Genetic repair and phenotypes after CRISPR/Cas9 targeted nuclease activity in the ADE2 gene of *Saccharomyces cerevisiae*

Isaac Roberts, March 18, 2023

METHODS

Genome Exploration and HDR Design

Our research began by examining the ADE2 gene of *Saccharomyces cerevisiae* and exploring its promoter region, coding region, and functional domains. This preliminary exploration served as a basis for our design of guide RNA and HDR template sequences for CRISPR-Cas9 targeted gene editing. The gRNA sequence we designed targets a region at the beginning of the ORF of the ADE2 gene. The HDR template we designed spans 50 base pairs upstream and downstream of our gRNA target region. With the goal of answering our research questions, we designed two HDR templates: HDR A which introduces six point mutations with various intended effects and HDR B which introduces the same six mutations as well as mutating the PAM sequence adjacent to the intended Cas9 cut site.

Yeast Transformation

In order to prepare for the transformation of our yeast colonies, we first ligated our gRNA sequence with plasmid PML104 which expresses Cas9 and contains a uracil selection marker for yeast. We transformed bacteria with this plasmid and allowed bacterial colonies to grow, allowing us to acquire many copies of our desired plasmid. We then extracted our recombinant plasmid from the bacteria and set up several types of transformations of our yeast. As a positive control, we transformed yeast cells with just pML104 to confer the uracil selectivity marker since we would be growing our yeast colonies on uracil-deficient media. Our negative control was yeast cell transformation with water. Experimental transformations for our project included a transformation with recombinant plasmid and separate transformations with the recombinant plasmid and our HDR templates A and B present.

Colony Data Collection

After an incubation period, we counted the colonies by phenotype for each of the five transformation types. Yeast genomic DNA from multiple experimental replicates of each transformation type were extracted and amplified with PCR, and then sequenced. The results of sequencing were subsequently analyzed with the intent of answering our research questions.

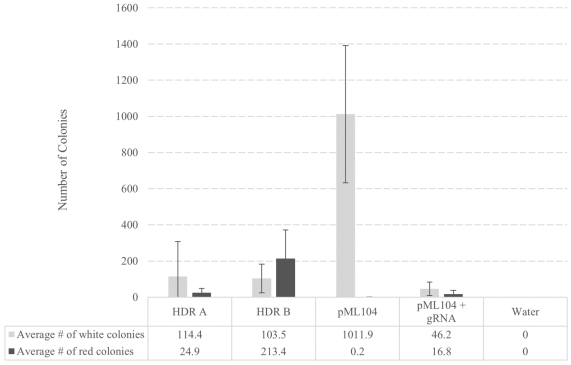
Bioinformatics Analysis of Colony Sequences

Sequencing analysis was done using trivial bioinformatics tools to classify sequences by genetic repair type. The primary tool used was alignment with affine gap penalties. Each sequence that was statistically significant was aligned to the known wild-type ADE2 gene sequence and repair type was determined by analyzing the indel frequency near the cut site and by assessing incorporation of mutations from our HDR templates. Sequences with insertions and/or deletions around the cut site were classified as having undergone NHEJ repair, whilst sequences that had incorporated mutations from our HDR templates and had no indels around the cut site were classified as having undergone HDR. Sequences with no mutations were classified as having not been cut by Cas9. In addition to classifying by repair, the number of HDR template designed mutations that were incorporated into the alignment was counted for each sequence.

RESULTS

Colony Counts of Transformation Types

The counts of the number of colonies of each type of transformation from our experiments revealed several preliminary trends in answering our research questions. Principally, we discovered that there was very high variance of colony counts even within the same transformation type for each of the experimental replicates. This is evidenced by the large error bars — which are representative of the standard deviation of the data (Figure 1). Additionally, transformation of yeast with our plasmid vector was more likely to result in LOF of the ADE2 gene (red phenotype) when HDR template B was present compared to when HDR template A or when no template were present (Figure 1). Furthermore, colony growth regardless of phenotype was more likely to occur in transformations with our designed gRNA sequence ligated to the plasmid vector when an HDR template was present (Figure 1).



