Supporting Information for

Tools

Screening antimicrobial peptides and probiotics using multiple deep learning and directed evolution strategies

Yu Zhang[a,†], Li-Hua Liu[a,b,†], Bo Xu[c,†], Zhiqian Zhang[a,†], Min Yang[a,†], Yiyang He[d,†], Jingjing Chen[e,†], Yang Zhang[a], Yucheng Hu[a], Xipeng Chen[a], Zitong Sun[a], Qijun Ge[a], Song Wu[a], Wei Lei[a], Kaizheng Li[a], Hua Cui[a], Gangzhu Yang[a], Xuemei Zhao[a], Man Wang[e], Jiaqi Xia[f,*], Zhen Cao[e,*], Ao Jiang[a,*], Yi-Rui Wu[a,*]

[a]*Tidetron Bioworks Technology (Guangzhou) Co., Ltd., Guangzhou Qianxiang Bioworks Co., Ltd., Guangzhou 510000, China*

[b]*Biology Department and Institute of Marine Sciences, College of Science, Shantou University, Shantou 515063, China*

[c]*School of Basic Medical Sciences, Hubei University of Science and Technology, Xianning 437100, China*

[d]*School of Education, Jianghan University, Wuhan 430056, China*

[e]*Yeasen Biotechnology (Shanghai) Co., Ltd., Shanghai 200000, China*

[f]*School of Basic Medicine, Jiamusi University, Jiamusi 154000, China*

[†]These authors made equal contributions to this work.

*Corresponding authors.

E-mail addresses: wuyirui@tidetronbio.com (Yi-Rui Wu), jiaqixia@whu.edu.cn (Jiaqi Xia Xia), jiangao@tidetronbio.com (Ao Jiang).

Received 4 January 2024; received in revised form 25 March 2024; accepted 6 May 2024

This PDF file includes:

Supplementary Figures. S1 to S11
Supplementary Tables S1 and S3
Legends for Supplementary tables S2, S4, and to S5

Other Supplementary Material for this manuscript includes the following:

