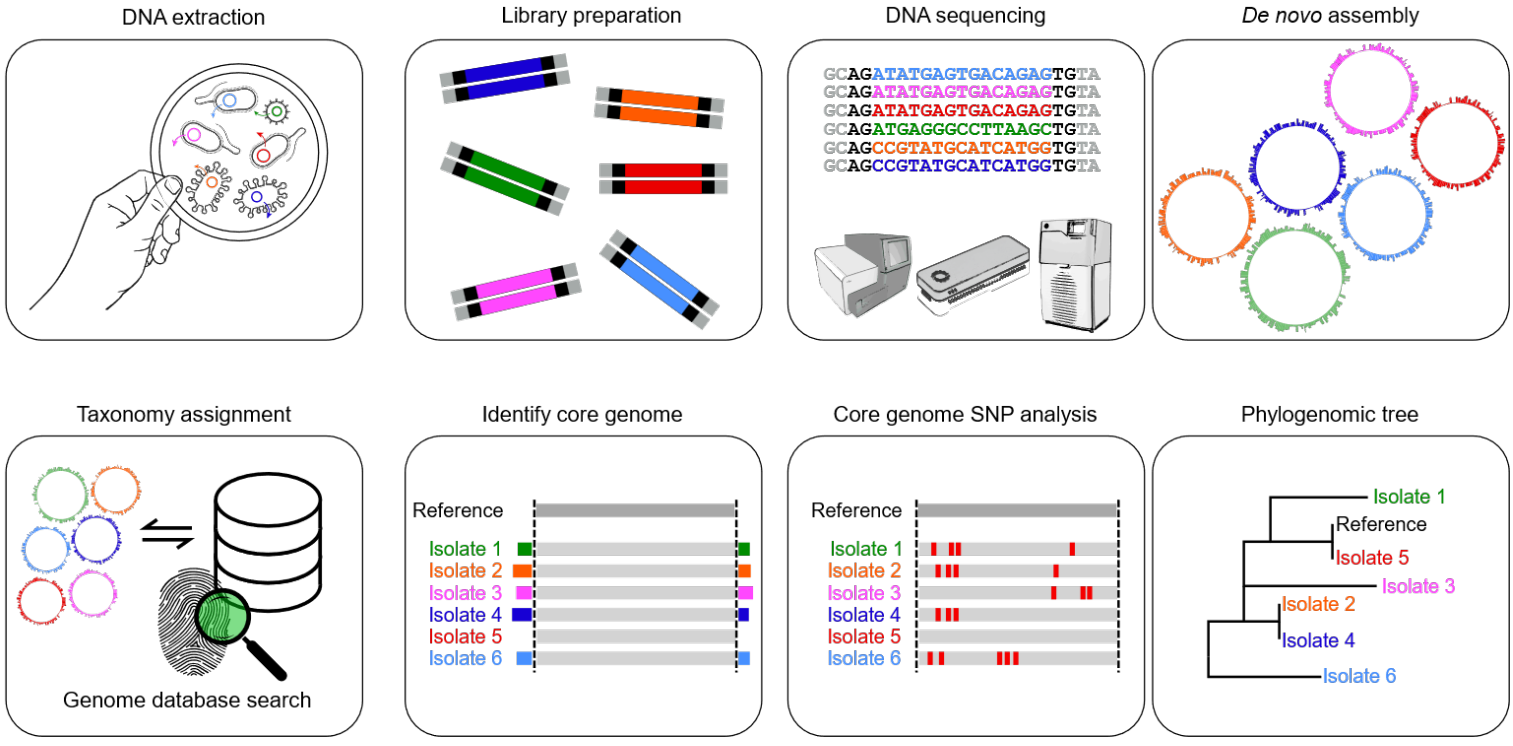
Microbial source tracking

Do you want **state-of-the-art** microbial source tracking?
Do you want to **improve your response time** and improve your HACCP system?

- Complete and unambiguous genomes provide a high-resolution foundation for multi-locus sequence typing (MLST, cgMLST), analysis of core genomic single-nucleotide polymorphisms (cgSNP), and other higher-order analyses.
- Option to store genomes in customer-specific databases at DNASense and compare previous and more recent outbreaks.
- Fast-track (≤ 7 days turn-around time) and ultra fast-track (≤ 3 days turn-around time) options are available.

DNASense provides complete **sample-to-answer** services for microbial source tracking based on the technology used for SARS-CoV-2 variant tracking

State-of-the-art workflow



Customized solutions

Our standard package includes: Optional pre- and post-project meetings with a DNASense specialist, DNA extraction, library preparation, sequencing, pre- and post-sequencing quality control, de novo assembly (or SNP calling), taxonomic profiling, cgSNP analysis, online access to raw data and result files and a detailed project report.

