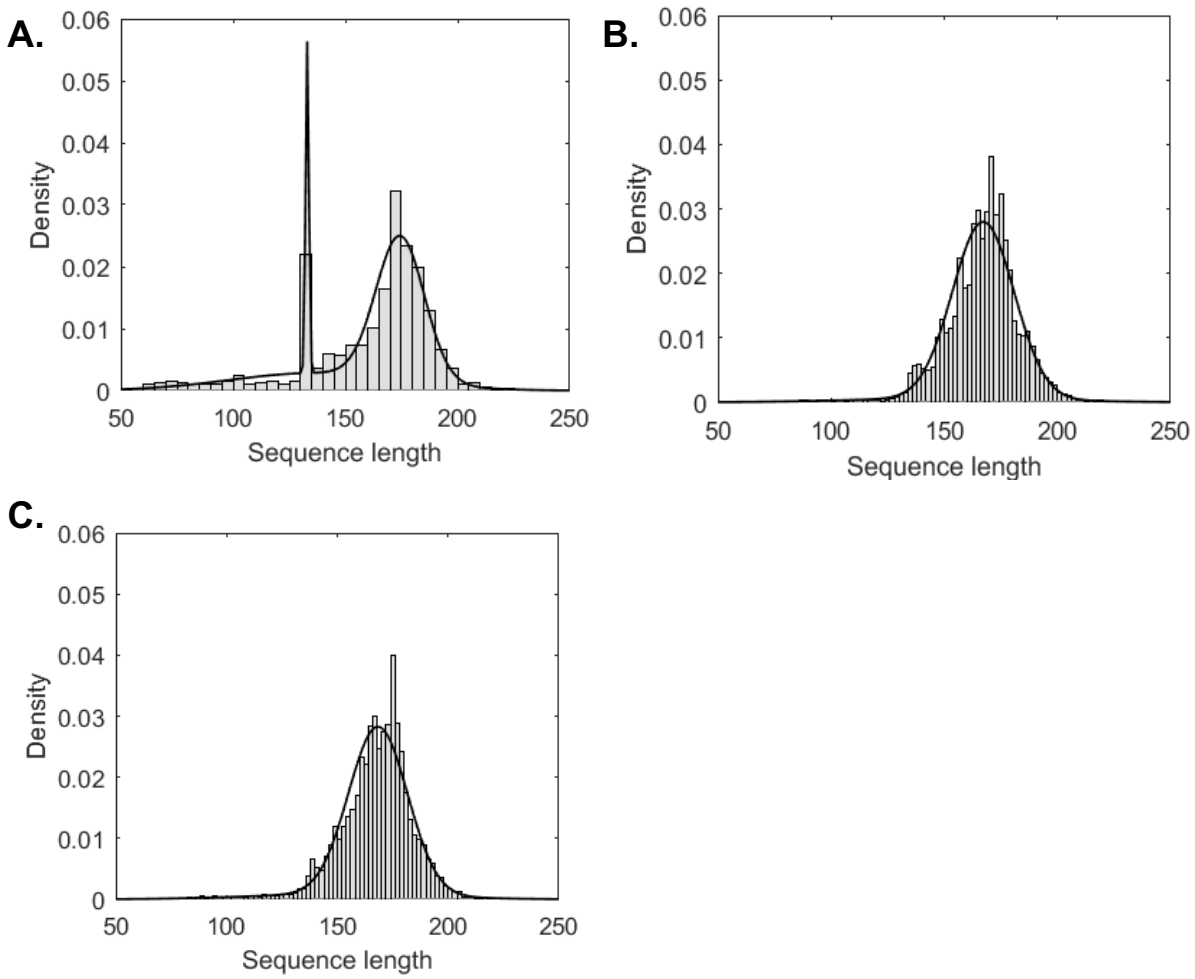


FIG S1 Two- or three-component Gaussian mixture models (GMMs) fitted to the distribution of sequence lengths resulting from a jackhmmer search of the UniProtKB database for homologs to the wild-type amidase domain of LysEFm5. (A) The taxonomy in the search was restricted to firmicutes only (three-component GMM). (B) The taxonomy in the search was restricted to bacteria only (two-component GMM). (C) The taxonomy in the search was not restricted (two-component GMM).