Supplementary Tables S2, S4, and to S5

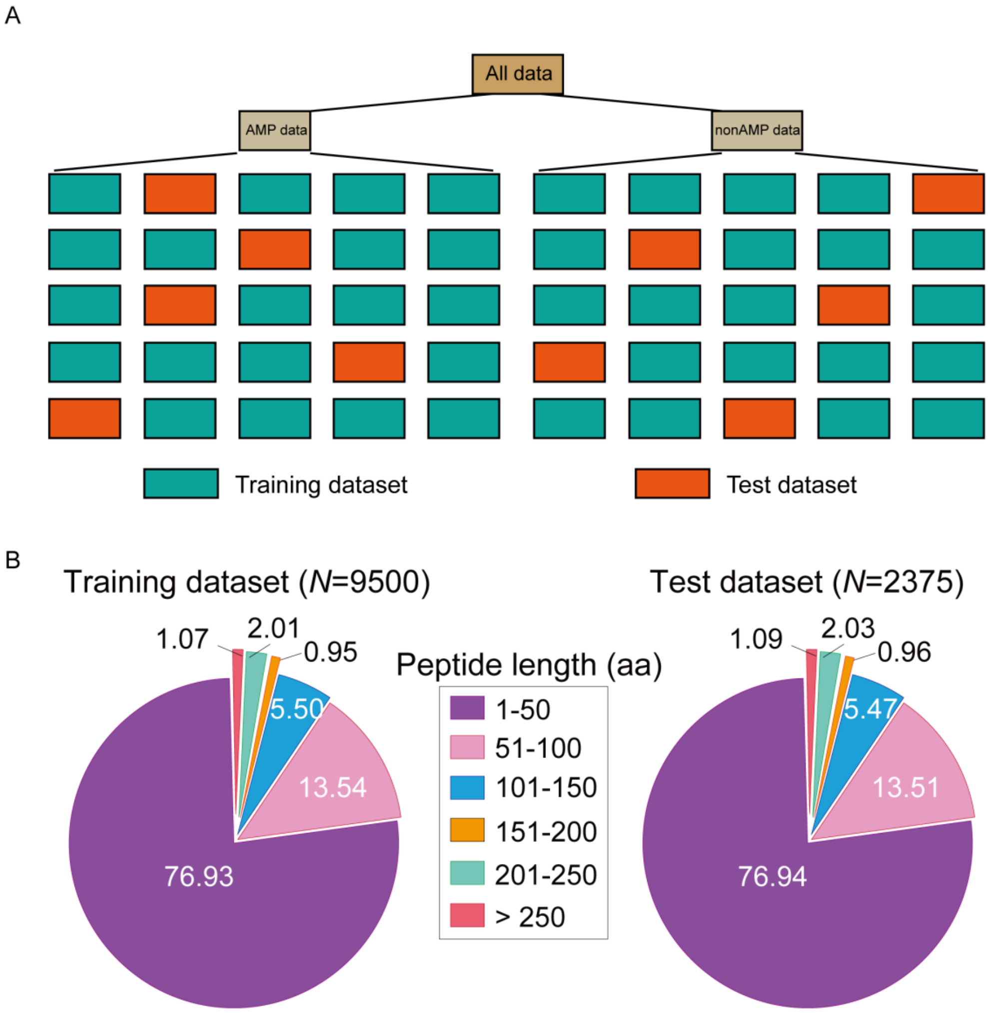

Figure S1 Dataset processing of training and test data. (A) All data was divided into training and test datasets by 4:1 proportion. (B) The length distribution of peptides in training and test datasets.

