

# Comparative Genomics of *Haemophilus ducreyi*; Insight into Genes related to Virulence, Disease and Defense

Rimon Mutsuddi <sup>1\*</sup> and Abdullah All Jaber<sup>1</sup>

<sup>1</sup>Department of Biochemistry and Microbiology,

North South University

Dhaka, Bangladesh

## ABSTRACT

*Haemophilus ducreyi*, a fastidious, gram-negative, coccobacillus bacteria of Pasteurellaceae family causes sexually transmitted disease known as Chancroid which results genital ulcer disease globally characterized by painful sores on the genitalia and also have been associated particularly with facilitating the transmission of human immunodeficiency virus (HIV). Currently, only a single reference strain (AE017143) are available. This study presents comparative analysis of ten draft genomes isolated from Children with Cutaneous Lesions in Vanuatu and Ghana. Computational analysis of draft sequences have triggered the genes related to virulence, disease and diagnosis. This overall function based analysis based on virulence, disease and diagnosis finally showed that during comparison between ten isolates with reference strain, four genes are common in all additional isolates except reference strain. Random analysis among additional ten isolates except reference strain revealed two genes which are common in three isolates but absolutely absent in all other isolates. Those genes were characterized and found to be functioned as Mercuric ion reductase and Mercury resistance operon regulatory protein. Gene gain/loss was also observed among different isolates in association of CRISPRs and prophage due to their prospective contribution in evolutionary change. This study on multiple strains of *Haemophilus ducreyi* revealed the biological system of this bacteria including the genes related to virulence and causation of disease which can be effective novel drug/vaccine targets in future prospects.

**Key Words:** ULCER, HAEMOPHILUS DUCREYI, COMPARATIVE GENOMICS, VIRULENCE

## 1. INTRODUCTION

*Haemophilus ducreyi*, which is referred to as one of the fastidious gram-negative bacterium, is responsible as a causative agent of chancroid. This chancroid is known a genital ulcer disease (GUD). The organism we are studying is usually spread during sexual intercourse through microabrasions, and the disease usually evinces as multiple painful superficial ulcers which is associated with inguinal lymphadenitis [1] As a result of this painful nature of the lesions, it is important to provide immediate treatment and asymptomatic carriage to the patient is therefore uncommon [2]. Not only as causing factor of GUD, *H. ducreyi* has also been found in several recent performed studies where it is revealed that it may be a major cause of chronic skin ulceration in children from developing countries [3-4]. Although to detect antigen, serology and performing genetic amplification methods are sometimes used to diagnose infections with *H. ducreyi*. Those genetic tests have greater sensitivity and they are not widely available. So cultures are currently considered the "gold standard" test. According to previous studies very little is known regarding *H. ducreyi* strains which are responsible for cutaneous lesions in children. For the better understanding of the genetic differences between genital and cutaneous strains of *Haemophilus ducreyi* according to the research from different geographic locations, recently some new studies have been performed which consists of whole-genome sequencing on cutaneous strains isolated in 2014 and 2015 from children in Vanuatu and Ghana during yaws surveys. On the study in Vanuatu and Ghana, researchers conducted the Whole-genome sequencing using the PacBio RSII platform (Pacific Biosciences, Menlo Park, CA, USA) with P6-C4 and P6 v2-C4 chemistry. A single-molecule real-time (SMRT) cell was used by them to sequence each genome, and de novo assembly of the genomes was conducted using the hierarchical genome assembly process (HGAP3, SMRT Analysis version 2.3.0) workflow, which included consensus-polishing using Quiver [5]. In our study those isolated strains from Vanuatu and Ghana were used in the comparative analysis with the reference strain based on the genomic approaches. Several online tools were used in this comparative analysis to identify significant virulence genes which can be the target factors for the development of novel drug or vaccine in future.

## 2. METHODS AND MATERIALS

### 2.1 Sequence Retrieval and Alignment of draft genomes

There is only a single reference strain can be found for this studied bacteria *Haemophilus ducreyi*. The complete sequence of this reference strain *Haemophilus ducreyi* 35000HP was downloaded from European Bioinformatics Database (<http://www.ebi.ac.uk/ena/data/view/AE017143>) under accession number AE017143. For the functional based comparison, the required draft genome sequences which were isolated from children with Cutaneous Lesions in Vanuatu and Ghana have been found and were downloaded from National Centre for Biotechnology Information (NCBI) under the bioproject "PRJNA318024" (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA318024/>) [6]. Then all of the draft genome sequences were submitted for the alignment purpose to Multiple Genome Alignment (MAUVE) [7] which further was necessary for the annotation and phylogenetic analysis study. Like other genome alignment methods, Mauve uses anchoring as a heuristic to speed alignment. Unlike other multiple genome alignment systems, Mauve's anchor selection method relaxes the assumption that the genomes under study are collinear. Instead, Mauve identifies and aligns regions of local collinearity called locally collinear blocks (LCBs). Each locally collinear block is a homologous region of sequence shared by two or more of the genomes under study, and does not contain any rearrangements of homologous sequence. The weight of an LCB provides a measure of confidence that it is a true genome rearrangement rather than a spurious match. By selecting a high minimum weight during alignment, it was successful to identify genome rearrangements that are very likely to exist.

### 2.2 Similarity Comparisons of draft genomes with reference strain

Beside the alignment of the draft sequences, all of those were also used in BLAST Ring Image Generator (BRIG) [8] to find out the similarity among the strains of *H. ducreyi*. It can create an image that demonstrates various prokaryote genome correlations, without a self-assertive utmost on the number of genomes compared. BRIG is fully capable of generating circular comparison images for prokaryote genomes, showing multiple genome comparisons in a single image, and displaying similarity between a reference genome of *H. ducreyi* in the centre against other query isolates as a set of concentric rings colored according to BLAST identity.

