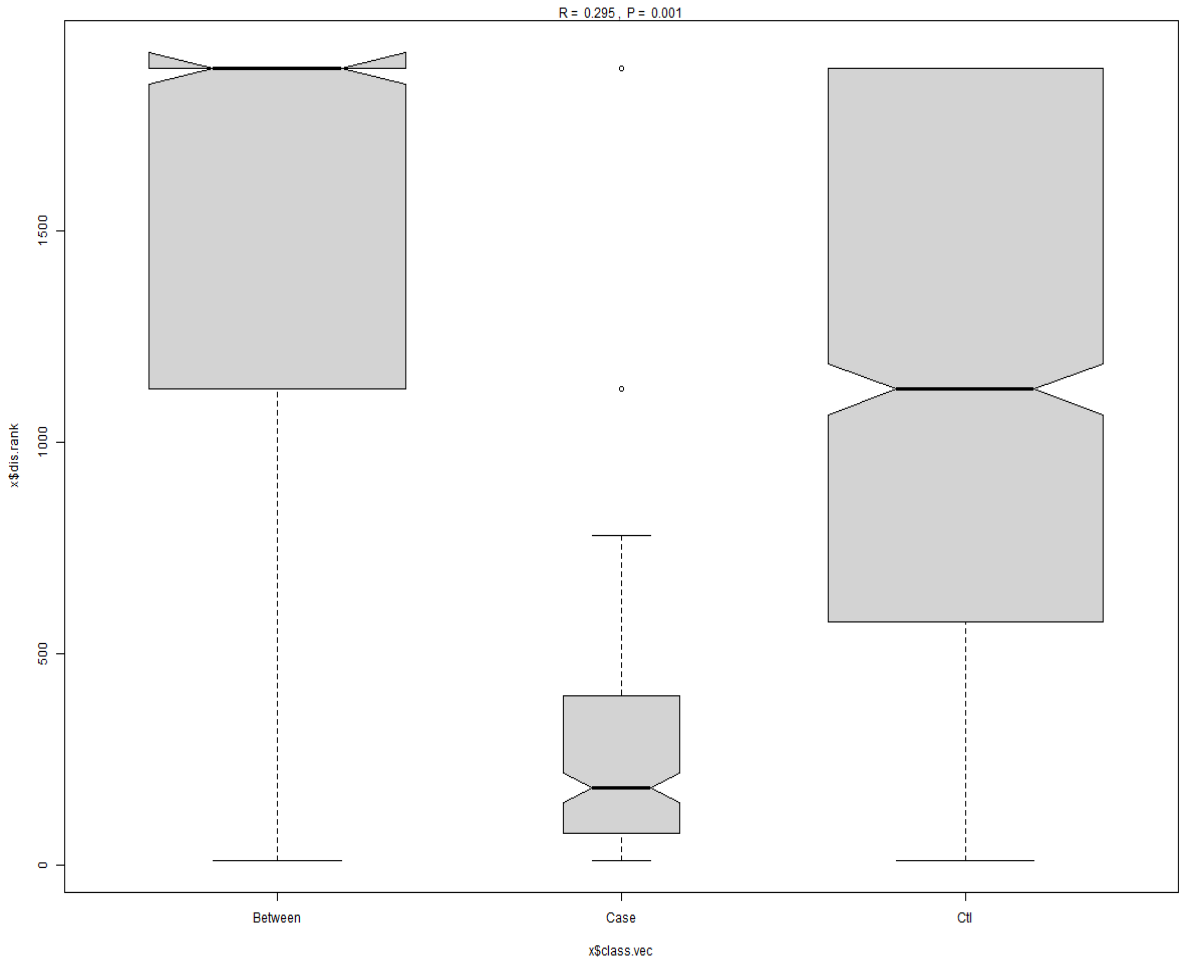
Supplementary Figure S7.6: Bar plot of relative abundance of bacteria at the family level



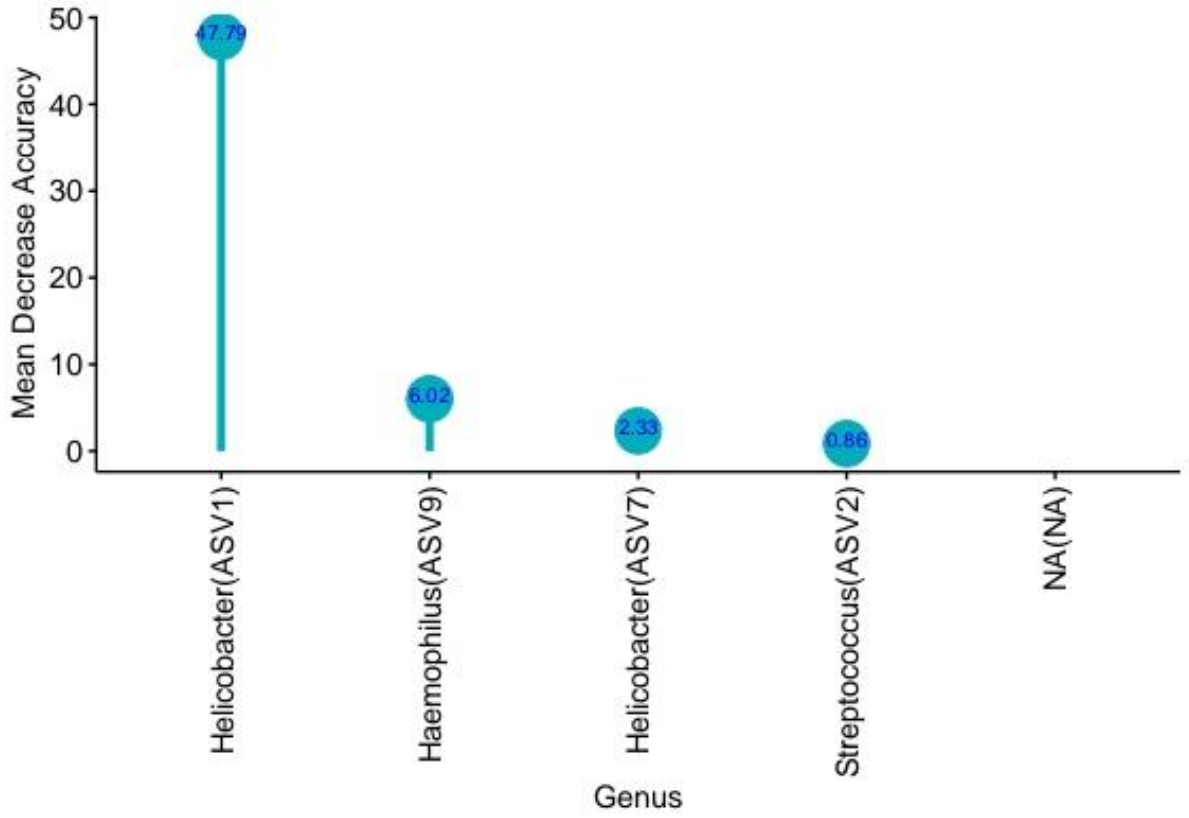
Supplementary Figure S7.7: Bar plot of relative abundance of bacteria at the species level