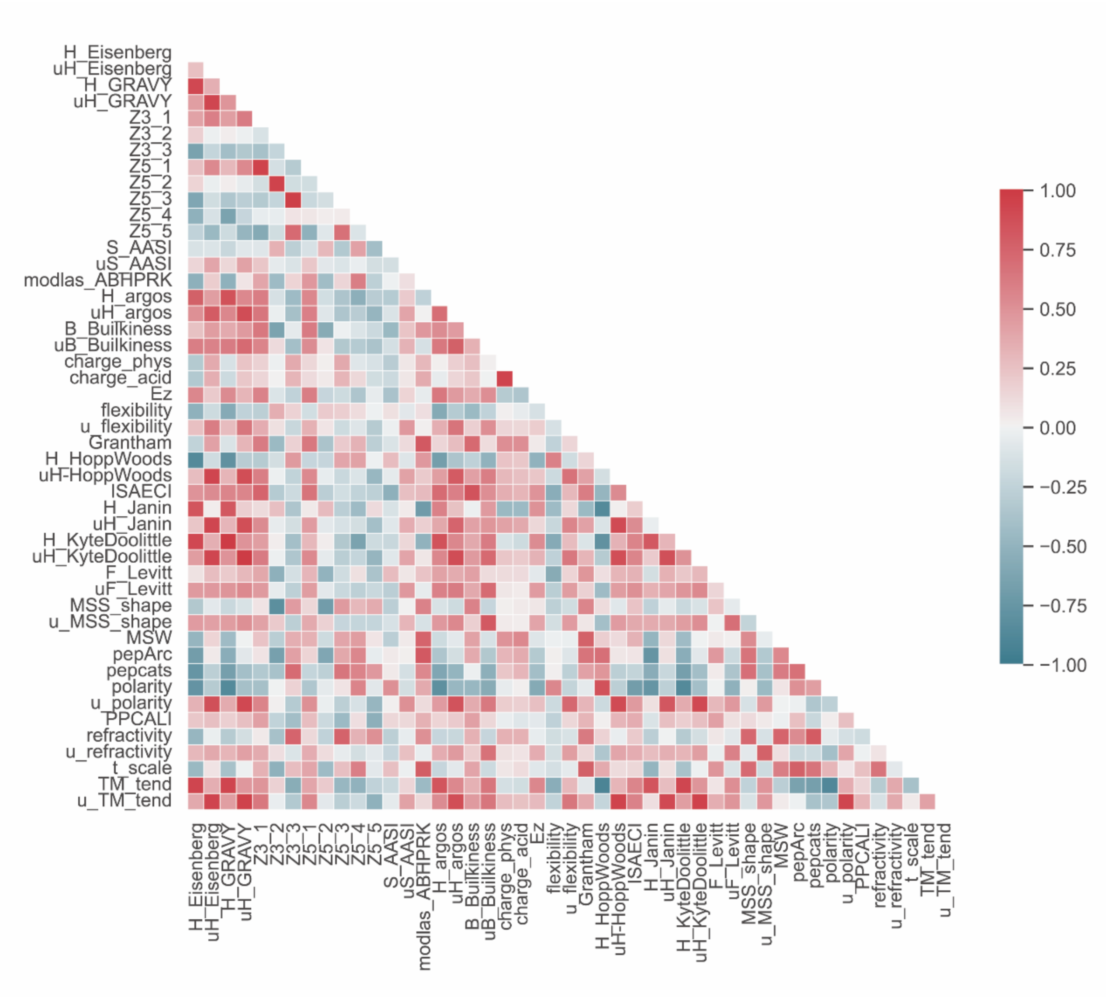

Figure S2 The correlation of AMPs characterization in COMDEL model.

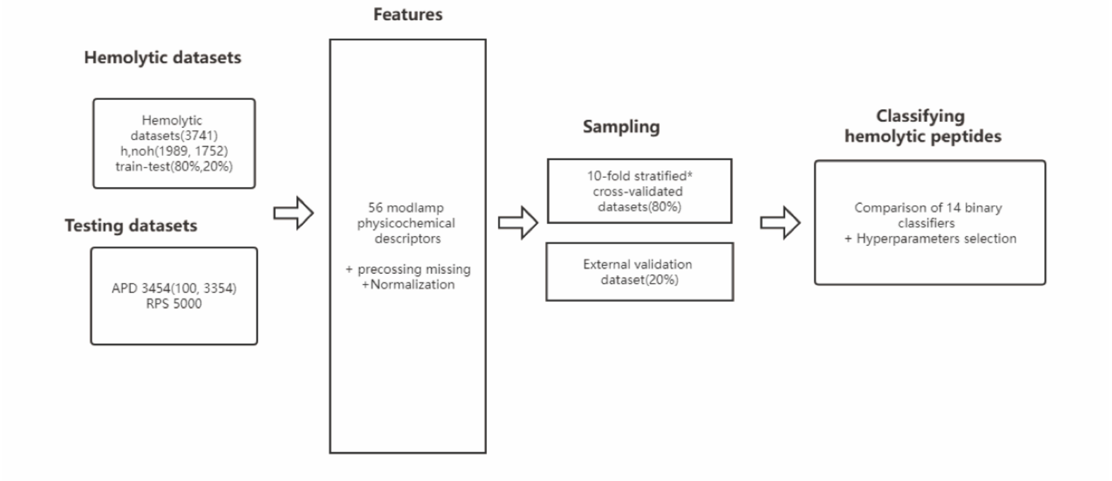

Figure S3 Hemolytic prediction process of peptides in COMDEL model.

