AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

COMMENTARY

100K Pathogen Genome Project

Bart C. Weimer

Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

ABSTRACT The 100K Pathogen Genome Project is producing draft and closed genome sequences from diverse pathogens. This project expanded globally to include a snapshot of global bacterial genome diversity. The genomes form a sequence database that has a variety of uses from systematics to public health.

Received 6 May 2017 **Accepted** 20 May 2017 **Published** 13 July 2017

Citation Weimer BC. 2017. 100K Pathogen Genome Project. Genome Announc 5:e00594-17. https://doi.org/10.1128/genomeA.00594-17.

Address correspondence to bcweimer@ucdavis.edu.

For the article discussed, see https://doi.org/10.1128/genomeA.00518-17.

The views expressed in this Commentary do not necessarily reflect the views of the journal or of ASM.

Whole-genome sequencing (WGS) has accelerated in recent years to produce over 300,000 public genomes. This progress has been, in part, fueled by lower-cost sequencing but more directly motivated by using WGS for food and public health applications. Early efforts to coordinate WGS began with the Lactic Acid Bacteria Consortium (1) that resulted in the reclassification of lactic acid bacteria (LAB) (2–5) and the release of genomes of *Lactococcus*, *Lactobacillus*, *Brevibacterium*, and *Bifidobacterium*. The 100K Pathogen Genome Project (http://100kgenomes.org; BioProject PRJNA186441) was established as an expansion of WGS for use in host-microbe interactions, public health, and genome ecology. The overall goal of the 100K Pathogen Genome Project is to produce high-quality draft genomes, as well as closed genomes of a variety of pathogens from food, animal disease, human disease, wildlife, and environmental reservoirs of those pathogens. It is also being used to inform accurate bacterial identification in metagenomic projects in food safety, where identification accuracy is of utmost importance.

To date, the 100K Pathogen Genome Project has released genomes from *Campylobacter* (6–11), *Shigella* (12), *Salmonella* (13–16), *Listeria* (6, 17), *Helicobacter* (18), and *Vibrio* (19) species, and more are in progress. The study by Kong et al. (15) is the largest release to date and is for *Salmonella*, with over 1,100 draft genomes from 185 serotypes and 130 untypeable isolates. The 100K Pathogen Genome Project uses standardized methods (20–26) to produce genomes with the number of contigs usually <300, and often under 50, with 50 to 100× coverage. Genomes of this quality allow confident measurement of diversity and functional characteristics; initial examples of this are represented in work by Chen et al. (27), who used closed *Salmonella* genomes to provide computational methods to estimate the epigenetic modification methylation status on a population scale as a link to possible function and *Listeria* in an attempt to link methylation status to virulence and risk (6). Initial examination of a variety of genomes provides insights into genome flexibility and rapid evolution on a microbial population scale that was inaccessible previously (28). Realizing that genome diversity is important for identification and functional capability, a global network of participants for the 100K Pathogen Genome Project was established with China, South Korea, and Mexico, with additional internationalization coprojects under way.

REFERENCES

1. Weimer B, Mills D. 2002. Enhancing foods with functional genomics. Food Technol 56:184.
2. Klaenhammer T, Altermann E, Arigoni F, Bolotin A, Breidt F, Broadbent J, Cano R, Chaillou S, Deutscher J, Gasson M, van de Guchte M, Guzzo J, Hartke A, Hawkins T, Hols P, Hutkins R, Kleerebezem M, Kok J, Kuipers O, Lubbers M, Maguin E, McKay L, Mills D, Nauta A, Overbeek R, Pel H, Pridmore D, Saier M, van Sinderen D, Sorokin A, Steele J, O'Sullivan D, de Vos W, Weimer B, Zagorec M, Siezen R. 2002. Discovering lactic acid bacteria by genomics. Antonie Van Leeuwenhoek 82:29–58.
3. Lee JH, Karamychev VN, Kozyavkin SA, Mills D, Pavlov AR, Pavlova NV, Polouchine NN, Richardson PM, Shakhova VV, Slesarev AI, Weimer B, O'Sullivan DJ. 2008. Comparative genomic analysis of the gut bacterium