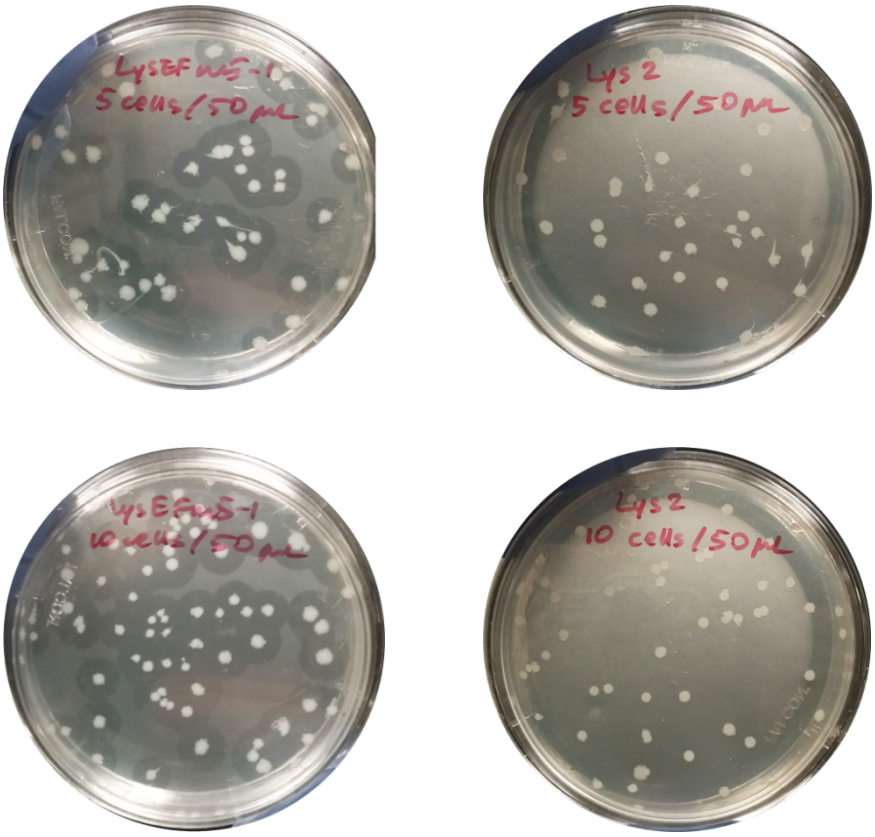


FIG S2 *E. coli* expressing a lysin with alternative specificity do not form halos on LB+kan/VRE/IPTG plates. *E.coli* containing a plasmid encoding LysEFm5 (left) and Lys2, a *C. perfringens*-specific lysin (right) were plated in an identical manner on LB+kan/VRE/IPTG plates at two densities (top: 0.1 cells/µL; bottom: 0.2 cells/µL). After 19 - 20 hr of incubation at 37°C, halos were observed on the LysEFm5 plates, but not on the Lys2 plates.

