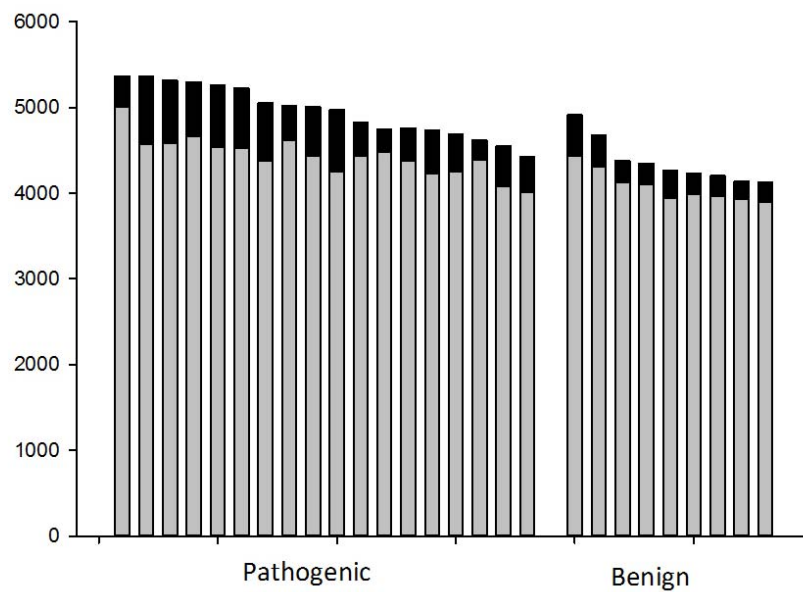


## Supplemental Figure 1



**Supplemental Figure 1.** Prophage content in pathogenic and benign *Escherichia* strains. Black, phage-derived genes; grey, non-phage-derived genes.

Supplemental Table 1. Bacterial genes (proteins),  
associated with virulence, that match POGs

<b>POG#</b>	<b>Protein</b>	<b>Reference (PMID)</b>
POG0073	Zona occludens toxin	<a href="#">16775751</a>
POG0111	Shiga toxin	10074068
POG0150	Zona occludens toxin	16775751
POG0156	ClpP	21855356
POG0157	ClpP	21855356
POG0158	ClpP	21855356
POG0234	Bactoprenol glucosyltransferase	19717593, 10637639
POG0304	ThyA	20347953
POG0306	DNA-binding protein Hu/La	21212121
POG0413	SopB	22325603, 18725540
POG0584	NrdG	21478338
POG0585	NrdD	21502590
POG0689	Panton-Valentine leukocidin	21524624
POG0733	N-6 DNA methylase	11179656, 16489347
POG1138	Shiga toxin	10074068
POG1139	Shiga toxin	10074068
POG1154	OspB	19017275
POG1199	ExsB	7635810
POG1935	Eae-like adhesion protein	10211442
POG2016	GalE	10768949
POG2038	Heme binding virulence factor	19564389
POG3134	KpsF	9383150
POG3603	6-PGD	16634799

## Supplemental Table 2: Assignment of the strains used in this work to pathogenic or non-pathogenic groups

Strain	Pathogenicity	Reference
<b>Escherichia spp.</b>		
Escherichia_coli_O26_H11_11368_uid41021	P	2230233
Escherichia_coli_O111_H_11128_uid41023	P	KEGG
Escherichia_coli_O157_H7_EC4115_uid59091	P	KEGG
Escherichia_coli_O157_H7_TW14359_uid59235	P	KEGG
Escherichia_coli_O157_H7_Sakai_uid57781	P	KEGG
Escherichia_coli_O103_H2_12009_uid41013	P	KEGG
Escherichia_coli_O157_H7_EDL933_uid57831	P	KEGG
Escherichia_coli_O55_H7_CB9615_uid46655	P	KEGG
Escherichia_coli_IAI39_uid59381	P	GOLD
Escherichia_coli_S88_uid62979	P	GOLD
Escherichia_coli_APEC_O1_uid58623	P	KEGG
Escherichia_coli_O127_H6_E2348_69_uid59343	P	KEGG
Escherichia_coli_UTI89_uid58541	P	GOLD
Escherichia_coli_UMN026_uid62981	P	GOLD
Escherichia_coli_CFT073_uid57915	P	11553558
Escherichia_coli_E24377A_uid58395	P	22215741
Escherichia_coli_536_uid58531	P	10894541
Escherichia_coli_ED1a_uid59379	NP	GOLD
Escherichia_coli_SE11_uid59425	NP	GOLD
Escherichia_coli_55989_uid59383	P	KEGG
Escherichia_fergusonii_ATCC_35469_uid59375	NP	22252367
Escherichia_coli_HS_uid58393	NP	GOLD
Escherichia_coli__BL21_GOLDd_DE3_pLysS_AG__uid59245	NP	GOLD
Escherichia_coli_IAI1_uid59377	NP	3
Escherichia_coli_B_REL606_uid58803	NP	GOLD
Escherichia_coli_K_12_substr__DH10B_uid58979	NP	GOLD
Escherichia_coli_K_12_substr__MG1655_uid57779	NP	GOLD
<b>Bacillus spp.</b>		
Bacillus_cereus_NC7401_uid82815	P	11883672
Bacillus_cereus_AH187_uid58753	P	GOLD
Bacillus_cytotoxicus_NVH_391_98_uid58317	P	GOLD
Bacillus_cereus_E33L_uid58103	P	16621833
Bacillus_licheniformis_ATCC_14580_uid58199	NP	15461803
Bacillus_licheniformis_ATCC_14580_uid58097	NP	15461803

