An Assessment of Different Genomic Approaches for Inferring Phylogeny of Listeria monocytogenes

Background/objectives: Whole genome sequencing (WGS) has proven to be a powerful subtyping tool for foodborne pathogenic bacteria like L. monocytogenes. The interests of genome-scale analysis for national surveillance, outbreak detection or source tracking has been largely documented. The genomic data however can be exploited with many different bioinformatics methods like single nucleotide polymorphism (SNP), core-genome multi locus sequence typing (cgMLST), whole-genome multi locus sequence typing (wgMLST) or multi locus predicted protein sequence typing (MLPST) on either core-genome (cgMLPST) or pan genome (wgMLPST). Currently, there are little comparisons studies of these different analytical approaches. Our objective was to assess and compare different genomic methods that can be implemented in order to cluster isolates of L monocytogenes.

Methods: The clustering methods were evaluated on a collection of 207 L. monocytogenes genomes of food origin representative of the genetic diversity of the Anses collection. The trees were then compared using robust statistical analyses. Results: The backward comparability between conventional typing methods and genomic methods revealed a near-perfect concordance. The importance of selecting a proper reference when calling SNPs was highlighted, although distances between strains remained identical. The analysis also revealed that the topology of the phylogenetic trees between wgMLST and cgMLST were remarkably similar. The comparison between SNP and cgMLST or SNP and wgMLST approaches showed that the topologies of phylogenetic trees were statistically similar with an almost equivalent clustering.

Conclusion: Our study revealed high concordance between wgMLST, cgMLST, and SNP approaches which are all suitable for typing of L. monocytogenes. The comparable clustering is an important observation considering that the two approaches have been variously implemented among reference laboratories.
Characterization and genetic variation of vibrio cholerae isolated from clinical and environmental sources in Thailand

Cholera is still an important public health problem in several countries, including Thailand. In this study, a collection of clinical and environmental V. cholerae serogroup O1, O139, and non-O1/non-O139 strains originating from Thailand (1983 to 2013) was characterized to determine phenotypic and genotypic traits and to investigate the genetic relatedness. Using a combination of conventional methods and whole genome sequencing (WGS), 78 V. cholerae strains were identified. WGS was used to determine the serogroup, biotype, virulence, mobile genetic elements, and antimicrobial resistance genes using online bioinformatics tools. In addition, phenotypic antimicrobial resistance was determined by the minimal inhibitory concentration (MIC) test. The 78 V. cholerae strains belonged to the following serogroups O1: (n = 44), O139 (n = 16) and non-O1/non-O139 (n = 18). Interestingly, we found that the typical El Tor O1 strains were the major cause of clinical cholera during 1983-2000 with two Classical O1 strains detected in 2000. In 2004-2010, the El Tor variant strains revealed genotypes of the Classical biotype possessing either only ctxB or both ctxB and rstR while they harbored tcpA of the El Tor biotype. Thirty O1 and eleven O139 clinical strains carried CTXφ (Cholera toxin) and tcpA as well four different pathogenic islands (PAIs). Beside non-O1/non-O139, the O1 environmental strains also presented chxA and Type Three Secretion System (TTSS). The in silico MultiLocus Sequence Typing (MLST) discriminated the O1 and O139 clinical strains from other serogroups and environmental strains. ST69 was dominant in the clinical strains belonging to the 7th pandemic clone. Non-O1/non-O139 and environmental strains showed various novel STs indicating genetic variation. Multidrug-resistant (MDR) strains were observed and conferred resistance to ampicillin, azithromycin, nalidixic acid, sulfamethoxazole, tetracycline, and trimethoprim and harboured variants of the SXT elements. For the first time since 1986, the presence of V. cholerae O1 Classical was reported causing cholera outbreaks in Thailand. In addition, we found that V. cholerae O1 El Tor variant and O139 were pre-dominating the pathogenic strains in Thailand. Using WGS and bioinformatic tools to analyze both historical and contemporary V. cholerae circulating in Thailand provided a more detailed understanding of the V. cholerae epidemiology, which ultimately could be applied for control measures and management of cholera in Thailand.
Draft Genome Sequence of Acinetobacter johnsonii C6, an Environmental Isolate Engaging in Interspecific Metabolic Interactions

Acinetobacter johnsonii C6 originates from creosote-polluted groundwater and performs ecological and evolutionary interactions with Pseudomonas putida in biofilms. The draft genome of A. johnsonii C6 is 3.7 Mbp and was shaped by mobile genetic elements. It reveals genes facilitating the biodegradation of aromatic hydrocarbons and resistance to antimicrobials and metals.

General information
State: Published
Organisations: National Food Institute, Research Group for Genomic Epidemiology, Novo Nordisk Foundation Center for Biosustainability, Infection Microbiology, Infection Microbiology
Authors: Kaas, R. S. (Intern), Mordhorst, H. (Intern), Leekitcharoenphon, P. (Intern), Jensen, J. D. (Intern), Haagensen, J. A. J. (Intern), Molin, S. (Intern), Pamp, S. J. (Intern)
Number of pages: 2
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Genome Announcements
Volume: 5
Issue number: 16
Article number: e00155-17
ISSN (Print): 2169-8287
Ratings:
BFI (2018): BFI-level 1
BFI (2017): BFI-level 1
Scopus rating (2016): CiteScore 0.41 SJR 0.217 SNIP 0.233
Web of Science (2016): Indexed yes
Scopus rating (2015): SJR 0.199 SNIP 0.077
Scopus rating (2014): SJR 0.218 SNIP 0.089
ISI indexed (2013): ISI indexed no
Original language: English
Electronic versions:
Kaas_Genome_Announcements_Draft_Genome_Sequence_of_Acinetobacter_johnsonii_C6_an_Environmental_Isolate_Engaging_in_Interspecific_Metabolic_Interactions_2017.pdf
DOIs:
10.1128/genomeA.00155-17

Bibliographical note
This is an openaccess article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Relations
Activities:
Tar-Eating Bacterial Duo may Transform Toxic Compounds into New Usable Materials
Source: PublicationPreSubmission
Source-ID: 131476392
Publication: Research - peer-review › Journal article – Annual report year: 2017

Erratum to: Evaluating next-generation sequencing for direct clinical diagnostics in diarrhoeal disease
Erratum to: Eur J Clin Microbiol Infect Dis.
DOI 10.1007/s10096-017-2947-2

Originally published article contains error.