### 2.3 Genome Assembly and Annotation

After completing the alignment process of draft sequences, Rapid Annotation using Subsystem Technology (RAST) [9] were used to annotate those sequences for function based comparison of all isolates with reference strain and also randomly among the draft sequences. This annotation server usually functioned as a fully automated service for annotating bacterial and archaeal genomes. The service identifies protein-encoding, rRNA and tRNA genes, assigning functions to the genes and predicts which subsystems are represented in the genome and those information were used to reconstruct the metabolic network. The annotated genome were browsed in an environment named SEED viewer [10] that supports comparative analysis with the annotated genomes.

### 2.4 Identification of prophage regions

The draft sequences additionally were utilized to recognize the prophage regions in the greater part of the arrangements. Bacterial genomes can contain a huge extent (>20%) of functional and non-functional bacteriophage genes. Thus, prophage arrangements can represent a significant fraction of the variety inside bacterial species or clades. The presence of prophage regions may enable a few microorganisms to secure anti-toxin protection and to exist in new ecological niches, those regions may also enable some factors for microorganism to become pathogenic because bacterial genome fragments can also be carried by phage particles and the lytic process is considered to be an important vehicle for horizontal gene transfer. Phage search tool (PHASTER) [11] was used to determine the prophage regions of our studied bacteria.

### 2.5 Identification of CRISPRs regions

All of the draft sequences along with the reference sequence of *H. ducreyi* were also uploaded to CRISPRFinder [12] server to identify the "Clustered regularly interspaced short palindromic repeats"(CRISPRs) regions. Those regions usually constitute a particular family of tandem repeats usually found in a wide range of prokaryotic genomes. The study of the analysis of CRISPR region were performed for identification and observation of the total gene gain or loss which may contribute in the evolutionary change of *Haemophilus ducreyi*.

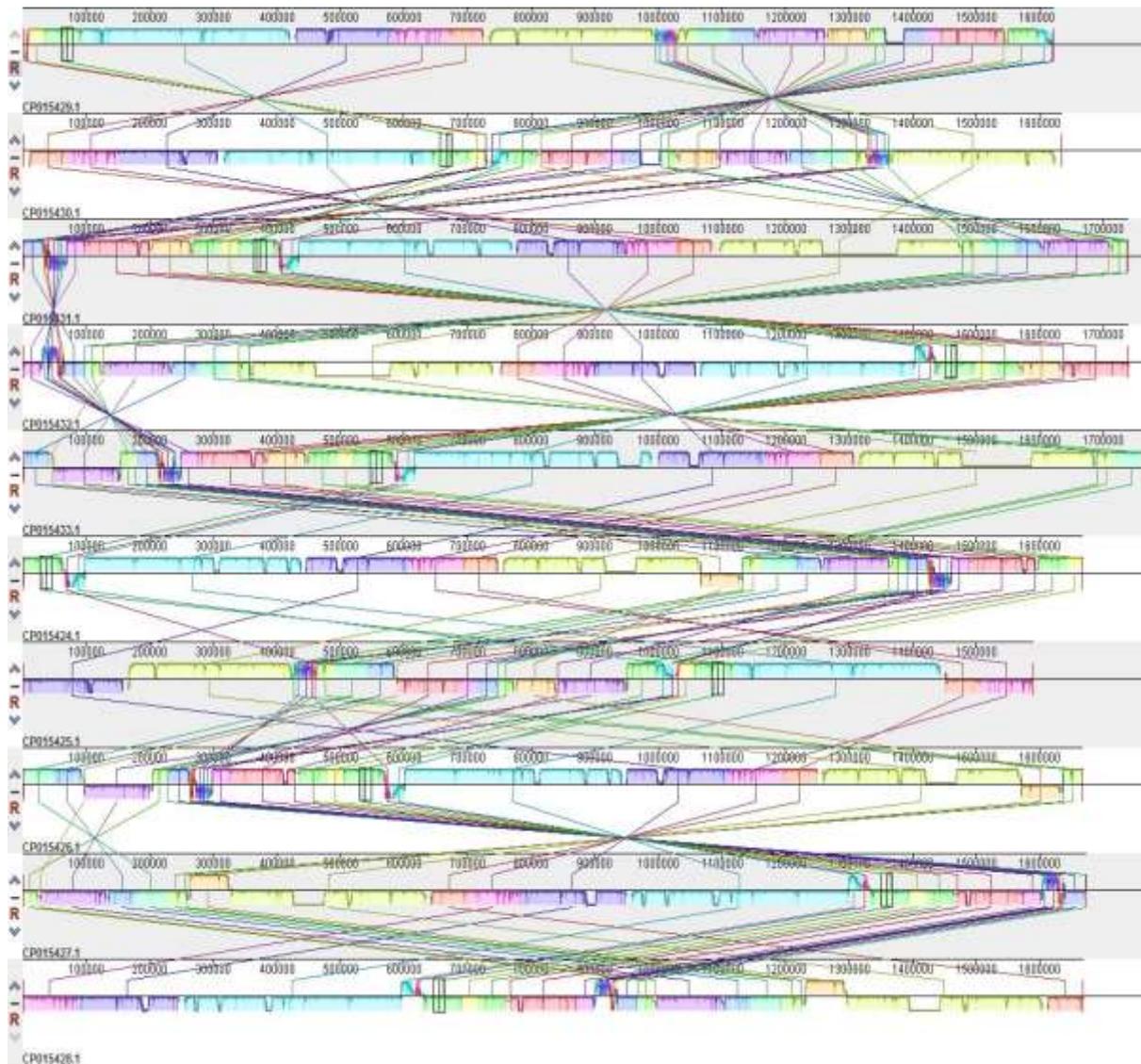
### 2.6 Phylogenetic analysis

In this study, phylogenetic analysis were performed using reference genome to visualize the rearrangement of virulence genes on that particular genome. For this purpose Molecular Evolutionary Genetics Analysis (MEGA) [13] was used to prepare the tree and Interactive tree of life (iTOL) v3 [14] was used to edit the overview of the tree.

