Construction of a zebra fish knock-out mutant lacking a conserved bacterial homolog, *yfiH*

M.Sc. Thesis
2017
Submitted to

CENTRAL DEPARTMENT OF BIOTECHNOLOGY
Tribhuvan university, Nepal

In partial fulfilment of requirement for the degree of
2 year M.Sc. in Biotechnology

By
ANKIT JAISWAL

Supervisor
Prof. Dr. Tilak R. Shrestha
Central Department of Biotechnology
Tribhuvan university
Kirtipur, Nepal

Co-supervisor
Dr. Manjula Reddy
Senior Principal Scientist
Centre for Cellular and Molecular Biology
Hyderabad, 500 007, India

ACKNOWLEGMENTS

I would like to express deepest gratitude to my supervisor Prof. Dr. Tilak R. Shrestha for his excellent guidance and for providing me his valuable suggestions throughout my thesis. It was because of his constant support and suggestions that, I had an opportunity to work at CCMB, a precious opportunity for me that was very helpful to increase my knowledge in both practical and theoretical aspects. I would also like to thank my external supervisor Dr. Manjula Reddy for providing me an excellent atmosphere to work during my thesis period in CCMB. I am grateful to Dr. Rakesh Kumar Mishra, Director of CCMB for providing me an opportunity to work in his research laboratory and for official appproval.

I am thankful to Prof. Rajani Malla, former HOD of Central Department of Biotechnology, Tribhuvan University for granting me the permission to complete my dissertation at CCMB.

Truly heartfelt thanks to my dissertation mentor, Gopal Kushwah for his constant support and ideas to troubleshoot the problems. I highly acknowledge his valuable guidance and constant supervision during the entire dissertation period. Special thanks go to my lab members Dr. k. Ravinder, L Saisree, Sadiya Parveen, Balaji Venkataraman, Nilanjan Som, Moneca Kaul, Pavan K. Chodisetti, Dr. Anju Bala, Parna Sah, Avinash Srivastava, Mukesh Yadav, Shagufta Khan, Rahul Sureka, Poornima Naidu, who were always willing to help and give their best suggestions. It is a matter of great privilege for me to express my deepest acknowledgements towards all those who have coordinated directly or indirectly towards the completion of thesis work.

I would also like to thank my Mother, Brothers and Sisters. They were always supporting me and encouraging me with their best wishes throughout my educational career.

LIST OF ABBREVIATIONS

ATCC	American type culture collection
BLAST	Basic local alignment search tool
Cas	CRISPR associated
CHIP-Seq	Chromatin immunoprecipitation sequencing
CRISPR	Clustered regularly interspaced short palindromic sequences
crRNA	CRISPR RNA
Cys	Cysteine
Da	Dalton
DEPC	Diethylpyrocarbonate
Dpf	Day post fertilization
DSBs	Double stranded breaks
DTT	Dithiothreiotol
EtBr	Ethidium bromide
FMF	Familial Mediterranean fever
GWAS	Genome wide association study
HDR	Homology directed repair
His	Histidine
Hpf	Hour post fertilization
JIA	Juvenile idiopathic arthritis
KO	Knock out
LiCl	Lithium chloride
NCBI	National centre for biotechnology information
ng	Nanogram
NHEJ	Non homologous end joining
NOMID	Neonatal onset multisystem inflammatory disease
NUC	Nuclease
PAM	Protospacer adjacent motif
REC	Recognition lobe

SAM	Spacer acquisition motif
sgRNA	Single guide RNA
SNP	Single nucleotide polymorphism
TALEN	Transcription activator like effector based nucleases
tracrRNA	trans-activating CRISPR RNA
TRAPS	Tumour necrosis factor associated periodic syndrome
µg	Microgram
µL	Microlitre
UTR	Untranslated region
ZFN	Zinc finger nucleases

CONTENTS

Chapters Page No.

List of figures

List of Tables

ABSTRACT

YfiH gene codes an uncharacterized conserved protein widely distributed in 37 species from bacteria to human. The function of *yfiH* gene is still unknown, but preliminary data from *Brevibacterium lactofermentum* shows that the gene is not essential for the growth and viability. In vertebrates, the *yfiH* homolog is termed as *LACC1*. *LACC1* encodes the enzyme laccase (multicopper oxidoreductase) domain; no precise information is there about this gene except that it has some implications in some autoimmune and auto-inflammatory conditions in human like Rheumatoid arthritis, Juvenile idiopathic arthritis, Crohn's disease and also leprosy. To know functional aspect of this gene it was knockout from zebrafish using CRISPR/Cas technique. CRISP Repeats are components of an immune system which protects many bacteria and archaea against foreign genetic elements. They function by targeting these elements in a sequence specific fashion, guiding the nuclease Cas9 to degrade them. The CRISPR-Cas system from *S. pyogenes* has been adapted and is used as an *in vivo* genome editing tool in a variety of organisms like zebrafish, mouse, and monkey and so on. In this project, we used it to knock out (KO) our gene of interest (*LACC1*) in zebrafish. This was achieved by co-injecting a gene-specific single guide RNA (sgRNA) and Cas9 nuclease mRNA in early zebrafish embryos. Cas9 is RNA guided nuclease and creates double stranded break at the site where the single guide RNA bind. The cleaved DNA is repaired by Non Homologous End Joining (NHEJ) repair mechanism, generating unpredictable indel mutations. The induced mutations were detected by performing T7 endonuclease assay and were further confirmed by sequencing the target exons of the *LACC1* gene.

The exon 1 of *LACC1* gene was successfully mutated using CRISPR-Cas9 technique. The guide RNA and Cas9 injected fishes were with deformed body with elongated body with curved tail and deformed head, showing the rate of deformity 6 in per 10 injected larvae. The phenotypes of the *LACC1*-deficient zebra fish mutants have to be further confirmed.

Keywords:

CRISPR Cas9, sgRNA, knockout, *yfiH*, *LACC1*.

CHAPTER 1
INTRODUCTION

1.1 Background

The transfer or movement of genetic material between organisms (unicellular and multicellular) other than vertical transmission is known as lateral gene transfer. Gene flow between species is a minor contributor to genetic makeup, and it was considered to occur only between closely related species (Salzberg, 2001). But when the genetics of microorganism was studied this theory got changed. When antibiotic resistance among infectious bacteria was first understood and involvement of plasmid was revealed, several researchers claimed that there is a significant role of lateral gene transfer in evolution. These studies suggest that the role of lateral gene transfer in prokaryotic evolution is much more than it was thought before (Doolittle, 1999). Homologues of 44 different plastids encoded proteins were identified as functional nuclear genes of chloroplast origin providing evidence for endosymbiotic gene transfer to the nucleus in plants, thus in eukaryotes lateral gene transfer include transfer from organelles genome into the nucleus (Martin *et al.,* 1998).

Analysis of human genome revealed that more than hundreds of human genes appear likely to be transferred from bacteria through horizontal gene transfer during course of evolution (International Human genome sequencing consortium, Nature 409, 860, 2001). This discovery was of great interest as it implied that the bacterial infections have led to permanent transfer of genes into their hosts. Different theories were given for this observation. One study suggested that bacteria might be manipulating the human genome for their own benefit. Such an event would require that genes be transferred into the germ cell lineage and also that the transferred gene be stably maintained in the host, either by insertion into the chromosome or as an extra chromosomal element (Salzberg, 2001).

1.2 *YfiH* homolog in vertebrates

Bacterial *yfiH* is one such gene among those genes thus identified to be laterally transferred into the human genome. *YfiH* protein is widely distributed in 37 species from bacteria to human. The function of *yfiH* gene is still unknown, but preliminary data from *Brevibacterium lactofermentum* shows that the gene is not essential for the growth and viability (Honrubia *et al.,* 2001). His-71, Cys-107 and His-124 are highly conserved in most *yfiH*-like proteins of bacteria and vertebrates (Kim *et al.,* 2006). In *Escherichia coli*, *yfiH* is a 243 amino acid protein and has a molecular weight of 26,338 Da. In vertebrates, the *yfiH* homolog is termed as *LACC1*. *LACC1* encodes the enzyme laccase (multicopper oxidoreductase) domain. In humans, it is located in chromosome 13 at position

13q14.11 starting from 43,879,284 bp to 43,893,933 bp. It has a total length of 14,650 bases. It codes for a 430 amino acids protein and has a molecular mass of 47780 Da. A review of ENCODE project data for *LACC1*, including expression array, DNA hypersensitivity and CHIP-Seq tracks provide no obvious insight into the biology of this enzyme (Wakil *et al.,*2014) .

LACC1 plays role in a number of disease conditions in humans including Rheumatoid arthritis, Juvenile idiopathic arthritis, Crohn's disease and also leprosy.

Laccases (benzenediol:oxygen oxidoreductase; EC 1.10.3.2) belong to group of polyphenol oxidases that have copper atoms in the catalytic centre and are usually called "blue multicopper oxidases". Mammalian plasma protein ceruloplasmin and ascorbate oxidases of plants are other member of this group. Characteristic blue colour of laccase is due to the copper atom. It catalyses the reduction of oxygen to water accompanied by oxidation of substrate. Laccase is a nonspecific enzyme and has broad substrate range that varies from one laccase to the other. It is an oxidase that oxidizes substrate like polyphenols, methoxy substituted phenols, aromatic deamines and a range of other compounds (Baldrian, 2006).

Blast analysis of *yfiH* protein of *E.coli* with zebrafish and human laccase domain containing protein revealed that the identities in between them was 30% in case of zebrafish and 28% in case of human.

```
Query  20   STRIGGVSLPPYDSLNLGAHCGDNPDH---VEENRKRLFAAGNLPSKPVWLEQVHGKDVL  76
            +TR GG+S  P  S           D    V+EN +RL  A     +  +  + H  + +
Sbjct  195  TTRTGGISYIPTLSSFNLFSSSKRRDPKVVVQENLRRLANAAGFNVEKFYRIKTHHSNDI  254

Query  77   NLTG--EPYASKRADASYSNTPGTVCAVMTADCLPVLFCNRAGTEVAAAHAGWRGLCAGV  134
             + G  EP      D   +N  G   A + ADC+P++F +         AHAGW+G   GV
Sbjct  255  WIMGRKEP---DSYDGITTNQRGVTIAALGADCIPIVFADPVKKACGVAHAGWKGTLLGV  311

Query  135  LEETVSC----FADNPENILAWLGPAIGPRAFEVGAEVREAFMAADAKASTAFIQHGDKY  190
               TV+     +  + E+I+  LGP++GP  F +  E  EAF          F
Sbjct  312  AMATVNAMIAEYGCSLEDIVVVLGPSVGPCCFTLPRESAEAFHNLHPACVQLF--DSPNP  369

Query  191  LADIYLLARQRLASVGV------EQIFGGDRCTYTENETFFSYRRD  230
              DI    R  L   G+      +Q    + CT    + FFS+ RD
Sbjct  370  CIDIRKATRILLEQGGILPQNIQDQNQDLNLCTSCHPDKFFSHVRD  415
```

Figure 1.1: Blast analysis of *E. coli yfiH* with human reference sequence. Middle protein sequence shows the percentage identity i.e. 28% between *yfiH*(Query) and *LACC1*(Subject).

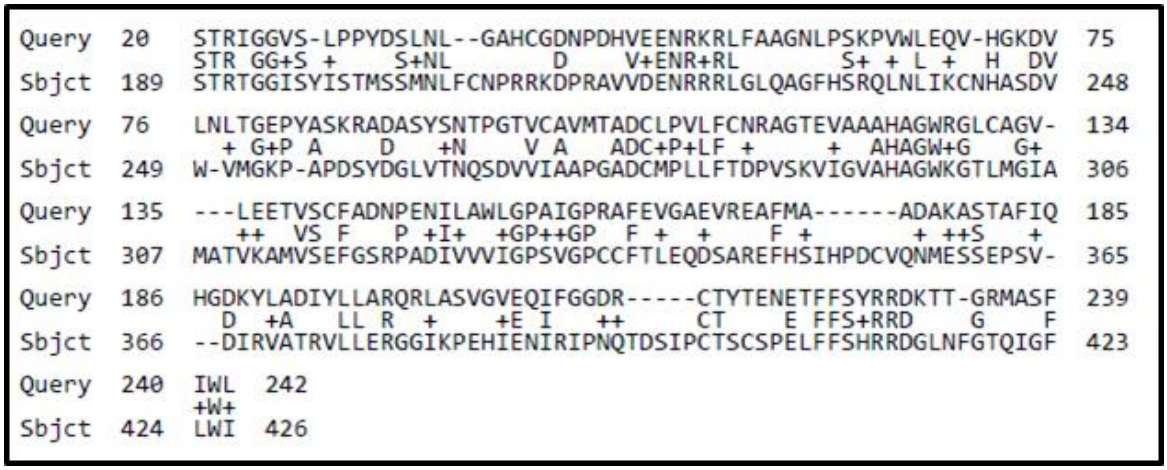


Figure 1.2: Blast analysis of *E. coli yfiH* with zebrafish reference sequence. Middle protein sequence showing percentage identity i.e. 30% between *yfiH*(Query) and *LACC1*(Subject).

1.3 CRISPR-Cas9 system

Clustered regularly interspaced palindromic repeats and CRISPR associated (Cas) genes are found in prokaryotes like bacteria and archaea. It provides adaptive immunity to the host; it recognizes and degrades foreign genetic material like bacteriophages and plasmids. In recent year, the Cas9 of type II CRISPR system from *Streptococcus pyogenes* has been used as a robust and efficient method of genome editing, Cas9 is a single RNA guided nuclease that recognizes the site where the sgRNA bind and creates double stranded break at the target site.

CRISPR regions were first observed in the K12 strain of *Escherichia coli* (Ishino *et al.,* 1987). CRISPR regions consist of similar-sized non-repetitive (spacer) DNA interspaced by short repetitive regions associated with Cas genes. CRISPR loci are a series of repeats of approximately 20 to 50 base pairs separated by specific non-repeat spacer sequences (Figure 1.3), in which the number of repeat-spacer units varies among organisms. These repeats were classified as a unique family of clustered repeats element present in more than 40% of bacteria and 90% of archaea (Mojica *et al.,* 2000). Upstream of the repeat-spacer regions of the CRISPR loci, there are three types of elements (Figure 1.3). Firstly, the leader sequence, which is an AT rich non-coding element composed of 300-500 base pairs (bp) and the promoter of this system. Secondly, the Cas genes which are only present in organisms containing the CRISPR loci and thirdly the trans activating CRISPR RNA which is a non-coding RNA that is required for the maturation of crRNA.

The CRISPR repetitive regions, some *Cas* genes as well as the leader sequence were found to be homologous between many prokaryote organisms. When the homology between the spacer sequences and bacteriophage and plasmid DNA elements (foreign

DNA) was detected a possible biological function as an immune system was hypothesized.

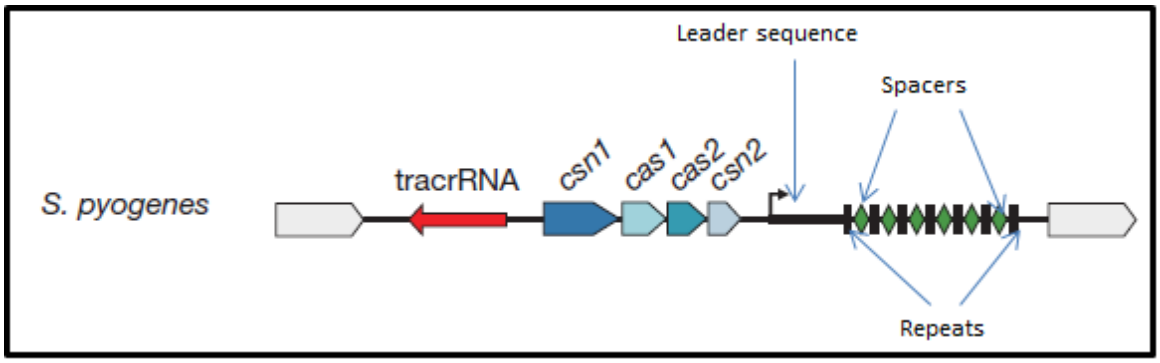


Figure 1.3: CRISPR loci of *Streptococcus pyogenes.* Csn1 is the Cas9 gene. (Deltcheva *et al.,* 2011).

In 2007, Deveau and Horvath shows that with each phage challenge there is a integration of new spacer in the CRISPR loci and the iterative addition of spacers increases the overall phage resistance of the host. Thus, they showed that the CRISPR Cas system provides adaptive immunity in the bacteria, similar to the adaptive immunity observed in higher eukaryotes. The broad existence of this system, present in 46% of the bacteria and more than 86% of the archaea, shows how important is this system to maintain the total fitness of the prokaryotes.

CRISPR Cas is a simple technique and only a single chimeric RNA is needed to guide the nuclease at the specific site. After the Cas9 nuclease activity, the DSBs created are repaired either by Non-Homologous End Joining (NHEJ) or Homology Directed Repair (HDR) mechanism.

Following the ground-breaking discovery of the CRISPR-Cas9 system, various experiments showed that this technique can be successfully used in human cell lines, bacteria and zebrafish to edit the genomes.

Efficient biallelic mutations in zebrafish and mice were generated showing phenotypes in F_0 generation; thus showing the capability of CRISPR-Cas9 system to generate knock-out (KO) strains (Sander and Joung, 2014). Studies have already proved that delivery of multiple guide RNAs can be done to generate mutations at various desired sites. CRISPR-Cas9 can not only be used for generating knockout mutant but also if simultaneous delivery of DNA is performed it stimulates HDR and leads to knock in strains (Hisano *et al.,* 2015).

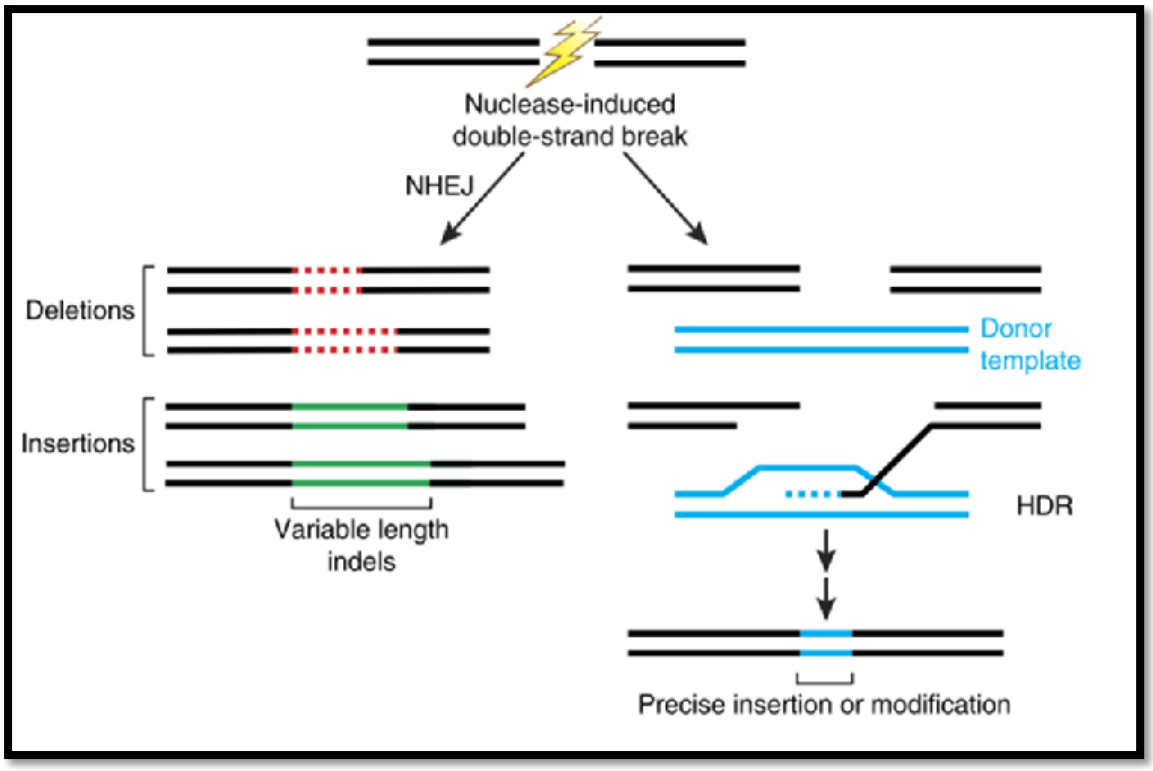


Figure 1.4: DSBs created by nuclease are repaired by NHEJ or HDR pathways. (Sander and joung, 2014).

1.4 Zebrafish

Zebrafish is a tropical fresh water fish native to Himalayan region and is a popular aquarium fish.

Kingdom: Animalia

Phylum: Chordata

Class: Actinopterygii

Order: Crypiniformes

Family: Cyprinidae

Genus: Danio

Species: rerio

It is gaining popularity in scientific field as a model organism for the study of functional biology. It is named for the five uniform pigmented, horizontal blue stripes on the side of the body which resembles the zebra's stripes. The male are torpedo shaped with gold stripes between blue stripes whereas the female have silver stripes instead of gold and have large white belly. The generation time for zebrafish is 3 months. The life span of zebrafish in captivity is around 2 to 3 years whereas in natural habitat it may extend up

to five years. Fertilization is ex utero and female lays more than 100 eggs in a single clutch. Fertilized eggs are transparent that make it a perfect model for research work (Dahm, 2006, Spence *et al.,* 2007).

The zebra embryo develops rapidly and all of the major organs are developed within 36 hpf. Thus, it is easy to screen them within short period of time if one is seeking a specific phenotype.

1.5 Use of zebrafish as animal model in scientific research

There are a number of reasons that make zebrafish, a good model to study the vertebrate functional biology. It is cost-effective, easy to maintain in large density, they lay more than hundreds of eggs in a single clutch, the embryo is completely transparent until 24 hpf and all of vertebrate organs are developed within 48 hpf. Therefore, in situ and in vivo embryonic development of normal and the mutant embryo can be easily observed. Also, a large number of the individuals can be analysed for the phenotypic changes within a short period of time.

About 71.4% of human genes have at least one obvious zebrafish orthologue. A list of human genes possessing at least one zebrafish orthologue was compared with 3176 genes bearing morbidity description that are listed on Online Mendelian Inheritance In Man and it was found that 82% of those morbid genes were related to at least one zebrafish orthologue. Same was done for 4023 human genes implicated in Genome Wide Association Studies (GWAS) and 76% of those genes were related to at least one zebrafish orthologue (Howe *et al.,* 2013). These work proved that the zebrafish is best model for studying disease condition in humans. In 2008 researchers from Boston children hospital developed a strain of zebra fish whose adult body have transparent skin. Thus, those strains are perfect to study disease like leukaemia and cancers as the detailed visualization of cellular activity, circulation, metastasis and other phenomena are easy (White *et al.,* 2008).

Functional biology refers to the branch of biology that studies the connections between genotype and phenotype. This involves perturbation of gene functions to check the effect and consequences on the other gene in the genome. Morpholinos i.e synthetic and stable oligonucleotides were used for many years for functional studies. But it has a large number of drawbacks like toxicity, off target activity and it does not lead to total loss of function as it act as transcript level not at gene level. So, the development of site specific nucleases replaces this technique as those came into limelight.

Though protein based nuclease like Zinc Finger Nucleases (ZFN) and Transcription Activator Like Effector-Based Nucleases (TALEN) were properly implemented in zebrafish

but low cost and robustness of CRISPR Cas9 technique make it an effective attractive and efficient method to edit zebrafish genome. Study have already proved that the mutations generated by this technique is heritable as it is pass to the germline at high rates with probability of biallelic mutations. Efficient multiplexing for the first time in zebrafish was done by using this technique only. By using CRISPR-Cas9 both knockout and knock in strains can be developed. Thus, the implementation of CRISPR Cas9 in zebrafish has opens a new dimensions to modify its genome and generate mutant lines.

1.6 Rationale of study

Before, only known fact about *LACC1* gene was that it has implications in some polygenic autoimmune and auto inflammatory disorders in human but recent studies suggest that mutation in this gene solely can lead to systemic juvenile idiopathic arthritis, still what is the precise function of this gene is unknown. Bacterial homolog of this gene i.e. *yfiH* function is also not clear, work done on *Brevibacterium lactofermentum* suggest that this gene has no role in growth and viability in bacteria. But recent work done in this gene (yet to be published) suggests that it plays role in peptidoglycan metabolism in bacteria. Thus, present study aims to know what function *LACC1* gene plays in human. For this zebrafish as a model organism is used and the *LACC1* gene is knock out from the zebrafish to understand the functional aspect of this gene by analysing the phenotypes and effect of this gene in proper growth and development of fish.

1.7 0bjectives

1.7.1 General objective

- To knockout *LACC1* gene in zebrafish using CRISPR/Cas9 system.

1.7.2 Specific objectives

- Designing the oligos for guide RNA synthesis.
- Cloning the oligos in pDR27 vector.
- In vitro transcription for guide RNA and Cas9 mRNA.
- Microinjection at one cell stage of zebrafish embryo.
- Analysis of mutation if any in the injected fishes.
- Phenotypic analysis of the mutants if any due to *LACC1* gene knockout.

1.8 Research hypothesis

Null hypothesis: There will be no affect in proper growth and development of the *LACC1* knockout mutants.

Alternative hypothesis: The growth and development in the *LACC1* knockout mutants will be affected.

CHAPTER 2
LITERATURE REVIEW

2.1 *YFIH* gene

In 1990s the concept of horizontal gene transfer was introduced to explain the observed incongruence in phylogenetic reconstructions using different genes. Horizontal gene transfer is an important factor that affects the evolution of Bacteria, Archaea and unicellular eukaryotes. Now evidences are rising that supports that the role of horizontal gene transfer is not confined to unicellular eukaryotes only but also in multicellular eukaryotes (Boto, 2010).

During vertebrate evolution, gene from bacteria were transferred to the vertebrate and analysis of human genome suggested that more than hundreds of bacterial gene were transferred to the human genome (International Human genome sequencing consortium, Nature 409, 860, 2001), among those genes thus identified, bacterial *yfiH* is one such gene that is laterally transferred into the human genome during course of evolution.

YfiH is an uncharacterized conserved protein that belongs to a vast protein family which contain conservative Pfam motif PF02578 (DUF152) (residues 29-243) and TIGR motif (TIGR00726) (residues 25-243).

YfiH protein is widely distributed in 37 species from bacteria to human and belongs to COG1496 consisting of 39 proteins (Kim *et al.,* 2006). The function of *YfiH* gene is still unknown, but study done on *Brevibacterium lactofermentum* show that when the *yfiH* gene located downstream of the cell division gene ftsQ and ftsZ was disrupted the phenotypes of the disrupted mutants were similar to that of wild type thus indicating that this gene have no role in cell growth and viability (Honrubia *et al., 2*001). His-71, Cys-107 and His-124 are highly conserved in most of *yfiH*-like proteins present in archae, bacteria, *Homo sapiens,* file fish, frog and mouse (Kim *et al.,* 2006). The homolog of the *yfiH* gene present in human is *LACC1,* no precise information is there about this gene except that it has some implications in some autoimmune and auto-inflammatory conditions in human.

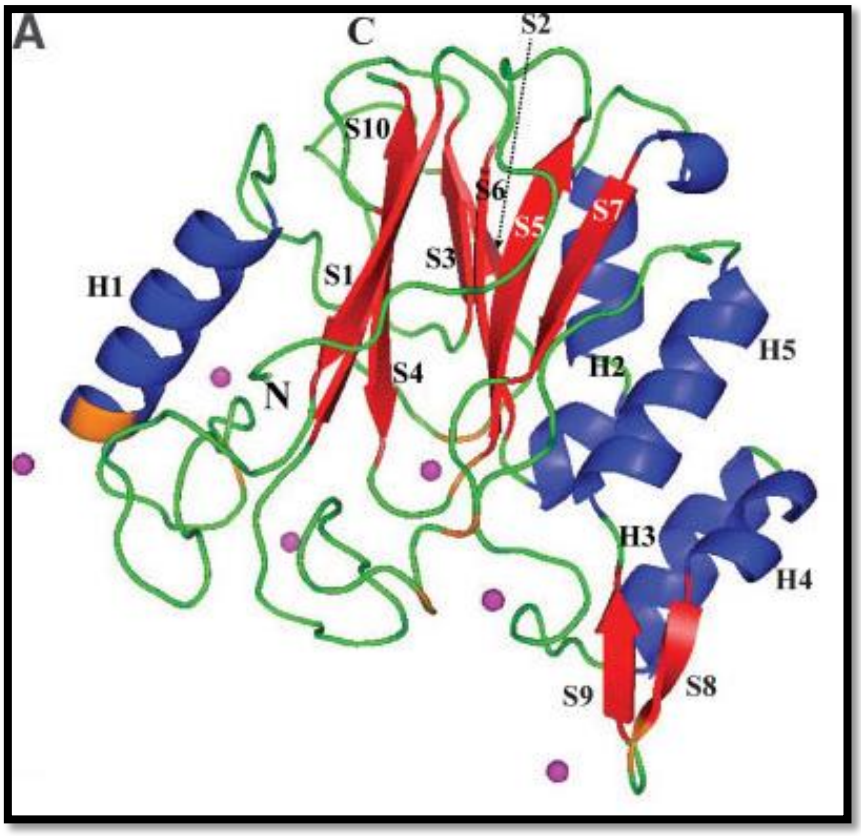


Figure 2.1. Crystal structure of *yfiH* protein. A: Ribbon diagram of *yfiH* protein. Helices are shown in blue, β sheets in red, and Zn^{2+} ions are indicated in magenta; orange indicates protein residues interacting with Zn^{2+} ions, among eight Zn^{2+} ions in the asymmetric unit, six Zn^{2+} ions are shown here, four of these are shared between the two protein molecules. The two Zn^{2+} ions not shown interact with the other protein molecule in the asymmetric unit. Secondary structure elements are also indicated in black. (Kim *et al.,* 2006).

2.2 *YfiH* gene homolog in human

YfiH is an interesting gene that is conserved in most eubacteria and also in vertebrates (from fishes to humans). The homolog of the *yfiH* gene present in human is *LACC1.*

LACC1 [laccase (Multicopper oxidoreductase) domain containing 1] is a protein coding gene. It is located in chromosome 13 and starts from 43,879,284 bp from p-terminal to 43,893,933 bp from p-terminal. It has size of 14650 bp and codes protein having 430 amino acids. No data are available for quaternary structure of this protein (http://www.genecards.org/cgi-bin/carddisp.pl?gene=LACC1). *LACC1* plays role in a number of auto inflammatory conditions in humans including Rheumatoid arthritis, Juvenile idiopathic arthritis, Crohn's disease and leprosy (Liu and Anderson, 2014, Wakil *et al.,* 2014, Sales‐Marques *et al.,* 2014). Previously it was thought that auto inflammatory condition are due to monogenic mutation only like in case of FMF (Familial Mediterranean Fever), TNF receptor associated periodic syndrome (TRAPS) and NOMID

(Neonatal Onset Multisystem inflammatory disease) but studies proved that auto inflammatory disorders can also happen due to polygenic disorder like in Crohn's disease, Type II Diabetes mellitus and systemic JIA (Juvenile Idiopathic Arthritis) {Masters *et al.,* 2009, Aksentijevich *et al.,* 2002}.

No significant associations have been made to date between systemic JIA and genes underlying monogenic auto inflammatory disorders like NLRP3, NOD2, MEFV or PSTPIP1 only exception was of Turkish patient with systemic JIA and MEFV mutation (Day *et al.,* 2008, Ayaz *et al.,* 2008).

Recent study done in Saudi Arabia of the patients of juvenile idiopathic arthritis (JIA) revealed that homoallelic missense mutation in *LACC1* leads to JIA (Wakil *et al.,* 2014). The p.C284R mutation is evolutionary invariant and part of copper reductase domain of *LACC1.* This study suggests that *LACC1* can also lead to auto inflammatory disorders if there is mutation in this gene solely (Wakil *et al.,* 2014). The first leprosy GWAS was conducted in a Chinese population, sample of 706 patients and 1225 controls were taken and it was found that there was a significant association between disease and single nucleotide polymorphisms in the genes *CCDC122-LACC1, NOD2, TNFSF15, HLA-DR-DQ* and *RIPK2* (Zhang *et al.,* 2009). A subsequent study was done in 492 patients in New Delhi and 382 patients in Kolkata, same was done for 273 cases and 221 controls from Mali West Africa and association between leprosy and SNPs at *C13orf31* and *CCCDC122* was found (Wong *et al.,* 2010). Subsequent work done on these genes proved that there is association between leprosy and Single Nucleotide Polymorphism (SNPs) at *CCDC122-LACC1* and *NOD2* (Sales-Marques *et al.,* 2014).

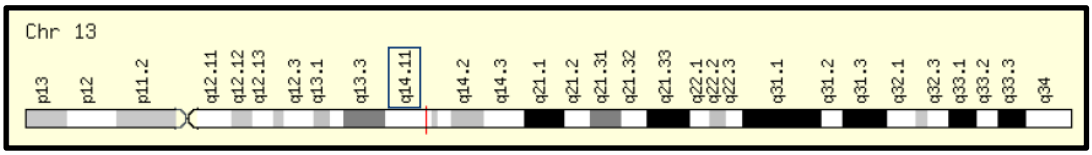


Figure 2.2: Genomic view for *LACC1* gene (shown in blue box) in humans. (http://www.genecards.org/cgi-bin/carddisp.pl?gene=*LACC1*).

