Spring 2024

Genetic Analysis of Hereditary Gingival Fibromatosis Associated Sos1 Missense Variants of Uncertain Significance in Caenorhabditis Elegans

Himani Patel
hpatel7@stu.jsu.edu

Genetic Analysis of Hereditary Gingival Fibromatosis Associated *Sos1* Missense Variants of Uncertain Significance in *Caenorhabditis Elegans*

A Thesis Submitted to the
Graduate Faculty of Jacksonville State University
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science
with a Major in Biology

By

Himani Ramesh Patel

Jacksonville, Alabama

May 3, 2024

Himani Ramesh Patel May 3, 2024

Abstract

Hereditary gingival fibromatosis (HGF) is a disease that can present as benign overgrowth of gingival tissue in the mouth. The overgrowth can enclose the entire mouth and teeth in severe cases or present itself in a concentrated area. Researchers have identified that mutations in the *SOS1* gene can be responsible for HGF. This disease can impair basic functions related to the mouth. Eating, smiling, speaking can all be affected. Additionally, excess inflammation can cause periodontal disease because of the difficulty in maintaining proper oral health. Periodontal disease can lead to severe bone loss which can lead to complete loss of function of the teeth in the mouth. To further investigate HGF, associated missense variants of uncertain significance (VUS) were analyzed to assess their potential impact on the *SOS1* gene through a series of bioinformatics experiments and a comparative study in the *C. elegans* model system. This project primarily focuses on the potential effects of *SOS1* VUS on gene structure and function in human *SOS1* and *C. elegans sos-1.* With the use of Benchling, a VUS was identified in a conserved locus across human, nematode, and 8 other species at *SOS1* c.1289A>G (p. Asp430Gly). Additional bioinformatics research was conducted to examine the conservation for additional VUS loci across all clinical classes of significance. PolyPhen-2 was utilized to further analyze the predicted level of damage for each variant. The VUS *SOS1* c.1289A>G (p. Asp430Gly) was identified as a likely pathogenic or damaging variant. DNA primers were designed and tested to amplify the VUS region in the *C.elegans sos-1* ortholog. Additionally, a single RNA guide was designed to target *sos-1*, following CRISPR-Cas9 microinjection for future use in experiments. The designed and tested molecular biology reagents will be used in downstream microinjections for creating a CRISPR-Cas9 generated *sos-1* VUS mutant strain.

This mutant VUS strain will allow for *in vivo* assessment of potential functional impact to observe if *sos-1* mutant VUS nematodes will show a different phenotype compared to N2 wildtype nematodes related to reproductive issues. This project helps shed light on the potential impact of a HGF associated *SOS1* VUS through *in silico* experimentation and aimed to generated reagents to be used in future experiments for further *in vivo* evaluation in *C. elegans*.

viii, (38)

Acknowledgements

Thank you to Dr. Turner for being such a great mentor and for guiding me throughout my project and for advocating for me in spaces where I needed support. Dr. Turner has always believed in me, and my capabilities, and she has given me a space to express myself and show my individuality while also making me feel like I belong in the field of science! I would also like to thank Dr. Moore and Dr. Murdock for being such great committee members and for guiding me along with providing valuable feedback. Additionally, thank you to the Jacksonville State University's Biology Department. I would also like to thank Evan Waters for being such a great peer and for helping me with my experiments. Would like to give a huge shoutout to my little nephew Dhyaan for always being a light in my life. Seeing him smile would always make my day better even during the tough days of being a graduate student. I also would like to thank all my friends and family including but not limited Abhay, Bhoomi, Yash, Kritika, Makaila, Vanessa, and Brittany for being such amazing support systems. Biggest thank you for my mom, Rita Patel. She truly is my role model, and she gives me strength and motivation to work hard. Thank you for the skills you have instilled in me and thank you for always encouraging me to push boundaries and for striving for success.

Himani Patel

Table of Contents

Page

List of Tables

Page

List of Figures

Page

Introduction

Genes serve as fundamental building blocks of heredity, embodying both physical and functional traits. Delving into the intricacies of their structure and operation not only elucidates the spectrum of genetic diversity but also sheds light on how this diversity influences health and susceptibility to disease. Genes are composed of deoxyribonucleic acid (DNA), both coding and noncoding regions known as exons and introns. Genetic variants or mutations that occur within an exon can lead to a DNA base pair change resulting in a downstream change in a codon's encoded amino acid. This is known as a missense variant and can lead to a change in gene function that can lead to gene dysfunction and disease. This project examines the condition hereditary gingival fibromatosis (HGF) through the lens of the human gene *SOS1* and the nematode ortholog *sos-1*.

Gingival fibromatosis is a condition that can be drug induced or caused by a genetic mutation in the *REST* or *SOS1* genes (Strzelec et al., 2021). For this project, the *SOS1* gene will be analyzed in association with HGF. HGF typically presents as benign overgrowth of the gingival tissue in the mouth. This fibrous overgrowth can encapsulate the entire mouth and cover all teeth in severe situations, or it can be present in a concentrated area of the mouth, where it looks like a keratin-filled nodule Figure 1. The gingival overgrowth can be visible as a light pink color and generally is not painful (Almiñana-Pastor et al., 2017). However, a patient that does have this condition can undergo physiological complications because of mastication, speech problems, and the inability to close the lips properly. The onset of HGF is slow and progressive and can take several years of gingival enlargement before the condition poses as an issue for a patient (Dani et al., 2015). Gingival overgrowth does not directly affect the bone, however, in severe cases it can cause periodontitis which can then later lead to bone damage and tissue loss

identified through radiographic examination (Gawron et al. 2017). It is very rare for both conditions, periodontitis and HGF to occur concomitantly, but if a patient has HGF and poor dental hygiene, a development of periodontitis can occur. This case occurs more frequently in younger patients (Bhatsange et al., 2021). Bone damage can specifically occur in the mandible or the maxilla, which can cause dysmorphia in the face (Garwon et al., 2017).

HGF is an autosomal dominant inherited disease. There are situations when HGF is passed down through autosomal recessive disease and X-linked inherited. This disease can be accompanied by additional symptoms such as inflammation, neoplasia, and hormonal imbalances (Majumder et al 2013). Chromosomes numbers 2, 4, and 5 are found with their specific genetic loci. Other genetic loci like 8, 14q, 19p, 19q, and Xq can also be linked to gingival fibromatosis (Majumder et al 2013).

Treatments of HGF can vary in different ways, there are traditional methods of surgery and new innovative technology. The most common treatment and procedure used for HGF is through a gingivectomy, where the overgrowth gingival is cut using a scapula while the patient is being treated with general anesthesia. Following this treatment, it is common for an oral surgeon to prescribe antibiotics for the patients with a chlorohexidine oral rinse for two weeks (Garwon et al., 2016). An alternative route to the treatment of HGF can be done with a laser, specifically the diode laser, where the overgrowth gingival tissue is burned off. The method of using a laser can be beneficial because it causes less bleeding for the patient and the healing time post-surgery is shorter than a traditional gingivectomy. The only concern with the laser would be heat damage to the gums, and it can be costly (Dureja et al., 2020). An important factor to consider regarding HGF is that it is inherited, but it can also be drug induced through medications such as

nifedipine, phenytoin and cyclosporine. In the situation of gingival fibromatosis as a drug side effect, treatment of the condition would be the same as HGF (Margiotti et al., 2017).

The *SOS1* gene is associated with the RAS pathway. Gingival fibromatosis is a rasopthy disease which can be categorized as a disease that is caused by changes in genes that deal with the Ras/nitrogen-activated protein kinase (Ras/MAPK) pathway (Modzelewska et. al, 2007). This pathway helps systems in the body to grow and prevent cancers. The *SOS1* gene is a dual-function guanine nucleotide exchange factor (GEF) that catalyzes the exchange on Ras and Rac. Due to the mutation of the *SOS1* gene in relation to HGF, it affects the growth of gum tissues in the mouth (Modzelewska et. al, 2007).

