**DOCTOR OF PHILOSOPHY (PhD)**

**RESEARCH PROPOSAL**

**On**

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

**By**

**Dr Francis Innocent Ekparolaguaziba**

**Supervisor:** Prof Mashiko Setshedi

**Co-supervisor:** Dr Mamadou Kaba

**ABSTRACT**

|  |  |
| --- | --- |
| Protocol title | Clinical factors, environmental risk factors and microbiome signatures of Helicobacter Pylori (*H. Pylori*) infection in a diverse South African cohort. |
| Hypothesis | 1. The so called “African enigma” is not true. 2. The clinical risk factors and histological activity in *H. Pylori* infected persons will provide clues to risk factors for the development of gastric cancer (GCA). 3. The gastric microbiome of patients infected with *H. Pylori* will have a more protective/anti-inflammatory signature against aggressive phenotypes of disease. |
| Aim | This study is aimed at evaluating the burden of *H. Pylori* infection and *H. Pylori* associated gastric cancer, as well as to evaluate the clinical and environmental risk factors and the microbiome signatures of *H. Pylori* infection using gastric biopsy samples. |
| Objective | 1. To determine if the enigma is true by doing a search of the literature of Sub-Saharan Africa (SSA) and comparing it to region of similar *H. Pylori* prevalence, and comparing the incidence of GCA, and associated risk factors. 2. Retrospectively, assess a cohort of patients with biopsy proven GCA with *H. Pylori* 16sRNA positive against a cohort with *H. Pylori* 16sRNA negative as control. Comparing demographic data, clinical risk factors, endoscopic and histologic pathology, with a view to assessing the contributions of *H. Pylori* in the development of GCA. 3. Prospectively, comparing a cohort of patients with H. Pylori and a *H. Pylori* negative control group, and comparing demographic data, clinical risk factors, endoscopic and histologic pathology, but also collect gastric biopsies for microbiome analysis for comparison in these two groups. 4. To establish a biorepository for blood specimens, for future genetic and immunological studies. |
| Study design | The study comprises of three (3) components:   1. A systematic review of the literature on the incidence and epidemiology of GCA in SSA compared to another region of similar H. Pylori prevalence. 2. A retrospective cohort study. 3. A prospective cohort study. |
| Study participants | The study will be carried out among adults in a South African cohort at a tertiary hospital:   1. For the retrospective study, folders of patients who are 18 years and above diagnosed with GCA will be retrieved for the study, with *H. Pylori* 16sRNA negative GCA cohort as a control. 2. For the prospective study patients with H. Pylori infection will be compared to H. Pylori negative controls for demographics, clinical features, endoscopy, histology and gastric microbiome. |
| Study site | Division of Gastroenterology, Groote Schuur Hospital, Cape Town |
| Data collected | Patient’s demographics, clinical-data, endoscopic and histologic findings for all cohorts (retrospective and prospective) and formalin fixed paraffin embedded (FFPE) gastric samples for the retrospective cohort.  Samples that will be collected for the prospective cohorts are:   1. Gastric biopsies for histology (normal clinical care) and additional “research” biopsies for the microbiome for 16srRNA. 2. Bloods (30ml) for storage for genetic and immunology studies in the future. |
| Significance of the study | This study aims to understand the so-called “African enigma”, to determine if indeed it exists, and if so, what are the factors accounting for this. These can in turn be exploited to advance preventative and therapeutic strategies to maintain or further reduce the low rates of GCA in our region. |
| Ethics | Ethical approval will be obtained from the University of Cape Town Human Research Ethics Committee. |
| Study personnel | **PhD candidate:** Dr Francis Innocent Ekparolaguaziba  **Primary supervisor:** Prof Mashiko Setshedi  **Co-supervisor:** Dr Mamadou Kaba |

**INTRODUCTION**

*Helicobacter Pylori* (*H. Pylori)* is a gram -negative organism affecting about 50% of the general population.(1-3) There are striking geographic differences in prevalence of *H. Pylori* in patients with peptic ulcer disease. It is ubiquitous in Africa with a prevalence of approximately 90% compared to 25% in developed countries.(4, 5) *H. Pylori* has a causal relationship with various diseases such as gastric ulcers, duodenal ulcers, non-ulcer dyspepsia, mucosal-associated gastric lymphoma and gastric cancer.(6) *H. Pylori* is estimated to be responsible for more than 66% of gastric cancers worldwide.(7) Despite a high prevalence of *H. Pylori* infection in African countries there is a disproportionately low incidence of gastric cancer.(8, 9) This is the so-called “African enigma” and is yet to be elucidated.