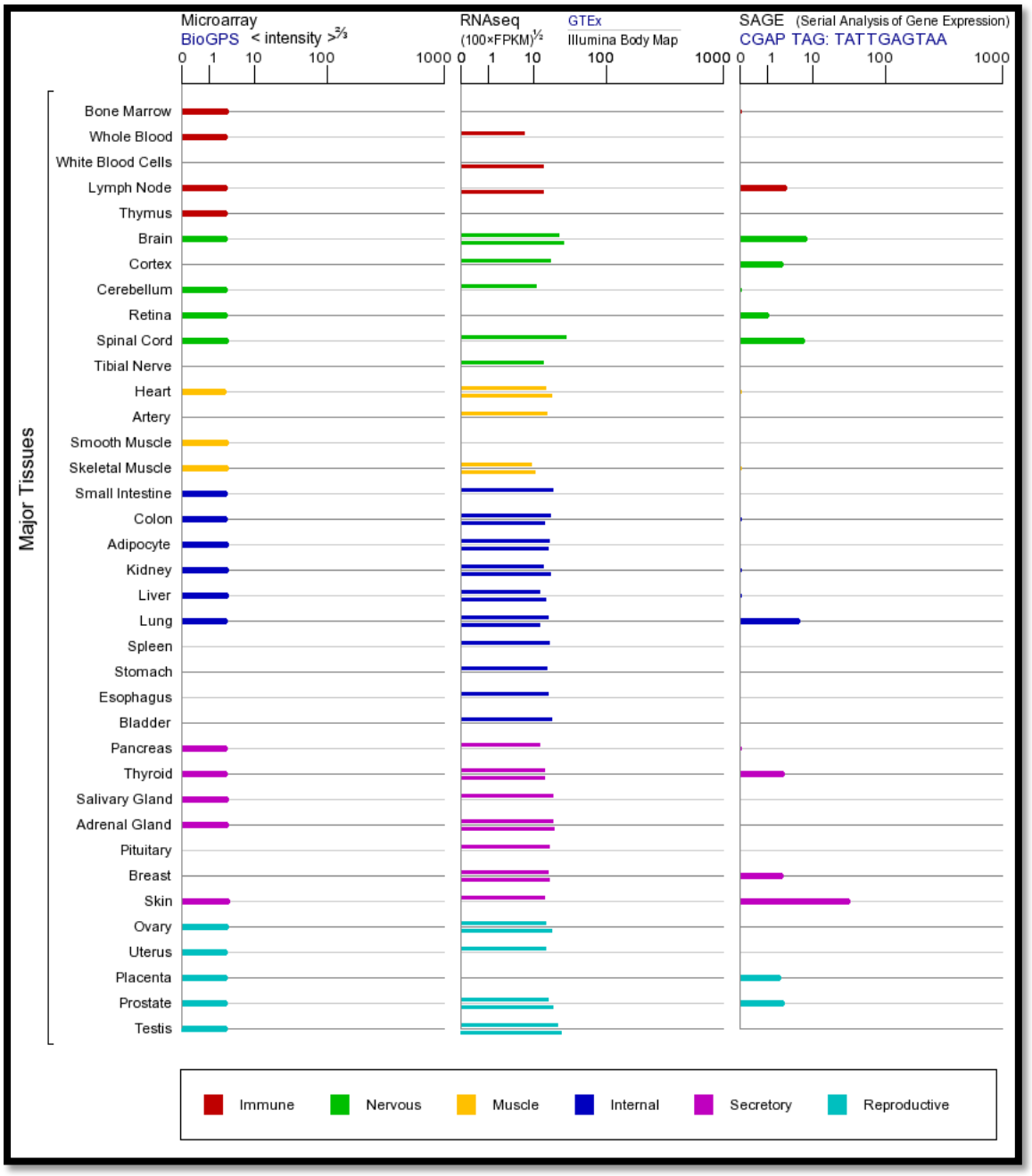


Figure 2.3: mRNA expression in normal human tissues for *LACC1* gene. (*http://www.genecards.org/cgi-bin/carddisp.pl?gene=LACC1*).

This gene is found to be overexpressed in testis, breast, spleen and stomach.

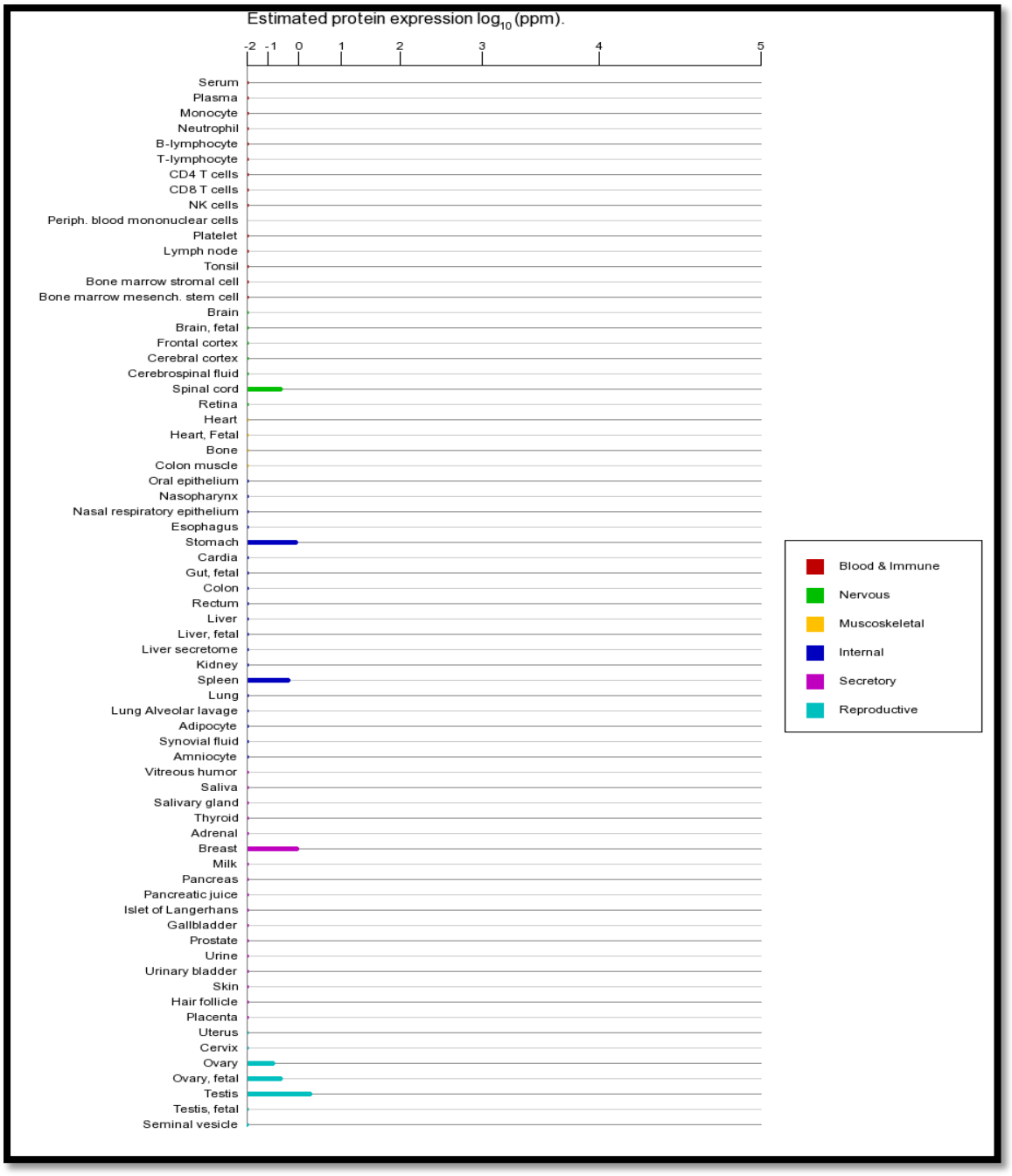


Figure2.4: Protein expression in normal tissues and cell lines in human. (*http://www.genecards.org/cgi-bin/carddisp.pl?gene=LACC1*).

2.3 Genome editing

Genome editing is a type of genetic engineering in which DNA is either deleted from the genome of organism or inserted into the genome of the organism using engineered nucleases or so called molecular scissors. These nucleases create double strand breaks at the target site which is repaired by non-homologous end joining or through homologous recombination resulting in targeted mutations. There are four genome editing tools available and used: Meganucleases, Zinc finger nucleases (ZFNs),

Transcription Activator like Effector-based Nucleases (TALEN) and the CRISPR-Cas system (Esvelt and Wang, 2014, Puchta and Fauser, 2013). The CRISPR Cas system is the most recent and cheap method of genome editing.

2.4 CRISPR/Cas9

The CRISPR story started in 1987 when Nakata and colleagues reported a curious set of 29 nucleotides that were arranged as direct repeats with 32 nucleotides as spacing, downstream of *iap* gene while they were studying the iap enzyme involved in isoenzyme conversion of alkaline phosphatase in *E .coli* (Ishino *et al.,* 1987).They found that unlike most repetitive elements that take form of tandem repeats like TALE repeats monomer, these repeats were interspaced by 32nt non repetitive sequence. With due course of time as more and more microbial genome were sequenced additional repeat elements were reported in archaea and bacteria. These repeats were classified as a unique family of clustered repeats element present in more than 40% of bacteria and 90% of archaea. It was named as Short Regularly Spaced Repeats, that were typically short partially palindromic sequences of 24-40 bp, containing inner and terminal repeats of 11 bp (Mojica *et al.,* 2000).These findings draws the interest of scientists in such microbial repeats.

The name CRISPR was coined by Jansen for those loci having directs repeats, with different size from 21 to 37 bp interspaced by similarly sized non-repetitive sequence. Also several cluster of CRISPR associated genes were identified to be well conserved and located adjacent to repeat elements. A common leader sequence of 300- 500 bp was also found at one end of the loci. The direct repeats and the leader sequences were similar within same species and dissimilar within different species (Jansen *et al.,* 2002). On the basis of these genes CRISPR Cas system was divided into three types. Type I and III contains multiple cas proteins that form complex with crRNA to facilitate recognition and destruction of nucleic acids (Bronus *et al.,* 2008).

Type II system has a significantly reduced number of Cas proteins. In 2005 a new insight about this system came when the analysis of spacer sequence separating the individual direct repeats was done, spacers regions were derive from transmissible genetic elements such as bacteriophage or conjugative plasmids (Mojica *et al.,* 2005; Pourcel *et al.,* 2005; Bolotin *et al.,* 2005). Tang *et al* 2002 shows that small non-messenger RNA were transcribed from CRISPR loci and Mojica *et al* 2005 shows that viruses were unable to infect archaeal cells carrying spacers, acquired from the virus protospacer sequences thus suggesting that the CRISPR loci plays role in immunity in bacteria and archaea.

After these studies it was clear that CRISPR loci plays role in immunity in bacteria and archaea but nobody knows how, the mechanism of this system was still to be revealed. Several hypothesis were given by people working on this field, one was that CRISPR

spacers act as small RNA guides to degrade viral transcripts in a RNAi like mechanisms, they suggested that the spacers region somehow act as prokaryotic siRNA by base pairing with the target mRNA and promote their degradation and lead to translational shutdown (Makarova *et al.,* 2006), other hypothesis given by Bolotin *et al.,* 2005 suggested that CRISPR spacers is transcribed as antisense RNA and inhibit phage propagation as studies already suggests that the antisense RNA inhibit phage propagation. First experimental evidence for the natural role of type II CRISPR system as an adaptive immunity system was given by Danisco, Horvarth and colleagues while working with the dairy production bacterial strain *Sterptococcus thermophiles* at the food ingredient company, demonstrating that CRISPR spacers dictate target specificity while some Cas enzymes is involved in spacer acquisition and some cas gene leads to phage DNA degradation through its nuclease activity (Barrangou *et al.,* 2007).

As more about CRISPR system was known researchers were more accelerated to find more about this system to unravel more details of each type of CRISPR system. Protospacer adjacent motif may direct the type II Cas9 nuclease to cleave DNA was suggested by Bolotin *et al.*, in 2005, Deveau *et al* produces mutations in protospacer adjacent motif (PAM) sequence of the *Streptococcus thermophilus* bacteriophage and demonstrated that PAM mutations in phage genomes circumvented CRISPR interference (Deveau *et al.,* 2007).

As everybody were in a race to know, what from this system can be programmed so that it can be used in genome modification, Moineau and colleagues inactivated the Cas5 (now known as Cas9) and Cas7 and showed that the inactivation of Cas5 confer resistance to phages but Cas7 mutation has no effect and bacteriophage DNA was cleaved thus they were the first to reveal that Cas9 is the only enzyme within the Cas gene cluster that mediates target DNA cleavage (Garneau *et al.,* 2010).

After that Charpentier and colleagues did differential RNA sequencing of *Streptococcus pyogenes* and revealed a key component in the biogenesis and processing of CRISPR RNA (crRNA) in type II CRISPR systems a noncoding trans activating crRNA (tracrRNA) with 24 nucleotide complementary to the repeat region of pre-crRNA, tracrRNA hybridizes with the crRNA to facilitate RNA guided targeting of Cas9 (Deltcheva *et al.,* 2011).

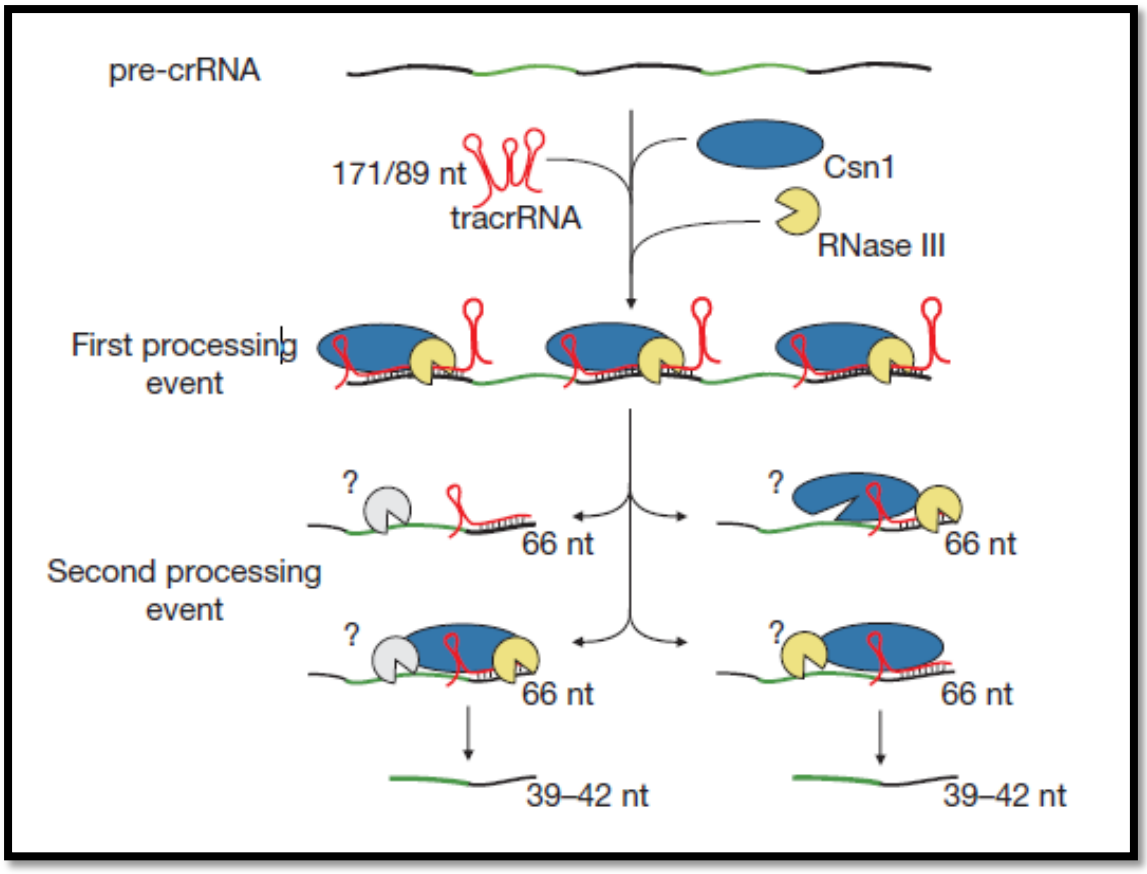


Figure 2.5: Model for tracrRNA-mediated crRNA maturation involving RNase III and Csn1. (Deltcheva *et al.,* 2011).

This study suggests that there are three key components in type II CRISPR Cas system i.e. Cas9, the mature crRNA and tracrRNA. By the time the use of programmable nuclease like ZFs and TALEs were gaining popularity in field of genome editing, but they have some complications so it was obvious to think that perhaps Cas9 could be used as an alternative to those system and a race was on to harness Cas9 for genome editing.

In 2011 work done by Siksnys and colleagues first demonstrated that the type II CRISPR system is transferrable, in that transplantation of the type II CRISPR locus from *Streptococcus thermophilus* into *Escherichia coli* provides heterologous protection against plasmid transformation and phage infection (Sapranauskas *et al.,* 2011). Biochemical characterization done by groups of Charpentier, Doudna, and Siksnys showed that purified Cas9 from *Streptococcus thermophilus* or *Streptococcus pyogenes* can be guided by crRNA to cleave DNA in vitro (Jinek *et al.,* 2012; Gasiunas *et al.,* 2012). A single chimeric RNA having sequence of crRNA and trcrRNA can be constructed to facilitate DNA cleavage by Cas9 in vitro (Jinek *et al.,* 2012).

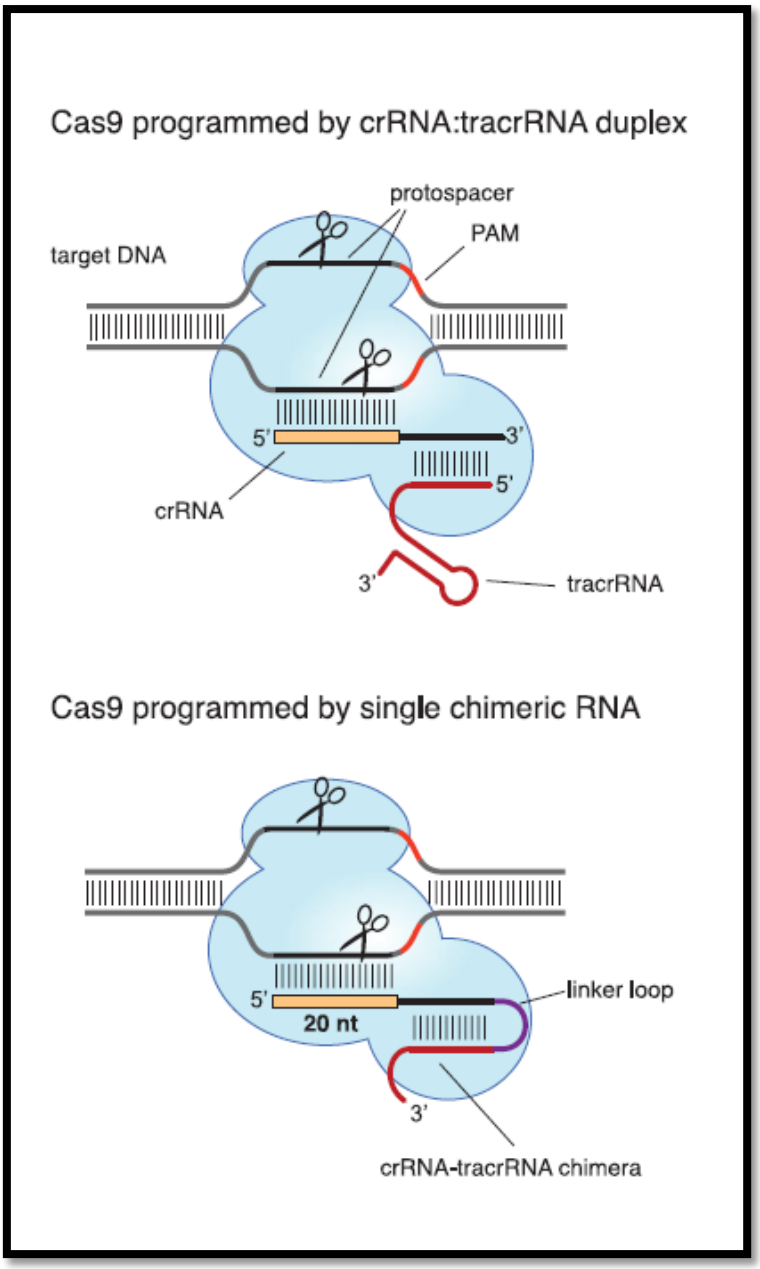


Figure 2.6: In type II CRISPR/Cas systems, Cas9 is guided by a two-RNA structure formed by activating tracrRNA and targeting crRNA to cleave site-specifically–targeted dsDNA. (Bottom) A chimeric RNA generated by fusing the 3′ end of crRNA to the 5′ end of tracrRNA. (Jinek *et al.,* 2012).

In 2013 work done by Cong et al and Mali et al showed how a single custom guide RNA can be used to direct Cas9 to create double stranded break at the target site, they successfully engineer type II CRISPR system from *Streptococcus thermophilus* for genome editing and successfully did multiplex genome editing in human and mouse cells by using their RNA guided editing tool (Cong *et al.,* 2013; Mali *et al.,* 2013). Heterologous expression of crRNA-tracrRNA hybrids as well as sgRNAs guides Cas9 to cleave at a specific site within the mammalian genome to stimulate NHEJ or HDR mediated genome editing.

2.4.1 Structural organization of Cas9

Cas9 proteins have two nuclease domains, RuvC and HNH domain. The HNH is a single nuclease domain whereas the RuvC domain is divided into three subdomains the RuvC I lies near to the N terminal region of Cas9 and RuvC II/III flanking the HNH domain near the middle of the protein. Structural studies were performed to know the exact structural mechanism of CRISPR Cas mediated DNA cleavage. Single particle electron microscopy of the *Streptococcus pyogenes* Cas9 (SpCas9) revealed a large structural rearrangement between apo-Cas9 unbound to DNA and Cas9 in complex with the crRNA and tracrRNA, forming a central channel to accommodate the RNA DNA heteroduplex (Jinek *et al.,* 2014).

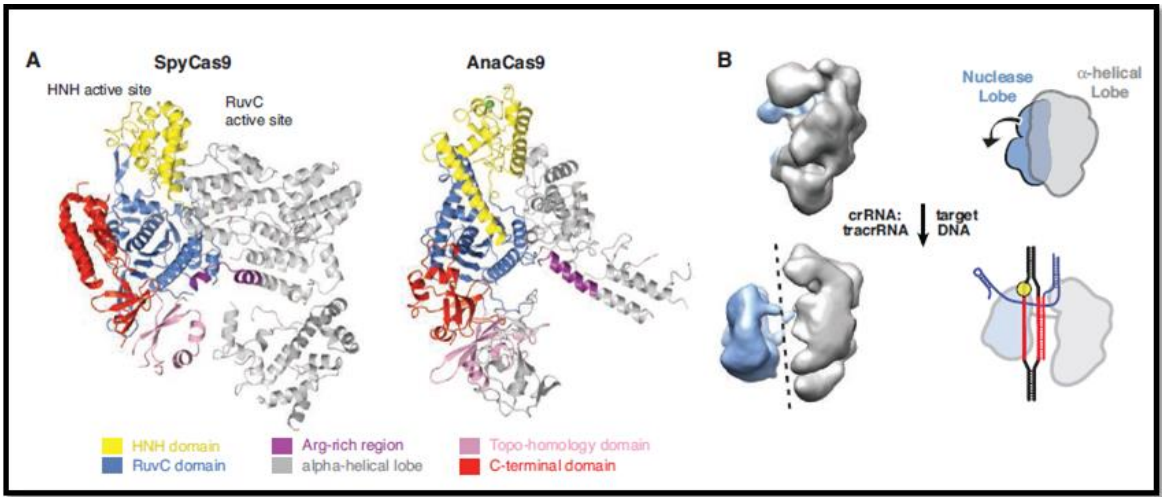


Figure 2.7: Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. (A) Crystal structures of *S. pyogenes* (SpyCas9) and *A. naeslundii* (AnaCas9) Cas9 proteins. (B)Left: Negative-stain EM reconstructions of apo- SpyCas9 (top) and SpyCas9-RNA-target DNA complex (bottom) show that nucleic acid binding causes a reorientation of the nuclease (blue) and α-helical (gray) lobes in SpyCas9. Right: Cartoon representations of the structures. tracrRNA, transactivating crRNA. (Jinek *et al.,* 2014).

High resolution structure of SpCas9 with complex with the RNA and complementary strand of the target DNA revealed the domain organization to compromise of a α-helical recognition lobe (REC) and nuclease (NUC) lobe consisting of the HNH domain, assembled RuvC subdomains and a carboxy-terminal domain responsible for PAM interaction (Nishimasu *et al.,* 2014). These studies suggested that when on unbound state of Cas9 the HNH domain active site is blocked by the RuvC domain and is positioned away from the REC lobe. So apo-Cas9 cannot bind and cleave to DNA on its own, sgRNA acts as a scaffold around which Cas9 fold and organizes the various domains.

Crystal structure of SpCas9 with complex with sgRNA reveals that an arginine rich bridge helix within the REC lobe is responsible for contacting the 3’ 8-12 nt of the RNA DNA

heteroduplex, mutational analysis demonstrated that mutation in the bridge helix markedly reduces the DNA cleavage (Nishimasu *et al.,* 2014). Cas9 is exclusively associated with type II CRISPR locus and serve as the important type II gene. On the basis of associated cas gene type II is further sub divided into three sub types IIA-IIC (Makarova *et al.,* 2011; Chylinski *et al.,* 2013).

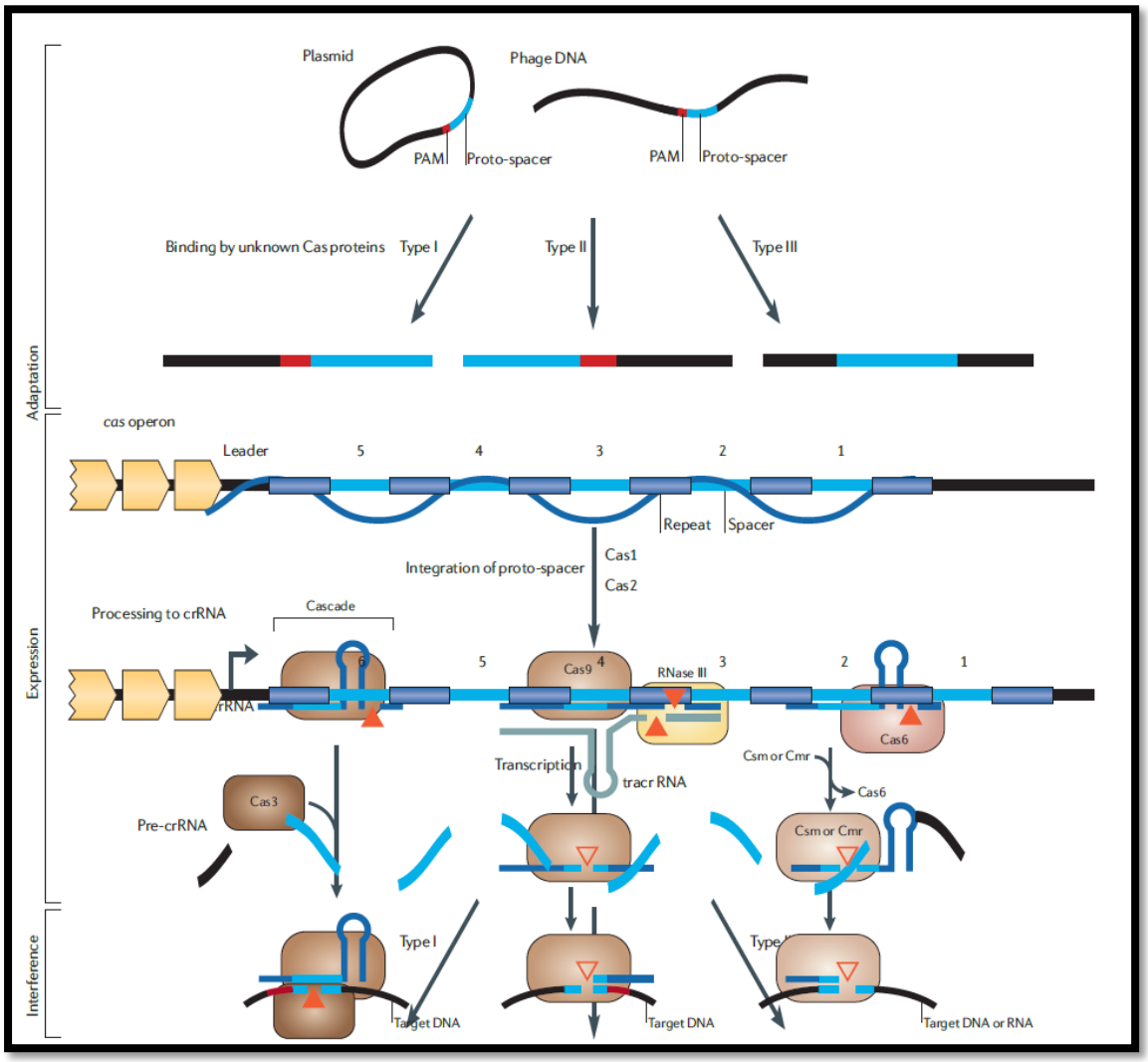


Figure 2.8: The three stages of CRISPR–Cas action. (Makarova *et al.,* 2011).

Type II CRISPR loci consist mainly of cas9, cas1 and cas2 genes, type IIC consists minimal number of genes whereas type IIA and IIB consists of additional csn2 and cas4 gene respectively (Chylinski *et al.,* 2013). Unlike type I and III CRISPR system which are found in both archaea and bacteria type II is found only in bacteria. The length of Cas9 varies in the three subtypes of the type II system this is due to the variable conservation of the REC domain. For example, the type IIC Steptococcus Cas9 is less compact and has smaller REC lobe with different orientation as compared to its ortholog in *Actinomycetes naeslundii* (Jinek *et al.,* 2014).

2.4.2 Protospacer Adjacent Motif (PAM)

Protospacer adjacent motif is critical feature of this system it plays role in determining self-versus non self as direct repeats don't contain PAM sites (Shah *et al.,* 2013). Stenberg et al in 2014 shows through single molecule imaging that Cas9-crRNA-tracrRNA complexes first associate with the PAM sequences throughout the genome and then proceed directionally toward the distal end of the target site, also PAM interactions trigger Cas9 catalytic activity.

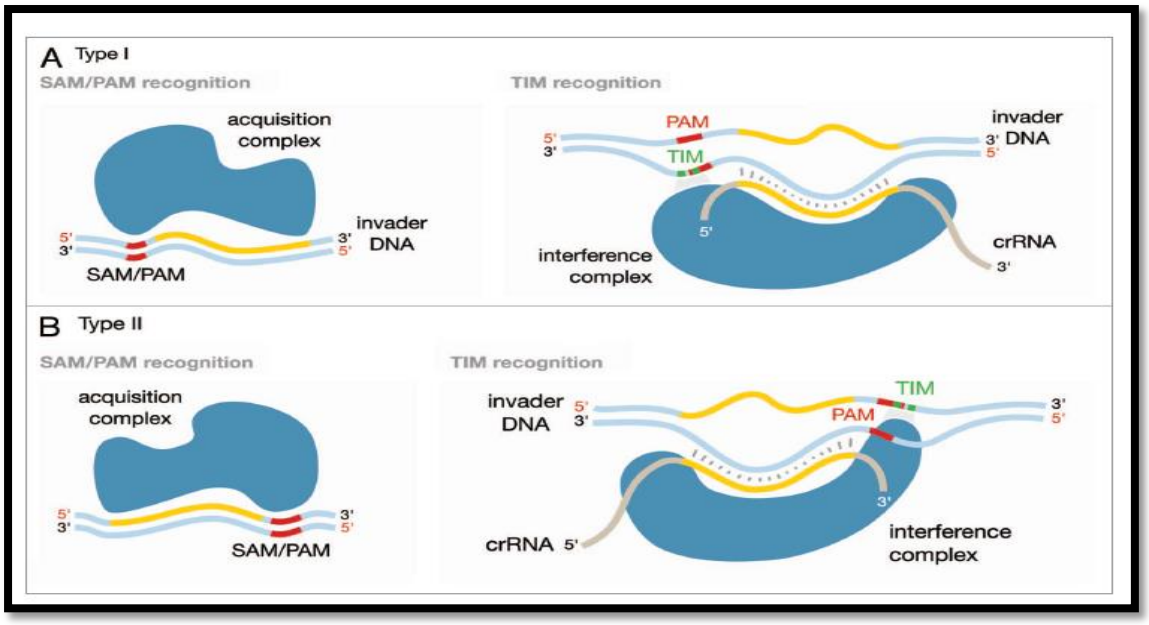


Figure 2.9 (A) The spacer acquisition motif (SAM) is recognized on the invader DNA by the Cas protein acquisition complex, which leads to the protospacer being excised by a putative ruler mechanism and reinserted into a CRISPR locus by another putative ruler mechanism. During interference by type I systems the target interference motif (TIM), on the crRNA-complementary DNA strand, is recognized by the Cas protein-crRNA complex where both TIM recognition and crRNA annealing are required for successful invader cleavage. (B) In type II systems, the SAM/PAM motif is inferred to be recognized by a mechanism related to the type I system but inverted on the dsDNA whereas TIM recognition occurs on the non-complementary DNA strand to the crRNA. (Shah *et al.,* 2013).

The complexity of PAM sequence also determines the overall DNA targeting space of Cas9. For example, the 5'NGG of SpCas9 allows it to target on average every 8 bp within the human genome (Hsu *et al.,* 2013). PAM are specific that means each Cas9 orthologs have their own specific sites, even within same species, such as 5' NAAGAAW is for *Streptococcus thermophilus* CRISPR1 (Deveau *et al.,* 2007) and 5' NGGNG for *Streptococcus thermophilus* CRISPR3 (Horvarth *et al.,* 2007). PAM specificity can be modified, for instance orthologous replacement of PAM interacting domain from *Streptococcus thermophilus* CRISPR Cas9 with the corresponding domain from

Streptococcus pyogenes Cas9 successfully altered PAM recognition from 5' NGGNG to 5' NGG (Nishimasu *et al.,* 2014).