### 3. RESULT AND DISCUSSION

#### 3.1 Retrieval of sequences and draft genomes alignment

Sequences of our studied bacterial strains were retrieved and were submitted for the Alignment for further annotation and phylogenetic analysis. MAUVE aligns genomes which identically discounting of their input order by determining multi-MUMs in subsets of the genomes and also calculating guide tree for the process of progressive alignment. The rest of the unaligned regions are often either unvaried, or lineage-specific regions acquiring by horizontal transfer. Our study represents the visualization of the genomes (Figure1) after completion of the alignment process which shows each contiguously colored region which is a locally collinear block, a region without rearrangement of homologous backbone sequence. LCBs below the line of genome's center are put in the reverse complement orientation in relation with the reference genome. Other lines between studied genomes trace each orthologous LCB through every genome. The visualized large gray regions within an LCB denote the presence of lineage-specific sequence at that site.

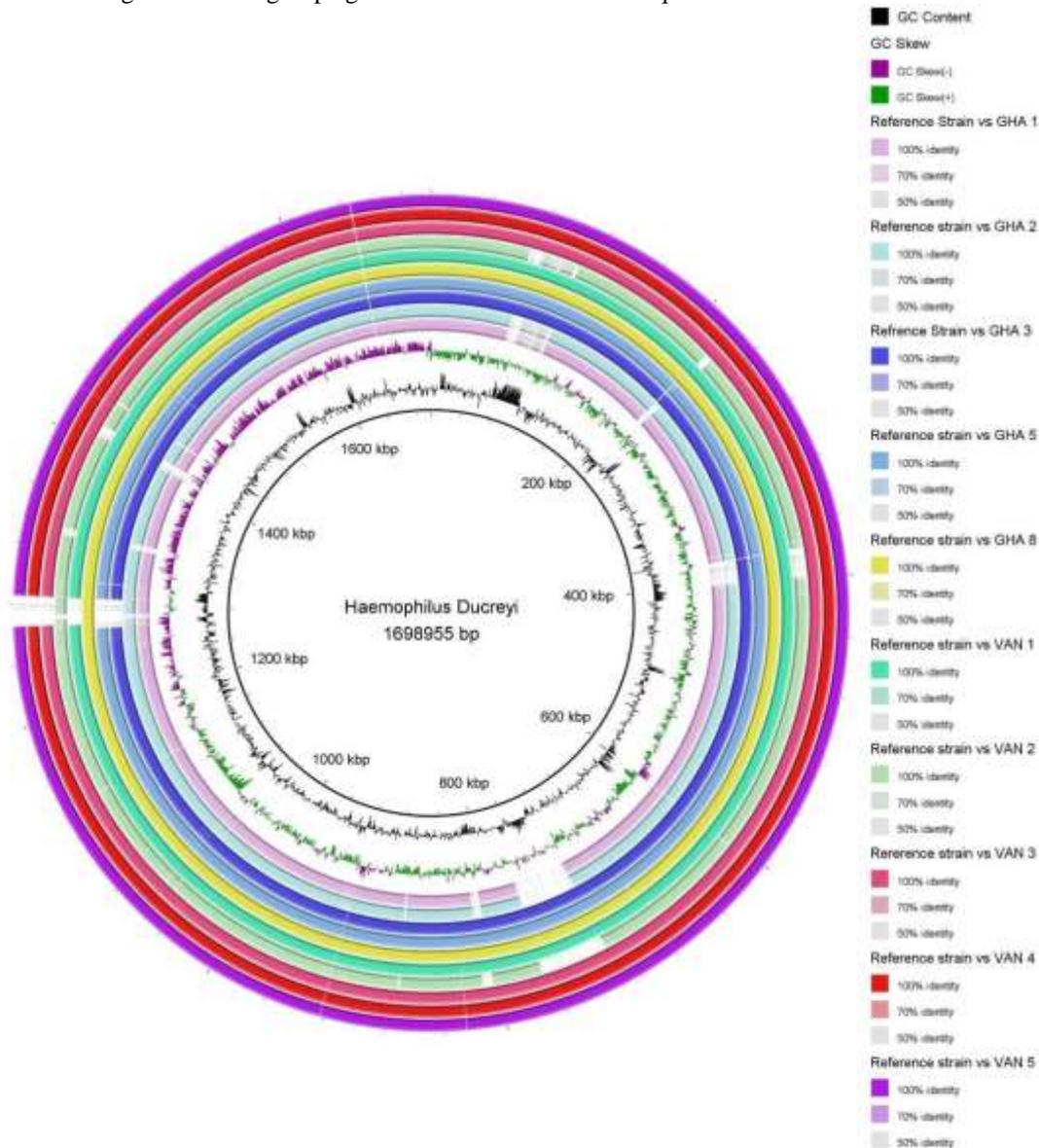


**Figure1: Visualization of the draft genomes after the alignment process.**

Each locally collinear block represents homologous region of sequence shared by ten studied draft genomes. It also involves extracted conserved backbone sequences which is defined as regions of the alignment containing >50 gap-free columns without any stretches of 50 or more consecutive gaps in any of the single genome sequence.

**3.2 Comparison of draft genomes:**

Visualization of genome correlations is important for deciding genotypic contrasts between firmly related prokaryotes. In our study we performed BLAST Ring Image Generator (BRIG) for the comparison of our draft genomes with the reference genome of *H. ducreyi*. The output figure (Figure2) represents the correlation between the reference genome and the multiple draft genomes. The innermost rings show GC skew (purple/green) and GC content (black). The remaining rings shows the blast comparison among the draft genomes against the simulated draft genome assembly. The varying color gradient of rings 4-13 indicates a BLAST match of a particular percentage identity, as shown in the key. BLAST matches was filtered according to a minimum percentage identity or E-value cut-off. These matches are calculated from the perspective of the reference sequence; consequently, regions that are absent from the reference genome but present in one or more of the query sequences will not be displayed. Data from different genomes can be accumulated into a single lane, which usually enables visualization of a large number of genomes and allows users to estimate genomes as a group against the central reference sequence.



