A method for authenticating the fidelity of *Cryptococcus neoformans* knockout collections

Ella Jacobs[1], Quigly Dragotakes[1], Madhura Kulkarni[1], Amanda Dziedzic[1], Anne Jedlicka[1], J. Marie Hardwick[1], Arturo Casadevall[1]

[1]W. Harry Feinstone Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

Corresponding author: acasade1@jhu.edu

Abstract

Gene knockout strain collections are important tools for discovery in microbiology. The only available genome-wide deletion collection for a human pathogenic fungus, *Cryptococcus neoformans,* is utilized widely for genetic studies. We uncovered mix-ups in the assembly of the commercially available *C. neoformans* deletion collection of ~6,000 unique strains acquired by our lab. While pursuing the characterization of a gene-of-interest, the corresponding deletion strain from the *C. neoformans* KO collection displayed several interesting phenotypes associated with virulence. However, RNAseq analysis identified transcripts for the putative knockout gene, and the absence of transcripts for a different knockout strain found in the same plate position in an earlier partial knockout collection, raising the possibility that plates from one collection were substituted for the other, indicating a misidentified knockout strain. This was supported by determining the size of the nourseothricin (NAT)-resistance cassette used to generate the two separate knockout libraries. Here we report that our KN99α collection is comprised of mixed plates from two independent libraries and present a simple authentication method that other investigators can use to distinguish the identities of these KO collections.

Introduction

In the field of microbiology, genome-wide gene deletion collections are important investigative tools to rapidly screen for genes associated with specific phenotypes. For studies of *Cryptococcus neoformans*, several gene deletion collections are available [1, 2], including the genome-wide knockout collection containing 6,840 strains generated in the Madhani laboratory, an arduous feat given the many challenges relative to manipulating the genome of the model yeast *Saccharomyces cerevisiae* [3-5].

The first Madhani *C. neoformans* gene knockout (KO) collection was generated in the H99W background strain (2008 collection of ~1300 KO strains) [6]. H99W, also known as "Wimp", is a laboratory adapted strain with attenuated virulence and melanization defects attributed to a 7 bp insertion in the *LMP1* gene [7, 8]. Given the severe virulence attenuation of the H99W strain, a subsequent genome-wide deletion collection was generated in the virulent KN99α background strain, arrayed in 96-well plates and distributed in three phases, designated as 2015 (22 plates), 2016 (20 plates) and 2020 (7 plates) collections containing one KO strain per well [9]. Knockout strains were generated by gene replacement/disruption via homology directed repair using gene-specific ~1 kb homology arms flanking a nourseothricin (NAT) cassette [9].

Our 2015 and 2016 KO collections were purchased from the Fungal Genetics Stock Center (FGSC) in March 2018 and the 2020 KO collection in March 2021. These purchased collections have proved useful for screening and identifying genes associated with non-lytic exocytosis [10] protein secretion [11], and cell death [12]. However, during the

characterization of several deletion strains selected based on proteomics datasets, we realized that many strains from our 2015 collection did not contain the expected gene disruption, despite being NAT resistant.

We tested plates of our 2015 collection and found that rather than individual well mix-ups or well-well contamination, whole plates from the KN99 collection appear to have been swapped with 2008 collection plates. The 2008 collection contains strains are arrayed in 14 96-well plates. Based on the strains we tested, only plates within our first 14 plates of the 2015 KO collection were impacted. Overall, a total of nine non-consecutive plates appeared to be directly swapped between the 2015 and 2008 collection, while no mix-ups were detected for the strains tested from the 2016 and 2020 collections. We developed a simple, cost-effective method of testing deletion strains purchased from the Fungal Genetics Stock Center (FGSC) by comparing NAT cassette sizes. This protocol will allow screening to quickly assess potential fidelity issues for follow-up.

Results

Targeting genomic loci of 2015 KO collection library strains via PCR

Using overlap PCR, the Madhani lab constructed deletion cassettes containing the NAT-resistance gene and regulatory sequences plus 1 kb homology arms. These cassettes were subsequently transformed and inserted into the *C. neoformans* gene of interest by homology directed repair (HDR) [9].

The cassette construction primers used by the Madhani lab to generate the KO library are available on the FGSC website (Fig 1A). Primer pairs W1+W2 and W5+W6 amplify the 5' and 3' recombination arms, respectively, derived from genomic DNA flanking the gene-of-interest. The primer pairs W3 + W4 amplified the NAT resistance cassette (Fig 1A). and allowed the fusion of homology arms and the resistance cassette (Fig 1A overhanging region). Of these primers, W1, W2, W5 and W6 are included in the FGSC spreadsheets while W3 and W4 are not.

To verify the presence of the NAT drug selection marker in the correct KO gene locus, we used modified primer sequences reported for the Madhani collection for genotyping. Primers W2 and W5 contain sequences with homology to both the NAT cassette and unique genomic DNA for each gene. We used primers adapted from W2 and W5, indicated as W2GC and W5GC. These correspond to the unique genomic sequences present in primers W2 and W5 used to amplify 1 kb homology arms during construction of the deletion collection by Madhani and colleagues. As such, the sequence of W2GC and W5GC differs for each gene (Fig 1B, Table 3).

In KO strains, if the gene was disrupted as intended, the W2GC and W5GC primer pair amplifies the NAT cassette and yields a PCR product (expected ~1.7 kb) that differs in size from the PCR product of the WT strain (Fig 1C). It is worth noting that this methodology only works if the gene of interest is different in size from the ~1.7kb expected NAT cassette. If not, further validation with a diagnostic restriction enzyme digestion, alternative primer design or sequencing of the amplified products is necessary. We compared the size of PCR products amplified between WT and KO strains at the genomic locus (Fig 1D, Table 2). We found that many strains did not contain a NAT cassette insertion at the expected locus (Table 2). For example, the PCR products amplified from the WT and CNAG_04891Δ targeting the genomic region of CNAG_04891 are the same size as the expected gene product, which is smaller than ~1.7kb (Fig 1D). This implies that there is no disruption in the CNAG_04891Δ strain within the CNAG_04891 locus. Conversely, when comparing the PCR products amplified from WT and CNAG_01047Δ targeting the CNAG_01047 locus, there is a size difference between the bands and the size of the CNAG_01047Δ PCR product is consistent with the expected NAT cassette insertion. The size difference also occurs when comparing the PCR products targeting the CNAG_05893 locus amplified from the WT and CNAG_05893Δ strains. Taken together, CNAG_04891Δ does not contain a NAT cassette disruption at the expected locus, while both CNAG_01047Δ and CNAG_05893Δ KO strains contain NAT cassette disruptions at the expected locus. Using the methodology outlined in panel C can be time consuming if many strains must be tested due to the varying annealing temperatures

and individualized primers required for each strain. Hence, a method to verify strains using the same primers, as we discovered below, would be useful.

Larger NAT cassette size correlated with strains lacking expected KO

We were curious whether the strains lacking a disruption at the expected locus, such as CNAG_04891*Δ,* contained a NAT cassette elsewhere in the genome. Within the W2 + W5 primer pairs, the same sequence appeared to be used in all strains to fuse the homology arms with the NAT cassette via overlap PCR. By designing primers that were the reverse complement of the W2 + W5 NAT cassette region, designated as W2NC and W5NC in figure 1 panel B, this primer pair could be used on any deletion strain to amplify the NAT cassette as it contains no locus-specific base pairs (Fig 2A, Table 1). As expected, using these primers with genomic DNA isolated from WT KN99α did not yield a PCR product due to the lack of NAT cassette (Fig 2B&C). A NAT cassette PCR product was successfully amplified from all deletion strains tested, including strains that did not contain the expected KO, as determined by the methodology outlined in figure 1. This implied the NAT cassette was present elsewhere in the genome.

Interestingly, when comparing amplified NAT cassettes PCR product size from 5 strains lacking the expected KO (CNAG_01019Δ, CNAG_02179Δ, CNAG_04981Δ, CNAG_02801Δ, and CNAG_04891Δ) to strains containing the expected KO, we noticed that in strains where the expected gene did not contain a disruption, the corresponding NAT cassette PCR product was always larger than expected (~2 kb rather than ~1.7 kb) (Table 2). As such, we found the larger NAT cassette size correlated with KO strains lacking the expected NAT cassette locus disruption.