The *SOS1* gene belongs to a set of genes known as Sons of Sevenless (SOS) that are exchange factors for RAS proteins that are important signal transduction regulators. This signal transduction regulating pathway plays a role in crucial biological processes. If a point mutation occurs in these types of proteins, it can create diseased changes affecting the signal transduction pathway, altering gene expression and protein structure and function (Baltanás et al., 2020). If gene expression is altered through a change in nucleotide sequence, a disease can become present in a patient (Bhatsange et al., 2021). When a single nucleotide changes in the *SOS1* gene, known as a missense mutation or point mutation, it can result in a pathogenic phenotype of gingival overgrowth (Bhatsange et al., 2021). Other mutations such as deletions and duplications are also present in the *SOS1* gene that aids in the disease HGF. The genetic variants explored in this project are missense variants of uncertain significance (VUS) associated with HGF. With this type of point mutation, the reading frame for the gene remains the same since only a single nucleotide is being altered resulting in a new encoded amino acid. The alteration of this single

nucleotide results in a new amino acid to be inserted into the gene's encoded protein that might alter the protein's structure and function.

In this experimental design, the model organism that will be studied is a *C. elegans*. There are many reasons as to why *C. elegans* are used as a model organism for experimentation. Firstly*, C. elegans* are very easy to maintain, they are roundworms that allow for investigations in a nonvertebrate organism, and they reproduce very quickly, which can allow data collecting to be a quicker process. *C. elegans* have a short reproductive cycle, they begin in a larval state and can reach the adult stage in about three days with a life span of two to three weeks. Adults have 959 somatic nuclei. There are two sexes for *C. elegans,* hermaphrodite and male. Hermaphrodites have both sperm and egg cells which allows these *C. elegans* to self-fertilize. These hermaphrodites can create hundreds of offspring. These organisms are wonderful organisms to study because despite their microscopic size, they have nervous systems, epidermal layers, muscular layers, an intestinal layer, and reproductive system. Up to 60% of genes in the *C. elegans* animal models have orthologs and strong homologues to mammals especially humans. This is great for studying human disease in the nematode models (Apfeld et. al, 2018).

With the knowledge of the connections and relationships between *C. elegans* and human genomes, scientists are using this animal model to screen variants and test to see if variants of uncertain significance are pathogenic or benign. For experimentation purposes, the HGF associated *SOS1* c.1448A>G (p. Asp 483Gly) VUS will be examined to see if it has any potential pathogenic or damaging potential (Hopkins et. al, 2023). If the human *SOS1* gene has a nematode ortholog, then the *SOS1* VUS location c.1448A>G (p. Asp 483Gly) will show conservation within the nematode *sos-1* gene. This results in HGF in humans, and if placed in the

context of the *sos-1* ortholog could potentially lead to reproductive issues in the mutant nematodes compared to the control N2 nematodes.

Materials and Methods

All experiments detailed in this study were conducted in accordance with established standard operating protocols and procedures as set forth by the research laboratory of Dr. Ashley Turner.

Identification of HGF Associated Variants within *SOS1* in Conserved Loci

HGF associated missense VUS previously identified in human patients were obtained from ClinVar (Landrum et al., 2018). The program Benchling ((Benchling [Biology Software]. (2022)) was used for sequence files and multiple sequence alignments (MSAs) to examine the conservation of each VUS locus. Coding DNA (cDNA) gene sequence files for human and *C. elegans* were initially obtained from Ensembl, imported into Benchling, and used to construct an MSA. Eventually, an MSA was carried out with ten species. The following gene IDs were used to import each species' cDNA file: human (ENSG00000115904), macaque (ENSMMUG00000020575), Pig (ENSSSCG00000008478), dog (ENSCAFG00845022915), mouse (ENSMUSG00000024241), rat (ENSRNOG00000007106), guinea pig (ENSCPOG00000002487), tropical clawed frog (ENSXETG00000008386), zebrafish (ENSDARG00000100503), and nematode (WBGene00004947). Each MSA was constructed through a consensus alignment utilizing the MAFFT alignment program (Katoh & Standley, 2013). Conservation at the DNA and amino acid levels were assessed.

PolyPhen-2 Analysis

PolyPhen-2 is a tool used to predict the possible impact of an amino acid change on human proteins. The SOS1 protein identifier (Q07889) and amino acid substitution position

(430) were input into the PolyPhen-2 online platform, along with the specific amino acid substitution (Aspartic Acid to Glycine) (Adzhubei et al., 2013). The human SOS1 protein sequence has 1,333 amino acids.

HGF Associated *SOS1* VUS Mutational Analysis

Mutational analysis was carried out for thirty HGF associated *SOS1* missense variants across all clinical classifications of significance ranging from benign to uncertain to pathogenic. Conservation for each variant locus was examined through comparing human *SOS1* and nematode *sos-1* through an MSA. The *SOS1* VUS c.1289A>G (p. Asp430Gly) was identified, and its surrounding DNA base pairs and amino acid residues examined.

Primer Design and Resuspension

Two DNA primer sets were designed using the design tool within Benchling ((Benchling [Biology Software] (2022)) to amplify the VUS locus in the *sos-1* gene. Primer sequences are provided in the following results. These primers were needed to help amplify the VUS of DNA for *sos-1* using N2 *C. elegans* DNA as template during polymerase chain reaction (PCR) experiments. Primers were obtained from IDT as dry, lyophilized oligos and resuspended initially to create a primer stock at 100 µM in TE buffer. Next a working primer dilution was created at 10 µM with molecular water and used for downstream PCR experiments.

Nematode Manipulation and DNA Digest

Nematodes were manipulated and transferred using a custom platinum wire nematode pick for strain maintenance for worm lysis. Genomic DNA was obtained by placing 10-20 N2 *C. elegans* adults in a PCR tube with 30 µL of lysis solution containing Proteinase K (Cat. # V302B, Promega) and incubated at 65°C for 2 hours, followed by 95°C for 10 minutes.

Polymerase Chain Reactions and Gel Electrophoresis

A small volume (1 µL) of the DNA digest solution was used directly as a template in 12.5 µL PCR reactions with 2x Taq Master Mix (Cat. # M0270S, New England Biology) and DNA primers at a 10 µM starting concentration. This PCR was completed with 1 µL forward primer, 1 µL reverse primer, 6.25 µL 2X Taq Master Mix, 12.5 µL nuclease-free water, and 1 µL variable. An initial PCR was carried out with an annealing temperature of 55°C. The following PCRs were conducted to optimize the annealing temperature for each primer set across a temperature gradient. The PCR reactions were resolved and analyzed on 6% polyacrylamide/TBE gels with a 100-base pair (bp) ladder (Cat. # G2101, Promega) and 1 kilobase (kb) ladder (Cat. # G5711, Promega) through electrophoresis. Gels were stained with GelRed (Cat. # 41003, Biotium) and observed and imaged with a UV gel documentation system.

CRISPR RNA Guide Design

CRISPR targets were identified surrounding the VUS location in the *sos-1* gene using Benchling ((Benchling [Biology Software]. (2022)). A single RNA guide was designed and selected to target the VUS region in *sos-1*

Results

Once the *SOS1* gene and genetic variants were identified through Ensembl and ClinVar, the mutational analysis was carried out. A large mutational analysis was conducted to analyze the conservation status for 30 HGF associated variants across human *SOS1* and *C. elegans sos-1*.