	HDR A	HDR B	pML104	pML104 + gRNA	Water
Average # of white colonies	114.4	103.5	1011.9	46.2	0
Average # of red colonies	24.9	213.4	0.2	16.8	0

Figure 1. Colony Growth of Different Transformation Types. The above results are representative of the average number of colored yeast colonies grown from five different transformation types. The average is taken from transformations from the experiments of ten different groups. The error bars indicate the standard deviation of the data.

Sequencing Results

The computational classification of the repair mechanism of the selected yeast sequences are summarized in Figure 2. Incorporation frequencies have also been calculated from the list of stored HDR-introduced mutations for each sequence analyzed. Notably, single point mutations occurred more frequently when their location in the template was closer to the cut site.

Color	Repair	Transformation Type					
		recombinant	HDR A	HDR B	HDR C	HDR D	total
white	NHEJ	6	6	0	2	0	14
	HDR	1	0	7	0	0	8
	no cut	23	28	7	6	10	74
red	NHEJ	21	37	2	11	11	82
	HDR	0	5	54	0	1	60
	no cut	1	2	0	0	0	3
total		52	78	70	19	22	241

Figure 2. Computational Classification of Sequencing Results. The above table represents the results of classifying the repair mechanism of each sequence via rudimentary bioinformatics analysis of the 353 colonies sequenced by our lab staff. Sequences were controlled by quality (QC score greater than 30) and by length (several groups submitted sequences that were too short for proper alignment). The 241 sequences meeting these controls were classified by repair type using computational analysis of the alignment of the sequence to the wild-type sequence using affine gap penalties.

Colony types	Incorporations	Colony count	Mutation Incorporation %
NHEJ-classified, HDR A transformed	89	43	34.5%
NHEJ-classified, HDR B transformed	3*	2	25.0%
HDR-classified, HDR A transformed	18	5	60.0%
HDR-classified, HDR B transformed	212	61	57.9%
Total NHEJ-classified	92	45	34.1%
Total HDR-classified	230	66	58.1%
Total HDR and NHEJ classified	322	111	48.3%

Figure 3. Incorporation Frequency of Different Colony Types. The designed mutation incorporations in a colony from transformations with HDR A or HDR B present were tracked if a Cas9-mediated cut was made. Then these colonies were partitioned into groups based on which repair mechanism they were classified as having undergone. For each of these groups, the mutation incorporation percentage was calculated by dividing the number of incorporations in a group by the number of colonies in a group, and finding the percentage by normalizing by six (for the six designed point mutations in the HDR templates).

Mutation	Effect	Position	Incorporation %
Mutn 1	A → G	-51	20.9%
Mutn 2	A → T	-46	29.5%
Mutn 3	T → C	-8	48.0%
Mutn 4	G → A	17	41.2%
Mutn 5	A → T	27	41.2%
Mutn 6	G → C	43	37.2%

Figure 4. Incorporation Frequency for Each Designed HDR Point Mutation. For each of the six positions in our HDR template where a point mutation was designed — as well as for the PAM site mutation in HDR template B — the frequency of incorporation was calculated across all sequences.

```
        *         *         *         *         *         *         *         *         *         *
186>CAAGTATGGATTCTAGAACAGTTGGTATATTAGGAGGGGGACAATTGGGACGTATGATTGTTGAGGCAGCAAACAGGCTCAACATTAAGACGGTAATACT>285
158>CAAGTATGGATTCTAGAACAGTTGGTATATTAGGAGGGGGACAATTGGGACGCATGATTGTTGAATCAGCAAACAGACTCAACATTTAGACGGT-ATACT>256
155>CAAGTATGGATTCTAGAACAGTTGGTATATTAGGAGGGGGACAATTGGG-CGCATGATTGTTGAGGCAGCAAACAGACTCAACATTTAGACGGTAATACT>253
151>CAAGTATGGATTCTTGAACAGTTGGTATATTAGGAGGGGGACAATTGG-ACGCATGATTGTTGAGGCAGCAAACAGGCTCAACATTAAGACGGTAATACT>249
157>CAAGTATGGGTTCTTGAACAGTTGGTATATTAGGAGGGGGACAATTGG-ACGCATGATTGTTGAGGCAGCAAACAGGCTCAACATTAAGACGGTAATACT>255
149>CAAGTATGGATTCTTGAACAGTTGGTATATTAGGAGGGGGACAATTGG-ACGCATGATTGTTGAGGCAGCAAACAGACTCAACATTTAGACGGTAATACT>247
```

