MLST+ (cgMLST) Strain Typing of Multidrug-Resistant Organisms (MDROs) Using Acuitas® Whole Genome Sequence Analysis

Weizhong Chang, Rossio Kersey, Alex Saeed, Vadim Sapiro, Terry Walker

OpGen, Inc., 708 Quince Orchard Road, Gaithersburg, MD 20878, USA.

Abstract

Quick and efficient strain typing of MDRO clinical isolates is crucial for the prevention of outbreaks. In this study we developed MLST+ (cgMLST) schemas to strain type closely related clinical isolates of eight Gram-negative species with the highest priority in healthcare facilities, by using Acuitas® Whole Genome Sequence Analysis with next generation sequencing technology.

We selected eight species of microbes based on their prominence and clinical relevance: *Escherichia coli, Pseudomonas aeruginosa, Klebsiella pneumonia, Acinetobacter baumanii, Enterobacter cloacae, Citrobacter freundii, Klebsiella oxytoca, and Serratia marcescens.*

To create a stable MLST+ (cgMLST) schema, a reference genome sequence and enough query genome sequences of that species are required. Among our eight selected species, the first four already had reference genomes suggested by Ridom™. For the remaining four species, we selected reference genome sequences by following the criteria provided by Ridom: the candidate reference genome must be finished, annotated, and accessible; the reference should ideally be constructed using Sanger sequencing; the reference isolate should be available from culture collections and DNA for sequencing must be available; and the reference isolate should preferably be the type strain or another well characterized strain of the species. Since the schema we developed will be used for strain typing MDROs, we always selected the strains from among human pathogens.

The query genome sequences were drawn from finished genome and scaffold sequences of each species from The National Center for Biotechnology Information (NCBI). However, there were not enough genome sequences for *Enterobacter cloacae, Citrobacter freundii, Klebsiella oxytoca*, *and Serratia marcescens* in NCBI to create stable MLST+ (cgMLST) schema; we supplemented those query sequences with our own assembled WGS to complete the query genome set.

To test these schemas, Illumina MiSeq data were generated on a total of 69 clinical isolates of these eight species, and the whole genome sequences were then assembled and analyzed. The results demonstrated that Acuitas Whole Genome Sequence Analysis MLST+ (cgMLST) can strain type these closely related clinical isolates of each species, evaluate evolutionary relationships among the isolates, and reveal the possibility of an outbreak occurrence.

In conclusion, we successfully created MLST+ schemas for eight species: *Escherichia coli, Pseudomonas aeruginosa, Klebsiella pneumonia, Acinetobacter baumanii, Enterobacter cloacae, Citrobacter freundii, Klebsiella oxytoca, and Serratia marcescens.* These schemas successfully strain typed closely related clinical isolates, demonstrating the utility of Acuitas Whole Genome Sequence Analysis for transmission investigations and outbreak prevention.

Objectives

To develop and validate MLST+ (cgMLST) for eight clinical significant species: *Escherichia coli, Pseudomonas aeruginosa, Klebsiella pneumonia, Acinetobacter baumanii, Enterobacter cloacae, Citrobacter freundii, Klebsiella oxytoca, and Serratia marcescens.* These schemas will be used to strain type closely related clinical isolates.

Methods and Materials

Define MLST+ (cgMLST) Schema

The MLST+ (cgMLST) targets of each species were defined at OpGen in conjunction with Seqsphere+ from Ridom.

1) **Selection of the reference genomes**

   We used chromosomal sequences of Ridom-recommended strains as reference genomes for *Escherichia coli, Pseudomonas aeruginosa*, *Klebsiella pneumonia*, *Acinetobacter baumanii.* For *Enterobacter cloacae, Citrobacter freundii, Klebsiella oxytoca, and Serratia marcescens.* we selected reference genome sequences by following the criteria provided by Ridom.

2) **Selection of query sequences**

   The query genome sequences were drawn from finished genome and scaffold sequences of each species from National Center for Biotechnology Information (NCBI). However, there were not enough genome sequences for *Enterobacter cloacae, Citrobacter freundii, Klebsiella oxytoca*, *and Serratia marcescens* in NCBI to create stable MLST+ schema; we supplemented those query sequences with our own assembled WGS to complete the query genome set.