Ten variants were selected from the 3 clinical classifications: pathogenic, uncertain, and benign. In Figure 2, in the mutational analysis in the *sos-1* gene, there are multiple boxes present. The red boxes are variants collected from ClinVar indicating known pathogenic variants, the green boxes are indicating benign variants, and the blue boxes are variants of uncertain significance. Table 1 displays three VUS significance with their locations with the exon in the *sos-1* gene. The VUS presented in Table 1 was further examined for PolyPhen-2 analysis. PolyPhen-2 is a tool used to estimate the potential impact of a genetic variant. A genetic variant is predicted with a score ranging between 0 and 1. Anything close to one means that the VUS is likely to present as damaging and any score closer to the number 0 indicates the variant is most likely to be benign. Based off this research, the *SOS1* VUS c.2344A>G (p. Ile782Val) was predicted to have the highest PolyPhen-2 score as seen in Figure 3. The score was 0.994 and this VUS is predicted to be damaging and pathogenic. Once choosing the VUS being studied in these experiments the next step was to carry out a MSA between 10 species. Table 2 depicts the 10 species MSA on the DNA base pair level and table 3 depicts those same 10 species orthologs on an amino acid level. Table 2 and table 3 show several species ranging from mammals to nonmammals. Humans are at the top of table 2 and table 3 since they are the most complex species on the MSA. The SOS1 VUS c.2344 A>G (p. Ile 782Val) genetic loci occurs within conserved DNA base pair and amino acid residue across all species except zebrafish. Nematodes are at the bottom of table 2 and 3 since they are the least complex species on the MSA.

Next, two DNA primer sets were designed and used to amplify the VUS locus within the *sos-1* gene. In Table 4, the designed primer sets are provided. The optimum PCR product range for design was 600 to 1,000 base pairs. These primers were used in PCR experiments. In the initial PCR, each primer set was tested at an annealing temperature of 55°C. The initial PCR as

depicted in Figure 4 shows the two primer sets. The first lane is a 100 bp ladder. Lanes two and three are the N2 *C. elegans* DNA from the DNA digest, lane three is the positive control N2 *C. elegans* and lane four is the water negative control. Little to no amplification was present for primer set 1 and specific and nonspecific amplification was present for primer set 2. After this, a gradient PCR was done to amplify the DNA with each primer set at different annealing temperatures to increase amplification and reduce nonspecific amplification present. Figures 5 shows primer set 1 with temperatures ranging from 55°C to 65°C to 65°C to 75°C. Figure 6 is associated with primer set two and there were also two annealing temperatures ranging from 55°C to 65°C to 65°C to 75°C.

Following the gradient PCR, a CRISPR Cas-9 RNA guide design was created to target the VUS region of the *SOS1* gene in the *C. elegans* model through CRISPR Cas-9 editing. Figure 7 depicts the single RNA guide that was created using Benchling ((Benchling [Biology Software]. (2022)). This RNA guide is complimentary to the binding site with the protospacer adjacent motif (PAM) of CGG that Cas9 will bind to cut the specified DNA target in the *SOS1* gene.

Conclusions

Following the identification of a *C. elegans* ortholog for the HGF associated *SOS1* gene, variants of uncertain significance were examined to test their potential impacts on gene structure and function. Evolutionary conservation of genetic variant loci was examined through mutational analysis across human and nematode orthologs through multiple sequence alignments. The mutational analysis was done to visualize and examine all VUS and other genetic variant classifications within the *SOS1* gene. This is helpful to see where there are concentrated areas of pathogenic variants, and where there is a concentration of benign variants. A MSA can be done to test the conservation of the VUS location between many distinct species. If this VUS region is conserved in many species, then that is additional data that can support the likelihood of this VUS being in an evolutionarily conserved region that is likely functionally important. This also helps to observe variants of uncertain significance and their locations on the *SOS1* gene and if their locations are evolutionary conserved across species. If the VUS is in an area where there is a high density of other pathogenic variants, there is a high probability that the VUS will also be pathogenic. This experiment as shown in Figure 2, depicts the VUS, known pathogenic and benign variants in the *SOS1* gene. The VUS that was chosen will be selected from an area of the gene where there are verified pathogenic variants in this area. This is done because that are of loci of the gene will most likely be damaging if the surrounding variant locations are shown to be damaging. Exons 10 and 14 show a high prevalence of HGF associated pathogenic variants with a higher density of pathogenic and likely pathogenic variants compared to surrounding exons. All data referred to as known benign or known pathogenic is provided through the platform ClinVar that evaluates reports human clinical variants. There are also several VUS present within

these exons, these three VUS located in exon 10 and 14 adjacent were moved forward for further evaluation through PolyPhen-2 analysis.

The PolyPhen-2 tool is an applicable software that will predict whether a genetic variant will be damaging or not damaging. The three VUS selected from the mutational analysis are shown in Table 1. The results from PolyPhen-2 indicated that the *SOS1* VUS c.1289 A>G (p. Asp430Gly) has the highest PolyPhen-2 score of 0.994 as shown as Figure 4 indicating it will be probably damaging. With the results of PolyPhen-2, the next step was further evaluating the evolutionary conservation of this VUS location and the *sos*-1 gene with additional species orthologs. Through MSA, there were ten species evaluated for evolutionary conservation at the VUS region. These organisms were human, macaque, pig, dog, mouse, rat, guinea pig, tropical clawed frog, zebrafish, and nematode. Through the MSA shown in Table 2, the VUS region is shown to be highly conserved amongst all organisms except for the zebrafish on a DNA and amino acid level. This analysis was done by alignment of the nucleotide, referenced as DNA level, and with the amino acids.

Following the MSA, Table 4 shows the primers designed to further this project. There were two primers created, this is due to find an optimal primer set for specific amplification of the VUS region in *sos-1*. When designing primers for PCR amplification, several factors must be carefully considered to ensure successful and specific amplification of the target DNA region. Key parameters include primer length, annealing and melting temperatures, GC content, and secondary structure. Primer length ideally falls within the range of 18 to 24 base pairs; shorter primers may amplify unintended regions, while longer ones may hinder-hybridization kinetics. Annealing temperature, typically set 5°C below the melting temperature, is critical for primer

specificity; lower temperatures risk nonspecific binding. It is essential for both forward and reverse primers to possess similar melting temperatures to facilitate efficient PCR with amplification of a single target PCR band. Moreover, GC content, representing the percentage of guanine and cytosine bases, should fall between 40% and 60%. To enhance primer stability and specificity, a GC clamp consisting of 2 to 3 guanine and cytosine bases at the 3' end is recommended. This configuration exploits the additional hydrogen bonding capacity of GC pairs, ensuring robust and specific binding to the target DNA sequence during PCR amplification. In PCR, the DNA sequence is heated to denature and separated into two pieces. After this step, an enzyme called Taq polymerase creates two new strands of DNA. A polymerase chain reaction is going to be used in this experiment to amplify the target DNA sequence of the *SOS1* gene. The exact portion of the DNA sequence that is going to be amplified is the variant location c.1289A>G (p. Asp430Gly) in the human *SOS1* gene and *sos-1* c.1448A>G (p. A483Gly) in the nematode which is what is being amplified. Once the DNA is amplified, the process of gel electrophoresis will be used to visually see the results of the PCR. Gel electrophoresis is a procedure used to isolate DNA parts as indicated by their size and allows to check and see whether the PCR successfully ran. Gel electrophoresis will use a current through the gel with the molecules being analyzed. The DNA fragments will separate based on their size. The base pairs that are smaller travel to the bottom of the gel plate. The base pairs that are lager tend to stay towards the top of the gel plate.

