

Ridom™ SeqSphere+ Software



Microbial Genomics For Outbreak-, Resistom-, Virulome-Investigation, and Real-time Surveillance

ridom
BIOINFORMATICS

Combining the best – Comparability of MLST and discriminatory power of PFGE

Ridom SeqSphere+ Software is designed for distributed work-groups (client/server model) and allows automatic processing and analyzing of NGS (e.g., Illumina, Ion Torrent, or PacBio) and Sanger capillary-electrophoresis sequence data. SeqSphere+ is the single solution for easy and automated microbial analysis; enabling your lab to employ whole genome microbial typing (*Nat Biotechnol.* **31**: 294, 2013).

Bacterial epidemiological typing generates isolate-specific geno- or phenotypic characters that can be used to elucidate the sources and routes of spread of bacteria. Typing may also be used to identify emerging pathogenic strains or clones within a species, including agents of bioterrorism (*Clin Microbiol Infect.* **13** [Suppl. 3]: 1, 2007).

A variety of methods have been developed to generate isolate-specific fingerprints for epidemiological genotyping. However, until recently there was no single typing method available to address all bacterial species, different population structures (monomorphic vs. panmictic), and all study types (e.g., evolutionary/phylogenetic, population genetics, or transmission chain/outbreak investigation). For decades, the common typing method was based on multi locus sequence typing (**MLST**) of 5-7 housekeeping genes, pulse field gel electrophoresis (**PFGE**), or multi locus VNTR analysis (**MLVA**). Although widely adopted, these methods either lack discriminatory power or are labor intensive and difficult to standardize.

Now, microbiologists everywhere can experience the genomic revolution that benchtop next generation sequencing (**NGS**) provides. With fast and affordable microbial whole genome shotgun sequencing (**WGS**) and automatized software analysis (**Fig. 1**), microbiologists can use genome-wide gene-by-gene allele calling of hundreds/thousands of genes (core genome MLST [**cgMLST**]) for genotyping (**Fig. 2**). cgMLST schemes consist of a fixed set of conserved genome-wide genes. Alleles are used instead of single nucleotide polymorphisms (SNP) or concatenated sequences to mitigate the effects of recombination and to enable for a global and public nomenclature (**cgMLST.org** service). The high discriminatory power of cgMLST, coupled with rapid and simple WGS workflow, makes this complete solution ideal for everyday usage in a routine laboratory, e.g., for prospective surveillance of multi-drug resistant bacteria in a hospital (*JCM* **54**: 2874, 2016). That is, **one disruptive, accurate, and highly reproducible technology** (*JCM* **55**: 908, 2017) **that fits any bacteria, any research topic, and any lab** is finally within reach for every microbiologist.

Ridom SeqSphere+

User friendly

No scripting or bioinformatics skills needed to process your data and run your analyses.

Highly automated workflow

Download pre-defined typing schemes or quickly create your own scheme based on a reference genome or allele libraries. Setup an assembling and processing pipeline to analyze hundreds of your samples without any further user intervention.

Global nomenclature

Submit your high quality allelic profiles and contribute to the world-wide unique cgMLST.org nomenclature service that ensures for a 'Molecular Typing Esperanto' derived from WGS data.

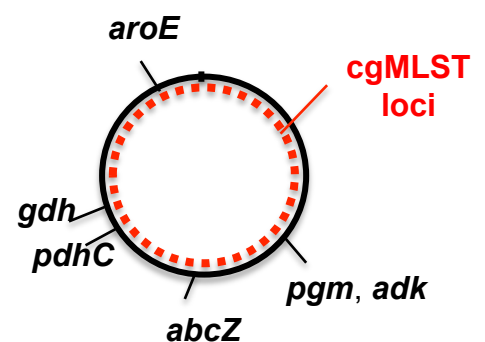


Figure 2. MLST of 6 genes versus genome-wide cgMLST. Shown are the 6 MLST targets for *N. meningitidis* and red dots across the genome that represent symbolically the cgMLST loci.

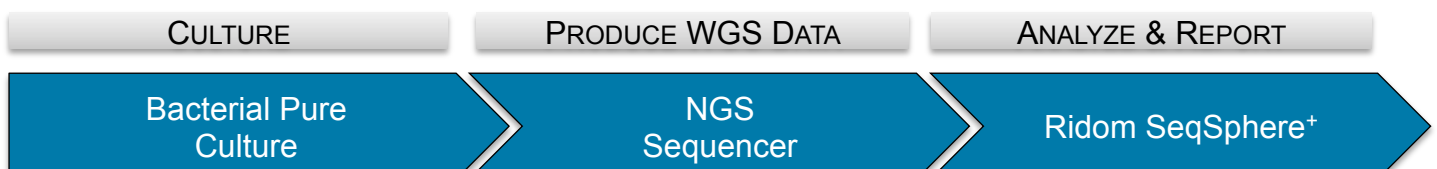


Figure 1. Microbial WGS next generation sequencing and genome-wide gene by gene (cgMLST) analysis workflow.

Elevate your microbial research to the next level — www.ridom.de

Ridom SeqSphere⁺ Key Features

DNA re-sequencing editor

Assemble raw reads (FASTQ) with integrated Velvet, SPAdes, and BWA algorithms. Read and analyze assembly files (ACE, BAM, FASTA). Autocorrect homo-polymer related InDel errors. Edit and analyze NGS or Sanger DNA sequence data.

