



# HHS Public Access

Author manuscript

*Environ Microbiol.* Author manuscript; available in PMC 2018 March 23.

Published in final edited form as:

*Environ Microbiol.* 2013 February ; 15(2): 307–312. doi:10.1111/j.1462-2920.2012.02886.x.

## Contribution of Phage-derived Genomic Islands to the Virulence of Facultative Bacterial Pathogens

Ben Busby<sup>#1,\*</sup>, David M. Kristensen<sup>#1</sup>, and Eugene V. Koonin<sup>1</sup>

<sup>1</sup> National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

<sup>#</sup> These authors contributed equally to this work.

### Abstract

Facultative pathogens have extremely dynamic pan-genomes, to a large extent derived from bacteriophages and other mobile elements. We developed a simple approach to identify phage-derived genomic islands and apply it to show that pathogens from diverse bacterial genera are significantly enriched in clustered phage-derived genes compared to related benign strains. These findings show that genome expansion by integration of prophages containing virulence factors is a major route of evolution of facultative bacterial pathogens.

---

Most of the known bacterial pathogens have closely related benign counterparts among environmental organisms. What makes a facultative pathogen different from its non-pathogenic relatives? The answer is complex and involves at least two major evolutionary routes. One is the reductive path whereby benign bacteria lose certain metabolic genes, which makes them dependent on their host for at least some essential nutrients. This path has been investigated in much detail and the current view holds that the environmental strains first convert to an obligate host-dependent lifestyle and then, during the massive genome reduction that follows, lose regulatory genes and become pathogenic (Andersson and Kurland, 1998; Moran *et al.*, 2001; Vissa and Brennan, 2001; Merhej *et al.*, 2009; Clarke, 2011; Merhej and Raoult, 2011; Belda *et al.*, 2012) Although this model might explain the origin of many obligate pathogens, it cannot be universal, as many bacterial pathogens retain the ability to live freely outside of their hosts, and their genome sizes actually tend to increase during the conversion to pathogenicity (Ho Sui *et al.* 2009, Botzman and Margalit, 2011).

For these facultative pathogens, the degree of damage inflicted on the host is ultimately determined by the interplay of the organism's genetic make-up and the environmental cues given by the host (Casadevall and Pirofsky, 2001). A critically important role is provided not by the core set of house-keeping genes but by so-called virulence factors, often encoded by foreign genes, horizontally acquired from other bacteria, and/or integrative elements such as prophages or plasmids. In many environments, phages far outnumber their host bacteria, potentially making the phage contribution to the evolution of the pathogen genomes

---

\* To whom correspondence should be addressed.

particularly important. It has been observed that that horizontal gene transfer is much more extensive in the human gut (an extremely ‘crowded’ habitat) than in other environments characterized to date (Smillie *et al.* 2011), although the contribution of prophages to this transfer is likely underrated (Reyes *et al.* 2012). Horizontally acquired genes often cluster together in bacterial genomes resulting in genomic islands that differ from the surrounding genes in their GC content, dinucleotide frequencies, codon usage, and other characteristics (Hacker and Kaper, 2000; Guy, 2006; Soares *et al.*, 2012).

Over a decade ago sequencing and comparison of complete genomes from pathogenic and laboratory strains of *Escherichia coli* revealed that the major differences between these strains were due to the insertion of prophages and other mobile elements, many of which have since undergone extensive decay (Perna *et al.*, 1998; 2001; Burland *et al.*, 1998; Plunkett *et al.*, 1999; Ohnishi *et al.*, 1999; Hayashi *et al.*, 2001; Welch *et al.*, 2002; Asadulghani *et al.*, 2009). Since that time, comparative analysis of the vastly increased collection of complete bacterial genomes has demonstrated that accretion of pathogenicity islands (a type of genomic island found primarily in pathogens) is a general mechanism by which many bacteria undergo conversion to pathogenicity (e.g. Hensel and Schmidt, 2008; Mohammadi-Barzelighi *et al.*, 2011; Trost *et al.*, 2012). Importantly, the dynamics of island accretion is much faster than the dynamics of either speciation or of changes in the core bacterial genome (Castillo-Ramirez *et al.*, 2011). Moreover, certain pathogenic strains of *E. coli* appear to have acquired their prophage insertions independently (Reid *et al.*, 2000), with several insertions specific to different isolates of O157:H7 (Eppinger *et al.*, 2011).