Supplemental Figure S7.8: Relative abundance at the Phylum level in both the *H. pylori* negative and *H. pylori* positive groups

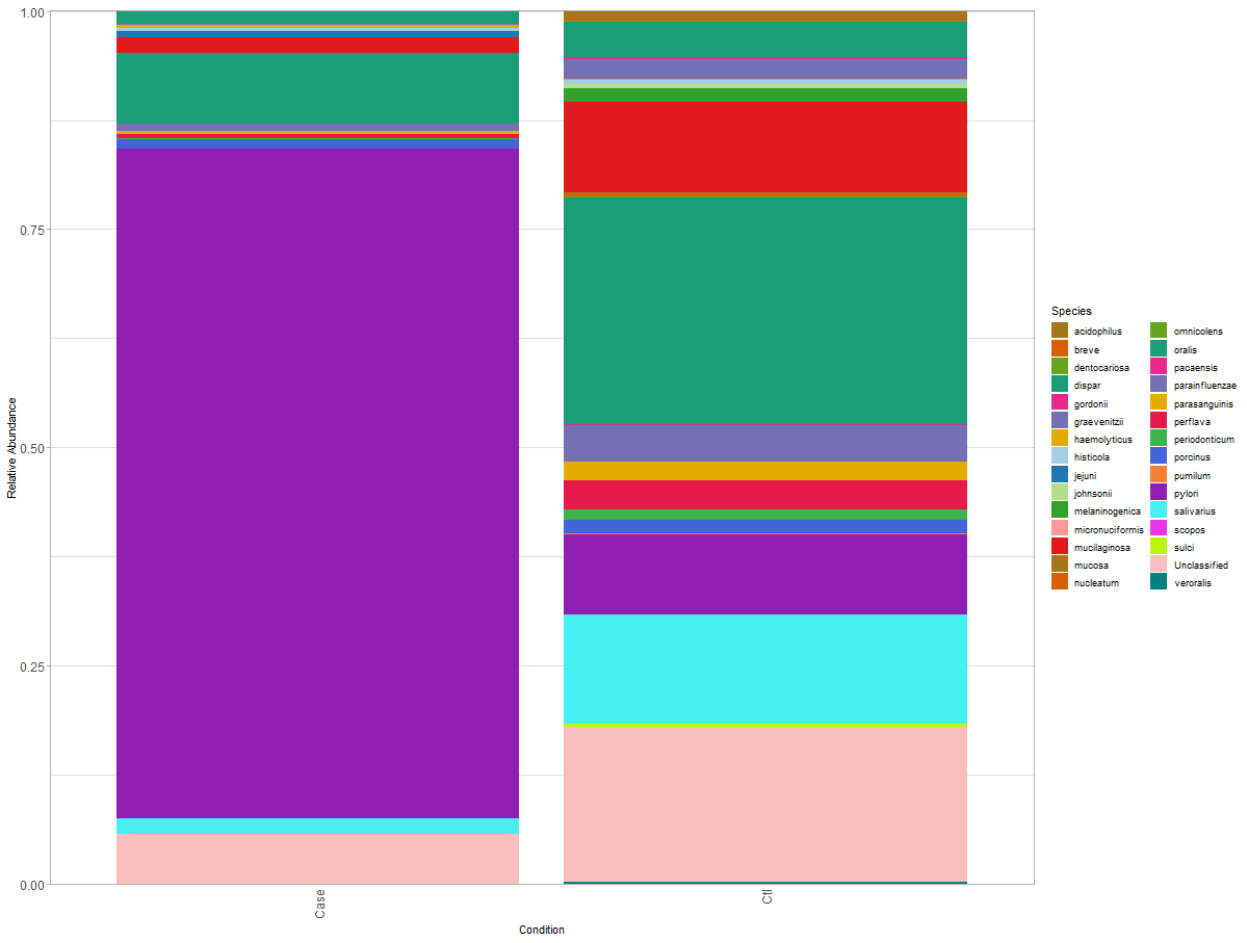


Supplementary Figure S7.9: Shows the Analysis of similarity (ANOSIM) between the cases and control

