**Assessing the relative functions of two putative collagenase genes extracted from a clinically relevant vaginal microbe**

1. **Introduction**

The human body possesses a complex network of microorganisms which play vital roles in defining and managing health and immunity1,2. In 2008, the United States National Institutes of Health (NIH) established the Human Microbiome Project (HMP) to obtain a deeper understanding of the genetic and functional characteristics of human-associated microbes1. The project evolved in 2014 to become the Integrative Human Microbiome Project (iHMP)2. Unlike phase one (HMP) which consisted of developing DNA datasets for large-scale computational analyses, phase two (iHMP) focuses on understanding the link between the human microbiome and health. Lab affiliates of the Vaginal Microbiome Consortium (VMC) at Virginia Commonwealth University contribute to these efforts by designing studies that attempt to identify microbial profiles in the vagina associated with adverse gynecological conditions such as Bacterial vaginosis (BV) and preterm birth.

In a recent study from the VMC, Fettweis et al (2014) characterizes *Candidatus* Mycoplasma gerirdii as a clinically relevant vaginal microbe found almost exclusively in women infected with *Trichomonas vaginalis* (the causative agent of Trichomoniasis). Due in part to its novelty, attempts to cultivate the bacteria in culture have proven unsuccessful. Thus far, computational analyses of Ca. Mycoplasma gerirdii’s genetic characteristics have been sufficient in providing vital information on the microbe’s metabolic strategies and environmental necessities3.

Genomic analyses of *Ca*. Mycoplasma gerirdii strain VCU-M1 indicate the presence of two putative collagenase genes belonging to the U32 peptidase family3. Sequence alignments to other U32 enzymes revealed low sequence homology (data not shown). Considering the intrinsic role of collagenases and other metal metalloproteinases in facilitating the onset of child labor4,5, this proposal seeks to determine the function of the proteins encoded by these genes. The execution of the following experiment may permit a deeper understanding of the potential clinical risks associated with Ca. Mycoplasma gerirdii and other vaginal microbes which exhibit collagenase activity.

1. **Experiment**

This experiment aims to determine if the open reading frames (ORF) identified in the Ca. Mycoplasma girerdii strain VCU-M1 encode enzymes which actively catalyze the breakdown of collagen. Despite both ORFs (*Ca.* Mycoplasma gerirdii ORF55 and ORF56) mapping to the U32 peptidase family and originating from the same microbe, pairwise alignments of protein sequences reveal only 17.5% percent homology. To determine the function and potential catalytic activity of both ORFs, each putative gene will be synthesized, cloned, and analyzed separately. Functional comparisons will be made only after results have been obtained for both genes.

1. *Genetic Synthesis*

*Ca.* Mycoplasma gerirdii utilizes UGA as a codon for tryptophan3. This differs from other bacteria which commonly use UGA to signal stops in translation6. To minimize translational errors incurred from these differences (i.e. early stops in sequence translation), genetic synthesis technology was employed. The GENEWIZ’s [PriorityGENE](https://web.genewiz.com/pg-40off-sb) synthesis service allows for the creation of codon optimized sequences. In a 2016 article by Zhou et al, researchers found that codon usage or codon bias impacts protein expression significantly7. Due to the degeneracy of the genetic code, a single amino acid can be encoded using multiple codons. Codon usage refers to the pattern and frequency in which organisms select different codons for use when synthesizing amino acids. The Zhou et al (2016) study found that genes containing frequently used codons showed higher levels of expression7. To maximize gene synthesis during cloning, GENEWIZ replaces codons present in the original sequences with those commonly used by *E. coli* and other major host organisms.

Once the genes are optimized and an order has been placed, GENEWIZ begins synthesizing the target genes by first designing single-stranded oligonucleotides which correspond to different regions of a full gene. These oligonucleotides function as building-blocks which, when assembled, create one continuous gene. Once the gene is created, it is inserted into the pUC57 plasmid vector and cloned within a host organism to produce up to 5 μL of the synthesized gene. Following this process, the gene inserts are checked for accuracy and quality using Sanger Sequencing. Once quality control protocols are completed, the synthesized genes are prepared for shipment. Included in the delivered package are the synthesized genes incorporated into pUC57 plasmids, a Certificate of Analysis (COA) which provides information about the plasmids and restriction digests.

1. *Cloning*



As its name suggests, the pFN29A His6HaloTag® T7 Flexi® plasmid consists of a T7 promoter which enables high gene expression and product, a N-terminal His6HaloTag® region which allows target protein purification, a TEV protease site for cleavage of the expressed protein from poly-His tag using a HaloTEV Protease, a barnase gene (replaced with gene insert), an ampicillin-resistance gene for selection of the plasmid, and a rrnB transcription terminator (see Figure 1).

