Bacterial typing using the Ion PGM System and Ridom SeqSphere+ Software

Key findings

• Core-genome multilocus sequence typing (MLST+), composed of 1,241 targets, was applied to the de novo assembly of three *N. meningitidis* community outbreak strains, resulting in improved strain discrimination compared with standard meningococcal typing (Figure 1).

• Fast, accurate whole-genome sequencing (WGS) on the Ion PGM™ System followed by typing of microbes using Ion Torrent™ Ridom™ SeqSphere+ Software can be performed in a rapid <24-hour workflow, from library construction to MLST+ typing results.

• Genome-wide bacterial strain typing, using a gene-by-gene (allelic) approach, better accounts for recombination events and the polymorphic nature of many bacterial species, while facilitating an easily expandable nomenclature and classification structure.

Processing of bacterial pathogens

The characterization of bacterial pathogens is a significant burden on clinical research microbiologists, with the requirements for speed, affordability, and accuracy in the characterization of pathogens becoming increasingly more demanding. Consider the multifaceted workflow of bacterial pathogen characterization: following specimen collection, isolates are cultured. Typically, culturing occurs on a selective medium that promotes the growth of a suspected pathogen, with colony growth and morphological characteristics being the first stage in determining identity. Further characterization to determine the species involves mass spectrometry as well as Gram staining and biochemical assays. Species identification then facilitates the appropriate susceptibility testing for a particular group of pathogens. The above process can take days for rapidly growing bacteria, or months for slow-growing species such as *Mycobacterium tuberculosis*. Subsequently, depending on the species, a subset of isolates can be selected for further bacterial typing. Here we present a rapid, simple workflow solution for bacterial typing by WGS that may be used in the future to replace current complex and labor-intensive techniques.

Figure 1. Minimum spanning trees for sequence targets obtained from de novo assembled WGS data of three community outbreak strains. (A) Sequence information from 32 targets (comprising MLST/eMLST/AST/4CMenB/AR) resulted in a single-allelic discrimination between strains DE9622 and DE9686/DE9938, with strains DE9686 and DE9938 indistinguishable. (B) Sequence information from 1,241 targets better resolved the three strains, with DE9686 separated by 13 and 22 alleles from DE9938 and DE9622, respectively.
Applications of bacterial typing
Bacterial typing can be considered the forensics of bacteria and is used to provide an isolate-specific molecular fingerprint. The molecular fingerprint can be used to identify different types of organisms within the same species and is useful in epidemiological applications that seek to understand the source and transmission route of bacterial spread. These investigations can be global in nature or can involve local surveillance within a region or a hospital. For instance, the Ion PGM System was utilized during investigation of a putative multidrug-resistant Escherichia coli, and used to resolve the relationships between strains of methicillin-resistant Staphylococcus aureus in a hospital setting [1,2].

The sequence information acquired through bacterial typing can aid in research to profile microbes for antibiotic resistance and surveillance for genes that encode vaccine antigens. Antibiotic resistance gene surveillance was assessed using the Ion PGM System for outbreak-associated Klebsiella pneumoniae isolates and teicoplanin resistance emerging during treatment of vanB-type vancomycin-resistant Enterococcus faecium infections [3,4].

Further, bacterial typing data can be used as evidence in medical legal cases that seek to understand the transmission of infections that result from contaminated food or beverages, or for tracking the origin of bacterial agents of bioterrorism such as the 2001 anthrax attacks in the United States.

MLST+ offers the comparability and standardization of MLST combined with the discriminatory power of PFGE
One current method for bacterial typing is MLST, the sequencing of 5–7 housekeeping genes with a genome-wide distribution. The combination of alleles for the target locus defines the sequence type (ST) for a particular strain, with phylogenetically related STs grouped in clonal complexes (CC). MLST and multilocus VNTR (variable nucleotide tandem repeat) analysis (MLVA), the analysis of short tandem repeats, are techniques that facilitate intra-laboratory data comparison and standardization through publicly available and expandable nomenclature for STs, but often lack sufficient discriminatory power at the sequence level for accurate identification of all bacterial strains. In contrast, another current typing method, pulsed field gel electrophoresis (PFGE), does offer adequate discriminatory power for bacterial strains but provides no specific DNA sequence information, is difficult to standardize across laboratories, and is labor-intensive.

Ion TorrentTM semiconductor sequencing was crucial in the rapid characterization of WGS of the enterohaemorrhagic E. coli O157:H7 outbreak in northeastern Germany, allowing researchers to respond quickly to a serious public health risk [5,6]. Standard MLST analysis was unable to differentiate this outbreak isolate from an unrelated isolate collected in 2010. However, analysis of WGS data indicated that the 2011 and 2010 isolates differed extensively in sequence, with important differences between the isolates as a result of homologous recombination events [7].

