



The role of bioinformatics in microbial analysis

UNIVERSITY OF VETERINARY MEDICINE, FOUNDATION, HANNOVER



founded in 1778

6 clinics
20 institutes



Institute of Food Quality and Food Safety

- **Department of Foodborne Zoonoses**
- Department of Milk Science
- Department of Meat technology
- Department of Food Microbiology

Department of Foodborne Zoonoses:

Functions:

- **Lectures of veterinary medicine and biology**
- **Services**
- **Research projects**

Functions:

→ Services

- Detection characterization of food borne pathogen
- Identification of animal species in meat and meat products
- Sex determination of the meat animal
- Detection of genetic modified organism GMOs
- Detection of Allegens like:
 - Wheat
 - Celery
 - Mustard
 - Nuts
- Quantitative detection of probiotics

Selected research projects

FOOD-PCR (EU-Project)

“Validation and standardization of diagnostic polymerase chain reaction for detection of food-borne pathogens- FOOD-PCR”

BSE-Risk material

„Novel molecular method for detection of bovine-specific central nervous system tissues as BSE risk material in meat and meat products “

Characterization of novel bacterial species of different genera from various habitats

Development of Loop-mediated isothermal amplification assays as a rapid method for the immediate detection of pathogens in the case of food infections and intoxications

The role of bioinformatics in microbial analysis

The role of bioinformatics in microbial analysis

Introduction

- **Definition of bioinformatics**
- **Importance in microbial research**
- **Examples**

Introduction to bioinformatics

The bioinformatics

Paulien Hogeweg and Ben Hesper coined it in 1970



Informatics

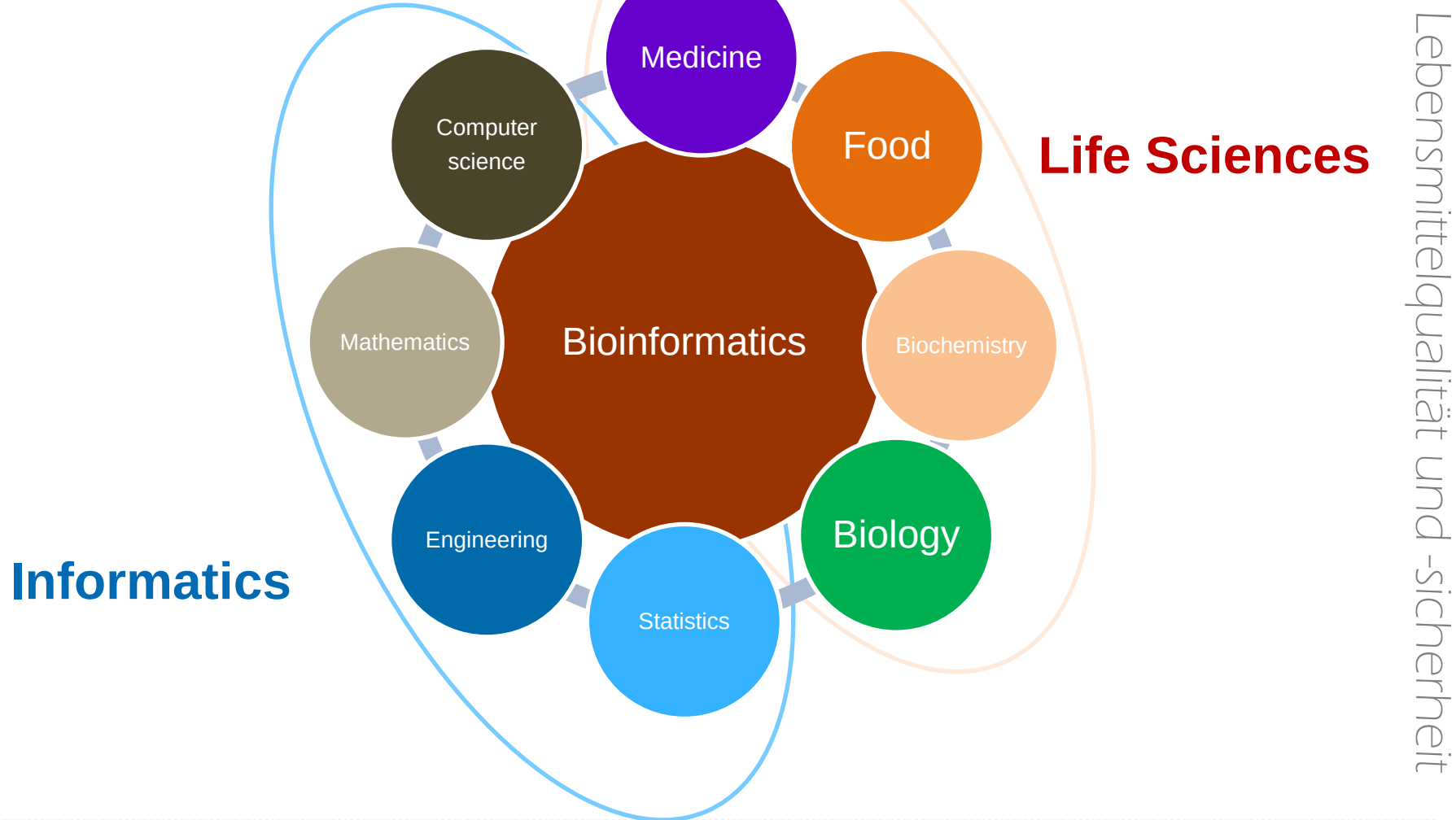


Life Sciences

Complex data from the life sciences can be **processed**, **presented** and **used** with the help of bioinformatics