To prepare the genes for insertion into the pFN29A His6HaloTag® T7 Flexi® plasmid vector, forward and reverse PCR primers were designed to append the SgfI and PmeI restriction sites (underlined in table) to the amplified PCR product (refer to Figure 2). Table 1 below provides all the primer sequences for both genes with restriction sites underlined. Following amplification of the target genes, PCR products were purified in 2 stages. The first stage purified all genetic material present in the PCR reaction, this process was achieved using spin column, several wash cycles and DNA elution buffer. The second stage separated plasmids from amplified genes by running the purified DNA from stage one on 0.8% agarose and using a scalpel to excise the band corresponding to the target gene directly from the gel. Once removed, the gel was dissolved, and the target genes were eluted using a method like the one used in stage one.



|  |  |  |
| --- | --- | --- |
| **Type** | **Primer Name** | **Sequence (5’ → 3’)** |
| Forward | ORF55\_F1 | NNNGCGATCGCNATGAAAATTATTGTTAGTCC |
| Forward | ORF56\_F2 | NNNGCGATCGCUNATGAGTAGAAGAATAGAATT |
| Reverse | ORF55\_R1 | NNNGTTTAAACNTCCTCTTTTATTACTCATC |
| Reverse | ORF56\_R2 | NNNGTTTAAACNCCACGCTTAATAATTAATT |

**Table 1.** Primer Designs. Forward primers contain the SgfI restriction site (5’- GCGATCGC - 3’) and reverse primers contain the PmeI site (5’- GTTTAAAC - 3’).

Once the target genes are purified, they are digested with Flexi® Enzyme Blend (SgfI & PmeI) to produce sticky ends, then ligated into pFN29A His6HaloTag® T7 Flexi® plasmids. Sanger sequencing is used to verify the accuracy of the insert. JM109 Competent Cells (*E. coli* host organism which contains a T7 RNA Polymerase gene) are fed the plasmid vectors using a heat shock method. The plasmid-exposed cells are grown on ampicillin treated agar plates where only cells containing the plasmid can survive. JM109 cells containing the Flexi® plasmid vector begin constitutively expressing the target gene upon the activation of their T7 RNA Polymerase gene with Isopropyl β-D-1-thiogalactopyranoside (IPTG).

Due to the presence of the N-terminal His6HaloTag® region in the Flexi® plasmid vector, all synthesized proteins are labeled with a poly-Histidine tag. These tags enable the easy purification of target proteins. 

Purification occurs when nitrogen atoms in the poly-His region of the target protein form ionic bonds with nickel ions magnetically bound to the reaction tube. When this interaction occurs, tagged proteins are retained in the tube while non-tagged molecules are washed away.

****

**Figure 4.** Removal of His6HaloTag® tags from protein of interest.

The purified proteins still contain their His6HaloTag® . To remove this tag, the proteins must be treated with TEV Protease. The enzyme functions by cleaving the connective sequence linking the target protein to the tag.

****

**Figure 5.** Fluorescently tagged substrates when exposed to enzyme.

Collagenase activity was determined using the EnzChek Gelatinase/Collagenase Assay Kit. The package came with a variety of fluorescently tagged substrates including collagen type I from bovine skin, collagen type III from human placenta, and gelatin from pig skin. The assay works by measuring changes in fluorescence. As a substrate is broken down into smaller fragments by a collagenase enzyme, more light is emitted. Using this method, collagenase activity is directly linked to the amount of fluorescence emitted from a given substrate after exposure to a proteolytic enzyme (i.e. high fluorescence corresponds to high enzymatic activity). Fluorescent activity can be measured and compared between the two ORFs using a fluorescence microplate reader. (Figure 6 displays sample results obtained from reader).

**III. Sample Results and Discussion**

**Figure 6.** Sample results from collagenase activity assay.

 In a 2004 article by Reza & Ashraf7, researchers determined that collagenase enzymes produced by *Prevotella bivia*, a relatively rare vaginal microbe, played a role in the induction of premature rupture of membranes (PROM) in female patients who experienced preterm birth. Considering that collagen is a major component of the amniotic membrane and human cervical cells induce labor by secreting collagenases (referred to as matrix-metalloproteinases), the current belief is that collagenase activity in vaginal microbiota plays a role in the induction of early labor in women. If collagenase activity is detected, the results of this experiment may ease some of the enigma surrounding the *Ca.* Mycoplasma gerirdii’s virulence as well as open some doors to larger-scale studies. If not, further steps or revisions to the experiment may need to be taken to ascertain the function of the genes or identify new genes of interest.

**Bibliography**

1. The NIH HMP Working Group (2009). Genome Res. 19(12): 2317–2323. doi: 10.1101/gr.096651.109
2. Integrative HMP (iHMP) Research Network Consortium (2014), Cell Host Microbe. 16(3):276-89. doi: 10.1016/j.chom.2014.08.014.
3. Fettweis et. al (2014). PLoS One. 9(10): e110943. Published online 2014 Oct 22. doi: 0.1371/journal.pone.0110943
4. Hampson, V., Liu, D., Billett, E., & Kirk, S. (1997).. BJOG: An International Journal of Obstetrics and Gynaecology, 104(9), 1087-1091. doi:10.1111/j.1471-0528.1997.tb12073.x
5. Rajabi, M. R., Dean, D. D., Beydoun, S. N., & Woessner, J. F. (1988). American Journal of Obstetrics and Gynecology, 159(4), 971-976. doi:10.1016/s0002-9378(88)80183-2.
6. Jones, O. W., & Nirenberg, M. W. (1962). Proceedings of the National Academy of Sciences, 48(12), 2115-2123. doi:10.1073/pnas.48.12.2115.
7. Zhou, Z., et al. 2016. Proceedings of the National Academy of Sciences of the United States of America. Published online before print September 26, 2016, doi: 10.1073/pnas.1606724113
8. Reza HD & Ashraf MM (2004). Medical Journal of the Islamic Republic of Iran. 18(1):61-66.