Characterization of yjdC, a Putative Antimutator Gene from Escherichia coli

Lucille Rich, Jess Moll, Brittany Niccum and Patricia L. Foster
Department of Biology, Indiana University, Bloomington, Indiana

Introduction
Mutations provide vital genetic variance that evolution can act upon, and variations in the mutation rate of an organism can result.

A previous study looked at mutations across the genome in proofreading deficient E. coli using a mutation accumulation experiment. 50 identical lines were derived from a founder strain. The lines were single colony purified every day until approximately 150 generations were obtained. At the end of the mutation accumulation experiment the genomes were purified and sequenced.

High mutation rates were expected due to the proofreading deficiency, but three lines exhibited a much lower rate. From these lines, mutations were analyzed to find a cause for the lowered mutation rates. Once a mutation in a gene was found as a possible cause the gene was deleted in the proofreading deficient strain background to confirm its antimutator phenotype.

Research Goal
To characterize genes that lower the mutation rate of E. coli.

Antimutator Lines
The mutation rates of lines from the MA experiment using the strains M163, M165, and M173 were calculated by dividing base pair substitutions by number of generations:

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Phenotype</th>
<th>Mutation Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>M163</td>
<td>mutD5</td>
<td>Proofreading deficient</td>
<td>9.4x10^{-5}</td>
</tr>
<tr>
<td>M165</td>
<td>ΔmutD mutD5</td>
<td>Proofreading &amp; Mismatch repair</td>
<td>1.7x10^{-5}</td>
</tr>
<tr>
<td>M173</td>
<td>ΔyjdC ΔmutD5</td>
<td>SOS off &amp; Proofreading deficient</td>
<td>6.4x10^{-5}</td>
</tr>
</tbody>
</table>

From here three lines with a lowered mutation rates, M163-51, M173-37, and M573-40, were chosen for investigation. Fluctuation assay results showed that their mutation rates were approximately 200 fold lower than their respective founders.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Estimated difference from average</th>
<th>Experimental difference from average</th>
</tr>
</thead>
<tbody>
<tr>
<td>M163-51</td>
<td>3.27x</td>
<td>200x</td>
</tr>
<tr>
<td>M173-37</td>
<td>3.27x</td>
<td>200x</td>
</tr>
<tr>
<td>M573-40</td>
<td>3.72x</td>
<td>200x</td>
</tr>
</tbody>
</table>

From the lines several mutations were identified as putative antimutator genes, genes that when mutated caused a lowering of the mutation rate. Each putative antimutator gene was researched, candidate antimutator genes were deleted in proofreading deficient strain backgrounds and mutation rates were estimated by using fluctuation assays. The gene yjdC was identified as an antimutator gene for further analysis.

Blot results identified YjdC as a putative TetR like transcriptional repressor.

Deletion of yjdC Lowers Mutation Rate in Proofreading Deficient Strain

To check for an antimutator phenotype in a proofreading deficient strain carrying the allele mutD5, yjdC was deleted. A fluctuation assay was used to estimate the mutation rate in 5 isolates carrying both a deletion of yjdC and mutD5. The results were compared to a strain carrying only the mutD5 allele.

- When yjdC is deleted in a proofreading deficient strain background, the mutation rate was reduced by an average of 5.7 fold in all five isolates.

Deletion of yjdC Lowers Mutation Rate in Mismatch Repair Deficient Strain

- We deleted yjdC in a strain with mismatch repair deleted. This was done to determine if YjdC is an antimutator in another highly mutation strain.
- The yjdC deletion showed a decrease in the mutation rate of 1.48 fold.
- These results indicate that YjdC lowers the mutation rate in multiple DNA repair deficient strains.

Deletion of yjdC Lowers Mutation Rate in Wild-Type

- We deleted yjdC in a wild-type background to determine if the mutation rate of wild-type E. coli could be lowered by a deletion of this gene.
- In a wild type strain the mutation rate was decreased by 1.62 fold.
- These results show that a deletion of yjdC has an antimutator phenotype in multiple strains with different mutation rates indicating a broad mechanism of lowering the mutation rate.