Figure 5. Multiple-Sequence Alignment of Red Colonies Transformed with HDR A. These five subsequences are from red colonies transformed with HDR A that were classified as having undergone homology directed repair. The top sequence is the wild-type sequence of the ADE2 gene with the gRNA target site, PAM sequence, and HDR-designed mutation locations highlighted. The first colony sequence (second row) was labeled as being transformed with HDR A, but it may have been misreported because it includes the PAM site edit that was only part of HDR B. Interestingly, the remaining four colony sequences all have a deletion somewhat far from the cut site but proximal to a point mutation that substitutes a thymine for a cytosine that is present in each of the sequences. The reason these colonies were not classified as NHEJ-repaired is because the indel in each of them is farther from the cut site than the threshold determined for the computational algorithm — the threshold was set to a ten base pair distance. However, it is still possible that the ADE2 gene in these colonies were repaired by NHEJ as the final repair step.

```
        *         *         *         *         *         *         *         *         *         *
198>CTAGAACAGTTGGTATATTAGGAGGGGGACAATTGGGACGTATGATTGTTGAGGCAGCAAACAGGCTCAACATTAAGACGGTAATACTAGATGCTGAAAA>297
177>||||||||||||||||||||||||||||||||||||||||||||||||   |##||||||||||#|||||||||#|||||||||||||||#|||||||||
177>CTAGAACAGTTGGTATATTAGGAGGGGGACAATTGGGACGTATGATTG---AATCAGCAAACAGACTCAACATTTAGACGGTAATACTAGGTGCTGAAAA>273
```

Figure 6. Alignment of Colony #2015 to the Wild-type ADE2 Sequence. Curiously, this colony (which was transformed with HDR B) had an ADE2 sequence that contained a deletion after the cut site (typical of NHEJ) but also introduced mutations from HDR template B including an edit of the PAM site.

---

*The three point mutation incorporations from NHEJ-classified colonies that were transformed by HDR B all came from the unexpected result of sequencing colony #2015 referenced in Figure 6. The other colony from this category in Figure 3 did not incorporate any point mutations, which we believe is the likely outcome if a larger dataset had afforded more colonies to this category.

DISCUSSION

Frequency of Point Mutation Incorporation

Our findings indicate that point mutations from a template oligo are more likely to be incorporated after Cas9-mediated cutting when they are closer to the cut site. This is evidenced by Figure 4, which shows the frequency of incorporation for each point mutation included in our HDR templates. When the mutations are sorted by absolute distance from the cut site — starting with the closest mutation — their corresponding incorporation frequencies are decreasing. The PAM site edits were only five base pairs away from the cut site and had an incorporation rate of 88.6% while the furthest mutation was 51 base pairs away from the cut site and had an incorporation rate of 20.9%. This is consistent with the biological mechanism for HDR repair, in which due to the nature of the binding of homology arms, minimizing cut-to-mutation distance of a point mutation increases the efficiency of incorporation of that mutation by HDR (Wang et al).