Primer set 1 in this experiment if referred to the one with the expected base pair length of 1,183. Primer set 2 in this experiment will be the one with the base pair length of 741. Primer set 1 has a penalty score of 0.373 which is preferred since a low penalty score can predict the primer pair’s on-target specificity. Primer set 2 has a penalty score of 7.854 which is relatively high and

could potentially indicate that this primer set may have trouble targeting the specific region of the gene where the VUS is located. Following the creation and resuspension of the primer sets, the next step of this project was to complete polymerase chain reactions. The initial PCR was done with testing primer set 1 and primer set 2. The initial PCR will including a positive control, a water negative control along with two master mixes for each primer set. The Taq2x reagent was created by Taq DNA polymerase, originating from the thermophilic bacterium *Thermus aquaticus*, a resilient enzyme known for its ability to endure high temperatures. Widely employed in PCR, it effectively amplifies DNA fragments. Notably, Taq polymerase can withstand the repetitive heating cycles up to 95°C required by PCR without experiencing substantial loss of activity. The PCR product from the initial PCR ran under the thermocycler in temperatures 55°C. Following this experiment, to visualize the results of the PCR, a gel electrophoresis is carried out. As seen in Figure 4, the first primer set does not seem to have amplification in any of the columns. The second primer set is labeled 2-1 to 2-4 and does have amplification in bands around the 1,000 base pair area. For the first primer set, there does seem to be some faint banding over the 1,500 base pair area. There is no amplification in the water negative controls, so this suggests there is no contamination within the reagents in this experiment. Additional PCR experiments were needed to identify and optimize amplification conditions.

Following this experiment, gradient PCR was carried out to try and find the optimal temperature where the primer sets amplify the single target band. In the gradient PCR for the first and second primer set, the same steps were used from the initial PCR to create master mix however there were 8 test tubes used to create the temperature gradient that contained the tested N2 DNA from the initial PCR. The gradient PCR experiment was done to find the optimal

annealing temperature for each primer set. The setup for this type of PCR closely resembles the previous experiment conducted at a controlled temperature. However, it involves a range of temperatures applied to the same primers. The desired banding pattern is determined by the number of base pairs visible in the primers and is assessed using the DNA ladder in the experiment. An amplified, single target band is desired from each primer set through PCR. If the primers successfully bind to the desired DNA region during the gradient PCR, then the specific temperatures and the range of temperatures used are analyzed to determine which temperature yields a more efficient amplification with fewer extraneous banding patterns from other amplicons. This experiment is also done to get single clean bands at the optimal temperature ranges.

For primer set 1and primer set 2, the pcr was done at temperatures 55°C to 65°C and 65°C to 75°C. Following these PCR experiment results, a gel electrophoresis was done to visualize the gradient PCR amplification from each primer set.

The gradient PCR in primer set 1 gel results as seen in Figure 5, the ladder in lane one is ripped, therefore this experiment can be redone again. For what the ladder visible top of the ladder can help discern, however, the 1,500 base pair band is still visible. The banding for the rest of the lanes, however, is significantly higher than the 1500 base pair band in the ladder. Due to this analysis, a 1 kb ladder was used further on in these experiments for primer set 1. The next gel that was done was with a 1 kb ladder at temperatures 55°C to 65°C. This gel electrophoresis was done with the same PCR product from the previous gel. The ladder and PCR products did not separate enough. This could be due to the gel not being run long enough along with the PCR products being run at a low temperature. To trouble shoot these errors, the next step was to

complete the gradient PCR once again, however, to complete it at a higher temperature. This temperature ran from the rang 65°C to 75°C. Additionally, the length of the gel experiment was expanded. The gel electrophoresis was run in the electrode box for roughly 2 and a half hours. Following this experiment, the gel electrophoresis with the 1 kb ladder did fully separate following the extended gel electrophoresis experiment, however, the product is seen to be very blurry which indicates that the products is melting or fading away due to the high voltage and length of the ran gel. This has led to the conclusion that this PCR and gel electrophoresis experiment needs to be changed to agarose gel to accommodate for the length of the expected band size along with the separation of the 1 kb ladder. For primer set 2 there were several gels ran with this experiment as well. The first gradient PCR was done at 55°C to 65°C. The results of the PCR did not have any banding in the water negative control lane which suggest there was no contamination while performing micropipeting during PCR and loading the lanes for the gel electrophoresis. However, there were some inconsistent bands amplifying. These bands are not in the same pattern as seen in the initial gel in Figure 4. The amplification in Figure 4 was around 1,000 base pairs and in the next gel the bands were amplifying around the same area however the bands are faint and are not present in all the lanes. To trouble shoot this error this experiment was repeated entirely. The gradient PCR was done once again at the same temperature of 55°C to 65°C. Following the repeat experiment, as seen in Figure 6, there is a much clearer and consistent banding in all lanes. The water negative control lane is clear which shows that there is no contamination in the micropipeting process. Lanes 2 is the lowest temperature of 55°C and lane 9 is the highest temperature is 65°C. There are many nonspecific bands in most of the lanes in this gel, especially at the lower temperature. The nonspecific bands could indicate there are other regions of nematode DNA being amplified at different temperatures. Lane 8, which is

64.3°C, is the darkest band with the least number of nonspecific bands, suggesting that this is the optimal annealing temperature needed for creating and using the RNA guides for CRISPR Cas-9.

Once the optimal temperature was identified for a primer set and perfecting the visualization of the location of the VUS region, the next step was creating a RNA guide that will be used during microinjection of CRISPR-Cas9 in downstream experiments. CRISPR-Cas9 is a gene editing tool that will be used during these experiments to mutate healthy nematodes to place the *SOS1* VUS into the *sos-1* nematode ortholog. CRISPR stands for clustered regularly interspaced short palindromic-associated repeats. This technology can be used to edit and manipulate any DNA sequence, such as with this nematode experiment or even to treat genetic conditions. This can be extremely specific to a single nucleotide which is what will be examined in this experiment. Cas9 cuts the targeted part of the DNA and can be followed by a single stranded DNA repair template that allows the natural DNA repair process to incorporate the new DNA donor repair sequence. The DNA repair mechanism can be performed by either the host's natural repair machinery or by using customized DNA sequences (Sharma et. al, 2021). The designed and optimized PCR assay can be used for screening for the edited mutant VUS nematodes to breed and create this strain.

Once editing *C. elegans* at the variant location c.1448A>G (p. Asp 483Gly) to create a point mutation, observations of wildtype N2 and mutant *sos-1* VUS nematodes could be carried out to examine the impact of the VUS on *sos-1* gene function. It is suspected that the mutant nematode will present with phenotypes associated with known null mutations in the *sos-1* gene. In this case, the changes would involve vulva development and changes in the reproduction of the hermaphrodite specifically causing the mutant C. *elegans* to be sterile. If the nematode does

not reproduce once mutated, that is confirmation that the VUS being studied is pathogenic to the *sos-1* gene and nematode health. Connecting this back to humans, this would suggest that the human clinical variant would also indeed be likely pathogenic and would be damaging to the gene *SOS1* causing the disease HGF since there is a nematode orthologue of this VUS. This research offers *in silico* analysis and the creation of reagents to be used for *in vivo* testing to explore the significance of a particular HGF associated *SOS1* VUS. By doing so, this study aims to illuminate its clinical relevance for both humans and nematodes, while also investigating the conservation of gene and protein structures and functions across various species. The methods and strategies used in this research can be wider usage beyond analyzing *SOS1* variants. Through establishing a framework for computer-based and real-world evaluations, scientists can modify these techniques to investigate diverse genetic variants linked to a range of diseases in various species. This adaptability underscores the broader applicability and importance of this experimental approach, extending beyond the specific of *SOS1* associated variations, but also to a magnitude of genes and diseases.