**Figure2: Similarity comparison of reference strain *Haemophilus ducreyi* 35000HP with several draft genomes isolated from Vanuatu and Ghana.**

The rings at the furthest in, show GC skew (purple/green) and GC content (black). The rest of the rings represents the blast comparison surrounded by the draft genomes against the simulated draft genome assembly. The differing color gradient of rings 4-13 demonstrates a BLAST match of a particular percentage identity.

**3.3 Annotations and function based comparisons:**

In this study, function based analysis using Rapid Annotation using Subsystem Technology (RAST) were performed to reveal unique genes from the reference strains and all other isolates. After random functional comparisons among all strains, certain number of significant genes have been found which are related to virulence, disease and defense. They are found in the

subcategory of resistance to antibiotics and toxic compound under the specific subsystem. The function of those unique genes are also found in this comparative analysis based on functional characterization. There are in total 1488 genes are found in the analysis of *Haemophilus ducreyi* 3500HP. From these total genes, 34 genes are found which shows virulence function after expression of those genes. After performing this overall function based comparison, we found an overall summary of the virulence genes (Table 1). This overview represents some of the genes which consists of virulence factor. There are eight genes were found which are completely unavailable in the reference strain. From this eight genes, the first four genes were found in all other isolated strains. There are also some genes which functioned as macrolide export ATP binding/ permease protein and macrolide efflux protein (MacA) were found from most of the draft genomes. The other remaining two genes which is referred as the significant genes are functioned as Mercuric reductase and Mercuric operon regulatory protein respectively. Those significant genes were found in just only three isolated strains which are GHA 3, GHA 5 and GHA 8.

**Table 1: Overview of the random function based comparison among the reference strains and all other draft genomes of isolated strains.**

The Mercuric ion Reductase and Mercuric resistance operon regulatory protein are absolutely absent in reference genome and only found in the isolated strains GHA3, GHA5 and GHA8. Two other functional genes named MacB and MacA which are present in most of the draft genomes but completely absent in reference strain *H. ducreyi* 3500HP.

| Category                      | Subcategory   | Subsystem   | Role   | Reference Starin | VAN1 | VAN2 | VAN3 | VAN4 | VAN5 | GHA1 | GHA2 | GHA3 | GHA5 | GHA8 |
|-------------------------------|---|---|--|------------------|------|------|------|------|------|------|------|------|------|------|
| Virulence Disease and Defense | Bacteriocin, ribosomally synthesized antibacterial Peptides | Colicin V and Bacteriocin Production cluster        | DedA protein   | NO               | YES  |
| Virulence Disease and Defense | Resistance to antibiotics and toxic compounds               | Beta-lactamase                                      | Negative Regulator Of Betalactemase  | NO               | YES  |
| Virulence Disease and Defense | Resistance to antibiotics and toxic compounds               | Cobalt-zinc-Cadmium Resistance                      | Probable Co/Zn/Cd efflux system membrane fusion  | NO               | YES  |
| Virulence Disease and Defense | Resistance to antibiotics and toxic compounds               | Multidrug Resistance Efflux Pumps                   | Multi antimicrobial extursion protein (Na <sup>(+)</sup> /drug antiporter),MATE family of MDR efflux pumps | NO               | YES  |
| Virulence Disease and Defense | Toxins and Super Antigen                                    | Cytolethal distending toxin of Campylobacter Jejuni | Cytolethal distending toxin subunit A  | YES              | YES  | NO   | YES  | YES  | YES  | NO   | NO   | YES  | YES  | YES  |
| Virulence Disease and Defense | Toxins and Super Antigen                                    | Cytolethal distending toxin of Campylobacter Jejuni | Cytolethal distending toxin subunit B  | YES              | YES  | NO   | YES  | YES  | YES  | NO   | NO   | YES  | YES  | YES  |
| Virulence Disease and Defense | Toxins And Super Antigen                                    | Cytolethal distending toxin of Campylobacter Jejuni | Cytolethal distending toxin subunit C  | YES              | YES  | NO   | YES  | YES  | YES  | NO   | NO   | YES  | YES  | YES  |
| Virulence Disease and Defense | Resistance to antibiotics and toxic compounds               | Mercuric reductase                                  | Mercuric ion Reductase (EC1.16.1.1)  | NO               | NO   | NO   | NO   | NO   | NO   | NO   | NO   | YES  | YES  | YES  |
| Virulence Disease and Defense | Resistance to antibiotics and toxic compounds               | Mercuric reductase                                  | Mercuric resistance operon regulatory protein  | NO               | NO   | NO   | NO   | NO   | NO   | NO   | NO   | YES  | YES  | YES  |
| Virulence Disease and Defense | Resistance to antibiotics and toxic compounds               | Multidrug Resistance Efflux Pumps                   | Macrolide export ATP binding/ permease protein MacB(EC 3.6.3.)   | NO               | YES  | NO   | YES  | YES  | YES  | NO   | NO   | YES  | YES  | YES  |
| Virulence Disease and Defense | Resistance to antibiotics and toxic compounds               | Multidrug Resistance Efflux Pumps                   | Macrolide specific efflux protein MacA   | NO               | YES  | NO   | YES  | YES  | YES  | NO   | NO   | YES  | YES  | YES  |

### 3.4 Analysis of the prophage region

In the analysis of our studied draft genomes by PHASTER, the prophage region shows that every eleven isolates consists of intact prophage region (Table 2). In this study it also shows that there is an incomplete region in the 35000P isolate which is found on the region position (855164-879548). Another incomplete prophage region is noticed on one of the eleven additional strains named GHA 1. The position of this region is 330-9153. Those some diversity can contribute in the alteration of lysogens to new hosts.