.

A direct plate swap between the 2015 and 2008 KO collections

We summarized these findings with a PCR experiment (Fig 3A). Following the 1kb+ ladder lane, the size differences of passed and failed NAT cassettes are compared in the first three wells. This strategy is outlined in figure 2 panel A. The remaining wells are PCR products from the genomic segment of primers W2 and W5 outlined in figure 1 panel B. Next, we amplified the genomic region of CNAG_02179 from KN99α and our 2015 KO collection plate 2 well H12 (Putative CNAG_02179Δ). Both PCR products are the same size, indicating this strain did not possess a disruption in CNAG_02179. Next, we checked this same strain for a disruption in CNAG_03019, the gene that would be disrupted if the 2015 plate was swapped with a 2008 plate 1 for 1. The KN99 PCR product was the expected size and 2015 KO collection plate 2 well H12 possessed a confirmed disruption in CNAG_03019.

Next, we compared our 2015 library plate 4 well A1 (Putative CNAG_01019Δ). This strain was also confirmed to lack a disruption in the CNAG_01019 genomic locus. If our plate 4 well A1 was swapped 1 for 1 with the 2008 library plate 4 in a direct swap, we would expect CNAG_04781 to be disrupted. Amplifying the genomic region of CNAG_04781 resulted in the expected product in KN99α WT, and a size change in the putative CNAG_01019Δ strain, consistent with the plates having been swapped. Plates where the NAT cassette was the smaller 1.7kb size, such as plate 10 well A4 (CNAG_04388Δ) did not possess disruptions in

the 2008 plate position KO (CNAG_01495). The strain in our 2015 plate 3 well D4 (putative CNAG_04981Δ) also possessed a disruption in the 2008 plate 3 well D4 gene CNAG_06252. Taken together, these results confirm that the larger NAT cassette size correlated with a direct plate swap between the 2015 and 2008 KO collections.

We tested many strains in our 2015 library using a mix of the genomic locus and NAT cassette size strategy (Table 2). The results are best interpreted to indicate that whole plates were directly swapped, not individual wells. The 2008 KO collection possesses 14 plates, and only plates within our first 14 plates of the 2015 KO collection were impacted. Overall, our plates 1-7, 9, and 14 appear to be directly swapped between the 2015 and 2008 libraries (Fig 4 A and B).

Additionally, it is worth noting that the 2015 W2 + W5 primers were used to amplify the suspected 2008 swapped gene disruptions (Fig 4A). The 2008 KO collection used different primers to KO target genes, therefore, can be expected to differ when comparing locus specific targets with 2015 library primers. For example, The PCR product amplified targeting the CNAG_03019 locus within the putative CNAG_02179Δ strain is larger than just the NAT cassette. This is explained by the insertion site of the 2008 version of this KO occurring ~300 bp after the start codon. These findings were confirmed by the RNAseq data (Fig 6B). For locus-specific confirmation, 2008-specific primers could be used, but we feel the size differences add further evidence to confirm the KO is in fact of 2008 origin. Ultimately, we

discovered that the larger NAT cassette size may be the result of a direct plate swap between the 2015 and 2008 KO collections.

Experimental work leading to the discovery of a library mix-up.

In a proteomics experiment, we infected mouse bone marrow derived macrophages (BMDMs) with M0, M1, M2 polarization states with *C. neoformans*. To characterize proteins secreted by *C. neoformans* during infection, infected macrophages were subjected to gentle lysis and proteomic analysis (Fig 5A). We identified CNAG_02179 as a candidate of interest for further study based on its significantly higher abundance in M1 polarized macrophages compared to M0 (Fig 5B). This gene was present in the 2015 KO collection and denoted as passing their internal PCR verification (available on the FGSC website). We pulled the CNAG_02179*Δ* strain from plate 2 well H12 of our 2015 collection (Fig 5C). Characterization of this strain revealed phenotypes that were very similar to the phenotypes of H99W. This included avirulence in the *Galleria mellonella* model of infection and melanization defects (Fig 5D&E).

In addition to our PCR based verification, we used data from an RNAseq experiment conducted on putative CNAG_02179Δ to uncover the plate swap. CNAG_02179*Δ* was pulled from plate 2 well H12 from our 2015 labeled library plates. The putative CNAG_02179*Δ* possessed RNAseq reads for the gene that was supposed to be disrupted (Fig 6A). To narrow down the list of potentially knocked out genes, we figured that knocked out gene would be significantly upregulated expression in KN99 vs. CNAG_02179Δ comparisons and would

exhibit no significant expression changes when comparing CNAG_02179Δ to itself in different conditions. The four highest differences between KO strain and control that satisfied these conditions were: CNAG_03019, CNAG_02959, CNAG_06092, and CNAG_02958. We wondered if the differences in NAT cassette size could be due to barcoding or other potential changes that would differ between KO collections. For example, the 2008 KO collection cassettes are barcoded [6]. Following this hunch, we looked at the 2008 library plate map, available in the supplemental information of the accompanying paper [6]. As expected, CNAG_03019*Δ* was the strain in plate 2 well H12 in the 2008 KO collection. This indicated that our 2015 KO collection plates were swapped with 2008 KO collection plates. The 2008-plate specific deletion cassette was further confirmed with RNAseq data showing the CNAG_03019 deletion begins ~300 bp after the start codon within the putative CNAG_02179*Δ*, supporting the PCR product size in figure 3 panel A (Fig 6B). The methodology comparing NAT cassette sizes between strains offers a quick, cost-effective methodology for testing the presence of a plate swap within a 2015 KO collection, as summarized in figure 7 (Fig 7A).

Discussion

Here, we describe an unfortunate plate mix-up involving cryptococcal gene KO libraries and demonstrate a cost-effective, time-saving method for authenticating the fidelity of *Cryptococcus neoformans* 2015 deletion KO collections. This will allow other labs to rapidly determine the status of their 2015 libraries. Using NAT cassette specific primers, not locus specific primers, to size comparison screening of the library proved a useful tool for initial screening. However, we did not test every strain and other sized NAT cassettes may exist within the library. Follow up with locus specific testing is essential to confirm that plates were swapped.

There are several limitations to the methods presented here. Firstly, we have found that locus specific primers were not specific enough to compare insertion sizes between 2008 and 2015 KO strains. The same insertion sites were not used between the 2008 and 2015 libraries to generate the same KO strains. Hence, care must be used when applying locus-specific primers to suspected 2008 swapped strains, as PCR product sizes will differ. An example of this can be seen when CNAG_03019*Δ* primers from the 2015 library are applied to the CNAG_03019*Δ* from the 2008 swapped plate. To confirm a 2008 swapped plate in a locus specific manner, we recommend using 2008 locus specific primers. Alternatively, be aware of the expected size differences when performing the experiment. Additionally, comparing NAT cassette sizes requires a 1.5% agarose gel because higher concentrations impede the separation of the PCR products. Finally, we have no information of where, when or how the plate swap occurred.

In summary, our experience is a cautionary tale of the foibles that can accompany the use of community resources like gene deletion libraries. Although these resources are powerful tools for discovery one must always maintain a certain skepticism when using them and validate any important finding by independent methods. We estimate that the lead authors in this paper lost approximately one year of research time in following up on preliminary data and then sorting out the mix-up with the libraries. Currently, these libraries are in many laboratories studying *C. neoformans* and continue to be useful tools for research. We are hopeful that our experiences and the validation roadmap described here will allow other laboratories to maximize their potential while avoiding false leads.

Materials and methods

Strains

Our lab purchased the 2015 and 2016 Madhani KO collections in March 2018 and the 2020 plate collections in March 2021.

Genomic DNA isolation and colony PCR.