Effect of PAM Site Editing on Mutation Incorporation

Mutating the PAM site during CRISPR-Cas9 gene editing is essential for ensuring efficient homology directed repair. In Figure 2, we see that 98% of colonies that were red and were transformed by HDR B were repaired via HDR whilst for red colonies transformed by HDR A, that percentage was only 11%. This difference is made even more stark when examining the five red HDR A sequences that were classified as having been repaired by HDR. Figure 5 shows how these five sequences appear to have not been repaired solely by HDR A — one of them was misreported and the other four are likely to have initially undergone HDR but then were terminally repaired by NHEJ. In fact, given the mutation incorporation frequency of supposed NHEJ-repaired HDR A colonies (Figure 3), it is likely that a majority of the red colonies transformed by HDR A were initially repaired by HDR before being repaired for the last time by NHEJ.

The reason these colonies could have undergone two different repair mechanisms is due to the repeatability of the Cas9 nuclease activity. If the PAM site remains unedited, Cas9 may recognize it again after a cut has been repaired and may execute another cut, necessitating another repair. In their study, Bao et al. attended to the problem of needing to terminate Cas9 cutting activity by including an 8 base pair deletion at the PAM sequence of their HDR template (Bao et al). Interestingly, this study also demonstrated the repeatability of the Cas9 cutting action by designing an efficient multi-gene editing CRISPR/Cas system in which Cas9 enzymes can cut at multiple sites in the yeast genome to disrupt multiple genes. In our research, the difference in mutation incorporation between colonies transformed with HDR A and those transformed with HDR B can be attributed to the fact that our designed HDR B template includes a two nucleotide substitution in the PAM sequence, which renders that site unrecognizable to Cas9 after homology directed repair has been enacted.

Frequency of ADE2 Loss of Function After Cas9 Activity

In the yeast ADE2 gene, LOF after a Cas9-mediated cut is more likely to occur in the presence of an HDR template that introduces a nonsense mutation and edits the PAM sequence adjacent to the Cas9 cut. Figure 1 shows that, on average, 67% of colonies transformed under these conditions (transformation with HDR B) experience LOF of the ADE2 gene compared to 18% (transformation of HDR A) and 27% (recombinant plasmid transformation) of colonies that aren't transformed with such an HDR template present. Either HDR or NHEJ can be responsible for loss of function to the ADE2 gene. In our HDR templates, point mutations designed to create early stop codons after incorporation are likely responsible for LOF after homology directed repair. Since the efficiency of HDR is improved by editing the PAM site (as demonstrated in the previous section), the higher observed frequency of LOF for transformations with HDR B compared to the other transformation types is expected. Meanwhile for sequences repaired by NHEJ, the likely culprit for LOF are frameshifts that are resultant from the insertions or deletions near the cut site that are characteristic of this repair mechanism. These ideas are echoed by the findings of Lian et al, whose summary of several CRISPR/Cas9 studies explains that the most efficient mediated gene knock-out options are introductions of stop codons or frameshifts to the target site (Lian et al).

REFERENCES

Bao, Zehua, Han Xiao, Jing Liang, Lu Zhang, Xiong Xiong, Ning Sun, Tong Si, and Huimin Zhao. "Homology-integratedCRISPR–Cas (HI-CRISPR) system for one-step multigene disruption in Saccharomyces cerevisiae." *ACS synthetic biology* 4, no. 5 (2015): 585-594.

Lian, Jiazhang, Mohammad HamediRad, and Huimin Zhao. "Advancing metabolic engineering of Saccharomyces cerevisiae using the CRISPR/Cas system." *Biotechnology Journal* 13, no. 9 (2018): 1700601.

Wang, Kankan, Xiaochun Tang, Yan Liu, Zicong Xie, Xiaodong Zou, Mengjing Li, Hongming Yuan, Hongsheng Ouyang, Huping Jiao, and Daxin Pang. "Efficient generation of orthologous point mutations in pigs via CRISPR-assisted ssODN-mediated homology-directed repair." *Molecular Therapy-Nucleic Acids* 5 (2016): e396.