General information
Evaluating next-generation sequencing for direct clinical diagnostics in diarrhoeal disease

The accurate microbiological diagnosis of diarrhoea involves numerous laboratory tests and, often, the pathogen is not identified in time to guide clinical management. With next-generation sequencing (NGS) becoming cheaper, it has huge potential in routine diagnostics. The aim of this study was to evaluate the potential of NGS-based diagnostics through direct sequencing of faecal samples. Fifty-eight clinical faecal samples were obtained from patients with diarrhoea as part of the routine diagnostics at Hvidovre University Hospital, Denmark. Ten samples from healthy individuals were also included. DNA was extracted from faecal samples and sequenced on the Illumina MiSeq system. Species distribution was determined with MGmapper and NGS-based diagnostic prediction was performed based on the relative abundance of pathogenic bacteria and Giardia and detection of pathogen-specific virulence genes. NGS-based diagnostic results were compared to conventional findings for 55 of the diarrhoeal samples; 38 conventionally positive for bacterial pathogens, two positive for Giardia, four positive for virus and 11 conventionally negative. The NGS-based approach enabled detection of the same bacterial pathogens as the classical approach in 34 of the 38 conventionally positive bacterial samples and predicted the responsible pathogens in five of the 11 conventionally negative samples. Overall, the NGS-based approach enabled pathogen detection comparable to conventional diagnostics and the approach has potential to be extended for the detection of all pathogens. At present, however, this approach is too expensive and time-consuming for routine diagnostics.

General information

State: Published
Organisations: National Food Institute, Research Group for Genomic Epidemiology, Department of Bio and Health Informatics, Genomic Epidemiology, Hvidovre University Hospital, University of Copenhagen
Authors: Joensen, K. G. (Intern), Engsbro, A. L. Ø. (Ekstern), Lukjancenko, O. (Intern), Kaas, R. S. (Intern), Lund, O. (Intern), Westh, H. (Ekstern), Aarestrup, F. M. (Intern)
Pages: 1325-1338
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information

Journal: European Journal of Clinical Microbiology & Infectious Diseases
Volume: 36
Issue number: 7
ISSN (Print): 0934-9723
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.81 SJR 1.289 SNIP 1.137
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.215 SNIP 1.144 CiteScore 2.62
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.205 SNIP 1.2 CiteScore 2.68
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.091 SNIP 1.047 CiteScore 2.63
ISI indexed (2013): ISI indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.136 SNIP 1.154 CiteScore 2.75
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.146 SNIP 1.115 CiteScore 2.69
ISI indexed (2011): ISI indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 1.242 SNIP 1.05
First detection of linezolid resistance due to the optrA gene in enterococci isolated from food products in Denmark

General information
State: Published
Organisations: National Food Institute, Division of Risk Assessment and Nutrition, Research Group for Genomic Epidemiology, Danish Veterinary and Food Administration
Authors: Cavaco, L. (Intern), Korsgaard, H. B. (Intern), Kaas, R. S. (Intern), Seyfarth, A. M. (Ekstern), Leekitcharoenphon, P. (Intern), Hendriksen, R. S. (Intern)
Number of pages: 2
Pages: 128-129
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Journal of Global Antimicrobial Resistance
Volume: 9
ISSN (Print): 2213-7165
Ratings:
Web of Science (2018): Indexed yes
Web of Science (2017): Indexed Yes
Scopus rating (2016): SJR 0.508 SNIP 0.474 CiteScore 1.13
Scopus rating (2015): SJR 0.431 SNIP 0.51 CiteScore 1.01
Web of Science (2015): Indexed yes
Scopus rating (2014): SJR 0.518 SNIP 0.657 CiteScore 1.17
Original language: English
Enterococcus faecalis, Enterococcus faecium, chloramphenicol, florfenicol, linezolid, optrA, tedizolid
DOIs:
10.1016/j.jgar.2017.04.001
Source: FindIt
Source-ID: 2358780092
Publication: Research › Letter – Annual report year: 2017

The CGE Tool Box
As whole genome sequence data of microorganisms are becoming easily accessible and cheap to produce, a transformation of the traditional methods used for typing, phenotyping and phylogenetic analysis of microorganisms is on the way. Following the anticipation that most clinical microbiological and food safety laboratories will soon have a
sequencer in use on a daily basis, there is a growing need for easy-to-use bioinformatics methods that can quickly convert the sequence data into useful information on, e.g., the type of bacteria, whether it is resistant towards any types of antibiotics, and whether it is part of an outbreak. The Center for Genomic Epidemiology, which is located at the Technical University of Denmark, has since its beginning in 2010 developed such bioinformatics methods and made them freely available as web-services. These web-services and their use is the focus of this chapter.

**General information**

State: Published
Organisations: Department of Systems Biology, Department of Bio and Health Informatics, Center for Biological Sequence Analysis, Genomic Epidemiology, National Food Institute, Research Group for Genomic Epidemiology, Immunoinformatics and Machine Learning, Metagenomics, Statens Serum Institute, Osaka University
Pages: 65-90
Publication date: 2017

**Host publication information**

Title of host publication: Applied Genomics of Foodborne Pathogens
Place of publication: Switzerland
Publisher: Springer
Chapter: 5
Main Research Area: Technical/natural sciences
Life Sciences, Food Microbiology, Food Science, Bioinformatics, Microbial Genetics and Genomics, Applied Microbiology, Whole genome sequencing, Web-services
DOIs: 10.1007/978-3-319-43751-4_5
Source: Findit
Source-ID: 2372561475
Publication: Research - peer-review › Book chapter – Annual report year: 2017

Fatal septicemia linked to transmission of MRSA clonal complex 398 in hospital and nursing home, Denmark

We describe 2 fatal cases of methicillin-resistant Staphylococcus aureus (MRSA) clonal complex 398 septicemia in persons who had no contact with livestock. Whole-genome sequencing of the isolated MRSA strains strongly suggest that both were of animal origin and that the patients had been infected through 2 independent person-to-person transmission chains.

**General information**

State: Published
Organisations: National Food Institute, Research Group for Genomic Epidemiology, Odense University Hospital, Hvidovre University Hospital
Authors: Nielsen, R. T. (Ekstern), Kemp, M. (Ekstern), Holm, A. (Ekstern), Skov, M. N. (Ekstern), Detlefsen, M. (Ekstern), Hasman, H. (Intern), Aarestrup, F. M. (Intern), Kaas, R. S. (Intern), Nielsen, J. B. (Ekstern), Westh, H. (Ekstern), Kolmos, H. J. (Ekstern)
Number of pages: 3
Pages: 900-902
Publication date: 2016
Main Research Area: Technical/natural sciences

**Publication information**

Journal: Emerging Infectious Diseases
Volume: 22
Issue number: 5
ISSN (Print): 1080-6040
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 4.92 SJR 3.305 SNIP 2.206
Web of Science (2016): Indexed yes
Is the Evolution of Salmonella enterica subsp. enterica Linked to Restriction-Modification Systems?