Introduction to bioinformatics

Tasks in bioinformatics



Focus of today's presentation

- **Connection to microbial analysis issues**
- **Relevant databases**
- **Characterisation of bacterial isolates**
- **Further analysis tools for sequence data**
 - **BioNumerics**
 - **GGDC / TYGS**
- **Next-generation sequencing (NGS) and Metagenomics**

Connection to microbial analysis issues

- Identification of new species (WGS)
- Species characterization
- Microbiological profile of samples (NGS)
- Tracking pathogens along the process chain
 - Input sources, prediction models
- Authenticity control of food
 - GMO, animal species differentiation



Focus of today's presentation

- **Connection to microbial analysis issues**
- **Relevant databases**
- **Characterisation of bacterial isolates**
- **Further analysis tools for sequence data**
 - **BioNumerics**
 - **GGDC / TYGS**
- **Next-generation sequencing (NGS) and Metagenomics**

Databases

Data obtained as part of research projects is usually stored in large databases and can thus be retrieved and processed for subsequent use.



Genome sequencing and bioinformatics

Databases

Examples

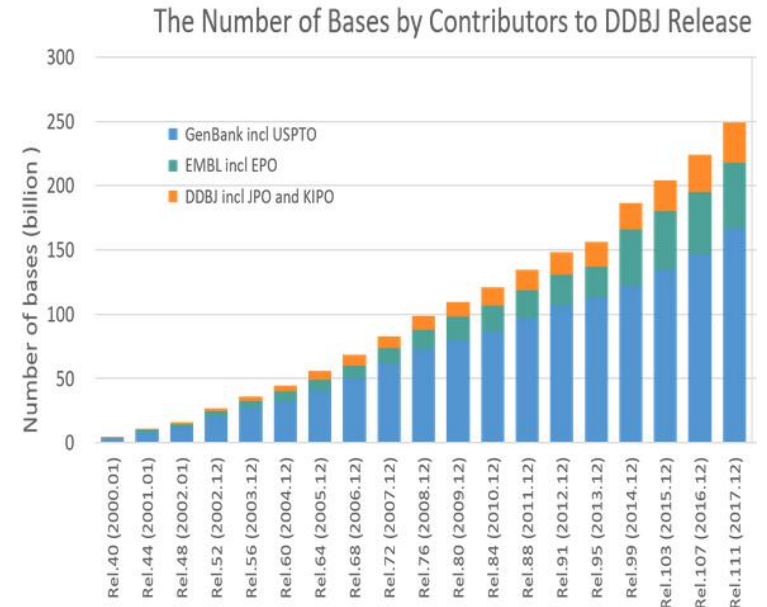
- SepsiT^{est}™ BLAST
- The European Molecular Biology Laboratory (EMBL)
- European Bioinformatics Institute (EBI)
- DNA Data Bank of Japan (DDBJ)
- The National Center for Biotechnology Information
 - NCBI databases (<https://www.ncbi.nlm.nih.gov/>)

SepsiT^{est}™ BLAST

Broad-range Detection & Identification of Bacteria, Yeast and Fungi



Genome sequencing and bioinformatics



International Nucleotide Sequence Database Collaboration (INSDC)
Bioinformation and DDBJ Center

Lebensmittelqualität und -sicherheit

Genome sequencing and bioinformatics

NCBI databases

<https://www.ncbi.nlm.nih.gov/>

An official website of the United States government [Here's how you know](#)

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

All Databases Search

NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

Welcome to NCBI
The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.
[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

Analyze
Identify an NCBI tool for your data analysis task

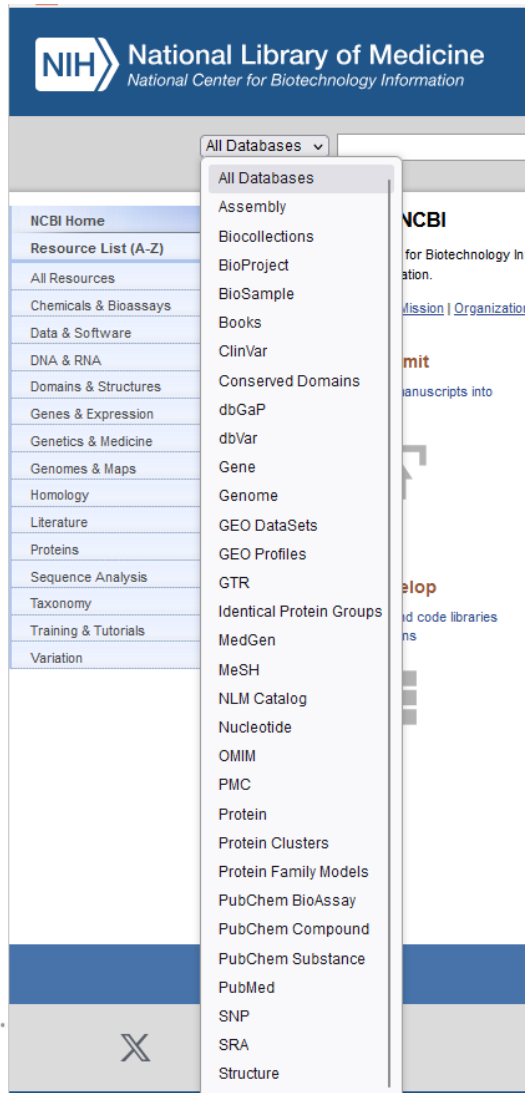
Research
Explore NCBI research and collaborative projects

Popular Resources
PubMed
Bookshelf
PubMed Central
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem

NCBI News & Blog
dbSNP Build 157 Release 18 Mar 2025
RefSNP (rs) exceed 1.2 billion records. We are pleased to announce the release of the Database of Single Nucleotide
RefSeq Release 229 is Now Available! 13 Mar 2025
Check out RefSeq release 229, now available online and from the FTP site. You can access RefSeq data through NCBI

Genome sequencing and bioinformatics

NCBI-databases



The search can be specified in more detail and limited to certain databases, e.g.

