RapidKO Production Analysis

Benjamin David

2024-04-25

Production Summary

Gene Knockouts

We attempted to create 2435 knockouts in *S. mutans*. In the following summary table, I define a successful knockout as observing both visible colonies and visible growth in liquid culture. If there is any doubt, I classify the knockout as unsuccessful.

	Knockouts	Attempts	Percent
Annotated Genes	1643	1966	83.6
Intergenic Regions	426	469	90.8
Total	2069	2435	85.0

The Quivey library targeted 1961 loci; however, they only attemped 1413 knockouts due to technical constraints. Their final library consisted of 1112 successful knockouts. The following table is a breakdown of how our library compares to the Quivey library for the subset of loci the Quivey library attempted.

Subset	Mutants
Both	993
Us Only	198
Quivey Only	119
Neither	103
Total	1413

Operon Modulation

We attempted to create `1728` operon modulated strains, `864` upregulatd and `864` downregulated, and successfully made `1564` of them. The table below shows the breakdown by modulation type:

	Down Grow	Down No-grow	**Total**
Up Grow	716	115	831
Up No-grow	17	16	33
Total	733	131	864

Technical Analysis

Failures by Enzyme

One technical concern we had in the production of the library was that some enzymes may result in higher than expected failure rates. Given a global failure rate of 0.127, we tested whether the observed failures are independent of the golden gate enzyme used. If the failures are independent of the enzyme used, we would expect the distribution of failures to follow the distribution of enzymes used in the library. We compared this expected distribution to the observed distribution of failures using a χ^2-goodness of fit test. As would seem fairly obvious just by looking at the data, the computed χ^2 test statistic suggests that the distribution of failures by enzyme follows the expected distribution (χ^2 value: 2.29, p-value: 0.515).

Enzyme	Counts	Observed Failures	Expected
BbsI	341	48	43.4
BsaI	3692	468	470.0
BsmBI/Esp3I	122	12	15.5
PaqCI	8	2	1.0

We were also concerned about the payload cassette used. The payload was determined both by the enzyme used and the gene's orientation in the genome. There is perhaps less of an argument to use the same independence assumption about the payloads as there is the enzymes. For example, if there is an entire operon that is essential on a particular strand, we would not expect failures by payload to be independent. Since I cannot think of a better alternative, I used the same hypothesis for the payloads. The computed χ^2 test statistic was 97.87, p-value: 0.

Payload	Counts	Observed Failures	Expected
KD_Payload_1	364	39	46.3
KD_Payload_2	398	71	50.7
KD_Payload_3	44	9	5.6
KD_Payload_4	33	7	4.2
KD_Payload_5	12	4	1.5
KD_Payload_6	13	1	1.7
KU_Payload_1	364	12	46.3
KU_Payload_2	398	17	50.7
KU_Payload_3	44	2	5.6
KU_Payload_4	33	1	4.2
KU_Payload_5	12	1	1.5
KU_Payload_6	13	0	1.7
Payload_1	896	154	114.1
Payload_2	1272	175	161.9
Payload_3	90	18	11.5
Payload_4	97	11	12.3
Payload_5	31	4	3.9
Payload_6	41	2	5.2
Payload_7	4	0	0.5
Payload_8	4	2	0.5