Salmonella enterica subsp. enterica bacteria are highly diverse foodborne pathogens that are subdivided into more than 1,500 serovars. The diversity is believed to result from mutational evolution, as well as intra- and interspecies recombination that potentially could be influenced by restriction-modification (RM) systems. The aim of this study was to investigate whether RM systems were linked to the evolution of Salmonella enterica subsp. enterica. The study included 221 Salmonella enterica enterica genomes, of which 68 were de novo sequenced and 153 were public available genomes from ENA. The data set covered 97 different serovars of Salmonella enterica subsp. enterica and an additional five genomes from four other Salmonella subspecies as an outgroup for constructing the phylogenetic trees. The phylogenetic trees were constructed based on multiple alignment of core genes, as well as the presence or absence of pangenes. The topology of the trees was compared to the presence of RM systems, antimicrobial resistance (AMR) genes, Salmonella pathogenicity islands (SPIs), and plasmid replicons. We did not observe any correlation between evolution and the RM systems in S. enterica subsp. enterica. However, sublineage correlations and serovar-specific patterns were observed. Additionally, we conclude that plasmid replicons, SPIs, and AMR were all better correlated to serovars than to RM systems. This study suggests a limited influence of RM systems on the evolution of Salmonella enterica subsp. enterica, which could be due to the conjugational mode of horizontal gene transfer in Salmonella. Thus, we conclude that other factors must be involved in shaping the evolution of bacteria.

The Lake Chad Basin, an Isolated and Persistent Reservoir of Vibrio cholerae O1: A Genomic Insight into the Outbreak in Cameroon, 2010

The prevalence of reported cholera was relatively low around the Lake Chad basin until 1991. Since then, cholera outbreaks have been reported every couple of years. The objective of this study was to investigate the 2010/2011 Vibrio cholerae outbreak in Cameroon to gain insight into the genomic make-up of the V. cholerae strains responsible for the outbreak. Twenty-four strains were isolated and whole genome sequenced. Known virulence genes, resistance genes and integrating conjugative element (ICE) elements were identified and annotated. A global phylogeny (378 genomes) was inferred using a single nucleotide polymorphism (SNP) analysis. The Cameroon outbreak was found to be clonal and clustered distant from the other African strains. In addition, a subset of the strains contained a deletion that was found in the ICE element causing less resistance. These results suggest that V. cholerae is endemic in the Lake Chad basin and different from other African strains.
The proficiency test (pilot) report of the global microbial identifier (GMI) initiative, year 2014

General information
State: Published
Organisations: National Food Institute, Research Group for Genomic Epidemiology, Technical University of Denmark, University of Sydney, New York State Department of Health, Hvidovre Hospital, Ben Gurion University, National Institute of Standards and Technology, Microbiologics, Inc., Public Health England, US Food & Drug Administration
Authors: Hendriksen, R. S. (Intern), Karlsmose Pedersen, S. (Intern), Larsen, M. V. (Intern), Neubert Pedersen, J. (Ekstern), Lukjancenko, O. (Intern), Kaas, R. S. (Intern), Leekitcharoenphon, P. (Intern), Bergmark, L. (Intern), Hansen, I. M. (Intern), Sintchenko, V. (Ekstern), Wolfgang, W. J. (Ekstern), Westh, H. T. (Ekstern), Moran-Gilad, J. (Ekstern), Hsiao, W. (Ekstern), Cuesta, I. (Ekstern), Barrera, J. (Ekstern), Zaballos, A. (Ekstern), Olson, N. D. (Ekstern), Beck, B. (Ekstern), Underwood, A. (Ekstern), Aarestrup, F. M. (Intern), Strain, E. (Ekstern), Pettengill, J. (Ekstern)
Number of pages: 23
Publication date: 2016

Detection of mcr-1 encoding plasmid-mediated colistin-resistant Escherichia coli isolates from human bloodstream infection and imported chicken meat, Denmark 2015

The plasmid-mediated colistin resistance gene, mcr-1, was detected in an Escherichia coli isolate from a Danish patient with bloodstream infection and in five E. coli isolates from imported chicken meat. One isolate from chicken meat belonged to the epidemic spreading sequence type ST131. In addition to IncI2*, an incX4 replicon was found to be linked to mcr-1. This report follows a recent detection of mcr-1 in E. coli from animals, food and humans in China.

General information
State: Published
Organisations: National Food Institute, Research Group for Genomic Epidemiology, State Serum Institute, Copenhagen University Hospital
Authors: Hasman, H. (Ekstern), Hammerum, A. M. (Ekstern), Hansen, F. (Ekstern), Hendriksen, R. S. (Intern), Olesen, B. (Ekstern), Agersø, Y. (Intern), Zankari, E. (Intern), Leekitcharoenphon, P. (Intern), Stegger, M. (Ekstern), Kaas, R. S. (Intern), Cavaco, L. (Intern), Hansen, D. S. (Ekstern), Aarestrup, F. M. (Intern), Skov, R. L. (Ekstern)
Number of pages: 5
Pages: 1-5
Publication date: 2015
Main Research Area: Technical/natural sciences
Genomic Dissection of Travel-Associated Extended-Spectrum-Beta-Lactamase-Producing Salmonella enterica Serovar Typhi Isolates Originating from the Philippines: a One-Off Occurrence or a Threat to Effective Treatment of Typhoid Fever?

One unreported case of extended-spectrum-beta-lactamase (ESBL)-producing Salmonella enterica serovar Typhi was identified, whole-genome sequence typed, among other analyses, and compared to other available genomes of S. Typhi. The reported strain was similar to a previously published strain harboring blaSHV-12 from the Philippines and likely part of an undetected outbreak, the first of ESBL-producing S. Typhi.

General information
State: Published
Organisations: National Food Institute, Division of Epidemiology and Microbial Genomics, Comparative Microbial Genomics, National Center for Emerging and Zoonotic Infectious Diseases, Thailand Ministry of Public Health, VU University Medical Centre, Norwegian Institute of Public Health
Number of pages: 4
Pages: 677-680
Publication date: 2015
Main Research Area: Technical/natural sciences

Publication information
Journal: Journal of Clinical Microbiology
Volume: 53
Issue number: 2
ISSN (Print): 0095-1137
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
Genomic Signature of Multidrug-Resistant Salmonella enterica Serovar Typhi Isolates Related to a Massive Outbreak in Zambia between 2010 and 2012.