The number, position and distribution of mismatches between guide RNA and target DNA determines the ability of Cas9 to tolerate mismatches (Fu *et al.,* 2013; Hsu *et al.,* 2013; Mali *et al.,* 2013; Pattanayak *et al.,* 2013). Since spCas9 also recognizes the PAM site with 5' NAG sequence so it is important to consider both NGG and NAG PAMs in off target analysis. Though there should be good interactions between DNA and sgRNA for Cas9 to act but the Cas9 can bind semi transiently with only short stretch of complementary sequence between RNA and DNA, thus suggesting that Cas9 have many off target sites but cleaves only a small fractions of them (Wu *et al.,* 2014). Concentrations of Cas9 also plays role in determining the ability of Cas9 to tolerate mismatches at the target site, if present in higher concentrations it can tolerate more mismatches leading to higher off target activity(Hsu *et al.,* 2013; Pattanayak *et al.,* 2013).

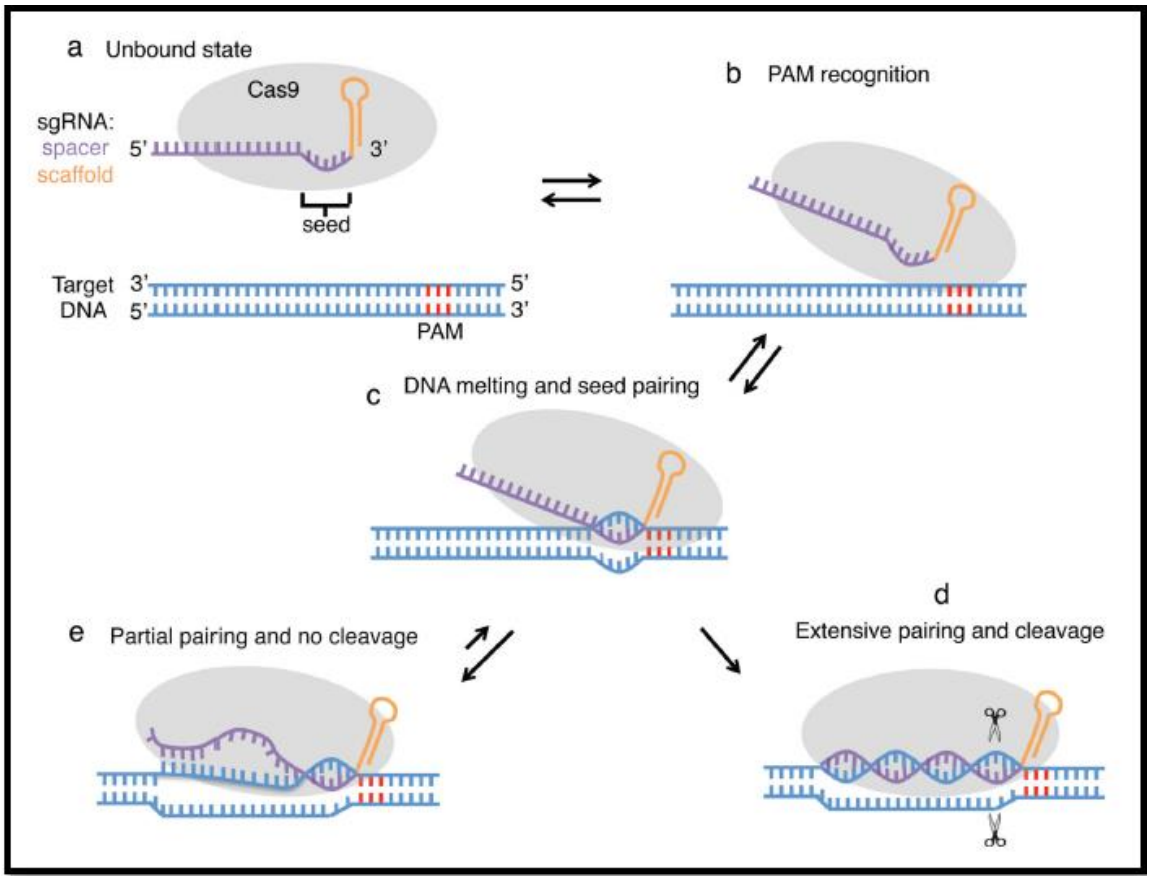


Figure 2.10: A model for Cas9 target binding and cleavage. (a) In the unbound state, Cas9 is loaded with sgRNA but not bound to DNA. The PAM region in the DNA is colored in red. (b) Recognition of the PAM by Cas9. (c) Cas9 melts the DNA target near the PAM to allow seed pairing. (d) If base pairing can be propagated to PAM-distal regions, the two Cas9 nuclease domains may be able to 'clamp' the target DNA and cleave it. (e) If only partial pairing occurs, there is no cleavage and Cas9 remains bound to the target. (Wu *et al.,* 2014).

2.5 Genome editing using CRISPR/Cas9

Cas9 from *Streptococcus pyogenes* has been widely used for efficient genome editing in a variety of species including rats, zebrafish, yeast, fruit fly, roundworm, mouse, common crops, pigs and monkey (Sander and Joung 2014).

Libraries for gRNA has been prepared using array based oligonucleotide synthesis; that can be used to encompass multiple gRNAs for every target gene in an organism. Whereas the short hairpin RNA libraries are used for generating knockdown mutant sgRNA libraries is used to generate knockout mutant (Sander and Joung 2014). Multiplex genome editing has also been successfully done in zebrafish and cynomolgus monkey using CRISPR/ Cas9 system (Kotani *et al.,* 2015; Niu *et al.,* 2014).

Either crRNA and tracrRNA (Cong *et al.,* 2013) or a custom guide RNA/single guided RNA (Mali *et al.,* 2013; Jinek et al., 2013; Cho *et al.,* 2013) can be used with the Cas9 for genome editing, as the crRNA or sgRNA contains a 20 nucleotide guide sequences complementary to the protospacer that directly bind to the target site. Protospacer adjacent motif (PAM) sequences present on the 3' end of the target sequence is the only requirement for the sgRNA to recognize and bind to the target site.

One main hallmark of the CRISPR Cas9 system is its ability to cleave multiple distinct target sites at once (Barrangou *et al.,* 2007; Garneau *et al.,* 2010) by converting a pre-crRNA containing many spacers into individual mature crRNA-tracrRNA complex. This unique ability of this system is used for multiplex genome editing. By using a number of sgRNA targeting different sites it is possible to edit multiple genes at once (Cong *et al.,* 2013). CRISPR/Cas9 technique is not only used for generating knockout mutants but as well it is used to generate knock in strains. If a donor vector having a gene of interest to be integrated with a flanking 40 bp homology arm to the target site is co injected with the sgRNA and Cas9, then there will be 60%-65% chances of integration of desired gene at the target site (Hisano *et al.,* 2015).

CHAPTER 3
MATERIALS AND METHODOLOGY

3.1 Preparation of media and solutions

3.1.1 Luria Bertani media: (for 1 litre)

10g Tryptone ,5g Yeast Extract ,10g NaCl was weighed and kept in 1 litre flask and water was added to make volume up to 1 litre. Then pH was adjusted to 7.2 and autoclaved at 121°C at 15 lbs pressure for 15-20 minutes for sterilization.

3.1.2 Luria Bertani agar

LB agar was prepared by making 1.5 % agar in LB broth and autoclaved at 121°C at 15 lbs pressure for 15-20 minutes for sterilization.

3.1.3 Kanamycin solution (50 mg/ml)

Three hundred seventy five mg of kanamycin was dissolved in 7.5 ml of fresh autoclaved water in the laminar hood. This was allowed to dissolve and then filtered through a 0.22 µm filter membrane. This was then dispensed into 1 ml aliquots and stored at -20°C.

3.1.4 LB kanamycin plates

LB Agar was melted by heating. It was allowed to cool to 45-50°C (Kanamycin degrades at high temperature). Kanamycin stock solution (50 mg/ml) was added such that the final concentration was 50 µg/ml in the LB Kan agar. This was poured into petri plates (~25 ml per plate) and allowed to solidify under sterile conditions (Laminar hood).

3.1.5 Preparation of Agarose gel

1.2 g of agarose was weighed and added to 100 ml of 1XTAE buffer in a 250 ml conical flask. The contents were heated till the agarose dissolved completely and a clear solution was formed. It was allowed to cool to a temperature of about 30-40°C. Once it reached optimum temperature, 4 µL of EtBr was added and mixed thoroughly. This was then poured in the casting tray set with an appropriate-sized comb and allowed to solidify.

3.2 Designing and cloning the oligonucleotides

3.2.1 Identification of zebrafish gene sequence to be targeted with CRISPR-Cas9 system

The target exon sequences of *LACC1* gene i.e., exon 1, exon 2 and exon 5 were retrieved from ensemble/UCSC genome browser. Thus, retrieved exon sequences were incorporated into the ZIFIT online CRISPR designing tool. The output results in 22 base pair of target sequences along with the forward and reverse oligonucleotides (oligos) for the target sequence. Those data of paired oligonucleotides were sent to Bioserve, India for oligos synthesis.

3.2.2 Annealing of the oligonucleotides

From thus ordered oligonucleotides 100 picomoles of stock oligonucleotides were prepared adding MQ in required volume as indicated by Bioserve, India. From stock of 100 pm of oligonucleotides, 10 pm of oligonucleotides were prepared by adding 50 µL of oligonucleotides in 450 µL of MQ.

50 µL from both forward and reverse oligonucleotides were added in eppendorf tube and incubated at 95^0C for 4 min in water bath and were allowed to cool at room temperature for 5 to 6 h for annealing of the oligonucleotides. The annealed oligonucleotides were stored at -20°C for further use.

3.2.3 Cloning the oligonucleotides in plasmid

The vector pDR274 was used for cloning the oligonucleotides since it has a guide RNA scaffold. For cloning the vector was digested with BsaI, the digestion reaction was performed as shown below in Table 3.1 and was incubated at 37^0C for 5 h.

Table 3.1 Restriction digestion setup for gRNA vecctor.

COMPONENT	VOLUME
pDR274	10 µL
10 X NEB buffer 3.1	2 µL
BsaI	2 µL
MQ	6 µL
Total	20 µL

After digestion of vector, ligation of the oligonucleotides in the vector was done. The oligonucleotides were ligated with the linearized vector making a ligation mixture as listed in Table 3.2 below.

Table 3.2 DNA ligation setup for ligation of oligos in vector.

REAGENTS	VOLUME
Linearized plasmid	2 µL
Annealed oligonucleotides	4 µL
10X T4 DNA ligase buffer	1 µL
T4 DNA ligase	1 µL

The ligation mixture thus prepared was incubated at 16^{o}C in water bath for 6-8 hour.

Characteristics: Vector size: 2147 bp. gRNA scaffold: 76 bp guide RNA scaffold for the Streptococcus pyogenes CRISPR/Cas9 system, T7 promoter: promoter for bacteriophage T7 RNA polymerase, cat promoter, kanamycin resistant gene, T7te terminator: T7 phage early transcription terminator, high-copy-number ColE1/pMB1/pBR322/pUC origin of replication.

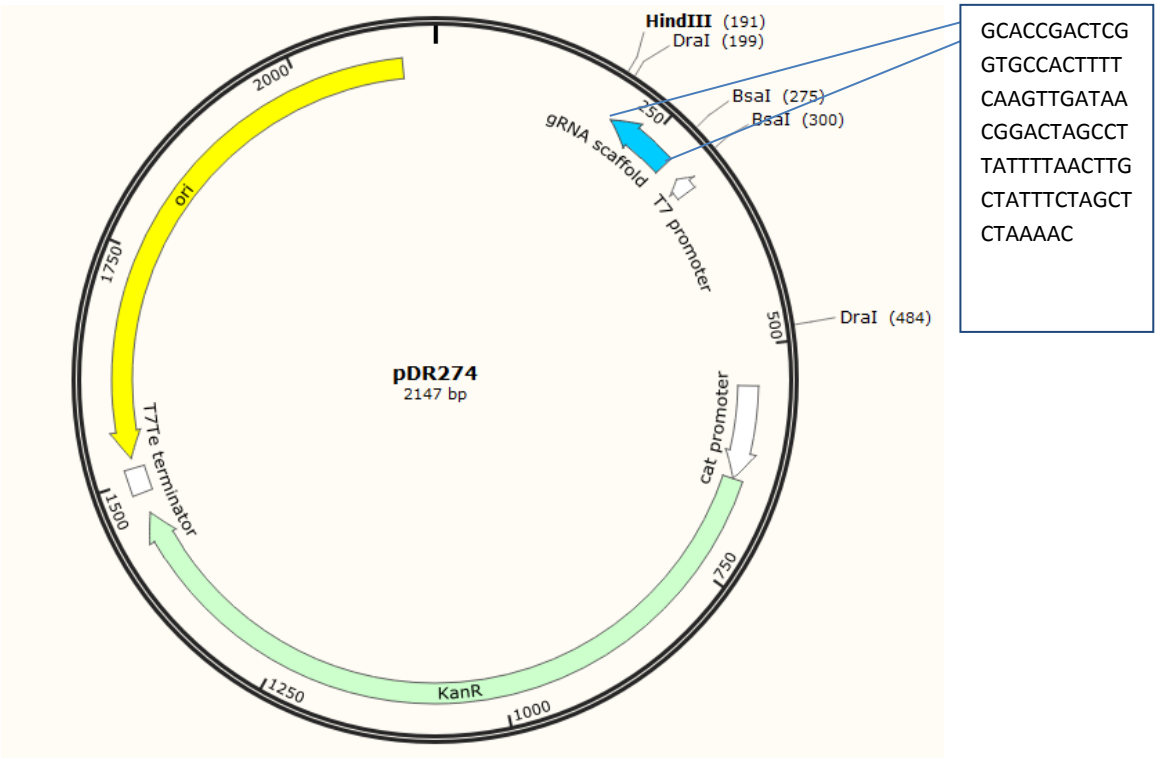


Figure 3.1: Map of vector pDR274. The box shows the sequence of guide RNA scaffold downstream of which target oligonucleotides are cloned.
(*http://www.snapgene.com/resources/plasmid_files/crispr_plasmids/pDR274/*).

3.2.4 Transformation into DH5α competent cells

After ligation of the oligonucleotides in the vector pDR274 the cloned vector were transformed into DH5α competent cells.

For this competent DH5α cells were thawed on ice and 1 µL of ligation mixture was added and mixed by tapping the tube at the bottom. Then, the tubes were incubated on ice for 30 minutes. After 30 min the tubes were taken out of the ice box and heat shock was given for exactly 90 seconds by keeping the tube in a waterbath at 42°C. Then those tubes were placed on ice for 3 minutes, 800 µL of LB broth was added into each tube and incubated at 37^{o}C for 45 minutes at 300 rpm. After 45 minutes of incubation it was centrifuged for 2 minutes at 5000 rpm. Seven hundred microliter of supernatant was discarded and the pellet was mixed in the remaining media using pipette. Then that was pipetted out and poured into LB-Kanamycin plates, using glass beads it was spread all over the media and was let dry for few minute. The plates were incubated at 37^{o}C overnight.

3.2.5 Identification of recombinant plasmids by colony PCR

After the growth of cells on the media colony PCR was performed. For this, reaction mixture for colony PCR was prepared as shown in Table 3.3 below:

Table 3.3 Composition of PCR reaction mix using Emerald master mix.

REAGENTS	VOLUME
M13 forward primer	1 µL
Forward oligo of the target site	1 µL
2X emerald master mixture	10 µL
MilliQ	8 µL
Total	20 µL

After preparation of the PCR master mixture, various transformed colonies were picked; sub cultured in LB kanamycin plate and was then added to the PCR mixture.

Table 3.4 PCR condition.

Initial Denaturation	Denaturation	Annealing	Extension	Final Extension	Hold
98 °C, 3 min	98 °C, 30 sec	62 °C, 10 sec	72 °C, 50 sec	72 °C, 10 min	4 °C
	35 cycles				

The PCR products were loaded on a 1.2% agarose gel.

3.2.6 Isolation of recombinant plasmids

The colonies that gave the positive colony PCR were again inoculated into LB media for plasmid isolation. Plasmids were isolated using MACHEREY NAGEL kit.

For this 1.5 mL of a saturated *E. coli* culture grown in LB was loaded in eppendorf tube and centrifuged for 30s at 11,000 x g. The supernatant was discarded. Then 250 µL of buffer A1 was added and pellet was dissolved completely by vortexing vigorously. After that 250 µL of buffer A2 was added and mixed by inverting the tube 6-8 times and was incubated at room temperature for 5 min. Three hundred µL of buffer A3 was added and mixed by inverting 6-8 times. Then it was centrifuged at 11000 g for 5 min. The clear lysate was then loaded in column and kept in collection tube and it was centrifuged at 11000g for 1 min. The flow through was discarded and 500 µL of buffer AW was added and centrifuged at 11000g for 1min. The flow through was discarded and 600 µL of buffer A4 supplemented with ethanol was added and centrifuged at 11000g for 1 min. Again the flow through was discarded and it was centrifuged at 11000 g to remove residual wash buffer. Then, the column was placed in 1.5ml of microcentrifuge tube and 20 µl of elution buffer was added, it was incubated at room temperature for 1 min and centrifuged at 11000g for 1 min. The column was discarded and eluted sample was kept at -20^0C.

3.2.7 Sequencing the samples and screening positive clones

Those purified plasmid samples were sent for sequencing. After sequencing the plasmid sequences were checked for positive clones. SnapGene Viewer software was used for sequence analysis of the plasmid for identification of plasmid with the insert.

3.3 RNA synthesis

Cas9 nuclease mRNA and sgRNA were prepared using m message m machine SP6 ultra kit and maxiscript T7 kit respectively. The cloned vector with oligonucleotides i.e., pDR274 was digested with HindIII and the pCS2 nCas9n vector was digested with NotI and these linearized plasmids were used for RNA synthesis. The digestion reaction mixture prepared as shown in Table 3.5 and were incubated at 37^0C for 4-5 h.

Table 3.5 Restriction digestion setup for vectors.

For cas9		For pDR274	
COMPONENT	VOLUME	COMPONENT	VOLUME
pCS2 nCas9n	10µL	pDR274	10 µL
NE Buffer 3.1	2 µL	NEBuffer 2.1	2 µL
Not I	2 µL	Hind III	2 µL
MQ	6 µL	MQ	6 µL
Total	20 µL	Total	20 µL

3.3.1 sgRNA synthesis

For sgRNA synthesis maxiscript T7 kit was used. At first 10X transcription buffer and ribonucleotide solutions were vortexed, microcentrifuged briefly and after that the ribonucleotide solution was kept in ice whereas the transcription buffer was kept at room temperature. After the solutions were thawed, 10X of transcription buffer was added on the tube after adding DEPC treated water and template. The reaction was assembled at room temperature because the spermidine can co-precipitate the template DNA if assembled on ice. The reaction was prepared as shown in Table 3.6 below:

Table 3.6 Composition for IVT of sgRNA.

COMPONENT	VOLUME
DEPC treated water	7 µL
Template DNA	5 µL
10X transcription buffer	2 µL
10 mM ATP	1 µL
10 mM CTP	1 µL
10 mM GTP	1 µL
10 mM UTP	1 µL
T7 enzyme mix	2 µL
Total	20 µL

After preparing the reaction mixture the tube was gently flicked and then was microcentrifuged to bring the solution to the bottom. Then, it was incubated at 37°C for 1 h. After incubation 1µL of turbo DNAase was added and mixed by gentle flicking the tube and the tube was incubated at 37°C for 15 min. Thirty µL of DEPC treated water was added to make the volume up to 50 µL and 5 µL of ammonium acetate was added and was vortexed to mix. Then 165 µL of 100% alcohol was added and kept at -20°C for 30 minutes. After the 30 minute of incubation in ice it was centrifuged for 30 min at maximum speed in cold centrifuge. The supernatant was carefully discarded and 300 µL of 70% alcohol was added. Then, it was again centrifuged at maximum speed for 10 min. After centrifugation the supernatant was carefully discarded and the tube was kept in upside down position to dry. After the tubes were dry, the pellet in each tube were dissolved in 20 µL of DEPC-treated water and stored at -80^{o}C.

3.3.2 Cas9 synthesis

For Cas9 synthesis m message m machine SP6 ultra kit was used. At first 10Xtranscription buffer and SP6 2X NTP/ARCA solutions were vortexed, microcentrifuged briefly and after that the ribonucleotide solution was kept on ice whereas the transcription buffer was kept at room temperature. After the solutions were thawed 10X of transcription buffer was added to the tube after adding DEPC-water and template. The reaction was assembled at room temperature as spermidine co-precipitates the template DNA if assembled on ice. The reaction was prepared as shown in Table 3.7 below:

Table 3.7 Composition for IVT of Cas9 mRNA.

COMPONENT	VOLUME
2X SP6 NTP/ARCA	10 µL
10X Reaction buffer	2 µL
DEPC treated water	1 µL
Template DNA	5 µL
SP6 enzyme mix	2 µL
Total	20 µL

After the preparation of the reaction mixture the tube was gently flicked and then microcentrifuged to bring the solution to the bottom. Then it was incubated at 37°C for 2 h.

After incubation 1µL of turbo DNAase was added and mixed by gentle flicking the tube and incubated at 37°C for 15 min.

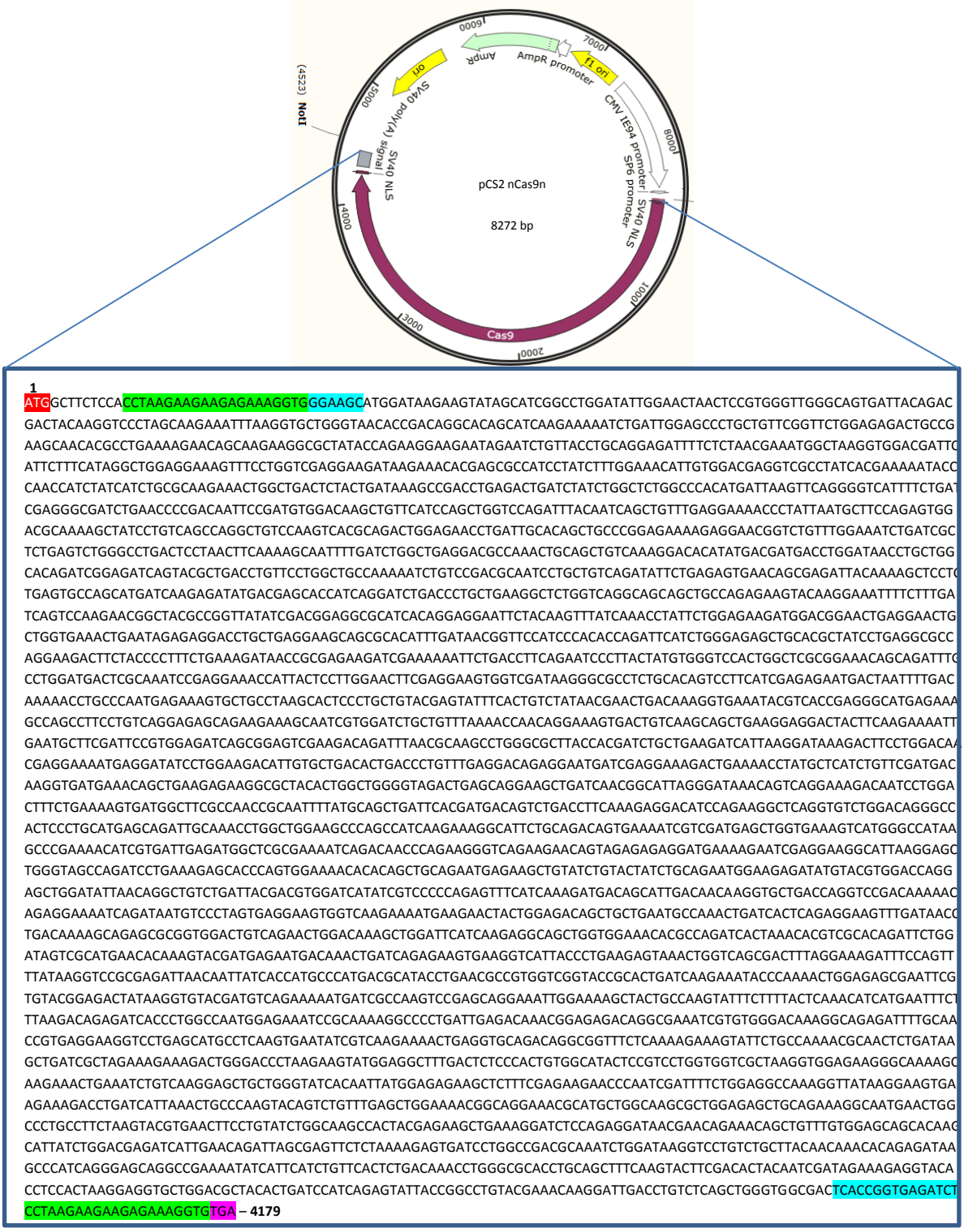
After DNAase treatment tailing reagents were added in following order

Table 3.8 Composition for tailing the RNA.

COMPONENT	VOLUME
DEPC treated water	36 µL
5X E-PAP buffer	20 µL
25 mM $MnCl_2$	10 µL
ATP solution	10 µL

After tailing reagents were added, 2.5 µL of the sample was preserved from the reaction mixture and 4 µL of E-PAP enzyme was added, and was incubated at 37°C for 45 min. Following incubation, 50 µL of LiCL was added, mixed thoroughly and stored at -30°C for 30 min. After incubation at -30°C the tube was centrifuged for 30 min at maximum speed in 4°C centrifuge. Then the supernatant was carefully discarded and 300 µL of 70% alcohol was added. After addition of alcohol it was again centrifuged at maximum speed for 10 min. Then, the supernatant was carefully discarded and the tube was kept in upside down position to dry the excess alcohol. After few minutes the tube was dry and the pellet was dissolved in 20 µL of DEPC-water and stored at -80°C.

Characteristics: Vector size: 8772 bp. SP6 promoter: promoter for bacteriophage SP6 RNA polymerase, SV 40 NLS, Cas9: Cas9 (Csn1) endonuclease from the *Streptococcus pyogenes* Type II CRISPR/Cas system, SV40 poly(A) signal, high-copy-number ColE1/pMB1/pBR322/pUC origin of replication, Ampicillin resistance gene (β lactamase), AmpR promoter, f1 origin of replication, CMV IE94 promoter.



1
ATGGCTTCTCCACCTAAGAAGAAGAGAAAGGTGGGAAGCATGGATAAGAAGTATAGCATCGGCCTGGATATTGGAACTAACTCCGTGGGTTGGGCAGTGATTACAGAC
GACTACAAGGTCCCTAGCAAGAAATTTAAGGTGCTGGGTAACACCGACAGGCACAGCATCAAGAAAAATCTGATTGGAGCCCTGCTGTTCGGTTCTGGAGAGACTGCCG
AAGCAACACGCCTGAAAAGAACAGCAAGAAGGCGCTATACCAGAAGGAAGAATAGAATCTGTTACCTGCAGGAGATTTTCTCTAACGAAATGGCTAAGGTGGACGATTC
ATTCTTTCATAGGCTGGAGGAAAGTTTCCTGGTCGAGGAAGATAAGAAACACGAGCGCCATCCTATCTTTGGAAACATTGTGGACGAGGTCGCCTATCACGAAAAATACC
CAACCATCTATCATCTGCGCAAGAAACTGGCTGACTCTACTGATAAAGCCGACCTGAGACTGATCTATCTGGCTCTGGCCCACATGATTAAGTTCAGGGGTCATTTTCTGAT
CGAGGGCGATCTGAACCCCGACAATTCCGATGTGGACAAGCTGTTCATCCAGCTGGTCCAGATTTACAATCAGCTGTTTGAGGAAAACCCTATTAATGCTTCCAGAGTGG
ACGCAAAAGCTATCCTGTCAGCCAGGCTGTCCAAGTCACGCAGACTGGAGAACCTGATTGCACAGCTGCCCGGAGAAAAGAGGAACGGTCTGTTTGGAAATCTGATCGC
TCTGAGTCTGGGCCTGACTCCTAACTTCAAAAGCAATTTTGATCTGGCTGAGGACGCCAAACTGCAGCTGTCAAAGGACACATATGACGATGACCTGGATAACCTGCTGG
CACAGATCGGAGATCAGTACGCTGACCTGTTCCTGGCTGCCAAAAATCTGTCCGACGCAATCCTGCTGTCAGATATTCTGAGAGTGAACAGCGAGATTACAAAAGCTCCTC
TGAGTGCCAGCATGATCAAGAGATATGACGAGCACCATCAGGATCTGACCCTGCTGAAGGCTCTGGTCAGGCAGCAGCTGCCAGAGAAGTACAAGGAAATTTTCTTTGA
TCAGTCCAAGAACGGCTACGCCGGTTATATCGACGGAGGCGCATCACAGGAGGAATTCTACAAGTTTATCAAACCTATTCTGGAGAAGATGGACGGAACTGAGGAACTG
CTGGTGAAACTGAATAGAGAGGACCTGCTGAGGAAGCAGCGCACATTTGATAACGGTTCCATCCCACACCAGATTCATCTGGGAGAGCTGCACGCTATCCTGAGGCGCC
AGGAAGACTTCTACCCCTTTCTGAAAGATAACCGCGAGAAGATCGAAAAAATTCTGACCTTCAGAATCCCTTACTATGTGGGTCCACTGGCTCGCGGAAACAGCAGATTTC
CCTGGATGACTCGCAAATCCGAGGAAACCATTACTCCTTGGAACTTCGAGGAAGTGGTCGATAAGGGCGCCTCTGCACAGTCCTTCATCGAGAGAATGACTAATTTTGAC
AAAAACCTGCCCAATGAGAAAGTGCTGCCTAAGCACTCCCTGCTGTACGAGTATTTCACTGTCTATAACGAACTGACAAAGGTGAAATACGTCACCGAGGGCATGAGAAA
GCCAGCCTTCCTGTCAGGAGAGCAGAAGAAAGCAATCGTGGATCTGCTGTTTAAAACCAACAGGAAAGTGACTGTCAAGCAGCTGAAGGAGGACTACTTCAAGAAAATT
GAATGCTTCGATTCCGTGGAGATCAGCGGAGTCGAAGACAGATTTAACGCAAGCCTGGGCGCTTACCACGATCTGCTGAAGATCATTAAGGATAAAGACTTCCTGGACAA
CGAGGAAAATGAGGATATCCTGGAAGACATTGTGCTGACACTGACCCTGTTTGAGGACAGAGGAATGATCGAGGAAAGACTGAAAACCTATGCTCATCTGTTCGATGAC
AAGGTGATGAAACAGCTGAAGAGAAGGCGCTACACTGGCTGGGGTAGACTGAGCAGGAAGCTGATCAACGGCATTAGGGATAAACAGTCAGGAAAGACAATCCTGGA
CTTTCTGAAAAGTGATGGCTTCGCCAACCGCAATTTTATGCAGCTGATTCACGATGACAGTCTGACCTTCAAAGAGGACATCCAGAAGGCTCAGGTGTCTGGACAGGGCC
ACTCCCTGCATGAGCAGATTGCAAACCTGGCTGGAAGCCCAGCCATCAAGAAAGGCATTCTGCAGACAGTGAAAATCGTCGATGAGCTGGTGAAAGTCATGGGCCATAA
GCCCGAAAACATCGTGATTGAGATGGCTCGCGAAAATCAGACAACCCAGAAGGGTCAGAAGAACAGTAGAGAGAGGATGAAAAGAATCGAGGAAGGCATTAAGGAGC
TGGGTAGCCAGATCCTGAAAGAGCACCCAGTGGAAAACACACAGCTGCAGAATGAGAAGCTGTATCTGTACTATCTGCAGAATGGAAGAGATATGTACGTGGACCAGG
AGCTGGATATTAACAGGCTGTCTGATTACGACGTGGATCATATCGTCCCCCAGAGTTTCATCAAAGATGACAGCATTGACAACAAGGTGCTGACCAGGTCCGACAAAAAC
AGAGGAAAATCAGATAATGTCCCTAGTGAGGAAGTGGTCAAGAAAATGAAGAACTACTGGAGACAGCTGCTGAATGCCAAACTGATCACTCAGAGGAAGTTTGATAACC
TGACAAAAGCAGAGCGCGGTGGACTGTCAGAACTGGACAAAGCTGGATTCATCAAGAGGCAGCTGGTGGAAACACGCCAGATCACTAAACACGTCGCACAGATTCTGG
ATAGTCGCATGAACACAAAGTACGATGAGAATGACAAACTGATCAGAGAAGTGAAGGTCATTACCCTGAAGAGTAAACTGGTCAGCGACTTTAGGAAAGATTTCCAGTT
TTATAAGGTCCGCGAGATTAACAATTATCACCATGCCCATGACGCATACCTGAACGCCGTGGTCGGTACCGCACTGATCAAGAAATACCCAAAACTGGAGAGCGAATTCG
TGTACGGAGACTATAAGGTGTACGATGTCAGAAAAATGATCGCCAAGTCCGAGCAGGAAATTGGAAAAGCTACTGCCAAGTATTTCTTTTACTCAAACATCATGAATTTCT
TTAAGACAGAGATCACCCTGGCCAATGGAGAAATCCGCAAAAGGCCCCTGATTGAGACAAACGGAGAGACAGGCGAAATCGTGTGGGACAAAGGCAGAGATTTTGCAA
CCGTGAGGAAGGTCCTGAGCATGCCTCAAGTGAATATCGTCAAGAAAACTGAGGTGCAGACAGGCGGTTTCTCAAAAGAAAGTATTCTGCCAAAACGCAACTCTGATAA
GCTGATCGCTAGAAAGAAAGACTGGGACCCTAAGAAGTATGGAGGCTTTGACTCTCCCACTGTGGCATACTCCGTCCTGGTGGTCGCTAAGGTGGAGAAGGGCAAAAGC
AAGAAACTGAAATCTGTCAAGGAGCTGCTGGGTATCACAATTATGGAGAGAAGCTCTTTCGAGAAGAACCCAATCGATTTTCTGGAGGCCAAAGGTTATAAGGAAGTGA
AGAAAGACCTGATCATTAAACTGCCCAAGTACAGTCTGTTTGAGCTGGAAAACGGCAGGAAACGCATGCTGGCAAGCGCTGGAGAGCTGCAGAAAGGCAATGAACTGG
CCCTGCCTTCTAAGTACGTGAACTTCCTGTATCTGGCAAGCCACTACGAGAAGCTGAAAGGATCTCCAGAGGATAACGAACAGAAACAGCTGTTTGTGGAGCAGCACAAG
CATTATCTGGACGAGATCATTGAACAGATTAGCGAGTTCTCTAAAAGAGTGATCCTGGCCGACGCAAATCTGGATAAGGTCCTGTCTGCTTACAACAAACACAGAGATAA
GCCCATCAGGGAGCAGGCCGAAAATATCATTCATCTGTTCACTCTGACAAACCTGGGCGCACCTGCAGCTTTCAAGTACTTCGACACTACAATCGATAGAAAGAGGTACA
CCTCCACTAAGGAGGTGCTGGACGCTACACTGATCCATCAGAGTATTACCGGCCTGTACGAAACAAGGATTGACCTGTCTCAGCTGGGTGGCGACTCACCGGTGAGATCT
CCTAAGAAGAAGAGAAAGGTGTGA – **4179**

Figure 3.2: Map of pcs2-ncas9n vector where Cas9 gene construct is of 4179 bp size. Red highlighted region shows start codon, green highlighted region shows SV40 Nuclear Localization Signal, plain region shows Cas9 sequence, blue region shows the flanking sequence between Cas9 and NLS sequence and pink region shows the stop codon. (*http://www.snapgene.com/resources/plasmid_files/crispr_plasmids/pCS2-nCas9n/*).