Bibliography

Abraham, K. K., Ambooken, M., Mangalathu, A. G., Jose, S. C., & Mathews, S. E. (2022). Idiopathic Gingival Fibromatosis in a Pediatric Patient. *Journal of dentistry for children (Chicago, Ill.)*, *89*(1), 41–45.

Almiñana-Pastor, P. J., Buitrago-Vera, P. J., Alpiste-Illueca, F. M., & Catalá-Pizarro, M. (2017). Hereditary gingival fibromatosis: Characteristics and treatment approach. *Journal of clinical and experimental dentistry*, *9*(4), e599–e602.

Apfeld, J., & Alper, S. (2018). What Can We Learn About Human Disease from the Nematode C. elegans? *Methods in molecular biology (Clifton, N.J.)*, *1706*, 53–75. https://doi.org/10.1007/978-1-4939-7471-9_4

Baltanás, F. C., Zarich, N., Rojas-Cabañeros, J. M., & Santos, E. (2020). SOS GEFs in health and disease. *Biochimica et biophysica acta. Reviews on cancer*, *1874*(2), 188445. https://doi.org/10.1016/j.bbcan.2020.188445

Bhatsange, A., Moharir, A., Mistry, V., & Japatti, S. (2021). Idiopathic gingival fibromatosis in association with aggressive periodontitis and candidal infection: A unique case report with 7-year follow-up. *Journal of Indian Society of Periodontology*, *25*(1), 65–69. https://doi.org/10.4103/jisp.jisp_522_19

Chen, W., & Chen, X. (1996). Gingival fibromatosis. *Journal of Tongji Medical University = Tong ji yi ke da xue xue bao*, *16*(1), 55–57. https://doi.org/10.1007/BF02889047

Costa, C., Braz, S. V., de Toledo, I. P., Martelli-Júnior, H., Mazzeu, J. F., Guerra, E., Coletta, R. D., & Acevedo, A. C. (2021). Syndromes with gingival fibromatosis: A systematic review. *Oral diseases*, *27*(4), 881–893. https://doi.org/10.1111/odi.13369

Cunha, J., Ramos, M., Regis, D. M., Sanchéz-Romero, C., de Andrade, M. E., Bezerra, B. T., &

de Albuquerque-Júnior, R. (2020). Generalized hereditary gingival fibromatosis in a child: clinical, histopathological and therapeutic aspects. *Autopsy & case reports*, *10*(1), e2020140.

Gandhi, M., Tandon, S., Sharma, M., & Vijay, A. (2018). Nonsyndromic Gingival Fibromatosis: A Rare Case Report. *International journal of clinical pediatric dentistry*, *11*(3), 250–253. https://doi.org/10.5005/jp-journals-10005-1521

Dani, N. H., Khanna, D. P., Bhatt, V. H., & Joshi, C. P. (2015). Idiopathic gingival fibromatosis. *Journal of Indian Society of Periodontology*, *19*(6), 698–700. https://doi.org/10.4103/0972-124X.162206

Dureja, D., Sharma, H., Agarwal, M., Srivastava, S., & Tandon, P. (2020). Hereditary Gingival Fibromatosis: A Report of a Rare Case in Siblings and Its Management Using Diode Laser. *Contemporary clinical dentistry*, *11*(3), 290–293. https://doi.org/10.4103/ccd.ccd_133_19

Gao, G., Tian, Q., Han, A., Yang, R., Shi, F., & Chen, D. (2021). Case Report: A Case of Hereditary Gingival Fibromatosis with a High Level of Human β Defensins in Gingival Epithelium. *Frontiers in immunology*, *12*, 763026. https://doi.org/10.3389/fimmu.2021.763026

Gawron, K., Bereta, G., Nowakowska, Z., Łazarz-Bartyzel, K., Potempa, J., Chomyszyn-Gajewska, M., Górska, R., & Plakwicz, P. (2017). Analysis of mutations in the SOS-1 gene in two Polish families with hereditary gingival fibromatosis. *Oral diseases*, *23*(7), 983–989. https://doi.org/10.1111/odi.12684

Gawron, K., Łazarz-Bartyzel, K., Potempa, J., & Chomyszyn-Gajewska, M. (2016). Gingival

fibromatosis: clinical, molecular, and therapeutic issues. *Orphanet journal of rare diseases*, *11*, 9. https://doi.org/10.1186/s13023-016-0395-1

Gawron, K., Ochała-Kłos, A., Nowakowska, Z., Bereta, G., Łazarz-Bartyzel, K., Grabiec, A. M., Plakwicz, P., Górska, R., Fertala, A., Chomyszyn-Gajewska, M., & Potempa, J. (2018). TIMP-1 is associated with collagen type I overproduction in hereditary gingival fibromatosis. *Oral diseases*, *24*(8), 1581–1590. https://doi.org/10.1111/odi.12938

Hart, T. C., Zhang, Y., Gorry, M. C., Hart, P. S., Cooper, M., Marazita, M. L., Marks, J. M., Cortelli, J. R., & Pallos, D. (2002). A mutation in the SOS1 gene causes hereditary gingival fibromatosis type 1. *American journal of human genetics*, *70*(4), 943–954. https://doi.org/10.1086/339689

Hazzaa, H. H., Gouda, O. M., Kamal, N. M., Ali, S., El Shiekh, M., & Tawfik, M. M. (2018). Expression of CD163 in hereditary gingival fibromatosis: A possible association with TGF-β1. *Journal of oral pathology & medicine: official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology*, *47*(3), 286–292. https://doi.org/10.1111/jop.12679

Hopkins, C. E., McCormick, K., Brock, T., Wood, M., Ruggiero, S., Mcbride, K., Kim,C., Lawson, J. A., Helbig, I., & Bainbridge, M. N. (2023). Clinical variants in Caenorhabditis elegans expressing human STXBP1 reveal a novel class of pathogenic. variants and classify variants of uncertain significance. Genetics in Medicine Open, 1(1), 100823. https://doi.org/10.1016/j.gimo.2023.100823

Kumari, A., Bansal, M. B., Asrani, K. H., & Yadav, A. (2021). Nonsyndromic with Recurrent Idiopathic Gingival Fibromatosis: A Rare Case Report. *International journal of clinical pediatric dentistry*, *14*(1), 158–160. https://doi.org/10.5005/jp-journals-10005-1933

Margiotti, K., Pascolini, G., Consoli, F., Guida, V., Di Bonaventura, C., Giallonardo, A. T., Pizzuti, A., & De Luca, A. (2017). Lack of pathogenic mutations in SOS1 gene in phenytoin-induced gingival overgrowth patients. *Archives of oral biology*, *80*, 160–163. https://doi.org/10.1016/j.archoralbio.2017.04.002

Majumder, Poulami et al. "The autosomal recessive inheritance of hereditary gingival fibromatosis." *Case reports in dentistry* vol. 2013 (2013): 432864. https://doi.org/10.1155/2013/432864

Mariani, P., Zhurakivska, K., Santoro, R., Laino, G., Russo, D., & Laino, L. (2021). Hereditary gingival fibromatosis associated with the missense mutation of the KCNK4 gene. *Oral surgery, oral medicine, oral pathology, and oral radiology*, *131*(6), e175–e182. https://doi.org/10.1016/j.oooo.2020.08.006

Modzelewska, K., Elgort, M. G., Huang, J., Jongeward, G., Lauritzen, A., Yoon, C. H., Sternberg, P. W., & Moghal, N. (2007). An activating mutation in *sos-1* identifies its Dbl domain as a critical inhibitor of the epidermal growth factor receptor pathway during Caenorhabditis elegans vulval development. *Molecular and cellular biology*, *27*(10), 3695–3707. https://doi.org/10.1128/MCB.01630-06