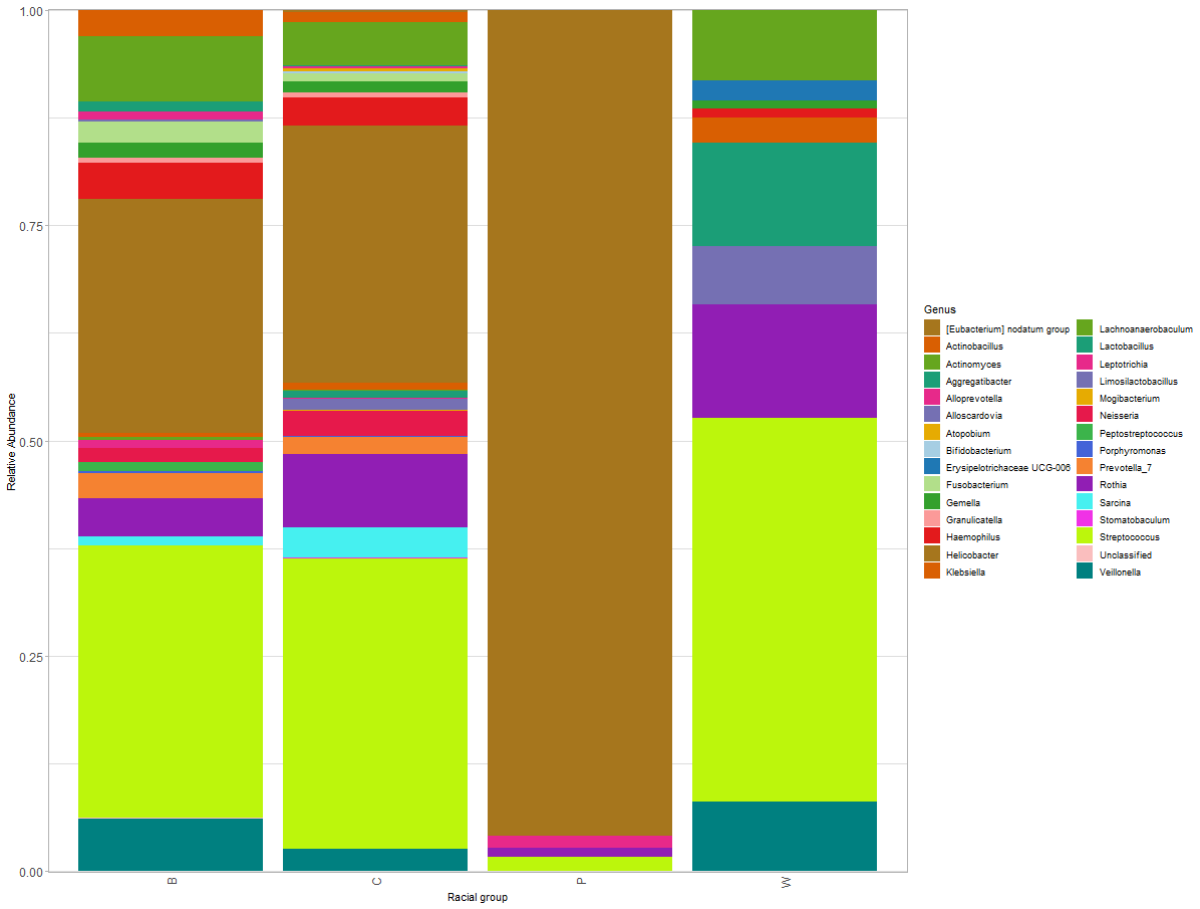


Supplementary Figure S7.10: The important differentially expressed bacteria/taxa to the microbial community at the Genus level

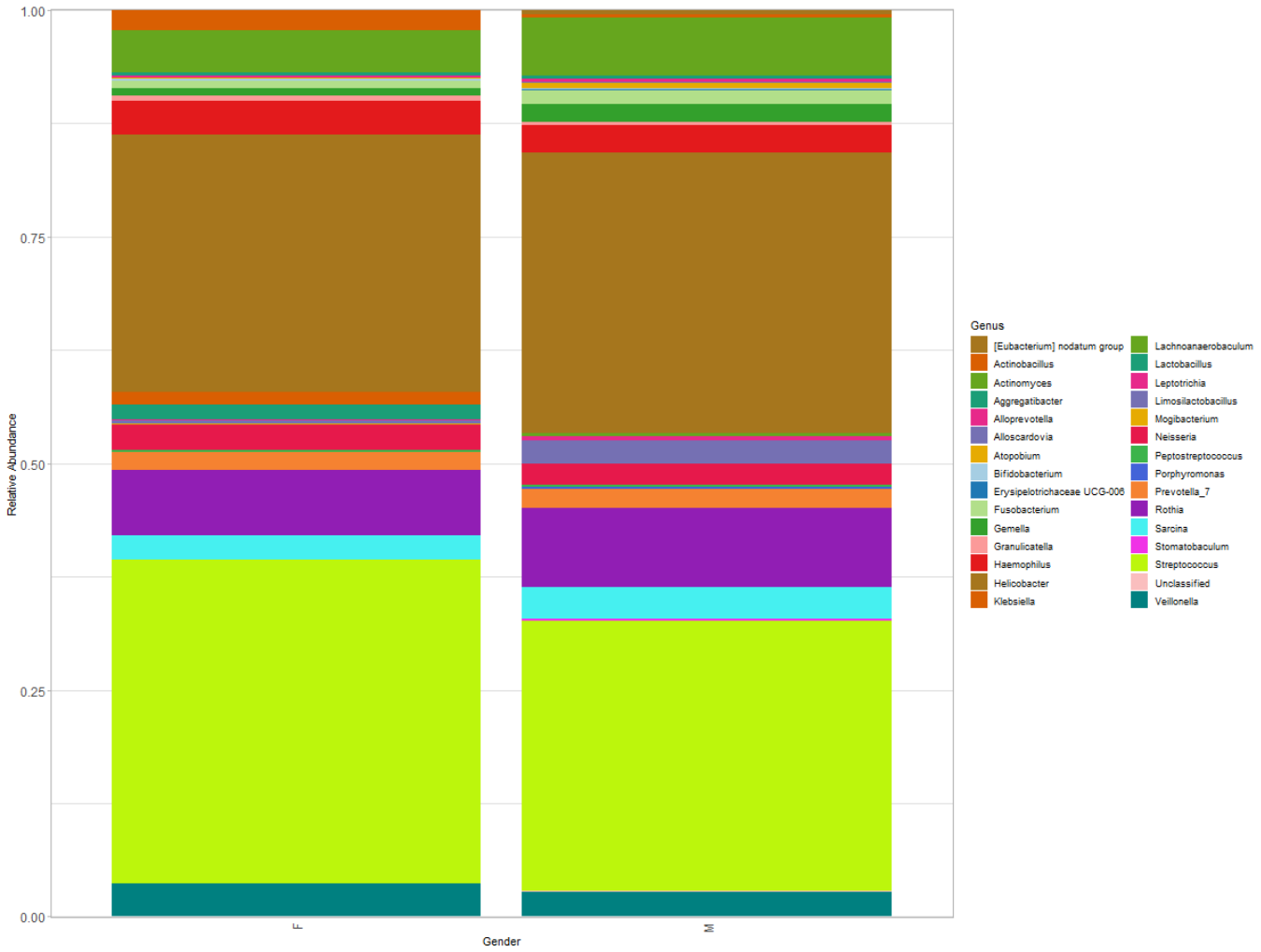
NB: ASV (Amplicon Sequence Variants), NA (Not Applicable)



