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

1. **Introduction**

The vaginal microbiome (VM) is believed to play a vital role in modulating a woman’s risk for preterm birth1. Preterm premature rupture of membranes (PPROM) is a pregnancy complication that occurs when the amniotic sac is punctured or degraded prior to 37 weeks of gestation. When this occurs, a woman is at an increased risk for preterm delivery and the health and survival of her unborn child becomes greatly compromised1,2. 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 PPROM 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 women3,4.

A recently characterized vaginal microbe, referred to as “*Candidatus* Mycoplasma girerdii”, has become a topic of interest due to its tight association with the sexually transmitted pathogen *Trichomonas vaginalis*2*.* Due to its obscurity and unresolved invasive properties, attempts to isolate the microbe in culture have proven unsuccessful. Thus far, computational analyses of “*Ca.* Mycoplasma gerirdii” have been sufficient in providing vital information on the microbe’s metabolic strategies and environmental necessities2. These analyses revealed the presence of two putative collagenase genes belonging to the U32 peptidase family3. The presence of these genes suggest “*Ca.* Mycoplasma girerdii” may have a pathogenic function associated with adverse reproductive outcomes.

1. **Experiment**

This experiment aims to determine if two genes identified in the “*Ca.* Mycoplasma girerdii” strain VCU-M1 encode for enzymes which actively catalyze the breakdown of collagen. Despite both genes mapping to the U32 peptidase family and originating from the same microbe, pairwise alignments of both protein sequences revealed only 17.5% percent homology. To determine the relative function and potential catalytic activity of both genes, all steps within the following experiment will be executed for both sequences separately. Comparisons will be made only after results have been obtained for both genes.

1. *Gene Synthesis*



**Figure 1.** DNA construct of synthesized genes.

*“Ca.* Mycoplasma girerdii” utilizes UGA as a codon for tryptophan2. 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), artificial gene synthesis technology will be employed.

In a 2016 article by Zhou et al, researchers found that species-specific codon bias impacts protein expression significantly6. Due to the degeneracy of the genetic code, a single amino acid can be encoded using multiple codons5. Codon bias 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 expression6. Considering these results, a codon optimization software will be used to replace codons present in the original sequences with those commonly used by *E. coli* and other major host organisms.

**

**Figure 2.** Gene synthesis service procedure. Image obtained from [www.geneuniversal.com/](http://www.geneuniversal.com/)

Once the genes are optimized and an order has been placed, target genes are synthesized by designing single-stranded oligonucleotides. These oligonucleotides function as building-blocks which, when assembled, create one continuous gene. Once the gene is created, it is inserted into a plasmid vector and cloned within a host organism to produce copies of synthesized gene. Following this process, the gene inserts are checked for accuracy and quality using sequencing technology and prepared for shipment. The final synthesized genes will consist of the target sequences flanked by highly-specific SgfI and PmeI restriction sites (refer to Figure 1).

1. *Cloning*

|  |  |  |
| --- | --- | --- |
| **Type** | **Primer Name** | **Sequence (5’ → 3’)** |
| Forward | ORF55\_F1 | NNNGCGATCGCN*ATGAAAATTATTGTTAGTCC* |
| Forward | ORF56\_F2 | NNNGCGATCGCUN*ATGAGTAGAAGAATAGAATT* |
| Reverse | ORF55\_R1 | NNNGTTTAAACN*TCCTCTTTTATTACTCATC* |
| Reverse | ORF56\_R2 | NNNGTTTAAACN*CCACGCTTAATAATTAATT* |

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

To prepare the genes for insertion into the recipient plasmid vector, forward and reverse PCR primers will be designed to amplify the target genes with their restriction sites. Each primer will consist of an additional sequence Table 1 below provides all the primer sequences for both genes with restriction sites underlined and nucleotides belonging to the original gene sequences italicized. PCR amplification will occur in three stages described as follows:

* 1. *Denaturation.* The temperature is set to 94°C. Under these conditions, hydrogen bonds between the plasmid’s template and parent strands break, allowing primers in the solution to interact with the individual DNA strands comprising the carrier plasmid vector.
	2. *Primer annealing*. The temperature is lowered to 55°C to enable the annealing of primers to their complementary binding sequence. Once primers have annealed to the target gene, Taq DNA polymerase is able to bind and polymerization is initiated.
	3. *Elongation by Taq Polymerase.* In this step, copies of the target gene are synthesized. The temperature is raised to 74°C to permit optimal Taq DNA Polymerase activity.