3.4 Microinjection of zebrafish embryos

For the preparation of mold used for aligning the zebrafish embryos, 1.2 g of agarose was weighed, dissolved in 60 ml of double distilled water and was poured in the cover of a petri plate and cast was kept over it and was let to solidify for few minutes.

For injection eggs were collected from fresh breeding tank using a glass pipette and were kept in a bowl. After that the eggs were washed two or more time to remove the debris which otherwise may block the needle during injection. Then, the eggs were aligned in the well as shown in Fig. 3.3. with the help of a dropper and brush. Excess water was removed from mold but was kept sufficient that won't let the eggs to dry. Injection mixture was prepared by mixing 2 µL of sgRNA for 1^{st} and 2^{nd} target site of exon1, 2.1 µL of sgRNA for 3^{rd} target site of exon 1, 1.25 µL of sgRNA for exon 5 and 2 µL of cas9 mRNA were added along with 0.5 µL of phenol red as tracer and pH indicator such that the final concentration of sgRNA and that of Cas9 mRNA in the solution was 17.5 ng/µL and 300 ng/µL respectively. Needle was taken out and was loaded with 5µL of mixture with a micro loader pipette. The needle was jerked to move the bolus to the needle tip until there were no bubbles in the needle. Then, the air source and microinjector (injectMan 4) was turned on and the needle was inserted into the microinjector. After that the needle was brought in the plane of view of microscope, high off the stage and was focused on the thinnest region of the tip. Using a sharp forceps the tip of the needle was pinched off such that the needle tip was narrow enough to pierce the chorion and yolk but still capable of delivering a consistent bead size. Then, the foot pedal was pressed to monitor the size of bead while breaking the tip of the needle. Depending on the size of the bead the injection pressure was maintained. Then, the needle was lowered toward the column of embryos holding the petridish with opposite hand. The surface of the chorion was pierced and was entered to the yolk in one smooth stroke. The injection mixture was injected into the yolk very carefully to avoid the injection of bubbles and stretching of the yolk which can be lethal for the embryos. While working down the column the pressure was maintained as necessary to maintain the consistent bead size. After injection the embryos were collected in a petridish using a slow stream of water. Those injected embryos were incubated at 28^{o} C in incubator for five days after that they were kept in room having proper condition for their growth. SgRNA and Cas9 mRNA were co-injected into one cell stage to ensure a high efficient delivery of the injected RNA to the embryo as well as to reduce mosaicism. For control, only Cas9 nuclease mRNA was injected into the embryos.

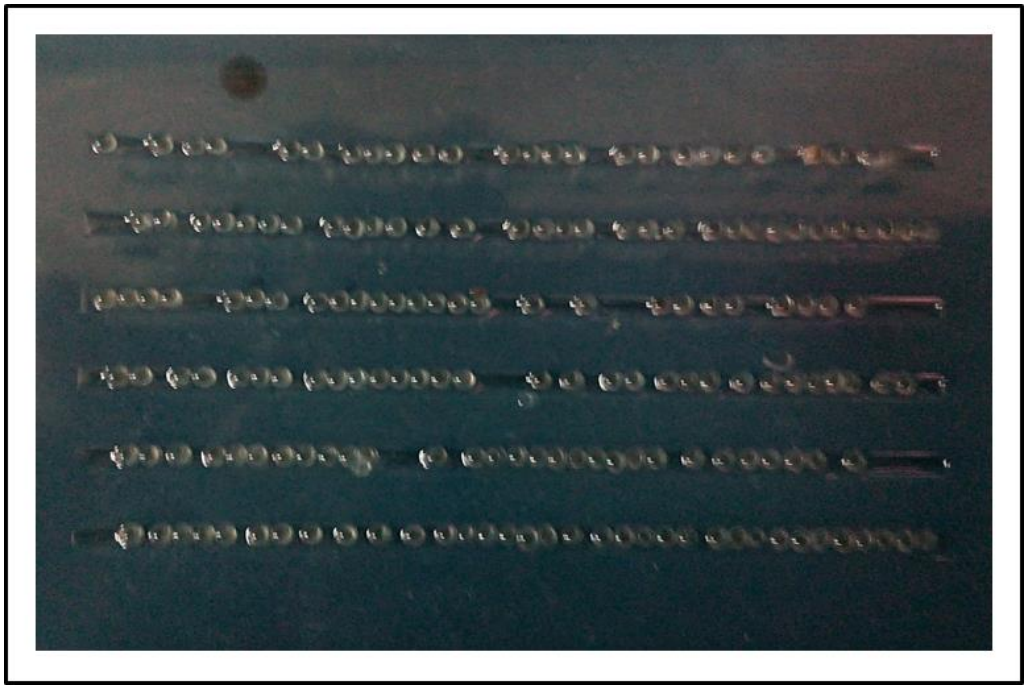


Figure 3.3 A: A close-up of zebrafish embryos aligned in six rows of the mold.

Figure 3.3 B: Two percent agarose mold in which the zebrafish embryos are aligned in the well prior to injection.

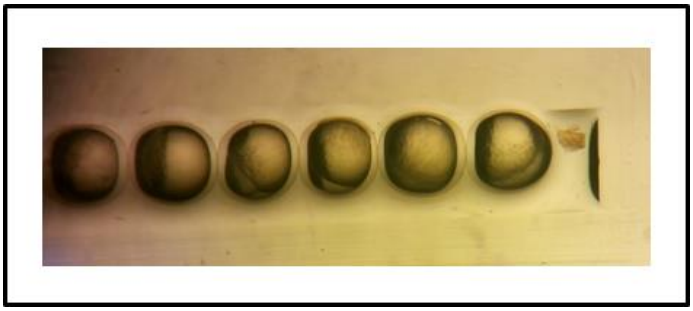


Figure 3.4: Single-cell zebrafish embryos in the mold when seen under microscope prior to microinjection.

3.5 Genomic DNA extraction

Genomic DNA from both injected and mocked 7dpf old larvae were isolated using MACHEREY NAGEL gDNA isolation kit. Ten samples were used for genomic DNA isolation in each case.

Single larvae was taken in an Eppendorf tube and 180 µL of lysis buffer T1 from MACHEREY NAGEL gDNA isolation kit and 25 µL of proteinase K was added from 20mg/ml of stock solution. It was incubated for 3 h at 56^{o}C in water bath. After lysis 200 µL of buffer B3 was added in each tube and vortexed vigorously. Then 1/10 volume of 3M of sodium acetate was added and then 3 volume of 100% alcohol was added and kept at -20^{o}C for 30 min.

After that the tubes were centrifuged at maximum speed in 4^{o}C centrifuge for 30 min. Supernatant was discarded and 300 µL of 70% alcohol was added in each tube and again centrifuged for 10 min at maximum speed. Again, supernatant was discarded and it was allowed to dry and after that 20 µL of MQ was added to dissolve the pellet in each tube. Then, all the samples were stored at -20^{o}C and gel electrophoresis was done.

3.6 PCR amplification of the target sequence

PCR was performed for the target sequence using gDNA isolated from injected and control larvae as template DNA. The primer for PCR was designed using PRIMER 3 input online tool. PCR mixture was prepared as shown in Table 3.9 below.

Table 3.9 Composition of PCR reaction mix using Emerald master mix.

REAGENTS	VOLUME
Forward primer	1 µL
Reverse primer	1 µL
Template DNA	2 µL
2X emerald master mixture	10 µL
MilliQ	6 µL
Total	20 µL

Table 3.10 PCR condition.

Initial Denaturation	**Denaturation**	**Annealing**	**Extension**	**Final Extension**	**Hold**
98 °C, 3 min	98 °C, 30 sec	62 °C, 10 sec	72 °C, 50 sec	72 °C, 10 min	4 °C
	35 cycles				

3.7 Checking the PCR products on an agarose gel and purification

The PCR products were loaded on a gel, electrophoresed and purified using MACHEREY NAGEL gel purification kit.

For this a clean scalpel was taken and the DNA fragments were excised from the gel. Thus, excised gels were kept in eppendorf tubes and 500 µL of NTI buffer was added in each tube. After addition of the buffer the tubes were incubated at 50°C for 10 to 20 min in a thermomixer with gentle mixing. Then, NucleoSpin Gel and PCR clean-up Column were placed into a Collection Tube (2 mL) and 700 µL of sample was loaded into the column. Then, collection tubes along with the column were centrifuged for 30 s at 11,000 X g. Flow-through was discarded and the columns were placed back into the collection tube. Seven hundred µL of BUFFER NT3 was added in each column and was centrifuged for 30 s at 11,000 X g. Flow-through was discarded and the columns were placed back into the collection tube. This step was repeated once again. After that the columns were centrifuged for 1min at 11,000 X g. Flow-through was discarded and the columns were placed into 1.5 ml microcentrifuge tube. Twenty µL elution buffer (NE) was added and incubated at room temperature (18–25°C) for 1 min, then centrifuged for 1 min at the same speed. Thus, purified PCR products were store at -20^{0}C.

3.8 Detection of CRISPR-Cas induced mutations

T7 endonuclease assay was done to check for the presence of mutations in the target exons. T7 endonuclease I recognizes and cleaves non-perfectly matched DNA, cruciform DNA structures, holiday structures or junctions, heteroduplex DNA. Since, both wild type and mutant exon of the gene will be amplified by PCR, the eluted product was denatured and renatured. The denaturation and renaturation condition is shown in the Table 3.11 below:

Table 3.11 Condition for denaturation and renaturation of PCR product done before T7 endonuclease assay.

TEMPERATURE	TIME	CYCLE
37	1 min	
50	1 min	
60	1 min	
72	2 min	1
85	2 min	
94	5 min	
90	2 min	
85	2 min	1
80	2 min	
75	2 min	
70	2 min	
65	2 min	1
55	2 min	
45	2 min	
35	2min	1
25	1 min	
15	1 min	
10	1 min	
4	1 min	

After denaturation and renaturation of the PCR product, T7 endonuclease assay was done, for that reaction mixture was prepared as shown in Table 3.12 and then the reaction mixture was incubated at 37^{o}C for 1 hour.

Table 3.12 T7 endonuclease assay setup.

REAGENTS	VOLUME
PCR product	5 µL
10X NEB buffer 2	1 µL
T7 endonuclease	1 µL
MilliQ	3 µL

Since, the T7 endonuclease is known to give false positive result; the amplicons were normally sequenced for confirmation.

3.9 Phenotypic observations

Fluorescence microscope (leica MZ10F) was used for phenotypic observations. Single larvae was taken at a time in a petriplate containing methylene blue water mixed with few drop of 20x tricaine (used as anaesthesia). Larvae were aligned in horizontal axis and images were taken.

3.10 Zebrafish culture and breeding

Adult zebrafish (*Danio rerio*) were grown in the fish facility with a 14 hours light/10 hours dark cycle. Zebrafish embryos used for injections were obtained from natural spawning.

Adult male and female were kept separate in a dark room at 4 'O' clock in the afternoon in different tanks. To increase total egg production, females to males were kept in a ratio of 2:1. Next day at 10 am the light was turn on and male and female fishes were kept in same tank where a net was there in middle of tank so that the fish can't get into the bottom and the egg can be collected easily. Undisturbed mating was allowed for approximately 20 minutes.

3.11 Construction of phylogenetic tree

Phylogenetic tree was constructed by using MEGA 7 software to show the evolutionary relationship between the *yfiH* gene and *LACC1* gene in different organism. Sequences were retrieved from National Centre For Biotechnology Information (NCBI) and phylogenetic tree was constructed using MEGA 7 software.

CHAPTER 4

RESULTS

4.1 Finding the target exon sequences

To knock out the *LACC1* gene; exon 1, exon 2 and exon 5 were selected as the target exons and mutations were created at these sites by using CRISPR Cas9 method of genome editing. Target exons sequences were retrieved from ensemble. The zebrafish *LACC1* gene diagrammatic representation along with UTR, exons and intron is shown below in Fig 4.1.

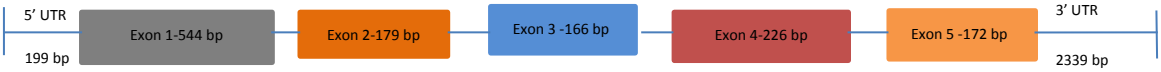


Figure 4.1: Diagrammatic representation of *LACC1* gene in Zebrafish showing 5' UTR, exons, introns and 3' UTR region.

The gene sequences of target exons are shown below, underline region shows the target sites.

ENSDARE00000811407: 18,565,059 -18,565,801 Danio rerio *LACC1* exon 1

CAGACAGGATGCTGGTCCGGTGAGGAGTTTCCACTGGGGGCGGGGCTTACACTCTTTTCAGGAAGTGTT
GGGACAGAGGAAATTTTATTTACGTGGTAAATAAAAAAGTTCCTCGTCTGCAACTCTTATAAACCAGGTGT
*GGTATGTTTGCACGTCCACATACTACATAACCTCCCAGAGTCTGCTTGTGACGGCAGAC***ATG**GCTGCAGTT
ACTCTTGTGGACCTCACAAACATCTGTTCTTCAGAAAGCGACATTTGCGTGAAAAACCGTATCCGCACATT
CTGTGCAGTGACCGAGAATGGCTCACAGGAGCCG**GTGTACTTCGTGGCGGATCC**CG**GGAGCTCCTTTGG**
GACGGACAGATTTCCGAGAAGCTCACTGCTTGAGATGTTTGTCGGTTTCAGCAATTCCCCACAGATCCTG
TGTAAAAGCTCTCTTTCGGCTTCGTTATATTCGTTCAAACAGGCCATAGACGAAAGCGATATTAGTAACGT
GAGGATAGTGACCTCGAGCCGCGGTAGAGGGGTGTTTCAGGTCTACCAGGAGCTGCTCTTCACTTCTGCT
TATAGTTTTGAGTACTGCATCATGTTTGACAACCTGTCGTGTGACTCCTGTCCTCCGGGTGACCGAGCAAA
CAGCAGCACTCCTGCAGAAGGTGACGCCAGACT**GGAGCTCAGCGCGTTCCTTC**AGCGCCTGCCAGCGGT
CAAAGGAGATATTACAGTTCTCAGATCAGCTGCCATTTCAG

ENSDARE00000864222: 18,568,335 -18,568,513 Danio rerio *LACC1* exon 2

ACTGCTTTGGCCAT**GGCTTCAGCACCCGTACCGG**AGGGATCTCCTACATCAGCACCATGAGCTCAATGAA
CCTCTTCTGCAACCCAAGGCGCAAAGACCCCAGAGCAGTGGTGGATGAAAATCGCCGGCGGCTCGGCCT
GCAGGCTGGATTTCATTCTCGTCAGTTAAACCTCATAAAG

ENSDARE00001105341: 18,576,672-18,579,182 Danio rerio *LACC1* exon5

AGTTCTGCTGGAAAGAGGTGGGATTAAACCTGAGCACATAGAGAACATCCGAATCCCAAACCAGACGGA
TTCTATACCATGCACATCCTGCAGTCCTGAGCTATTCTTCTCCCATA**GGAGAGACGGACTCAACTTT**GGGA
CACAGATTGGGTTCCTGTGGATCAAGCCC**TGA**

By using ZIFIT sgRNA online designing tools the oligonucleotides for the sgRNA synthesis were designed. The size of exon 1 was 743 bp along with 5' Untranslated region (UTR) whereas that of exon 2 was 179 bp and that of exon 5 was 2511 along with 3' UTR. But the actual size of exon 5 excluding the 3' UTR was only 172 bp long which is shown above and these sequences were only used for designing oligonucleotides. The target sequence and oligos are shown in Table 4.1 below.

Table 4.1: Target sequences of the exons and the oligonucleotides designed using ZIFIT online tool.

Target name	Target sequence (5' to 3')	Oligonucleotide (5' to 3') (Forward and Reverse)
Lacc1[exon 1.1]	GGATCCGCCACGAAGTACAC Reverse strand	TAGGATCCGCCACGAAGTACAC AAACGTGTACTTCGTGGCGGAT
Lacc1[exon 1.2]	GGAGCTCCTTTGGGACGGAC Forward strand	TAGGTCGTCCCAAAGGAGCTCC AAACGGAGCTCCTTTGGGACGA
Lacc1[exon 1.3]	GGAGCTCAGCGCGTTCCTTC Forward strand	TAGGACGCGCTGAGCTCCAGTC AAACGACTGGAGCTCAGCGCGT
Lacc1[exon 2]	GGCTTCAGCACCCGTACCGG Forward strand	TAGGCTTCAGCACCCGTACCGG AAACCCGGTACGGGTGCTGAAG
Lacc1[exon 5]	GGAGAGACGGACTCAACTTT Forward strand	TAGGAGAGACGGACTCAACTTT AAACAAAGTTGAGTCCGTCTCT

The target sequences are 20 bp long having PAM site at i.e., NGG sequence. The underlined sequence in the target site shows the PAM site. The 1st target site of exon 1 was the reverse strand whereas for all others site it was the forward strand. For exon 1, three oligonucleotide pairs where designed whereas only one oligonucleotide pair for exon 2 and exon 5 was designed.

4.2 Construction of sgRNA clones for in vitro synthesis of sgRNA

Thus, designed oligonucleotides were annealed and were cloned in the vector pDR274. The vector pDR274 was used because it has a gRNA scaffold. The vector was digested with BsaI at 37^{o}C for 5 hour. After digestion cloning of the 22 nucleotides of oligos was done. Ligation was done for 6-8 hour in 16^{o}C as described in method. It was possible to clone all the annealed oligonucleotide pairs in pDR274 vector except that of exon 2. After cloning, plasmid were transformed in DH5α competent cells and cultured on LB agar plate with kanamycin. After that colony PCR was performed using the transformed cell as template using emerald PCR master mix, at the same time those colonies used for PCR were sub cultured in kanamycin LB agar plates. The PCR product was run on 1.2%

agarose gel containing 0.5µg/ml of ethidium bromide (Fig 4.2- Fig 4.3). Since, M13 primer binding site and the site where the oligonucleotides were cloned were around 300 bp apart so the PCR products were 300 bp long.

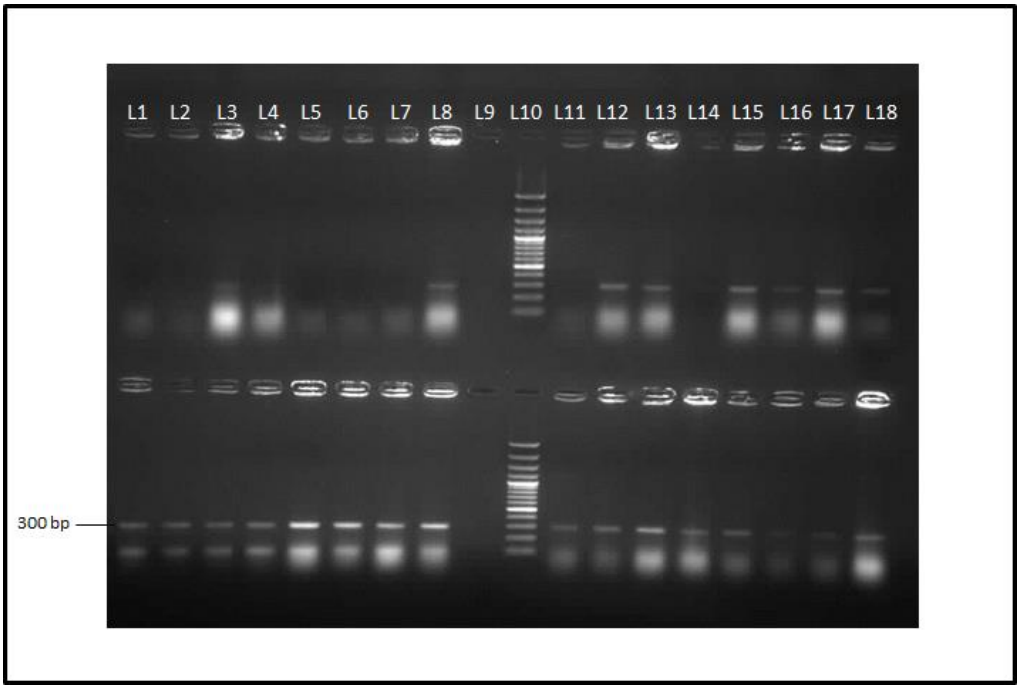


Figure 4.2: Agarose gel (1.2%) electrophoresis of colony PCR of transformants for exon 1 and exon 5 of *LACC1* gene amplified using emerald master mixtures. Up L1 to L9 and L11 to L18 colony PCR of exon 5 oligonucleotides cloned colonies, L10 100 bp ladder. Down L1 to L9 and L11 to L18 colony PCR of exon 1 oligonucleotides cloned colonies, L10 Generuler 100 bp ladder.

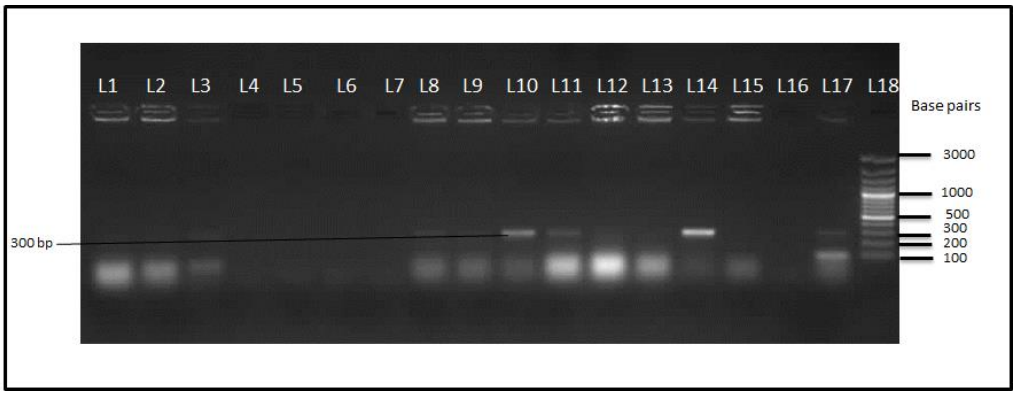


Figure 4.3: Agarose gel (1.2%) electrophoresis of colony PCR of transformants for exon 2 of *LACC1* gene amplified using emerald master mixtures. L1 to L17 colony PCR of exon 2 oligonucleotides cloned colonies, L18 Generuler 100 bp ladder.

Then, the colonies that gave the positive colony PCR were again inoculated in LB broth and plasmids were isolated using Macherey nagel plasmid isolation kit. Those isolated plasmids were run on 0.8% agarose gel (Fig 4.4-4.6).

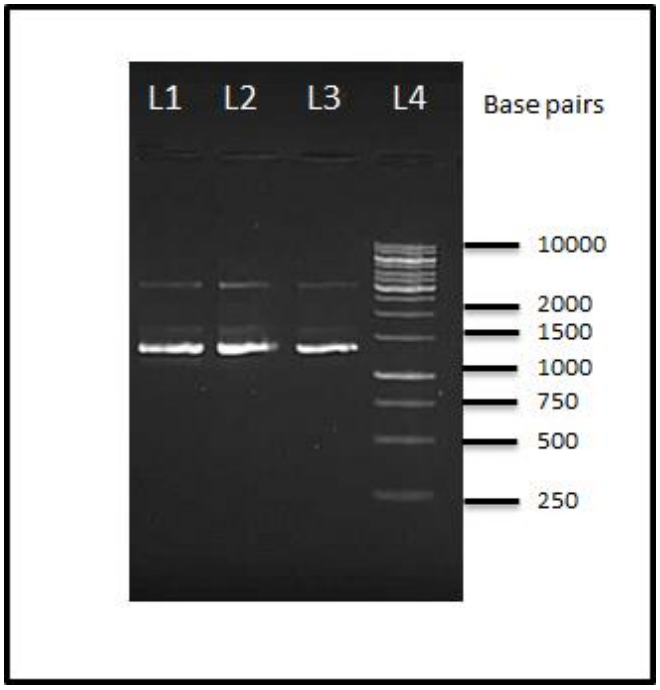


Figure 4.4: Agarose gel (0.8 %) electrophoresis for screening of the isolated plasmids from colonies that have exon 2 oligonucleotides cloned in them and that gave positive colony PCR. L1, L2, L3 plasmids and L4 1 kb Generuler ladder.

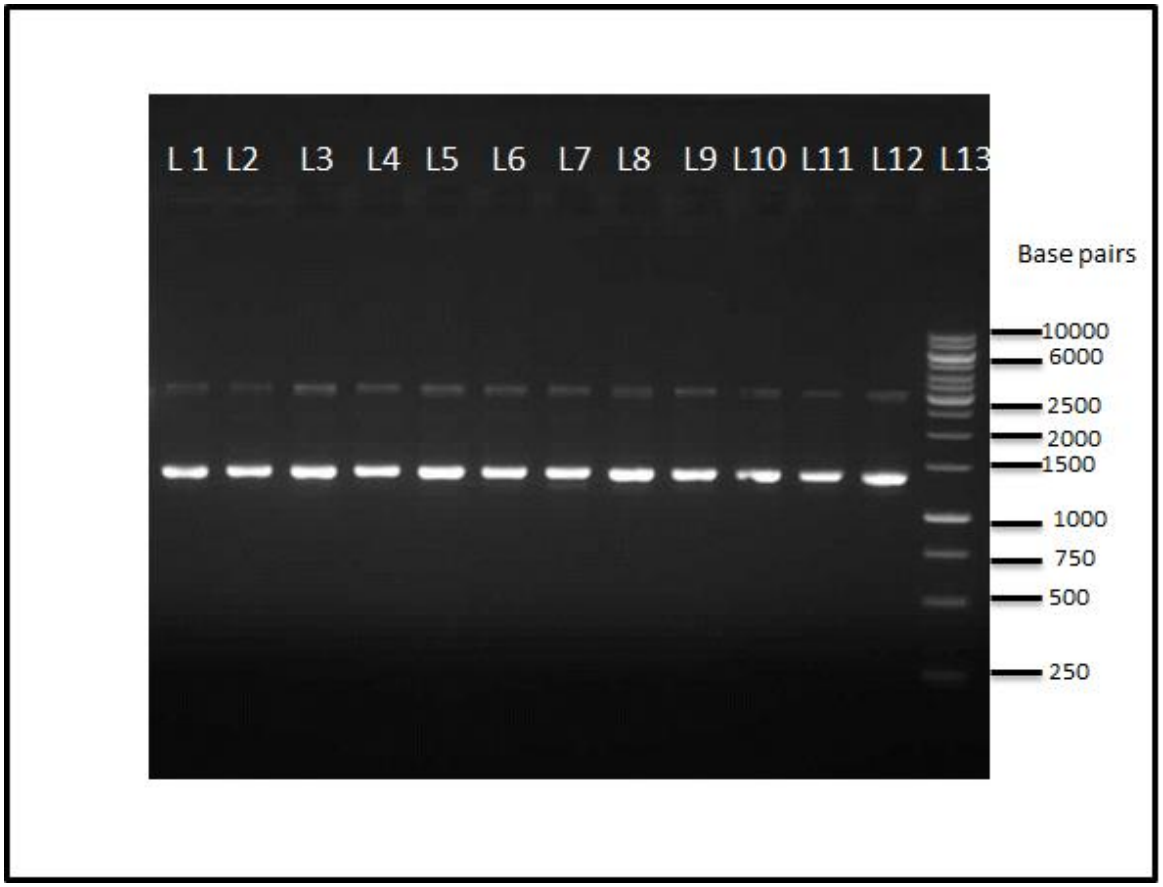


Figure 4.5: Agarose gel (0.8 %) electrophoresis for screening of the isolated plasmids from colonies that have exon 1 oligonucleotides cloned in them and that gave positive colony PCR. L1-L12 plasmids and L13 1 kb Generuler ladder.

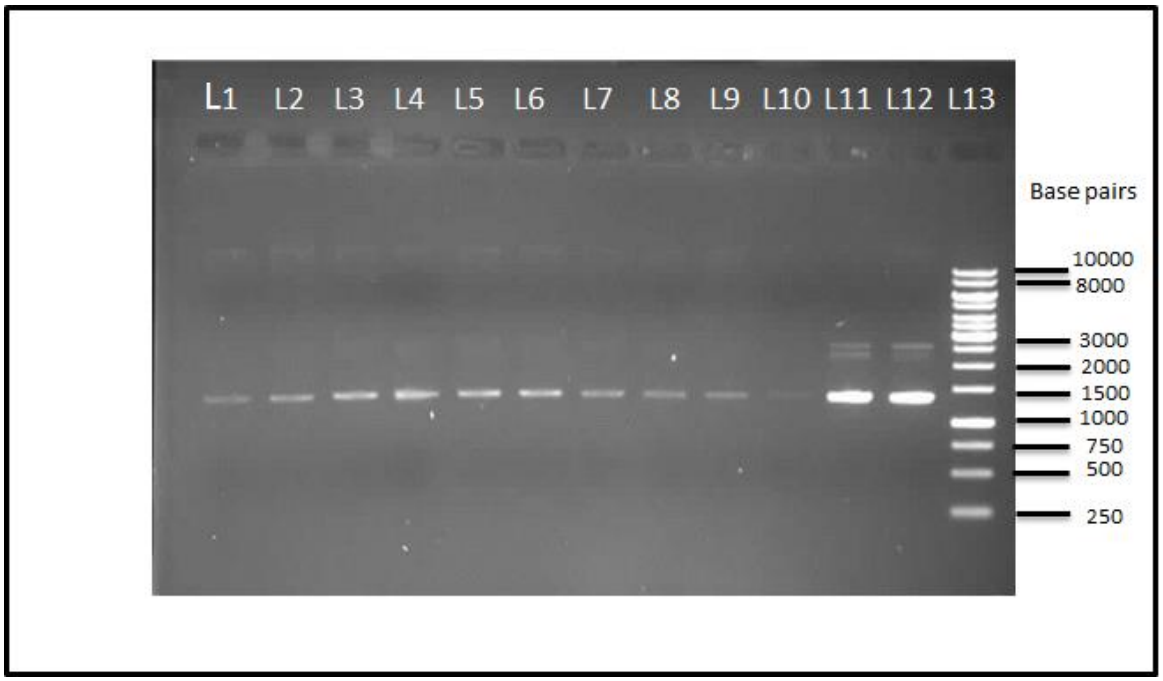


Figure 4.6: Agarose gel (0.8 %) electrophoresis for screening of the isolated plasmids from colonies that have exon 5 oligonucleotides cloned in them and that gave positive colony PCR. L1-L12 plasmids and L13 1 kb Generuler ladder.

After plasmid isolation those plasmid were sent for sequencing to Bioserve, India. Then those sequences were analysed using SnapGeneViewer. The analysis result showed 22 bases of oligonucleotides (highlighted) cloned in the vector pDR274 (Fig 4.7 to Fig 4.10).

Figure 4.7: Chromatogram showing oligonucleotide for target site 1st of exon 1 cloned in vector.

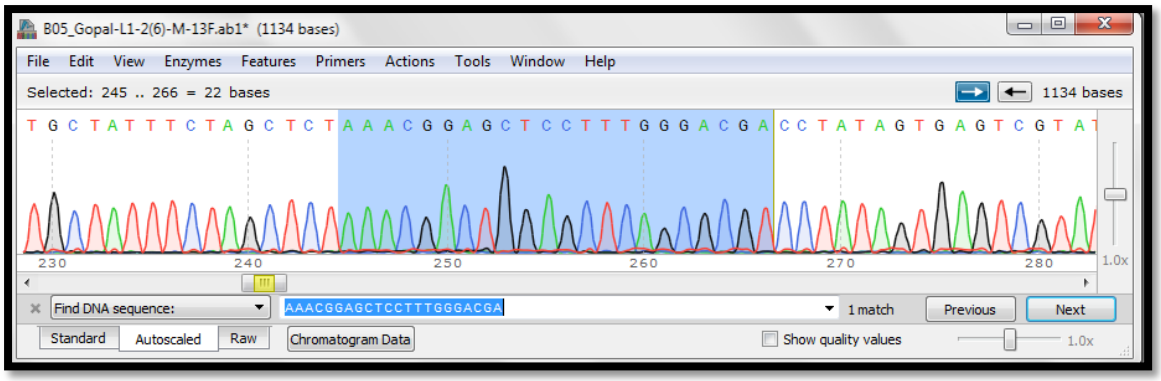


Figure 4.8: Chromatogram showing oligonucleotide for target site 2nd of exon 1 cloned in vector.