Combining the easy-to-implement, cost-effective, scalable microbial genome sequencing workflow of the Ion PGM System with the automated Ridom SeqSphere+ Software enables researchers to leverage the entire genome and use hundreds to thousands of genes for bacterial typing. As a result, MLST+ has greater discriminatory power when compared to MLST and MLVA, and the burdensome issues associated with PFGE are alleviated, resulting in a workflow solution ideal for routine as well as outbreak microbial monitoring.

Experimental design
Neisseria meningitidis, a gram-negative commensal bacterium, is one of the etiological agents for bacterial meningitis. This invasive disease is of importance due to the substantial fatality rate of infected young; there are an estimated 50,000 deaths annually worldwide of children under the age of 5 years. Further, the spread of N. meningitidis has an increased risk of outbreaks and even the possibility of epidemics. As a consequence, many industrialized countries have established surveillance laboratories commonly employing MLST to type seven housekeeping genes (abcZ, adk, fumC, gdh, pgi, pgl, rpmC). Typing of N. meningitidis by MLST can be sufficient to define the clonality of a strain, while the inclusion of 13 additional housekeeping genes—known as extended MLST (eMLST)—is often used to resolve phylogenetic relationships. Further testing of loci (pomA and ploB) is often performed to associate sequence information to antibiotic resistance (AR). Serogroup and antigen sequence typing (AST) is performed for PorA, PorB, and FetA, as well as sequencing genes that encode targets for protein-based vaccine development (e.g., 4CMenB vaccine).

To highlight the MLST+ workflow, we assessed three strains from a long-lasting meningococcal serogroup B community outbreak using the Ion PGM System and Ridom SeqSphere+ Software (Supplementary Information). The strains DE9622, DE9686, and DE9938 were isolated in 2003 and 2004 as part of an outbreak caused by strains of ST-42 (ST-41/44 cl) with PorA variable region 1 (VR1) being 7-2, VR2 being 4, and FetA being F1-5 [8].

Results
Using the Ion PGM™ Sequencing 200 Kit v2 and the Ion 318™ Chip v2, sequence information from 32 targets—comprising MLST, eMLST, AST, 4CMenB, and AR—was used in the initial analysis to replicate previous results on the Ion PGM System with an earlier version of the 200-base chemistry [9]. Two additional genetic (porA-porB, partials) and three protein (PorA/FetA/FHbp) sequence typing results were inferred from the 32 targets using de novo assembled WGS data from the three community outbreak meningococcal serogroup B strains. The results indicated a single allele discriminating between strains DE9622 and DE9686/DE9938 (Figure 1A). Strains DE9686 and DE9938 were independently confirmed by Sanger sequencing [9]. Typing by allele instead of SNPs ensures that this intragenic recombination event was counted as a single mutation event between DE9622 and the other two strains (Figure 3). Bacterial strain typing by a unique combination of target alleles on a genome-wide, gene-by-gene approach rather than an SNP-based approach better accounts for intragenic recombination events and the polymorphic nature of many bacterial species [10]. Further, SNP selection approaches can be subjective and difficult to reproduce and standardize.

Figure 2. Comparison table generated for community outbreak strains DE9622, DE9686, and DE9938. Twelve out of thirteen eMLST targets were identical, with sample DE9622 differing from the other two strains for the carB target.

Figure 3. A comparison of the carB gene target between community outbreak strains DE9622, DE9686, and DE9938. A single intragenic recombination event of 2,253 bp in carB comprises 14 SNPs between DE9622 and DE9686/DE9938.
as well as to add to an expandable nomenclature. Similarly, other genome comparison methods, such as average nucleotide identity/k-mer and pairwise genome comparison, are not additively expandable as classification approaches.

The same de novo assembled WGS data from the three community outbreak strains were analyzed for MLST+ of 1,241 targets. Following autocorrection of homopolymer insertion and deletion (indel) errors, this highly discriminatory approach was used for typing, with discernment between the three outbreak stains significantly improved by MLST+ compared with 32-target typing (Figure 1B). In addition, MLST+ analysis of N. meningitidis strains provides sequence information for retrospective sequence data mining.

Conclusions
The Ion PGM System, in combination with Ridom SeqSphere+ Software, provides a set of reagents and software to enable rapid bacterial typing in less than a day. MLST+ dramatically increases the discriminatory power of bacterial typing while eliminating the time-consuming PCR and Sanger sequencing used to generate traditional MLST data. Features of the Ion PGM System that support replacement of other complex typing techniques with a single efficient workflow include scalability, affordability, simplicity, and speed.

The democratizing attributes of the Ion PGM System will empower a shift to real-time surveillance from the current focus on retrospective outbreak investigation. With prospective epidemiology a possibility through the networking of small- and medium-sized laboratories, the Ion PGM System will truly facilitate the opportunity of global surveillance and early-warning outbreak detection needed for an increasingly connected world. The Ion PGM System and Ridom SeqSphere+ Software enable MLST+, a single method for any bacteria, any study type, and any research lab.