**Table 2: Analysis of the prophage regions of reference strain and all other draft genomes of *H. ducreyi*.**

It represents the details of PHASTER's completeness overview. A prophage region is considered to be incomplete if its completeness score is less than 60 and questionable if the score is between 60 and 90. Intact status appears if the score is above 90. Two incomplete regions were found on the reference genome of *H. ducreyi* 35000HP and GHA1 respectively.

| Strain   | Region | Length | Status       | Total Proteins | Putative package                          | GC %   | Region Position |
|----------|--------|--------|--------------|----------------|---|--------|-----------------|
| 35000 HP | 1      | 31.1Kb | intact       | 54             | PHAGE_Escher_D108_NC_013594 (12)          | 48.19% | 75767-106940    |
|          | 2      | 29.8KB | questionable | 43             | PHAGE_HAEMOP_SuMu_NC_019455(18)           | 42.01% | 382034-411901   |
|          | 3      | 24.3KB | Incomplete   | 27             | PHAGE_Megavi_chiliensis_NC_016072(4)      | 36.51% | 855164-879548   |
|          | 4      | 41.6KB | Intact       | 57             | PHAGE_Pseudo_B3_NC_006548(22)             | 41.15% | 1262080-1303775 |
| VAN 1    | 1      | 38.6KB | Intact       | 54             | PHAGE_Mannhe_vB_MhM_3927AP2_NC_028766(23) | 41.88% | 57402-96022     |
|          | 2      | 35.6KB | Intact       | 58             | PHAGE_Escher_D108_NC_013594 (16)          | 47.54% | 1422531-1458223 |
| VAN 2    | 1      | 34.1KB | Intact       | 54             | PHAGE_Escher_D108_NC_013594 (17)          | 42.74% | 428386-462502   |
|          | 2      | 35.7KB | Intact       | 57             | PHAGE_Pseudo_vB_PaeS_PM105_NC_028667(21)  | 42.12% | 996670-1032448  |
| VAN 3    | 1      | 35.6KB | Intact       | 57             | PHAGE_Escher_D108_NC_013594 (16)          | 47.54% | 259644-295336   |
|          | 2      | 38.6KB | Intact       | 54             | PHAGE_Mannhe_vB_MhM_3927AP2_NC_028766(23) | 41.88% | 561606-600226   |
| VAN 4    | 1      | 38.6KB | Intact       | 54             | PHAGE_Mannhe_vB_MhM_3927AP2_NC_028766(23) | 41.88% | 1298902-1337522 |
|          | 2      | 35.6KB | Intact       | 58             | PHAGE_Escher_D108_NC_013594 (16)          | 47.54% | 1604174-1639866 |
| VAN 5    | 1      | 38.6KB | Intact       | 54             | PHAGE_Mannhe_vB_MhM_3927AP2_NC_028766(23) | 41.88% | 594890-633510   |
|          | 2      | 35.6KB | Intact       | 59             | PHAGE_Escher_D108_NC_013594 (16)          | 47.54% | 900156-935848   |
| GHA 1    | 1      | 8.8KB  | Incomplete   | 18             | PHAGE_Burkho_phiE255_NC_009237(7)         | 41.07% | 330-9153        |
|          | 2      | 34.1KB | Intact       | 54             | PHAGE_Escher_D108_NC_013594 (17)          | 42.74% | 997434-1031550  |
|          | 3      | 25.2KB | Intact       | 35             | PHAGE_Pseudo_vB_PaeS_PM105_NC_028667(16)  | 42.84% | 1595533-1620819 |
| GHA 2    | 1      | 35.7KB | Intact       | 57             | PHAGE_Pseudo_vB_PaeS_PM105_NC_028667(21)  | 42.11% | 726950-762726   |
|          | 2      | 34.1KB | Intact       | 53             | PHAGE_Escher_D108_NC_013594 (17)          | 42.74% | 1326674-1360788 |
| GHA 3    | 1      | 35.8KB | Intact       | 56             | PHAGE_Escher_D108_NC_013594 (16)          | 47.54% | 30543-66347     |
|          | 2      | 35.9KB | Intact       | 52             | PHAGE_Mannhe_vB_MhM_3927AP2_NC_028766(23) | 42.27% | 397144-433048   |
| GHA 5    | 1      | 35.8KB | Intact       | 56             | PHAGE_Escher_D108_NC_013594 (16)          | 47.54% | 30186-55990     |
|          | 2      | 35.9KB | Intact       | 52             | PHAGE_Mannhe_vB_MhM_3927AP2_NC_028766(23) | 42.27% | 1401983-1437887 |
| GHA 8    | 1      | 35.6KB | Intact       | 58             | PHAGE_Escher_D108_NC_013594 (16)          | 47.51% | 209682-245343   |
|          | 2      | 38.6KB | Intact       | 55             | PHAGE_Mannhe_vB_MhM_3927AP2_NC_028766(23) | 41.90% | 576746-615397   |
|          | 3      | 37KB   | Intact       | 55             | PHAGE_Pseudo_vB_PaeS_PM105_NC_028667(24)  | 41.53% | 935338-972356   |

### 3.5 Analysis of the CRISPRs regions

In Archeae and Bacteria, the repeated elements called CRISPRs for "clustered regularly interspaced short palindromic repeats" are believed to participate in the defense against viruses. After submitting any genome in CRISPRFinder, it detects the clusters which are divided into two groups: confirmed CRISPRs (>=3DRs) usually stored in the database and the small questionable clusters (2 or 3 DRs) are analyzed by using blast process of their conserved region (DR) against the approved DRs. Clusters which do not fit

those criteria are usually deleted and the rest are kept as questionable. In the draft genomes of *H. ducreyi*, there were no confirmed CRSPRs regions found. There were just only 13 questionable CRISPR found after analysis all of the genome sequences. Table 3 represents the finding of the possible CRISPR from draft genomes *H. ducreyi*.