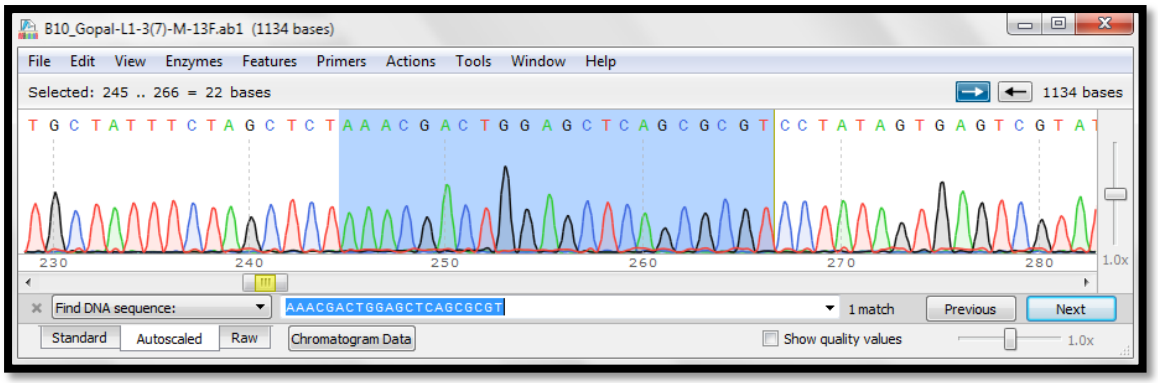


Figure 4.9: Chromatogram showing oligonucleotide for target site 3rd of exon 1 cloned in vector.

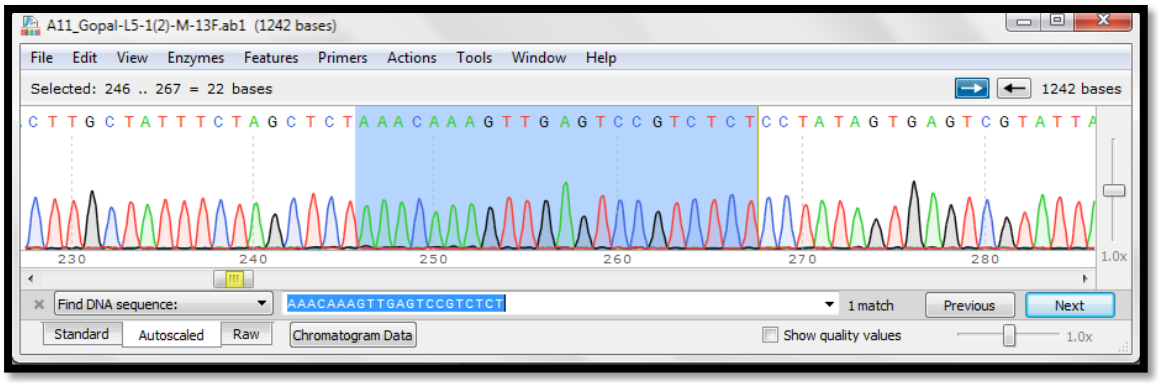


Figure 4.10: Chromatogram showing oligonucleotide for target site of exon 5 cloned in vector.

After confirmation through sequencing, those plasmids having oligonucleotides cloned in them were used for guide RNA synthesis. The recombinant plasmids were named as pDR-Sp E1.1 for the vector pDR274 having gRNA scaffold for *Streptococcus pyogenes* Cas9 and oligonucleotides for target site 1st of exon 1, in the same way other recombinant vector with oligonucleotides were named as pDR-Sp E1.2, pDR-Sp E1.3, pDR-Sp E5 respectively. The recombinant vectors with oligonucleotides cloned in them is shown below (Fig 4.11).

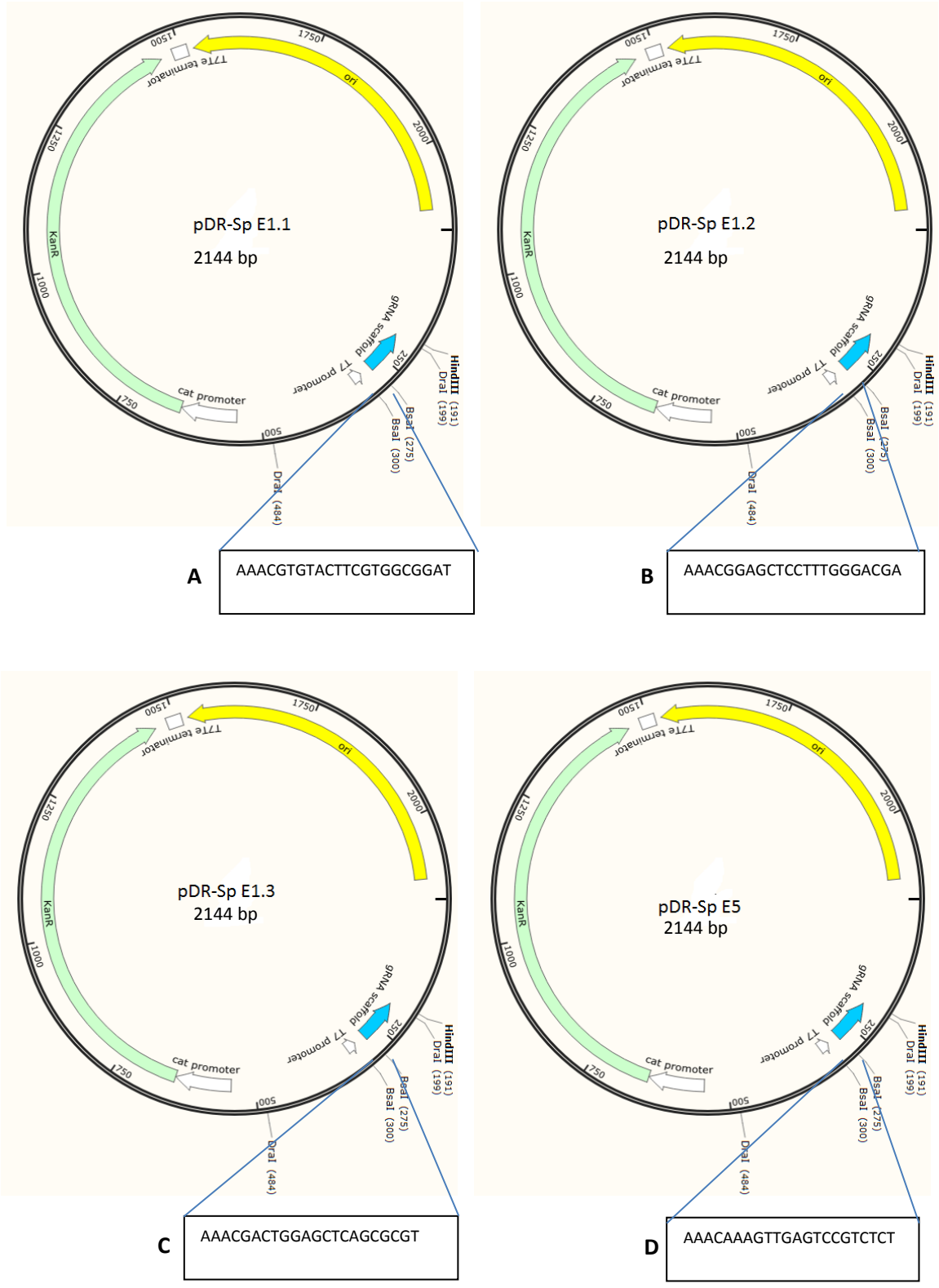


Fig 4.11: A) Recombinant plasmid with the oligonucleotide for 1st target site of exon 1. B) Recombinant plasmid with the oligonucleotide for 2nd target site of exon 1. C) Recombinant plasmid with the oligonucleotide for 3rd target site of exon 1. D) Recombinant plasmid with the oligonucleotide for target site of exon 5. Box shows the sequence of the oligonucleotide clone in the vector.

4.3 In vitro synthesis of sgRNA and Cas9 mRNA

After isolation of plasmid cloned with the oligonucleotides those plasmids were subjected to digestion using HindIII restriction enzyme for 4-5 hour at 37^{o}C. HindIII was used because it has restriction site just upstream of gRNA scaffold. The digested cloned pDR274 vector was run on 1.2% agarose gel (Fig 4.12). Then those plasmids were purified using Macherey nagel gel purification kit. For Cas9 mRNA the vector pCS2 nCas9n was digested with restriction enzyme NotI for 4-5 hour at 37^{o}C. NotI was used because it has only one restriction site just after the polyA signal.

In vitro transcription (IVT) was done to prepare three sgRNA for three target sites of exon 1 and one sgRNA for exon 5, the protocol used was described in Methods (Fig 4.13). IVT of Cas9 was also done after linearization of the plasmid pCs2 nCas9n (Fig 4.13). The concentration of Cas9 mRNA and sgRNA was calculated using nanodrop and is shown below:

Table 4.2: Guide RNA and Cas9 mRNA concentration as measured by nanodrop.

RNA	CONCENTRATION
Cas9 mRNA	1.5µg/µL
sgRNA for exon 1 1st target site	86ng/µL
sgRNA for exon 1 2nd target site	80ng/µL
sgRNA for exon 1 3rd target site	84ng/µL
sgRNA for exon 5 1st target site	140ng/µL

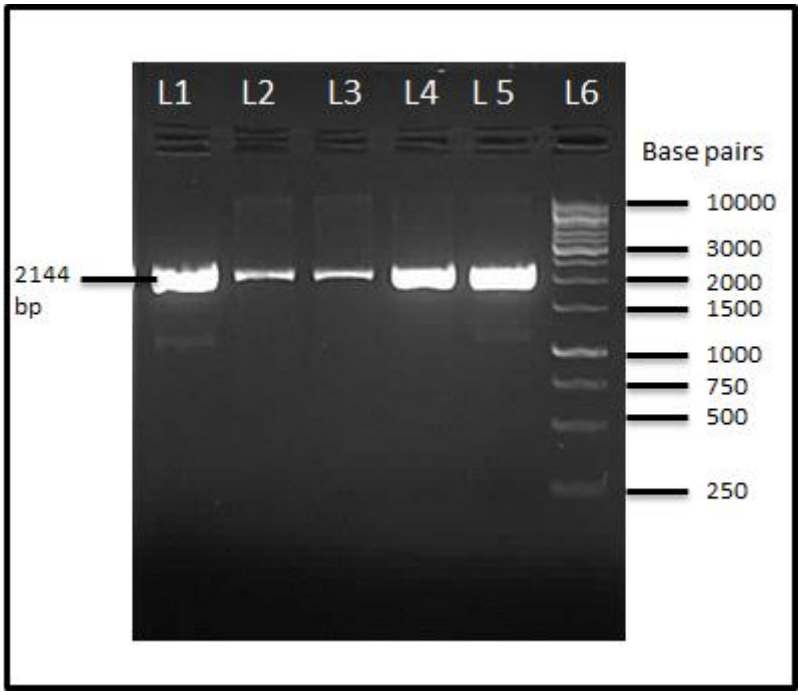
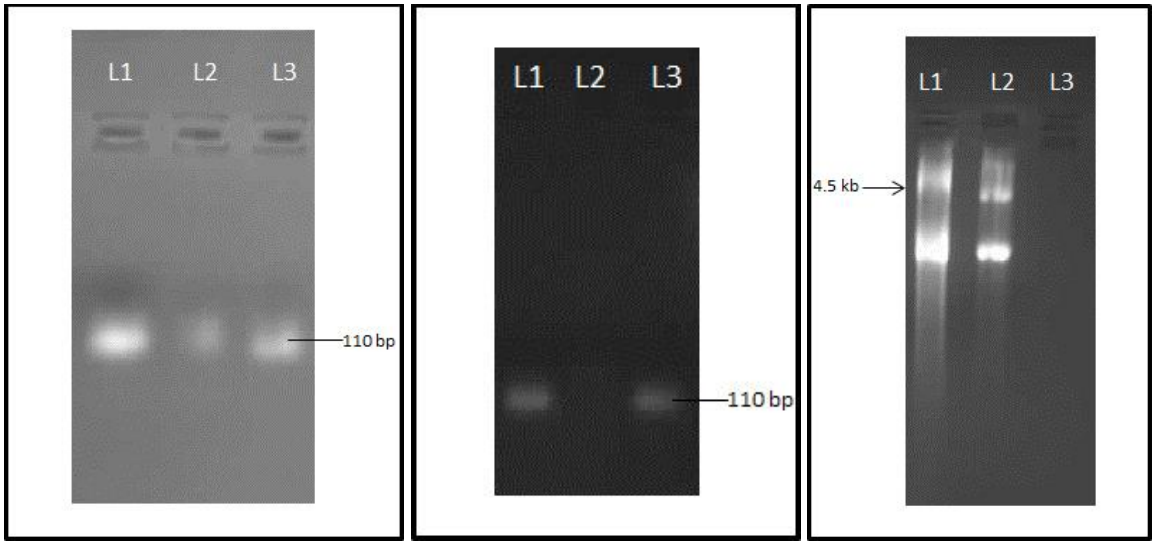


Figure 4.12: Agarose gel (0.8 %) electrophoresis for screening of the isolated plasmids (pDR274) after digestion with HindIII that have oligonucleotides (22bp) for target exons cloned in them. L1 cloned plasmid (pDR-Sp E1.1) with oligonucleotides for 1st target site of exon 1 of *LACC1*. L2 cloned plasmid (pDR-Sp E1.2) with oligonucleotides for 2nd target site of exon 1 of *LACC1*. L3 cloned plasmid (pDR-Sp E1.3) with oligonucleotides for 3rd target site of exon 1 of *LACC1*. L4 and L5 cloned plasmid (pDR-Sp E5) with oligonucleotides for target site of exon 5. L6 Generuler 1 kb ladder.



A **B** **C**

Figure 4.13: 1.5% agarose gel electrophoresis of RNA. a) L1 sgRNA(98bp) for 1st target site of exon 1, L2 sgRNA(98bp) for 2nd target site of exon 1 and L3 sgRNA(98bp) for 3rd target site of exon 1. b) L1 and L3 sgRNA(98bp) for exon 5. c) L1 Cas9 mRNA with polyA tail and L2 Cas9 mRNA without polyA tail.

4.4 Microinjection

After synthesis of sgRNA and Cas9 mRNA, injection mixture was prepared mixing sgRNA, Cas9 mRNA and phenol red as tracer and pH indicator. Injection was performed at one cell stage and approximately 1 to 3nl of the injection mixture was injected in around 200 embryos. For control only Cas9 mRNA was injected in 200 embryos, since Cas9 alone can't bind and create double stranded break in the genome. After microinjection the injected embryos were collected in new sterile petriplates and were incubated at 28^{o}C in incubator for 5 days after that they were kept in room that have optimum temperature for their growth.

Figure 4.14: Micromanipulator used for microinjection.

4.5 Genomic DNA isolation and PCR of the target exons

For analysis of mutation genotyping was performed. For that genomic DNA was isolated from 10 injected and 10 control 7dpf old larvae (Fig 4.15) using Macherey Nagel DNA isolation kit. Those larvae that were showing some deformities were used for gDNA isolation. After genomic DNA isolation, PCR was performed for the target exons.

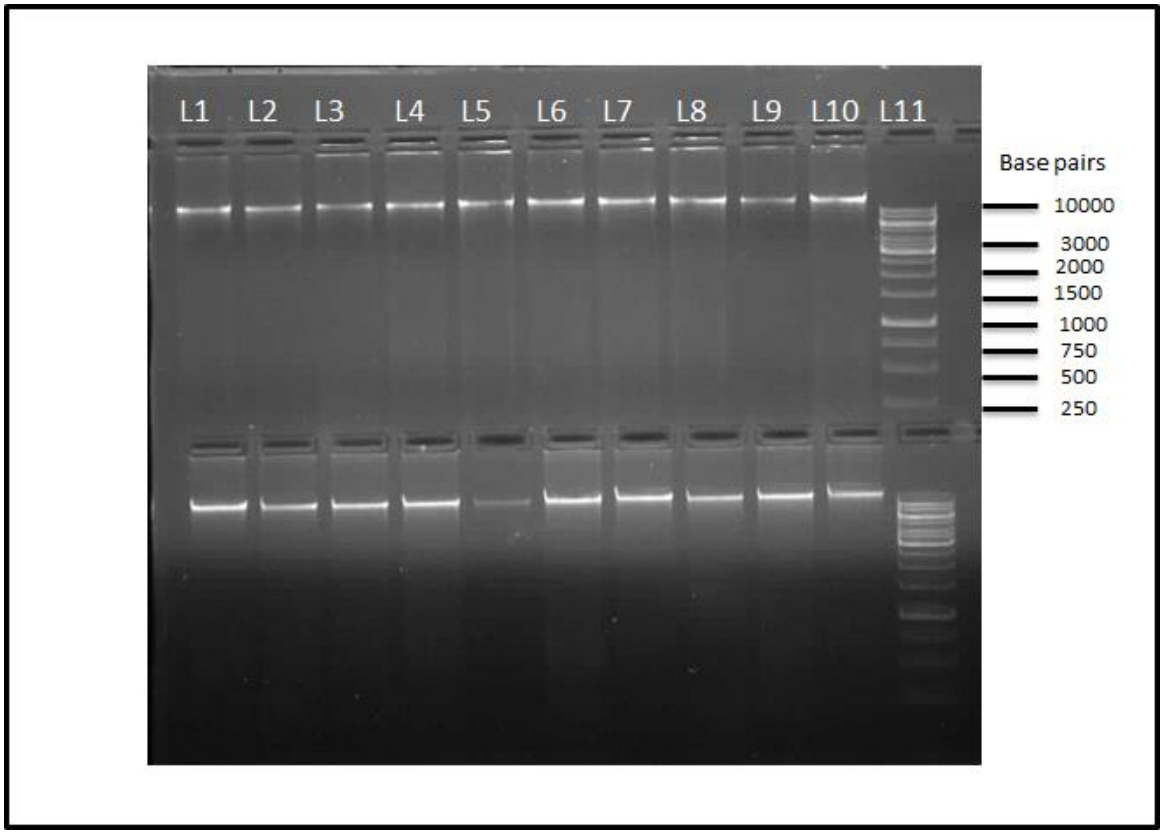


Figure 4.15: Agarose gel (0.8%) electrophoresis of genomic DNA isolated from 7dpf old injected and mocked larvae. Up L1 to L10 genomic DNA of 7dpf old larvae developed from embryo injected with sgRNA and Cas9 mRNA for *LACC1* target sites, L11 Generuler 1kb ladder. Down L1 to L10 genomic DNA of 7 dpf old larvae injected with only Cas9 mRNA, L11 Generuler 1kb ladder.

For PCR, primers were designed using primer 3 input online tools for amplification of the target exon 1, exon 2 and exon 5 of *LACC1* gene. The list of primers for amplification of exon 1, exon 2 and exon 5 of *LACC1* gene are shown in Table 4.3.The target exons i.e. exon 1 and exon 5 were successfully amplified using Emerald 2X master mix (Fig 4.16-Fig 4.19). Since, for exon 1, primer were designed using the of 5' UTR sequences, exon 1 sequences and intron 1 sequences so the expected PCR product was 728 bp long and for designing primer for exon 5, exon 5 sequences and sequence of 3' UTR was also used so the expected PCR product was 179 bp.

Table 4.3: Primers designed for PCR amplification of exon 1, exon 2 and exon 5 of *LACC1* gene.

Exon	Forward primer	Reverse primer
1	CTCCCAGAGTCTGCTTGTGA	AGACAAAGGTTACCGGTTTCT
2	CTGCTGCTGGAGTGTTTCAA	GGCTAGTGGTCTTTTCCTGC
5	TGCTGGAAAGAGGTGGGATT	TTTTGTGCATTCGGTCAGGG

Figure 4.16: Agarose gel electrophoresis (1.2%) of PCR product of exon 5 of *LACC1* gene using gDNA of 7dpf old larvae developed from embryo injected with sgRNA and Cas9 mRNA. L1-L10 PCR product amplified using Emerald master mixture, L11 NEB 100 bp ladder.

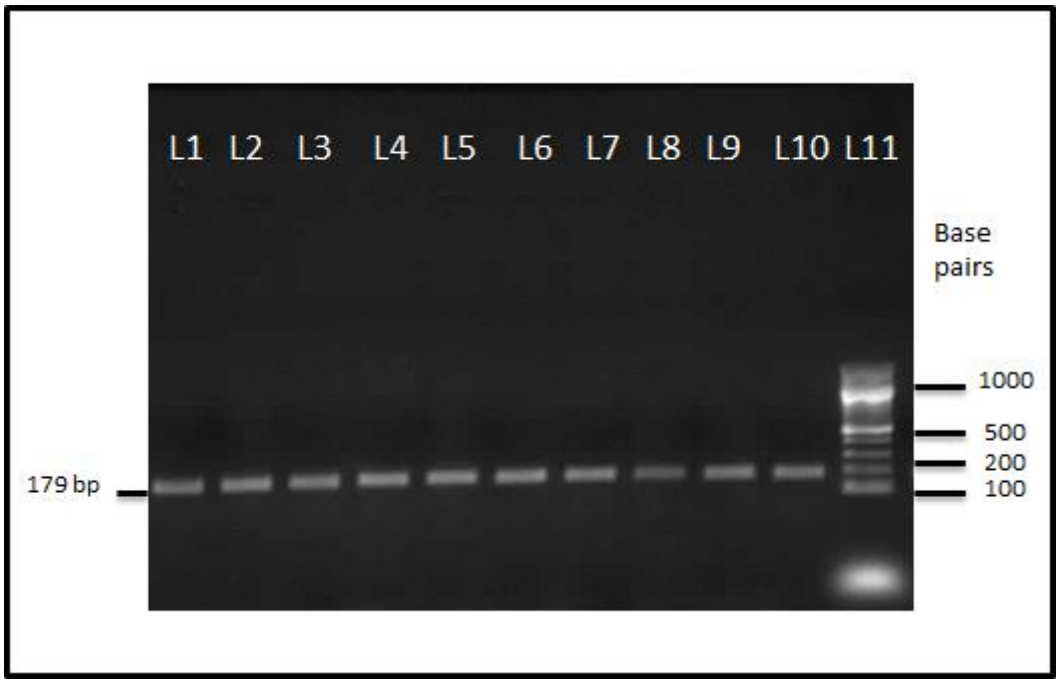


Figure 4.17: Agarose gel electrophoresis (1.2%) of PCR product of exon 5 of *LACC1* gene using gDNA of 7dpf old larvae developed from embryo injected with only Cas9 mRNA. L1-L10 PCR product amplified using Emerald master mixture, L11 NEB 100 bp ladder.

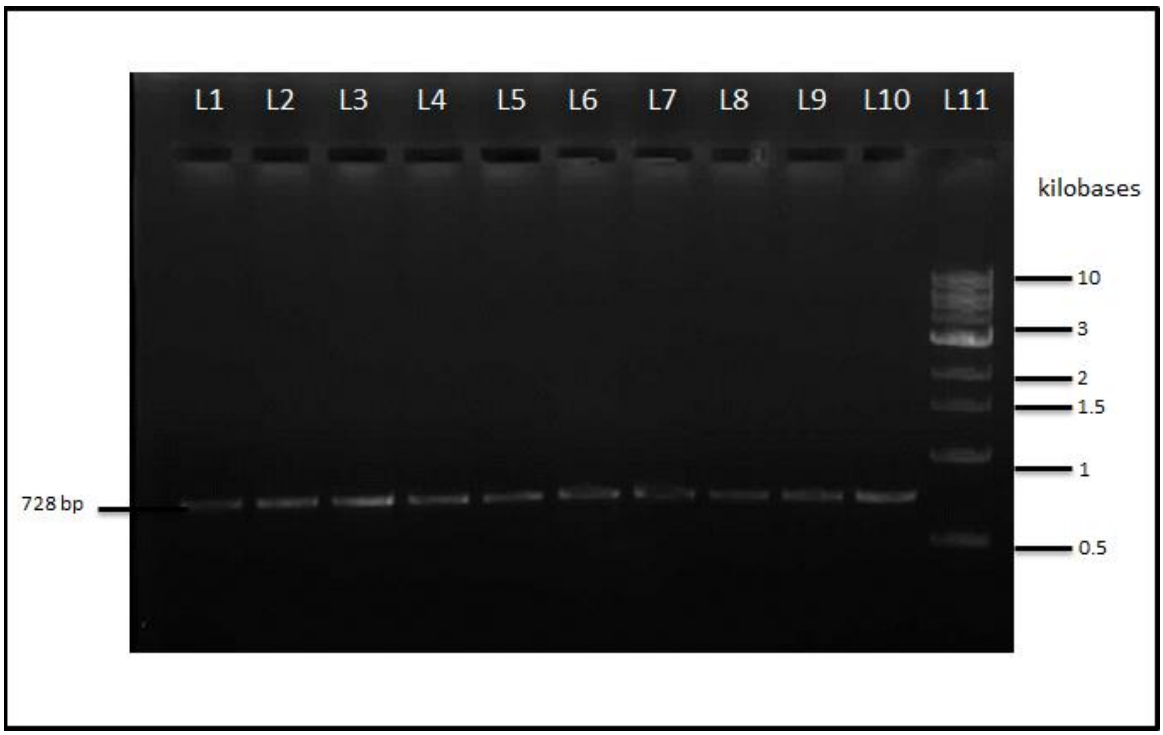


Figure 4.18: Agarose gel electrophoresis (1.2%) of PCR product of exon 1 of *LACC1* gene using gDNA from 7dpf old larvae developed from embryo injected with sgRNA and Cas9 mRNA. L1-L10 PCR product amplified using Emerald master mixture, L11 NEB 1 kb ladder.

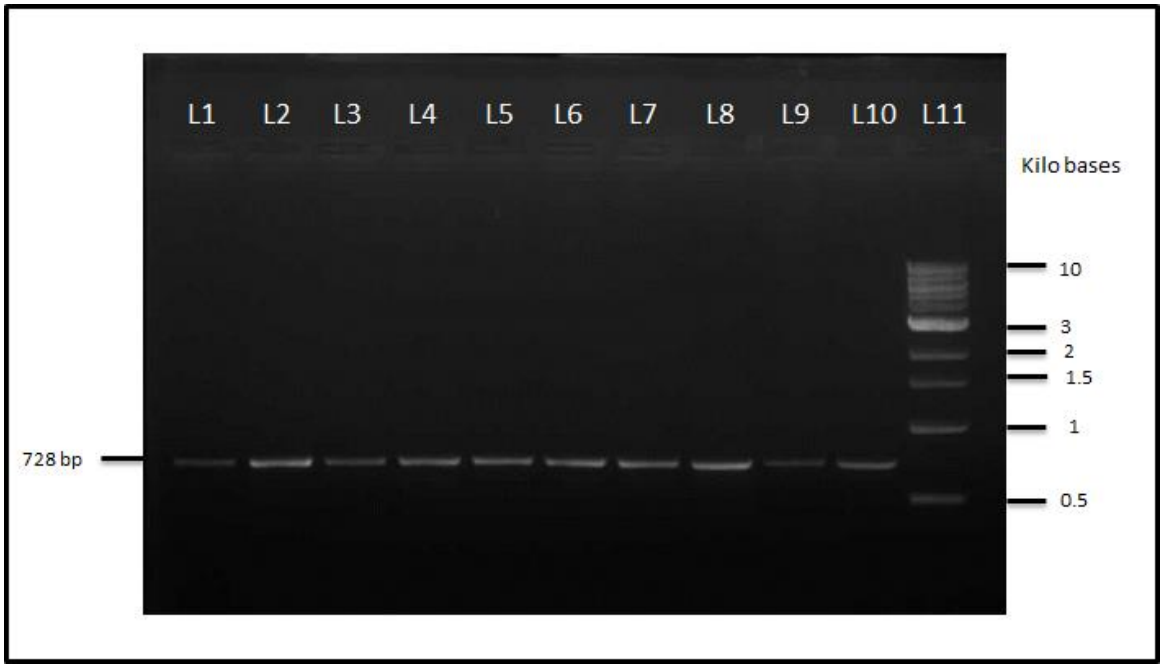


Figure 4.19: Agarose gel electrophoresis (1.2%) of PCR product of exon 1 of *LACC1* gene using gDNA from 7dpf old larvae developed from embryo injected with only Cas9 mRNA. L1-L10 PCR product amplified using Emerald master mixture, L11 NEB 1 kb ladder.

4.6 T7 endonuclease assay

T7 endonuclease assay was done as a preliminary test for detecting mutation in the target exons. For this the PCR products were denatured and renatured. This was done because PCR can amplify both mutant and wild type allele of the gene. After denaturation and renaturation of the PCR product there were three possibilities 1) wild type will renature with wild type 2) wild type will renature with mutant type and 3) mutant type will anneal with mutant type. Since, T7 endonuclease enzyme acts only when there is a mismatch between the sequences so denaturation and renaturation is a crucial step and should be done before performing T7 endonuclease assay.

T7 endonuclease assay for both exon 1 and exon 5 showed positive assays as there were more than one band in the gel thus indicating mutation in the target exons (Fig 4.20- Fig 4.21). Since, T7 endonuclease assay gives false positive result due to single nucleotide polymorphism so it is not a reliable method for determining mutations. Thus, the sample i.e. purified PCR product were send for sequencing to Bioserve, India for further analysis.

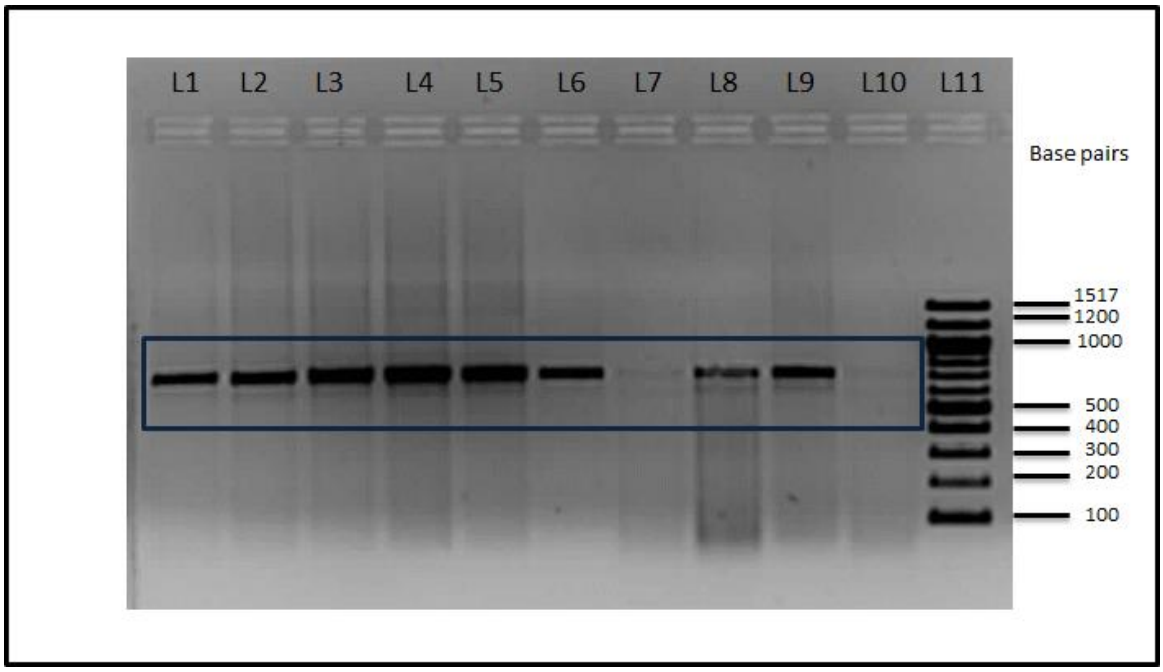


Figure 4.20: Agarose gel(1.2%) electrophoresis of *LACC1* exon 1 PCR product after T7 endonuclease assay. L1 to L10 PCR product of exon 1 after T7 endonuclease assay, L11 NEB 100 bp ladder. Those on the boxes shows positive assay.

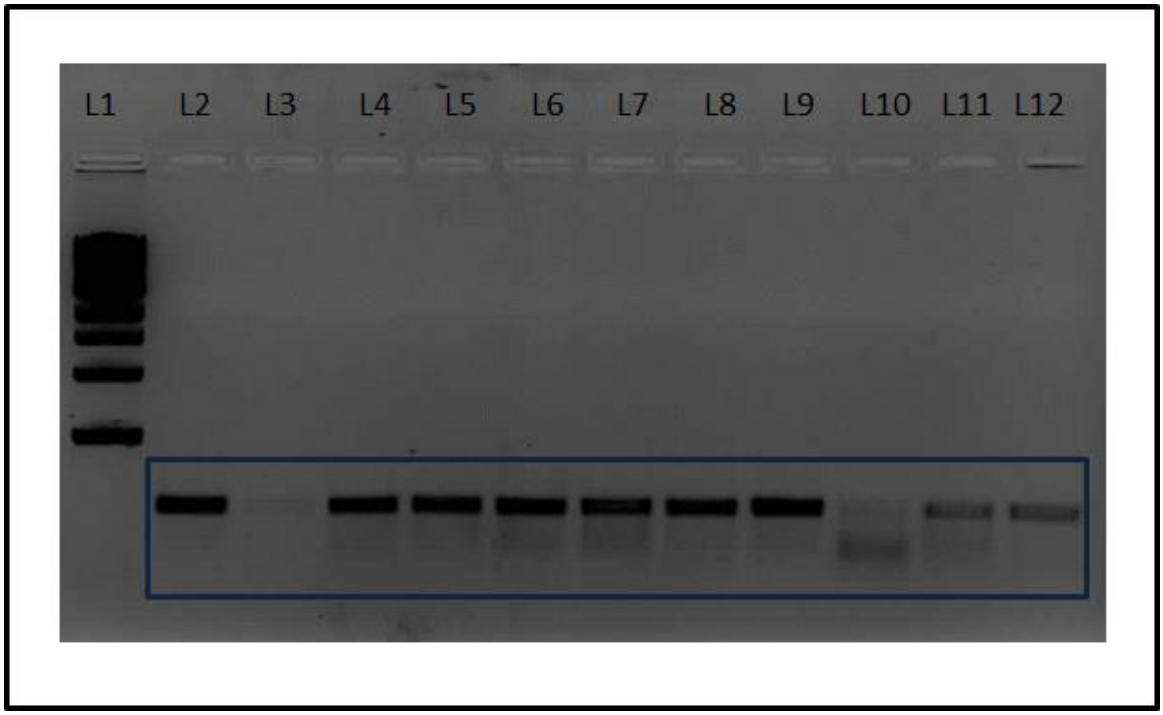


Figure 4.21: Agarose gel(1.2%) electrophoresis of *LACC1* exon 5 PCR product after T7 endonuclease assay. L1 to L10 PCR product of exon 5 after T7 endonuclease assay, L11 NEB 100 bp ladder. Those on the boxes shows positive assay.