Sharma, Garima et al. "CRISPR-Cas9: A Preclinical and Clinical Perspective for the Treatment of Human Diseases." *Molecular therapy: the journal of the American Society of Gene Therapy* vol. 29,2 (2021): 571-586. doi: 10.1016/j.ymthe.2020.09.028

Strzelec, K., Dziedzic, A., Łazarz-Bartyzel, K., Grabiec, A. M., Gutmajster, E., Kaczmarzyk, T., Plakwicz, P., & Gawron, K. (2021). Clinics and genetic background of hereditary gingival fibromatosis. *Orphanet journal of rare diseases*, *16*(1), 492. https://doi.org/10.1186/s13023-021-02104-9

Appendix A

Tables

Table 1

Analysis of 3 VUS in the SOS1 gene associated with HGF from ClinVar

SOS1 Variant	Exon	Protein Change	Reported Patient Conditions	Conserved Loci (DNA/amino acid)
c.2344A>G (p. Ile782Val)	14	I782V, I775V	Raso Pathy, Fibromatosis, gingival	Yes / Yes
c.2253T>G (p. Asn751Lys)	14	N751K, N744K	Cardiovascular phenotype, not provided, Noonan syndrome, Fibromatosis, gingival	Yes / Yes
c.1289A>G (p. Asp430Gly)	10	D423G, D430G	Noonan syndrome, Fibromatosis, gingival	Yes / Yes

Table 2

Multiple sequence alignment for SOS1 VUS at DNA level across ten species orthologs

Common Name	**Scientific Name**	**MSA (DNA base pairs)**
HGF VUS	*Homo sapiens*	AAG AAT ATT G**G**T GGT TGG GAG
Human	*Homo sapiens*	AAG AAT ATT G**A**T GGT TGG GAG
Macaque	*Macaca mulatta*	AAG AAT ATT G**A**T GGT TGG GAG
Pig	*Sus scrofa*	AAG AAT ATT G**A**T GGT TGG GAG
Dog	*Canis lupus familiaris*	AAG AAT ATT G**A**T GGT TGG GAG
Mouse	*Mus musculus*	AAG AAC ATT G**A**T GGC TGG GAG
Rat	*Rattus norvegicus*	AAG AAC ATT G**A**T GGT TGG GAG
Guinea Pig	*Cavia porcellus*	AAG AAT ATT G**A**T GGT TGG GAG
Tropical Clawed Frog	*Xenopus tropicalis*	AAG AAC ATT G**A**T GGA TGG GAG
Zebrafish	*Danio rerio*	--- --- --- --- --- --- ---
Nematode	*Caenorhabditis elegans*	TCG TCA ATT G**A**T GGA TTT ACC

Table 3

Multiple sequence alignment for SOS1 VUS at amino acid level across ten species orthologs

Common Name	Scientific Name	MSA (amino acid residues)
HGF VUS	*Homo sapiens*	K K M N E I Q K N I **G** G W E G K D I G Q C
Human	*Homo sapiens*	K K M N E I Q K N I **D** G W E G K D I G Q C
Macaque	*Macaca mulatta*	K K M N E I Q K N I **D** G W E G K D I G Q C
Pig	*Sus scrofa*	K K M N E I Q K N I **D** G W E G K D I G Q C
Dog	*Canis lupus familiaris*	K K M N E I Q K N I **D** G W E G K D I G Q C
Mouse	*Mus musculus*	K K M N E I Q K N I **D** G W E G K D I G Q C
Rat	*Rattus norvegicus*	K K M N E I Q K N I **D** G W E G K D I G Q C
Guinea Pig	*Cavia porcellus*	K K M N E I Q K N I **D** G W E G K D I G Q C
Tropical Clawed Frog	*Xenopus tropicalis*	K K M N E I Q K N I **D** G W E G K D I G Q C
Zebrafish	*Danio rerio*	- -
Nematode	*Caenorhabditis elegans*	K R L N E I Q S S I **D** G F T G S P I G K T

Table 4

DNA primer design for targeting the SOS1 VUS locus within sos-1

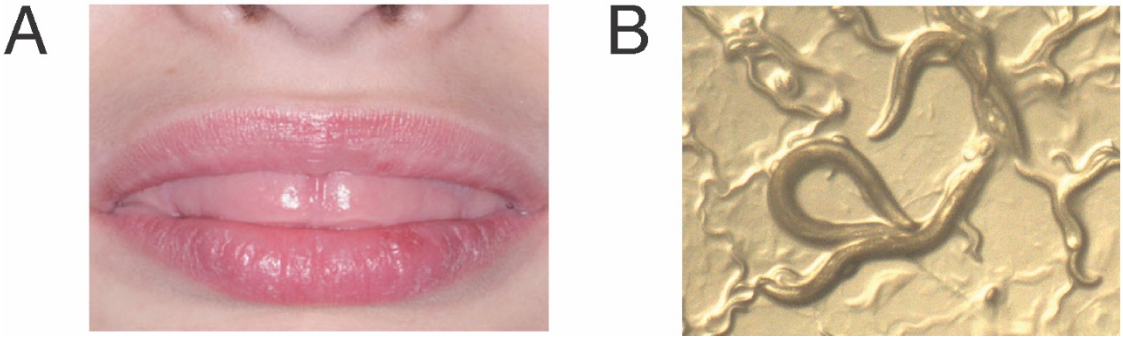
Primer Set	Targeted VUS	Forward Sequence	Sequence 2 (Reverse)	Length (bp)
1	*SOS1* c.1289A>G (p. Asp430Gly)	5' GGCAAACCGGCA ATTTTCCGCA 3'	5' CCGGTGTCGTTTG CAGAGCACT 3'	1183
2	*SOS1* c.1448A>G (p. Asp483Gly)	5' TCCGGCAAATCG GCATATTGGC 3'	5' TCAAACATGCGA GAAAACGGTT 3'	741

Appendix B

Figures

Figure 1

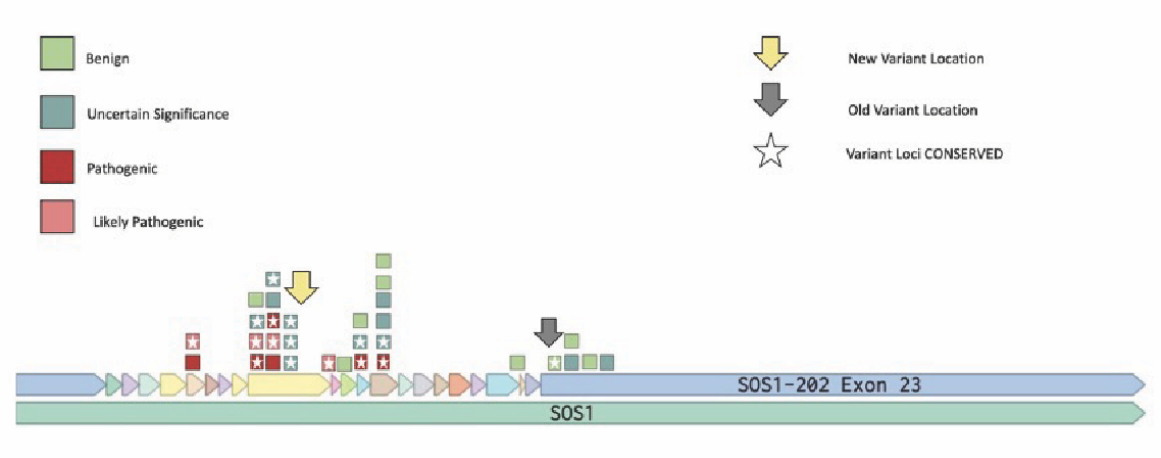
Image of hereditary gingival fibromatosis associated characteristics in a human patient and microscopy of N2 C. elegans hermaphrodites



A patient experiencing lack of eruption of front teeth and swollen appearance of the gingival tissue, preventing the closure of the lips (A). N2 hermaphrodite *C. elegans* nematodes across multiple life stages imaged at 25x with a dissecting microscope and a Motic camera (B). HGF patient photo adapted from a source (Alminana-Paster *et al*., 2017).