After integration, prophages and other mobile elements undergo extensive decay, making recognition of the genomic (pathogenicity) islands a technically challenging task. Also, the genomes of phages harbor the greatest amount of genetic diversity on the planet, with an estimated 90% still remaining unknown which limits the utility of approaches that rely on similarity to known genes (Angly *et al.*, 2006). To evaluate the contribution of pathogenicity islands, specifically intact and degraded prophages, to the virulence of facultative bacterial pathogens, we compared the characteristics of islands detected by several independent methods of prophage detection (Fouts, 2006; Leplae *et al.*, 2010; Akhter *et al.*, 2012). We also employed the recently updated collection of Phage Orthologous Groups (POGs, available at <ftp://ftp.ncbi.nlm.nih.gov/pub/kristensen/>) which represent a collection of evolutionarily conserved gene families observed in at least 3 distinct complete phage genomes (Kristensen *et al.*, 2011) to detect clusters of genes transferred from phages to bacteria.

Figure 1 shows the mapping of genes homologous to POGs [identified by profile PSI-BLAST (Altschul *et al.*, 1997) searches] within the genomes of two of the first *E. coli* strains to be sequenced, the pathogenic O157:H7 Sakai (Watanabe *et al.*, 1999), and the benign laboratory strain K-12 MG1655 (Blattner *et al.*, 1997). Clusters of phage genes that likely represent inserted prophage regions are shown in pink, and prophages predicted by several other methods are shown for comparison, including (inner boxes, in black) known prophages identified by manual inspection of the genomes (Hayashi *et al.*, 2001, Asadulghani *et al.*, 2009). The regions highlighted in purple are those of the completely automated heuristic program Phage\_Finder (Fouts, 2006); in brown are the Prophinder results available in the

ACLAME database (Leplae *et al.*, 2010), which relies on similarity to known phage genes and examination of integration sites and performs automated genome analysis followed by manual curation; and in blue are the results from the newer PhiSpy tool (Akhter *et al.*, 2012) that combines similarity- and composition-based strategies.

Phage genes that appear in clusters are more likely to have arisen by insertion of a prophage than isolated phage gene homologs, even if much of the prophage has since degraded. Even without using sophisticated techniques employed by many prophage identification methods (e.g., examining integration sites *etc.*), clusters of POGs manage to reproduce the general locations of all of the known prophages in O157:H7 Sakai, with most clusters containing at least one phage-specific gene (phage structural proteins, terminase, integrase *etc.*). Good agreement is also found with known prophages in K-12 except for the cryptic CP4 regions (which might not be prophage-derived) as well as one additional region previously annotated as a “prophage-like” region (Asadulghani *et al.*, 2009). It has been shown that many of the prophages in O157:H7 Sakai, despite having obvious defects, retained the ability to be excised from the host genome, replicate, have their DNA packaged, be released from the cell, and in a few cases even productively infect other bacteria, often with the help of other prophages present in the same cell, thus leading to an overall potentiation of horizontal gene transfer (Asadulghani *et al.*, 2009). Thus, many of the phage gene clusters detected by the POG-based approach are likely to contribute to the evolution of pathogenicity islands even if they are not complete prophages.

Obviously, the pathogenic O157:H7 Sakai genome contains many more clusters of phage genes as well as genes encoding virulence factors associated with phage genes than the genome of the benign laboratory strain K-12 (Figure 1). We defined a set of phage-associated virulence factors by performing literature searches for each POG, looking for proteins whose differential expression (mostly knockout) had a minimal effect on growth in other environments but measurably affected pathogenicity (Brown *et al.*, 2012) (see Supplementary Table 1). Acquisition of pathogenicity islands seems to be a general trend among the pathogenic *E. coli* strains, and among pathogens from other bacterial phyla as well, often bringing additional secretion systems (such as the Type III system that is rarely observed in non-pathogenic *E. coli*), toxins, adhesins, invasins, modulins, and many other virulence genes that modulate host cell activities, such as providing pathogens with the ability to invade the epithelial cell layer (Hacker and Kaper, 2000).