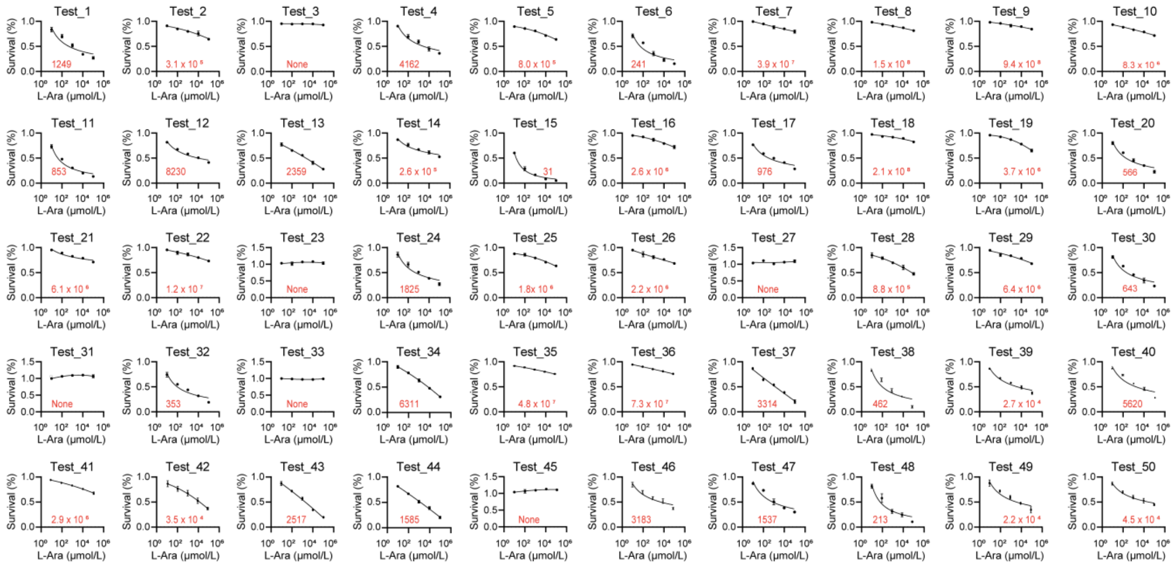

Figure S4 The survival curve of *E. coli* DH5α caused by the expression intensities of peptide candidates. The expression of peptide candidate was under the contron of arabinose operon using pBAD18 vector. The minimum inhibitory concentration reached by 50% (MIC_{50}) of L-arabinose was calculated using dose-response (inhibition) equation. The calculated MIC_{50} was marked in red.

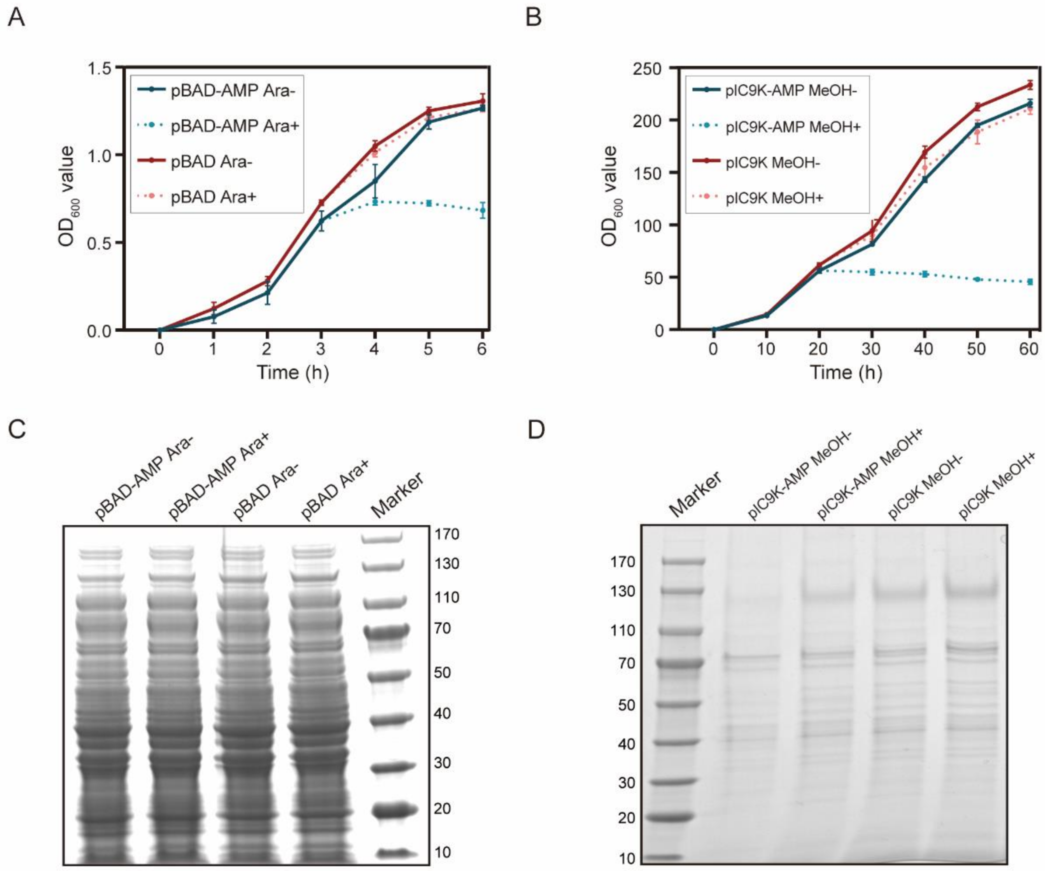

Figure S5 Antimicrobial peptides expressed in *E. coli* and *Pichia yeast*. (A) The OD_{600} value changes of *E. coli* DH5*α* containing AMP or empty vector under the condition with or without inducer L- arabinose. (B) The OD_{600} value changes of *Pichia yeast* containing AMP or empty vector under the condition with or without inducer methanol. (C) SDS-PAGE detected AMP expression in *E. coli* BL21(DE3). (D) SDS-PAGE detected AMP expression in *Pichia yeast*.