Validation of MLST+ (cgMLST) Schema

1) **Clinical Samples**

   Total of 69 clinical isolates of eight species: *Escherichia coli, Pseudomonas aeruginosa, Klebsiella pneumonia, Acinetobacter baumanii, Enterobacter cloacae, Citrobacter freundii, Klebsiella oxytoca, and Serratia marcescens* have been chosen for the validating of the MLST+ (cgMLST) schemas. Besides these NGS data generated at OpGen, we also use five outbreak E.coli strains (*E.coli O104:H4*, *BMC Research Notes* 2011, 4:533) to test the ability of the schema to detect outbreak.

2) **Testing of the cgMLST with Assembled Whole Genome Sequences**

   The whole genome sequences (WGS) of clinical isolates were assembled from Illumina MiSeq using assembler Velvet. These assembled WGS were used to test the MLST+ (cgMLST) created in OpGen..

Results

1. **Define MLST+ (cgMLST) Schemas for Acuitas whole genome sequence**. Eight species' MLST+ (cgMLST) schemas have been created in this project. Reference genomes for *Escherichia coli, Pseudomonas aeruginosa, Klebsiella pneumonia* and *Acinetobacter baumanii have been recommended by Ridom* For the remaining four species (*Enterobacter cloacae, Citrobacter freundii, Klebsiella oxytoca,* and *Serratia marcescens*), we selected reference genome sequences by following the criteria provided by Ridom: the candidate reference genome must be finished, annotated, and accessible; the reference should ideally be constructed using Sanger sequencing; the reference isolate should be available from culture collections and DNA for sequencing must be available; and the reference isolate should preferably be the type strain or another well characterized strain of the species. Since the schema we developed will be used for strain typing MDROs, we always selected the strains from among human pathogens. The final target number in the MLST+ (cgMLST) and number of isolates used to validate the schemas for each species are listed in Table 1.

Table 1. Development and validation of MLST+ (cgMLST) Schemas for Acuitas Whole Genome Sequence Analysis.

Organism	Number of Targets	Number of Isolates for Validation
E. coli / STEC	2300	14
P. Aeruginosa	3694	4
K. pneumoniae	3,197	20
A. baumanii	1855	4
E. cloacae	1110	8
C. freundii	1776	8
K. Oxytoca	3136	5
S. marcescens	2599	6

2. **Acuitas Whole Genome Sequence Analysis MLST+ (cgMLST) Provides High Resolution Strain Typing of Clinical Isolates:** We have tested all eight MLST+ (cgMLST) with whole genome sequences of clinical isolates for each species. Strain typing results of ten isolates of Klebsiella (nine *Klebsiella pneumonia* and one *Klebsiella oxytoca*) were shown in Figure 1. MLST+ (cgMLST) results (A) are consistent with the results from traditional MLST (B) with higher resolution.

