

final_paper

Ahmed Al-Tohamy

12/8/2021

Mapping of global transcriptional regulator MftR binding loci in *Burkholderia thailandensis* using ChIP-seq

Introduction

Bacterial diseases infect individuals throughout the history of mankind with fatal consequences. Finding an efficient treatment against infectious bacterial diseases is a challenging endeavor. Ancient civilizations used antimicrobial agents to overcome the spread of such maladies [1]. It was reported that in ancient Egypt and ancient Greek, certain types of fungi were used to cure infections [2]. In the late 1880, a Nobel prize laureate Paul Ehrlich developed the first dye based synthetic antibiotic which was called later arsphenamine [3]. In 1928, Sir Alexander Fleming discovered the first effective antibiotic which he called the fungal extract which is known now as penicillin [4].

Since the discovery of penicillin, more than 100 antibiotics were discovered, each has a certain mechanism of action [5]. Different antibiotics target certain vital processes that interfere with the bacterial cellular functions. Antibiotic molecular targets include interfering with protein synthesis, cell wall construction, DNA synthesis and folic acid synthesis [6]. However, bacteria develop defensive mechanisms to resist antibiotics and survive. The misuse of antibiotics worldwide accelerated the evolution of bacterial resistance which has been clearly observed in the last couple of decades. In 2015, it was reported that more than 33 thousand persons died from antibiotic resistance bacteria [7].

One defense mechanism that bacteria use to fight antibiotics and other toxins is efflux pumps. Efflux pumps are transmembrane proteins that are localized in the cytoplasmic membranes in all bacterial cells. Some efflux pumps are specific to one antibiotic while others are multidrug efflux pumps [8]. There are five main families of efflux pumps which are ATP binding cassette family, small multidrug resistance family, multidrug and toxin extrusion family, major facilitator superfamily and resistance nodulation cell division family [9]. In *Burkholderia thailandensis*, a major facilitator transport protein (MFTP) efflux pump was discovered which is regulated by MftR global transcription factor [10]. MftR is a global regulator of secondary metabolite in *Burkholderia thailandensis* [11]. Also, MftR was reported to regulate approximately 400 genes in *Burkholderia thailandensis* [11].

In the current study we aim to study the binding sites of global transcription regulator MftR in *Burkholderia thailandensis*. In order to do that we are going to use ChIP-seq technique. Chromatin immunoprecipitation (ChIP) is a technique which is used to study protein-nucleic acid interaction in a specific matter. In ChIP, first protein of interest is crosslinked with nucleic acid which is followed by sonication to fragment nucleic acid and finally a specific antibody is used to enrich nucleic acid binding fragments [12].

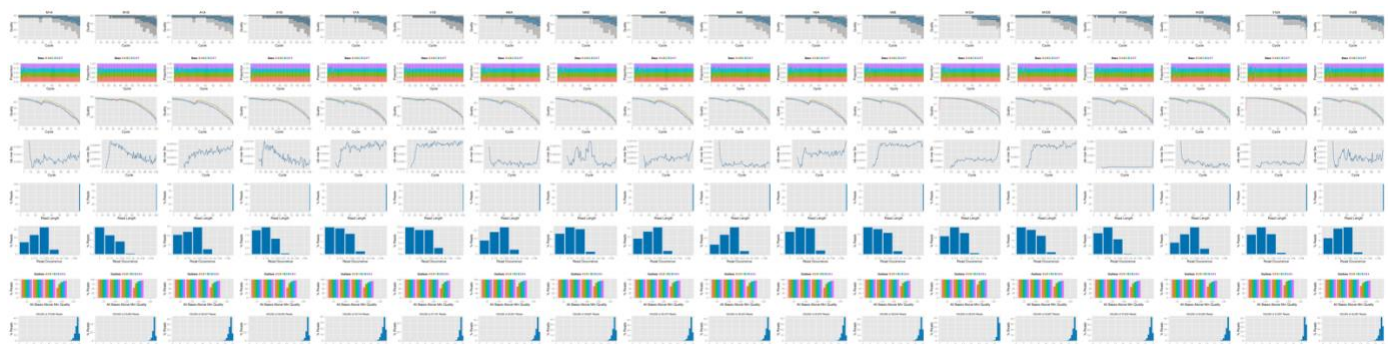
ChIP can be coupled with PCR, qPCR or Next-generation sequencing. In the current study, we are going to couple ChIP with illumina novaseq 6000 next generation sequencing. By using ChIP-seq, we should be able to identify genes under direct regulation of MftR. Methods and Materials Currently I am still working on the wet lab experiments. I will use 3 samples of wild type *Burkholderia thailandensis* and 3 samples of disrupted *mftR* strain. I have searched through NCBI Gene Expression Omnibus (GEO) database and found quite similar dataset under GEO accession (GSE61334). They performed ChIP-seq for the same bacterial species with different strain. Link to download the data:
<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61334>

- For this course, I have adapted R pipeline from `systemPipeRdata` package that can be downloaded from <https://github.com/tgirke/systemPipeRdata/blob/master/vignettes/systemPipeChIPseq.Rmd> with minor modifications
- The R code can be accessed through `RChIP_seq_code`