Pipeline for automated sequence analysis

Define and start a pipeline to trim by quality, down-sample, assemble, analyze, and type hundreds of NGS data at once automatically; e.g. by fetching raw reads from the benchtop sequencer as soon as data are generated. Define quality criteria (e.g., % of good cgMLST targets) for sample success. Correct handling of repeated samples is built-in to the pipeline.

Characterize bacteria

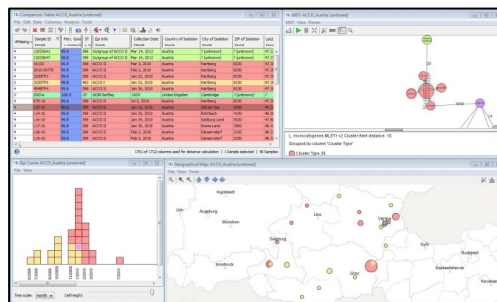
Type bacteria automatically with user defined QC parameters (e.g., coverage or frame shift). Genome-wide allele and SNP calling from WGS data either on core genome and/or accessory genome level is implemented. Use public schemes for genotyping or define them by using the incorporated cgMLST Target Definer. User-defined virulome and resistome analysis is also supported.

Database

Store, search, retrieve, export, and create reports from experiment, epidemiologic, and DNA sequence data stored in an integrated database. Data fields are compliant with the meta-data requirements of the EBI European Nucleotide Archive (ENA). Compare new sequence entries against stored data. Get automatically cluster alerts of possible outbreaks. Manage and backup all data (sequence and epi-data). Download from NCBI complete/draft genomes or SRA reads. Submit with just one click raw read- and epi-data to EBI ENA.

Analytical tools

Visualize place, time, 'person', and type dimensions in a comparison table with built-in GIS, epi-curve, and phylogenetic tree (among others the minimum spanning tree algorithm is supported) functionality (**Fig. 3**). All dimension views are inter-linked and exportable in publication quality format (SVG and EMF). Store tree coloring, topology, and selection of samples in a comparison table snapshot. Search for group-specific SNPs, e.g., for designing an outbreak-strain specific PCR screening assay.



Community

- Download certified typing schemes online.
- Rapidly and easily share typing schemes with others.
- Contribute optional to the world-wide expanding publicly available cgMLST.org server of allele nomenclature.
- Request a non-public epidemiologic data server for a closed group of users.

Security/Privacy

- Encryption (SSL) of all data in transmission.
- Configurable anonymization of meta-data like place and time before making data public (cgMLST.org or ENA submission).
- Various configurable user roles, user groups, and access controls.
- Audit trail functionality (who did, what and when).
- Optional database and backup encryption.

Figure 3. Place, time, 'person' and type dimensions visualized with SeqSphere⁺. Screenshot of *Listeria monocytogenes* outbreak data from [JCM 53: 2869, 2015](#).

Demonstration videos, step-by-step tutorials with demonstration data, and free trial version are available at www.ridom.de

Minimum requirements for evaluation: Microsoft Windows 64-bit, 8 GB RAM. Client and server can be installed on the same computer.

Computer Specifications (recommended)

Client	Microsoft Windows 64-bit or Linux 64-bit, quad-core processor, 16-32 GB RAM, 500 MB hard disk (HD) , and Internet connection (via ports 80 & 443) required for license activation and to use public Ridom cgMLST Nomenclature Server and Task Template Sphere.
Server	Microsoft Windows 64-bit or Linux 64-bit, quad-core processor, 16 GB RAM, 1-2 TB HD (with default settings 10,000 samples require approx. 1 TB on a single hard disk), for database backups and optional storage of assembly files (ACE/BAM) a separate server HD or NAS is highly desirable.

Please note that most computation is done on the client computer (e.g., assembling). Therefore, computational power and amount of RAM is more important on the client than on the server computer. Client and server communicate by default via port 8064 (configurable) over the Intra-/Internet. - An unlimited number of client computers can process the same pipeline script for **speeding-up** the analysis for example of a MiSeq run. For **scaling-up** simply also add another client computer (one client computer can easily cope with the output of one benchtop sequencer run in 24/7 mode). To use the power of a **Linux-Cluster** the SeqSphere⁺ client pipeline can be started in command-line mode. SPAdes is only supported by the Linux-Client. For Illumina data the in SeqSphere⁺ integrated Velvet or SPAdes *de novo* assemblers are recommended. For Ion Torrent and for Pacific Biosciences data the usage of SPAdes or HGAP combined with the Celera assembler are the options of choice, respectively.

Licensing

Standard license 1 Year / 2 User Accounts: 2,300 EUR (for academic/government users).

Prices for more user accounts, other time periods, and commercial licenses are available on request.

The license is limited to a **time period** and a number of named user accounts on the server. During the license period an **unlimited number of samples can be analyzed**. Each **named user account** allows to login once at the same time. If somebody else logs in with the same account, a read-only mode is activated for this session. The Pipeline Mode also runs with a named user account but can run concurrently. So the same user can be logged in in the normal mode, and can run several pipelines at the same time. Client software can be installed on an unlimited number of computers. The license period starts with the activation of the license. The license will expire after the defined period and the software runs with read-only access unless a new license is ordered. Read-only access still allows to view and export existing data. License includes **updates** and **free technical support** for installation and troubleshooting of the software within the time period.

FOR RESEARCH USE ONLY.

© 2017 Ridom GmbH, Münster Germany (document version 4.0). The trademarks mentioned herein are the property of Ridom GmbH or their respective owners.