Retrospectively, we investigated the epidemiology of a massive Salmonella enterica serovar Typhi outbreak in Zambia during 2010 to 2012. Ninety-four isolates were susceptibility tested by MIC determinations. Whole-genome sequence typing (WGST) of 33 isolates and bioinformatic analysis identified the multilocus sequence type (MLST), haplotype, plasmid replicon, antimicrobial resistance genes, and genetic relatedness by single nucleotide polymorphism (SNP) analysis and genomic deletions. The outbreak affected 2,040 patients, with a fatality rate of 0.5%. Most (83.0%) isolates were multidrug resistant (MDR). The isolates belonged to MLST ST1 and a new variant of the haplotype, H58B. Most isolates contained a chromosomally translocated region containing seven antimicrobial resistance genes, catA1, blaTEM-1, su1, su2, strA, and strB, and fragments of the incompatibility group Q1 (IncQ1) plasmid replicon, the class 1 integron, and the mer operon. The genomic analysis revealed 415 SNP differences overall and 35 deletions among 33 of the isolates subjected to whole-genome sequencing. In comparison with other genomes of H58, the Zambian isolates separated from genomes from Central Africa and India by 34 and 52 SNPs, respectively. The phylogenetic analysis indicates that 32 of the 33 isolates sequenced belonged to a tight clonal group distinct from other H58 genomes included in the study. The small numbers of SNPs identified within this group are consistent with the short-term transmission that can be expected over a period of 2 years. The phylogenetic analysis and deletions suggest that a single MDR clone was responsible for the outbreak, during which occasional other S. Typhi lineages, including sensitive ones, continued to cocirculate. The common view is that the emerging global S. Typhi haplotype, H58B, containing the MDR IncH1 plasmid is responsible for the majority of typhoid infections in Asia and sub-Saharan Africa; we found that a new variant of the haplotype harboring a chromosomally translocated region containing the MDR islands of IncH1 plasmid has emerged in Zambia. This could change the perception of the term “classical MDR typhoid” currently being solely associated with the IncH1 plasmid. It might be more common than presently thought that S. Typhi haplotype H58B harbors the IncH1 plasmid or a chromosomally translocated MDR region or both.
Issue number: 1
ISSN (Print): 0095-1137
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.57 SJR 2.14 SNIP 1.417
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 2.204 SNIP 1.448 CiteScore 3.56
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 2.205 SNIP 1.538 CiteScore 3.84
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 2.414 SNIP 1.646 CiteScore 4.18
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 2.114 SNIP 1.632 CiteScore 4.11
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 2.336 SNIP 1.698 CiteScore 4.27
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 2.303 SNIP 1.727
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 2.173 SNIP 1.694
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 2.239 SNIP 1.621
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 2.202 SNIP 1.689
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 2.187 SNIP 1.642
Web of Science (2006): Indexed yes
Scopus rating (2005): SJR 2.012 SNIP 1.655
Web of Science (2005): Indexed yes
Scopus rating (2004): SJR 1.678 SNIP 1.701
Web of Science (2004): Indexed yes
Scopus rating (2003): SJR 1.845 SNIP 1.855
Web of Science (2003): Indexed yes
Scopus rating (2002): SJR 1.947 SNIP 1.722
Web of Science (2002): Indexed yes
Scopus rating (2001): SJR 2.076 SNIP 1.808
Web of Science (2001): Indexed yes
Scopus rating (2000): SJR 1.945 SNIP 1.938
Web of Science (2000): Indexed yes
Evaluation of whole genome sequencing for outbreak detection of Salmonella enterica

Salmonella enterica is a common cause of minor and large food borne outbreaks. To achieve successful and nearly ‘real-time’ monitoring and identification of outbreaks, reliable sub-typing is essential. Whole genome sequencing (WGS) shows great promises for using as a routine epidemiological typing tool. Here we evaluate WGS for typing of S. Typhimurium including different approaches for analyzing and comparing the data. A collection of 34 S. Typhimurium isolates was sequenced. This consisted of 18 isolates from six outbreaks and 16 epidemiologically unrelated background strains. In addition, 8 S. Enteritidis and 5 S. Derby were also sequenced and used for comparison. A number of different bioinformatics approaches were applied on the data; including pan-genome tree, k-mer tree, nucleotide difference tree and SNP tree. The outcome of each approach was evaluated in relation to the association of the isolates to specific outbreaks. The pan-genome tree clustered 65% of the S. Typhimurium isolates according to the pre-defined epidemiology, the k-mer tree 88%, the nucleotide difference tree 100% and the SNP tree 100% of the strains within S. Typhimurium. The resulting outcome of the four phylogenetic analyses were also compared to PFGE reveling that WGS typing achieved the greater performance than the traditional method. In conclusion, for S. Typhimurium, SNP analysis and nucleotide difference approach of WGS data seem to be the superior methods for epidemiological typing compared to other phylogenetic analytic approaches that may be used on WGS. These approaches were also superior to the more classical typing method, PFGE. Our study also indicates that WGS alone is insufficient to determine whether strains are related or un-related to outbreaks. This still requires the combination of epidemiological data and whole genome sequencing results.
Genome-Wide High-Throughput Screening to Investigate Essential Genes Involved in Methicillin-Resistant Staphylococcus aureus Sequence Type 398 Survival

Livestock-associated methicillin-resistant Staphylococcus aureus (LA-MRSA) Sequence Type 398 (ST398) is an opportunistic pathogen that is able to colonize and cause disease in several animal species including humans. To better understand the adaptation, evolution, transmission and pathogenic capacity, further investigations into the importance of the different genes harboured by LA-MRSA ST398 are required. In this study we generated a genome-wide transposon mutant library in an LA-MRSA ST398 isolate to evaluate genes important for bacterial survival in laboratory and host-specific environments. The transposon mutant library consisted of approximately 1 million mutants with around 140,000 unique insertion sites and an average number of unique inserts per gene of 44.8. We identified LA-MRSA ST398 essential genes comparable to other high-throughput S. aureus essential gene studies. As ST398 is the most common MRSA isolated from pigs, the transposon mutant library was screened in whole porcine blood. Twenty-four genes were specifically identified as important for bacterial survival in porcine blood. Mutations in 23 of these genes resulted in attenuated bacterial fitness. Seven of the 23 genes were of unknown function, whereas 16 genes were annotated with functions predominantly related to carbon metabolism, pH shock and a variety of regulations and only indirectly to virulence factors. Mutations in one gene of unknown function resulted in a hypercompetitive mutant. Further evaluation of these genes is required to determine their specific relevance in blood survival.