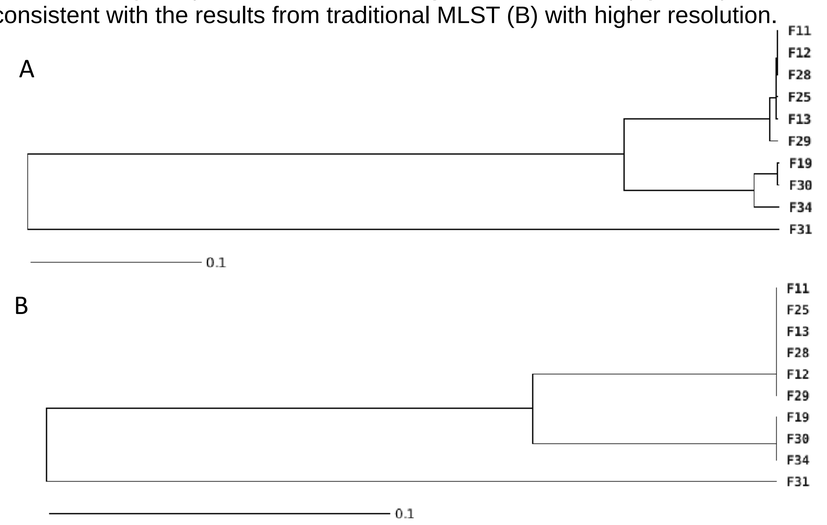

Figure 1. Strain Typing of *Klebsiella* with MLST+ (cgMLST) and MLST. (A) MLST+ (cgMLST) provide consistent strain typing results comparing to traditional MLST (B) with higher resolution: MLST+ (cgMLST) further differentiate strains in two MLST clusters into sub-clusters.

2. **Validation of MLST+ (cgMLST) Schemas with Outbreak Strains.** Five German 2011 *E.coli O104:H4* outbreak strains and six unrelated clinical isolates of *E.coli* were included in this study. Acuitas Whole Genome Analysis MLST+ (cgMLST) correctly placed these outbreak strains in one cluster with a little difference between them (Figure 2).

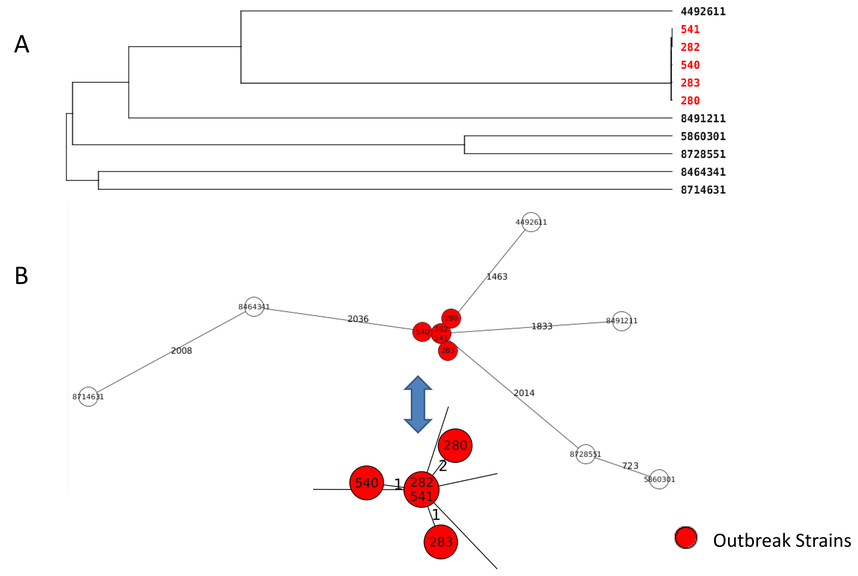

Figure 2 Analysis of German 2011 *E.coli O104:H4* Outbreak with Acuitas Sequence Analysis MLST+ (cgMLST). Whole genome sequences (WGS) were assembled from Illumina Miseq data for five *E.coli O104:H4* outbreak strains downloaded from NCBI and six other unrelated strains generated at OpGen, Inc. The assembled WGS were then analyzed with MLST+ (cgMLST) schema developed at OpGen, Inc. The results were visualized with dendrogram (A) and minimum spanning tree (B). The results demonstrate MLST+ (cgMLST) can identify outbreak strains (in red). MLST+ (cgMLST) can resolved very closely clinical isolates. In this case, MLST+ (cgMLST) can differentiate these outbreak strains by less or equal to three genes different out of total 2300 core genomic genes.

Conclusion

1. MLST+ schemas for eight species: *Escherichia coli, Pseudomonas aeruginosa, Klebsiella pneumonia, Acinetobacter baumanii, Enterobacter cloacae, Citrobacter freundii, Klebsiella oxytoca, and Serratia marcescens* have been successfully created.
2. These schemas successfully strain typed closely related clinical isolates, demonstrating the utility of Acuitas Whole Genome Sequence Analysis for transmission investigations and outbreak prevention.

Acuitas Whole Genome Sequence Analysis is for research use only and not for use in diagnostic procedures.