To investigate this trend of acquisition of pathogenicity islands in diverse bacteria, we compiled a collection of 107 completely-sequenced genomes from organisms that retain the capability to live freely outside of their hosts, and represent 4 diverse genera from two different phyla, namely 27 *Escherichia* and 21 *Pseudomonas* (both *Gammaproteobacteria*), 21 *Burkholderia* (*Betaproteobacteria*), and 38 *Bacillus* (*Firmicutes*) strains. This data set includes roughly equal numbers of pathogens and non-pathogens (53 and 54, respectively), and represents several diverse examples of relatively recent and parallel conversions to pathogenicity. For each organism, pathogenicity status was assigned based on the data available in KEGG, GOLD, BioProject, BacMap, and Human Oral Microbiome databases (Kanehisa, 2012; Pagani *et al.*, 2012; Barrett *et al.*, 2012; Sayers *et al.*, 2012; Cruz *et al.*, 2012; Chen *et al.*, 2010) or in the DOE Joint Genome Institute Portal (Grigoriev, *et al.*,

2012) and ascertained from literature searches for the specific strains involved (see Supplementary Table 2). To identify phage-derived genes in these genera, we used the procedure outlined above. Clusters were defined as at least 7 consecutive windows (a window is at least 4 POG matches in a succession of 10 genes), allowing for up to 16 genes without POG matches between them.

Figure 2 shows the comparison of several characteristics potentially related to pathogenicity in the 4 genera examined. Similar to O157:H7 Sakai, the genomes of the other 16 *E. coli* pathogens have significantly more (predicted) protein-coding genes on average than benign strains of *Escherichia* (9 *E. coli* and 1 *E. fergusonii*) (Figure 2A). To a large extent, the additional genes in the pathogens are accounted for by phage-derived genes as measured by matches to POGs (Supplemental Figure 1). Most of these phage-derived genes are organized in clusters that correspond to integrated prophages or their decayed remnants (Figure 1) that are considered to have been acquired independently by different *E. coli* strains (Reid *et al.*, 2000). The same trend generally holds for pathogens from other genera although certain non-pathogenic *Burkholderia* spp. possess large genomes (Figure 2A). Typically, POG profiles yield significantly more BLAST matches per genome for pathogens than for the related benign strains (Figure 2B) even when normalized for genome size (Figure 2C). Furthermore, the pathogen genomes show a stronger tendency for clustering of phage-derived genes (Figure 2D, E), possibly from more recent insertion of prophages, and in 3 of the 4 genera, these clusters tend to be larger (Figure 2F), compatible with the same conclusion. Finally, in some pathogens, these clusters are enriched in presumably phage-derived virulence factors (Figure 2G, H). All the trends toward enrichment for phage-derived genes and (putative) prophages were far more pronounced for the pathogenic strains of *Escherichia* than for the bacteria from other genera (Figure 2). Although it is impossible to rule out that evolution of pathogenicity in *Escherichia* differs from that in other bacteria, we find it much more likely that the stronger trends are a result of the far better characterization of phages that infect *E. coli* (21% of all sequenced phages with known hosts infect *Escherichia*, compared to <10% for each of the other genera (Kristensen *et al.*, 2011)). Should that be the case, the results obtained for *Escherichia* should be considered better representative of the actual processes accompanying the evolution of pathogenicity in free-living bacteria than the results for the other genera.

The results of the present analysis indicate that genome expansion by integration of prophages containing virulence factors is a major route of evolution by which free-living bacteria convert from a benign state to pathogenicity. A striking example outside of these 4 genera analyzed here is the normally commensal *Neisseria meningitidis* that, upon chromosomal integration of a phage, converts to a pathogenic form that can cross the blood-brain barrier and kill a previously healthy individual within hours (Bille *et al.*, 2005). Indeed, it has been more than half a century now since it was first discovered that the integration of a prophage can convert previously benign *Corynebacterium diphtheria* bacteria into a pathogenic form (Freeman, 1951; Uchida *et al.*, 1971). Vibriophages carrying genes necessary for cholera epidemics represent yet another well-studied example (Sen and Ghosh, 2005). As genome sequencing inexorably continues, genomic datasets are getting too large and complex to use brute force algorithms, so simple and efficient approaches like the one

applied here for phage gene detection are becoming especially useful for detecting trends across these large datasets.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

The authors would like to thank Andrew Edwards, Sivakumar Kannan, Alexander Lobkovsky, Kira Makarova, Nobuto Takeuchi, Yuri Wolf and Natalya Yutin for helpful discussions. The authors' research is supported by the Intramural Research Program of the US National Institutes of Health at the National Library of Medicine.