General information
State: Published
Organisations: National Food Institute, Division of Epidemiology and Microbial Genomics, Comparative Microbial Genomics, University of Cambridge
Authors: Christiansen, M. T. (Intern), Kaas, R. S. (Intern), Chaudhuri, R. R. (Ekstern), Holmes, M. A. (Ekstern), Hasman, H. (Intern), Aarestrup, F. M. (Intern)
Number of pages: 13
Publication date: 2014
Main Research Area: Technical/natural sciences

Publication information
Journal: PLOS ONE
Volume: 9
Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes

Most current approaches for analyzing metagenomic data rely on comparisons to reference genomes, but the microbial diversity of many environments extends far beyond what is covered by reference databases. De novo segregation of complex metagenomic data into specific biological entities, such as particular bacterial strains or viruses, remains a largely unsolved problem. Here we present a method, based on binning co-abundant genes across a series of metagenomic samples, that enables comprehensive discovery of new microbial organisms, viruses and co-inherited genetic entities and aids assembly of microbial genomes without the need for reference sequences. We demonstrate the method on data from 396 human gut microbiome samples and identify 7,381 co-abundance gene groups (CAGs), including 741 metagenomic species (MGS). We use these to assemble 238 high-quality microbial genomes and identify affiliations between MGS and hundreds of viruses or genetic entities. Our method provides the means for comprehensive profiling of the diversity within complex metagenomic samples.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Novo Nordisk Foundation Center for Biosustainability, Center for Biological sequence analysis, Comparative Microbial Genomics, National Food Institute, Division of Epidemiology and Microbial Genomics, INRA Institut National de La Recherche Agronomique, South China University of Technology, European Molecular Biology Laboratory, Centre National de la Recherche Scientifique, University of Southern Denmark, University Hospital Vall d’Hebron, University of Copenhagen, Vrije Universiteit Brussel, Beijing Genomics Institute Hong Kong, Wageningen IMARES, Tokyo Institute of Technology
Pages: 822-828
Publication date: 2014
Main Research Area: Technical/natural sciences

Publication information
Journal: Nature Biotechnology
Volume: 32
Issue number: 8
ISSN (Print): 1087-0156
Ratings:
BFI (2018): BFI-level 3
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 13.16 SJR 20.253 SNIP 6.303
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 17.892 SNIP 5.505 CiteScore 11.88
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 16.443 SNIP 5.433 CiteScore 11.4
Web of Science (2014): Indexed yes
Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli.

Fast and accurate identification and typing of pathogens are essential for effective surveillance and outbreak detection. The current routine procedure is based on a variety of techniques, making the procedure laborious, time-consuming, and expensive. With whole-genome sequencing (WGS) becoming cheaper, it has huge potential in both diagnostics and routine surveillance. The aim of this study was to perform a real-time evaluation of WGS for routine typing and surveillance of verocytotoxin-producing Escherichia coli (VTEC). In Denmark, the Statens Serum Institut (SSI) routinely receives all suspected VTEC isolates. During a 7-week period in the fall of 2012, all incoming isolates were concurrently subjected to WGS using IonTorrent PGM. Real-time bioinformatics analysis was performed using web-tools (www.genomicepidemiology.org) for species determination, multilocus sequence type (MLST) typing, and determination of phylogenetic relationship, and a specific VirulenceFinder for detection of E. coli virulence genes was developed as part of this study. In total, 46 suspected VTEC isolates were characterized in parallel during the study. VirulenceFinder proved successful in detecting virulence genes included in routine typing, explicitly verocytotoxin 1 (vtx1), verocytotoxin 2 (vtx2), and intimin (eae), and also detected additional virulence genes. VirulenceFinder is also a robust method for assigning verocytotoxin (vtx) subtypes. A real-time clustering of isolates in agreement with the epidemiology was established from WGS, enabling discrimination between sporadic and outbreak isolates. Overall, WGS typing produced results faster and at a lower cost than the current routine. Therefore, WGS typing is a superior alternative to conventional typing strategies. This approach may also be applied to typing and surveillance of other pathogens.
Solving the Problem of Comparing Whole Bacterial Genomes across Different Sequencing Platforms

Whole genome sequencing (WGS) shows great potential for real-time monitoring and identification of infectious disease outbreaks. However, rapid and reliable comparison of data generated in multiple laboratories and using multiple technologies is essential. So far studies have focused on using one technology because each technology has a systematic bias making integration of data generated from different platforms difficult. We developed two different procedures for identifying variable sites and inferring phylogenies in WGS data across multiple platforms. The methods were evaluated on three bacterial data sets and sequenced on three different platforms (Illumina, 454, Ion Torrent). We show that the methods are able to overcome the systematic biases caused by the sequencers and infer the expected phylogenies. It is concluded that the cause of the success of these new procedures is due to a validation of all informative sites that are included in the analysis. The procedures are available as web tools.

General information
State: Published
Organisations: Comparative Microbial Genomics, National Food Institute, Division of Epidemiology and Microbial Genomics, Department of Systems Biology, Center for Biological Sequence Analysis, Immunological Bioinformatics
Authors: Kaas, R. S. (Intern), Leekitcharoenphon, P. (Intern), Aarestrup, F. M. (Intern), Lund, O. (Intern)
Number of pages: 8
Publication date: 2014
Main Research Area: Technical/natural sciences

Publication information
Journal: PLOS ONE
Volume: 9
Issue number: 8
Article number: e104984
ISSN (Print): 1932-6203
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.11 SJR 1.201 SNIP 1.092
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.414 SNIP 1.131 CiteScore 3.32
Whole Genome Epidemiological Typing of Escherichia coli

*Escherichia coli* (*E. coli*) is of huge importance in global health both as a commensal organism living within its host or as a pathogen causing millions of infections each year. Infections occur both sporadically and as outbreaks with sometimes up to thousands of infected people. To limit the number of infections it is important to monitor pathogenic *E. coli* in order to detect outbreaks as quickly as possible and find the source of the outbreak. The effectiveness of monitoring and tracking of pathogens is very dependent on the typing methods that are employed. Classical typing methods employed for *E. coli* are in general expensive and to some extent unreliable. Next generation sequencing has quickly become a tool widely available and has enabled even smaller laboratories to do whole genome sequencing (WGS). Having the entire genome available provides the opportunity to create the ultimate typing method. This PhD thesis attempts to take the first steps toward such a method.

In *Kaas I* all publicly available *E. coli* genomes sequenced (186) are analyzed. 1,702 core genes were found in all genomes. 3,051 genes were found in 95% of the genomes. The pan genome was found to consist of 16,373 genes. The overall phylogeny was inferred from the core genome and also set into context of the *Escherichia* genus. The variance within each gene cluster was calculated in order to compare the variance between genes and possibly identify typing targets for further study. The variance scores calculated was also used to compare the three MLST schemes that exist for *E. coli*.