Restoring yjdC Increases the Mutation Rate in Antimutator Line

- Wild-type yjdC replaced the mutant yjdC in the M173-40 antimutator line.
- One isolate showed a significant 2 fold increase in the mutation rate of M173-40 yjdC.
- Since the mutation rate did not increase to the founder strain’s mutation rate, we can assume that there are multiple genes that affect the mutation rate.

Conclusions and Future Directions
Not only does a deletion of yjdC lower the mutation rate in multiple DNA repair deficient strains, but it has a broad mechanism of lowering the mutation rate in E. coli. However, when wild-type yjdC was replaced the mutant yjdC in the M173-40 antimutator line, the mutation rate did not increase to the expected mutation rate of a proofreading deficient strain. Therefore we can assume that multiple genes are acting together as antimutators in M173-40, and not yjdC alone. RNA-seq data shows that yjdC represses 33 genes, and at least one of these might be a factor in lowering the mutation rate.

To further investigate yjdC’s affect on mutation rate, strains with deletions of genes repressed by yjdC will be compared to mutD5 in fluctuation assays, to analyze their mutation rates.

Acknowledgements
Ellen Popperl, Matthew Gruenberger, and Jose Tovar
This work was supported by NIH Training Grant 5T32 GM007757 to P.L.F.
Undergraduate Research Mentorship Program (Bo-L.H. & C.A. Morin)

ΔyjdC Represses 33 Genes

- RNA-seq was performed on a strain carrying mutD5 and a strain carrying ΔyjdC mutD5. RNA was isolated from two identical samples of each strain at the log phase growth. RNA was then sequenced. The resulting sequences were mapped to wild-type E. coli MGL655. After mapping the expression levels were calculated.
- After expression levels were calculated, DEseq was used to determine the differential expression between the two strains.
- The heat map to the left is a representation of 42 genes that are at least 1.5 log fold higher in one of the conditions with a p-value < 0.05.
- Each row shows the normalized expression as a heat map ordered by the log fold differential expression. Each gene expression is represented by the Z-score (observed transcripts per million (TPM) minus the row mean TPM, divided by the standard deviation of the rows (TPM)).
- The legend for the 2-scores is represented at the bottom of the box.
- Because yjdC is a putative transcriptional repressor, most of the genes that are differentially expressed are expressed at a higher level in the strain carrying ΔyjdC mutD5.
- These results will be further analyzed to determine if any of the genes regulated by YjdC show mutator phenotypes when deleted.

Z-score

The results were compared to the conditions with a p-value < 0.05.

ΔyjdC elicits a decrease in the expression of genes that may be involved in the repair of DNA.

Z-score

The Z-score is calculated as follows:

\[ Z = \frac{\text{TPM}_{\text{observed}} - \text{TPM}_{\text{mean}}}{\text{TPM}_{\text{std deviation}}} \]

where TPM_{observed} is the observed TPM, TPM_{mean} is the mean TPM for a particular gene, and TPM_{std deviation} is the standard deviation of the TPM for all genes.

The Z-score is a measure of how many standard deviations a given gene's expression level is from the mean expression level of all genes.

For example, if a gene has a Z-score of 2, it means that its expression level is 2 standard deviations above the mean expression level of all genes.

Z-score

The Z-score is a measure of how many standard deviations a given gene's expression level is from the mean expression level of all genes.

For example, if a gene has a Z-score of 2, it means that its expression level is 2 standard deviations above the mean expression level of all genes.

Z-score

The Z-score is calculated as follows:

\[ Z = \frac{\text{TPM}_{\text{observed}} - \text{TPM}_{\text{mean}}}{\text{TPM}_{\text{std deviation}}} \]

where TPM_{observed} is the observed TPM, TPM_{mean} is the mean TPM for a particular gene, and TPM_{std deviation} is the standard deviation of the TPM for all genes.

The Z-score is a measure of how many standard deviations a given gene's expression level is from the mean expression level of all genes.

For example, if a gene has a Z-score of 2, it means that its expression level is 2 standard deviations above the mean expression level of all genes.