Results The typical pipeline for ChIP-seq analysis included standard steps. First of all, quality control is performed on fastq file which are generated from sequencing machine. After that, over expressed sequences are trimmed from both ends. Then *Burkholderia cenocepacia* reference genome was downloaded from <https://www.ncbi.nlm.nih.gov/genome/?term=Burkholderia+cenocepacia>. then fastq files were aligned to the reference genome. To identify protein-DNA, peak calling was applied to the aligned reads. Finally, Motif Enrichment Analysis was performed. FASTQ quality report which will be used later to filter DNA sequences.

```
``{r fastq_report, eval=FALSE}
fqlist <- seeFastq(fastq=infile1(trim), batchsize=100000, klength=8)
pdf("./results/fastqReport.pdf", height=18, width=4*length(fqlist))
seeFastqPlot(fqlist)
dev.off()
``

```



“FASTQ quality control metrics”

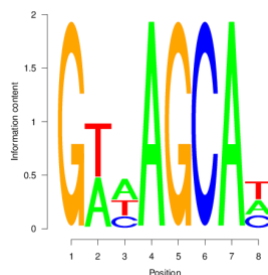
Relative to wildtype *B. cenocepacia* K56-2, the only genomic region that was found to be enriched by the ChIP-seq analysis of the CzcR-FLAG strain mapped to nucleotide coordinates 787809-798988 on chromosome 2.

```
## bcf
```

```
args <- loadWorkflow(targets = './results/targets_report_bcf.txt',  
                    wf_file = "combine.cwl", input_file = "varseq.yml", dir_path = dir_path)  
  
args <- renderWF(args, inputvars = c(FileName1 = "_FILE1_", SampleName =  
'_SampleName_'))  
  
varlist <- sapply(names(subsetWF(args[1:2], slot='output', subset=1, index=1)),  
                 function(x) as.character(read.delim(subsetWF(args[1:2], slot='output', subset=1,  
index=1)[x])$VARID))  
  
vennset_bcf <- overLapper(varlist, type="vennsets")  
  
pdf("./results/vennplot_var.pdf")  
  
vennPlot(list(vennset_gatk, vennset_bcf), mymain="", mysub="GATK: red; BCFtools: blue",  
colmode=2, ccol=c("red", "blue"))  
  
dev.off()  
``````  


```

Graphical representation of the sequence conservation of transcription factor binding site



## Discussion:

The aim of this study is to identify genes that are under direct control of MftR transcription factor in *Burkholderia thailandensis*. I have tested 4 genes that were differentially expressed in disrupted MftR strain using ChIP-PCR and I identified 2 genes under direct control of MftR “unpublished data”. One gene is called btaR2 which is an ATP-dependent transcription regulator that regulate quorum sensing. The other gene is called BhcM which is a transcription factor which regulates Histone deacetylase inhibitors. This preliminary result gave us confidence to proceed with ChIP-seq experiment.

The data we used is for CzcR transcription factor which is essential for heavy metal resistance and virulence in *Burkholderia cenocepacia*. From the results of the ChIP-seq analysis, the only binding site of CzcR transcription factor mapped to nucleotide coordinates 787809-798988 on chromosome 2. That means only one gene is under direct control of CzcR but there might be other genes that are under indirect control of CzcR.

## References

- [1] Forrest, R.D., 1982. Early history of wound treatment. *Journal of the Royal Society of Medicine*, 75(3), p.198.
- [2] Kambey, J.L., Farrell, A.P. and Bendell-Young, L.I., 2001. Influence of illegal gold mining on mercury levels in fish of North Sulawesi's Minahasa Peninsula, (Indonesia). *Environmental Pollution*, 114(3), pp.299-302.
- [3] Bosch, F. and Rosich, L., 2008. The contributions of Paul Ehrlich to pharmacology: a tribute on the occasion of the centenary of his Nobel Prize. *Pharmacology*, 82(3), pp.171-179.
- [4] Colebrook, L., 1956. Alexander Fleming, 1881-1955.
- [5] Khardori, N., Stevaux, C. and Ripley, K., 2020. Antibiotics: from the beginning to the future: part 1. *The Indian Journal of Pediatrics*, 87(1), pp.39-42.
- [6] Bollenbach, T., 2015. Antimicrobial interactions: mechanisms and implications for drug discovery and resistance evolution. *Current opinion in microbiology*, 27, pp.1-9.
- [7] Robertson, J., Vlahović-Palčevski, V., Iwamoto, K., Högberg, L.D., Godman, B., Monnet, D.L., Garner, S., Weist, K., Strauss, R., Vandael, E. and Ivanov, I.N., 2021. Variations in the consumption of antimicrobial medicines in the European region, 2014–2018: Findings and implications from ESAC-Net and WHO Europe. *Frontiers in pharmacology*, 12, p.727.
- [8] Bay, D.C. and Turner, R.J., 2016. Small multidrug resistance efflux pumps. In *Efflux-mediated antimicrobial resistance in bacteria* (pp. 45-71). Adis, Cham.

- [9] Sharma, A., Gupta, V.K. and Pathania, R., 2019. Efflux pump inhibitors for bacterial pathogens: From bench to bedside. *The Indian journal of medical research*, 149(2), p.129.
- [10] Grove, A., 2010. Urate-responsive MarR homologs from *Burkholderia*. *Molecular biosystems*, 6(11), pp.2133-2142.
- [11] Gupta, A., Bedre, R., Thapa, S.S., Sabrin, A., Wang, G., Dassanayake, M. and Grove, A., 2017. Global awakening of cryptic biosynthetic gene clusters in *Burkholderia thailandensis*. *ACS chemical biology*, 12(12), pp.3012-3021. [12] Mardis, E.R., 2007. ChIP-seq: welcome to the new frontier. *Nature methods*, 4(8), pp.613-614.