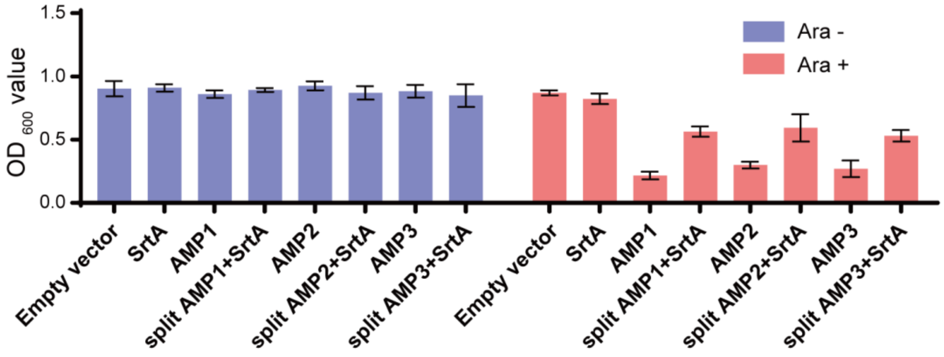

Figure S6 Peptide ligation activity verification of SrtA through AMP effect on strain growth.

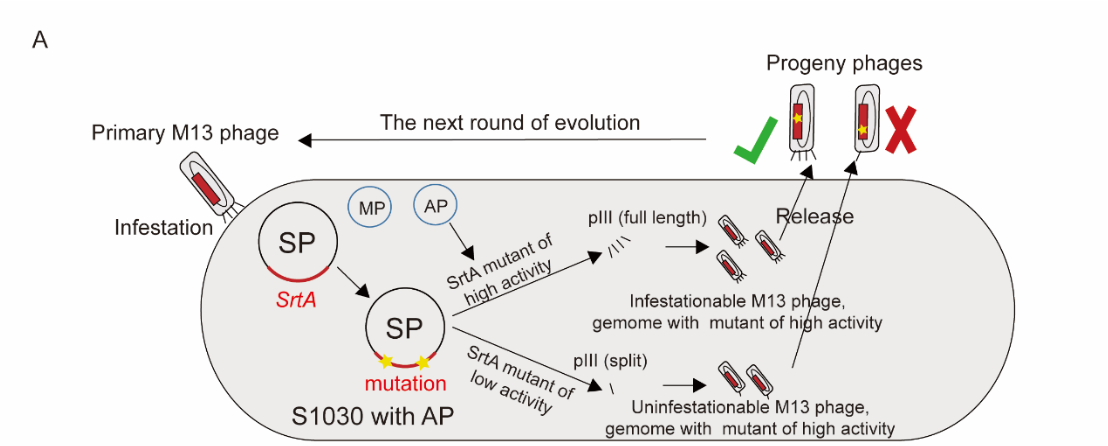

Figure S7 Process diagram of PANCE for SrtA. The activity of SrtA mutants on SP is higher, the more intact pIII proteins are produced, and more infective bacteriophages can be packaged.

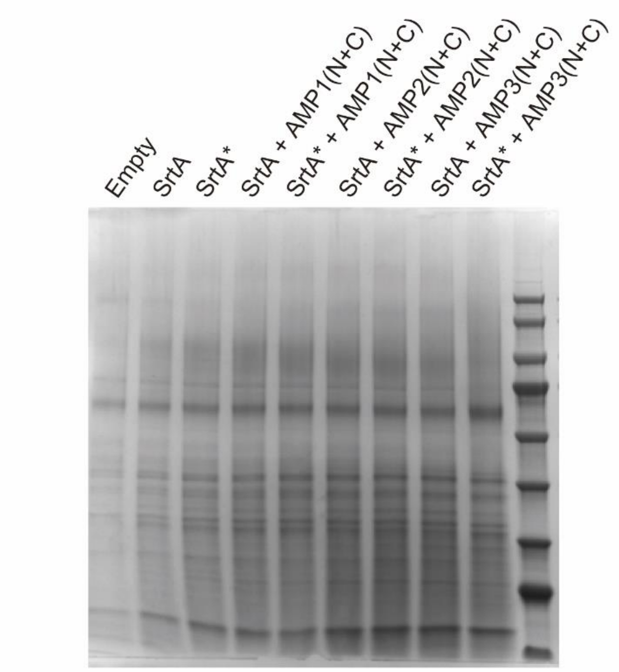

Figure S8 The yield of AMP ligated by SrtA WT or mutant verified by SDS-PAGE.