- Nucleotide database
- PubMed database
- Taxonomy database

Genome sequencing and bioinformatics

Key Databases and Their Size NCBI-databases

GenBank (NCBI's DNA Sequence Database)

- Current Size: Over 2 trillion nucleotides from hundreds of millions of sequences.
- Growth: Data doubles approximately every 18 months.

1 Petabyte (PB) = 1.000 Terabyte (TB)

1 Petabyte (PB) = 1.000.000 Gigabyte (GB)

1 Petabyte (PB) = 1.000.000.000 Megabyte (MB)

Genome sequencing and bioinformatics

Key Databases and Their Size NCBI-databases

PubMed (Biomedical Literature Database)

- **Number of Articles:** Over **36 million** citations.
- **Annual Growth:** More than **1 million** new citations are added yearly.
- **Daily Queries:** Over **3.5 million** searches per day.



Genomsequenzierung und Bioinformatik

NCBI-Datenbanken

Search NCBI × Search

Results found in 30 databases

GENOME ASSEMBLY Was this helpful? 👍 👎

ASM19603v1
Listeria monocytogenes EGD-e (firmicutes)
 European Consortium (April 2011)
 RefSeq GCF_000196035.1
[PubMed \(1\)](#)

[Entrez Genome](#) [BLAST](#) [Download](#)

Use NCBI Datasets for bulk downloading of genome sequence and annotation data.
[NCBI Datasets](#) [Command-line tool](#) [API documentation](#)

Assembly statistics +

Search using "All databases"

Literature	
Bookshelf	485
MeSH	47
NLM Catalog	65
PubMed	20,858
PubMed Central	44,086

Genes	
Gene	7,108
GEO DataSets	2,562
GEO Profiles	57,589
HomoloGene	2
PopSet	443

Proteins	
Conserved Domains	58
Identical Protein Groups	1,307,141
Protein	9,566,127
Protein Family Models	237
Structure	527

Genomes	
Assembly	39,302
BioCollections	0
BioProject	1,024
BioSample	52,792
Genome	1
Nucleotide	344,990
SRA	55,466
Taxonomy	1

Clinical	
ClinicalTrials.gov	26
ClinVar	0
dbGaP	1
dbSNP	0
dbVar	0
GTR	0
MedGen	8
OMIM	75

PubChem	
BioAssays	1,003
Compounds	1
Pathways	247
Substances	8

Genomsequenzierung und Bioinformatik

NCBI-Datenbank

Items: 1 to 20 of 344990

<< First < Prev Page 1

- ☐ [Listeria monocytogenes strain NH1, complete genome.](#)
1. 3,002,491 bp circular DNA
 Accession: NZ_CP021325.1 GI: 1358004970
[Assembly](#) [BioProject](#) [BioSample](#) [Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

Publication data, information on the sequence (accession no., origin, length, methods, etc.) and sequence data can be called up for each entry, downloaded and entered into other programmes

NCBI Resources How To

Nucleotide Nucleotide Advanced

COVID-19 Information
[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevent](#)

GenBank

Due to the large size of this record, sequence and annotated features are not shown. Use the "Customize view"

Listeria monocytogenes strain NH1, complete genome.

NCBI Reference Sequence: NZ_CP021325.1
[FASTA](#) [Graphics](#)

Go to:

LOCUS	NZ_CP021325	3002491 bp	DNA	circular	CON 10-JAN-2021
DEFINITION	Listeria monocytogenes strain NH1 chromosome, complete genome.				
ACCESSION	NZ_CP021325				
VERSION	NZ_CP021325.1				
DBLINK	BioProject: PRJNA224116 BioSample: SAMN06920120 Assembly: GCF_002969195.1				
KEYWORDS	RefSeq.				
SOURCE	Listeria monocytogenes				
ORGANISM	Listeria monocytogenes Bacteria; Firmicutes; Bacilli; Bacillales; Listeriaceae; Listeria.				
REFERENCE	1 (bases 1 to 3002491)				
AUTHORS	Yan, H., Wang, Y., Zhou, W. and Shi, L.				
TITLE	Whole genome analysis of Listeria monocytogenes isolate from quick-frozen food				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 3002491)				
AUTHORS	Yan, H., Wang, Y., Zhou, W. and Shi, L.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-MAY-2017) School of Food Science and Engineering, South China University of Technology, Wushan Road, Guangzhou, Guangdong 510641, China				
COMMENT	REFSEQ INFORMATION: The reference sequence is identical to CP021325.1 . The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/ Bacteria and source DNA available from School of Food Science and				

Genomsequenzierung und Bioinformatik

NCBI-Datenbank

National Library of Medicine
National Center for Biotechnology Information

Log in

Nucleotide

Species

Animals (261)
Plants (91)
Fungi (327)
Protists (170)
Bacteria (3,676)
Viruses (85)
Customize ...

Molecule types

genomic DNA/RNA (4,602)
mRNA (7)
Customize ...

Source databases

INSDC (GenBank) (2,621)
RefSeq (1,992)
Customize ...