It quickly became clear that single nucleotide polymorphism (SNP) analysis was becoming the method of choice for inferring the phylogeny of bacterial outbreaks. However, the method remained unavailable to many people due to technical obstacles. In *Kaas II* we describe the SNP method and the validation behind a web server that we set up in order to overcome some of the technical obstacles faced by many people and thereby making the method more available. The
method briefly, calls SNPs against a specified reference sequence, creates an alignment (pseudosequence) of all the SNPs, and uses the maximum likelihood (ML) method to create a tree. The most important detail in the method is the assumption made about “missing” SNPs. Meaning SNPs called in one strain but not in another. It was assumed that SNPs not found in a position was due to that nucleotide being identical to the one in the reference sequence. The assumption is in general valid if all the genomes compared are closely related and the sequencing data is of good quality.

In Kaas III we sought to overcome the assumption mentioned above but most important of all we wanted to create a method that could handle sequence data obtained from different sequencing technologies. The method from Kaas II was completely rewritten and a new web server (CSI Phylogeny) was published that could handle sequence data of all kinds and no longer made assumptions about missing SNPs. Very briefly, the method differs from Kaas II mainly by validating all the locations in all the genomes in which a SNP has been called in any genome. In parallel to the development of a new SNP method another method was also developed that briefly, relies on counting nucleotide differences (ND) between each genome pair, while also validating each position analyzed and ignoring the positions that cannot be validated thereby creating a distance matrix that is used as input to an UPGMA method that creates the final phylogeny. The ND method was also implemented as a web server and published.

If whole genome sequencing is to be used for routine monitoring and tracking of E. coli pathogens, it is crucial to have an idea of how large the difference is between isolates from the same outbreak, compared to the difference to other non-outbreak isolates, in order to do reliable distinctions. In Kaas IV we analyzed ten different outbreaks. Seven of the outbreaks were sequenced for the study and three of the outbreaks were obtained from published studies. Several background isolates that resembled the outbreak isolates were also sequenced. Five different bioinformatic methods were evaluated against the 10 outbreaks. The five different methods were based on SNP, ND, core genes, k-mers, and average nucleotide identity (ANI). Only the ANI method was not able to cluster all outbreaks correctly. The pairwise distance between all isolates were also calculated by each method and compared. Most methods showed lower distance between isolates in the same outbreak compared to the background strains, but only the SNP method was able to set one common threshold for outbreak isolates versus non-outbreak isolates for the entire dataset.

Whole genome sequencing is a powerful but also a rather new tool. This PhD thesis has hopefully shed some light on how we can continue development of whole genome sequence typing and also made WGS more available to a broader audience.
more SNPs over time.
Conclusion: In general full agreement between WGS and PFGE was seen. WGS has higher resolution and are able to
discriminate between isolates of same PFGE type.

General information
State: Published
Organisations: Comparative Microbial Genomics, National Food Institute, Division of Epidemiology and Microbial
Genomics, Statens Serum Institut, Copenhagen University Hospital
Authors: Hansen, D. S. (Ekstern), Kaas, R. S. (Intern), Nielsen, E. M. (Ekstern), Aarestrup, F. M. (Intern), Hasman, H.
(Intern)
Number of pages: 1
Pages: 154
Publication date: 2013

Host publication information
Title of host publication: 10th International Meeting on Microbial Epidemiological Markers (IMMEM-10) - Abstract Book
Place of publication: Paris, France
Main Research Area: Technical/natural sciences
Conference: 10th International Meeting on Microbial Epidemiological Markers, Paris, France, 02/10/2013 - 02/10/2013
Publication: Research - peer-review › Conference abstract in proceedings – Annual report year: 2013

Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial
susceptibility testing
Objectives: Antimicrobial susceptibility testing of bacterial isolates is essential for clinical diagnosis, to detect emerging
problems and to guide empirical treatment. Current phenotypic procedures are sometimes associated with mistakes and
may require further genetic testing. Whole-genome sequencing (WGS) may soon be within reach even for routine
surveillance and clinical diagnostics. The aim of this study was to evaluate WGS as a routine tool for surveillance of
antimicrobial resistance compared with current phenotypic procedures.
Methods: Antimicrobial susceptibility tests were performed on 200 isolates originating from Danish pigs, covering four
bacterial species. Genomic DNA was purified from all isolates and sequenced as paired-end reads on the Illumina
platform. The web servers ResFinder and MLST (www.genomicepidemiology.org) were used to
identify acquired antimicrobial resistance genes and MLST types (where MLST stands for multilocus sequence typing). ResFinder results were compared with phenotypic antimicrobial susceptibility testing results using EUCAST
epidemiological cut-off values and MLST types.
Results: A total of 3051 different phenotypic tests were performed; 482 led to the categorizing of isolates as resistant and
2569 as susceptible. Seven cases of disagreement between tested and predicted susceptibility were observed, six of
which were related to spectinomycin resistance in Escherichia coli. Correlation between MLST type and resistance profiles
was only observed in Salmonella Typhimurium, where isolates belonging to sequence type (ST) 34 were more resistant
than ST19 isolates.
Conclusions: High concordance (99.74%) between phenotypic and predicted antimicrobial susceptibility was observed. Thus, antimicrobial resistance testing based on WGS is an alternative to conventional phenotypic methods.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, National Food Institute, Division of Epidemiology and Microbial
Genomics, Center for Systems Microbiology, Division of Food Microbiology, Department of Systems Biology
Authors: Zankari, E. (Intern), Hasman, H. (Intern), Kaas, R. S. (Intern), Seyfarth, A. M. (Intern), Agersø, Y. (Intern), Lund,
O. (Intern), Larsen, M. V. (Intern), Aarestrup, F. M. (Intern)
Pages: 771-777
Publication date: 2013
Main Research Area: Technical/natural sciences

Publication information
Journal: Journal of Antimicrobial Chemotherapy
Volume: 68
Issue number: 4
ISSN (Print): 0305-7453
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 4.21 SJR 2.24 SNIP 1.527
Veillonella, Firmicutes: Microbes disguised as Gram negatives