Bifidobacterium longum reveals loci susceptible to deletion during pure culture growth. BMC Genomics 9:247. https://doi.org/10.1186/1471-2164-9-247.
4. LoCascio RG, Desai P, Sela DA, Weimer B, Mills DA. 2010. Broad conservation of milk utilization genes in *Bifidobacterium longum* subsp. *infantis* as revealed by comparative genomic hybridization. Appl Environ Microbiol 76:7373–7381. https://doi.org/10.1128/AEM.00675-10.
5. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N, Shakhova V, Grigoriev I, Lou Y, Rohksar D, Lucas S, Huang K, Goodstein DM, Hawkins T, Plengvidhya V, Welker D, Hughes J, Goh Y, Benson A, Baldwin K, Lee JH, Díaz-Muñiz I, Dosti B, Smeianov V, Wechter W, Barabote R, Lorca G, Altermann E, Barrangou R, Ganesan B, Xie Y, Rawsthorne H, Tamir D, Parker C, Breidt F, Broadbent J, Hutkins R, O'Sullivan D, Steele J, Unlu G, Saier M, Klaenhammer T, Richardson P, Kozyavkin S, Weimer B, Mills D. 2006. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci U S A 103:15611–15616. https://doi.org/10.1073/pnas.0607117103.
6. Chen P, den Bakker HC, Korlach J, Kong N, Storey DB, Paxinos EE, Ashby M, Clark T, Luong K, Wiedmann M, Weimer BC. 2017. Comparative genomics reveals the diversity of restriction-modification systems and DNA methylation sites in *Listeria monocytogenes*. Appl Environ Microbiol 83:e02091-02016. https://doi.org/10.1128/AEM.02091-16.
7. Taff CC, Weis AM, Wheeler S, Hinton MG, Weimer BC, Barker CM, Jones M, Logsdon R, Smith WA, Boyce WM, Townsend AK. 2016. Influence of host ecology and behavior on *Campylobacter jejuni* prevalence and environmental contamination risk in a synanthropic wild bird species. Appl Environ Microbiol 82:4811–4820. https://doi.org/10.1128/AEM.01456-16.
8. Weis AM, Clothier KA, Huang BC, Kong N, Weimer BC. 2016. Draft genome sequences of *Campylobacter jejuni* strains that cause abortion in livestock. Genome Announc 4(6):e01324-01316. https://doi.org/10.1128/genomeA.01324-16.
9. Weis AM, Clothier KA, Huang BC, Kong N, Weimer BC. 2017. Draft genome sequence of multidrug-resistant abortive *Campylobacter jejuni* from northern California. Genome Announc 5(15):e00171-17. https://doi.org/10.1128/genomeA.00171-17.
10. Weis AM, Huang BC, Storey DB, Kong N, Chen P, Arabyan N, Gilpin B, Mason C, Townsend AK, Smith WA, Byrne BA, Taff CC, Weimer BC. 2017. Large-scale release of *Campylobacter* draft genomes: resources for food safety and public health from the 100K Pathogen Genome Project. Genome Announc 5(1):e00925-16. https://doi.org/10.1128/genomeA.00925-16.
11. Weis AM, Storey DB, Taff CC, Townsend AK, Huang BC, Kong NT, Clothier KA, Spinner A, Byrne BA, Weimer BC. 2016. Genomic comparisons and zoonotic potential of *Campylobacter* between birds, primates, and livestock. Appl Environ Microbiol 82:7165–7175. https://doi.org/10.1128/AEM.01746-16.
12. Weis AM, Gilpin B, Huang BC, Kong N, Chen P, Weimer BC. 2017. *Shigella* draft genome sequences: resources for food safety and public health. Genome Announc 5(16):e00176-17. https://doi.org/10.1128/genomeA.00176-17.
13. Arabyan N, Park D, Foutouhi S, Weis AM, Huang BC, Williams CC, Desai P, Shah J, Jeannotte R, Kong N, Lebrilla CB, Weimer BC. 2016. *Salmonella* degrades the host glycocalyx leading to altered infection and glycan remodeling. Sci Rep 6:29525. https://doi.org/10.1038/srep29525.
14. Deng X, Desai PT, den Bakker HC, Mikoleit M, Tolar B, Trees E, Hendriksen RS, Frye JG, Porwollik S, Weimer BC, Wiedmann M, Weinstock GM, Fields PI, McClelland M. 2014. Genomic epidemiology of *Salmonella enterica* serotype Enteritidis based on population structure of prevalent lineages. Emerg Infect Dis 20:1481–1489. https://doi.org/10.3201/eid2009.131095.
15. Kong N, Davis M, Arabyan N, Huang BC, Weis AM, Chen P, Thao K, Ng W, Chin N, Foutouhi S, Foutouhi A, Kaufman J, Xie Y, Storey DB, Weimer BC. 2017. Draft genome sequences of 1,183 *Salmonella* strains from the 100K Pathogen Genome Project. Genome Announc 5(28):e00518-17. https://doi.org/10.1128/genomeA.00518-17.