Intuitive and simplified microbial typing analysis
Ridom SeqSphere+ Software removes the complexity of microbial typing, enabling MLST+ sequencing projects. Bioinformatics are made simple with this user-friendly software that requires no scripting skills and can be automated using pre-designed and customizable templates to coordinate all the data processing steps necessary to complete the analysis of hundreds of strains. Further, users can participate in an expanding nomenclature service by submitting high-quality sequences to a publicly available database of worldwide epidemiological data that seeks to establish a common molecular language for whole-genome bacterial typing.

Adaptable microbial typing solution to suit a wide range of bacteria and study types
Unlike current bacterial isolate typing techniques, WGS and Ridom SeqSphere+ Software can be applied to all bacterial population structures, which can vary from populations exhibiting low levels of genetic diversity with near clonality for genes at different loci (monomorphic), to those with high levels of genetic diversity that display a nearly random association between loci (panmictic).

In addition, Ridom SeqSphere+ Software can be applied to a wide variety of study types, ranging from outbreak/transmission investigations to population genetic studies, as well as to understanding the phylogenetic relationships between bacterial species.

Features
• Automated typing of bacteria with user-defined quality parameters such as coverage, intragenic stop codons, and frameshift detection using public or self-defined query libraries and task templates
• Designed for both individual labs and distributed workgroups (client/server model), it facilitates automatic processing and analyzing of next-generation sequencing (NGS) and Sanger capillary electrophoresis (CE) sequence data
• Edits and analyzes NGS-derived de novo (.ace) and reference-mapped assemblies (.bam) with incorporated auto-correction of homopolymer-related indel errors. Additionally assembles, edits, and analyzes Sanger CE sequencing data obtained from traditional MLST
• Analytical tools for epidemiological, evolutionary, and functional analysis, with cluster and data visualization using minimum spanning or UPGMA/neighbor joining trees
• Integrated database for the storage, retrieval, and export of data, with the ability to search new sequence entries against stored data and create reports from your epidemiological and DNA sequencing data. Data fields are compliant with the metadata requirements of external databases such as NCBI BioSample
• Security enabled with encryption (SSL) of data in transmission; configurable user roles, user groups, and access controls; audit trail functionality
References

Methods
Genomic DNA from three outbreak strains was used to prepare fragment libraries using the Ion Xpress™ Plus Fragment Library Kit (Figure S1). Libraries generated from these samples were clonally amplified with the Ion PGM™ Template OT2 200 Kit and the Ion OneTouch™ 2 System for template preparation and enrichment prior to chip loading. Sequencing was performed using the Ion 318 Chip v2 and the Ion PGM Sequencing 200 Kit v2, with a total sequencing time of 4.4 hours. Conversion of raw signal to base calls was performed on the Torrent Server using Torrent Suite™ Software, which incorporates algorithmic improvements in both raw accuracy and variant calling, even in challenging regions such as long homopolymers.

Output files were used for subsequent de novo assembly using MIRA v3.9.4 on the Torrent Server as a Torrent Suite™ plug-in. While slower than reference-guided approaches, de novo assembly is hypothesis-free and is frequently used for bacterial species with high levels of genetic diversity that display a nearly random association between loci (panmictic species such as H. pylori) and is required for the identification of mobile elements and plasmids, which frequently encode virulence factors and AR genes. Benchmarking experiments indicate that the MIRA assembler performs consistently better than other methods with Ion PGM™ data (Figure S2). Subsequent to de novo assembly, Ridom SeqSphere+ Software analysis required a further 2 minutes per genome using a computer with 16 GB of RAM, which is computing power that is affordable and readily available to most scientists.

Figure S1. Microbial-genome sequencing workflow using the Ion PGM System and Ridom SeqSphere+ Software. Bacterial typing can be performed in a rapid <24-hour workflow from library construction to mapped results. After culturing and gDNA extraction, library construction can be accomplished in 2 hours using the Ion Xpress Plus Fragment Library Kit. Automated template preparation can then be performed in 6–7 hours using the Ion OneTouch 2 System. On the Ion PGM Sequencer, 400 bp sequencing runs are completed in just 3.7 hours with the Ion 314™ Chip and 7.3 hours with the Ion 318 Chip. Primary data analysis is performed using Torrent Suite Software, and de novo assembly is accomplished in 4–5 hours, depending on the organism and computer, using the MIRA assembler plug-in. Following assembly, Ridom SeqSphere+ Software provides an easy-to-use interface for rapid, automated MLST+ typing in 2–30 minutes, depending on the organism and computer.

Figure S2. Benchmarking experiments comparing the performance of de novo assembly algorithms. Shown are 200-base and 400-base chemistry reads using Ion PGM sequencing kits for S. aureus (COL strain) and E. coli (Sakai strain) as well as 400-base chemistry reads for M. tuberculosis (H37). Note the sequence read data were down-sampled to 40X coverage as supported by data generated for a recent NGS platform comparison [11]. Contig NGA50 values indicate that MIRA and Newbler consistently perform better than other assembly methods.