The Firmicutes represent a major component of the intestinal microflora. The intestinal Firmicutes are a large, diverse group of organisms, many of which are poorly characterized due to their anaerobic growth requirements. Although most
Firmicutes are Gram positive, members of the class Negativicutes, including the genus Veillonella, stain Gram negative. Veillonella are among the most abundant organisms of the oral and intestinal microflora of animals and humans, in spite of being strict anaerobes. In this work, the genomes of 24 Negativicutes, including eight Veillonella spp., are compared to 20 other Firmicutes genomes; a further 101 prokaryotic genomes were included, covering 26 phyla. Thus a total of 145 prokaryotic genomes were analyzed by various methods to investigate the apparent conflict of the Veillonella Gram stain and their taxonomic position within the Firmicutes. Comparison of the genome sequences confirms that the Negativicutes are distantly related to Clostridium spp., based on 16S rRNA, complete genomic DNA sequences, and a consensus tree based on conserved proteins. The genus Veillonella is relatively homogeneous: inter-genus pairwise comparison identifies at least 1,350 shared proteins, although less than half of these are found in any given Clostridium genome. Only 27 proteins are found conserved in all analyzed prokaryote genomes. Veillonella has distinct metabolic properties, and significant similarities to genomes of Proteobacteria are not detected, with the exception of a shared LPS biosynthesis pathway. The clade within the class Negativicutes to which the genus Veillonella belongs exhibits unique properties, most of which are in common with Gram-positives and some with Gram negatives. They are only distantly related to Clostridia, but are even less closely related to Gram-negative species. Though the Negativicutes stain Gram-negative and possess two membranes, the genome and proteome analysis presented here confirm their place within the (mainly) Gram positive phylum of the Firmicutes. Further studies are required to unveil the evolutionary history of the Veillonella and other Negativicutes.
Draft Genome Sequence of the Yeast Pachysolen tannophilus CBS 4044/NRRL Y-2460

A draft genome sequence of the yeast Pachysolen tannophilus CBS 4044/NRRL Y-2460 is presented. The organism has the potential to be developed as a cell factory for biorefineries due to its ability to utilize waste feedstocks. The sequenced genome size was 12,238,196 bp, consisting of 34 scaffolds. A total of 4,463 genes from 5,346 predicted open reading frames were annotated with function.

General information
State: Published
Organisations: Department of Systems Biology, Center for Microbial Biotechnology, Center for Systems Microbiology, National Food Institute, Division of Epidemiology and Microbial Genomics, Division of Microbiology and Risk Assessment
Authors: Liu, X. (Intern), Kaas, R. S. (Intern), Jensen, P. R. (Intern), Workman, M. (Intern)
Pages: 827
Publication date: 2012
Main Research Area: Technical/natural sciences

Publication information
Journal: Eukaryotic Cell (Online Edition)
Volume: 11
Issue number: 6
ISSN (Print): 1535-9786
Ratings:
Web of Science (2018): Indexed yes
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3 SJR 1.543 SNIP 1.053
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.878 SNIP 0.858 CiteScore 3.12
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.801 SNIP 0.902 CiteScore 3.13
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 2.062 SNIP 1.003 CiteScore 3.58
ISI indexed (2013): ISI indexed no
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 2.148 SNIP 1.008 CiteScore 3.81
ISI indexed (2012): ISI indexed no
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 2.026 SNIP 0.934 CiteScore 3.71
ISI indexed (2011): ISI indexed no
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 2.112 SNIP 0.942
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 2.267 SNIP 0.948
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 2.632 SNIP 0.936
Scopus rating (2007): SJR 2.458 SNIP 0.883
Scopus rating (2006): SJR 2.667 SNIP 0.912
Scopus rating (2005): SJR 3.013 SNIP 1.001
Scopus rating (2004): SJR 2.64 SNIP 1.032
Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes

Background

Escherichia coli exists in commensal and pathogenic forms. By measuring the variation of individual genes across more than a hundred sequenced genomes, gene variation can be studied in detail, including the number of mutations found for any given gene. This knowledge will be useful for creating better phylogenies, for determination of molecular clocks and for improved typing techniques.

Results

We find 3,051 gene clusters/families present in at least 95% of the genomes and 1,702 gene clusters present in 100% of the genomes. The former 'soft core' of about 3,000 gene families is perhaps more biologically relevant, especially considering that many of these genome sequences are draft quality. The E. coli pan-genome for this set of isolates contains 16,373 gene clusters.

A core-gene tree, based on alignment and a pan-genome tree based on gene presence/absence, maps the relatedness of the 186 sequenced E. coli genomes. The core-gene tree displays high confidence and divides the E. coli strains into the observed MLST type clades and also separates defined phylotypes.

Conclusion

The results of comparing a large and diverse E. coli dataset support the theory that reliable and good resolution phylogenies can be inferred from the core-genome. The results further suggest that the resolution at the isolate level may, subsequently be improved by targeting more variable genes. The use of whole genome sequencing will make it possible to eliminate, or at least reduce, the need for several typing steps used in traditional epidemiology.
European freshwater VHSV genotype Ia isolates divide into two distinct subpopulations

Viral haemorrhagic septicaemia (VHS), caused by the novirhabdovirus VHSV, often leads to significant economic losses to European rainbow trout production. The virus isolates are divided into 4 distinct genotypes with additional subgroups including sublineage Ia, isolates of which are the main source of outbreaks in European rainbow trout farming. A significant portion of Danish rainbow trout farms have been considered endemically infected with VHSV since the first disease outbreak was observed in the 1950s. However, following a series of sanitary programs starting in 1965, VHSV has not been detected in Denmark since January 2009. Full-length G-genes of all Danish VHSV isolates that were submitted for diagnostic analyses in the period 2004−2009 were sequenced and analysed. All 58 Danish isolates from rainbow trout grouped with sublineage Ia isolates. Furthermore, VHSV isolates from infected Danish freshwater catchments appear to have evolved into a distinct clade within sublineage Ia, herein designated clade la-1, whereas trout
isolates originating from other continental European countries cluster in another distinct clade, designated clade Ia-2. In addition, phylogenetic analyses indicate that VHSV Ia-1 strains have caused a few outbreaks in Germany and the UK. It is likely that viruses have been transmitted from infected site(s) out of the Danish environment, although a direct transmission pathway has not been identified. Furthermore, VHSV Ia-2 isolates seem to have been transmitted to Denmark at least once. Interestingly, one viral isolate possibly persisted in a Danish watershed for nearly 4 yr without detection whereas other subclades of VHSV isolates appear to have been eliminated, probably because of implemented eradication procedures.

**General information**

State: Published
Organisations: National Veterinary Institute, Section of Fish Diseases, Division of Poultry, Fish and Fur Animals, Division of Epidemiology and Microbial Genomics, Center for Systems Microbiology, National Food Institute, Division of Microbiology and Risk Assessment, Danish Veterinary and Food Administration, Cefas
Pages: 23-35
Publication date: 2012
Main Research Area: Technical/natural sciences

**Publication information**

Journal: Diseases of Aquatic Organisms
Volume: 99
Issue number: 1
ISSN (Print): 0177-5103
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 1.95 SJR 0.858 SNIP 0.929
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 0.949 SNIP 0.935 CiteScore 1.96
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 0.889 SNIP 0.881 CiteScore 1.86
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 0.812 SNIP 0.918 CiteScore 1.77
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 0.912 SNIP 1.092 CiteScore 2.04
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.11 SNIP 1.165 CiteScore 2.29
ISI indexed (2011): ISI indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 0.91 SNIP 0.951
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 0.889 SNIP 0.99
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 0.859 SNIP 0.998
snpTree - a web-server to identify and construct SNP trees from whole genome sequence data.