4.7 Phenotypic observations

In all of the injected embryos, there were wild type-like larvae and the larvae with deformities. They showed phenotypes such as elongated body length with curved tails and pericardial edema. Images were taken using fluorescence microscope. The larvae were analysed on daily basis after they were four days old. The larvae were kept in a petriplate having methylene blue water and tricaine. Larvae were aligned in the horizontal axis and images were taken. After that the larvae were washed 2 to 3 times with methylene blue water to remove excess of tricaine and were again kept in petriplates with methylene blue water.

There were some deformed larvae in the control as well but they were not showing the phenotype like the injected ones rather they have swollen stomach region (Appendix 11), also the number of deformed larvae were comparatively low with comparison to the injected larvae i.e. 15% in control and 60% in those injected with guide RNA and Cas9 mRNA. All of the larvae showing some deformities were dead within 9dpf, the death rate in injected and control larvae is shown in Table 4.4 below.

Table 4.4: Death rate in 100 larvae larvae developed from embryo injected with sgRNA for *LACC1* and Cas9 mRNA in comparison to the control larvae developed from embryo injected with only Cas9 mRNA (mocked).

DPF	sgRNA for *LACC1* and Cas9 mRNA injected larvae (100)			Mocked larvae (100)		
	Alive	Dead	Death rate (approximately)	Alive	Dead	Death rate (approximately)
1	92	8	8%	95	5	5%
2	79	13	14%	85	10	11%
3	69	10	13%	77	8	9%
4	59	10	15%	75	2	3%
5	52	7	12%	74	1	1%
7	39	13	25%	64	10	14%
8	28	11	28%	57	7	11%
10	17	11	39%	48	9	16%

The death rate in larvae developed from embryo injected with sgRNA for *LACC1* and Cas9 mRNA were relatively more in comparision with larvae developed from embryo injected with only Cas9 mRNA. In the case of larvae developed from embryo injected with sgRNA for *LACC1* and Cas9 mRNA, majority of them were deformed and were dying prematurely due to bacterial and protozoal infection though excessive care was taken. In case of mocked larvae deformity was less and the death rate was also less since the larvae were totally healthy and not developing any infection in majority of the cases. Bar diagram showing percentage death rate in both the cases is shown below (Fig 4.22-Fig4.23).

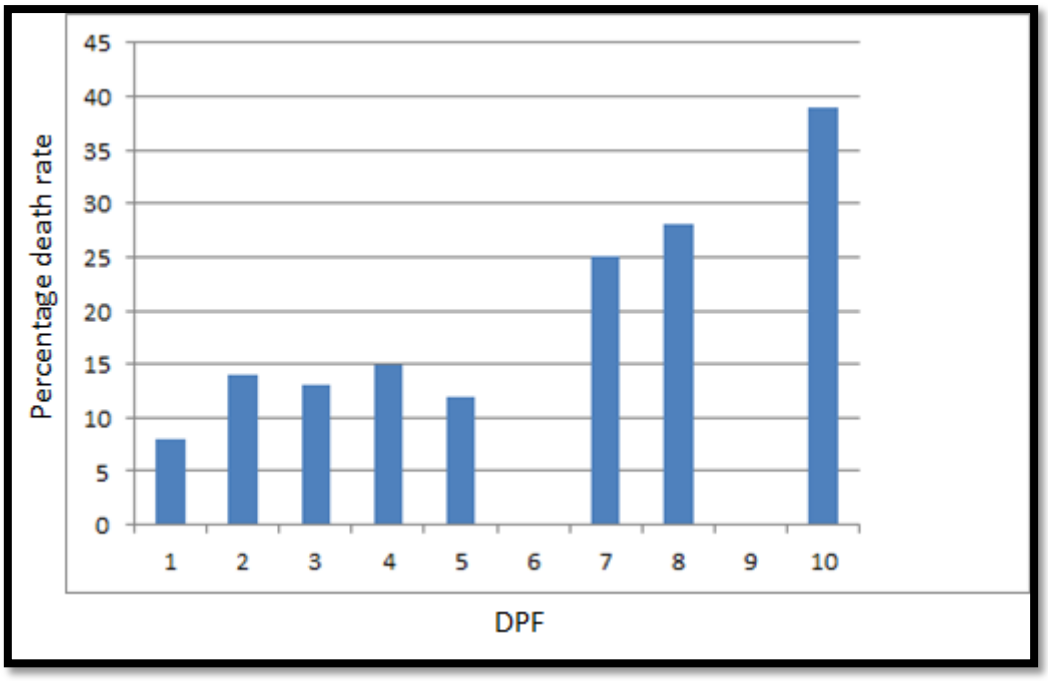


Figure 4.22: Bar diagram showing percentage death rate in larvae injected with sgRNA for *LACC1* and Cas9 mRNA. DPF- Day post fertilization.

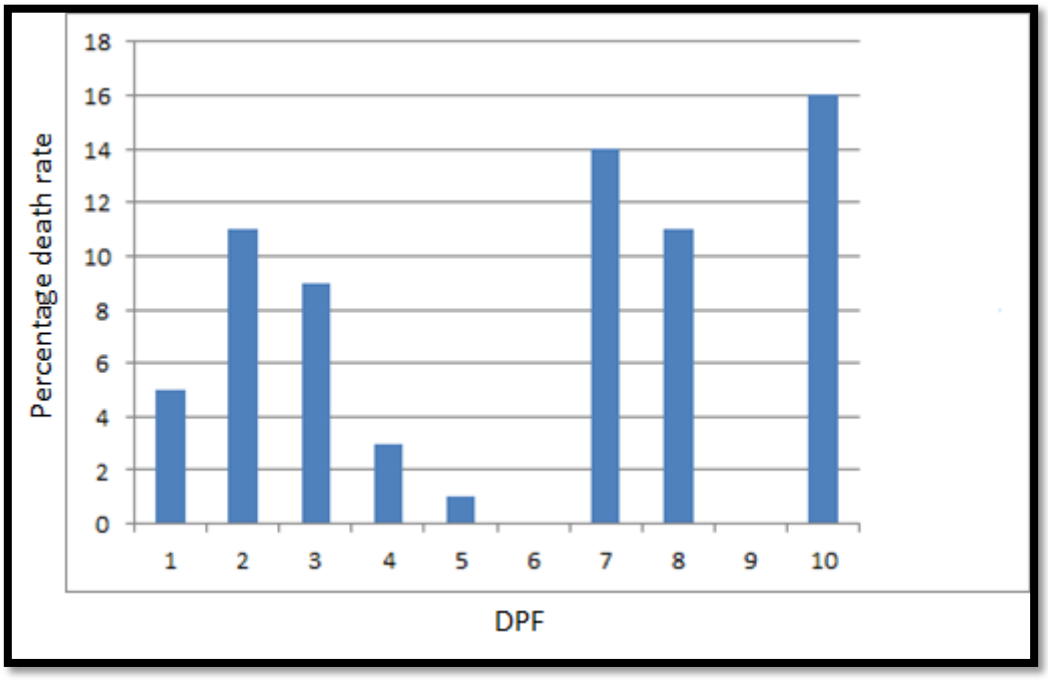


Figure 4.23: Bar diagram showing percentage death rate in larvae injected with only Cas9 mRNA (control). DPF- Day post fertilization.

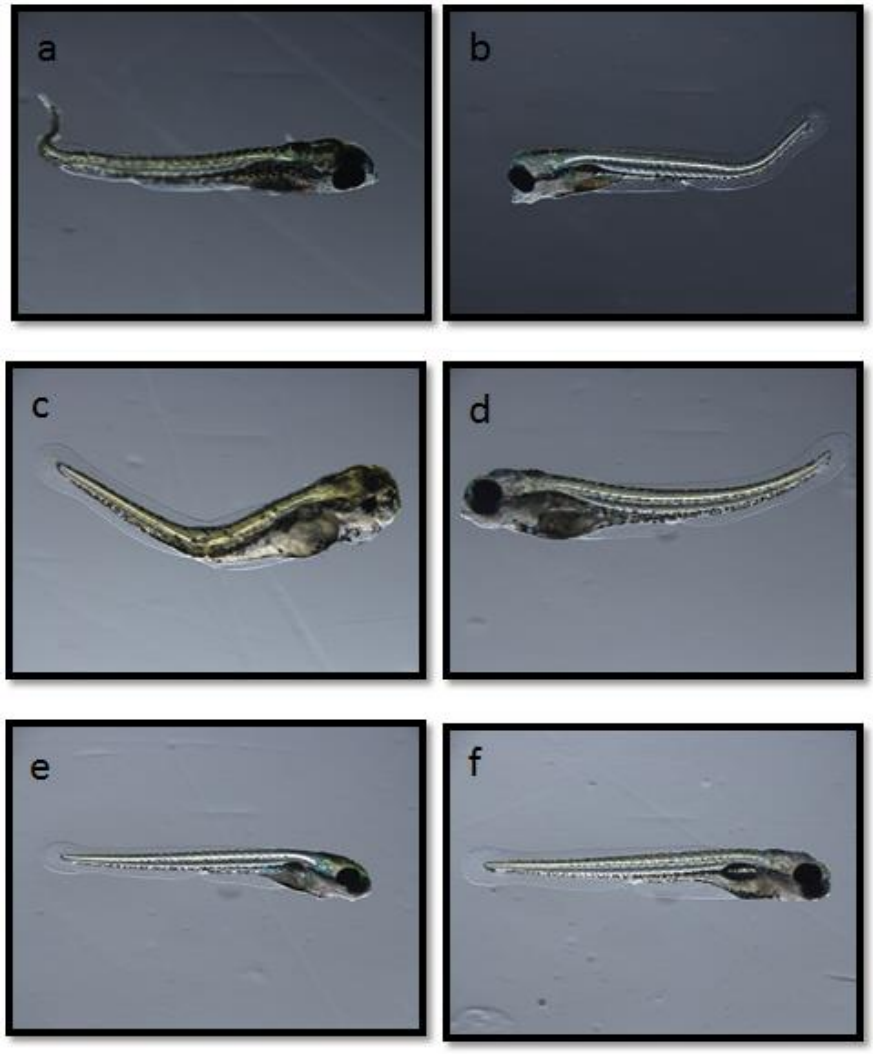


Figure 4.24: a) 7dpf old larvae developed from embryo injected with sgRNA for *LACC1* and Cas9 mRNA with elongated body and curved tail. b) 7dpf old larvae developed from embryo injected with sgRNA for *LACC1* and Cas9 mRNA with deformed head and curved posterior part. c) 7dpf old larvae developed from embryo injected with sgRNA for *LACC1* and Cas9 mRNA with head deformities and curved body. d) 7dpf old larvae developed from embryo injected with sgRNA for *LACC1* and Cas9 mRNA with curved body. e) and f) 7dpf old mocked larvae developed from embryo injected with only Cas9 mRNA.

4.8 Detection of mutations- Genotyping

Primers for genotyping were designed using the PRIMER 3 input online tool. Some of the injected larvae and control larvae were taken and genomic DNA was isolated from them and target exons were amplified and were sent for sequencing at Bioserve, India. Of the 10 injected samples sent for sequencing, 4 of them showed mutations at the target sites for exon 1 but not for exon five. There was no mutation in the exon 1 and exon 5 of control larvae. The sequences of injected larvae were aligned with the sequence of the controls using A Plasmid Editor (APE) software (Fig 4.25). The efficiency of the different sgRNA is shown below in Table 4.5.

Table 4.5: Showing efficiency of sgRNA used for generating knockout.

Target	Amount of sgRNA injected	Embryos	Embryos with mutation	%
Exon1	17.5ng/µL	10	4	40
Exon 5	17.5ng/µL	10	0	0

The sequences of injected larvae were aligned with the sequence of the controls using the APE software.

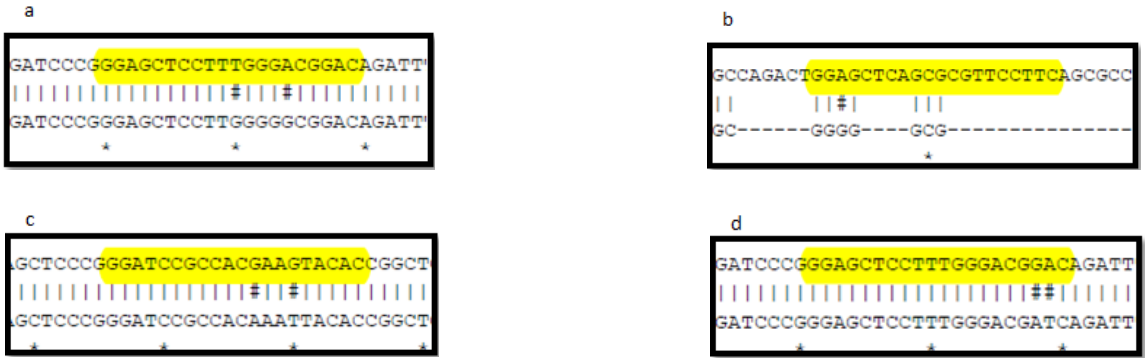


Figure 4.25: a) Sequence alignment between control (Upper sequence) and sgRNA injected larvae (down sequence) shows mutation at 2nd target site (highlighted region) of exon 1. # indicates that base **T** has been substituted with base **G** and **A** has been substituted with **G**. b) Sequence alignment between control (Upper sequence) and injected larvae(down sequence) shows mutation at 3rd target site of exon 1 (highlighted region). There is about 6 bp deletion in upstream of target site, more than 10 bp deletion at the target site and downstream of it. c) Sequence alignment between control (Upper sequence) and injected larvae (down sequence) shows mutation at 1st target site (highlighted region) of exon 1. # indicates that base **G** has been substituted with base **A** and **G** has been substituted with **T**. d) Sequence alignment between control (Upper sequence) and sgRNA injected larvae(down sequence) shows mutation at 2nd target site (highlighted region) of exon 1. # indicates that base **G** has been substituted with base **A** and **A** has been substituted with **T**.

Some bases were substituted with others in case of 'a', 'c', and 'd'. The reason may be that the Cas9 nuclease would have created double stranded break and the default mechanism of DNA repair (NHEJ or HDR) would have try to repair it but since these are error prone they were unable to completely repair the DSB created site with the same base sequences. In case of 'b' there is some base deletion upstream, downstream and at the target site. Since, in this case the deletion is large, so the default repair mechanism was able to repair very small part resulting in only some base insertion at the target site and large deletion.

Chromatogram showing homozygous and heterozygous mutation at target sites of exon 1 of *LACC1* gene is shown below (Fig 4.26-Fig 4.29).

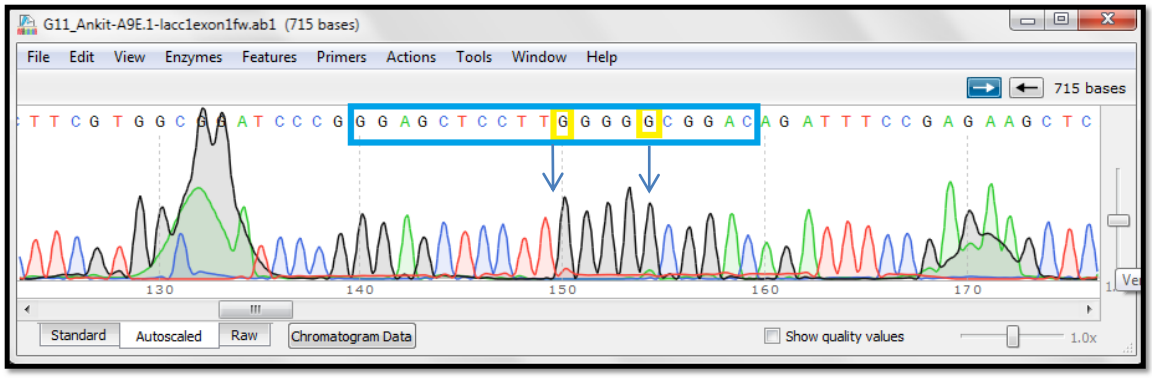


Fig 4.26: Chromatogram showing mutation at the 2nd target site of exon 1 shown in blue box and the mutated base shown in yellow box. The mutation is homozygous (T replaced with G at position 150 and A replaced with G at position 154) in this case since there is only one peak (shown by arrow).

Fig 4.27: Chromatogram showing mutation at the 3rd target site of exon 1. The 1st yellow box shows the region where around 6 bases have deleted between the two bases and the 2nd yellow box shows the region where 4 bases have been deleted between the two boxes.

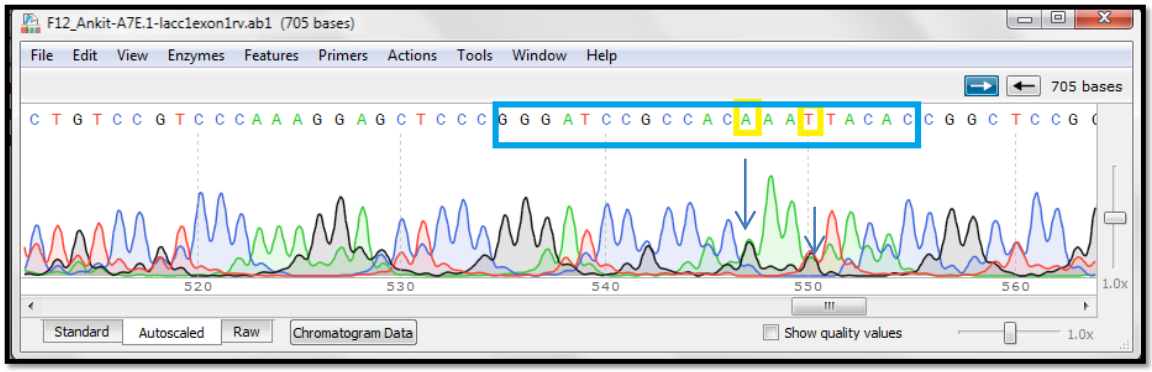


Fig 4.28: Chromatogram showing mutation at the 1st target site of exon 1 shown in blue box and the mutated base shown in yellow box. The mutation is homozygous (G replaced with A and G replaced with T) in this case since there is only one peak (shown by arrow).

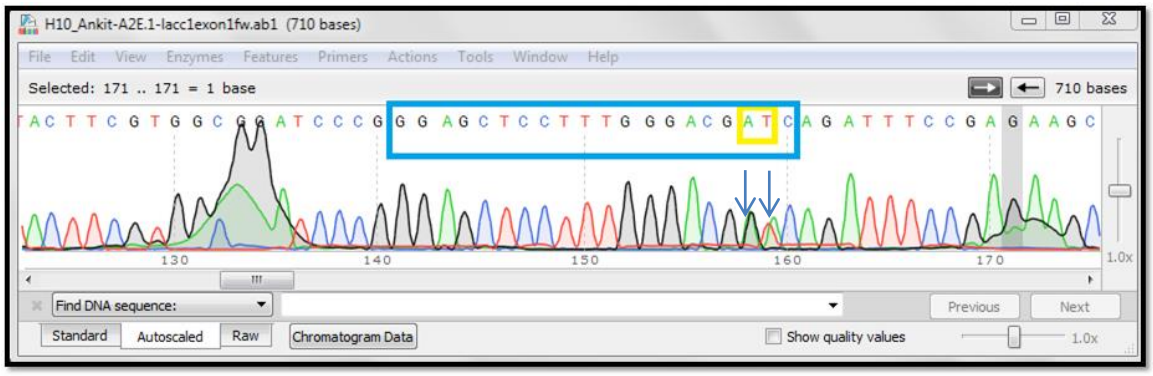


Fig 4.29: Chromatogram showing homozygous (G replaced with A) mutation in case of base A since there is one peak (shown by arrow) and heterozygous (A replaced with T) in case of base T since there is two peaks (shown by arrow) at the 2nd target site of exon 1 shown in blue box and the mutated base shown in yellow box.

In the Fig 4.26 and 4.28 there is homozygous mutation in the target site since there is only one peak in the mutated base but in case of Fig 4.27 there is deletion between the bases and there is noise as well, may be due to improper sequencing so it is difficult to say about homozygous or heterozygous mutation. In case of Fig 4.29 there is homozygous mutation in case of base A and heterozygous in case of base T since there are two peaks at the base T shown by arrow.

4.9 Zebrafish culture and breeding of the fishes injected with sgRNA and Cas9

For collection of embryos for microinjection normal adult healthy zebra fishes were used. Three to four breeding were set at a time. Next day more than 200 embryos were there in each breeding set and those embryos were used for microinjection. Microinjection was performed within 45 minutes from the time of embryos collection at one cell stage. Those which were in 2nd cell were not used and again fresh embryos that were in one cell staged were used. After injection those embryos were kept in petridish with methylene blue. Methylene blue water was changed on daily basis to avoid any infection. Till 5th day the larvae were kept in incubator at 28^{o}C that is optimum temperature for the growth of zebrafish. After that the larvae were kept in a room that have optimum condition maintained for the proper growth of fishes. Those larvae were feed after 4dpf once at time. After they were 1 month old they were feed two times in a day.

To maintain the larvae injected with sgRNA and Cas9 mRNA to adult stage, genotyping assay was performed only for some selected larvae. Due to lack of time it was not possible to perform any assay to check mutation in those live adult fishes but in due time they were breed to analyse the phenotypes of the F1 generation. Since, the fishes were fed twice a day there were more female than male fishes. So, after that they were fed only once in a day for 2-3 days and then were fed once on alternative days thus

leading to sex reversal in female. This happens in zebrafish in stress condition but exact mechanism is still unknown. After that, two set of breeding were set of the injected adult fishes in a gap of three days. But there were no embryos in the first two breeding set.

4.10 Phylogenetic tree construction using MEGA 7 software and analysis

Phylogenetic analysis using distance-based Neighbour joining method in MEGA 7 was carried out. The phylogenetic tree constructed using gene and protein sequence of *yfiH* and *LACC1* to show evolutionary relationship between different organisms is shown below.

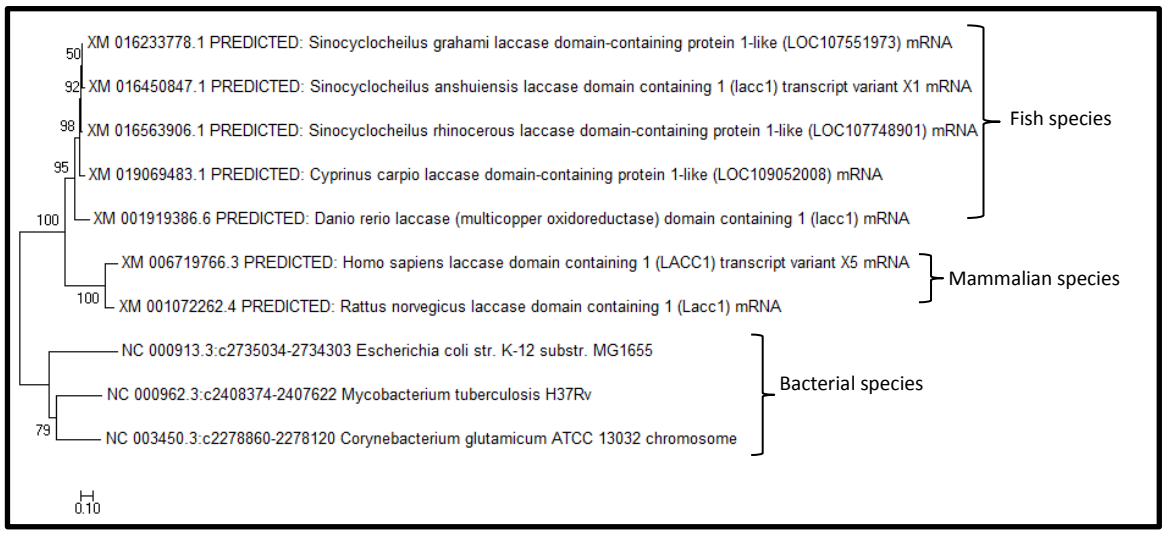


Fig 4.30: Evolutionary relationships based on *yfiH* gene sequence of bacteria and *LACC1* mRNA sequence of vertebrates using Neighbour-joining method. Numbers given at the main branches refer to bootstrap proportions among 1,000 bootstrap replicates.

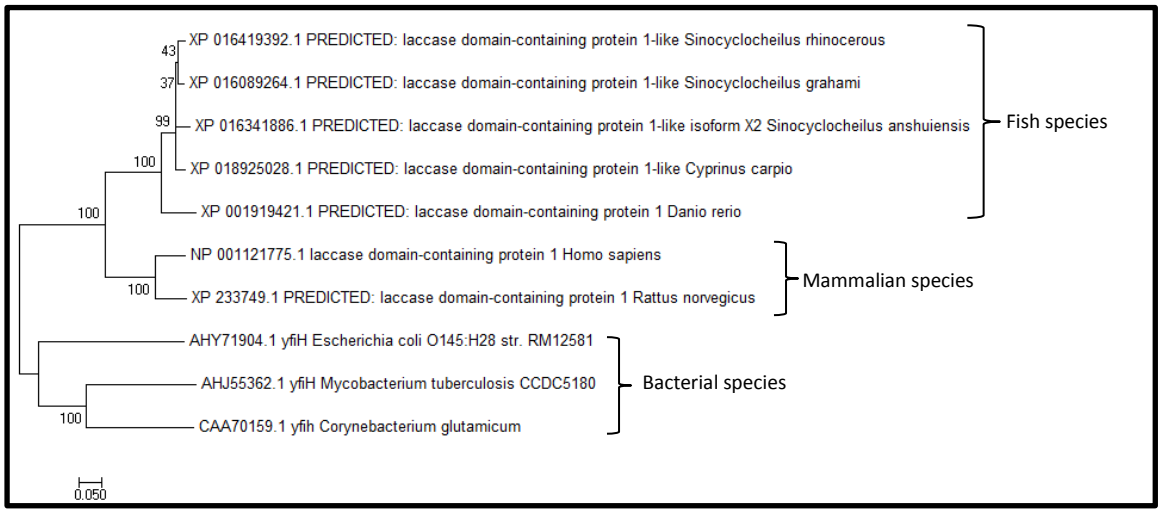


Fig 4.31: Evolutionary relationships based on *yfiH* protein sequence of bacteria and *LACC1* protein sequence of vertebrates using Neighbour-joining method. Numbers given at the main branches refer to bootstrap proportions among 1,000 bootstrap replicates.

The tree obtained was the bootstrap consensus tree of 1000 replicates. The tree was drawn to scale, with branch lengths (next to the branches) in same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum composite likelihood method in case of mRNA (vertebrates), gene sequences (bacteria) and p-distance method in case of protein sequence.

Three clusters were formed in both the cases that include bacterial species cluster, mammalian cluster and fish cluster. In fish cluster the closest species to *Danio rerio* was *Cyprinus* species and then was *Sinocyclocheilus* species. In mammalian species the closest species to *H. sapiens* was *R. norvegicus.* In bacterial cluster *M. tuberculosis* and *C. glutamicum* were closely related than *E. coli.*

Overall analysis of phylogenetic tree shows that there is evolutionary relationship between bacterial *yfiH* and *LACC1* of vertebrates. This is an already proven fact concluded by analysing the microbial and vertebrates genome.

CHAPTER 5
DISCUSSION

5.1 CRISPR/Cas and its implication in genome editing

CRISPR loci are present in bacteria and archaea and provide adaptive immunity in the prokaryotes. Cas9 from *Streptococcus pyogenes* has been successfully used in recent years as an efficient method for genome editing. It is cheap and its efficiency is same as that of TALEN and ZFNs and in some cases the efficiency of CRISPR Cas surpasses TALEN and ZFN method of genome editing (Cong *et al.,* 2013, Hwang *et al.,* 2013, Cho *et al.,* 2013). ZFN is quite expensive and difficult to design (Carroll, 2011). TALEN are cheap than ZFN but are much longer molecules and requires laborious cloning stages (Sander and joung, 2014). CRISPR Cas9 on the other hand overcomes each of these problems; it is cheap and short length of sgRNA also counter the complication associated with longer and highly repetitive TALEN encoding vector. In contrast to less specific protein and DNA interaction CRISPR uses simple Watson crick base pairing between sgRNA and target DNA sites (Garneau *et al.,* 2010, Gasiunas *et al.,* 2012). The use of bioinformatics tools and analysis to identify the best target region, molecular cloning steps and synthesis of sgRNA and Cas9 are very simplified in this system. In CRISPR Cas9 system of genome editing single Cas9 can be used for different sgRNA but in case of TALEN a new protein is engineered in each case.

The CRISPR-Cas9 genome editing technique was successfully used to generate knockout mutant for *LACC1* gene. However, it was only possible to create mutation in a single exon of the target gene. It appears that the zebra fish larvae injected with the sgRNA where showing some deformities and were dying prematurely (at 7-9 dpf) compared to the control larvae. Although this observation needs to be confirmed, it is interesting. In addition, the presence of mutations in these larvae has to be confirmed because it was not possible to analyse all the larvae and only some of them were used for genotyping and other analysis. Tentatively, results suggest that *LACC1* gene is crucial for development of zebra fish and mutants with a defective gene die due to multiple defects in body shape and organ development.

5.2 *LACC1* gene and its role in some autoimmune and auto-inflammatory disorders in human

LACC1 a gene of unknown function is a protein coding gene. No precise information is available till now that can explain anything about this gene function. The bacterial homolog of this gene i.e. *YfiH* has also been studied but it function is also not known. Work done on *Brevibacterium lactofermentum* suggests that this gene has no role in maintaining the viability and cell growth (Honrubia *et al.,* 2001). But recent research (yet

to be published) done on *E.coli* was different, it was found that this gene plays role in peptidoglycan metabolism in bacteria which is quite interesting.

Auto inflammatory disorders can be both polygenic and monogenic. Implication of *LACC1* gene in polygenic auto inflammatory disorders like Crohn's disease, Leprosy and Rheumatoid arthritis have already been proved (Liu *et al.,* 2014, Wakil *et al.,* 2014, Sales‐Marques *et al.,* 2014). No data was available about this gene solely leading to any auto inflammatory disorders by its own until the research done on 13 patients of juvenile idiopathic arthritis (JIA) in Saudi Arabia shows that a single mutation at exon 4 of *LACC1* gene leads to juvenile idiopathic arthritis. In all of the patients, the conserved cysteine present at 284th position was replaced by an arginine residue, thus implicating that mutations in this gene can solely lead to some autoimmune and auto inflammatory conditions. Mutation at p.C284R which is an evolutionary variant, leads to systemic juvenile idiopathic arthritis (Wakil *et al.,* 2014).

5.3 Efficient monoallelic and biallelic mutations

CRISPR Cas9 technique is efficient method for genome editing and induces efficient monoallelic and at some rate biallelic mutations as well (Whitworth *et al.,* 2014).

It was possible to create biallelic mutation at 1st and 2nd target site of exon 1 *LACC1* gene. Usually injected zebrafish embryos in fact are heterozygous so to create homozygous mutation the founder fishes should be grown and genotyped to identify two fishes having heterozygous mutation so they can be crossed to reach F1 progeny. But in this case there was biallelic mutation; Since Cas9 can induce biallelic mutation as well. Though the mutation was biallelic but since no functional study of *LACC1* gene using knockdown and knockout assay has been done before so it can't be assume that the phenotypes are due to the loss of function of this gene until unless western blotting or other assays are done.

T7 endonuclease assay was used as a preliminary method for detection of mutation in the samples (D'Agostino *et al.,* 2015). T7 endonuclease is an enzyme that cleaves non-perfectly matched DNA. Since, it is easy and less time consuming so it is used for detecting mutation in first phase of analysis of mutants in this kind of study. If more than one band is seen in gel after T7 endonuclease assay one can say that there is mutation due to which T7 endonuclease enzyme have cleaved the product. Though it is frequently used for detecting mutation but it is not a reliable method because it gives false positive results due to single nucleotide polymorphism.

5.4 Off-target activity and efficiency

Results showed that sgRNA Cas9 mediated mutagenesis was only in one exon i.e., exon 1 of the *LACC1* gene, since some sgRNA are more efficient in creating mutations than the

others. There may be some unknown mechanism that effect the overall activity of sgRNA Cas9 complex, which when revealed can explain why certain sgRNA complex with nucleotides are more efficient than the others.

Since, three sgRNA were used for targeting three sites in exon 1 of *LACC1* gene so using a number of sgRNA for single target can increases the chances of mutation in the target site. But it was not possible to use the same idea for exon 5 as the online tool designed only one 22 bp of oligos that was most efficient in targeting the exon 5, others oligos were least efficient and had greater off target activity. So, only one sgRNA was used for targeting exon 5 of *LACC1.*

The off-target activity of Cas9 is one of the major drawbacks of this system and should be kept in mind while using this technique. Since, Cas9 can tolerate some mismatches at the target site and also Cas9 can act at sites of partial homology between the target site and sgRNA (Fu *et al.,* 2013; Hsu *et al.,* 2013; Mali *et al.,* 2013b; Pattanayak *et al.,* 2013), there will be chances of a number of off-target mutations. Therefore, while designing the sgRNA only those with high specificity and efficiency should be selected. Also off target sites followed by 5' NAG protospacer adjacent motif can also lead to off target cleavage demonstrating the importance of considering both NGG and NAG protospacer adjacent motif in off target analysis (Jiang *et al.,* 2013; Hsu *et al.,* 2013). The mutagenesis at on-target and off-target genome locations could lead to the phenotype which would not only be a consequence of the intended mutated gene but also a synergistic effect of several mutations in genome making it difficult to conclude whether the phenotype shown by mutants is due to the KO of target gene alone or not.