Bacillus_subtilis_168_uid57675	NP	Bioproject
Bacillus_amyloliquefaciens_DSM_7_uid53535	NP	GOLD
Bacillus_cereus_G9842_uid58759	P	GOLD
Bacillus_cereus_Q1_uid58529	NP	GOLD
Bacillus_cereus_ATCC_14579_uid57975	P	Bioproject
Bacillus_cereus_ATCC_10987_uid57673	P	Bioproject
Bacillus_thuringiensis_Al_Hakam_uid58795	P	GOLD
Bacillus_cereus_AH820_uid58751	P	GOLD
Bacillus_weiheinstephanensis_KBAB4_uid58315	NP	BacMap
Bacillus_subtilis_BSn5_uid62463	NP	21317323
Bacillus_cereus_03BB102_uid59299	P	16954272
Bacillus_subtilis_spizizenii_TU_B_10_uid73967	NP	HOMD
Bacillus_anthraxis_CDC_684_uid59303	P	21962024
Bacillus_anthraxis_Sterne_uid58091	NP	5
Bacillus_anthraxis_A0248_uid59385	P	GOLD
Bacillus_anthraxis_Ames_Ancestor_uid58083	P	18952800
Bacillus_cereus_B4264_uid58757	P	10747152
Bacillus_subtilis_spizizenii_W23_uid51879	NP	18723616
Bacillus_thuringiensis_serovar_konkukian_97_27_uid58089	P	9650985
Bacillus_thuringiensis_BMB171_uid49135	P	KEGG
Bacillus_cereus_F837_76_uid83611	P	22374959
Bacillus_megaterium_QM_B1551_uid15862	NP	8025666
Bacillus_amyloliquefaciens_FZB42_uid58271	NP	21811596
Bacillus_selenitireducens_MLS10_uid49513	NP	9871015
Bacillus_cellulosilyticus_DSM_2522_uid43329	NP	GOLD
Bacillus_clausii_KSM_K16_uid58237	NP	GOLD
Bacillus_coagulans_2_6_uid68053	NP	GOLD
Bacillus_halodurans_C_125_uid57791	NP	GOLD
Bacillus_pumilus_SAFR_032_uid59017	NP	GOLD
Bacillus_megaterium_DSM319_uid48371	NP	GOLD
Bacillus_coagulans_36D1_uid54335	NP	GOLD
Bacillus_pseudofirmus_OF4_uid45847	NP	GOLD
<b>Pseudomonas spp.</b>		
Pseudomonas_aeruginosa_LESB58_uid59275	P	GOLD
Pseudomonas_aeruginosa_PA7_uid58627	P	20107499
Pseudomonas_fluorescens_Pf_5_uid57937	NP	GOLD
Pseudomonas_putida_GB_1_uid58735	NP	JGI
Pseudomonas_syringae_B728a_uid57931	P	12670984
Pseudomonas_putida_KT2440_uid57843	NP	GOLD
Pseudomonas_fluorescens_SBW25_uid62971	NP	GOLD