Background
The advances and decreasing economical cost of whole genome sequencing (WGS), will soon make this technology available for routine infectious disease epidemiology. In epidemiological studies, outbreak isolates have very little diversity and require extensive genomic analysis to differentiate and classify isolates. One of the successfully and broadly used methods is analysis of single nucleotide polymorphisms (SNPs). Currently, there are different tools and methods to identify SNPs including various options and cut-off values. Furthermore, all current methods require bioinformatic skills. Thus, we lack a standard and simple automatic tool to determine SNPs and construct phylogenetic tree from WGS data.

Results
Here we introduce snpTree, a server for online-automatic SNPs analysis. This tool is composed of different SNPs analysis suites, perl and python scripts. snpTree can identify SNPs and construct phylogenetic trees from WGS as well as from assembled genomes or contigs. WGS data in fastq format are aligned to reference genomes by BWA while contigs in fasta format are processed by Nucmer. SNPs are concatenated based on position on reference genome and a tree is constructed from concatenated SNPs using FastTree and a perl script. The online server was implemented by HTML, Java and python script.

The server was evaluated using four published bacterial WGS data sets (V. cholerae, S. aureus CC398, S. Typhimurium and M. tuberculosis). The evaluation results for the first three cases was consistent and concordant for both raw reads and assembled genomes. In the latter case the original publication involved extensive filtering of SNPs, which could not be repeated using snpTree.

Conclusions
The snpTree server is an easy to use option for rapid standardised and automatic SNP analysis in epidemiological studies also for users with limited bioinformatic experience. The web server is freely accessible at http://www.cbs.dtu.dk/services/snpTree-1.0/.

General information
State: Published
Organisations: National Food Institute, Division of Epidemiology and Microbial Genomics, Center for Systems Microbiology, Department of Systems Biology, Center for Biological Sequence Analysis, Division of Microbiology and Risk Assessment
Authors: Leekitcharoenphon, P. (Intern), Kaas, R. S. (Intern), Thomsen, M. C. F. (Intern), Rundsten, C. F. (Intern), Rasmussen, S. (Intern), Aarestrup, F. M. (Intern)
Number of pages: 8
Pages: S6
Publication date: 2012
ABSTRACT: Cholera continues to be an important cause of human infections, and outbreaks are often observed after natural disasters, such as the one following the 2010 earthquake in Haiti. Once the cholera outbreak was confirmed, rumors spread that the disease was brought to Haiti by a battalion of Nepalese soldiers serving as United Nations peacekeepers. This possible connection has never been confirmed. We used whole-genome sequence typing (WGST), pulsed-field gel electrophoresis (PFGE), and antimicrobial susceptibility testing to characterize 24 recent Vibrio cholerae isolates from Nepal and evaluate the suggested epidemiological link with the Haitian outbreak. The isolates were obtained from 30 July to 1 November 2010 from five different districts in Nepal. We compared the 24 genomes to 10 previously sequenced V. cholerae isolates, including 3 from the Haitian outbreak (began July 2010). Antimicrobial susceptibility and PFGE patterns were consistent with an epidemiological link between the isolates from Nepal and Haiti. WGST showed that all 24 V. cholerae isolates from Nepal belonged to a single monophyletic group that also contained isolates from Bangladesh and Haiti. The Nepalese isolates were divided into four closely related clusters. One cluster contained three Nepalese isolates and three Haitian isolates that were almost identical, with only 1- or 2-bp differences. Results in this study are consistent with Nepal as the origin of the Haitian outbreak. This highlights how rapidly infectious diseases might be transmitted globally through international travel and how public health officials need advanced molecular tools along with standard epidemiological analyses to quickly determine the sources of outbreaks. IMPORTANCE Cholera is one of the ancient classical diseases and particularly prone to cause major outbreaks following major natural disasters, such as earthquakes and hurricanes, where the normal separation between sewage and drinking water is destroyed. This was the case following the 2010 earthquake in Haiti. Rumors spread that the disease was brought to Haiti by a battalion of Nepalese soldiers serving as United Nations peacekeepers. This possible connection has never been confirmed. Sequencing the genomes of bacteria can give detailed information on whether isolates from different sites share a common origin. We used this technology to sequence isolates of Vibrio cholerae from Nepal, identify single-nucleotide polymorphisms (SNPs), and compare these high-resolution genotypes to the complete genome sequences of isolates from the Haiti outbreak. We provide support for the hypothesis that the isolates were brought to Haiti from Nepal. IMPORTANCE Cholera is one of the ancient classical diseases and particularly prone to cause major outbreaks following major natural disasters, such as earthquakes and hurricanes, where the normal separation between sewage and drinking water is destroyed. This was the case following the 2010 earthquake in Haiti. Rumors spread that the disease was brought to Haiti by a battalion of Nepalese soldiers serving as United Nations peacekeepers. This possible connection has never been confirmed. Sequencing the genomes of bacteria can give detailed information on whether isolates from different sites share a common origin. We used this technology to sequence isolates of Vibrio cholerae from Nepal, identify single-nucleotide polymorphisms (SNPs), and compare these high-resolution genotypes to the complete genome sequences of isolates from the Haiti outbreak. We provide support for the hypothesis that the isolates were brought to Haiti from Nepal.
Projects:

Detection of Pathogens and Antimicrobial Resistance in Microbiomes

National Food Institute
Period: 01/10/2015 → 30/12/2018
Number of participants: 4
Phd Student:
Poulsen, Casper Sahl (Intern)
Supervisor:
Kaas, Rolf Sommer (Intern)
Pamp, Sünje Johanna (Intern)
Main Supervisor:
Aarestrup, Frank Møller (Intern)

Financing sources
Source: Internal funding (public)
Name of research programme: Samfinansieret - Andet
Project: PhD

Whole Genome Epidemiological Typing Of Escherichia Coli

National Food Institute
Period: 01/11/2010 → 26/01/2015
Number of participants: 7
Phd Student:
Kaas, Rolf Sommer (Intern)
Supervisor:
Lund, Ole (Intern)
Ussery, David (Intern)
Main Supervisor:
Aarestrup, Frank Møller (Intern)
Examiner:
Hendriksen, Rene S. (Intern)
Krogh, Anders Stærmose (Intern)
Underwood, Anthony (Ekstern)

**Financing sources**
Source: Internal funding (public)
Name of research programme: Institut stipendie (DTU) Samf.
Project: PhD