Figure 2

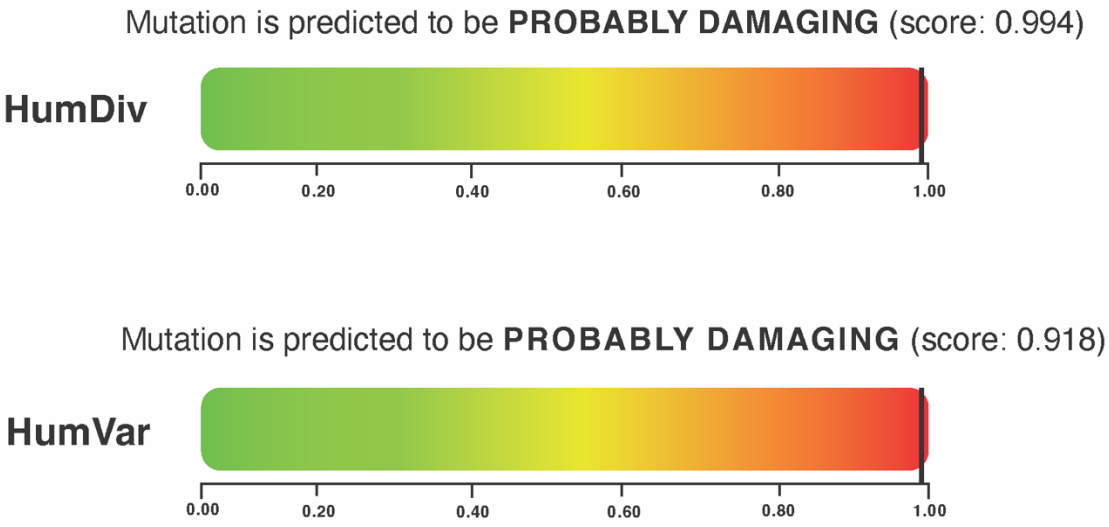
Mutational analysis of HGF associated SOS1 missense variants



Thirty HGF associated *SOS1* missense variants were examined for evolutionary conservation of the variant loci across human and nematode through multiple sequence alignments utilizing Benchling. Ten genetic variants were selected from each clinical significance classification ranging from benign to uncertain to pathogenic from ClinVar. This is a summary graphic of these combined bioinformatics analyses.

Figure 3

Results of the PolyPhen-2 analysis predicting the potential impact of the SOS1 VUS c.1289A>G (p. Asp430Gly) amino acid substitution on the human SOS1 protein



HumDiv and HumVar scores are reported from the PolyPhen-2 analysis.

Figure 4

Initial PCR results of the two designed DNA primer sets amplifying the SOS1 VUS locus in sos-1 through gel electrophoresis

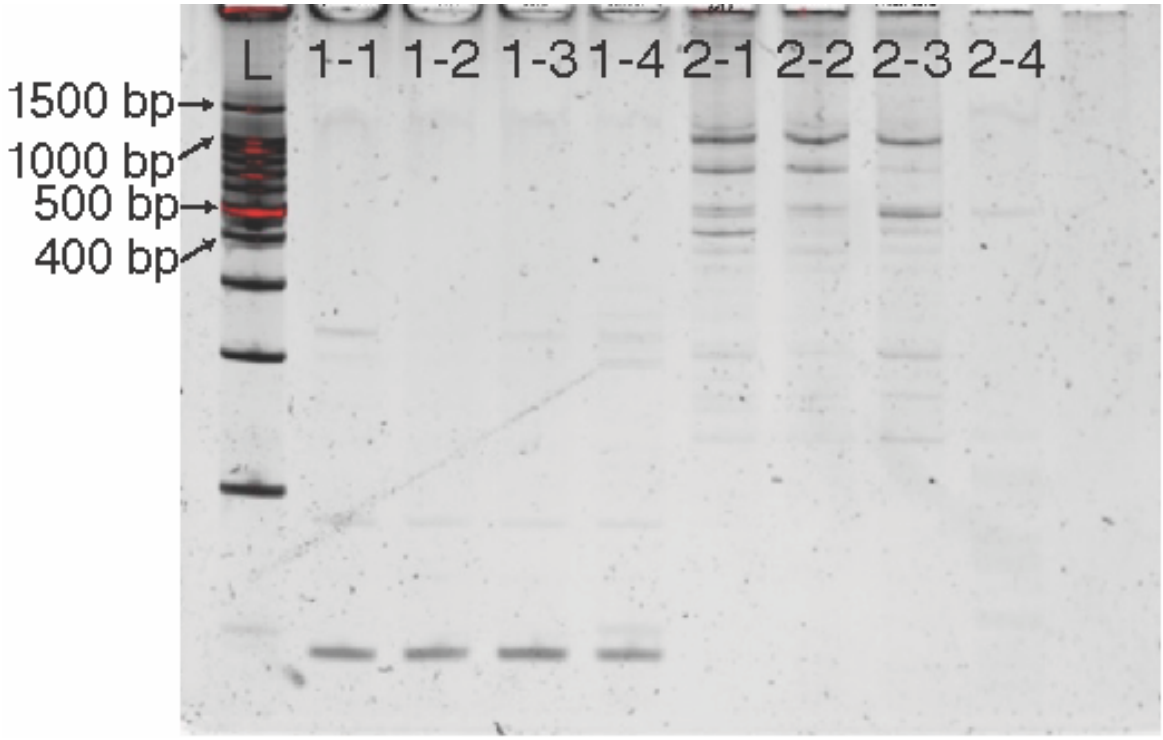


Image of Gel Red stained polyacrylamide gel showing initial PCR amplification for each primer set using a 55 ºC annealing temperature. Wildtype N2 *C. elegans* DNA template was used across four PCR reactions for each primer set. L = 100 bp ladder, 1-1/1-2/1-3/1-4 = primer set 1 PCR products, 2-1/2-2/2-3/2-4 = primer set 2 PCR products. The first two lanes contain digested nematode DNA, the third lane contains positive DNA control, and fourth lane contains no template control of water.

Figure 5

PCR results and annealing temperature optimization of primer set 1 amplification of the SOS1 VUS locus in sos-1 through gel electrophoresis