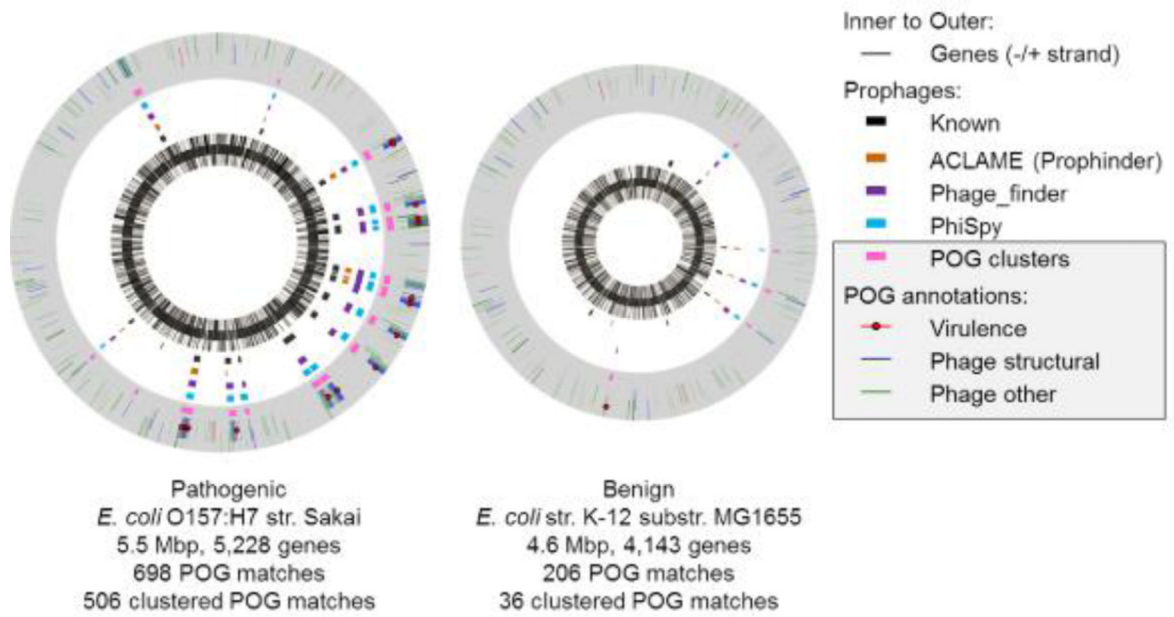
## References

- Akhter S, Aziz RK, Edwards RA. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* 2012 in press. doi: 10.1093/nar/gks406.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25:3389–3402. [PubMed: 9254694]
- Andersson SG, Kurland CG. Reductive evolution of resident genomes. *Trends Microbiol.* 1998; 6:263–268. [PubMed: 9717214]
- Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, et al. The marine viromes of four oceanic regions. *PLoS Biol.* 2006; 4:e368. [PubMed: 17090214]
- Asadulghani M, Ogura Y, Ooka T, Itoh T, Sawaguchi A, Iguchi A, et al. The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog.* 2009; 5:e1000408. [PubMed: 19412337]
- Barrett T, Clark K, Gevorgyan R, Gorelenkov V, Gribov E, Karsch-Mizrachi I, et al. BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. *Nucleic Acids Res.* 2012; 40:D57–D63. [PubMed: 22139929]
- Belda E, Silva FJ, Pereto J, Moya A. Metabolic networks of *Sodalis glossinidius*: a systems biology approach to reductive evolution. *PLoS One.* 2012; 7:e30652. [PubMed: 22292008]
- Bille E, Zahar JR, Perrin A, Morelle S, Kriz P, Jolley KA, et al. A chromosomally integrated bacteriophage in invasive meningococci. *J Exp Med.* 2005; 201:1905–1913. [PubMed: 15967821]
- Blattner FR, Plunkett G III, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of *Escherichia coli* K-12. *Science.* 1997; 277:1453–1462. [PubMed: 9278503]
- Botzman M, Margalit H. Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome Biol.* 2011; 12:R109. [PubMed: 22032172]
- Brown SP, Cornforth DM, Mideo N. Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. *Trends Microbiol.* 2012; 20:336–342. [PubMed: 22564248]
- Burland V, Shao Y, Perna NT, Plunkett G, Sofia HJ, Blattner FR. The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. *Nucleic Acids Res.* 1998; 26:4196–4204. [PubMed: 9722640]
- Casadevall A, Pirofski LA. Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease. *Infect Immun.* 2000; 68:6511–6518. [PubMed: 11083759]
- Castillo-Ramirez S, Harris SR, Holden MT, He M, Parkhill J, Bentley SD, Feil EJ. The impact of recombination on dN/dS within recently emerged bacterial clones. *PLoS Pathog.* 2011; 7:e1002129. [PubMed: 21779170]
- Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database (Oxford).* 2010:baq013. 2010. [PubMed: 20624719]