The mechanisms of *H. Pylori* chronic gastritis-dysplasia-metaplasia-cancer sequence are not known but are thought to be related to *H. Pylori* virulence factors (e.g., cagA, vacA and others)(10), the host inflammatory response as well as environmental factors,(11) importantly, western diets, smoking and alcohol had been positively associated with gastric cancer(12, 13).

More recently, the role of the microbiome in H. Pylori-related pathogenesis is being increasingly recognized. Patients with H. Pylori infection have faecal microbiota alterations with an increase in diversity.(14) Additionally, those with higher *H. Pylori* antigen stool tests showed a harmful microbial phenotype; this may predispose to the development of gastric cancer.(14) In another study dysbiotic microbiota in *H. Pylori* positive gastric biopsies was associated with chronic atrophic gastritis and intestinal metaplasia/dysplasia(15), lending credence to the theory that microbiota may play a role in *H. Pylori* oncogenesis.

Previous studies locally have evaluated *H. Pylori* virulence factors, but to our knowledge no studies locally have evaluated the role of the microbiota. In order to further understand the oncogenesis of

*H.* *Pylor*i and be able to predict which *H. Pylori* infected patients are at risk of progressing to gastric cancer, further mechanistic studies are required.

**THE RATIONALE OF THE STUDY**

The rationale of this study is to define and correlate baseline clinico-pathological characteristics and the microbial signature of *H. Pylori* to further understand pathogenetic mechanisms of gastric cancer. This is being performed as a proof of concept study.

**HYPOTHESIS**

We hypothesize that

1. The so called “African enigma” is not true.
2. The clinical risk factors and histological activity in *H. Pylori* infected persons will provide clues to risk factors for the development of gastric cancer (GCA).
3. The gastric microbiome of patients infected with *H. Pylori* will have a more “protective”/anti-inflammatory signature against aggressive phenotypes of disease such as GCA.

**AIM OF THE STUDY**

1. Evaluating the clinical and environmental risk factors, as well as microbiome signatures of *H. Pylori* infection in a South African population, compared to *H. Pylori* negative controls.
2. To determine the risk factors and the contribution of *H. Pylori* towards the aetiopathogenesis of GCA in SSA.

**OBJECTIVES**

1. To determine if the Africa enigma is true by doing a search of the literature of Sub-Saharan Africa (SSA) and comparing it to region of similar H. Pylori prevalence, and comparing the incidence of GCA, and associated risk factors.
2. Retrospectively determine the burden of gastric cancer in Groote Schuur Hospital and its associated predisposing factors, comparing a cohort of patients with biopsy proven GCA with 16sRNA *H. Pylori* positive against a cohort with *H. Pylori* 16sRNA negative as control group and also characterizing the clinical presentation (symptoms, risk factors, endoscopic findings and histology), with a view to determine the contributions of *H. Pylori* in the development of gastric cancer.
3. To prospectively investigate a local cohort of patients with *H. Pylori* infection compared to non-infected control by characterizing:
4. The clinical presentation (symptoms, risk factors, endoscopy findings and histology)
5. Microbial signatures using gastric samples
6. To establish a biorepository for blood for future genetic and immunological studies.

**MATERIALS AND METHODS**

**Study design and location**

1. To conduct a systematic review of the burden of *H. Pylori* associated gastric cancer in Sun-Saharan Africa and compare it to another region with similar epidemiological prevalence of *H. Pylori.*
2. Retrospectively review a cohort of patients with biopsy proven GCA with H. Pylori 16sRNA positive (n=5) against a cohort of *H. Pylori* 16sRNA negative as control group (n=5). Comparing demographic data, clinical risk factors, endoscopic and histologic pathology, with a view to assessing the contributions of *H. Pylori* in the development of GCA.
3. To prospectively investigate a local cohort of patients with H. Pylori infection compared to non-infected controls by characterizing: the clinical presentation (symptoms, risk factors, endoscopy findings and histology), and microbial signatures using gastric samples. We will enrol patients consecutively over the 3-year period, as they present. Due to cost restrictions, we will only analyze 20 samples each from this group for the microbiome analysis.
4. The study will be conducted in the Division of Gastroenterology, Groote Schuur hospital, Cape Town.