Sequence Type

Nucleotide (4,613)

Genetic compartments

Chloroplast (20)
Mitochondrion (81)
Plasmid (6)
Plastid (20)

Sequence length

Custom range...

Summary ▾ 20 per page ▾ Sort by Default order ▾

Items: 1 to 20 of 4613

<< First < Prev Page 1 of 231 Next > Last >>

☐ [Schizophyllum radiatum clone HSH. Mosul internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence](#)
591 bp linear DNA
Accession: OQ709247.1 GI: 2470442365
[Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Meloidogyne javanica isolate Mosul small subunit ribosomal RNA gene, partial sequence](#)
513 bp linear DNA
Accession: PP273501.1 GI: 2668910702
[Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Ganoderma sinense strain No. Ri. AI-1- Mosul small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence](#)
648 bp linear DNA
Accession: OQ379018.1 GI: 2439901348
[Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

Filters: [Manage Filters](#)

Results by taxon

Top Organisms [Tree](#)
Sinorhizobium medicae (351)
Enterococcus faecalis (340)
Citrobacter farmeri (340)
Enterobacter cloacae (310)
Morganella morganii (145)
All other taxa (3127)
More...

Find related data

Database:

Find items

Search details

Mosul[All Fields]

Search

See more...

Genomsequenzierung und Bioinformatik

NCBI-Datenbank

 An official website of the United States government [Here's how you know](#)


National Library of Medicine
National Center for Biotechnology Information

Log in

BLAST®

Home Recent Results Saved Strategies Help

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

NEWS

Mon, 17 Mar 2025
Improvements include upgrading to GCP Artifact Registry and better handling of job completion status in kubernetes version 1.30+.
ElasticBLAST 1.4.0 is now available! [More BLAST news...](#)

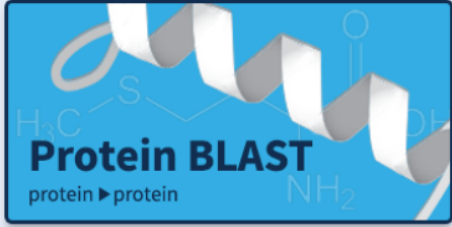
Web BLAST



Nucleotide BLAST
nucleotide ► nucleotide

blastx
translated nucleotide ► protein

tblastn
protein ► translated nucleotide



Protein BLAST
protein ► protein

BLAST Genomes

Genome sequencing and bioinformatics

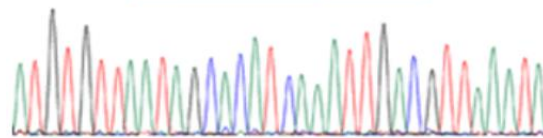
Pathogenomix

Infection intelligence

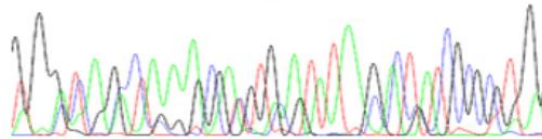
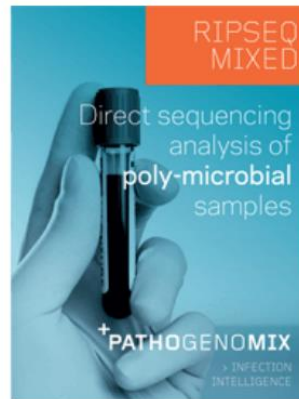


Lebensmittelsicherheit und -sicherheit

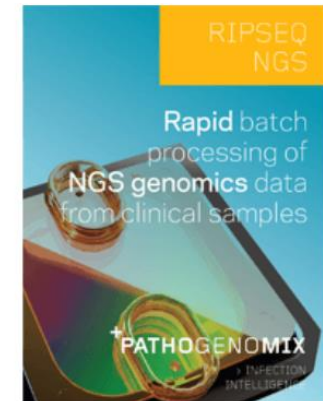
Rapid & accurate pathogen ID



Sanger Mono



Sanger Poly



NGS Targeted

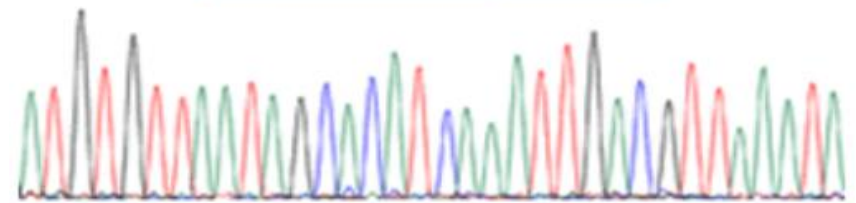
Pathogenomix

Infection intelligence

Rapid & accurate pathogen ID

Using this bulk analysis you can save a lot of time on your pure samples - with bulk upload, bulk analysis against any of our curated reference databases, automatic confidence status flags and automatic report generation.

You do not have to sift through hundreds, or even thousands of references in GenBank anymore!