Since, sgRNA online designing tool showed the off target sites so analysis of these sites can be done to ensure that the phenotype shown is due to sole activity of knockout of on target site only. Potential off targets can be computationally determine by searching genomic sequences with high sequence similarity to desired locus, whole genome sequencing can also be done. Although oligos with highest specificity and efficiency were used but off target activity can't be totally ruled out. The concentration of Cas9 also plays role in increasing the off-target activity of the Cas9, the higher the concentration of Cas9, the chances of it to tolerate the mismatches increases (Hsu *et al.,* 2013; Pattanayak *et al.,* 2013). The optimum concentration of Cas9 to be used is 300ng/µl.

5.5 Evolutionary analysis

Phylogenetic analysis helps to study evolutionary history, development and relationship among group of organism. Closely related species invariably get clustered in same clade. Phylogenetic tree was constructed to show the evolutionary relationship between different organisms by using the *yfiH* and *LACC1* gene and protein sequences.

Human and microbial genome analysis has already proved that genes from bacteria had transferred to humans during vertebrate evolution (International Human genome sequencing consortium, Nature 409, 860, 2001). So, phylogenetic tree was constructed using Mega 7 software and it was found that there is some kind of evolutionary relationship in *yfiH* gene found in bacteria and *LACC1* gene found in vertebrates.

5.6 Maintaining the zebrafish

Maintaining those injected embryos was a crucial step and was done very carefully. While capturing images the larvae were kept in tricaine containing water for very short period of time because prolonged exposure may lead to death. The larvae were washed 2-3 times after taking images to remove excess tricaine, because if unwashed it can lead to deformities. Though no written fact is available till know that proves that tricaine lead to deformities in zebrafish but no risk was taken and tricaine was removed as far as possible.

Zebrafish were kept in methylene blue containing water because it acts as antifungal and antibacterial agent in some extent. Methylene blue water was changed every day and when zebrafish were fed excess care was taken and the methylene blue water was removed twice a day to remove the debris and faecal material which if not removed can lead to bacterial and fungal infection.

5.7 Limitations

The ability to create mutation at any target site by using just sgRNA complementary to target site and Cas9 holds an enormous advantage in functional biology. But more about this system should be studied and revealed in order to minimize the risk of off-target effects when generating mutations, and as a result draw the correct conclusions about protein functions while interpreting by observing the phenotypes of mutants. The fact that the effectiveness of mutagenesis in certain regions cannot be predicted due to efficiency of sgRNA is also a limitation, with a range of effectiveness from zero, as in the examples of exon 5 of *LACC1*.

Since, no previous data of phenotype shown by knockdown or knockout mutant was available so comparison was not possible that would have eases the analysis of mutants. Genotyping is a crucial step for determining the effectiveness of this system. In this study, T7 endonuclease assay and sequence analysis was performed to find the mutations, but other techniques such as High Resolution Melting Analysis (HRMA), fragment length analysis can also be used to determine the mutagenesis (D'Agostino et al., 2015). It was not possible to maintain the deformed larvae for generating F1 generation for analysis because they were dying due to infection within 9 dpf.

5.8 An on-going project

Since, this kind of study has not been done before for *LACC1* gene so further analysis needs to be done before drawing any conclusion. Knockdown mutants can be generated using Morpholinos to check if they show same kind of phenotype or not.

New combination of sgRNA and Cas9 for exon 5 can be used to generate mutation in exon 5 and the same can be done again and again for exon 1 as well, before reaching to any conclusion. Breeding was set for generating the F1 embryos from the injected 3 month old fish but there was no embryo in the 1^{st} and 2^{nd} breeding. So, this can be repeated, injected fish can be crossed to generate F1 embryo. Thus, generated F1 generation can be used for phenotypic analysis. Western blotting can be done for the *LACC1* protein to confirm that the mutant thus generated yield full loss of function.

CHAPTER 6

SUMMARY

During the course of evolution genes from prokaryotes had been transferred to eukaryotes. Among 100 of genes thus transferred our studies aim at *YfiH* gene, a gene of unknown function that has been transferred from prokaryotes to eukaryotes through lateral gene transfer. *YfiH* is an uncharacterized protein that is widely distributed from bacteria to human. The function of the *yfiH* gene is not known, same is the case for its homolog in human that is *LACC1*. To show the evolutionary relationship between different organism with regard to *yfiH* and *LACC1,* gene and protein sequences of *yfiH* and *LACC1* was used to create the phylogenetic tree using MEGA 7 software. Analysis of phylogenetic tree shows that there is evolutionary relationship between different organisms (bacteria-vertebrates) in regard with *yfiH* and *LACC1*.

LACC1 [laccase (Multicopper oxidoreductase) domain containing 1] is a protein coding gene. A review of ENCODE project data for *LACC1*, including expression array, DNA hypersensitivity and CHIP-Seq tracks provide no obvious insight into the biology of this enzyme. GWA studies suggest that SNPs in this gene along with other gene can lead to disease condition like leprosy, Crohn's disease, Rheumatoid arthritis and others in human. Recent study done in patient of Juvenile idiopathic arthritis suggests that mutation in exon 4 of this gene can solely lead to JIA. But what is the role of this gene is unknown. So, to know something about *LACC1* gene, this gene was knockout from zebrafish using CRISPR/ Cas9 technique. Zebrafish is the best model organism to study vertebrate functional biology. The homology between human and zebrafish genome, ease of maintenance and cost effectiveness are the reasons because of which zebrafish is gaining popularity as a model organism to study the role of a gene in disease condition in human. CRIPSR/Cas9 is a simple technique that is cheap and requires only sgRNA and Cas9 nuclease to create double stranded break at the target site. The double stranded breaks created are either repair by NHEJ or HDR leading to insertion or deletion in the target site. The target exons were exon 1, exon 2 and exon 5 of the *LACC1* gene. Oligos for exon 1 and exon 5 were cloned in the vector pDR274 but it was possible to create mutation in exon 1 but not in exon 5. T7 endonuclease assay and genotyping, of the target exons were the methods used to detect mutation. Elongated body and curved tail were the phenotypes shown by the mutants and they were dying prematurely at 7-9dpf. It is just a preliminary study and repetition of this work need to be done. Those fishes that were not showing any deformities were alive and further analysis of those can be done. The founder fishes can be breed and F1 generation can be analysed. Our study done in this gene suggests that it is required for proper growth and development of the zebrafish.

CHAPTER 7

CONCLUSION

It was possible to achieve one aim of the project, since only the sgRNA-Cas9 mutagenesis in exon 1 of target gene was obtained but not for exon 5. This study supports the previous reported cases of high occurrence of Cas9-derived mutagenesis for some guide RNA but not for the other. It was not possible to generate F1 generation from the injected fishes due to lack of time. The present study gave some data, not profound but a preliminary data that put some light in the role of *LACC1* gene. This study suggests that the *LAAC1* gene may plays role in proper growth and development of the zebrafish since the zebrafish mutants were dying due to multiple defects in body shapes and organ development. The mutants were sick and have irregular movement and they were very prone to infection as compared to the control. This work is just a start and more substantiate research need to be done to prove any of the statement. Recent studies show that this gene plays role in auto inflammatory condition in human like rheumatoid arthritis, systemic juvenile arthritis, leprosy and so on but it proper function is still unknown. Since, functional biology of this gene has not been studied before, so repetition of this work is required before drawing any conclusion. Despite some problems, it is sure that CRISPR-Cas genome editing technique will help to understand the functional aspects of genes like *LACC1* which is a key gene that leads to auto-inflammatory and autoimmune disease conditions in human.

Recommendations:

1. The present study was focused to know the effect of *LACC1* gene knockout in zebrafish, and what kind of phenotype is shown by the knockout mutants. The next step will be to knockdown the gene to see if same phenotypes are shown by the knockdown mutants as well and perform western blotting to see if the protein is produce in the mutants or not.
2. With the combined data of knockout and knockdown mutants functional aspect of the *LACC1* gene can be known.

CHAPTER 8
REFERENCES

Aksentijevich, I., Nowak, M., Mallah, M., Chae, J. J., Watford, W. T., Hofmann, S. R. Goldbach-Mansky, R. (2002). De novoCIAS1 mutations, cytokine activation, and evidence for genetic heterogeneity in patients with neonatal-onset multisystem inflammatory disease (NOMID): A new member of the expanding family of pyrin-associated autoinflammatory diseases. Arthritis & Rheumatism, 46(12), 3340-3348. doi:10.1002/art.10688.

Ayaz, N. A., Ozen, S., Bilginer, Y., Erguven, M., Taskiran, E., Yilmaz, E., Bakkaloglu, A. (2008). MEFV mutations in systemic onset juvenile idiopathic arthritis. Rheumatology, 48(1), 23-25. doi:10.1093/rheumatology/ken409.

Baldrian, P. (2006). Fungal laccases – occurrence and properties. FEMS Microbiology Reviews, 30(2), 215-242. doi:10.1111/j.1574-4976.2005.00010.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science, 315(5819), 1709-1712. doi:10.1126/science.1138140.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551–2561.

Boto, L. (2010). Horizontal Gene Transfer in Evolution. Encyclopedia of Life Sciences. doi:10.1002/9780470015902.a0022835.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Oost, J. V. (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. Science, 321(5891), 960-964. doi:10.1126/science.1159689.

Carroll, D. (2011). Genome Engineering With Zinc-Finger Nucleases. Genetics, 188(4), 773-782. doi:10.1534/genetics.111.131433.

Cho, S. W., Kim, S., Kim, J. M., & Kim, J. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nature Biotechnology, 31(3), 230-232. doi:10.1038/nbt.2507.

Chylinski, K., Rhun, A. L., & Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biology, 10(5), 726-737. doi:10.4161/rna.24321.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science, 339(6121), 819-823. doi:10.1126/science.1231143.

D'Agostino, Ylenia, Annamaria Locascio, Filomena Ristoratore, Paolo Sordino, Antonietta Spagnuolo, Marco Borra, and Salvatore D'Aniello. "A Rapid and Cheap Methodology for CRISPR/Cas9 Zebrafish Mutant Screening." Molecular Biotechnology 58.1 (2015): 73-78.

Dahm, R. (2006). The Zebrafish Exposed. American Scientist, 94(5), 446. doi:10.1511/2006.61.446.

Day, T. G., Ramanan, A. V., Hinks, A., Lamb, R., Packham, J., Wise, C., Donn, R. P. (2008). Autoinflammatory genes and susceptibility to psoriatic juvenile idiopathic arthritis. Arthritis & Rheumatism, 58(7), 2142-2146. doi:10.1002/art.23604.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature, 471(7340), 602-607. doi:10.1038/nature09886.

Deveau, H., Barrangou, R., Garneau, J. E., Labonte, J., Fremaux, C., Boyaval, P., Moineau, S. (2007). Phage Response to CRISPR-Encoded Resistance in Streptococcus thermophilus. Journal of Bacteriology, 190(4), 1390-1400. doi:10.1128/jb.01412-07.

Doolittle, W. F. (1999). Phylogenetic Classification and the Universal Tree. Science, 284(5423), 2124-2128. doi:10.1126/science.284.5423.2124.

Esvelt, K. M., & Wang, H. H. (2014). Genome-scale engineering for systems and synthetic biology. Molecular Systems Biology, 9(1), 641-641. doi:10.1038/msb.2012.66.

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature Biotechnology, 31(9), 822-826. doi:10.1038/nbt.2623.

Garneau, J. E., Dupuis, M., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature, 468(7320), 67-71. doi:10.1038/nature09523.

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences, 109(39). doi:10.1073/pnas.1208507109.

Hisano, Y., Sakuma, T., Nakade, S., Ohga, R., Ota, S., Okamoto, H., Kawahara, A. (2015). Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. Scientific Reports, 5, 8841. doi:10.1038/srep08841.

Honrubia, M., Ramos, A., & Gil, J. (2001). The cell division genes ftsQ and ftsZ , but not the three downstream open reading frames YFIH, ORF5 and ORF6, are essential for growth and viability in Brevibacterium lactofermentum ATCC 13869. Molecular Genetics and Genomics, 265(6), 1022-1030. doi:10.1007/s004380100497.

Horvath, P., Romero, D. A., Coute-Monvoisin, A., Richards, M., Deveau, H., Moineau, S., Barrangou, R. (2007). Diversity, Activity, and Evolution of CRISPR Loci in Streptococcus thermophilus. Journal of Bacteriology, 190(4), 1401-1412. doi:10.1128/jb.01415-07.

Howe, k., Clark, M., Torrance, J., (2013). The zebrafish reference genome sequence and its relationship to the human genome. Nature doi:10.1038/nature12111.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature Biotechnology, 31(9), 827-832. doi:10.1038/nbt.2647.

<http://www.genecards.org/cgi-bin/carddisp.pl?gene=LACC1.>

Hwang, Woong Y., Yanfang Fu, Deepak Reyon, Morgan L. Maeder, Shengdar Q. Tsai, Jeffry D. Sander, Randall T. Peterson, J-R Joanna Yeh, and J. Keith Joung. "Efficient genome editing in zebrafish using a CRISPR-Cas system." Nature Biotechnology 31.3 (2013): 227-29.

International Human Genome Sequencing Consortium, *Nature* 409, 860 (2001).

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. Journal of Bacteriology, 169(12), 5429-5433. doi:10.1128/jb.169.12.5429-5433.1987.

Jansen, R., Embden, J. D., Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Molecular Microbiology, 43(6), 1565-1575. doi:10.1046/j.1365-2958.2002.02839.

Jiang, Wenyan, David Bikard, David Cox, Feng Zhang, and Luciano A. Marraffini. "RNA-guided editing of bacterial genomes using CRISPR-Cas systems." Nature Biotechnology 31.3 (2013): 233-39.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science, 337(6096), 816-821. doi:10.1126/science.1225829.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. ELife, 2. doi:10.7554/elife.00471.

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Doudna, J. A. (2014). Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. Science, 343(6176), 1247997-1247997. doi:10.1126/science.1247997.

Kim, Y., Maltseva, N., Dementieva, I., Collart, F., Holzle, D., & Joachimiak, A. (2006). Crystal structure of hypothetical protein YfiH from Shigella flexneri at 2 Å resolution. Proteins: Structure, Function, and Bioinformatics, 63(4), 1097-1101. doi:10.1002/prot.20589.

Kotani, H., Taimatsu, K., Ohga, R., Ota, S., & Kawahara, A. (2015). Efficient Multiple Genome Modifications Induced by the crRNAs, tracrRNA and Cas9 Protein Complex in Zebrafish. Plos One, 10(5). doi:10.1371/journal.pone.0128319.

Liu, J. Z., & Anderson, C. A. (2014). Genetic studies of Crohn's disease: Past, present and future. Best Practice & Research Clinical Gastroenterology, 28(3), 373-386. doi:10.1016/j.bpg.2014.04.009.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V.(2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct 1, 7.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., et al. (2011). Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. 9, 467–477.

Mali, Prashant., John Aach, P. Benjamin Stranges, Kevin M. Esvelt, Mark Moosburner, Sriram Kosuri, Luhan Yang, and George M. Church. "CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering." Nature Biotechnology 31.9 (2013b): 833-38.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., Dicarlo, J. E., Church, G. M. (2013). RNA-Guided Human Genome Engineering via Cas9. Science, 339(6121), 823-826. doi:10.1126/science.1232033.

Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M., Kowallik, k. V. (1998). Gene transfer to the nucleus and evolution of chloroplast. *Nature* 393, 162.

Masters, S. L., Simon, A., Aksentijevich, I., & Kastner, D. L. (2009). Horror Autoinflammaticus: The Molecular Pathophysiology of Autoinflammatory Disease. Annual Review of Immunology, 27(1), 621-668. doi:10.1146/annurev.immunol.25.022106.141627.

Mojica, F. J., Diez-Villasenor, C., Garcı´a-Mart Nez, J., & Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic

Elements. Journal of Molecular Evolution, 60(2), 174-182. doi:10.1007/s00239-004-0046-3.

Mojica, F. J., Diez-Villasenor, C., Soria, E., & Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Molecular Microbiology, 36(1), 244-246. doi:10.1046/j.1365-2958.2000.01838.

Nishimasu, H., Ishitani, R., & Nureki, O. (2014). Crystal structure of Streptococcus pyogenes Cas9 in complex with guide RNA and target DNA. doi:10.2210/pdb4oo8/pdb.

Niu, Y., Shen, B., Cui, Y., Chen, Y., Wang, J., Wang, L., Sha, J. (2014). Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-Mediated Gene Targeting in One-Cell Embryos. Cell, 156(4), 836-843. doi:10.1016/j.cell.2014.01.027.

Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., & Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nature Biotechnology, 31(9), 839-843. doi:10.1038/nbt.2673.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. (2005). Microbiology, 151(3), 653-663. doi:10.1099/mic.0.27437-0.

Puchta, H., & Fauser, F. (2013). Gene targeting in plants: 25 years later. The International Journal of Developmental Biology, 57(6-7-8), 629-637. doi:10.1387/ijdb.130194hp.

Sales-Marques, C., Salomão, H., Fava, V. M., Alvarado-Arnez, L. E., Amaral, E. P., Cardoso, C. C., Latini, A. C. (2014). NOD2 and CCDC122-LACC1 genes are associated with leprosy susceptibility in Brazilians. Human Genetics, 133(12), 1525-1532. doi:10.1007/s00439-014-1502-9.

Salzberg, S. L. (2001). Microbial Genes in the Human Genome: Lateral Transfer or Gene Loss? Science, 292(5523), 1903-1906. doi:10.1126/science.1061036.

Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nature Biotechnology, 32(4), 347-355. doi:10.1038/nbt.2842.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Research, 39(21), 9275-9282. doi:10.1093/nar/gkr606.

Shah, S. A., Erdmann, S., Mojica, F. J., & Garrett, R. A. (2013). Protospacer recognition motifs. RNA Biology, 10(5), 891-899. doi:10.4161/rna.23764.

Spence, R., Gerlach, G., Lawrence, C., & Smith, C. (2007). The behaviour and ecology of the zebrafish, Danio rerio. Biological Reviews, 83(1), 13-34. doi:10.1111/j.1469-185x.2007.00030.

Tang, T., Bachellerie, J., Rozhdestvensky, T., Bortolin, M., Huber, H., Drungowski, M., Huttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. Proceedings of the National Academy of Sciences, 99(11), 7536-7541. doi:10.1073/pnas.112047299.

Wakil, S. M., Monies, D. M., Abouelhoda, M., Al-Tassan, N., Al-Dusery, H., Naim, E. A., Al-Mayouf, S. (2014). Association of a Mutation in LACC1 With a Monogenic Form of Systemic Juvenile Idiopathic Arthritis. Arthritis & Rheumatology, 67(1), 288-295. doi:10.1002/art.38877.

White, R. M., Sessa, A., Burke, C., Bowman, T., Leblanc, J., Ceol, C., Zon, L. I. (2008). Transparent Adult Zebrafish as a Tool for In Vivo Transplantation Analysis. Cell Stem Cell, 2(2), 183-189. doi:10.1016/j.stem.2007.11.002.

Whitworth, K. M., K. Lee, J. A. Benne, B. P. Beaton, L. D. Spate, S. L. Murphy, M. S. Samuel, J. Mao, C. O'gorman, E. M. Walters, C. N. Murphy, J. Driver, A. Mileham, D. Mclaren, K. D. Wells, and R. S. Prather. "Use of the CRISPR/Cas9 System to Produce Genetically Engineered Pigs from In Vitro-Derived Oocytes and Embryos." Biology of Reproduction 91.3 (2014): 78.

Wong SH, Hill AV, Vannberg FO, India–Africa–United Kingdom Leprosy Genetics C (2010). Genomewide association study of leprosy. N Engl J Med 362:1446–1447. doi:10.1056/NEMc1001451 (author reply 1447–1448).

Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C., Hsu, P. D., Dadon, D. B., Sharp, P. A. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nature Biotechnology, 32(7), 670-676. doi:10.1038/nbt.2889.

Zhang, FR., Huang, W., Chen, SM., Sun, LD., Liu, H., Li, Y., Cui, Y., Yan, XX., Yang, HT., Yang, RD., Chu, TS., Zhang, C., Zhang, L., Hang, JW, Liu, JJ. (2009). Genomewide association study of leprosy. N Engl J Med 361:2609–2618. doi:10.1056/NEJMoa0903753.

Appendix

Appeendix 1. Oligonucleotides for guide RNA synthesis

Seq NAME	Sequences	T_m	ODU	Conc in 200µL of water			Vol in µl added to get 100pm/ µl or 100 µM conc
				pm/µL	nm/µL	µg/ µL	
Lacc1 exon 1.1 Fw	TAGGATCCGCCACG AAGTACAC	57	10.04	208.3	.2083	1.3	416.6
Lacc1 exon 1.1 Rv	AAACGTGTACTTCG TGGCGGAT	55	11.22	235.05	.23505	1.5	470.1
Lacc1 exon 1.2 Fw	TAGGTCGTCCCAAA GGAGCTCC	59	10.36	224.23	.22423	1.5	448.5
Lacc1 exon 1.2 Rv	AAACGGAGCTCCTT TGGGACGA	57	11.24	230.21	.23021	1.5	460.4
Lacc1 exon 1.3 Fw	TAGGACGCGCTGA GCTCCAGTC	60	10.56	231.76	.23176	1.5	463.5
Lacc1 exon 1.3 Rv	AAACGACTGGAGC TCAGCGCGT	59	10.74	221.2	.2212	1.4	442.4
Lacc1 exon 5.1 Fw	TAGGAGAGACGGA CTCAACTTT	53	11.28	226.82	.22682	1.5	453.6
Lacc1 exon 5.1 Rv	AAACAAAGTTGAG TCCGTCTCT	51	11.2	233.25	.23325	1.5	466.5
Lacc1 exon 2.1 Fw	TAGGCTTCAGCACC CGTACCGG	60	10.98	246.34	.24634	1.6	492.7
Lacc1 exon 2.1 Rv	AAACCCGGTACGG GTGCTGAAG	59	12.82	258.75	.25875	1.7	517.5

Appendix 2. Primers for genotyping

Seq NAME	Sequences	T_m	ODU	Conc in 200µL of water			Vol in µl added to get 100pm/ µl or 100 µM conc
				pm/µL	nm/µL	µg/ µL	
Lacc1 exon 1 Fw	CTCCCAGAGTCTG CTTGTGA	54	11.36	286.36	.28636	1.7	572.7
Lacc1 exon 1 Rv	AGACAAAGGTTA CCGGTTTCT	50	12.16	264.52	.26452	1.7	529
Lacc1 exon 2 Fw	CTGCTGCTGGAG TGTTTCAA	52	10.88	265.81	.26581	1.6	531.6
Lacc1 exon 2 Rv	GGCTAGTGGTCT TTTCCTGC	54	10.18	268.3	.2683	1.6	536.6
Lacc1 exon 5 Fw	TGCTGGAAAGAG GTGGGATT	52	12.34	264.65	.26465	1.6	529.3
Lacc1 exon 5 Rv	TTTTGTGCATTCG GTCAGGG	52	10.62	261.81	.26181	1.6	523.6

Appendix 3. NEBuffer 2.1

1X Buffer Components

50mM NaCl

10mM Tris HCl

10mM $MgCl_2$

100µg/ml BSA

pH 7.9@25°C

Appendix 4. NEBuffer 2

1X Buffer Components

50mM NaCl

10mM Tris HCl

10mM $MgCl_2$

1mM DTT

pH 7.9@25°C

Appendix 5. NEBuffer 3.1

1X Buffer Components

100mM NaCl

50mM Tris HCl

10mM $MgCl_2$

100µg/ml BSA

pH 7.9@25°C

Appendix 6.T4 DNA Ligase Reaction Buffer

1X Buffer Components

50mM Tris HCl

10mM $MgCl_2$

1mM ATP

10mM DTT

pH 7.5@25°C

Appendix 7.

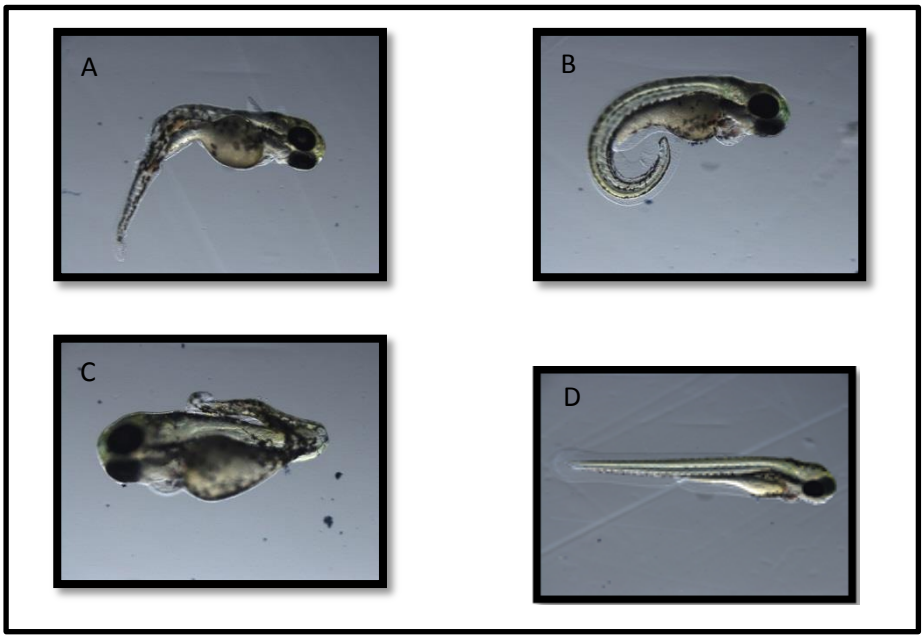


Figure: Unanalysed larvae developed from embryo injected with Cas9 mRNA and sgRNA. A) Four dpf old larvae with body bend at the middle. B) Four dpf old larvae with curved tail. C) Four dpf old larvae with curved tail. D) Four dpf old larvae injected with only Cas9 mRNA.

Appendix 8.

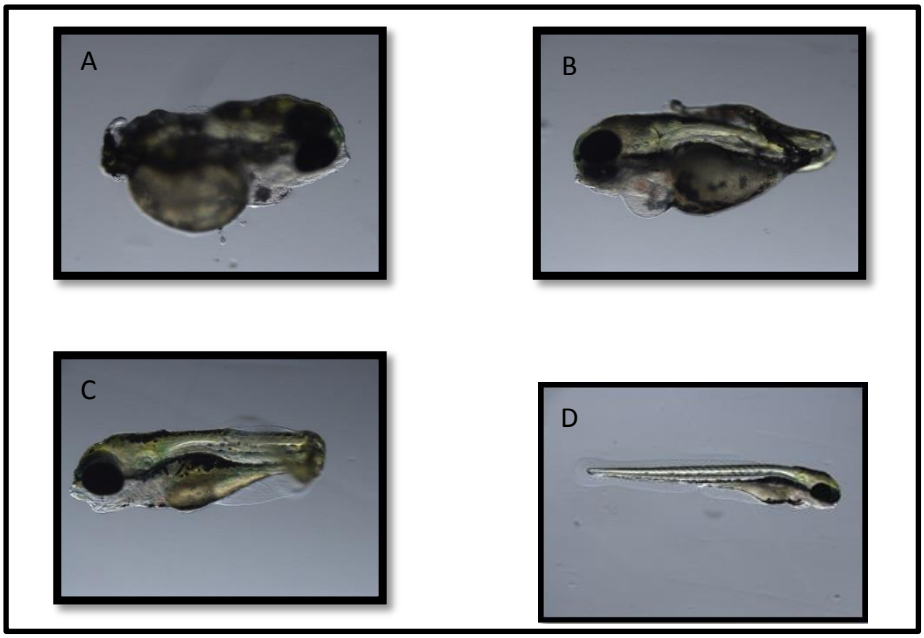


Figure: Unanalysed larvae developed from embryo injected with Cas9 mRNA and sgRNA. A) Five dpf old larvae with totally deformed posterior end. B) Five dpf old larvae with deformed posterior end. C) Five dpf old larvae with curved tail. D) Five dpf old larvae injected with only Cas9 mRNA.

Appendix 9.

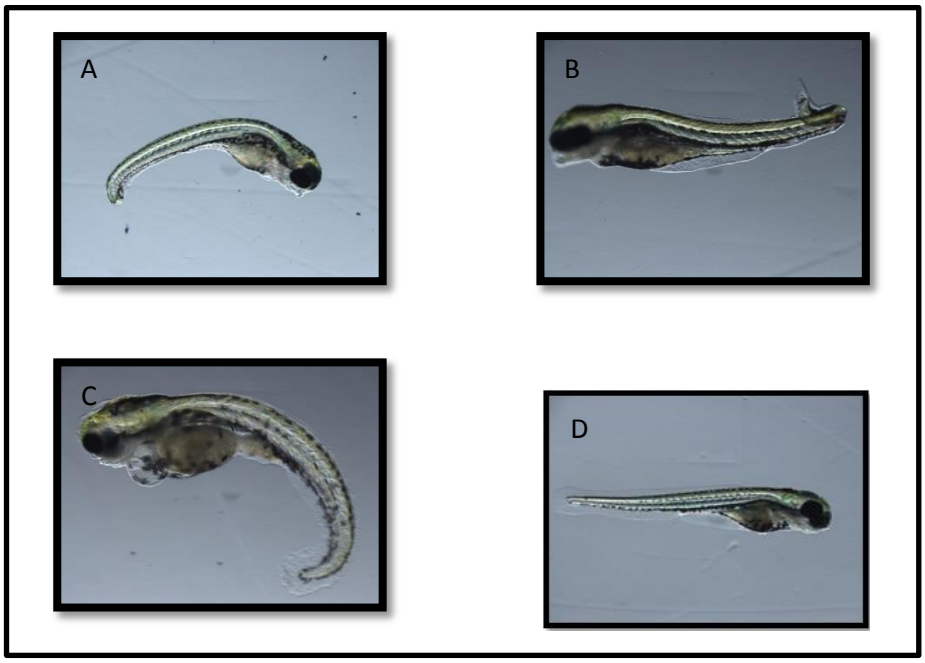


Figure: Unanalysed larvae developed from embryo injected with Cas9 mRNA and sgRNA. A) Six dpf old larvae with body bend at the anterior region. B) Six dpf old larvae with deformed head and tail. C) Six dpf old larvae with deformed head and tail. D) Six dpf old larvae injected with only Cas9 mRNA.

Appendix 10.

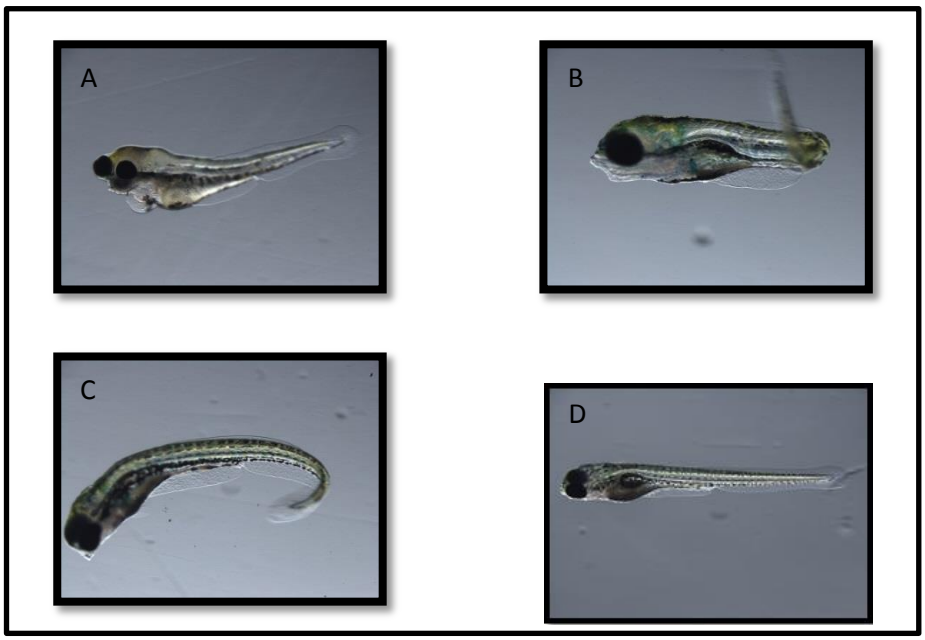


Figure: Unanalysed larvae developed from embryo injected with Cas9 mRNA and sgRNA. A) Seven dpf old larvae with deformed head. B) Seven dpf old larvae with deformed tail. C) Seven dpf old larvae with bend anterior end and curved tail. D) Seven dpf old larvae injected with only Cas9 mRNA.

Appendix 11.

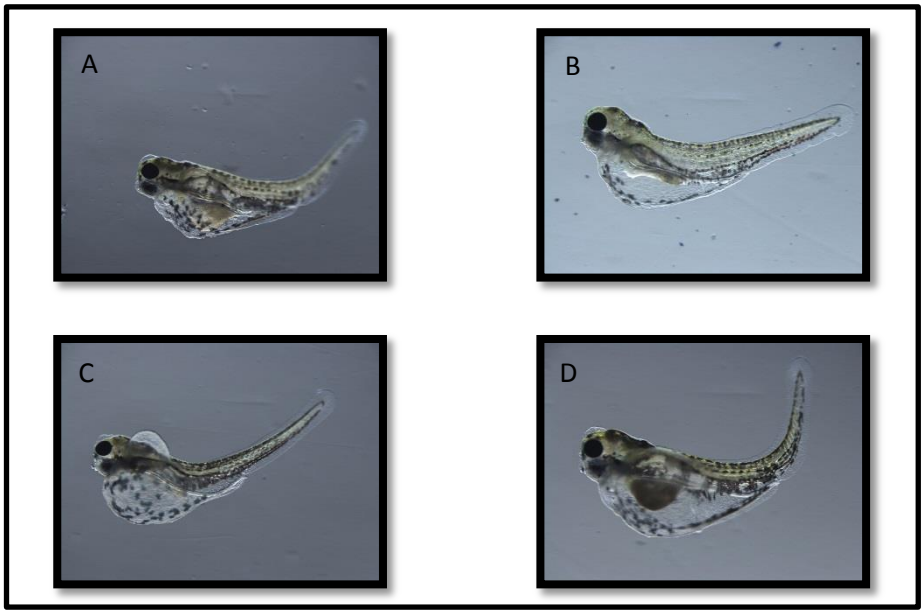


Figure: Common deformities (swollen stomach region) shown by the larvae developed from embryo injected with only Cas9 mRNA. A) 4DPF old larvae. B) 5DPF old larvae. C) 6DPF old larvae. D) 7 DPF old larvae.