**Study Population**

1. All publications for Sub-Saharan Africa with biopsy proven GCA for adults of 18 years and above published in last eleven (11) years (June 2010 and June 2021), will be included in the systematic review. A literature search will be done on Pubmed/Medline, Embase and the African data base.
2. The retrospective study cohort will include adults of 18 years and above with confirmed diagnoses of gastric cancer in Groote Schuur hospital.
3. The prospective cohort study will include a South African cohort of patients who are 18 years and above attending the gastroenterology clinic at Groote Schuur hospital, requiring a gastroscopy for the appropriate clinical indications who consent to participating in the study.

**Inclusion criteria**

1. All adults of 18 years.
2. Written consent.

**Exclusion criteria**

1. Patients under 18 years.
2. Pregnant women.
3. Patients with an acute gastrointestinal bleed.
4. Gravely ill patients.
5. No consent.

**Participant withdrawal**

Participants in the prospective cohort study will be reminded through on-going informed consent that the quality of their clinical care will not be adversely affected if they decline to participate in or withdraw from this study.

**Participant recruitment**

Searched articles from Pubmed/Medline, Embase and African data base which met the inclusion criteria will be included in the systematic review. Regarding the retrospective cohort study, patient’s folders with biopsy proven gastric cancer will be retrieved and data collected accordingly both from the folder records and NHLS.

For the prospective study, patients attending the new patient and follow up clinic or endoscopy open access service undergoing a gastroscopy for appropriate clinical indications, will be approached. They will be given verbal information and asked if they wish to partake in the study. If so, more detailed information about the study, including written information will be given to the patient. Thereafter the patient will be guided through an informed consent document, which they will sign, and a copy given to them. The original copy will be kept for our records at endoscopy, patients with gastric pathology i.e. gastritis, gastric and/or duodenal ulcers will have biopsies taken, for histopathology (routine practice) and microbiome analysis (research). There is no follow up appointment for the research as this is a cross-sectional study, but the patient will be followed up according to their routine management.

**Data and sample collection**

Clinical data for both cohorts to be collected includes patients' history of (smoking, alcohol use, diet, proton pump inhibitor (PPI) use), and demographics (age, sex). Bio-data i.e. weight, height, endoscopic findings (gastritis, gastric ulcers, mass lesions/location, duodenitis, duodenal ulcer or any abnormal looking area) and histologic findings (density of H. Pylori, degree of inflammatory activity, atrophy, and metaplasia (Sydney grading for chronic gastritis). Other samples that will be collected are:

1. Stored gastric biopsy specimens of GCA or controls to be retrieved for 16SrRNA analysis for the retrospective cohort. Cut sections of the FFPE will be obtained and treated with xylene to remove the paraffin, followed by centrifugation of the FFPE. DNA will be isolated from the FFPE using shotgun metagenomics sequencing. A control of the embedding process constituting a block of paraffin tissue will serve to distinguish the bacteria associated with the paraffin inclusion from the microbial signatures of the tissue.
2. Gastric biopsies for histology (normal clinical care, a total of 5) and additional “research” biopsies for the microbiome for 16SrRNA. Biopsied tissues will be snap-frozen in cryo-tubes in liquid nitrogen instantly after collection and transported in dry ice to the laboratory where it will be stored at – 800C in a designated freezer. The samples will subsequently be processed in batches.
3. Bloods (30ml) for storage for genetic studies and immunological assays in the future.

Blood will be collected as follows: 2.5ml Paxgene DNA tube, 2.5ml Paxgene RNA tube and 25ml ACD/SST tubes. Paxgene DNA and RNA tubes will be stored at -20**°**C. Samples will be batched, and the DNA and RNA will be extracted using QIAGEN PAX gene blood DNA and RNA extraction kits.Extracted samples will be quantified using a Nanodrop Spectrophotometer prior to storage at -80**°**C. Paxgene or equivalent tubes/kits will be used. ACD/SST tubes will be stored at -80**°**C for further studies.

**Data safety and monitoring**

Initial available specific data points will be collected and directly transcribed digitally by the enrolling physician. For each participant, an identification study number will be assigned, and results will be stored in an excel spread sheet. Immediate access to the database will be given only to the specified researchers. Any further research to be performed using this data set will be completed in accordance with guidelines set by ethical approval. The principal investigator will be given password-protected access to the database. Macroscopic endoscopic findings, histology results, and gastric microbiome data will be added as they are available.