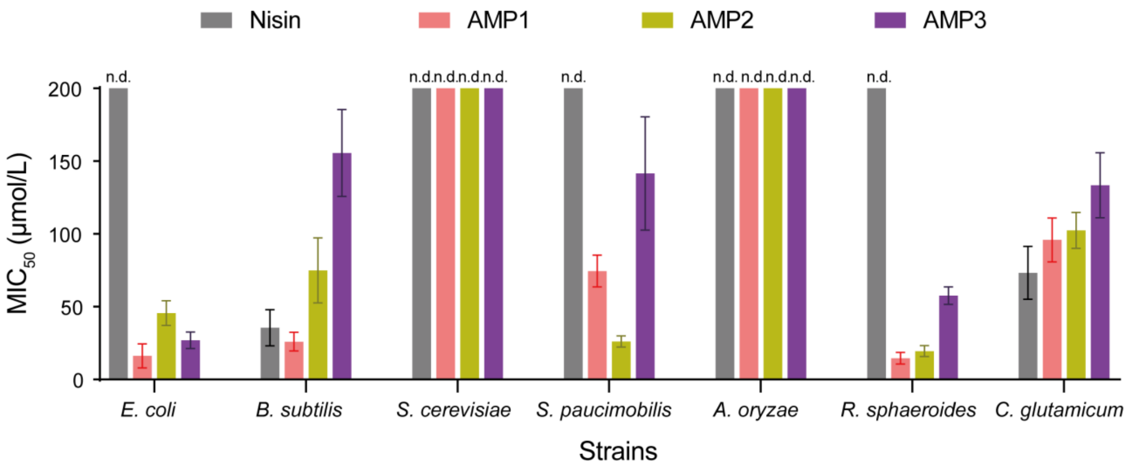

Figure S9 Bacteriostatic test of three purified AMPs. The minimum inhibitory concentration (MIC_{50}) was calculated using the dose-response (inhibition) equation. The n.d. marked on the column means the MIC_{50} value is greater than 200 μmol/L.

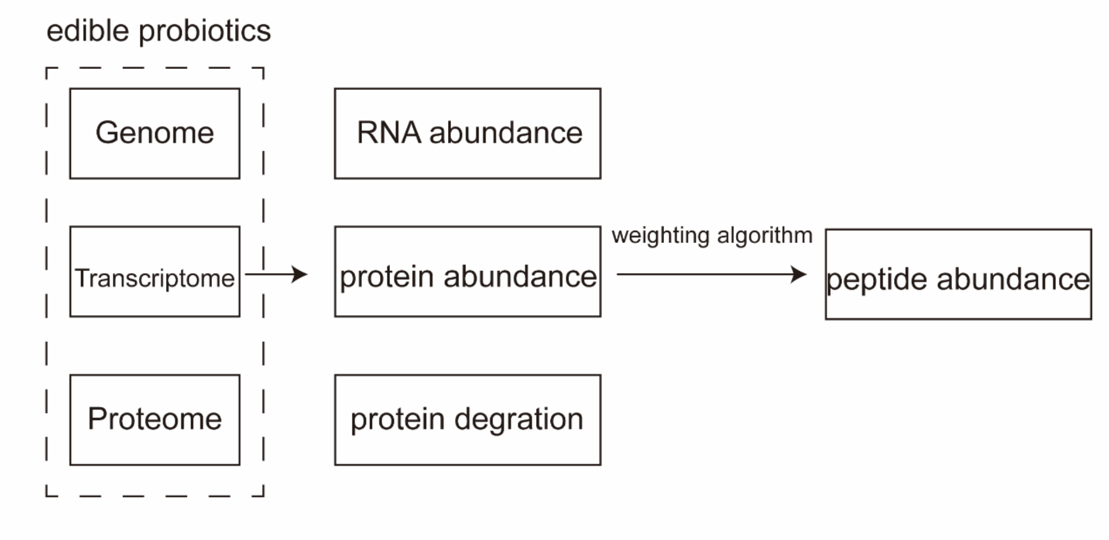

Figure S10 Peptide expression abundance processing for edible probiotics.