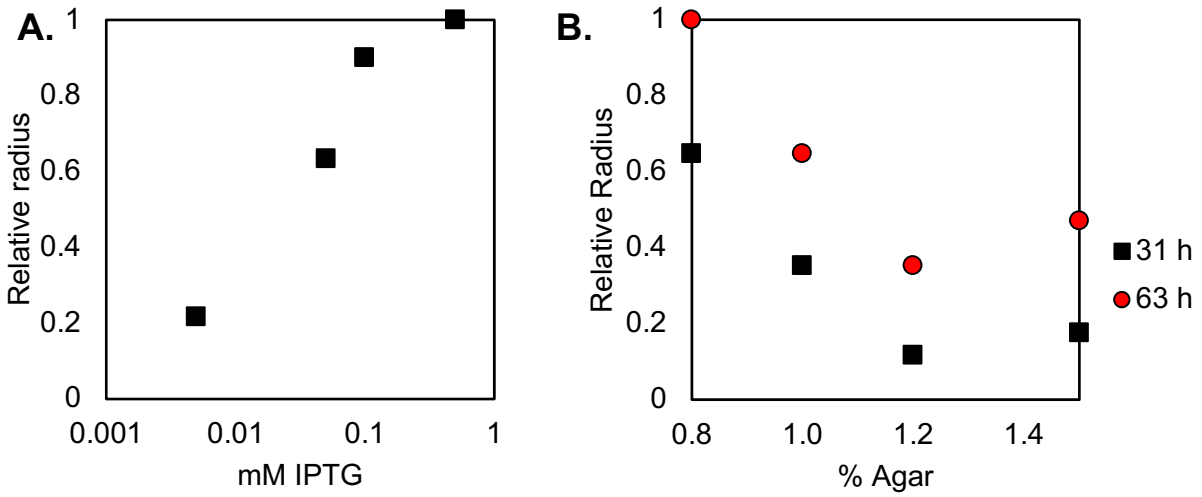
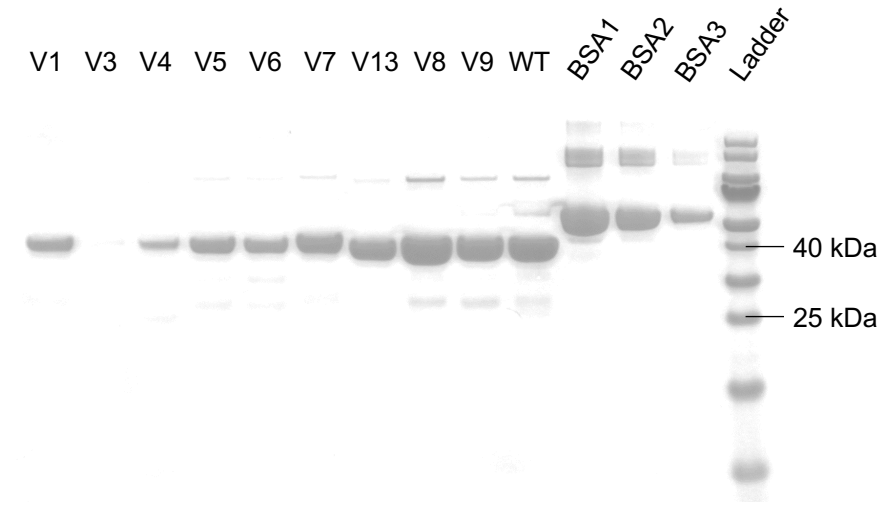
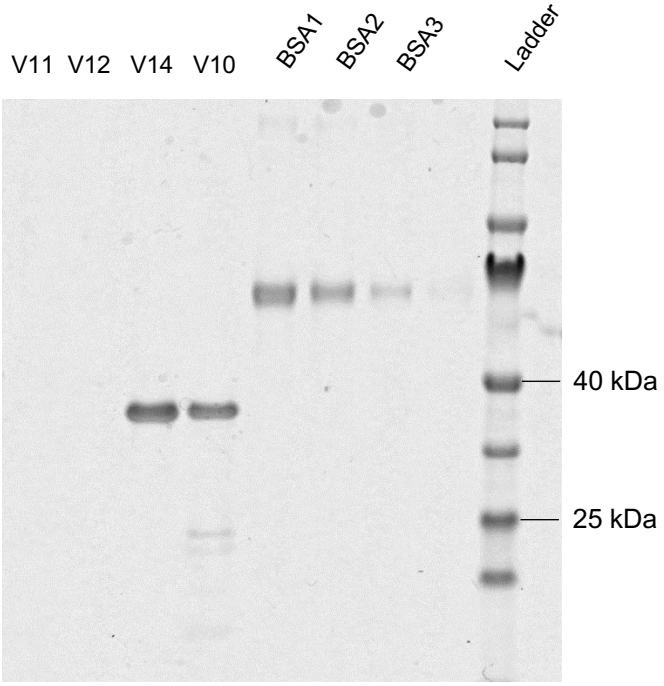


FIG S3 Effect of IPTG, agar content, and time on halo radius. Expression-competent *E. coli* containing pET:LysEFm5 were plated on LB + kan agar plates made with 2 w/v% LB, 50 µg/mL kan, 0.05 v/v% of stock autoclaved VRE (~0.3 g/mL cell material suspended in PBS), deionized water, and (A) 0.005, 0.05, 0.1, and 0.5 mM IPTG (and 1.0 w/v% agar) or (B) 0.8, 1.0, 1.2, and 1.5 w/v% agar (and 0.05 mM IPTG) in 15 mL of total volume. Measurements were taken 19 hr after the start of incubation at 37°C for the IPTG-variable plates, and 31 and 63 hr after the start of incubation at 37°C for the agar percentage-variable plates. The radius of the halo around each colony was measured and standardized according to the largest radius measured in that group. For both groups, *n* = 1, as this experiment was conducted only to determine an appropriate set of assay conditions. A direct relationship between IPTG concentration/time and halo size, and an inverse relationship between agar w/v% and halo size was observed. It was determined that a final IPTG concentration of 0.05 mM, 1.0 w.v% agar, and an incubation time of 19 hr or greater was sufficient for the purposes of this assay.



Variant	Purity
1	0.99
3	N/A
4	0.97
5	0.97
6	0.96
7	0.98
8	0.90
9	0.93
13	0.99
WT	0.91



Variant	Purity
10	0.85
11	N/A
12	N/A
14	1.00

FIG S4 SDS-PAGE used to determine variant concentrations. The bands of variants at ~37 kDa correspond to the expected size of LysEFm5.