- Clarke IN. Evolution of *Chlamydia trachomatis*. Ann N Y Acad Sci. 2011; 1230:E11–E18. [PubMed: 22239534]
- Cruz J, Liu Y, Liang Y, Zhou Y, Wilson M, Dennis JJ, et al. BacMap: an up-to-date electronic atlas of annotated bacterial genomes. Nucleic Acids Res. 2012; 40:D599–D604. [PubMed: 22135301]
- Eppinger M, Mammel MK, Leclerc JE, Ravel J, Cebula TA. Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. Proc Natl Acad Sci U S A. 2011; 108:20142–20147. [PubMed: 22135463]
- Fouts DE. *Phage\_Finder*: automated identification and classification of prophage regions in complete bacterial genome sequences. Nucleic Acids Res. 2006; 34:5839–5851. [PubMed: 17062630]
- Freeman VJ. Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. J Bacteriol. 1951; 61:675–688. [PubMed: 14850426]
- Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D, et al. The Genome Portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res. 2012; 40:D26–D32. [PubMed: 22110030]
- Guy L. Identification and characterization of pathogenicity and other genomic islands using base composition analyses. Future Microbiol. 2006; 1:309–316. [PubMed: 17661643]
- Hacker J, Kaper JB. Pathogenicity islands and the evolution of microbes. Annu Rev Microbiol. 2000; 54:641–679. [PubMed: 11018140]
- Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, et al. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. 2001; 8:11–22. [PubMed: 11258796]
- Hensel, M., Schmidt, H. Horizontal gene transfer in the evolution of pathogenesis. Cambridge University Press; Cambridge: 2008.
- Ho Sui SJ, Fedynak A, Hsiao WW, Langille MG, Brinkman FS. The association of virulence factors with genomic islands. PLoS One. 2009; 4:e8094. [PubMed: 19956607]
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012; 40:D109–D114. [PubMed: 22080510]
- Kristensen DM, Cai X, Mushegian A. Evolutionarily conserved orthologous families in phages are relatively rare in their prokaryotic hosts. J Bacteriol. 2011; 193:1806–1814. [PubMed: 21317336]
- Leplae R, Lima-Mendez G, Toussaint A. ACLAME: a CLAssification of Mobile genetic Elements, update 2010. Nucleic Acids Res. 2010; 38:D57–D61. [PubMed: 19933762]
- Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct. 2009; 4:13. [PubMed: 19361336]
- Merhej V, Raoult D. Rickettsial evolution in the light of comparative genomics. Biol Rev Camb Philos Soc. 2011; 86:379–405. [PubMed: 20716256]
- Mohammadi-Barzelighi H, Bakhshi B, Rastegar LA, Pourshafie MR. Characterization of pathogenicity island prophage in clinical and environmental strains of *Vibrio cholerae*. J Med Microbiol. 2011; 60:1742–1749. [PubMed: 21852526]
- Moran NA, Mira A. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. Genome Biol. 2001; 2:RESEARCH0054. [PubMed: 11790257]
- Ohnishi M, Tanaka C, Kuhara S, Ishii K, Hattori M, Kurokawa K, et al. Chromosome of the enterohemorrhagic *Escherichia coli* O157:H7; comparative analysis with K-12 MG1655 revealed the acquisition of a large amount of foreign DNAs. DNA Res. 1999; 6:361–368. [PubMed: 10691128]
- Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, et al. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res. 2012; 40:D571–D579. [PubMed: 22135293]
- Perna NT, Mayhew GF, Posfai G, Elliott S, Donnenberg MS, Kaper JB, Blattner FR. Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. Infect Immun. 1998; 66:3810–3817. [PubMed: 9673266]
- Perna NT, Plunkett G III, Burland V, Mau B, Glasner JD, Rose DJ, et al. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature. 2001; 409:529–533. [PubMed: 11206551]