**Data analysis**

All data exploration and analysis will be done in Stata (Version 13.1; Stata Corp, College Station, Texas, USA) ®. Descriptive statistics of mean, median, Standard deviation, quartile and interquartile ranges will be used to characterize the sample in terms of history, demographics, biodata, clinical presentation, endoscopic and histologic findings. Frequencies and percentages will be used for categorical variables. Non continuous variables will be represented using bar charts, pie charts and frequency distribution, while for continuous variables, means and (± standard deviations) will be used for normally distributed continuous variables, and- medians and (interquartile ranges) for skewed data. The means will be compared using two-sample t-test or its non-parametric equivalent where appropriate. Categorical variables will be compared using nonparametric tests- Chi-square and scattered box. A cross tabulation will be used to analyse the relationship of the histologic types and outcome. For all, test of significance and the P-values will be calculated. P-values of <0.05 will be considered statistically significant. A multivariate logistic regression model will be used to identify factors associated with GCA. A p<0.05 will be considered statistically significant while 95% confidence intervals will be used to determine the precision of estimates.

## LABORATORY PROCEDURES

**Gastroscopy**

This will be recorded as per usual in terms of macroscopic abnormal descriptions of the gastric mucosa.

Biopsy of abnormal-looking mucosa (gastritis, gastric ulceration or other abnormal endoscopic features) will be done**.**

**Histopathology**

This will be performed according to approved hospital protocols.

**Microbiome testing**

Gastric bacterial DNA from the biopsied specimen will be extracted using specific protocol on an automated platform, the QI symphony SP instrument (QIAGEN, Hombrechtikon, Switzerland) as previously

Described.(16) The V4 hypervariable region of bacterial 16S ribosomal RNA (16S rRNA) will be

sequenced in paired-end modus (2 \* 150 base pair) on MiSeq platform (Illumina, Inc.). The

resulting paired reads will be assembled using PANDAseq v 2.7 to generate an amplicon size

of 250 base pairs.(17) To test for inter-and intra-sequencing run variation, 6 samples will be

sequenced multiple times across and within different sequencing runs for quality control

purposes. Assembled reads will be demultiplexed and processed by the quantitative insights

into microbial ecology (QIIME v1.8.0) pipeline using the default parameters of QIIME.(18)

Chimeric sequences will be identified de novo and reference based and then removed using

usearch 61.(19) The non-chimeric sequences will be then clustered into operational taxonomic

units at 97.0% sequence similarity using a closed reference-based picking approach with

UCLUST software against Greengenes database 13\_8 of bacterial 16S rRNA sequences.(20)

Samples with less than 10,000 reads after quality filtering will be removed from the analysis.

Alpha diversity will be determined after rarefication at a depth of 10,000 reads per sample using

Chao1, phylogenetic diversity whole tree (PD whole tree), Simpson, Shannon, and observed

species indices. A filtering step will be performed such that only operational taxonomic units

with a prevalence. 5% will be included in analyses (R software v2.14.1; http://CRAN.Rproject.

org) based on grouped operational taxonomic units with the same taxonomic assignment at phylum, family, and genus level.

## STUDY MANAGEMENT

## Ethics

Ethical approval will be obtained from the university of Cape Town Human Research Ethics Committee (HREC). Written consent will be obtained from the patients. Confidentiality will be maintained, therefore names and addresses of patient will not be published and data will be secured with a password and the computer will be kept in a locked office.

Approval of the protocol and consent forms will be obtained before any participants are enrolled. Any amendments to the protocol or informed consents will require review and approval from the HREC, before any change is implemented.

### Benefits

### Participants may not benefit directly from the study. However, by participating, the results may benefit other patients in the future by improving our understanding of H. Pylori infection and factors involved in its progression to gastric adenocarcinoma.