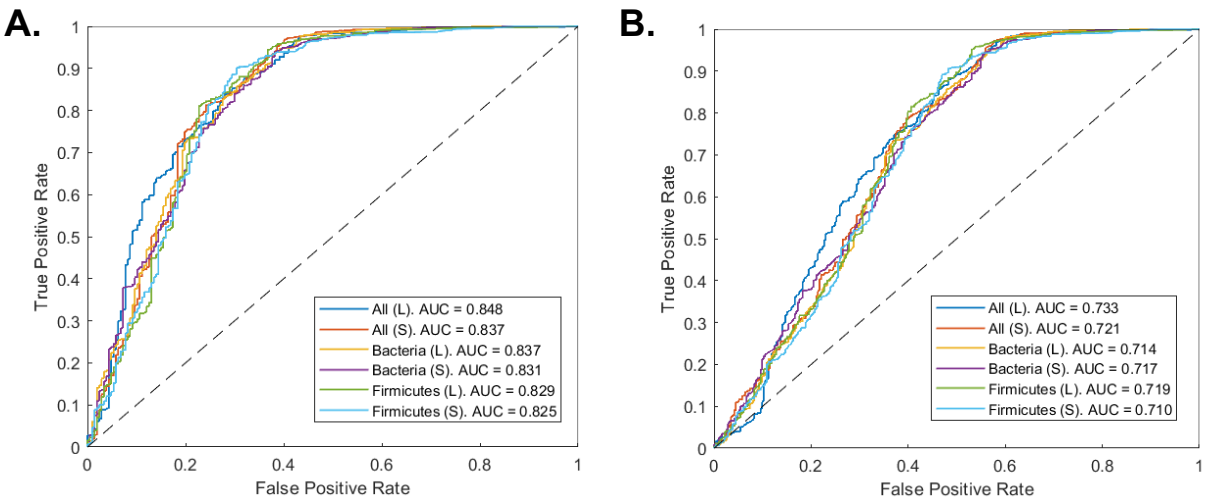


FIG S5 Excluding sequences that lack a majority of inactive or active designations improves the classification accuracy of the statistical fitness. ROC curves that quantify the binary classification ability of the statistical fitness parameter are plotted for each starting MSA (Table 1), provided that all 1731 sequences experimentally read at least 100 times are considered in the analysis (including those with the same number of active and inactive observations). (A) Results when sequences with an active fraction of 0.5 are considered active. The AUC ranges from 0.825 – 0.848, depending on the starting MSA. (B) Results when sequences with an active fraction of 0.5 are considered inactive. The AUC ranges from 0.710 – 0.733. Because on the order of hundreds, and up to a thousand, individual colonies were plucked per library, it was expected that there would be contamination from neighboring cell material some fraction of the time. Therefore, observing the same number of active and inactive observations was ultimately attributed to the contamination of one or more bins as a result of human error, leading to the dismissal of these sequences from the analysis. When sequences with an active fraction of 0.5 are excluded entirely, the AUC ranges from 0.840 – 0.894 (Fig. 8).

TABLE S1 List of reactions and primers used to create NNK libraries 1-9

Lib.	Rxn	Pos. 1	Pos. 2	Pos. 3	Fragment size [bp]	Primer 1	Primer 2
1	1-1	33	40		132	PRIM01	PRIM03
	1-2				936	PRIM12	PRIM02
2	2-1	40	47		153	PRIM01	PRIM04
	2-2				915	PRIM13	PRIM02
3	3-1	45	87		147	PRIM01	PRIM05
	3-2				150	PRIM21	PRIM24
	3-3				774	PRIM14	PRIM02
4	4-1	32	38		126	PRIM01	PRIM06
	4-2				939	PRIM15	PRIM02
5	5-1	33	45		147	PRIM01	PRIM07
	5-2				936	PRIM16	PRIM02
6	6-1	47	91		153	PRIM01	PRIM08
	6-2				156	PRIM17	PRIM22
	6-3				762	PRIM25	PRIM02
7	7-1	34	35		117	PRIM01	PRIM09
	7-2				933	PRIM18	PRIM02
8	8-1	74	83		261	PRIM01	PRIM10
	8-2				813	PRIM19	PRIM02
9	9-1	33	40	87	132	PRIM01	PRIM11
	9-2				186	PRIM20	PRIM23
	9-3				774	PRIM26	PRIM02

