

Supplementary

A computational framework based on ensemble deep neural networks for essential genes identification

Supplementary Table S1. Comparative performance among different ensemble models

Ensemble model	Sens	Spec	Acc	MCC	AUC
MLP + CNN	42.3	92.5	76.1	0.418	0.811
MLP + CNN + SVM	50.5	90.2	77.3	0.452	0.814
MLP + CNN + SVM + RF	45.6	90.2	75.7	0.409	0.754
MLP + CNN + SVM + RF + kNN	50.5	87.9	75.7	0.417	0.813
CNN + SVM	43.7	90.7	75.4	0.398	0.754
CNN + SVM + RF	39.4	92.5	75.2	0.391	0.797
CNN + SVM + RF + kNN	39.8	93	75.7	0.403	0.756
SVM + RF	36.1	93.5	74.8	0.376	0.774
SVM + RF + kNN	35.9	93.9	75.1	0.384	0.796
RF + kNN	41.3	88.8	73.4	0.347	0.764

Supplementary Table S2. Detail information and predictive accuracy of all species datasets

Species	Essential genes	Our Predicted	Our Accuracy	Previous work
<i>Bacillus subtilis</i> 168	271	260	95.94	47.62
<i>Vibrio cholerae</i> N16961	779	439	56.35	-
<i>Streptococcus pneumoniae</i>	244	154	63.11	53.57
<i>Helicobacter pylori</i> 26695	323	148	45.82	26.89
<i>Mycobacterium tuberculosis</i> H37Rv	614	436	71	19.7
<i>Salmonella typhimurium</i> LT2	230	118	51.30	27.65
<i>Acinetobacter baylyi</i> ADP1	499	125	25.05	58.44
<i>Pseudomonas aeruginosa</i> UCBPP-PA14	335	281	83.88	18.21

<i>Salmonella enterica</i> serovar Typhi	353	211	59.77	55.36
<i>Staphylococcus aureus</i> NCTC 8325	351	56	15.95	62.61
<i>Escherichia coli</i> MG1655 I	609	308	50.57	52.34
<i>Escherichia coli</i> MG1655 II	296	142	47.97	52.34
<i>Caulobacter crescentus</i>	480	396	82.5	61.58
<i>Streptococcus sanguinis</i>	218	172	78.9	52.33
<i>Porphyromonas gingivalis</i> ATCC 33277	463	455	98.27	55.95
<i>Bacteroides thetaiotaomicron</i> VPI-5482	325	197	60.62	35.08
<i>Burkholderia thailandensis</i> E264	406	362	89.16	47.39
<i>Mycobacterium tuberculosis</i> H37Rv II	771	566	73.41	19.7
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhimurium str. 140285	105	56	53.33	20.29
<i>Mycobacterium tuberculosis</i> H37Rv III	687	472	68.7	19.7
<i>Sphingomonas wittichii</i> RW1	535	359	67.1	43.55
<i>Shewanella oneidensis</i> MR-1	403	277	68.73	46.75
<i>Pseudomonas aeruginosa</i> PAO1	117	94	80.34	47.39
<i>Salmonella enterica</i> serovar Typhimurium SL1344	353	187	52.97	43.63
<i>Salmonella enterica</i> serovar Typhi Ty2	358	212	59.22	55.36
<i>Bacteroides fragilis</i> 638R	547	320	58.5	30.16
<i>Burkholderia pseudomallei</i> K96243	505	503	99.6	43.17
<i>Pseudomonas aeruginosa</i> PAO1	336	334	99.4	47.39
<i>Streptococcus pyogenes</i> MGAS5448	227	106	46.7	43.15
<i>Streptococcus pyogenes</i> NZ131	241	98	40.66	37.76
<i>Porphyromonas gingivalis</i> ATCC 33277	281	277	98.58	55.95
<i>Synechococcus elongatus</i> PCC 7942	682	606	88.86	-
<i>Rhodopseudomonas palustris</i> CGA009	522	416	79.69	-
<i>Streptococcus agalactiae</i> A909	317	90	28.39	-
<i>Acinetobacter baumannii</i> ATCC 17978	458	150	32.75	-

<i>Acinetobacter baumannii</i> ATCC 17978	156	46	29.49	-
<i>Agrobacterium fabrum</i> str. C58	361	326	90.3	-
<i>Brevundimonas subvibrioides</i> ATCC 15264	412	310	75.24	-
<i>Bacillus thuringiensis</i> BMB171	516	150	29.07	-
<i>Escherichia coli</i> ST131 strain EC958	315	134	42.54	-
<i>Pseudomonas aeruginosa</i> PAO1	551	464	84.21	47.39
<i>Burkholderia cenocepacia</i> K56-2	508	382	75.2	-
<i>Streptococcus mutans</i> UA159	197	61	30.96	-
<i>Neisseria gonorrhoeae</i> MS11	751	475	63.25	-
<i>Escherichia coli</i> O157:H7	1265	428	33.83	-
<i>Ralstonia solanacearum</i> GMI1000	465	383	82.37	-
<i>Streptococcus suis</i>	361	202	55.96	-
<i>Mycobacterium avium</i> subsp. <i>hominissuis</i> strain MAC109	230	157	68.26	-
<i>Providencia stuartii</i> strain BE2467	496	211	42.54	-
<i>Staphylococcus aureus</i> subsp. <i>aureus</i> MSSA476	305	49	16.07	-
<i>Burkholderia cenocepacia</i> J2315	383	235	61.36	-
<i>Vibrio cholerae</i> C6706	343	215	62.68	39.66
<i>Mycoplasma pneumoniae</i>	342	302	88.3	-

'-': dataset was not conducted in the previous work

Supplementary Figure S1. Detail explanation of generating different n-gram levels for DNA sequence (e.g., 'ATGAC' and n-gram levels is from 1 to 3).