Following amplification of the target genes, PCR products must be purified in two stages. The first stage requires the purification of all genetic material present in the PCR reaction and the second requires the extraction of the target genes from all other genetic material. Using a standard DNA purification process, the PCR products will be mixed with binding buffer to encourage binding of DNA to silica beads within to the elution column. Once the DNA and buffer mixture are combined, they will be centrifuged through the elution column. Following this process, nucleic acids will be retained inside the elution column, while all other molecular material (i.e. protein fragments, impurities, etc.) are washed away. Once the lysate is collected, the bound DNA is eluted from the column using elution buffer. The elution buffer functions by disrupting intermolecular interactions between negatively charged DNA and positively charged chaotropic salts adhered to the silica beads. In stage two, the target genes are isolated by running the purified DNA sample from stage one on a 0.8% agarose gel. During this process, negatively-charged DNA will migrate towards the positively charged electrode at the end of the gel well. This movement is driven by an electrochemical gradient formed by loading the negatively charged DNA samples on one side and the electrode on the other. The rate and distance of migration by DNA is limited by size in which shorter fragments are able to migrate farther while larger ones are unable. The band containing the amplified target genes should be ~1 kB in size. After identifying a band corresponding to this size (band size can be recognized only after loading a ladder or control alongside the DNA sample), a scalpel will be used to excise the band directly from the gel. Once removed, the gel is dissolved, and the target genes will be eluted using the DNA purification method described in stage one.



**Figure 3.** Expression vector used for host cell transformation.

Once the target genes are purified, they are digested using a proprietary blend of restriction enzymes (SgfI & PmeI) to produce sticky ends. After the creation of these ends, the target genes are ligated into recipient plasmids. The expression vector used for cloning will consists of a T7 promoter which enables high gene expression, a N-terminal His6Halo region which allows for target protein purification, a TEV protease site for cleavage of the expressed protein from poly-His tag using a HaloTEV Protease, and an ampicillin-resistance gene for selection. Once inserted, sequencing technology will be used to verify the accuracy of the insert (this step will be done through a sequencing facility). One the gene inserts are confirmed, JM109 Competent Cells (*E. coli* host organism which contains a T7 RNA Polymerase gene) are transformed with the expression 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. This survival is conferred by the presence of the ampicillin resistance gene (*AmpR)* in our expression vector.

Following incubation, and the subsequent establishment of plasmid-carrying JM109 colonies, target gene expression will be induced upon the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to the growing culture. The addition of IPTG derepresses the expression of the host organism’s T7 RNA Polymerase gene. The resulting binding of T7 RNA polymerase to the T7 promoter of the inserted plasmid enables the steady expression of the target gene.

1. *Protein Purification*



**Figure 4.** Protein purification process. Image modified from Aulakh S., Kim K.H., Paetzel M. (2015) Expression and Purification of the Individual Bam Components BamB–E. In: Buchanan S., Noinaj N. (eds) The BAM Complex. Methods in Molecular Biology, vol 1329. Humana Press, New York, NY

Once an even growth of cells is observed, the cells are scraped from their plate and suspended in washing buffer. The suspended cells are then centrifuged until and cell pellet is formed. The supernatant (fluid around pellet) is removed, and the pellet is resuspended in lysis buffer. Lysis buffers function by degrading the cell membrane of host cells. The breakdown on the membrane causes cellular lysis which results in the formation of a protein-rich lysate. This lysate contains a mixture of proteins and enzymes, most of which are not the desired target protein.

Due to the presence of the N-terminal His6Halo region in the plasmid vector (refer to Figure 3), all synthesized proteins are labeled with a poly-Histidine tag (these tags are present at the N-terminus of the folded proteins). Protein purification occurs when nitrogen atoms in the poly-His region of the target protein form ionic bonds with nickel coated metallic beads bound to the side of a magnetic reaction tube. When this interaction occurs, tagged proteins are retained in the tube while non-tagged molecules are washed away (refer to figure 4). The tagged proteins are later eluted away. While purified (separated from non-tagged proteins), the proteins will still possess their His6Halo tags. To remove the tags, they must be treated with TEV Protease. The enzyme functions by cleaving the connective sequence linking the target protein to the tag. Once the tags are removed the proteins can be purified by running the resulting solution (purified protein and cleaved histidine tags) through the nickel column once more. During this step, the newly cleaved histidine tags will bind to the nickel ions while the purified protein will be able to flow through.