Sanger Mono

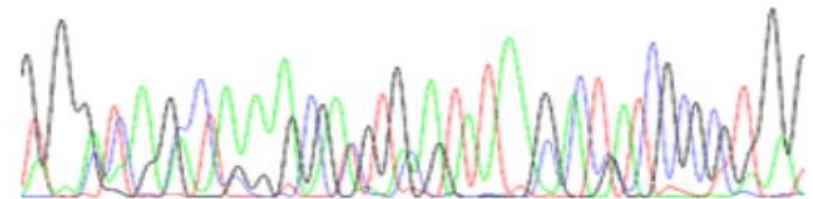
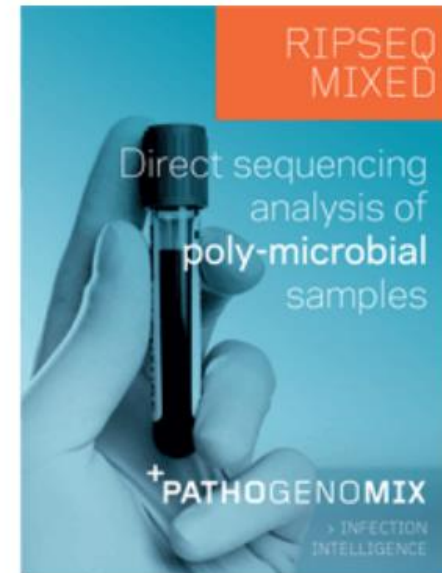
Pathogenomix

Infection intelligence

Rapid & accurate pathogen ID

No matter if you have a pure sample, or you have a more complex sample resulting in a mixed chromatogram - Pathogenomix Sanger Poly can also analyze samples you otherwise would have to discard or at best culture and separate, and not knowing if you missed anything.

No need to re-culture or discard a sample!



Sanger Poly

Pathogenomix

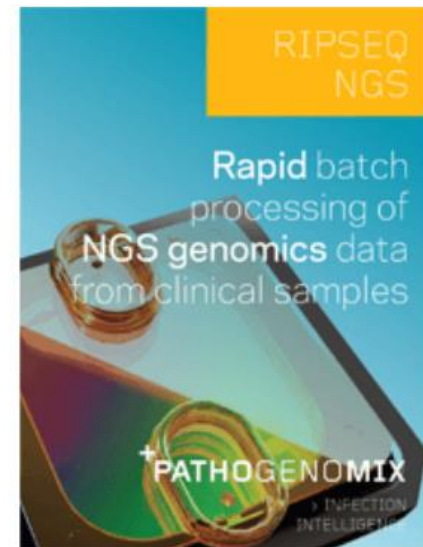
Infection intelligence



Rapid & accurate pathogen ID

Using this novel algorithms Pathogenomix NGS will process your targeted NGS sequence file locally, and reduce the information needed to upload by a factor of 1000:1, while still retaining resolution and specificity. This means no lengthy file upload, and no costly big data file storage. After upload, our cloud based software will analyze the uploaded reads against any of our curated reference databases - no matter the target.

Complete NGS analysis in less than 5 minutes! Including reporting!



NGS Targeted

Focus of today's presentation

- **Connection to microbial analysis issues**
- **Relevant databases**
- **Characterisation of bacterial isolates**
- **Further analysis tools for sequence data**
 - **BioNumerics**
 - **GGDC / TYGS**
- **Next-generation sequencing (NGS) and Metagenomics**

The role of bioinformatics in microbiological analysis

Characterisation of bacterial isolates

Characterisation of bacterial isolates

Fine typing and differentiation of microorganisms

The determination of the similarity or relationship of strains of a species, e.g. for epidemiological questions.

Example:

Between 1 and 30 July, *E. coli* O157:H7 was isolated in a city from:

- **A 3-year-old child with haemolytic uraemic syndrome**
- **faeces from cull cows**
- **beef**
- **Lettuce**

**The laboratory observed that all 4 isolates have the same biochemical profile.
Do you conclude that all 4 isolates are the same and could have come from the same source ?**

Characterisation of bacterial isolates

Fine typing and differentiation of microorganisms

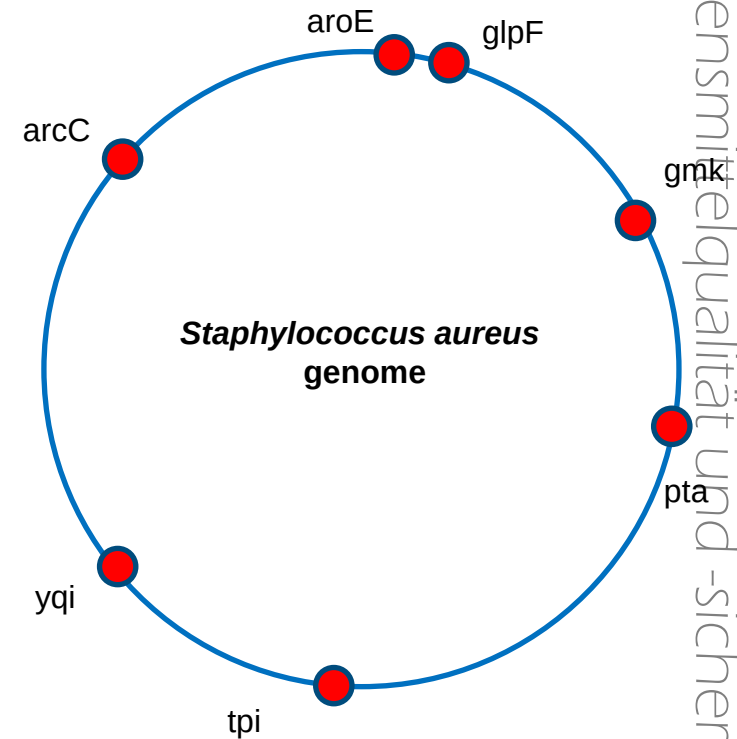
A number of methods have been developed:

- Pulsed field gel electrophoresis (PFGE)
- Multilocus sequence typing (MLST)
- Multilocus VNTR Analysis (MLVA)
-
-
- Next-Generation-Sequencing (NGS)