16. Arabyan N, Weis AM, Huang BC, Weimer BC. 2017. Implication of sialidases in *Salmonella* infection: genome release of sialidase knockout strains from *Salmonella enterica* serovar Typhimurium LT2. Genome Announc 5(19):e00341-17. https://doi.org/10.1128/genomeA.00341-17.
17. Chen P, Kong N, Huang B, Thao K, Ng W, Storey DB, Arabyan N, Foutouhi A, Foutouhi S, Weimer BC. 2017. 100K Pathogen Genome Project: 306 *Listeria* draft genome sequences for food safety and public health. Genome Announc 5(6):e00967-16. https://doi.org/10.1128/genomeA.00967-16.
18. Draper JL, Hansen LM, Bernick DL, Abedrabbo S, Underwood JG, Kong N, Huang BC, Weis AM, Weimer BC, van Vliet AH, Pourmand N, Solnick JV, Karplus K, Ottemann KM. 2017. Fallacy of the unique genome: sequence diversity within single *Helicobacter pylori* strains. mBio 8:e02321-16. https://doi.org/10.1128/mBio.02321-16.
19. Lüdeke CH, Kong N, Weimer BC, Fischer M, Jones JL. 2015. Complete genome sequences of a clinical isolate and an environmental isolate of *Vibrio parahaemolyticus*. Genome Announc 3(2):e00216-15. https://doi.org/10.1128/genomeA.00216-15.
20. Jeannotte R, Lee E, Arabyan N, Kong N, Thao K, Huang C, Kelly L, Weimer BC. 2014. Optimization of Covaris settings for shearing bacterial genomic DNA by focused ultrasonication and analysis using Agilent 2200 TapeStation. Agilent Technologies technical report. Agilent Technologies, Santa Clara, CA. http://cn.agilent.com/cs/library/applications/5991-5075EN.pdf.
21. Jeannotte R, Lee E, Kong ND, Ng W, Weimer BC. 2014. High-throughput analysis of foodborne bacterial genomic DNA using Agilent 2200 TapeStation and genomic DNA ScreenTape system. Agilent Technologies application note. Agilent Technologies, Santa Clara, CA. https://www.agilent.com/cs/library/applications/5991-4003EN.pdf.
22. Kong N, Ng W, Foutouhi A, Huang BC, Kelly L, Weimer BC. 2014. Quality control of library construction pipeline for PacBio SMRTbell 10 kb library using an Agilent 2200 TapeStation. Agilent Technologies, Santa Clara, CA. https://doi.org/10.13140/RG.2.1.4339.4644.
23. Kong N, Ng W, Cai L, Leonardo A, Weimer BC. 2014. Integrating the DNA integrity number (DIN) to assess genomic DNA (gDNA) quality control using the Agilent 2200 TapeStation system. Agilent Technologies application note. Agilent Technologies, Santa Clara, CA. https://doi.org/10.13140/RG.2.1.3616.8409.
24. Kong N, Thao K, Huang C, Appel M, Lappin S, Knapp L, Kelly L, Weimer BC. 2014. Automated library construction using KAPA library preparation kits on the Agilent NGS workstation yields high-quality libraries for whole-genome sequencing on the Illumina platform. Agilent Technologies application note. Agilent Technologies, Santa Clara, CA. http://www.agilent.com/cs/library/applications/5991-4296EN.pdf.
25. Kong N, Ng W, Thao K, Agulto R, Weis A, Kim KS, Korlach J, Hickey L, Kelly L, Lappin S, Weimer BC. 2017. Automation of PacBio SMRTbell NGS library preparation for bacterial genome sequencing. Stand Genomic Sci 12:27. https://doi.org/10.1186/s40793-017-0239-1.
26. Miller B, van Rooyen B, Whitehorn H, Jones P, Ranik M, van der Walt E, Appel M, Kong N, Huang C, Storey D, Weimer BC. 2015. A novel, single-tube enzymatic fragmentation and library construction method enables fast turnaround times and improved data quality for microbial whole-genome sequencing. Application note. Kapa Biosystems, Wilmington, MA. https://doi.org/10.13140/RG.2.1.4534.3440.
27. Chen P, Jeannotte R, Weimer BC. 2014. Exploring bacterial epigenomics in the next-generation sequencing era: a new approach for an emerging frontier. Trends Microbiol 22:292–300. https://doi.org/10.1016/j.tim.2014.03.005.
28. Weimer BC, Storey DB, Elkins CA, Baker RC, Markwell P, Chambliss DD, Edlund SB, Kaufman JH. 2016. Defining the food microbiome for authentication, safety, and process management. IBM J Res Dev 60:1:1–1:13. https://doi.org/10.1147/JRD.2016.2582598.