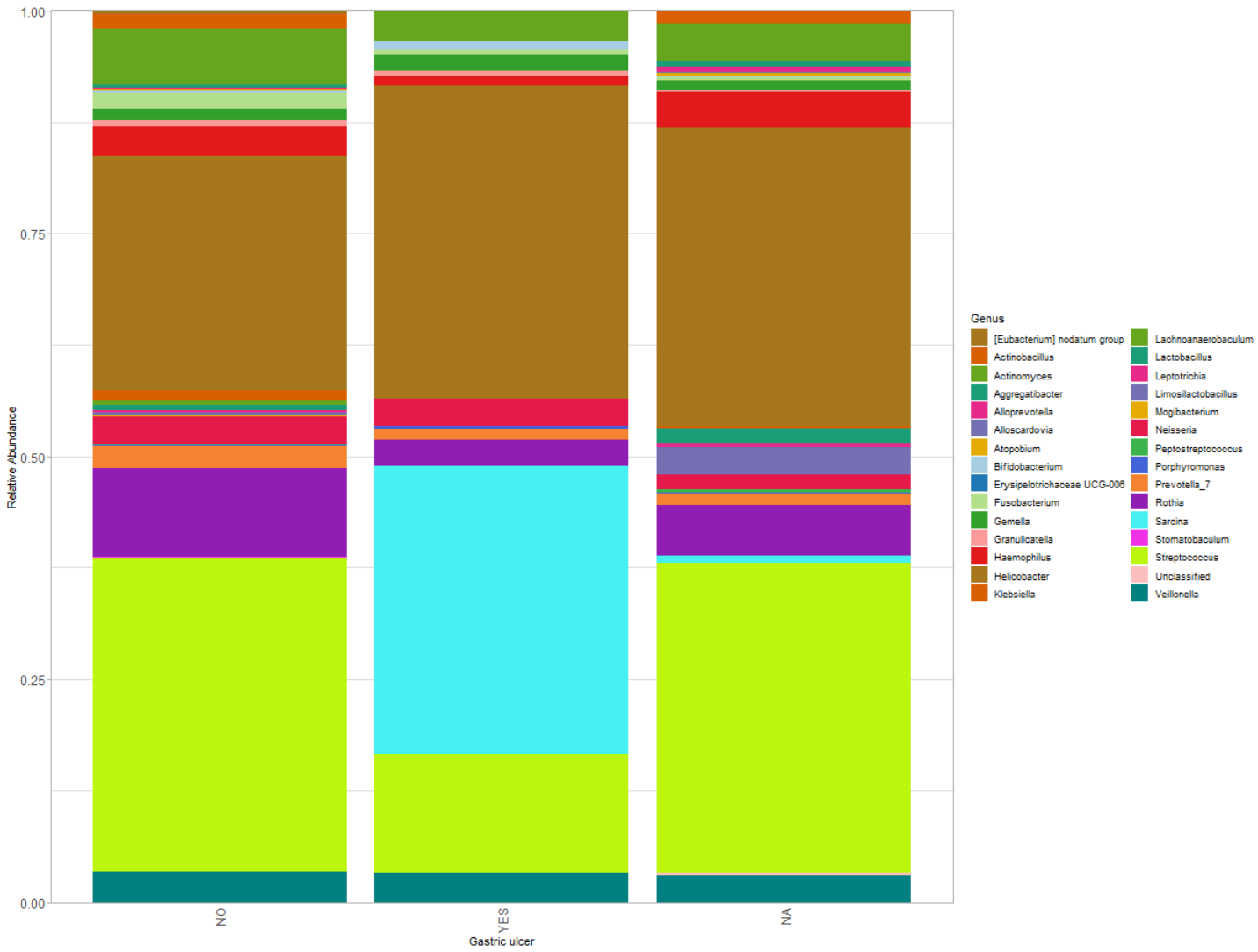
Supplementary Figure S7.11: Taxa differentially abundant between cases and controls based on the Linear Discriminant Analysis Effect Size effects (LEfSe) at the species level



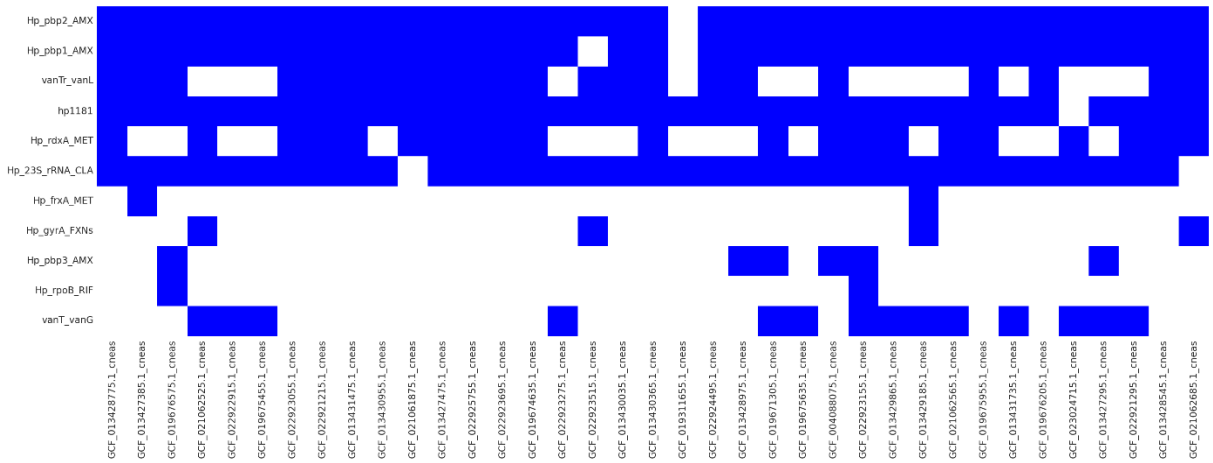
Supplementary Figure S7.12: A bar plot showing the correlation between the racial groups and the gastric microbiome in a South African cohort