**Table 3: Overview of the analysis of CRSPRs regions in the draft genomes of *H. ducreyi*.**

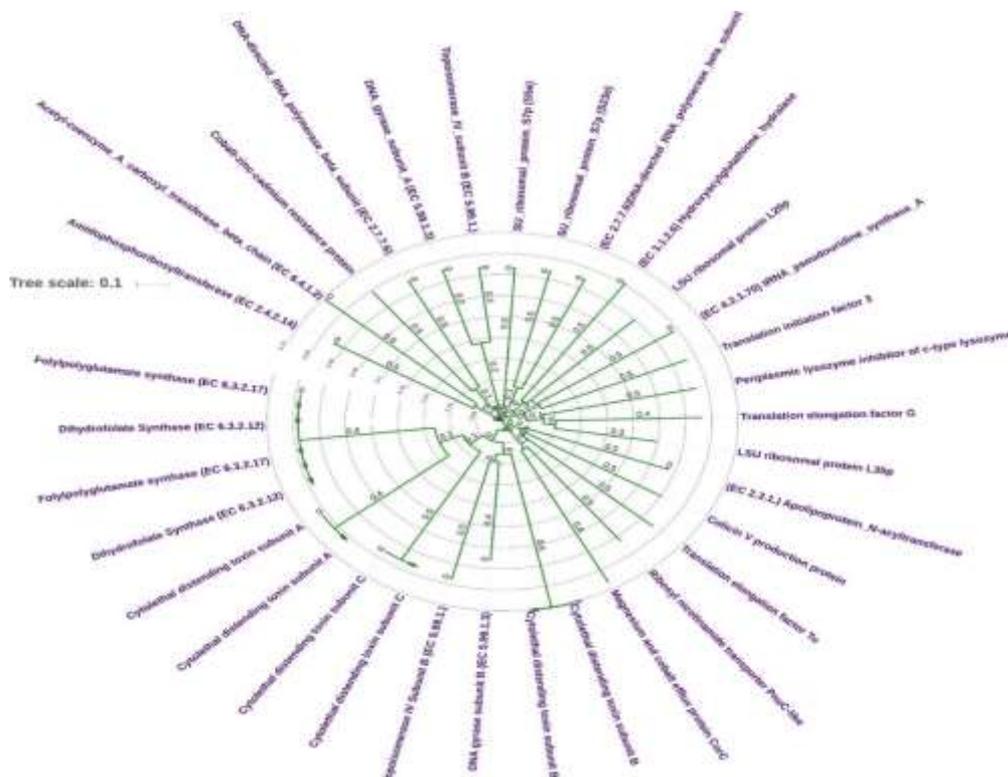
There are 13 questionable CRISPRs were found on several draft genomes of *H. ducreyi* and no confirmed CRISPRs was found in any of the draft genome of our studied bacteria.

- Confirmed CRISPRs = 0
- Questionable CRISPRs = 13

| Strain | CRISPR Start Position | CRISPR end Position | CRISPR Length | DR Consensus            | DR Length | Number of spacers | Status       |
|--------|-----------------------|---------------------|---------------|-------------------------|-----------|-------------------|--------------|
| VAN 1  | 757352...             | 757683...           | 331           | TTAGCTTGGCTGGTAATATTTAA | 23        | 3                 | Questionable |
| VAN 3  | 1261561...            | 1261892...          | 331           | TTAGCTTGGCTGGTAATATTTAA | 23        | 3                 | Questionable |
| VAN 4  | 631665...             | 631748...           | 83            | TTAAATATTACCAGCCAAGGTAA | 23        | 1                 | Questionable |
|        | 631908...             | 631996...           | 88            | TTAAATATTACCAGCCAAGGTAA | 23        | 1                 | Questionable |
|        | 960812...             | 960960...           | 148           | ACATTCTTAGTCACGTTTAAGGT | 23        | 2                 | Questionable |
| VAN 5  | 256795...             | 256943...           | 148           | ACATTCTTAGTCACGTTTAAGGT | 23        | 2                 | Questionable |
|        | 1600697...            | 1600780...          | 83            | TTAAATATTACCAGCCAAGGTAA | 23        | 1                 |              |
|        | 1600940...            | 1601028...          | 88            | TTAAATATTACCAGCCAAGGTAA | 23        | 1                 |              |
| GHA 3  | 1098392...            | 1098723...          | 331           | TTAGCTTGGCTGGTAATATTTAA | 23        | 3                 | Questionable |
| GHA 5  | 736356...             | 736439...           | 83            | TTAAATATTACCAGCCAAGGTAA | 23        | 1                 | Questionable |
|        | 736599...             | 736687...           | 88            | TTAAATATTACCAGCCAAGGTAA | 23        | 1                 |              |
|        | 1068894...            | 1069042...          | 148           | ACATTCTTAGTCACGTTTAAGGT | 23        | 2                 |              |
| GHA 8  | 1317657...            | 1317988...          | 331           | TTAGCTTGGCTGGTAATATTTAA | 23        | 3                 | Questionable |

**3.6 Phylogenetic Analysis**

In our study, phylogenetic analysis were performed in a different approach mainly to visualize the tree of 31 virulence genes arrangement in the draft genome of *Haemophilus ducreyi*. After identification of overall genes in the isolate, protein sequences were submitted to Molecular Evolutionary Genetics Analysis (MEGA) software which gave an overview of the phylogeny of the virulence genes present in *H. ducreyi*. The corresponding genes were labelled according to their functional roles in the tree representation under default tree scale 0.1. (Figure3)



**Figure3: Overview of virulence genes through phylogenetic analysis.**

There were overall 31 functional genes found in the *H. ducreyi* 35000HP that shows virulence properties and the above figure represents their distributions in the genome along with the position and distances calculated by using MEGA .