Genomic DNA isolation was performed using the CTAB isolation method [13, 14]. Strains were cultured in 50 mL YPD, centrifuged and frozen overnight at 80 °C. The frozen cell pellet was powdered using glass beads and a bead beater. 10 mL CTAB extraction buffer (100 mM Tris pH7.5, 0.7 M NaCl, 10 mM EDTA, 1% CTAB, 1% beta-mercaptoethanol) was added to the culture pellet and vortexed. After 30 min of incubation at 65 °C, tubes were cooled in ice water and an equal volume of chloroform was added. The solution was mixed by inversion for 1 min and centrifuged for 10 min at max speed (RT). The top layer was transferred to a clean tube and an equal volume of isopropanol was added. The solution was mixed by inversion for 1 min and centrifuged for 5 min at max speed (RT). The supernatant was decanted off and 1 ml fresh 70% EtOH was added to the sample. After 5 min of centrifugation, the supernatant was decanted off and the pellet was resuspended in 100 µL sterile molecular biology water.

For further testing of the entire library, we utilized colony PCR methodology [15]. In this method, small amounts of each strain are streaked directly into a PCR tube and microwaved for 2 min to release DNA. PCR mastermix and primers were added directly to the tube containing microwaved fungal cells.

Phusion PCR mastermix was used to amplify genomic loci and NAT cassettes. Due to the high annealing temperature of the NAT cassettes, a 2-step PCR protocol (combined annealing/extension) was utilized. All reactions were run on a 1.5% agarose gel. Higher percentages of agarose gels will impede size dependent comparisons.

Proteomics

Whole cell BMDM lysates were collected from plates using non-enzymatic cell lifter and pelleted at 500 *g* for 5 min. Cells were then lysed by resuspension in 10 mM HEPES buffer and passed through 26 ¾ gauge syringes to ensure mammalian cell lysis. Protein concentrations were determined by BSA and sent for mass spectroscopy protein identification.

Protein Digestion: Protein extracts were buffer exchanged using SP3 paramagnetic beads (GE Healthcare) (Hughes CS et al, Nature Protocols 2019 14(1):68-85). Briefly, protein samples (20 ug) were brought up to 100 uL with 10 mM TEAB + 1% SDS and disulfide bonds reduced with 10 uL of 50 mM dithiothreitol for 1 hour at 60C. Samples were cooled to RT and pH adjusted to ~7.5, followed by alkylation with 10 uL of 100 mM iodoacetamide in the dark at RT for 15 minutes. Next, 100 ug (2 uL of 50 ug/uL) SP3 beads were added to the samples, followed by 120 uL 100% ethanol. Samples were incubated at RT with shaking for 5 minutes. Following protein binding, beads were washed with 180 uL 80% ethanol three times. Proteins were digested on-bead with 2ng trypsin (Pierce) in 100uL 25mM TEAB buffer at 37C overnight. Resulting peptides were separated from the beads using a magnetic tube holder. Supernatants containing peptides were acidify and desalted on u-HLB Oasis plates. Peptides were eluted with 60% acetonitrile/0.1%TFA and dried using vacuum centrifugation

Isobaric Mass Tag Labeling: Each of the 12 dried peptide samples were labeled with one of the unique TMTpro 18-plex reagents (Thermo Fisher, Lot ????????) according to the manufacturer's instructions. All TMT labeled peptide samples were combined and dried by vacuum centrifugation.

Peptide fractionation: The combined TMT-labeled peptides were re-constituted to 2 mL in 10 mM TEAB in water and loaded on a XBridge C18 Guard Column (5 µm, 2.1 x 10 mm, Waters) at 250 µL/min for 8 min prior to fractionation on a XBridge C18 Column (5 µm, 2.1 x 100 mm column (Waters) using a 0 to 90% acetonitrile in 10 mM TEAB gradient over 85 min at 250 µL/min on an Agilent 1200 series capillary HPLC with a micro-fraction collector.

Eighty-four 250 ul fractions were collected and concatenated into 24 fractions according to Wang et al 2011 (Proteomics 11: 2019-2026. PMC3120047).

Mass Spectrometry analysis: TMT labeled peptides in each fraction were analyzed by nanoflow reverse phase chromatography coupled with tandem mass spectrometry (nLCMS/MS) on an Orbitrap-Fusion Lumos mass spectrometer (Thermo Fisher Scientific) interfaced with an EasyLC1000 UPLC. Peptides will be separated on a 15 cm × 75 µm i.d. self-packed fused silica columns with ProntoSIL-120-5-C18 H column 3 µm, 120 Å (BISCHOFF) using an 2-90% acetonitrile gradient over 85 minutes in 0.1% formic acid at 300 nl per min and electrosprayed through a 1 µm emitter tip (New Objective) at 2500 V. Survey scans (MS) of precursor ions were acquired with a 2 second cycle time from 375-1500 m/z at 120,000 resolution at 200 m/z with automatic gain control (AGC) at 4e5 and a 50 ms maximum injection time. Top 15 precursor ions were individually isolated within 0.7 m/z by data dependent monitoring and 15s dynamic exclusion, and fragmented using an HCD activation collision energy 39. Fragmentation spectra (MS/MS) were acquired using a 1e5 AGC and 118 ms maximum injection time (IT) at 50,000 resolution.

Data analysis: Fragmentation spectra were processed by Proteome Discoverer (v2.5, ThermoFisher Scientific) and searched with Mascot v.2.8.0 (Matrix Science, London, UK) against RefSeq2021_204 database. Search criteria included trypsin enzyme, two missed cleavage, 5 ppm precursor mass tolerance, 0.01 Da fragment mass tolerance, with TMTpro on N-terminus and carbamidomethylation on C as fixed modifications and TMTpro on K, deamidation on N or Q as variable modifications. Peptide identifications from the Mascot searches were processed within PD2.5 using Percolator at a 5% False Discovery Rate confidence threshold, based on an auto-concatenated decoy database search. Peptide spectral matches (PSMs) were filtered for Isolation Interference <30%. Relative protein abundances of identified proteins were determined in PD2.4 from the normalized median ratio of TMT reporter ions from the top 30 most abundant proteins identified. ANOVA method was used to calculate the p-values of mean protein ratios for the biological replicates set up using a non-nested (or unpaired) design. Z-score transformation of normalized protein abundances from a quantitative proteomics analysis using isobaric mass tags was applied before performing the hierarchical clustering based on Euclidean distance and complete (furthest neighbors) linkage. The horizontal dendogram shows the proteins in samples that clustered together. Technical variation in ratios from our mass spectrometry analysis is less than 10% (Herbrich et al 2013 J Proteome Res 12: 594).

RNAseq

The yeasts were brought up to 1mL of Trizol and homogenized in the FastPrep 24 (MP Bio) with Lysing Matrix C Fast Prep tubes according to manufacturer's protocol with minor modifications, followed by RNA extraction using the PureLink RNA Mini kit with on-column DNase treatment (ThermoFisher). RNA-Seq Libraries were prepared using the Universal Plus mRNA-Seq Library prep kit (Tecan Genomics) incorporating unique dual indexes.

Libraries were assessed for quality by High Sensitivity D5000 ScreenTape on the 4200 TapeStation (Agilent Technologies). Quantification was performed with NuQuant reagent and by Qubit High Sensitivity DNA assay, on Qubit 4 and Qubit Flex Fluorometers (Tecan Genomics/ThermoFisher).
Libraries were diluted and an equimolar pool was prepared, according to manufacturer's protocol for appropriate sequencer. An Illumina iSeq Sequencer with iSeq100 i1 reagent V2 300 cycle kit was used for the final quality assessment of the library pool. For deep mRNA sequencing, a 200 cycle (2 x 100 bp) Illumina NovaSeq 6000 S1 run was performed at Johns Hopkins Genomics, Genetic Resources Core Facility, RRID:SCR_018669. mRNA-seq data was analyzed with Partek Flow NGS Software as follows: pre-alignment QA/QC; alignment to *C. neoformans* Reference Index using STAR 2.7.8a; post- alignment QA/QC; quantification of gene counts to annotation model (Partek E/M); filter and normalization of gene counts; identification and comparison of differentially expressed genes with GSA (gene specific analysis). mRNA-seq data was aligned to *C. neoformans* using Bowtie 2 and chromosome view was used to visualize aligned reads to the putative CNAG_02179*Δ* strain vs WT KN99 strain (Partek Flow). Genome: NCBI: GCF_000149245.1_CNA3.
All sequence files and sample information have been deposited at NCBI Sequence Read Archive, NCBI BioProject: insert accession number here.