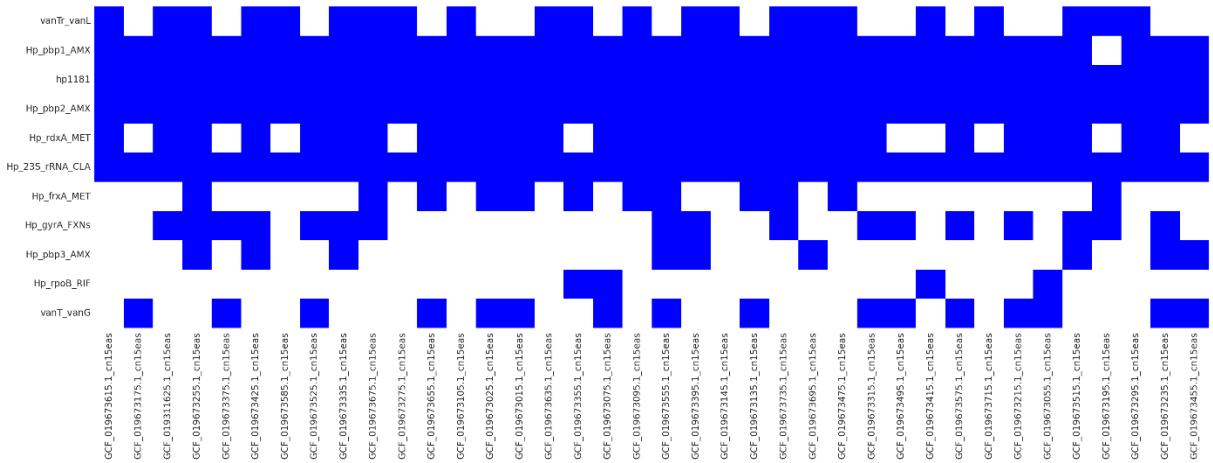
Supplementary Figure S7.13: A bar plot showing the correlation between gender and the gastric microbiome in a South African cohort



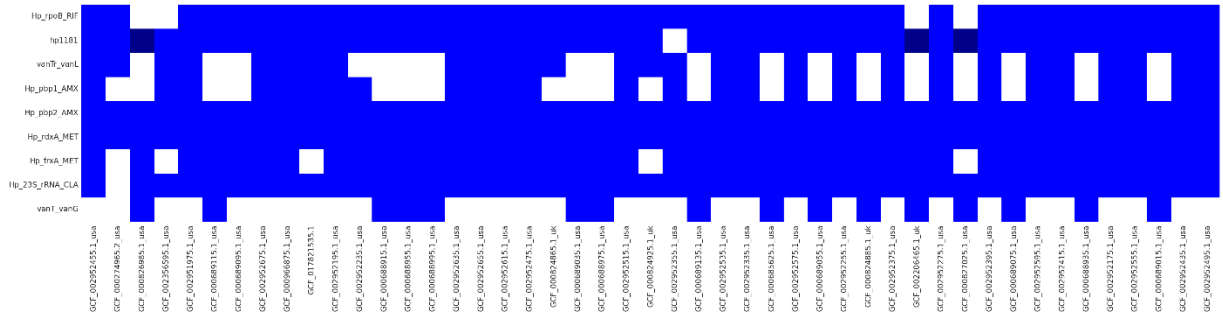
Supplementary Figure S7.14: A bar plot showing the correlation between patients with or without gastric ulcer and the gastric microbiome in a South African cohort