#### 4. DISCUSSION

After successful completion of the comparative analysis of *Haemophilus ducreyi* reference strain and ten draft genomes of different isolates isolated from children with cutaneous lesions in Ghana and Vanuatu, fruitful result has been found to reveal the biological system of this bacteria and its significant genes which is related to virulence, disease and diagnosis. After the functional based comparison there are some significant virulence genes found which are completely absent in the reference strain of *Haemophilus ducreyi*. There are also two unique genes are found which are absolutely absent in most of the isolates including the reference strain. DedA protein was found after the comparative analysis among all isolates but it is completely absent in reference strain. DedA protein belongs to a family which is highly conserved and ancient family of membrane proteins with significant number of representatives in most of the sequenced genome, including eucarya, archaea and bacteria. Mutants of bacterial DedA family provides such interesting phenotypes as temperature sensitivity, cell division effects, elevated envelop-related stress responses, altered membrane lipid composition and loss of proton motive force [15]. In bacteria like *Borrelia burgdorferi*, *Escherichia coli*, DedA protein family is one of the important component. Some other roles of this DedA protein are also found in multicellular organism. Many DedA genes are annotated “SNARE-associated Golgi protein” [15]. Overall function of this protein has some vital implications, not only for the elucidation of a potential drug target but also for the insight which may be provided into the process of protein evolution. This protein was also found in most of the draft isolates but are not found in the reference strain. This system membrane fusion protein is found under the subsystem of Cobalt-zinc-cadmium resistance. These are part of trans-envelope protein complexes, which seem to detoxify the periplasm by export of toxic metal cations from the periplasm to the outside. RND proteins belonging to a protein family of heavy metal exporters. Together with many efflux systems that detoxify the cytoplasm, regulators and possible metal-binding proteins, RND proteins mediate an efficient defense against transition metal cations. The cobalt-zinc-cadmium resistance determinant is often located on a plasmid. The *czc*, *chr*, *ncc* and *mer* genes that are responsible for resistance to Zn, Cr, Ni and Hg, respectively. Systems for heavy metal efflux are homologous and arose by duplication of a *czc*ICAB core determinant. The exact function of efflux complex can be sometimes predicted from the context of genes localized in proximity to an operon for a metal efflux complex. This protein is also absent in reference strain but are commonly available in all other compared isolates. This protein is found under the subsystem of Multidrug Resistance Efflux Pumps. Bacteria can resist to antibiotics by active exportation mediated by membrane transporters called efflux pumps. These proteins might be specific of a class of antibiotics and also responsible for multidrug resistance (MDR) [16]. Efflux pumps requires energy which can be provided by trans-membrane electrochemical gradient of protons (MFS, RND, SMR families) or sodium ions (MATE family) [17] or by ATP hydrolysis (ABC family). Efflux transporters usually are organized as multicomponent systems in Gram-negative bacteria [18] and in which the efflux pump located in the inner membrane works in conjunction with a periplasmic fusion protein and an outer membrane factor. The most frequently encountered pumps are of the RND-type such as *AcrB* in *Escherichia coli* or *MexB* in *Pseudomonas aeruginosa*. Efflux is solely mediated by the pump protein in Gram-positive bacteria, so described with MFS pumps such as *NorA* or *QacA* in *Staphylococcus aureus* and *PmrA* in *Streptococcus pneumoniae*. Efflux transporters have also been described in mycobacteria. At present it is understood that these efflux pumps also have a physiological role. They can confer resistance to natural substances produced by the host, including bile, hormones and host-defense molecules [19]. Some efflux pumps of the resistance nodulation division (RND) family also have been shown to play role in the colonization and the persistence of bacteria in the host. So it has been confirmed that they have roles in bacterial pathogenicity. Mercuric reductase (*merA*), is referred as oxidoreductase enzyme also flavoprotein functioned as catalysis to reduce  $Hg^{2+}$  to  $Hg^0$ . A significant output for further drug target for the following mercuric ion reductase is present in only three isolates in order to compare overall outcomes. Many eubacteria which are found mostly in the cytoplasm, serve to convert toxic mercury ions into relatively inert elemental mercury [20]. *MerA* is a broadly distributed system which plays a critical role in biogeochemistry and mercury detoxification [20]. Also *MerA* is an excellent early warning signal for  $Zn^{2+}$  inhibition which indicates during *MerA* exposure to non-inhibitory  $Zn^{2+}$  levels [21]. Mercuric mercury ( $Hg(II)$ ) is a highly toxic and mobile element in nature. Due to its high affinity for protein sulfhydryl groups and having an adverse effect on biology since Earth's oxygenation approx 2.4 billion years ago, which destabilize protein structure and decrease enzyme activity, finally decreased the organismal fitness. *Mer* operon consists of the homodimeric flavin-dependent disulfide oxidoreductase enzyme mercuric reductase (*MerA*). The  $Hg$  resistance system is encoded for by the *mer* operon also encode for organomercury lyase (*MerB*), a periplasmic  $Hg(II)$  scavenging protein (*MerP*), two regulatory proteins (*MerR*, *MerD*) and more inner membrane spanning proteins (*MerT*, *MerC*, *MerE*, *MerF*, *MerG*) that transport  $Hg(II)$  to the cytoplasm where it is reduced by *MerA*. Overall expression of *mer* is modulate by one of regulatory proteins *MerR*, which acts as a transcriptional activator or repressor in the presence or absence of  $Hg(II)$ , respectively [22]. *MerR* operon is largely passed around on a few plasmids and sequenced them to have the most intact clusters of the mercury resistance proteins. The viruses that have integrated their genomes into the genome of a bacterial host are termed as prophage. Prophage genome condition may vary from fully intact to form different infective particles to a remnant state where only a few phases of genes persist [23]. After analyzing the reference sequences of *Haemophilus ducreyi* and draft sequences, many prophage regions and significant variations were found which ensured by the result of recognition of intact,

incomplete and some questionable regions. There are no confirmed CRISPR found after performing the analysis of draft sequences. Only some expected regions found which can have significant effect in the evolution of the *Haemophilus ducreyi*.