Funding

We acknowledge the Madhani Lab and the National Institutes of Health funding (R01AI100272) for the creation of the KO collection resource. This work was supported by the National Institutes of Health (R01HL059842, 5R01AI033774, 5R37AI033142, and 5R01AI052733 to AC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements
We would like to thank the Madhani Lab for their assistance in getting to the bottom of this issue and for providing the correct replacement strains to correct our papers.

References

1. Lee, K.T., et al., *Systematic functional analysis of kinases in the fungal pathogen Cryptococcus neoformans*. Nat Commun, 2016. **7**: p. 12766.
2. Jung, K.W., et al., *Systematic functional profiling of transcription factor networks in Cryptococcus neoformans*. Nat Commun, 2015. **6**: p. 6757.
3. Baker Brachmann, C., et al., *Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications*. Yeast, 1998. **14**(2): p. 115-132.
4. Winzeler, E.A., et al., *Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis*. Science, 1999. **285**(5429): p. 901-906.
5. Giaever, G., et al., *Functional profiling of the Saccharomyces cerevisiae genome*. Nature, 2002. **418**(6896): p. 387-391.
6. Liu, O.W., et al., *Systematic Genetic Analysis of Virulence in the Human Fungal Pathogen Cryptococcus neoformans*. Cell, 2008. **135**(1): p. 174-188.
7. Janbon, G., et al., *Analysis of the genome and transcriptome of Cryptococcus neoformans var. grubii reveals complex RNA expression and microevolution leading to virulence attenuation*. PLoS Genet, 2014. **10**(4): p. e1004261.
8. Arras, S.D.M., et al., *Convergent microevolution of Cryptococcus neoformans hypervirulence in the laboratory and the clinic*. Scientific Reports, 2017. **7**(1): p. 17918.
9. Chun, C.D. and H.D. Madhani, *Applying genetics and molecular biology to the study of the human pathogen Cryptococcus neoformans*. Methods Enzymol, 2010. **470**: p. 797-831.
10. Dragotakes, Q., et al., *Bet-hedging antimicrobial strategies in macrophage phagosome acidification drive the dynamics of Cryptococcus neoformans intracellular escape mechanisms*. PLOS Pathogens, 2022. **18**(7): p. e1010697.
11. Jung, E.H., et al., *Cryptococcus neoformans releases proteins during intracellular residence that affect the outcome of the fungal-macrophage interaction*. Microlife, 2022. **3**: p. uqac015.
12. Stolp, Z.D., et al., *Yeast cell death pathway requiring AP-3 vesicle trafficking leads to vacuole/lysosome membrane permeabilization*. Cell Rep, 2022. **39**(2): p. 110647.
13. Pitkin, J.W., D.G. Panaccione, and J.D. Walton, *A putative cyclic peptide efflux pump encoded by the TOXA gene of the plant-pathogenic fungus Cochliobolus carbonum*. Microbiology (Reading), 1996. **142 (Pt 6)**: p. 1557-1565.
14. Velegraki, A., et al., *Rapid extraction of fungal DNA from clinical samples for PCR amplification*. Sabouraudia, 1999. **37**(1): p. 69-73.
15. Huang, M.Y., et al., *Short homology-directed repair using optimized Cas9 in the pathogen Cryptococcus neoformans enables rapid gene deletion and tagging*. Genetics, 2021. **220**(1).

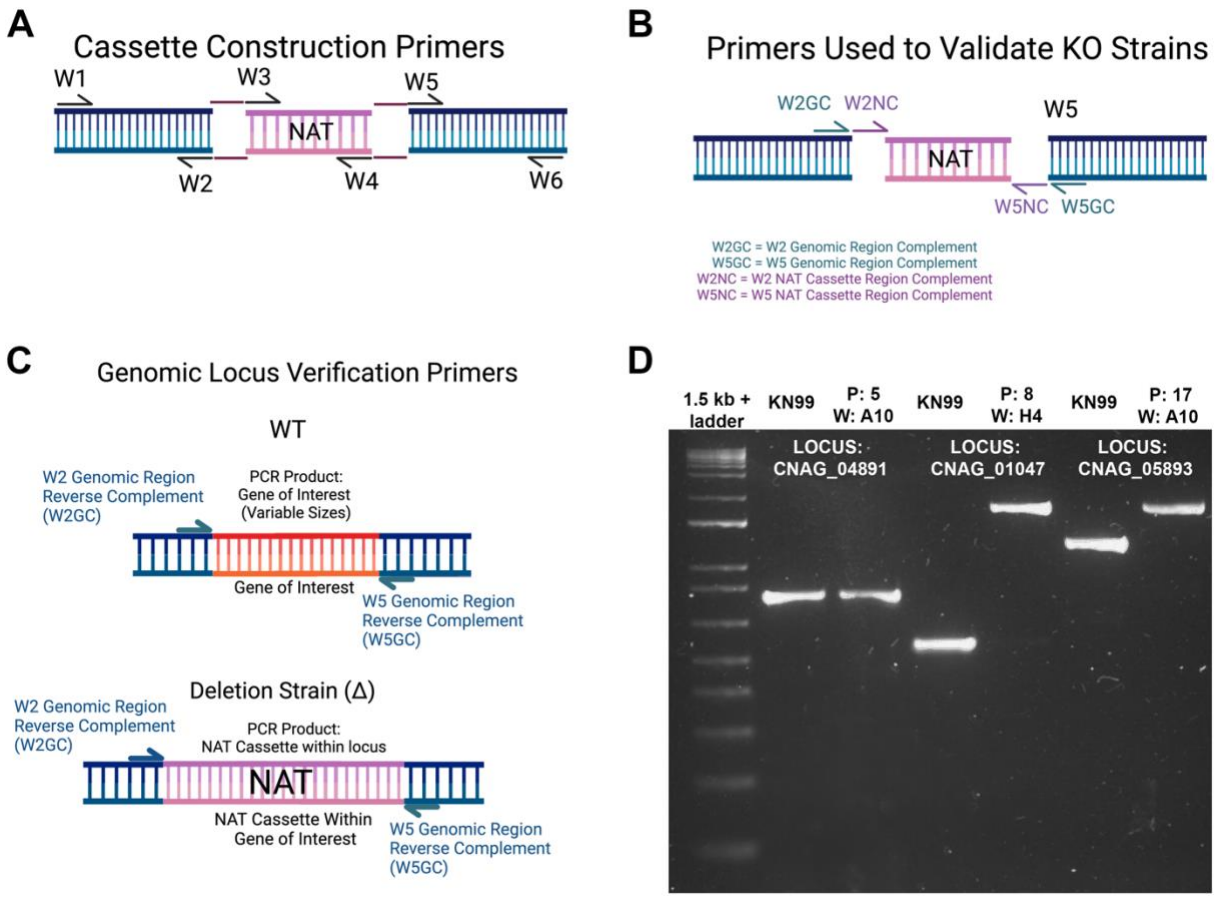
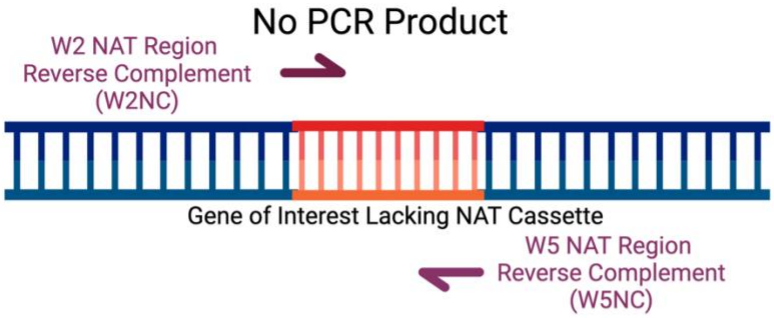


Figure 1 PCR validation of KO collection strains targeting the genomic locus
A) Deletion cassette construction adapted from [9]: W1, W2, W5, and W6 primers are publicly available for each strain on the FGSC website. W3 and W4 primers are not publicly available. Created with BioRender.
B) Strain verification schematic. The reverse complement of the genomic region of W2 (W2GC) and W5 (W5GC) were used to confirm deletion fidelity at each locus.
C) In WT KN99, these primer pairs amplify the gene of interest. In KO strains, these primer pairs will only amplify the NAT cassette.
C) Example genomic locus strain verification 1.5% agarose gel using the methodology in panel B. Primers targeted the reverse complement of the genomic locus specific sequence in the W2 (W2GC) and W5 (W5GC). K-denoted lanes contain WT (KN99) PCR products and Δ-denoted lanes contain putative deletion strain PCR products. Numbers refer to the CNAG locus corresponding to the W2 W5 sequence location. CNAG_04891*Δ* did not contain a NAT cassette within CNAG_04891, as the WT and deletion strain PCR products are the same size. Both CNAG_01047*Δ* and CNAG_05893*Δ* contain NAT cassettes in the expected locus.