**Figure 5.** Removal of His6Halo tags from protein of interest.

****

**Figure 6.** Fluorescently tagged substrates when exposed to enzyme.

Once the purified protein is obtained, collagenase activity can be determined using the EnzChek Gelatinase/Collagenase Assay Kit(). The substrates provided in the kit are modified to contain numerous fluorescein groups within their peptide chains. Fluorescent activity from these fluorescein groups is controlled by the presence of quencher sequences. When these quencher sites are adjacent to fluorescein groups, fluorescence is limited. When the substrate is treated with an active enzyme, fluorescence activity increases due to the separation of the fluorescent groups from their respective quencher sequences. 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). Once this interaction has occurred (i.e. treat EnzChek substrates with purified proteolytic enzyme) fluorescent activity can be measured and analyzed using a fluorescence microplate reader.

**III. Discussion**

Sample results of the collagenase activity assay consist of graphs containing peak enzymatic activity (i.e. fluorescence) for each gene when given a different substrate. Through this process, results for enzymatic activity and substrate specificity can be obtained from both genes and compared. If collagenase activity is detected in both genes, the results of this experiment may open some doors to larger-scale studies focused on exploring the presumed link between “*Ca.* Mycoplasma girerdii”. If collagenase activity is not observed, it would be important to continue testing for other potential virulence factors. An alternate option available would be to modify the current experiment so that rather than analyzing the genes separately, both are combined to form an operon-like construct as seen in the yrpAB operon of *Yersinia ruckeri*8*.* The operon, which was characterized in 2014was found to contain two putative collagenase genes also belonging to the U32 peptidase family. Irregardless of the results obtained from this experiment, there remain many different opportunities to expand on the insights obtained.

**Bibliography**

1. Fettweis et. al (2019). The vaginal microbiome and preterm birth. [Nat Med.](https://www.ncbi.nlm.nih.gov/pubmed/31142849) 2019 Jun;25(6):1012-1021. doi: 10.1038/s41591-019-0450-2.
2. Fettweis et. al (2014). An Emerging Mycoplasma Associated with Trichomoniasis, Vaginal Infection and Disease. PLoS One. 9(10): e110943. Published online 2014 Oct 22. doi: 0.1371/journal.pone.0110943
3. Hampson, V., Liu, D., Billett, E., & Kirk, S. (1997). Amniotic membrane collagen content and type distribution in women with preterm premature rupture of the membranes in pregnancy. BJOG: An International Journal of Obstetrics and Gynaecology, 104(9), 1087-1091. doi:10.1111/j.1471-0528.1997.tb12073.x
4. Rajabi, M. R., Dean, D. D., Beydoun, S. N., & Woessner, J. F. (1988). Elevated tissue levels of collagenase during dilation of uterine cervix in human parturition. American Journal of Obstetrics and Gynecology, 159(4), 971-976. doi:10.1016/s0002-9378(88)80183-2.
5. Jones, O. W., & Nirenberg, M. W. (1962). Qualitative Study of RNA Codewords. Proceedings of the National Academy of Sciences, 48(12), 2115-2123. doi:10.1073/pnas.48.12.2115.
6. Zhou, Z., et al. (2016). Codon usage is an important determinant of gene expression levels largely through its effects on transcription. 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
7. Reza HD & Ashraf MM (2004). Collagenase activity in Prevotella bivia isolated from patients with premature rupture of membranes. Medical Journal of the Islamic Republic of Iran. 18(1):61-66.
8. Navais, R., Méndez, J., Pérez-Pascual, D., Cascales, D., & Guijarro, J. A. (2014). The yrpAB operon of *Yersinia ruckeri* encoding two putative U32 peptidases is involved in virulence and induced under microaerobic conditions. *Virulence*, *5*(5), 619-624. doi:10.4161/viru.29363.