### Risks

### The study carries the risks as described in the study consent forms.

Biopsies and blood collection:

1. Biopsies will be collected in accordance with routine clinical care during gastroscopy. The patient’s clinician will explain the risks of biopsy as part of standard clinical procedure as there is no difference between a study directed gastroscopy and a normal symptom directed gastroscopy performed as standard care of practice. Consent for clinical and research-related biopsies will be obtained separately.
2. Should an unlikely adverse event arise that is due to the gastroscopy and additional biopsies the University of Cape Town has insurance to cover all reasonable medical expenses. An application for a generic insurance policy has been requested from the University of Town.
3. Blood collection will equally be done in accordance with the standard of care. It bears negligible risk; however, the procedure and risk will be explained to the patient, and should there be any, it will be covered by the same university of Cape Town insurance policy

**Confidentiality**

Confidentiality will be maintained by omitting the participant’s names and surnames from the data capture sheet. Data specific to ‘patient by patient’ can be accessed from the laboratory by the attending physician in situations when the information is required for best care, as per standard practice for that patient. To protect patient confidentiality:

1. Access to the registry is password-protected and will only be accessed by investigators on this study.
2. Data extracted from the registry to SPSS / Stata® will be anonymous and patient details will only be identifiable from their hospital folder number.

There is no formal data and safety-monitoring plan for this observational study given that there is no therapeutic intervention. Serious clinical adverse events will be documented and reported as required by the applicable ethics committee and hospital management.

### Privacy and confidentiality

### The study team will maintain clinical and laboratory data in a manner that ensures patient confidentiality. All study personnel have passed human subject protection courses. Each participant will be assigned a participant identification number to be used for all study data. Links to the patient ID will be maintained in a link log and stored in a secure password-protected file, only accessible by the principle investigator.

### 

### Reimbursement for Participation

### There will be no reimbursement for participation in the cohort study because those who are attending for their routine clinical procedures will be recruited.

### 

### Research-related Injury

### In the unlikely event that an injury occurs deemed related to the study procedure, the investigator will access the University of Cape Town’s no-fault insurance policy, to cover the treatment of injuries.

**Work plan**

This work will be conducted over a period of 3 years. Recruitment of patients will be done in the first one and half year, while analysis will be done in the last six months of the second year. The last year is for write up of reports/presentation.

**Dissemination of Study Results**

Investigators will present and publish data generated by this study at national and international scientific conferences and in peer-reviewed publications. Study results will be shared locally with colleagues at Groote Schuur Hospital and other institutions.

**BUDGET**

See attached budget (2 pages) for DNA extraction and sequencing. We aim to analyse 10 samples of GCA (5 cases, 5 controls) and 40 samples of H. Pylori in the prospective study (20 cases, 20 controls). Total budget estimated is R 99 401.53

**REFERENCES**

1.Tonkic A, Tonkic M, Lehours P, Mégraud F. Epidemiology and Diagnosis of H elicobacter pylori Infection. Helicobacter. 2012;17:1-8.

2.Malfertheiner P, Megraud F, O'morain C, Gisbert J, Kuipers E, Axon A, et al. Management of Helicobacter pylori infection—the Maastricht V/Florence consensus report. Gut. 2017;66(1):6-30.

3.Mitchell HM. Epidemiology of infection. Helicobacter pylori: physiology and genetics. 2001:7-18.

4.Bardhan PK. Epidemiological features of Helicobacter pylori infection in developing countries. Clinical infectious diseases. 1997;25(5):973-8.

5.Zamani M, Ebrahimtabar F, Zamani V, Miller W, Alizadeh‐Navaei R, Shokri‐Shirvani J, et al. Systematic review with meta‐analysis: the worldwide prevalence of Helicobacter pylori infection. Alimentary pharmacology & therapeutics. 2018;47(7):868-76.

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

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

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

9.Ghoshal UC, Chaturvedi R, Correa P. The enigma of Helicobacter pylori infection and gastric cancer. Indian Journal of Gastroenterology. 2010;29(3):95-100.

10.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.

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

12.Nishino Y, Inoue M, Tsuji I, Wakai K, Nagata C, Mizoue T, et al. Tobacco smoking and gastric cancer 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.

13.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.

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

15.Guo Y, Zhang Y, Gerhard M, Gao J-J, Mejias-Luque R, Zhang L, et al. Effect of Helicobacter pylori on gastrointestinal microbiota: a population-based study in Linqu, a high-risk area of gastric cancer. Gut. 2020;69(9):1598-607.

16.Claassen S, du Toit E, Kaba M, Moodley C, Zar HJ, Nicol MP. A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples. Journal of microbiological methods. 2013;94(2):103-10.

17.Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. BMC bioinformatics. 2012;13(1):1-7.

18.Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nature methods. 2010;7(5):335-6.

19.Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460-1.

20.DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and environmental microbiology. 2006;72(7):5069-72.