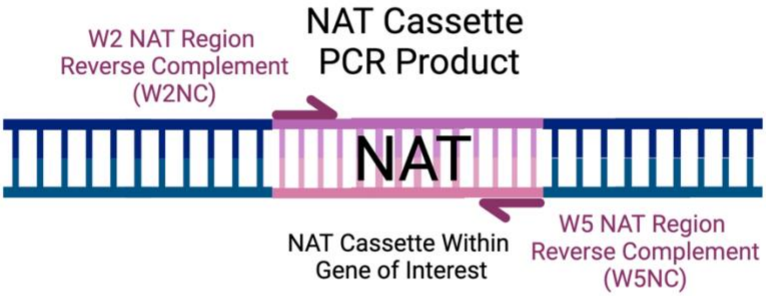
A

NAT Cassette Verification Primers

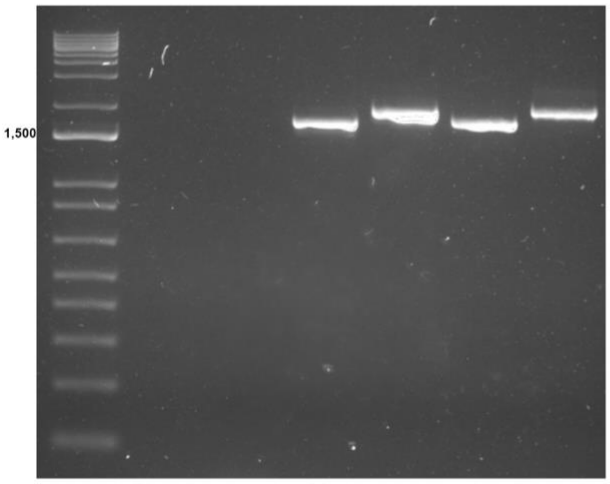
WT



Deletion Strain (Δ)



B



C

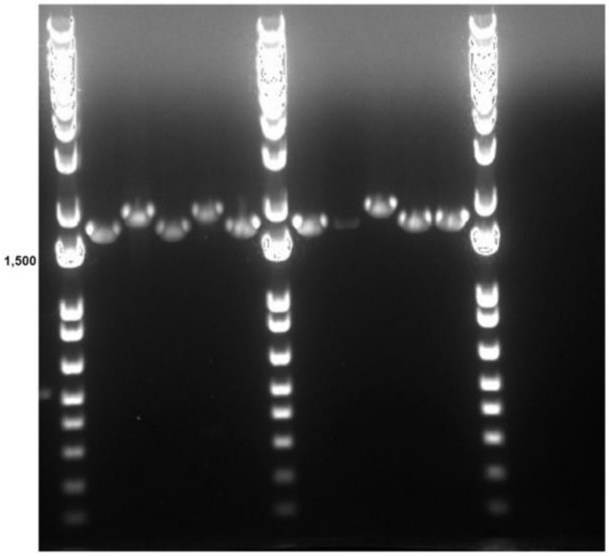


Figure 2: Comparing NAT cassette sizes between KO strains reveal size differences that correlate with locus-specific insertion failures

A) NAT cassette verification methodology. The reverse complement of the NAT cassette-specific region of W2 (W2NC) and W5 (W5NC) were used to confirm the presence of the NAT cassette in putative KO strains. In WT KN99α, these primer pairs will not amplify a product due to a lack of insertion cassette. In KO strains, these primer pairs will only amplify the NAT cassette in any KO strain as there are no genomic-locus specific base pairs.

B) NAT cassette size comparisons. KN99α and a no template control show lack of amplification. NAT cassettes PCR products from CNAG_02179*Δ* and CNAG_01019*Δ,* two strains that failed to have expected locus-specific disruptions, were larger in size than KO strains with correct KOs.

C) Amplification of NAT cassette sequences by PCR across KO strains. This further confirmed the larger NAT cassette size correlated with locus-specific PCR failures. Note the larger NAT cassette in CNAG_04981*Δ* corresponds to the locus-specific PCR failure in figure 2 panel C. Both CNAG_01047*Δ* and CNAG_05893*Δ* contain smaller NAT cassettes and passed the PCR test from figure 2 panel C. We found the differences in NAT cassette size to be predictive of a KO strain's locus-specific failure.

A

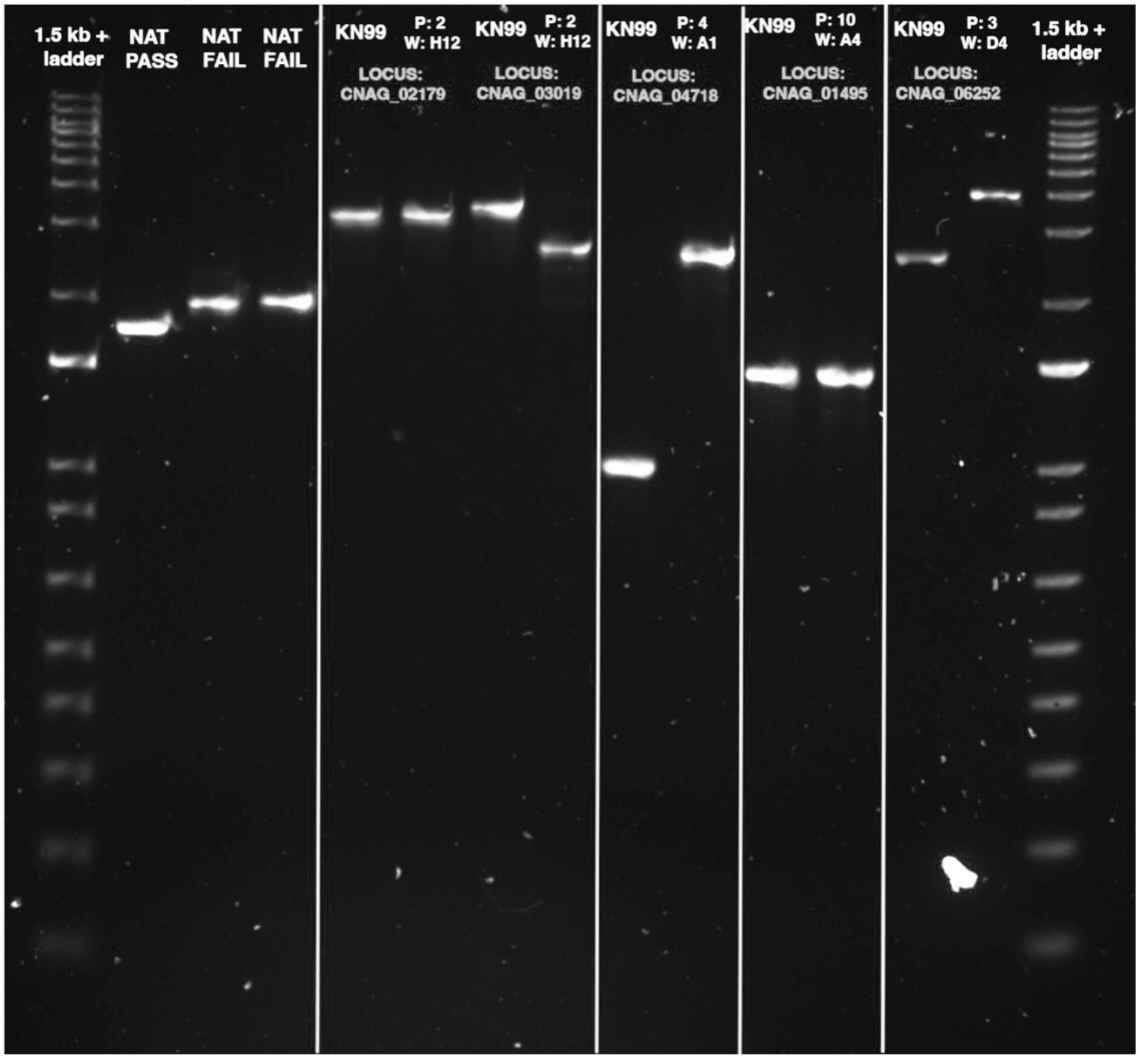


Figure Error! No text of specified style in document.**: Summary PCR and confirmation of 2015 and 2008 direct plate swaps.**