TABLE S2 NNK primer sequence identities

Primer Name	Library	Sequence identity (5’ → 3’)[a]
PRIM01	1	AAGAAGGAGATATACATATGGTTGAG
PRIM02	1	CAGTGATGATGGTGATGGTGGCATCCNNNNNNNN NNTTATTAATGGTGGTGATGGTG
PRIM03	1	CGCCGCAAGGCGMNNTGCTTCTTGTTTTGCAGTM NNATTACCCCAAGT
PRIM12	1	ACTTGGGGTAATNNKACTGCAAAACAAGAAGCAN NKCGCCTTGCGGCG
PRIM04	2	GGCCAGCTGGTTMNNATTCATCGCCGCAAGGCG MNNTGCTTCTTGTTT
PRIM13	2	AAACAAGAAGCANNKCGCCTTGCGGCGATGAATN NKAACCAGCTGGCC
PRIM05	3	GGTTATTATTMNNCGCCGCAAGG
PRIM14	3	TGGTAATATGAACTATNNKGGATATGAAGTCTGTG
PRIM21	3	CCTTGCGGCGNNKAATAATAACC
PRIM24	3	CACAGACTTCATATCCMNNATAGTTCATATTACCA
PRIM06	4	AAGGCGAGTTGCMNNTTGTTTTGCAGTTGAMNNA CCCCAAGTATT
PRIM15	4	AATACTTGGGGTNNKTCAACTGCAAAACAANNKGC AACTCGCCTT
PRIM07	5	GGTTATTATTMNNCGCCGCAAGGCGAGTTGCTTC TTGTTTTGCAGTMNNATTACCCCAAG
PRIM16	5	AATACTTGGGGTNNKTCAACTGCAAAACAANNKGC AACTCGCCTT
PRIM08	6	CAGCTGGTTMNNATTCATCGCC
PRIM17	6	GGCGATGAATNNKAACCAGCTG
PRIM22	6	CGTTGCCACAMNNTTCATATCCG
PRIM25	6	CGGATATGAANNKTGTGGCAACG

PRIM09	7	TGCTTCTTGTTTMNNMNNTGAATTACCCCA
PRIM18	7	TGGGGTAATTCANNKNNKAAACAAGAAGCA
PRIM10	8	GATATAGTTCATMNNACCATCGCCATTGGCAGTGT GCCAMNNACCATTGAACGT
PRIM19	8	ACGTTCAATGGTNNKTGGCACACTGCCAATGGCG ATGGTNNKATGAACTATATC
PRIM11	9	CGCCGCAAGGCGSNNTGCTTCTTGTTTTGCAGTG NBATTACCCCAAGT
PRIM20	9	ACTTGGGGTAATVNCACTGCAAAACAAGAAGCAN NSCGCCTTGCGGCG
PRIM23	9	GACTTCATATCCTAGATAGTTCATATT
PRIM26	9	AATATGAACTATCTAGGATATGAAGTC

[a]N = G, T, A, or C; V = G, T, or A; B = G, T, or C; M = A or C; R = A or G; W = A or T; S = G or C; K = G or T.

TABLE S3 High-throughput sequencing primer identities.

Primer Name	Sequence identity (5' → 3')[a]
FA1	TTTCCCTACACGACGCTCTTCCGATCTNNNNGTCGTTTTTCATAA TACTTGGGGT
FA2	TTTCCCTACACGACGCTCTTCCGATCTNNNNNGTCGTTTTTCATA ATACTTGGGGT
FA3	TTTCCCTACACGACGCTCTTCCGATCTNNNNNNGTCGTTTTTCAT AATACTTGGGGT
FA4	TTTCCCTACACGACGCTCTTCCGATCTNNNNNNNGTCGTTTTTCA TAATACTTGGGGT
FA5	TTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNGTCGTTTTT CATAATACTTGGGGT
RA1	GTTCAGACGTGTGCTCTTCCGATCTNNNNAGTCTGATCGTTGCC ACA
RA2	GTTCAGACGTGTGCTCTTCCGATCTNNNNNAGTCTGATCGTTGC CACA
RA3	GTTCAGACGTGTGCTCTTCCGATCTNNNNNNAGTCTGATCGTTG CCACA
RA4	GTTCAGACGTGTGCTCTTCCGATCTNNNNNNNAGTCTGATCGTT GCCACA
RA5	GTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNAGTCTGATCGT TGCCACA
FB	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA CGCTCTTCCGATCT
RB1	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT
RB2	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCT
RB3	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT
RB4	CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT
RB5	CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT
RB6	CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCT

[a]N = G, T, A, or C.