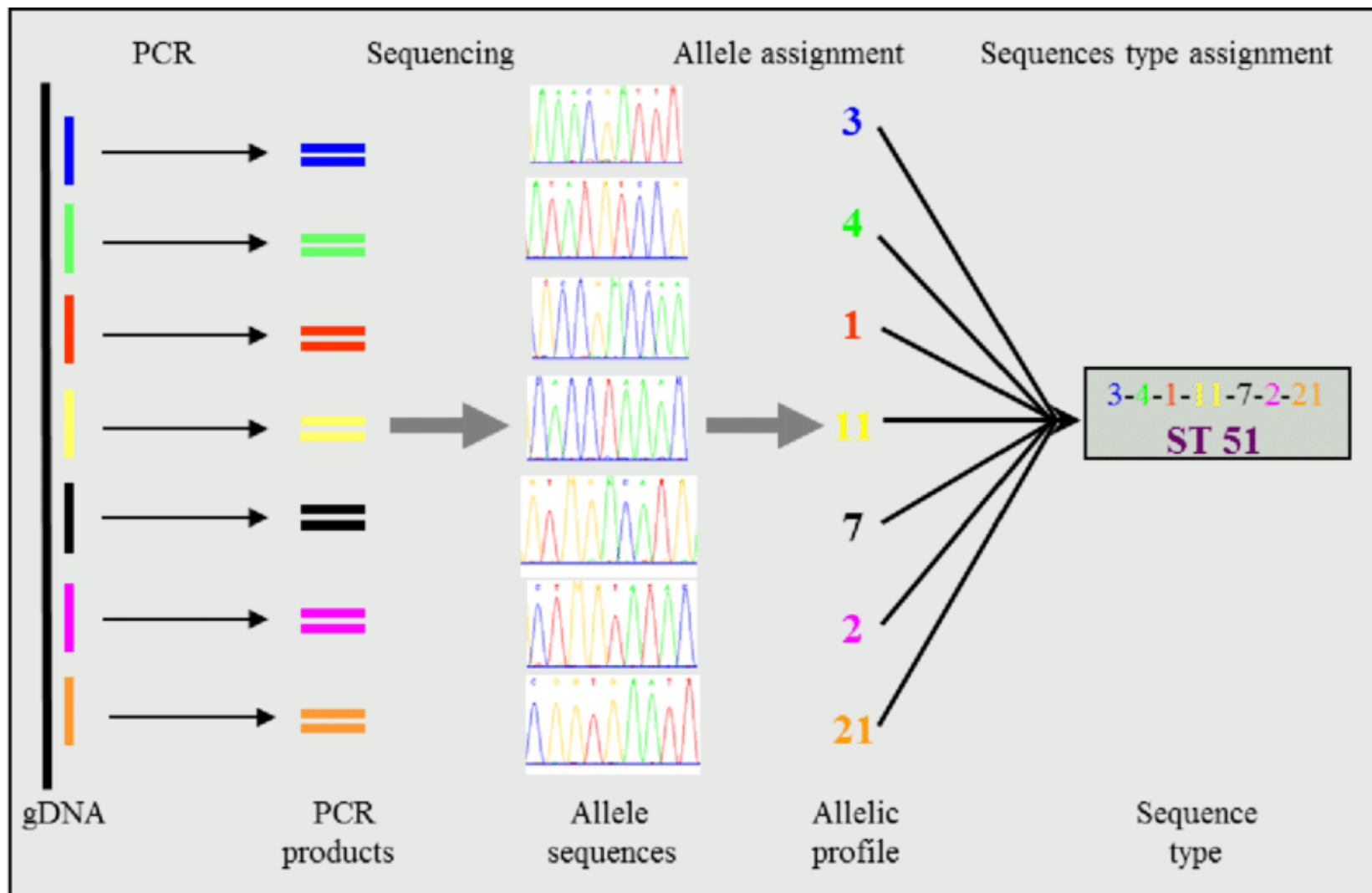
Multilocus sequence typing (MLST)

The *S. aureus* MLST scheme utilises internal fragments of the following seven housekeeping genes

- *arcC* (Carbamate kinase)
- *aroE* (Shikimate dehydrogenase)
- *glpF* (Glycerol kinase)
- *gmk* (Guanylate kinase)
- *pta* (Phosphate acetyltransferase)
- *tpi* (Triosephosphate isomerase)
- *yqi* (Acetyl coenzyme A acetyltransferase)



Multilocus sequence typing (MLST)



Werner Ruppitsch 2016

Multilocus sequence typing (MLST)

<https://pubmlst.org/>

PubMLST Public databases for molecular typing and microbial genome diversity [MY ACCOUNT](#)

[HOME](#) [ORGANISMS](#) [SPECIES ID](#) [ABOUT US](#) [UPDATES](#)

A collection of open-access, curated databases that integrate population sequence data with provenance and phenotype information for over 140 different microbial species and genera.

42,020,197 ALLELES	1,848,936 ISOLATES	1,342,184 GENOMES
-----------------------	-----------------------	----------------------

Organisms search

The role of bioinformatics in microbiological analysis



Manage. Analyze. Discover.