A) 1.5% agarose gel comparing PCR product sizes. L to R: 1 kb + ladder. The next three lanes are example NAT cassettes amplified using the strategy outlined in figure 2.
The following lanes contain PCR products amplified using the locus-specific strategy outlined in figure 1.
Labels indicate KN99α WT or 2015 plate and well position used for amplification. Locus labels indicate the genomic locus of the primer pairs.
Plate 2 well H12 should contain CNAG_02179Δ. However, the PCR product targeting CNAG_02179 is the same size for WT and the strain in plate 2 well H12, indicating that this gene is not disrupted with a NAT cassette. However, testing this same strain at

CNAG_03019, the locus that would be disrupted if the 2015 plate 2 were swapped with a 2008 plate 2, shows a PCR product size difference between KN99α WT and the strain, indicating that this strain is a 2008 KO of CNAG_03019, not a 2015 KO of CNAG_02179.

In the following two lanes, plate 4 well A1, the expected KO gene is CNAG_01019. We previously found that CNAG_01019 was not disrupted in this strain. Here we show that CNAG_04718, the locus that would be disrupted if plate 4 was directly swapped with a 2008 plate, shows a size difference between PCR products of KN99α WT and the strain in plate 4 well A1. This indicates that this strain is a 2008 KO of CNAG_04718, not a 2015 KO of CNAG_01019.

Finally, in the final two lanes, we previously found that the strain within plate 3 well D4 did not have a KO at the expected locus of CNAG_04981. Here we show that CNAG_06252, the locus that would be disrupted if plate 3 was directly swapped with a 2008 plate, shows a size difference between PCR products of KN99α WT and the strain in plate 3 well D4. This indicates that the strain in plate 3 well D4 is a 2008 KO of CNAG_06252, not a 2015 KO of CNAG_04981.

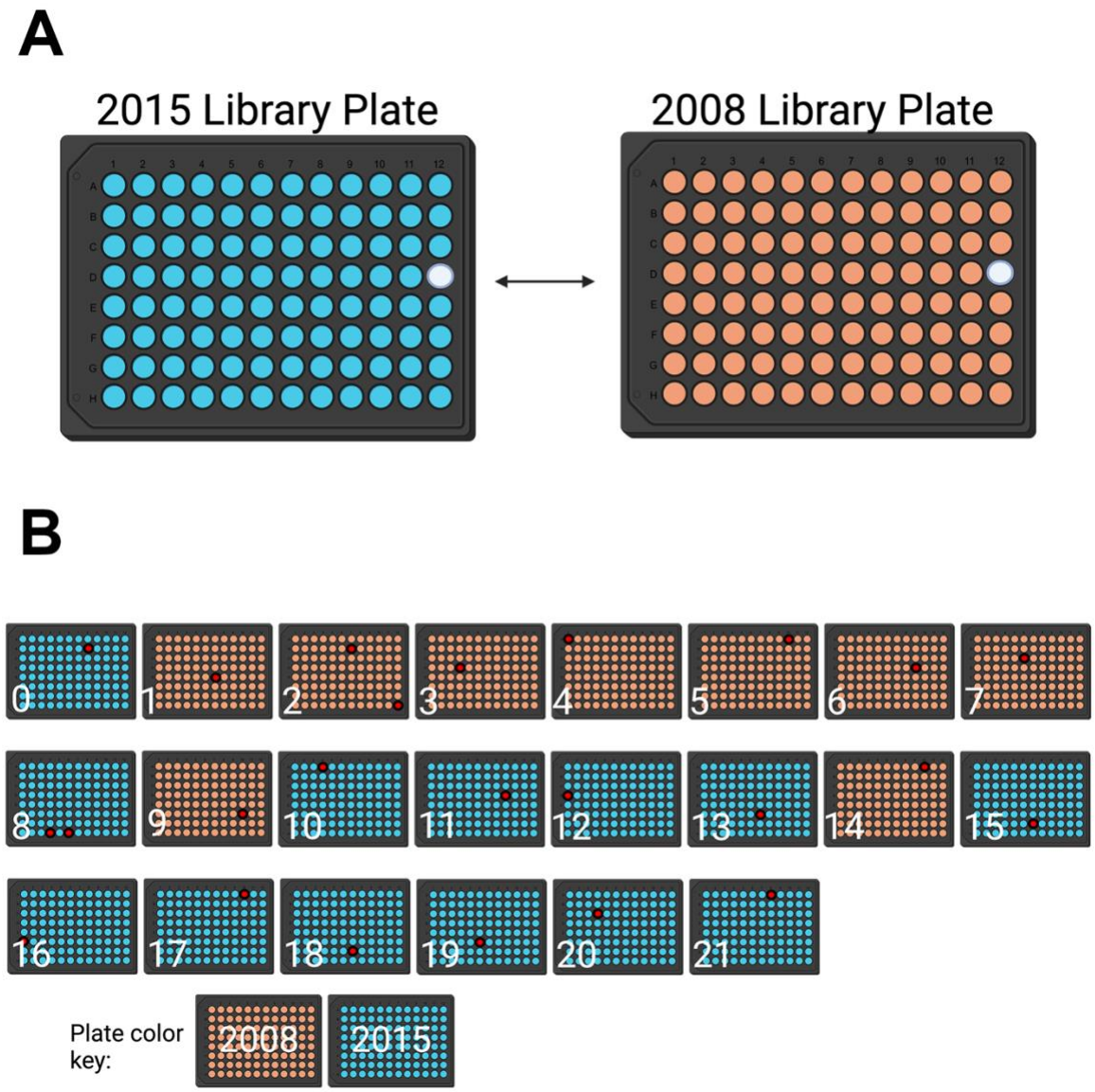


Figure 4 Plates 1-7, 9, and 14 are directly swapped
A) Based on our findings whole 2015 plates were directly swapped with 2008 plates, not individual wells **B)** Of the 22 plates in our 2015 KO collection, only the first 14 plates were impacted. Specifically, plates 1, 2, 3, 4, 5, 6, 7, 9, and 14 were directly swapped. Orange plates indicate 2008 swapped plates. Blue plates indicate 2015 library plates. Red wells indicate tested wells (Further information in Table 2)