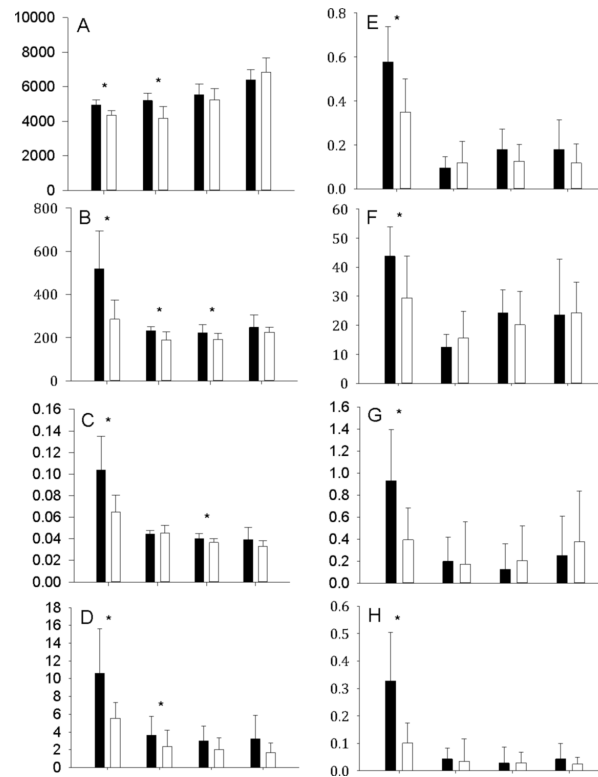


- Plunkett G III, Rose DJ, Durfee TJ, Blattner FR. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. *J Bacteriol.* 1999; 181:1767–1778. [PubMed: 10074068]
- Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS. Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature.* 2000; 406:64–67. [PubMed: 10894541]
- Reyes A, Semenkovich NP, Whiteson K, Rohwer F, Gordon JI. Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat Rev Microbiol.* 2012; 10:607–617. [PubMed: 22864264]
- Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2012; 40:D13–D25. [PubMed: 22140104]
- Sen A, Ghosh AN. New *Vibrio cholerae* O1 biotype ElTor bacteriophages. *Virology.* 2005; 2:28. [PubMed: 15823200]
- Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature.* 2011; 480:241–244. [PubMed: 22037308]
- Soares SC, Abreu VA, Ramos RT, Cerdeira L, Silva A, Baumbach J, et al. PIPS: pathogenicity island prediction software. *PLoS One.* 2012; 7:e30848. [PubMed: 22355329]
- Trost E, Blom J, Soares SC, Huang IH, Al-Dilaimi A, Schroder J, et al. Pangenomic study of *Corynebacterium diphtheriae* that provides insights into the genomic diversity of pathogenic isolates from cases of classical diphtheria, endocarditis, and pneumonia. *J Bacteriol.* 2012; 194:3199–3215. [PubMed: 22505676]
- Uchida T, Gill DM, Pappenheimer AM Jr. Mutation in the structural gene for diphtheria toxin carried by temperate phage. *Nat New Biol.* 1971; 233:8–11. [PubMed: 4999827]
- Vissa VD, Brennan PJ. The genome of *Mycobacterium leprae*: a minimal mycobacterial gene set. *Genome Biol.* 2001; 2:REVIEWS1023. [PubMed: 11532219]
- Watanabe Y, Ozasa K, Mermin JH, Griffin PM, Masuda K, Imashuku S, Sawada T. Factory outbreak of *Escherichia coli* O157:H7 infection in Japan. *Emerg Infect Dis.* 1999; 5:424–428. [PubMed: 10341179]
- Welch RA, Burland V, Plunkett G III, Redford P, Roesch P, Rasko D, et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A.* 2002; 99:17020–17024. [PubMed: 12471157]



**Fig. 1.** Intact and degraded prophage regions identified by several approaches in the pathogenic *Escherichia coli* O157:H7 strain Sakai (left) and the benign laboratory strain K-12 MG1655 (right).





**Fig. 2.** Significant differences between the pathogenic strains of four genera and the benign strains of the same species. For all panels, each two-column set represents (from left to right) *Escherichia* (27 genomes), *Bacillus* (38 genomes), *Pseudomonas* (21 genomes), and *Burkholderia* (21 genomes); pathogenic strains are in black, benign strains are in white. Asterisks indicate p-value < 0.05 by Student's or Welch's t-test, if the sets had equal or unequal variance, respectively. The panels are as follows: A, genes per genome (excluding plasmids); B, POG matches per genome; C, POG matches per gene; D, clusters per genome; E, percent of POG matches in clusters; F, cluster size; G, (known) virulence factors per cluster; H, percent of (known) virulence factors in clusters.