[SOFTWARE](#)
[SERVICES](#)
[ORDER](#)
[SUPPORT](#)
[DOWNLOAD](#)

[HOME](#) > [SOFTWARE](#) > BIONUMERICS: A UNIQUE SOFTWARE PLATFORM

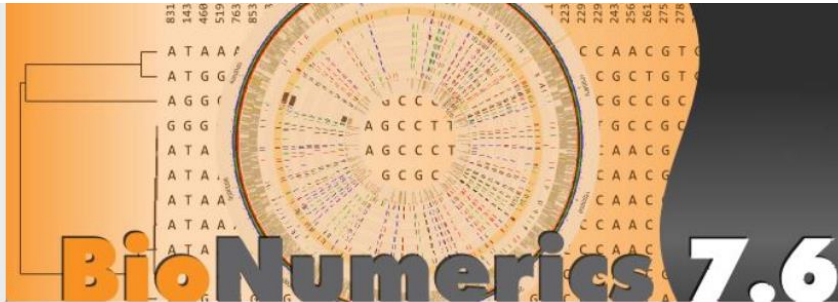


BioNumerics
@BioNumerics

Startseite

- Info
- Fotos
- Bewertungen
- BioNumerics
- Veranstaltungen
- Beiträge
- Community

[Seite erstellen](#)



BioNumerics 7.6

Gefällt mir Teilen ...

E-Mail senden Nachricht

Fotos

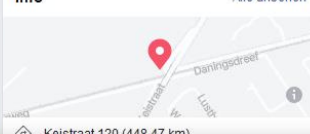


BioNumerics
Lokales Unternehmen in Sint-Martens-Latem
Durchgehend geöffnet

Community [Alle ansehen](#)

- 667 Personen gefällt das
- 680 Personen haben das abonniert
- 14 Besuche

Info [Alle ansehen](#)



Keistraat 120 (448,47 km)

[GET A FREE TRIAL](#)
[REQUEST A QUOTE](#)

• BIONUMERICS

- Applications
- Licensing options
- Configure now
- Other packages

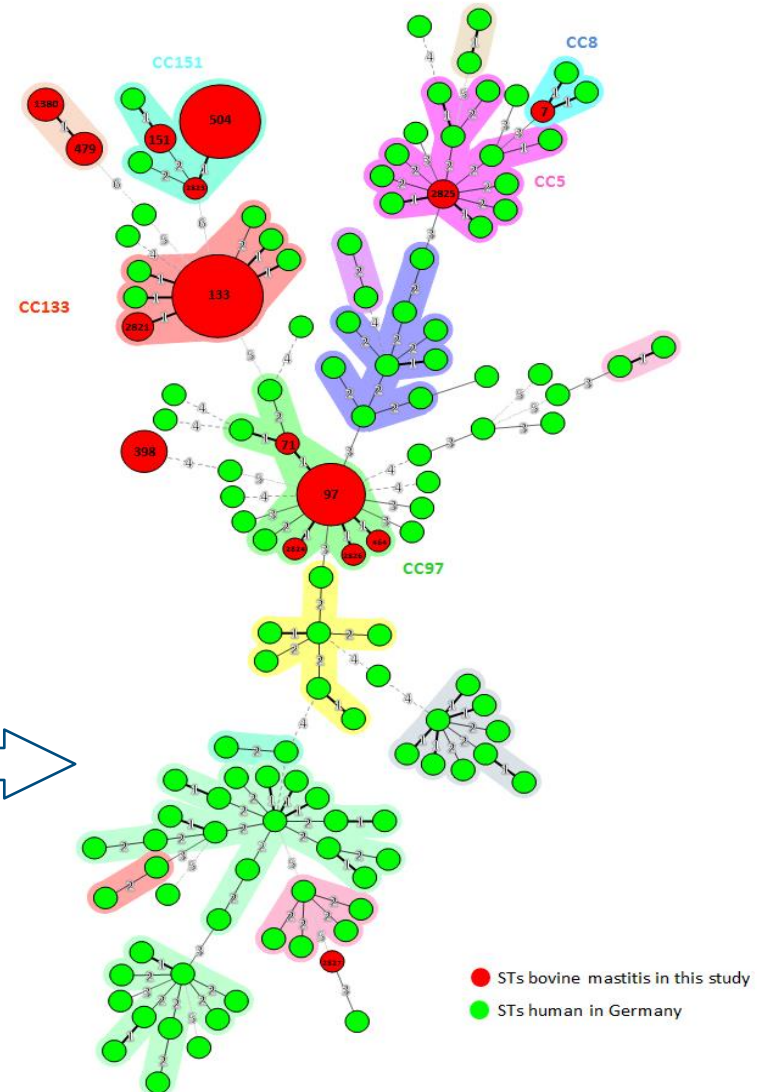
Lebensmittelqualität und -sicherheit

Multilocus sequence typing (MLST)

Staphylococcus aureus

MLST Typing of *Staphylococcus aureus* isolates associated with bovine milk

Minimum spanning tree of the *S. aureus* bovine mastitis isolates in this study (n = 70) combined with isolates from a human in Germany (n = 118). All the *S. aureus* isolates were analyzed by MLST. The tree was constructed by using BioNumerics version 7.5 allowed a presumptive coefficient. The clusters were created from two types within two changes of neighboring distance.



Sheet OH, Grabowski NT, Klein G, Reich F, Abdulmawjood A (2019).

The role of bioinformatics in food analysis

Further analysis tools for sequence data

Further analysis tools for sequence data

GGDC - Genome-to-Genome Distance Calculator (DSMZ)

DSMZ-German Collection of Microorganisms and Cell Cultures GmbH

Freely available at: <http://ggdc.dsmz.de/ggdc.php#>

GGDC Genome-to-Genome Distance Calculator 2.1

Submission Form

Want to analyze [viral genomes](#) or [single genes](#) instead? The new platform [TYGS](#) offers complete genome-based t

Local alignment tool

BLAST+ (recommended)

Optional bootstrap replicate-based confidence intervals (not recommended)

Query accession(s)

CP059964.1

Query genome

Durchsuchen... Keine Datei ausgewählt.