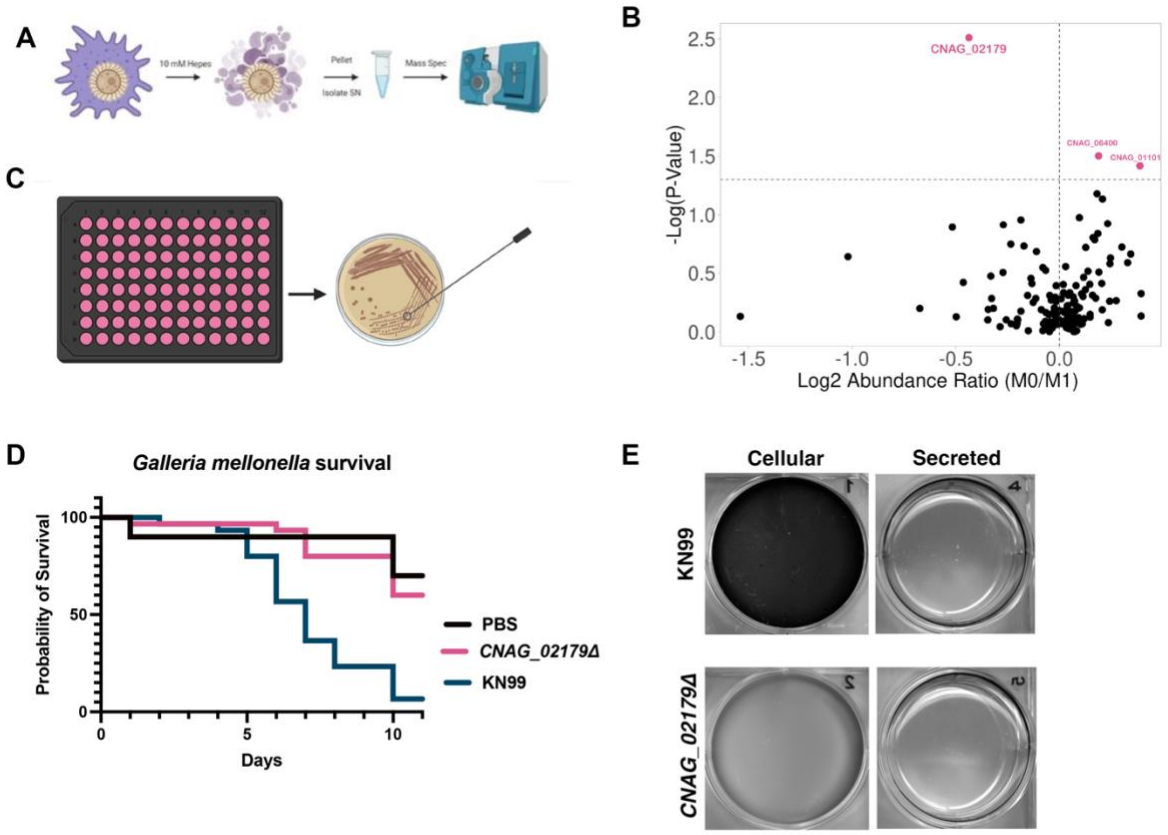


Figure 5 Characterization of proteomics candidates exhibited phenotypes similar to H99W.

A) Proteomics experiment schematic

B) Comparing protein abundance in M0 vs. M1 polarization states identified CNAG_02179 as a candidate of interest

C) CNAG_02179*Δ* was selected from the 2015 KO collection plate 2 well H12.

D) Virulence phenotypes of the putative CNAG_02179*Δ* strain mirrored H99W. Virulence in the *Galleria mellonella* model of infection (n=30/group, $1x10^5$ infection).

E) Melanization defects in the putative CNAG_02179*Δ* strain compared to KN99.

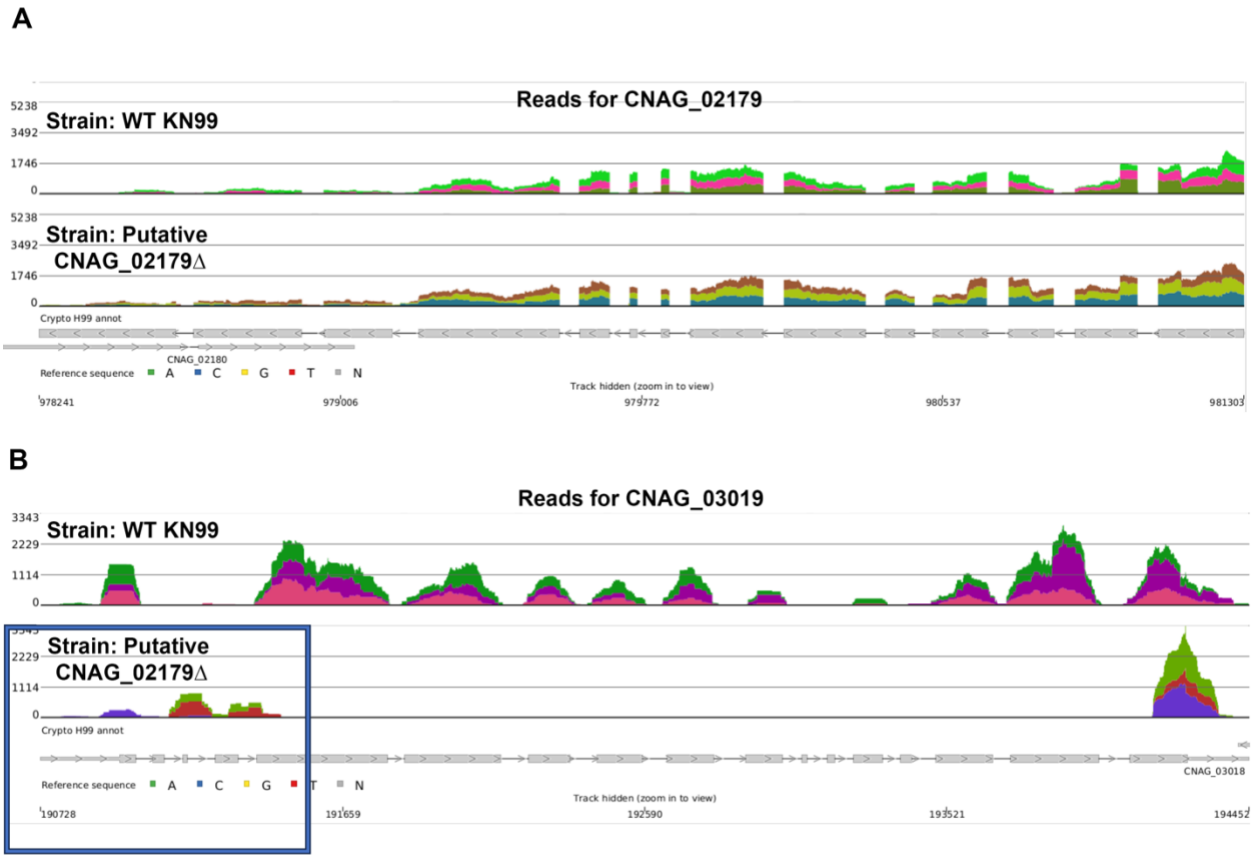


Figure 6: RNAseq reads for the putative CNAG_02179Δ strain revealed RNAseq reads were present for CNAG_02179 and absent at CNAG_03019 (the 2008 plate expected KO).

A) RNAseq data comparing the coverage of CNAG_02179 in the WT KN99 strain (top) to the strain in our 2015 library plate 4 well H12 strain (putative CNAG_02179Δ) (bottom). Expression levels were identical in the WT and KO strain confirming no disruption in the expected locus.

B) The RNAseq reads for CNAG_03019, the gene that would be knocked out if the plate was swapped directly with the 2008 library plate 4 well H12 strain in WT (top) and plate 4 well H12 strain (putative CNAG_02179Δ) confirming CNAG_03019 is the actual knocked out gene.