Appendix 12.

CAGACAGGATGCTGGTCCGGTGAGGAGTTTCCACTGGGGGCGGGGCTTACACTCTTTTCAGGAAGTGTTGGGACAGAGGAAATTTTATTTACGTGG
TAAATAAAAAAGTTCCTCGTCTGCAACTCTTATAAACCAGGTGTGGTATGTTTGCACGTCCACATACTACATAACCTCCCAGAGTCTGCTTGTGACGG
*CAGAC***ATGGCTGCAGTTACTCTTGTGGACCTCACAAACATCTGTTCTTCAGAAAGCGACATTTGCGTGAAAAACCGTATCCGCACATTCTGTGCAG**
TGACCGAGAATGGCTCACAGGAGCCGGTGTACTTCGTGGCGGATCCCGGGAGCTCCTTTGGGACGATCAGATTTCCGAGAAGCTCACTGCTTGA
GATGTTTGTCGGTTTCAGCAATTCCCCACAGATCCTGTGTAAAAGCTCTCTTTCGGCTTCGTTATATTCGTTCAAACAGGCCATAGACGAAAGCGA
TATTAGTAACGTGAGGATAGTGACCTCGAGCCGCGGTAGAGGGGTGTTTCAGGTCTACCAGGAGCTGCTCTTCACTTCTGCTTATAGTTTTGAGT
ACTGCATCATGTTTGACAACCTGTCGTGTGACTCCTGTCCTCCGGGTGACCGAGCAAACAGCAGCACTCCTGCAGAAGGTGACGCCAGACTGGAG
CTCAGCGCGTTCCTTCAGCGCCTGCCAGCGGTCAAAGGAGATATTACAGTTCTCAGATCAGCTGCCATTTCAGGTTAGTTGTATTTGACTTAGTTAG
ATAATTTACATCAAAATGAAAATTTAATCACTGCTTTTTTTACCAGCAAGTATTACTGATTATGATTTTTTGTTTTGTTGAACACAAAGGTAGATATTTT
GAAGAATGTTAGAAACCGGTAACCTTTGTCTTTTATAGTAGGGAAAGCAAATGCTATGGGCATCAGTGGTTACTATATATATATATATATTTTTTT
TTTTTCCCTTTATTATTGCATTCCCCATTATTAACAAGCAAAGTGATAACTTTTAGCTTGAAGATAATTGCTTCTTCAACACATCATTCATTGTGAAAAA
CACCAACATGTGTTAAAACACTGGCTCAGCTTACCCCATAGTATGTGGCTCATTATGGCCCATAGTTGACAAAAACTAGTTTGCTTATTAGATCTTGA
TCTATTTGTTTAGATGTATCACCTTTTACTAATGACAGGAAAATTTTACGGATGGTGACCACTTCTCCCTTATGCTTCTCCTTTACATGGTCCAGCAAA
ATTAATGTATAACTCCAAAATGTATGATCACATGACCAAAAATAATATAATGGTTGCCTAGATACAGGGGGTGGGTCAACTTACCCACTGGCTCATC
TACCTCACTCACCCCCTCCCCCAATATGATCACACTAAGCTTAGTTTGTGTACTCCTTAAGTGAAACTAACCAACTAACTAATGAATTCATTTGACGGA
TGGAATAATTTAATTGGGACAACACATTTTTTTTTATTCTTATATGATCTTTGGTTTTGTAAAGCCAGAAAAAATAATGTAAAAAATATATTAAATAAT
AGTTTATGAAATGGGAAGTTGGTATAAGAAAAAGATCAAAAAAGGAAAAAAATAATTTATTGACCTAAATGTAGCTCATACTAAAGGGTATCCAAG
TTTTAGGAACAACAAATAATAACTTGACTTCTAGTTGATTATTTGGTATCAGAAATGTCTGTTATGAAAGGCAAAGGCCTCTAGATTACGCTTATTTT
ACCTAAATAAAATATAATCATGCCTTGATTTTTGATTATTTAATTAGGACAGTAAAGTCTGACTTTGCTTAGACAAAAGTCTTGTCACTTAACAGAAAT
AATGTACAGTATAGAATATAAAGTCATGGTGCAGTGGAAAAACAATTGATATTGCGTATGACTCAAATAACTTAGTCATCTGAAATGGCAAAGGGA
GTGTTCTTGCAGGACTCACAGAGTTCATCAGGATATGTTAGATTTGTCTTCAAGGCCTCCTCCTTCATCTTACCCCAGACATGCTCAATAATGTTCATG
TCTGGTGACTGGGCTGGCTAATTCTGGAGCACCTTGACCTTCTTTATTACATATACATATATTATACATGTTGTGTGTGTGTGTGTGTGTGTGCCT
AATTAGATGTGGATCTTGTCATTTTTATGGCCAATGCAATGTAAGATGATATTTAGATTGTAATGTACAGCAATGACTTATAAAGATACATATAAACA
TCTGTGGTCACTGTATCTTGTATATGATAGCCTCTATGTCTGAAAAGAATGAATTTCACTTAGGTATTTAAACTCCTATAAGGTATCACAAGTTTTATG
CACGTTAAAAATGAACATGTATAACAAAGCACTTTGTGTTTAATAATTACCTTGGCATTAGATTAAAGACACTAGTTAACAATGTTGGGGGTAATGC
ATTAAAAGTAATAATATTATTACTTTTATAAGTAACAAGTAATGCATTACTTAAAAAAATGAAGTAAAAATATTTTAGTTACTTTTTGAAAAAGAAAT
GTAATGCGAGTTACTATTTAGTTTAATTAATTAGCTTTAAAAAATTGCTAAAATAAAATTAGCATAGTCACATTAAATCACACACTCAATGAGAGAAT
GCCCTCCCCGGGGGAACAGACAGACAAAGATTTGTGCTATGGAAGTATTCTCATTATTTTTAACACATCGAAGAAAAGGAAAAGGGGTTTGTCTCA
ATAGACAGTGACTAATAAGGCAAAAACTACAAAGCAGACACATTGCTTCAAACTTTCATTACAGCGCATGTAGCTCCGGGGGTCAGGAAGTTCCCAG
AATATAACATTATATTGAATTCACAATATACTAAACAATATACTAAATTTGATATGTTTTTTTTTATGGTAAACGAAGGTTATACAGTTCATTGTTAAT
CGCAGAGCTAAAATTTAAATAATATTGCCTAAATAAATAAATAAATAAATAAACAAACAAACAAACAAATAAATAAAAACCTGCAAGTTCTGAAAGA
GATCAAGCCTCACCCAGGTAAAACAAAAAAGGAACGCAAAAGTAACTCAAAAGTAACATAACACAATACTTTCCATAAAAAAGTAACTTAGTAATG
CAACTAATTACCTTTCTGGGGAGTAATGCAATAATTAAACGCATTACTTTCAAAAGTAACTTTCCCCAACACTGCTGCTGCTGGAGTGTTTCAAATTC
ACTGGCTAAGAGAGTGACATGTGTCATCATGTTATAAATAAACGCCTCTACTCTTTCACAG**ACTGCTTTGGCCATGGCTTCAGCACCCGTACCGGAG**
GGATCTCCTACATCAGCACCATGAGCTCAATGAACCTCTTCTGCAACCCAAGGCGCAAAGACCCCAGAGCAGTGGTGGATGAAAATCGCCGGCG
GCTCGGCCTGCAGGCTGGATTTCATTCTCGTCAGTTAAACCTCATAAAGGTAATTTACCATAAGGAGCATACTCTATTGACCTTATTCTTAAGTAAAT
AACAAAAACAAAAACATCATAGCATCGAAAAAATTATTTTACAAAAAAACTTAAAGCATGTGCAGGAAAAGACCACTAGCCATAAAAGAGCGCACC
AGACTACTAGCCAGCGTCTCGCGTTTCTGTTGCACTGCTGTTTATTGTGTTTCATTTAAAGAAAACCGTGTCTAACTGTGTATTCTGTTGTTGCCTGTT
TTAGTGTGTTCACGTATCCGTCTGCTATTCATGTCAATCCATTTATCTGTTTCATTCATTTATGTTGTTTACAAGGCTCTGACGAAGCGCACGGGTCAG
CAGGTCACGTGACATACATTTTGCAACTCTTGTGTATAGTACATTAGGAACGAACGCCAGCCCTTGTACTTTTTTTCATAAATAGAGTAGGATAGTTC
CAAGTGCACAATTACGCACAATACTGTTGTTAAGACTTTAAAGTGTAGCTCTAGGGGTTGTTACATGCCTGATGTTAAATGTTTAGAAGTGGGAGTA
ACAGTGAAAAAAGTTAAAAAAAAAAAAAAAAAAAATGTTGGTTGGGGTGAGTGTGGGTCAGTCCCCTGCCTTCACAACAACACACCAAATTCAAGTT
AAATAATTTATTAACAAAATGTACAATAGTATAACAAAACTGGGAAGCAAAGATAGACTTCAGGAGGGGGAGAGGCCAAAACAAACAAAAAGTAC
ACTCTAGCTAGTTAGGGAGTAAAACTGTCTAAACTGACAAAGTAAAATTAAATCAACAACAGAAATGTACACTAACTCCCTAACTAAACATAAACAA
GAGAAAAAGGGTTTTAGAAATAAAATGGCACACACCTCCCTATGGCTTCCAAACACAAAGTTAGTTTTAAAAAAAAAAAAAAAAAAAGAGCTCAAAT
GGCTTCAGAGGCTTTAACAACAGATTTTAACTTCAACAAGTTAAACACAACAGCACCAATAATCTTTCTCAATAGAGCACCCAGGCAGGAGAGAGAC
CACACTCAGGAAGCAAGCTTCAATGTGGCTTGCAACTGGCAGGCTTTTAAGCTTGCCGGGCAACGAGTAGCAGGTGCAATCAATCAACTTCCTCTCC
TCCACACCTAGCTGAGAAACAGGTTGCAGGAACGGAGAGGGAGAGGGAGAGAGAGAAACCAAACACAATAATACATTCAAAACACCATGGAATG
TAACAGGGTTAAAGAGGTAACATTTTGATGTGACTTAATTTGTTGGGAGTTGGTGGCAAAATAGCAAACCGACAACCAAATATGACGAAAAAGGG
CATTTATTTACAAAAGTAAAAAAGAATTCAACAGCAAAAATAATATGTACAAGTAGAAGTAGCTCACAAAAAAACAAAAACATTAATAAATCAATTT
TAAGGGAAAGAGGATTTCAACCAGCGCCAAATAAATAACTGTTCATAAATAACTGTATAAGGGAGTCTGTCTGAGTCAGATTTCTAAATGTAAACG
CGGCCTCGCTGATTAATCTTTCACATTTCTGTGTAATGCTTCACGTATTGATCACACGTCTTTGTCATTGACTTTTAG**TGCAACCATGCCAGTGATGTG**
TGGGTGATGGGCAAACCTGCCCCGGACAGTTACGATGGACTGGTGACCAATCAATCAGACGTTGTCATTGCAGCACCAGGGGCTGATTGTATGC
CATTGCTCTTCACTGACCCTGTATCAAAGGTCATTGGAGTTGCGCATGCAGGTCAGTCATCACACTGTAAAATCTATTTTAACACTCTTACTATTTCC
TACCAGTGTGGCAACATGGATTTATGAATTGTGAATCCCTTGTTTTGGTCATTGATCAATTGAAATAATTTAAAATGAATAAATTGGGATTGGTGATT
TTCATTTAAAGGGATAGTTCACCCAAAAATGACTATTTAAAAAGTAACAAATAGTGAGTAAATTGTCAATTTCAGGTAAAAAAAAAAACTATCCCTTT
AATATTGCCTGATGGTATTTCTATATTGTCACACAGTTTTGGGTTAAACTCACTACTACTTAAATGCTTTCATTGCATTTTAGCAGTCATAAAGCTTATT
GCAAAAAAAGGATAACAAAGGAATAATATTGTCACAGTTGACTTTGTCCTGAAGGGTCAATTTACCTAAAAAATGACAATTCGGTCAGTGAAAAGT
TTTTAAAAAAAAATGTTGGTTGGGGTGAGTGTGGGTCAGTCCCCTGCCTTCACAACAACACACCAAATTCAAGTTAAATAATTTATTAACAAAATGT
ACAATAGTATAACAAAACTGGGAAGCAAAAATAGACTTCAGGAGGGGGAGAGGCCAAAACAAACAAAAAGTACACTCTAGCTAGTTAGGGAGTA

5′ UTR 199 bp

Exon 1 544 bp

Intron 1 2533 bp

Exon 2 179 bp

Intron 2 1576 bp

Exon 3 166 bp

AAACTGTCTAAACTGACAAAGTAAAATTAAATCAACAACAGAAATGTATACTAACTCCCTAACTAAACATAAACAAGAGAAAAAGGGTTTTAGAAAT
AAAATGGCACCCACCTCCCTACGGCTTCCAATCACAAAGTTAGTTTTCAAAAAAAAAAAAAAAAAAGAGCTCAAATGGCTTCAGAGGCTTTAACAACA
GATTTTAACTTCAACAAGTTAAACACAACAGCACCAATAATCTTTATCAATAGAGCACCCAGGCAGGAGAGAGACCACACTCAGGAAGCAAGCTTC
AATGTGGCTTCAATTATACTTCATTACTTGTTCATAATCCGTTAGAGATTCTTTTTGTTTTGTCAGACACAAAGGAAGATATTCTGTCATATAATGTTG
TAAAAAAAAAAAAACAGTCATCGACTTCCACAGTGGGTTTTTCCTATATGGATGTCAATGGCTGCTTTTTCACCCCAACATCCTTAAGAATATATTGT
TTGGTCACATTTTATTTTGATGGTCCGTTTTTTATTTAAGTTACATTGCATCTACATGCCAGCTAATTCTCATCAGATTATCAGTTGACTGTTAGGTTGG
GATTAGGGTTAGTGTTAGTTGACATGTACTTGCAGAGTTTCTAACAGTCAGTTAAATGTCTGTTGAAGGAGCAGTATCAACATATTAAGCAGTCTAC
TAATACTCAAATGCACCATCAAAATAAAGCGTTACCTCTCGTTTTGTGTTCAACAGATATACAAATATATATTTAAATTTAATTGAACTTAAATGAAAT
AAATTAAGGTTTATATTGAATTAATTTGCTTAAAATAAAAATTCTCATTTCAATCTCCTAGAGTCAAATATGACAACATTTATTAGTGTTTTTTATTATT
ATTTTCACAAATACTACAAAAAATGTATGCAAAAATACAAATACTGCAAAAAAAGATTTGTATGTATATATACATAAACTCACCAGCCACCAAAAGA
AATAGCAAAAACATCCTAAATCAGATCAAAAGCTTGAACAAAATGAATGAAAGATTATATTAAATAAAATTTTTTTAAGCAAAATAAAACTTTTATTC
TACTTACCTAAATACTGCAAATATATAATTATAATAAAAAAAATTATAAATCAAATAAAAAACTTAAAGTATTTTTCAAAAAAAAAAAAAGTCAAATA
TGACAATGACATTGTCTCAGTGTTTTTAGTATTATTTACACAAATGCTGCTCAAATATATAATTACATAATAGAAATAGAAAAGAAACCTAAACCAAA
TAAATTAAGATTGAGTGAAATTAATTTTCTTTAAACAAAATAAAACTTTTCTCAATCTTTTGGACACAAATATTACAACATTTTCATAGTGTTTTTTATT
TACTTTATTTTCCCACATACTGTTGAATTATAATATAAACTCCTTTCTTTACTATTTATTAATACATAATGCCTCTAAATTGGGCTGCACAATATATAGTT
TCAGCATCAATATAGCAATCCTTCGCAATAGTCACATTGCAGGATATATGTAATGTTAAGTTTGTGGTACAATAATTTGATTGTGTTTTTGATGCCTG
TGATGGTATAATGATTTTAAAAGCGTTCAGGCATATGAATTTGTGCCATTTGTGACTTTAACAGCATTTAATTTTATTGAATTATTTTTATCATTTAGTT
ACTGTACTTAAATACTGTTAGACTTGGGTTTGAGTAAAACATTTTATTTCAATTGTATCTTTTATTGTATTTAGAGATTGTAATGCATATGAAATATGA
AAATCCTCACAGAAATGTACTGCAAATAGCACAGACAAAATGTGTTAGGGAACCCCAAAAGATTAGATTTTATGGAGATGCTAATACATTCATTCAT
TCACTCATTTTCTTGTCAGCTTAGTCCCTTTATTAATCCATGGTTGCCACACCGGAATGAACCGCCAACTTATCCAGCACATTTTTACGCAGCGGATGC
CCTTCCAGCCGCAACCCATCTCTGGGAATCACACACACACATTCACACACACACTCATACCCTACAGACAATTTAGCCCTCCCAATTCACCTGTACTGC
ATGTCTTTGGACTGTGGGGTAAACCGGAGCACTCGGAGGAAACCTACGTGAACACAGGGAGAACATGCAAACTCCACACAGAAACACCAACTGAG
CCGAGGATCGAACCAGCGACCCAGCAACCATCTTGCTGTGAGGCAACAGCACTACCTACTGCGCCACTGCCTCGCCCTATTATTTATTTATTATTTTT
TGAGGGGTAGCTGTGCCCCAGTAGAGCTTTATGTCTAGCAACACCCTTGCTTAGAGCTGTTCCTCAACACATGTAAATACAGAAATGAATATGAAAT
ATGAAAATCCTCACAGAAATGTACTGCAAATAGCACAGACCACAATGAAAATATTTCACGGGACCCTAAAAATTCTATAGATTGTTTATTAGATTTTA
GGGAGATGCATTCCCACTGCAGAATCAGCCCAATAAATCTAACATTATAAATTATTTTCAAATAGAGATAGTAGGACATCTGGTGAAATTGTTTTCAT
AATCACAATAAATATCCAAATAAATAAAATTGCAATGTTAAATTTTTCCAATACCGAGCATCCCTACCTCCAAATACCAAGCCATCAAAGATGCTACA
GTATGCATATATAAGTTGTCTATCTTGCGTGTCACAG**GCTGGAAGGGGACCCTGATGGGAATCGCCATGGCAACAGTAAAAGCCATGGTGTCAGA**
GTTTGGCAGCCGCCCTGCAGACATTGTGGTTGTGATTGGGCCTTCTGTGGGACCCTGCTGCTTCACTCTGGAGCAGGACTCGGCCAGAGAGTTTC
ATTCCATACACCCGGACTGTGTGCAAAACATGGAGTCATCCGAGCCTTCTGTCGACATCAGGGTCGCTACTAGGTAAGAATCAGAGGTCAGCCAG
AGGAGTCCTGCTTTGACTTTAACTGTACACCGGAGACTGTTTTATTGATCTCAGTGTCCCTCAAGTAGGTCACGTCATTCCCTGAAGTGACGCACAGT
AATGAACACTTCGAAATGAAGCTGATTTACAGCATCTCTATCTAGAATATTTGCAGAAGCGGAAAAGACGGATCATTGCTCAAGGGAGAAATGGCT
ATTTAGATTTAAAGTGCTTTTTTCGTCTTGATTCTAGTCCAAATGTCTACAAATTCTAAAACCAAGCAGCATTTTCTAAATAAGCAAAACATATTATGT
TGTTTTCAGAAATAAATTGACAATAATAAGAGTGTTTTTTTTCTTTTTTTTTCTAAAAAGAGCTAAATTATCTGCCAATGGAGTAAGCAAAATAATCTT
GTTTTCAATTTTGCTAAGCCTATTGGCACATTATTTTGCCTGTGTTAAGGAAAAAAACTTCATTTCGATATTATTTTCTGACTTGTCTAGAAAATGCTTC
TTGATTTAATAATTTTTTGGATAATTGGACTTGAAACAAGACTTTGTTGCAGCAGTAATTTGCGAGACACAATAATTCAATATTAGGTCCTACATAAA
AACAATTTTTACATTGTAACAACTGATCAAAAAAGTCATAACAATTTGTTTTTTAACAGCCATTCTTACACTTATTTGCTGAGTGTGGATATTTTGCTG
AATCGTGGATAAATAAATATGACTCTGCCGTTTGTGTCAGCAAGTAACTTTCTAAATGAATAGTTCATTGAATCATTTATTCAAACTATTTAAAATTTT
ACTAAATAGGTCCAAAATCAGTTTTTCTTTTATGAGCAATATCATTATAATAATATGTAAAGATCATCTTTTATGAATATATTTTGTAATTTTTTAACTG
TAAATATATCAAAACTCAATTTTTGATTAGTAACATGCATTGCTAAGCATTAATTTAAACCATTTTAATGGTAAGTTTTTACAATATTCTGATTTTTTGA
AACGTCAATTTCCAGATCAAAAAACTGTAAATTATAAAGTTGAAGTCAGAATTATTAGACCTCCTGAATTATTTTCTGTTCAACGGAGAGAAGATTTT
TTTTCAACACATTTCTAAACATAATAGTTTTAATAACTCATTTCTAATAACTGATTTATTTTATCTTTACCATAATGACAGTAAATAATATTTTACTATAT
ATTTATTAAGACACTTCTATACAGCTCAAAGTGACATTTAAAGGATTAACTAGGTTAATTAGGTTAACTGGGCAGGTTAGTGTAATTAGGCAAGTCA
TTGTATAACGATGGTTTGTTCTGTAGACTATCGAAAATGTATATAGCTTAAAGGGGCTAATGATTTTGACTGTAAATCGAGTTTTAAAAAATTAAAAA
CTGCTTTTATTTTTAAACAAATAAGACTTTCTCTTGAAGACTAAATATTATTAGACATACTGTTAAAATTTCCTTGCTCTGTTAAACATCATTTGGGAAA
AATTTTTAAAAGAAAAAAAATTCAAAAGGGGGCTAATAATGCTGACTTCAACTGTGTATATATATATATATATATATATATATATATTACCATAT
ATGAGTTATATAAGACAAACTGAATTAAACATGGAAGTTTGAAGGAAAAAGACCAAAACAGTATTTCCCTGGTAATTAGTGCTATTTACAACTTGAA
ACTATTAATTAACACTTACCCATAAATACTTTTATTTAAATATTTGGCAAAGATTATTTGCGAATAATCGCATCCAAAATAAAAGTTTGTTTTGACATA
ACAGCAAATATATTTGTATGTGCTGTGTATATTTATTATATAGGATTCAGCTGGTAGGGCATCCAGCGTAAAACATGCTGGATAAGTTGGTGGTTCA
TTCCGCTGTGGTAACCCCTGATTAATAAATTCAAATCAATTCAATTTGTTTATTGTCATGAGCACAGTAAGGAAACACAATTCCCTGTACAATAAAAT
TCTTACTTTGCCGTCCACAGTGAAGTCATAAACACAACATACACATCATACACGACATACACATAAACACAGATATGTACAAAGAAGTGACAGAAGT
ATAATTATAAATTGTAAATATATAAATATCTATGAATGTAAACAATATAGAATACTGATTGTAGAGTAATGAAAACAGACGTAAATAACGTGTAAAG
GGACTGAGCCGAATGATAATGCATGAATGAATATACACAATACACACAGGCATGCTTATATTTGGGGAACAAATATATTTAGATATAAAATATATAT
GAATATGGTACCATTTTTTATACATTTAACAAAATAGTATTACATATCCATTGCAAATAAAAAATAATACACACAAATATATTGTAAAGACAAACTTTT
ATTTTGGCTGCAATTTAATCTTTGACAAGCACTAATAAATATTTGTATTTATTATTTGTACAGTTAAAAGAGTCAACAATGCACATAAACTTTCTAGAT
CTATATAACCTCATCATTTTTATTGTTTACATTATTTTACACTTATTTAAATGTCCTTCTGTGTTCTTATTGTTAGGTATGTATGTATGTATGTATGTATG
TATGTATGTACCCTGAGCTTCTGAAAAAACCCAAAAAATGCCCTGTGTGTTTGCGCACGCTCAAACCTTATTCTGATTCTGAAAAGTAAACTAAAATA
AGTTTGCAGTTTTTATGTACATTATATATATTATTAAATATATTTGTATTGTTTTACTGTATTTCTTTTAAATGTCCTATAAAATTTATCCTAATTTTATTA
zGTATTACTATTATTAAAAGCCTATTCTTATAAGACATGATTGTTGAACCGCAACATGTCAACCAACCAATCATGCTTTTGTCTTGTAG**AGTTCTGCTG**
GAAAGAGGTGGGATTAAACCTGAGCACATAGAGAACATCCGAATCCCAAACCAGACGGATTCTATACCATGCACATCCTGCAGTCCTGAGCTAT
TCTTCTCCCATAGGAGAGACGGACTCAACTTTGGGACACAGATTGGGTTCCTGTGGATCAAGCCCTGA*CCGAATGCACAAAAGCTTTAGATATAAT*
TGAAGTCAGAATTAATATTAGCCCTCCTGTGATTTTTGTTTTCTTCTTCTTCTTTTTTTTCCCAAATGAAGTTTCACAGAGAAAGGAATTTTCACAGTAT
GTCTGATAATATTTTTTCTTCTGGAAAAGGTCTAATTTGTTTTATTTCTGTTAGAATAAAAGCAGTTTTTATAAAAACCATTTTAAGGTCAATATTATTA

Intron 3
3318 bp

Exon 4
226 bp

Intron 4
2872 bp

Exon 5
172 bp

GCCTGTTTAAGCTATATATTTTTGGATTGTCTACAAAACAAACCAGTGTAATAAAATTACTTGCCTAATTACCCTAACTTGCCTTATAAATCTAGTTAA
GCCTTTAAATGTCACTTTAAGCTGTATAGAAATGTCTTAAAAACATACATGGGGCTAATAATTCTGACTTCAATTGTAGTACTTCAAATGACAGTACAT
TCATTTATTTATATAGTTAACTAATTAGTATTGAATTACCCTACTCGCAACATGTGACACTTTTTTTTTACTTTTTTAGCTTACAAAATAGACTTTTACAG
TATCTTATATCAAAGTCTAATGTATGTAGATGTAAGATAGTCATCATTTTTTTTATCAGCACTTTTAACAAACTAACATGTATATAATATTACCTATGTA
AATAGTGAGAAAACATTTTTAATAAGTATAGATATTTCATATAATTTTTGCAGCATTATTCAACAAAAAAGGAAACTAAACAGGCTTGGAACAAGCG
AAGAGTGAGTAAATGATTATTTTTTTAGGGTGAACTATCCCTTTTAAAACTTACTTGAAAGCTTTTTTGTTGTTTGCATTTTGTTTGTCTAGATTTTCAA
GAGATTGTTTATTAGATTTTGCCAGTCCTTAGGATAATGGTGACACTGTTGTTGACATTTAATAAATAAGTTTTTATAAGCCCATTTACAGCTATGGAA
ATAAAATTTCTCTGGATTTACTGAAATTAGTAGGTATGTGTGTAATTGTTTTTTTATTCTGTACAAATTTAAAAAATTATAATATTGTTCTACTATATTG
TTTTGAGTTCAATATTGTTCAAAGTTCTACTATATTGTTCTAAGATAAAATGTTGATATTGGGTTCTCTACAAAATATAAATATAAACATATCAGTAAA
CATGGCATCTTTTATTTGTCATTTTCCACATTTGTCAAAAGTCAAGACTTTTATTTCCAATTTAAAAGAAATAATATCGGCAATTAAAAAAAAAAACCC
TAAATTACACTTTACTCAAAGGCATAACTGTGAATTGTAAATCCAGTGAAATGACATATATCTTAATTTGCATGTTAAATTTGGAAAACTAAGGGCTA
GTTGTATATTGCAAGTCAATTTTATTTTTTTATGTTAATTTGTTGAAAAATAAATTAAACATTAAATTAATATTTATTATTATATTTATATATAAATTTTA
CATTTTTGAACAAAATCCTGCCAAATTATAAAACTATGATGGTATTAAAGGACTGGTCTATTATTAAAGGACAGTAAAATTAAATGTATAATGGAAG
CTAAGCTACTAAAATAAATGTCAATAATAATCTAATTCAGAAATAGGCTGTTTCCCAGTGTTGGGTTGCAGCTGGAAGGGCATCCGGTGCATAAAAC
TTATGCTGGATAAGTTGGCGGTTCATTCTGCTGTGGCAATCCTTGATTAATAAAGGAACTAAGTCGAAAAGAAAATGAAGGAATGAGAAATAGCCT
GTTTATATCTTGCTGGTTTAAAAAAAAATATGGCTGTACTGTAACATGATGAATGTACACTAGAGGGAGATAATAGATTTATTTGTTGTCTCTACGTT
TATTTTTTTTTATAGTCTATGGATCTTTGTATTGTTTGACCACTATATTTATGATTTTTAAATCTGACTGTTTATTAATTAAAACTGCACAAAATATATCA
CATACAAACCTTCTTCAATGTCATTTCAAACCTAGGCTACAGTGAATTGACAGTTGTTTAAAAATTGTGTGCATTTAAATGATGTATAGTAAAGTTAAT
CATACATTACATGTCATTCCACTCAGGGGCCAATTTTTAAATGTGGGCCGGTCAGTTTTTTTTAAAATATGTATTGTTTTTAAGTGCAGACAGCTTTTA
TCTCTTGCTAGATGTGATATTATGATGATGATGATATTAATAATAATAATAAATGTTAAGCATTTGGCCCAATCGGCAACTGCTGCCTGGTTTCAAGA
TGGGCCGAACCAATCTGAGGTTACCCATACATAATCAAAATCTGTCAAGAATGTCATATTATTTCACCATCTGTAGACTAAATAGTGAATGTGCGTGT
CTGTGGGTGGGGCTTTTTCGGTGTGGTAATGTGGAC

3' UTR
2339 bp

LACC1 gene sequence of zebrafish (14,125 bp) with 5' UTR, exons, introns and 3' UTR regions. The underline bold sequences are the 5 exonic regions. The italicized sequences are the 5' and 3' untranslated regions**.** The sequences with the arrow above them are the primer binding site.

Appendix 13.

ATGAGTAAGCTGATTGTCCCGCAGTGGCCGCAGCCAAAAGGTGTTGCGGCCTGTAGCTCCACTCGTATCGGCGGCGTGAGCTTGCCCCCGTATGAC
TCACTCAACCTCGGTGCCCATTGTGGCGATAACCCGGATCACGTTGAGGAGAATCGCAAGCGACTTTTTGCTGCGGGCAATTTGCCTTCTAAACCGG
TCTGGCTTGAGCAGGTACACGGCAAAGATGTGCTTAAGCTCACTGGCGAACCTTATGCCTCAAAACGGGCGGATGCCTCTTATAGCAATACGCCCG
GCACGGTTTGCGCAGTGATGACTGCCGACTGCCTCCCTGTGCTGTTTTGCAATCGAGCGGGAACGGAAGTCGCCGCCGCTCATGCTGGCTGGCGTG
GACTGTGCGCAGGCGTGCTGGAAGAGACGGTTTCCTGTTTTGCTGATAATCCGGAAAATATTCTCGCCTGGTTAGGGCCGGCAATTGGTCCACGCG
CGTTCGAAGTGGGGGGGGAGGTTCGCGAGGCGTTTATGGCAGTAGACGCTAAAGCAAGTGCAGCTTTCATTCAGCATGGTGATAAGTATCTGGCG
GATATTTATCAGCTTGCCCGGCAGCGTCTGGCGAACGTGGGTGTTGAGCAAATTTTCGGCGGCGACCGTTGTACATATACGGAAAATGAGACTTTC
TTCTCTTATCGTCGCGACAAGACCACCGGTCGTATGGCAAGTTTCATTTGGCTGATATAA

YfiH gene sequence of *Escherichia coli.*

Appendix 14.

Table showing the list of the organism, mRNA and gene name along with the NCBI accession number for the sequences, used for generating the phylogenetic tree.

S.No	**Species name**	**mRNA**	**NCBI accession number**
1.	*Sinocyclocheilus rhinocerous*	*LACC1*	XM_016563906.1
2.	*Sinocyclocheilus grahami*	*LACC1*	XM_016233778.1
3.	*Sinocyclocheilus anshuiensis*	*LACC1*	XM_016450847.1
4.	*Cyprinus carpio*	*LACC1*	XM_019069483.1
5.	*Homo sapiens*	*LACC1*	XM_006719766.3
6.	*Danio rerio*	*LACC1*	XM_001919386.6
7.	*Rattus norvegicus*	*LACC1*	XM_001072262.4
S.No	**Organism**	**Gene**	**NCBI accession number**
1.	*Mycobacterium tuberculosis*	*yfiH*	NC_000962.3
2.	*Escherichia coli*	*yfiH*	NC_000913.3
3.	*Corynebacterium glutamicum*	*yfiH*	NC_003450.3

Table showing the list of the organism, protein name along with the NCBI accession number for the sequences, used for generating the phylogenetic tree.

S.No	**Species name**	**Protein**	**NCBI accession number**
1.	*Sinocyclocheilus rhinocerous*	*LACC1*	XP_016419392.1
2.	*Sinocyclocheilus graham*	*LACC1*	XP_016089264.1
3.	*Sinocyclocheilus anshuiensis*	*LACC1*	XP_016341886.1
4.	*Cyprinus carpio*	*LACC1*	XP_018925028.1
5.	*Homo sapiens*	*LACC1*	NP_001121775.1
6.	*Danio rerio*	*LACC1*	XP_001919421.1
7.	*Rattus norvegicus*	*LACC1*	XP_233749.1
8.	*Mycobacterium tuberculosis*	*yfiH*	AHJ55362.1
9.	*Escherichia coli*	*yfiH*	AHY71904.1
10.	*Corynebacterium glutamicum*	*yfiH*	CAA70159.1