Reference accessions

CP017868
CP046243
CP063083
CP049264

Allowed no. accessions/FASTA files: 75

Reference genomes

Durchsuchen... Keine Dateien ausgewählt.

Contact details

antonia.kreitlow@tiho-hannover.de

Submit

Current slot usage

32%

Further analysis tools for sequence data

TYGS - Type (Strain) Genome Server (DSMZ)

Freely available at: https://tygs.dsmz.de/user_requests/new

Submit Your TYGS job

Specify up to 20 genome files and/or accession IDs at once

GenBank / FASTA file(s):

Durchsuchen... Keine Dateien ausgewählt.

GenBank accession(s):

- CP059964
- CP017868
- CP046243
- CP059682
- CP049264

☒ Restrict job to above genome(s)?

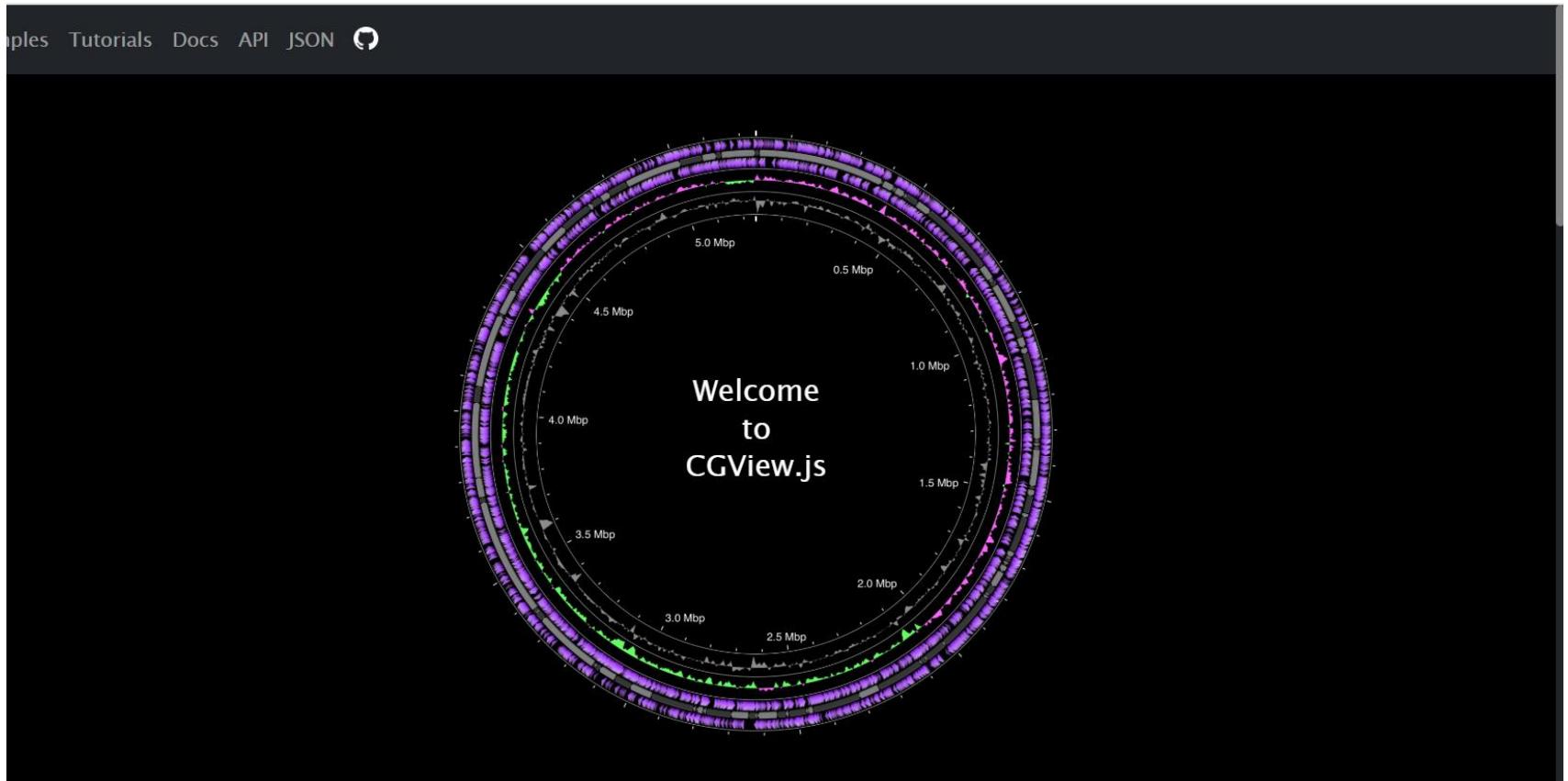
Provide contact details



© 2016-2021 Leibniz Institute DSMZ · [Privacy](#) · [Terms](#) · [Changelog](#) · v269

[Back to top](#)

[CGView.js](https://js.cgview.ca/)



<https://js.cgview.ca/index.html>

<https://js.cgview.ca/examples/example-plasmid.html>

Challenges and Future Perspectives

- ❖ **Data storage and management issues**
- ❖ **Need for advanced computational methods**
- ❖ **Ethical and regulatory concerns**
- ❖ **Future trends in microbial bioinformatics**

Thank you for your attention