<i>Pseudomonas putida</i> _S16_uid68747	NP	GOLD
<i>Pseudomonas syringae phaseolicola</i> _1448A_uid58099	P	16159782
<i>Pseudomonas syringae tomato</i> _DC3000_uid57967	P	GOLD
<i>Pseudomonas putida</i> _W619_uid58651	NP	GOLD
<i>Pseudomonas entomophila</i> _L48_uid58639	P	16699499
<i>Pseudomonas fluorescens</i> _Pf0_1_uid57591	NP	GOLD
<i>Pseudomonas aeruginosa</i> _PAO1_uid57945	P	10984043
<i>Pseudomonas brassicacearum</i> _NFM421_uid66303	NP	21515771
<i>Pseudomonas aeruginosa</i> _UCBPP_PA14_uid57977	P	11115892
<i>Pseudomonas putida</i> _F1_uid58355	NP	GOLD
<i>Pseudomonas mendocina</i> _NK_01_uid66299	NP	3
<i>Pseudomonas stutzeri</i> _A1501_uid58641	NP	GOLD
<i>Pseudomonas stutzeri</i> _ATCC_17588__LMG_11199_uid68749	NP	3
<i>Pseudomonas mendocina</i> _ymp_uid58723	NP	GOLD
<b>Burkholderia spp.</b>		
<i>Burkholderia vietnamiensis</i> _G4_uid58075	P	22187530
<i>Burkholderia glumae</i> _BGR1_uid59397	P	19329631
<i>Burkholderia cenocepacia</i> _J2315_uid57953	P	18931103
<i>Burkholderia gladioli</i> _BSR3_uid66301	P	GOLD
<i>Burkholderia thailandensis</i> _E264_uid58081	NP	GOLD
<i>Burkholderia pseudomallei</i> _K96243_uid57733	P	1
<i>Burkholderia pseudomallei</i> _1710b_uid58391	P	20368977
<i>Burkholderia cenocepacia</i> _MC0_3_uid58769	NP	20090946
<i>Burkholderia phymatum</i> _STM815_uid58699	NP	JGI
<i>Burkholderia</i> _YI23_uid81081	NP	22275096
<i>Burkholderia xenovorans</i> _LB400_uid57823	NP	GOLD
<i>Burkholderia ambifaria</i> _AMMD_uid58303	NP	JGI
<i>Burkholderia ambifaria</i> _MC40_6_uid58701	P	GOLD
<i>Burkholderia</i> _CCGE1001_uid42975	NP	4
<i>Burkholderia</i> _CCGE1002_uid42523	NP	4
<i>Burkholderia phytovirans</i> _PsJN_uid58729	NP	21942451
<i>Burkholderia pseudomallei</i> _668_uid58389	P	2
<i>Burkholderia pseudomallei</i> _1106a_uid58515	P	2
<i>Burkholderia cenocepacia</i> _HI2424_uid58369	NP	20090946
<i>Burkholderia</i> _CCGE1003_uid46253	NP	4
<i>Burkholderia cenocepacia</i> _AU_1054_uid58371	P	20090946

Pathogenicity status according to:

K = KEGG database (Kanehisa, 2012)

GOLD = Genomes Online Database (Pagani *et al.*, 2012)

Bioproject = NCBI Bioproject (Barrett *et al.*, 2012)

BacMap = BacMap database (Cruz *et al.*, 2012)

HOMD = Human Oral Microbiome Database (Chen *et al.*, 2010)

JGI = DOE Joint Genome Institute Portal (Grigoriev, *et al.*, 2012)

[#] = PubMedID (Sayers *et al.*, 2012)

1 = [http://www.bordeninstitute.army.mil/published\\_volumes/biological\\_warfare/BW-ch07.pdf](http://www.bordeninstitute.army.mil/published_volumes/biological_warfare/BW-ch07.pdf)

2 = <http://pathema.jcvi.org/cgi-bin/Burkholderia/shared/HtmlPage.cgi?page=strains>

GSCID = JCVI Genomic Sequencing Center for Infectious Diseases

3 = Not isolated from a person with consistent symptoms

4 = Coenye and Vandamme. *Burkholderia*: Molecular Microbiology and Genomics. 2006.

5 = The use of anthrax vaccines prepared from avirulent (uncapsulated) variants of *Bacillus anthracis*. Onderstepoort J. Vet. Sci. An. Ind. 13:307–312.

## Supplemental Methods

### **Comparing new sequences to the POG collection:**

1. Download the BLASTable POG database (4 files) from the <ftp://ftp.ncbi.nlm.nih.gov/pub/kristensen/extendedPOGs-10/blastdb/> folder
2. Download the stand-alone BLAST program from <ftp://ftp.ncbi.nlm.nih.gov/blast/>
3. Compare new proteins against POG using BLASTP with default parameters (PSI-BLAST with up to 3 iterations might produce better results).
4. In the output, initially consider the hits with the expect values of  $<10^{-5}$  (low stringency) and  $10^{-20}$  (high stringency).