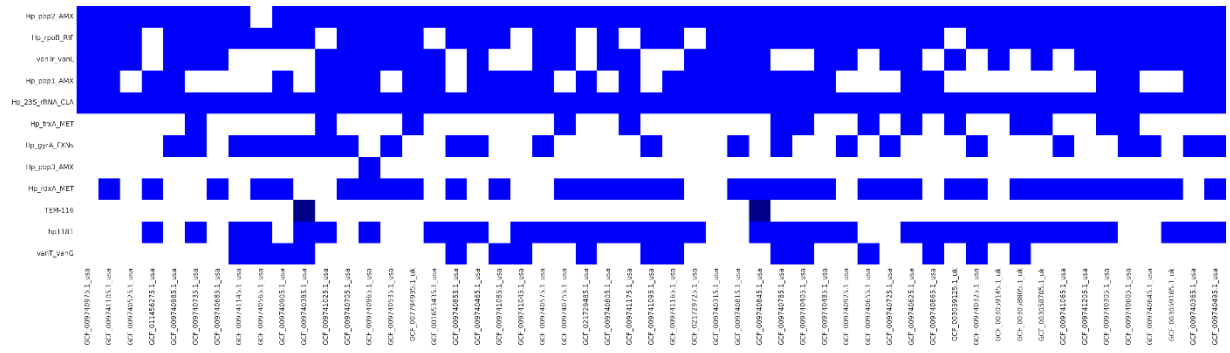
AMR genes for China



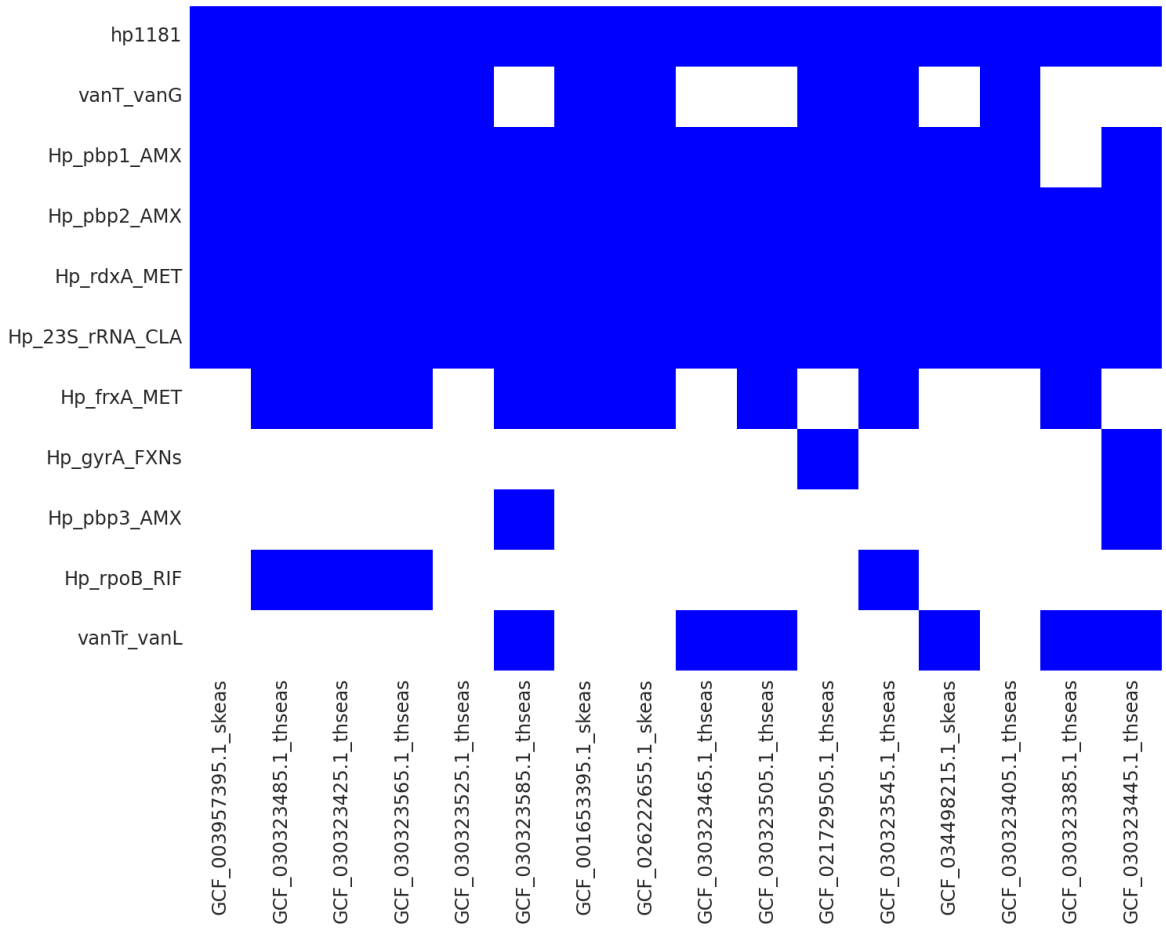
AMR genes for the Mongols



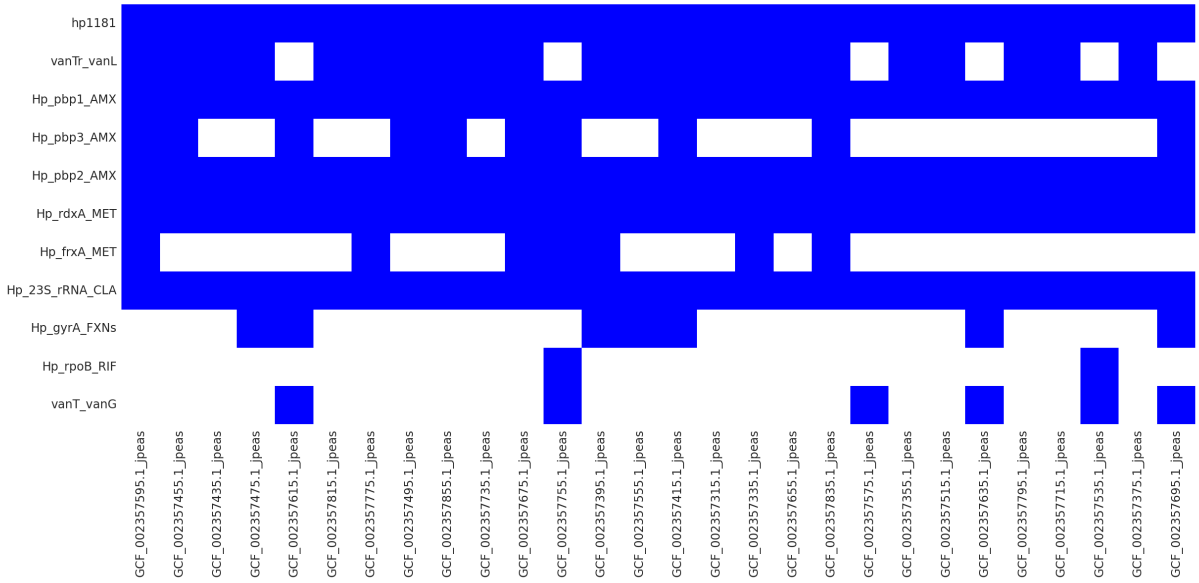
AMR genes for USA and UK 1



AMR genes for the USA and UK 2



AMR Genes for Thailand and South Koreans



AMR genes for Japan

Supplementary Figure S7.15: Heatmap showing the presence/absence of AMR genes in the USA, Europe (UK), and East Asian countries (Thailand, South Korea, Japan, Mongols and China)

APPENDICES:

Study Protocol Information Leaflet

Clinical factors, environmental risk factors, and microbiome signatures of *Helicobacter pylori* in a South African cohort

Information to patients and consent form

PART A: INFORMATION

Introduction

Helicobacter pylori is a common infection (bug) that affects your stomach. It usually causes few symptoms but may cause stomach irritation that leads to pain, bloating and general discomfort. In some individuals, if not treated on time, it can lead to changes in the stomach that puts the person at risk of developing cancer of the stomach. In our opinion we believe apart from this bug *(H. Pylori*), other factors also contribute to the changes that eventually leads to cancer of the stomach. It is possible that even you can be having that bug, and therefore needs to be investigated and treated accordingly, so that your risk of developing stomach cancer is reduced. This bug can be treated with antibiotics combined with tablets that reduces the level of the acid in the stomach. Though this bug can cause stomach cancer, not everyone having this bug will eventually develop stomach cancer. It therefore means that there are other factors that contributes to the development of this cancer.