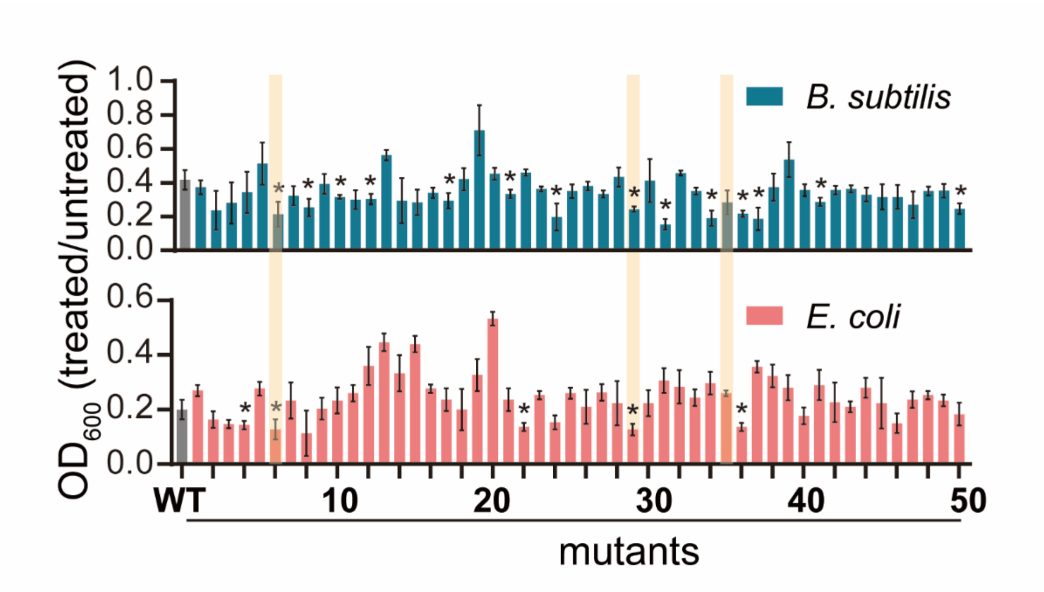

Figure S11 Verification of antibacterial activity of probiotic mutant fermentation product. An asterisk indicates that the data is significantly different.

Table S1. Strains and plasmids used in this study.

Strains	Description	Source
S1030	Strain for PANCE	Addgene
DH5*α*	Cloning strain	TransGen
BL21(DE3)	Protein expression strain	TransGen
B. subtilis 168	Protein expression strain	Biofeng
S.cerevisiae S288C	Antimicrobial test	ATCC
S.paucimobilis	Antimicrobial test	ATCC
A. oryzae	Antimicrobial test	ATCC
R. sphaeroides	Antimicrobial test	ATCC
C. glutamicum	Antimicrobial test	ATCC
L. plantarum	Antimicrobial test	ATCC
B. longum	Antimicrobial test	ATCC
B. bifidum	Antimicrobial test	ATCC
L. reuteri	Antimicrobial test	ATCC
L. lactis	Antimicrobial test	ATCC
pBAD18	Protein expression plasmid	Biofeng
pBAD18-COMDEL_u (1-10)	The expression plasmid for 10 COMDEL_u peptides	this study
pBAD18-COMDEL_e (1-10)	The expression plasmid for 10 COMDEL_e peptides	this study
pBAD18-150N	The 150 bp random library	this study
pBAD18-AMP (1-10)	The expression plasmid for 10 AMPs screened by High-throughput technology	this study
pBAD18-AMP (11-60)	The expression plasmid for 50 AMPs to test the accuracy of COMDEL	this study
MP4	Mutagenesis plasmid for PANCE	Addgene
pJC175e	gIII expression vector for PANCE	Addgene
M13 phage	M13 phage DNA	Zymostar
M13 phage for SrtA	SP for PANCE	this study
AP-split-gIII	AP for PANCE	this study