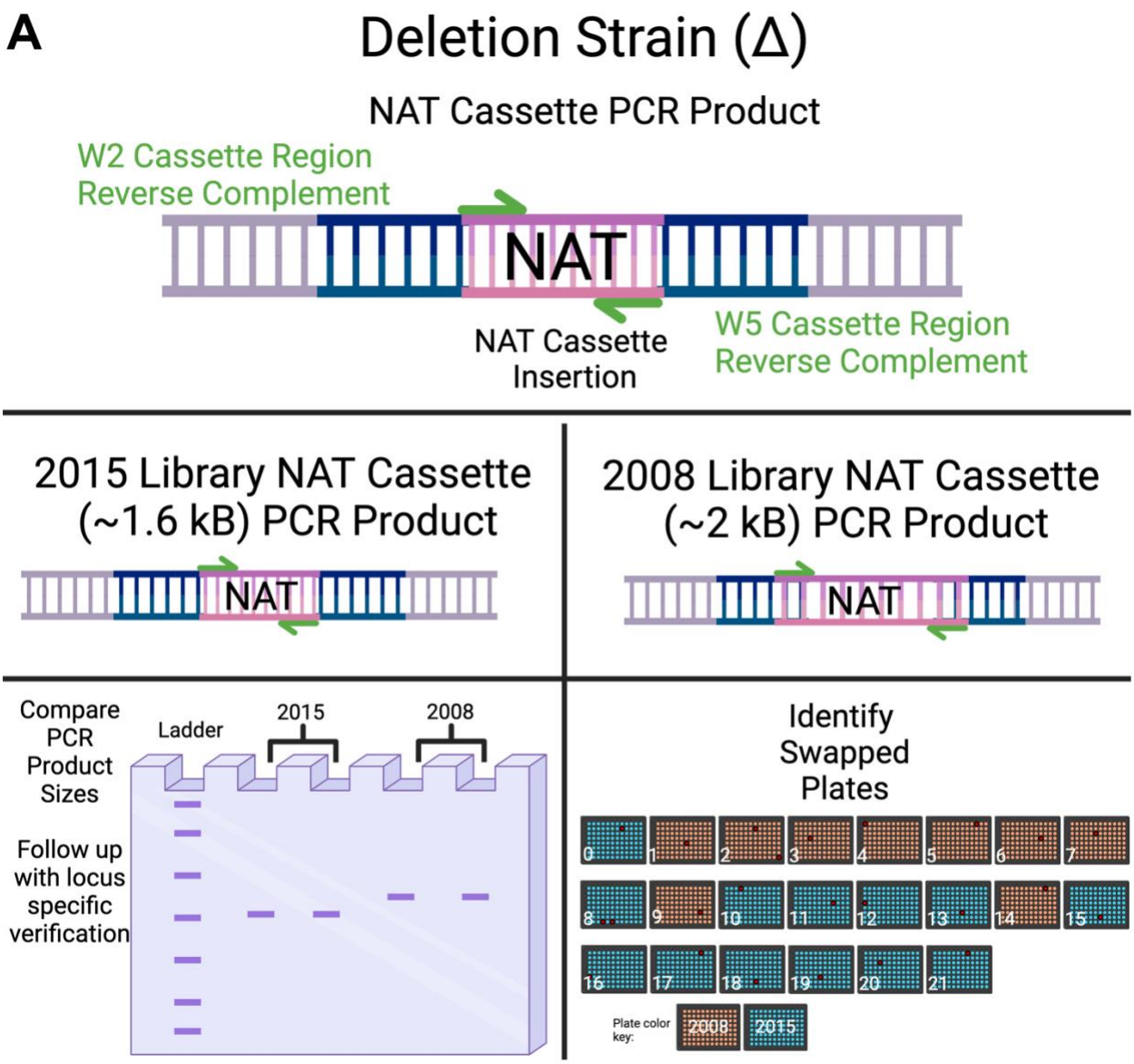


Figure 7: Graphical summary of library fidelity methodology
A) In the process of uncovering the swapped plates within our 2015 KO collection, we found PCR amplifying 2008 library NAT cassettes and 2015 library NAT cassettes yielded different insertion sizes when compared by gel electrophoresis (~1.8 kb vs ~1.7 kb). By utilizing NAT cassette amplifying primers, strains can be compared regardless of genomic locus insertion site. This cost-effective simple method allows for comparisons across library plates to determine if a swap occurred using the same primers. Further verification using genomic DNA specific primers is necessary to fully confirm a swap.

Tables:

	W2 NAT Region RC (W2NC)	W5 NAT Region RC (W5NC)
NAT Cassette Primers (5->3)	gttggatccgctgctaggcgcgccgtg	gttggatgcagggatgcggccgctgac

Table 1: NAT cassette PCR primers.

CNAG	2015 Library Plate	2015 Library Well	Locus-Specific Verification	NAT Cassette Predicted Status	NAT Cassette Size
CNAG_06104Δ	0	B8	KO PRESENT	SMALLER (2015)	~1.7 kb
CNAG_02435Δ	1	E7	NT	LARGER (2008)	~2 kb
CNAG_02179Δ	2	H12	KO ABSENT	LARGER (2008)	~2 kb
CNAG_02801Δ	2	B7	KO ABSENT	LARGER (2008)	~2 kb
CNAG_04981Δ	3	D4	KO ABSENT	LARGER (2008)	~2 kb
CNAG_01019Δ	4	A1	KO ABSENT	LARGER (2008)	~2 kb
CNAG_04891Δ	5	A10	KO ABSENT	LARGER (2008)	~2 kb
CNAG_02685Δ	6	D9	NT	LARGER (2008)	~2kb
CNAG_04122Δ	7	C6	NT	LARGER (2008)	~2 kb
CNAG_01047Δ	8	H4	KO PRESENT	SMALLER (2015)	~1.7 kb
CNAG_03019Δ	8	H6	NT	SMALLER (2015)	~1.7 kb
CNAG_00306Δ	9	F10	NT	LARGER (2008)	~2 kb

CNAG_04388Δ	10	A4	KO PRESENT	SMALLER (2015)	~1.7 kb
CNAG_04321Δ	11	D9	KO PRESENT		
CNAG_07638Δ	12	D1	KO PRESENT	SMALLER (2015)	~1.7 kb
CNAG_02489Δ	13	F7	NT	SMALLER (2015)	~1.7 kb
CNAG_05592Δ	14	A10	NT	LARGER	~2 kb
CNAG_02503Δ	15	G7	KO PRESENT	SMALLER (2015)	~1.7 kb
CNAG_03729Δ	16	F1	NT	SMALLER (2015)	~1.7 kb
CNAG_05893Δ	17	A10	KO PRESENT	SMALLER (2015)	~1.7 kb
CNAG_00064Δ	18	G7	NT	SMALLER (2015)	~1.7 kb
CNAG_06384Δ	19	F6	NT	SMALLER (2015)	~1.7 kb

CNAG_00006Δ	20	C4	NT	SMALLER (2015)	~1.7 kb
CNAG_00308Δ	21	A8	NT	SMALLER (2015)	~1.7 kb

Table 2: 2015 KO collection strains tested using a mix of locus specific primers and NAT cassette size differences.
NT indicates not tested.

CNAG	2015 Plate	2015 Well	Locus-Specific F Primer	Locus-Specific R Primer
CNAG_06104Δ	0	B8	CTGCATCCCCATCCATCACATC	GGTCAGTGACTTTCGACATCTC
CNAG_02179Δ	2	H12	AACATCGTATAGTACAGCAAAC	AAAATTCCGTGGGAATATGTAA
CNAG_02801Δ	2	B7	CCACATCACACAAGACATCAAA	CATGCACGATCAATCTATGAAA
CNAG_04981Δ	3	D4	CTGCGATTCTAAGCCGGTCAAT	AACATTCCTTGAAATCAACCGC
CNAG_01019Δ	4	A1	TCAGGAACTCTATCTAATCGAA	CGAATATGTAGTAACGGTGCGC
CNAG_04891Δ	5	A10	GAATCAAATAATAGACCATACACTTACAGTCATA	CACACTCTCCTCCGAGCATCAT
CNAG_01047Δ	8	H4	ACGAGCTAGAATCATATAAGTTCAGAAAC	ACAATCCACTAACCGATCAAGG
CNAG_03019Δ	8	H6	CTGTATTCTGCCCTAACCCG	CAACAATTTCCTAACGGTGACC
CNAG_04388Δ	10	A4	TTACACAGATCTCACTTCCAAA	GTTTACACATTGTTCATAAAGC
CNAG_04321Δ	11	D9	ATTACTATCCATTCATCCTACA	TTCCCAAAGCAACAAGACTAGC
CNAG_07638Δ	12	D1	ATCCCATACGCCCCAGCAAGAA	TACACATTGCAAGAAGACAGAAAAAAAAAGTGACT
CNAG_02503Δ	15	G7	TTACATTCCACACTAATCCAAC	CTCCATATTACAGCTAATGCCT
CNAG_05893Δ	17	A10	CATTACACTCGCACTCGCCATA	ACGGAATTATAATCTCTTCTAGAGAGAC

Table 3: Genomic locus primers