Purpose of the study

The main purpose of the study is for us to better understand how the Helicobacter pylori bug combined with the other factors (like smoking, diet, alcohol, etc) leads to cancer of the stomach. The manner at which this bug (H. Pylori) in the presence of these other risk factors causes stomach cancer is not well known. This makes it difficult for us to find out who will develop this cancer and even how to find out early and give appropriate treatment.

It is important to emphasize that, the fact that you are been selected for this study does not mean that you have cancer, or that you are at increased risk. We are only checking for the presence of the bug called *H. pylori* which commonly causes irritation of the stomach wall; if present, you will be treated accordingly.

What will my participation in this project involve?

- We will ask you to answer some questions for us that will help us in the study.
- We want you to also allow us to take blood from you. This blood is not going to be analysed in this study. It will be stored for future studies that we will plan to understand this bug

better. See below under "**What will happen to my samples?"**

- Place your past, current and future medical record information into the study database.
- Take 4 extra stomach biopsies (small pieces of the lining of the stomach) at time of your gastroscopy to help us understand the infection better.

By participating in this study, we will be able to better understand the type of bug causing infections in the stomach to help us provide better treatment for patients like you in the future. Up to 100 people will be in this study.

Will participation affect how you are treated for your condition?

Your treatment will not be affected by participation in the project. You will be treated by a group of specialists who are well trained in looking after Helicobacter pylori stomach infections. Your treatment and clinic follow up will be unaffected by whether you participate in the study or not.

What are the possible risks of my participation in the study?

There are very low risks of physical injury associated with your participation in this study. Gastroscopy, a procedure to look at your stomach lining with a camera and to take biopsies, will be part of your normal clinical care. Your doctor will explain the risks of this procedure to you. If you agree to participate in this study, your doctor will take 4 additional biopsies for this study at the same time. It is unlikely but possible that additional biopsies would increase your risk of complications. If the complication is due to the additional biopsies, then any additional costs of therapy would be covered by the study insurance.

Also, taking blood from your arm has very low risk for injury, though you will experience small needle stick at the time of sticking in the needle, and some may experience small redness at the site of blood withdrawal.

Participation does involve the possible risk that information about your health might become known to individuals other than your usual healthcare providers. We will make every effort to preserve your medical record confidentiality. Only the medical staff involved in the study will have access to this information. We will remove personal identifiers (for example, your name, address, contact details) from information stored about you on computer before it is analysed for research.

What are the possible benefits of my participation in the study?

It is unlikely that you will receive any direct benefit as a result of your participation in the study. However, the information obtained will be used to improve our knowledge and treatment of Helicobacter pylori stomach infections and we believe this knowledge will benefit patients with similar infections in the future.

What other choices do I have besides this study?

Taking part in this study is completely voluntary. You may choose not to take part in this study or leave this study at any time. You will be treated the same no matter what you decide.

Will my medical aid or I be charged for my participation in the study?

The cost of your medical care will be determined by the hospital and medical aid assessment processes. There will be no additional costs to you or your medical aid to participate in the study.

Will I be paid for my participation in the study?

You will not receive any money to participate in the study as you will be treated the same no matter what you decide.

What information from my medical record will be placed into the study?

All of your age, sex, place of birth, family history of gastric cancer and smoking habits, alcohol intake, your type of diet recorded onto a computer which is what we call keeping a registry.

After all of your clinic follow up visits, we may continue to place your medical record information into the registry, however, you may withdraw your permission for participation in the registry at any time. Your medical record information contained within the registry may be used for research purposes for an indefinite period of time.

Will information about me be kept private?

Researchers will take every reasonable step to protect the privacy of your health information and to prevent misuse of this information. You will not be identified by name or any other way in any publication about this study. Efforts will be made to keep your personal information private, but we cannot guarantee complete confidentiality. Your personal information may be released if required by law.

Access to your identifiable medical record information contained within the registry will be limited to investigators and staff associated with this study. Your medical and research records may be reviewed by regulatory authorities who ensure your safety while in the study.

Is my participation in the study voluntary?

Your participation in the study is completely voluntary. Whether or not you agree to participate in this study will have no effect on your current or future medical care.

Can I change my mind about taking part in this study?

You may withdraw at any time your consent for further participation in the study and your permission for us to use your medical record information further. You may have all your information removed from the Project Registry.

What will happen to my samples?

During the study your stomach biopsies will be collected. Your samples will be stored at a repository (a facility that stores samples). The samples can only be released to a researcher who has been approved by the primary researcher of this study. Your privacy will be protected because

a randomly assigned number that does not have any of your personal information will identify each specimen in the repository.

The bloods taken will be stored for future analysis of interleukins and cytokines (immunological assay) related to *H. Pylori* infection and DNA sequencing (genetic studies) to identify the changes associated with the risk in the development of gastric cancer among those who are infected with *H. Pylori*, with the aim of understanding all the contributions of disease related to this bug. For this study we have ethics approval ONLY for storage of your samples. No genetic or other tests will be performed. In the future if we want to use your samples, we will have to write a proposal to the ethics committee, explaining precisely what we will do with the stored blood. **Therefore, the use of your stored blood at that time will be pending ethics approval**. After ethics approval, we will also require you to give specific consent for this, at that time, which will be entirely voluntary.

What if Something Goes Wrong?

The University of Cape Town (UCT) undertakes that in the event of you suffering any significant deterioration in health or well-being, or from any unexpected sensitivity or toxicity, that is caused by your participation in the study, it will provide immediate medical care. UCT has appropriate insurance cover to provide prompt payment of compensation for any trial-related injury according to the guidelines outlined by the Association of the British Pharmaceutical Industry, ABPI 1991. Broadly-speaking, the ABPI guidelines recommend that the insured company (UCT), without legal commitment, should compensate you without you having to prove that UCT is at fault. An injury is considered trial-related if, and to the extent that, it is caused by study activities. You must notify the study doctor immediately of any side effects and/or injuries during the trial, whether they are research-related or other related complications.

UCT reserves the right not to provide compensation if, and to the extent that, your injury came about because you chose not to follow the instructions that you were given while you were taking part in the study. Your right in law to claim compensation for injury where you prove negligence is not affected. Copies of these guidelines are available on request.

Who should I contact if I have questions?