## 5. CONCLUSION

After the overall studies using multiple tools for functional based comparison among the reference strains and all other strains, there are finally some significant virulence genes have been found as described. The identification of prophage region and the identification of the CRISPR regions also reveals some significant factors as well. The detailed studies related with those significant genes can also be performed in future perspective which may show a positive way to decrease the severity of the disease caused by this *Haemophilus ducreyi* bacterium. After successful comparison of multiple strains of *Haemophilus ducreyi* using variety of genomic approaches, the biological system of this bacteria has been revealed which also represents those genes related to virulence and causation of disease and those can be used further as diagnosis purpose as well. In addition, the comparative analysis based on all other functional categories will be performed as future aspects to better understand the epidemiology, pathogenesis, evolution, and prevention of *H. ducreyi*-associated CU and to reveal the whole biological system of this bacterium which can be effective way to prevent the associated diseases caused by this causative agent.

## ACKNOWLEDGMENT

We would like to thank every professors and research staff of our department for their valuable guidelines and constant support to fulfil our research work.

## REFERENCES

1. DA L (2003) Chancroid: clinical manifestations, diagnosis, and management. *Sex Transm Infect* 79:68–71. <https://doi:10.1136/sti.79.1.68>
2. Schmid G, Steen R, N'Dowa F (2005) Control of bacterial sexually transmitted diseases in the developing world is possible. *Clin Infect Dis* 41:1313–5. <https://doi:10.1086/496987>
3. Marks M, Chi K-H, Vahi V, et al (2014) *Haemophilus ducreyi* associated with skin ulcers among children, Solomon Islands. *Emerg Infect Dis* 20:1705–7. <https://doi:10.3201/eid2010.140573>
4. Mitjà O, Houinei W, Moses P, et al (2015) Mass Treatment with Single-Dose Azithromycin for Yaws. *N Engl J Med* 372:703–710. <https://doi:10.1056/NEJMoa1408586>
5. Chin CS, Alexander DH, Marks P, et al (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi:10.1038/nmeth.2474>
6. Pillay A, Katz SS, Abrams AJ, et al (2016) Complete Genome Sequences of 11 *Haemophilus ducreyi* Isolates from Children with Cutaneous Lesions in Vanuatu and Ghana. *Genome Announc* 4:e00459-16. <https://doi:10.1128/genomeA.00459-16>
7. Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi:10.1101/gr.2289704>
8. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12(1):402. <https://doi:10.1186/1471-2164-12-402>
9. Aziz RK, Bartels D, Best A, et al (2008) The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics* 9 <https://doi:10.1186/1471-2164-9-75>
10. Aziz RK, Devoid S, Disz T, et al (2012) SEED Servers: High-Performance Access to the SEED Genomes, Annotations, and Metabolic Models. *PLoS One* 7. <https://doi:10.1371/journal.pone.0048053>
11. Arndt D, Grant JR, Marcu A, et al (2016) PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44:W16–W21. <https://doi:10.1093/nar/gkw387>
12. Grissa I, Vergnaud G, Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8. <https://doi:10.1186/1471-2105-8-172>
13. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9:299–306. <https://doi:10.1093/bib/bbn017>
14. Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44:W242–W245. <https://doi:10.1093/nar/gkw290>
15. Doerrler WT, Sikdar R, Kumar S, Boughner LA (2013) New functions for the ancient DedA membrane protein family. *J Bacteriol* 195:3–11. <https://doi:10.1128/JB.01006-12>

16. Saier MH, Paulsen IT (2001). Phylogeny of multidrug transporters. *Semin Cell Dev Biol* [Internet];12(3):205–13. [https://10.1006/scdb.2000.0246](https://doi.org/10.1006/scdb.2000.0246)
17. Kuroda T, Tsuchiya T (2009). Multidrug efflux transporters in the MATE family. *Biochimica et Biophysica Acta - Proteins and Proteomics*. Vol. 1794: p. 763–8. [https://10.1016/j.bbapap.2008.11.012](https://doi.org/10.1016/j.bbapap.2008.11.012)
18. Cattoir, Vincent.(2005). [Efflux-mediated antibiotics resistance in bacteria]. *Pathologiebiologie*. 52:607-16. [https://10.1016/j.patbio.2004.09.001](https://doi.org/10.1016/j.patbio.2004.09.001)
19. Piddock LJ (2006). Multidrug-resistance efflux pumps - not just for resistance. *Nat Rev Microbiol*. Aug;4(8):629-36. [https://10.1038/nrmicro1464](https://doi.org/10.1038/nrmicro1464)
20. Barkay, T; Miller, SM; Summers, AO (2003). "Bacterial mercury resistance from atoms to ecosystems.". *FEMS Microbiology Reviews*. 27 (2–3): 355–84. [https://10.1016/S0168-6445\(03\)00046-9](https://doi.org/10.1016/S0168-6445(03)00046-9)
21. Radniecki, Tyler & Semprini, Lewis & E Dolan ((2009). Expression of merA, amoA and hao in Continuously Cultured *Nitrosomonas europaea* Cells Exposed to Zinc Chloride Additions. *Biotechnology and bioengineering*. Mark. 102. 546-53. [https://10.1002/bit.22069](https://doi.org/10.1002/bit.22069)
22. Boyd ES, Barkay T (2012) The mercury resistance operon: From an origin in a geothermal environment to an efficient detoxification machine. *Front Microbiol* 3. [https://doi: 10.3389/fmicb.2012.00349](https://doi.org/10.3389/fmicb.2012.00349)
23. Hertel R, Rodríguez DP, Hollensteiner J, et al (2015) Genome-based identification of active prophage regions by next generation sequencing in *Bacillus licheniformis* DSM13. *PLoS One* 10. [https://doi: 10.1371/journal.pone.0120759](https://doi.org/10.1371/journal.pone.0120759)