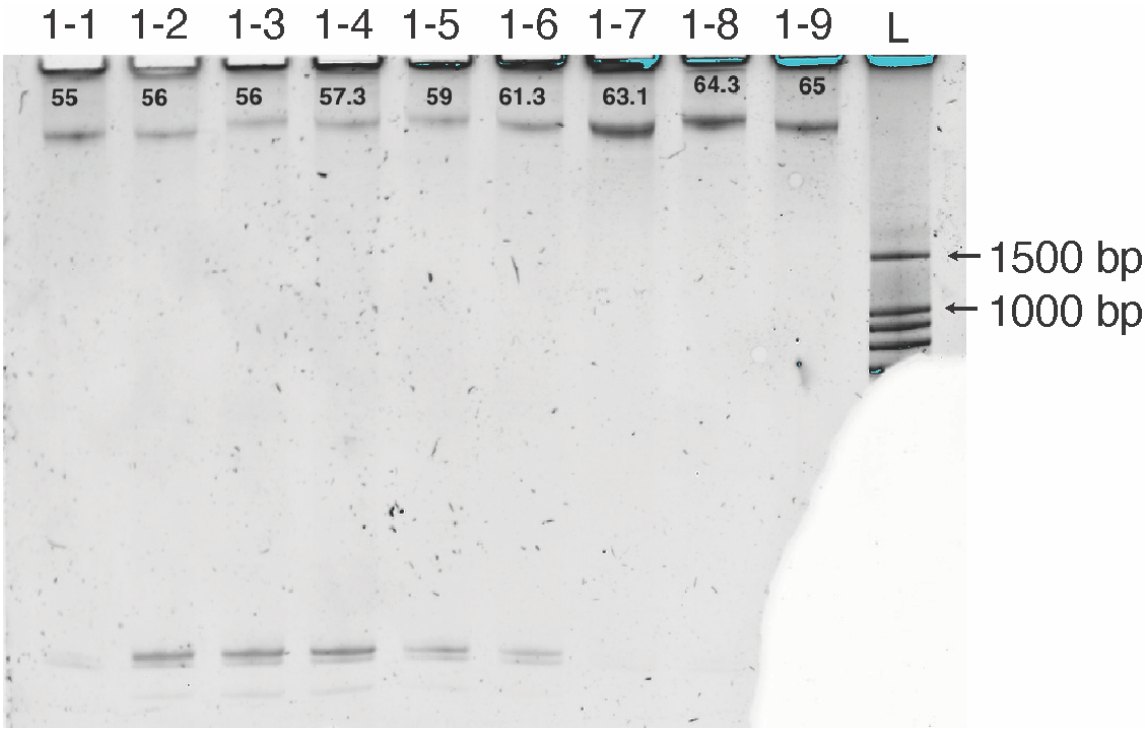


Image of GelRed stained polyacrylamide gel showing PCR amplification for primer set 1 using a thermal gradient for the annealing temperature. Wildtype N2 *C. elegans* DNA template was used for eight PCR reactions across a thermal gradient. Annealing temperatures tested in each lane for primer set 1 include: 1-1 = 55 ºC, 1-2/1-3 = 56 ºC, 1-4 = 57.3 ºC, 1-5 = 59 ºC, 1-6 = 61.3 ºC, 1-7 = 63.1 ºC, 1-8 = 64.3 ºC, 1-9 = 65 ºC, L = 100 bp ladder.

Figure 6

PCR results and annealing temperature optimization of primer set 2 amplifications of the SOS1 VUS locus in sos-1 through gel electrophoresis

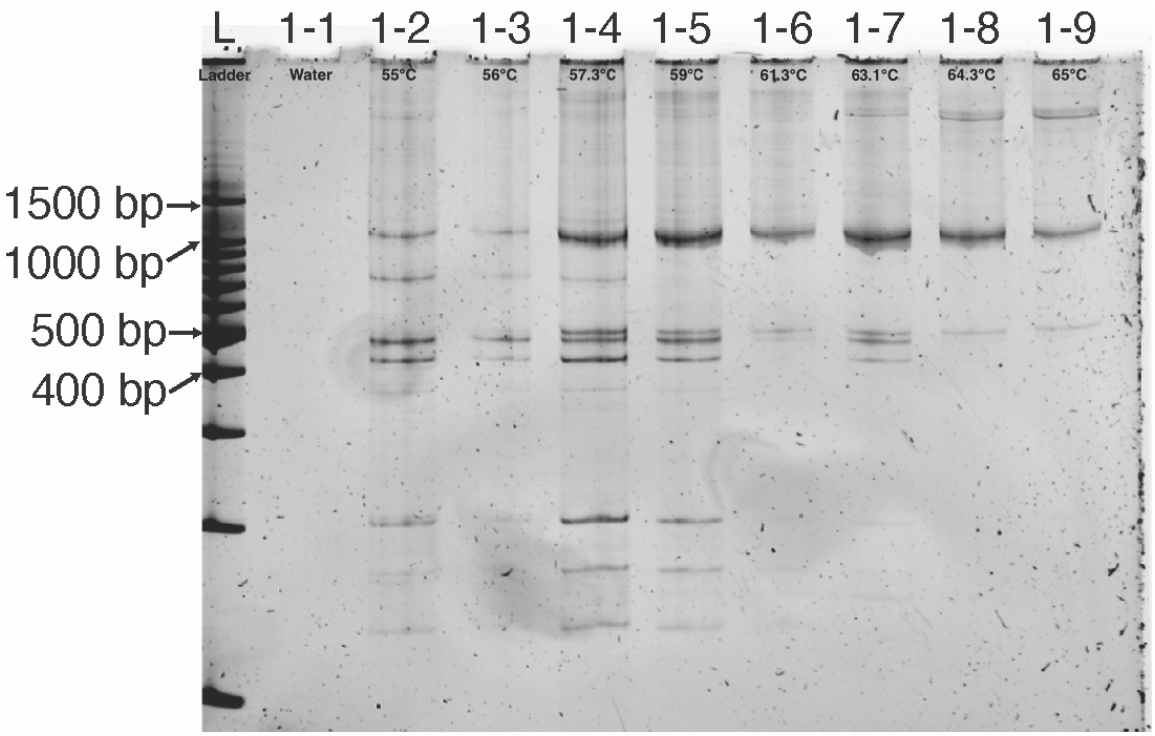
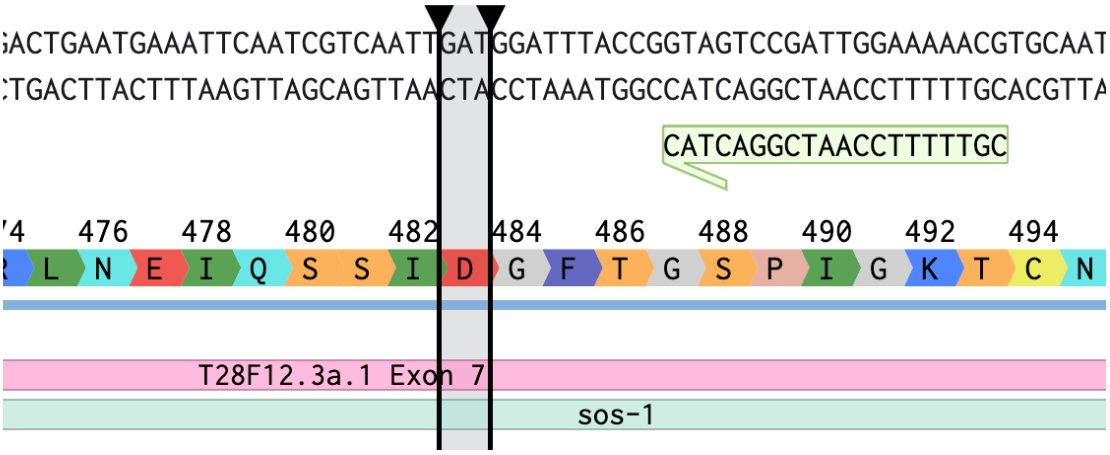


Image of GelRed stained polyacrylamide gel showing PCR amplification for primer set 2 using a thermal gradient for the annealing temperature. Wildtype N2 *C. elegans* DNA template was used for eight PCR reactions across a thermal gradient. Annealing temperatures tested in each lane for primer set 2 include: L = 100 bp ladder, 1-1 = water (no template control), 1-2 = 55 ºC, 1-3 = 56 ºC, 1-4 = 57.3 ºC, 1-5 = 59 ºC, 1-6 = 61.3 ºC, 1-7 = 63.1 ºC, 1-8 = 64.3 ºC, 1-9 = 65 ºC.

Figure 7

CRISPR targeting in the sos-1 gene



Schematic displaying the *sos-1* gene, the *SOS1* VUS loci (highlighted) and the CRISPR targeting region flanking that region (green arrow). Image obtained from design and analysis within Benchling.