If you have any other questions of if you want to leave this study, please ask the study staff. You may also contact

- Prof Setshedi M or Dr Innocent Francis, Telephone: (+27) 021 404 3040

For questions about your rights as a research subject or if you feel you have been harmed by the research, contact:

- University of Cape Town Human Research Ethics Committee, Prof Marc Blockman. Telephone: (+27) 021 406 6338.

PART B: CONSENT TO PARTICIPATE IN THE STUDY

I have read the foregoing information, or it has been read to me. I have had the opportunity to ask questions about it and any questions that I have been asked was answered to my satisfaction. I consent voluntarily to participate as a participant in this research. I consent to being interviewed, examined and have bloods taken.
Print Name of Participant: _____
Signature of Participant: _____
Date: _____

If unable to read or write
I have witnessed the accurate reading of the consent form to the potential participant, and the individual has had the opportunity to ask questions. I confirm that the individual has given consent freely.

Print name of witness: _____ AND Thumb print of participant
Signature of witness: _____
Date: _____

Statement by the researcher/person taking consent
I have accurately read out the information sheet to the potential participant, and to the best of my ability made sure that the participant understands what the study is about and the procedures that will be followed.
I confirm that the participant was given an opportunity to ask questions about the study, and all the questions asked by the participant have been answered correctly and to the best of my ability.
I confirm that the individual has not been coerced into giving consent, and the consent has been given freely and voluntarily.

A copy of this document has been provided to the participant.

Print Name of Researcher/person taking the consent: _____

Signature of Researcher /person taking the consent: _____

Date: _____

PART C: CONSENT FOR BLOOD TO STORED/GASTRIC BIOPSY SAMPLES TO BE COLLECTED

Additional Consent to [Clinical factors, environmental risk factors, and microbiome signatures of *Helicobacter Pylori* in a South African cohort] for storage and future use of samples.

This Statement of Consent consists of two parts:

- Information Sheet (to share information about unused samples with you)
- Certificate of Consent (to record your agreement) You will be given a copy of the full Statement of Consent

Part II. Certificate of Consent

If any of the biopsied tissue I have provided for this research project is unused or leftover when the project is completed (Tick one choice from each of the following boxes)

o I wish my [gastric biopsy tissue] sample to be destroyed immediately.

o I want my [gastric biopsy] sample to be destroyed after ____ years.

o I give permission for my [Biopsied sample] to be stored indefinitely

AND

o I give permission for my (blood/gastric biopsy) sample to be stored and used in future research but only on the same subject as the current research project: [Clinical factors, environmental risk factors, and microbiome signatures of Helicobacter Pylori in a South African cohort]

o I give my permission for my [blood/gastric biopsy] sample to be stored and used in future research of any type which has been properly approved

o I give permission for my [blood/gastric biopsy] sample to be stored and used in future research except for research about [Clinical factors, environmental risk factors, and microbiome signatures of Helicobacter Pylori in a South African cohort]

AND

o I want my identity to be removed from my (blood/gastric biopsy) sample.

o I want my identity to be kept with my (blood/gastric biopsy) sample.

I have read the information, or it has been read to me. I have had the opportunity to ask questions about it and my questions have been answered to my satisfaction. I consent voluntarily to have my samples stored in the manner and for the purpose indicated above.

Print Name of Participant_____

Signature of Participant _____

Date _____

Day/month/year

If illiterate

A literate witness must sign (if possible, this person should be selected by the participant and should have no connection to the research team). Participants who are illiterate should include their thumbprint as well.

I have witnessed the accurate reading of the consent form to the potential participant, and the individual has had the opportunity to ask questions. I confirm that the individual has given consent freely.

Print name of witness_____ AND Thumb print of participant

Signature of witness _____

Date _____ Day/month/year

Statement by the researcher/person taking consent

I have accurately read out the information sheet to the potential participant, and to the best of my ability made sure that the participant understands that the following will be done:

1. Blood sample will be taken from the participants
2. Extra gastric biopsies will be taken from the participant
3. **Both specimens will be stored for future studies in this cohort of patient as an extension of this study. Prior to analyzing these stored samples, a project approval will be obtained from the ethics committee**.

I confirm that the participant was given an opportunity to ask questions about the nature and manner of storage of the samples, and all the questions asked by the participant have been answered correctly and to the best of my ability. I confirm that the individual has not been coerced into giving consent, and the consent has been given freely and voluntarily. A copy of this ICF has been provided to the participant.

Print Name of Researcher/person taking the consent_____

Signature of Researcher /person taking the consent_____

Date _____ Day/month/year

DATA COLLECTION TABLE FOR THE PROSPECTIVE STUDY ON THE CLINICAL FEATURES, ENVIRONMENTAL RISK FACTORS AND THE MICROBIOME SIGNATURES OF *H. PYLORI* IN A SOUTH AFRICAN COHORT, USING GROOTE HOSPITAL IN CAPE TOWN AS A STUDY CENTRE

Data collected	Helicobacter Pylori positive	Helicobacter Pylori Negative
Biodata		
Age		
Gender		
History of presenting complaint		
Co-morbidities		
Clinical factors		
Alcohol intake	Yes/No	Yes/No
Smoking history	Yes/No	Yes/No
Use of NSAIDS	Yes/No	Yes/No
PPIs	Yes/No	Yes/No
Diet		
Vegetarian diet	Yes/No	Yes/No
Examination findings		
Height (metres)		
Weight (Kg)		
BMI (kg/m^2)		
Indications for gastroscopy (Epigastric pain, weight loss, iron deficiency anaemia, melena etc.)		
Laboratory findings		
Haemoglobin level		
Mean corpuscular volume		
Iron levels Transferrin %Saturation		
Endoscopic findings		

Gastritis	Yes/No	Yes/No
Duodenal ulcers	Yes/No	Yes/No
Gastric Ulcers	Yes/No	Yes/No
Others (specified)		
Rapid Urea test	Positive/Negative	Positive/Negative
Histological findings		
Degree of Atrophy Site (Antral or Corpus) Presence/density of H. Pylori		
Microbiome data on gastric Biopsy (H. Pylori 16SRNA Present	Yes/No	Yes/No
Other species (list)		