pBAD-GFP	GFP expression plasmid	this study
pBAD-SrtA	SrtA expression plasmid	this study
pBAD-SrtA*	SrtA* expression plasmid	this study
pBAD-AMP1	AMP1 expression plasmid	this study
pBAD-AMP1 (N+C)	Split AMP1 expression plasmid	this study
pBAD-AMP2	AMP2 expression plasmid	this study
pBAD-AMP2 (N+C)	Split AMP2 expression plasmid	this study
pBAD-AMP3	AMP3 expression plasmid	this study
pBAD-AMP3 (N+C)	Split AMP3 expression plasmid	this study
pT7-AMP1	AMP1 expression plasmid in the CFAS system	this study
pT7-AMP2	AMP2 expression plasmid in the CFAS system	this study
pT7-AMP3	AMP3 expression plasmid in the CFAS system	this study
pT7-GFP1-10	GFP1-10 expression plasmid in *E. coli* BL21(DE3)	this study
pTH01-GFP11	GFP11 expression plasmid in *B. subtilis* 168	this study

Table S3. Primes used in this study.

Primer	Sequence (5’ – 3’)
SrtA-F	atgccagatctgggtacccaagc
SrtA-R	ttatttgacttctgtagctacaaagattttacg
M13-F	tctttgtagctacagaagtcaaataatttccgtcaatatttaccttccctccctcaatcgg
M13-R	ggtttagcttgggtacccagatctggcatgttgaaaatctccaaaaaaaaaggctccaaaagg
M13-check-F	taaagacaaaagggcgacattcaaccg
M13-check-R	ctgtctttcgctgctgagggtgacg
Split-gIII-F	taattttgtttaactttaagaaggagatataccatggattttgattatgaaaagatggc
Split-gIII-R	tatatctccttcttaaagttaaacaaaattattttaaccggaaccagagccaccaccgg
GFP-F	atgaggaaaggagaagaacttttcactgg
GFP10-R	ttaaaggacagggccatcgccaattggag
pT7GFP-F	aattggcgatggccctgtcctttaaggatcctaactcgagcaccaccaccaccaccactgagatcc
pT7GFP -R	ccagtgaaaagttcttctcctttcctcatatggctgccgcgcggcaccaggccgctgc
GFP11-F	atcatatggttttgcacgaatatgttaatgccgccggtataacttaaggatcctctagagtcgacgtcc
GFP-R	ggcattaacatattcgtgcaaaaccatatgatctctcatcctcctttaattgggaattgttatccgctc
pBAD-F	atgtatgtatatctccttcttaaagttaaacaaaggtaccgagc
pBAD-R	gtcgacctgcaggcatgcaagcttggctg
pT7-F	ggtatatctccttcttaaagttaaacaaaattatttctagaggggaattg
pT7-R	gatccggctgctaacaaagcccgaaaggaagc
150N-F	cggtacctttgtttaactttaagaaggag
150N-R	ccgccaaaacagccaagcttgc
NGS-F	acactctttccctacacgacgctcttccgatctcggtacctttgtttaactttaagaaggag
NGS-R	gtgactggagttcagacgtgtgctcttccgatcccgccaaaacagccaagcttgc
pIC9K-F	catcgtttggatccttcgaataattagttg
pIC9K-R	ttcgccttagacatgactgttcctcagttc
Split- AMP1-F	taacagtaaggaggaaaaaaaaatgggcgttaactccgtggacatcttcaacg
Split-AMP1-R	catttttttttcctccttactgttaagagcctttgcactgagagcacagaacc
Split- AMP2-F	taacagtaaggaggaaaaaaaaatgggcctgaaaggtaacggttgcactg

Split- AMP2-R	catttttttttcctccttactgttagtactgttgacgcattttgctacggtag
Split- AMP3-F	taacagtaaggaggaaaaaaaaatgggctgtgctcgtgcctcctctttcgacc
Split- AMP3-R	catttttttttcctccttactgttaagcacatgctgcacacggcacccaagc
AMP-F	gtttagcttgggtacccagatctggcatttttttttcctccttactgttaccaagcttgcatgcctgcaggtcgactta
AMP-R	aaaacgtaaaatctttgtagctacagaagtcaaataactgttttggcggatgagagaag

Caption sentences for Supplementary Table S2, S4, and S5

Table S2. The AMP candidates used in this study.
Table S4. The high-throughput AMP screening data.
Table S5. AMP candidates from soybean (*Glycine Max*) and corn (*Zea Mays*) predicted by COMDEL and other four existing models.