Add-on services (non-exhaustive list): Structural variant (SV) analysis, multi-locus sequencing typing (MLST), on-site sequencing, fast turn-around time.

info@dnasense.com | +45 71 99 20 20

Working with the DNASense team

- Extensive experience from hundreds of projects and challenging samples
- Detailed documentation and full method transparency
- State-of-the-art sample preparation, DNA sequencing and bioinformatics
- Extensive expert consultant services

Encompassing report with actionable results

1 Project scope

DNASense received 17 samples from Dr. Rosalind McGiver at Food and Beverages LLC. The aim of the project is to conduct a core genome single nucleotide polymorphism (cgSNP) analysis to measure the relatedness of *Escherichia coli* isolates (see figure 1). DNASense extracted high-quality DNA from 17 samples (agar plate scrapings), prepared DNA sequencing libraries and sequenced these using the Oxford Nanopore Technologies long-read DNA sequencing technology. The sequencing data was used for *de novo* genome assembly and whole genome-based classification. Isolates classified as *Escherichia coli* were subjected to a cgSNP analysis, along with isolates from previous contamination outbreaks (2018 and 2019).

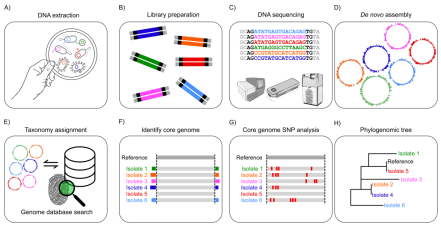


Figure 1: ***Core genome SNP analysis Workflow****. A core genome single nucleotide polymorphism (cgSNP) analysis workflow typically starts with the extraction of DNA from pure culture isolates but direct DNA extraction from source material is also possible* ***(A)****. A DNA sequencing library is prepared from native (Oxford Nanopore Technologies) or synthetic (Illumina) DNA and sequenced on a compatible platform* ***(B-C)****. The generated sequencing data is subjected to quality control filtering to ensure that only high quality data is processed. Quality-filtered data is used to generate highly contiguous de novo assemblies* ***(D)****, which are classified against a state-of-the-art genome taxonomy database* ***(E)****. Strains belonging to the same species are then subjected to a cgSNP analysis. A core genome (DNA segments shared by all isolates) is then identified* ***(F)****. SNPs are subsequently called across the core genome alignment* ***(G)*** *and concatenated to generate a SNP position-specific alignment and ultimately a phylogenomic tree* ***(H)****.*

2 Project summary

The core genome SNP analysis indicates the the *Reactor contaminant* (isolate AD178) can be traced back to Room AD178 (Toilet facility, level 1, room 78). It is further noted that it closely resembles the strains isolated from reactor AX1 (2018 outbreak) and AX4 (2019 outbreak).

genomes (i.e circular) or highly contiguous assemblies with chromosome-size contigs. 6 out of the 17 (35.3 %) samples were contaminated above a 5 % contamination threshold.

Table 2: ***Assembly statistics****. **Sampling ID** and **Sample name** denote customer-assigned sample identification nomenclature. **Sequencing ID** denotes the sequencing barcodes assigned by DNASense. **Assembly size (Mbp)** and **No. of contigs** denote the size of the final assembly in megabases and the number of contiguous DNA elements associated with it, respectively. **N50 (Mbp)** denotes that half of the data is contained within contigs of length N50 or greater. **Compl. (%)** is the estimated genome completeness based on the presence or absence of essential lineage-specific marker genes. **Contam. (%)** is the estimated genome contamination based on the presence of multiple single-copy marker genes. The **SH index** denotes the strain heterogeneity index (0-100). It reflects the relatedness of the assembled contigs as determined from the number of multi-copy marker pairs which exceed a specified amino acid identity threshold. A high percentage of contamination implies that multiple single-copy marker genes are present (i.e. from several organisms) and if this corresponds with a high strain heterogeneity, it indicates that the marker genes belong to closely related taxa (strains or closely related species).*

Sampling ID	Sample name	Sequencing ID	Assembly size (Mbp)	No. of contigs	N50 (Mb)	Compl. (%)	Contam. (%)	SH index
A1D178	Reactor contaminant	barcode01	4.91	14	3.39	99.69	0.32	14.29
A1D179	2018 outbreak strain	barcode02	4.52	3	4.26	99.99	0.44	12.50
A1D180	2019 outbreak strain	barcode03	4.51	5	4.33	99.69	0.14	50.00
A1D181	Isolate 1 (Toilet)	barcode04	18.13	116	4.55	100.00	199.31	32.05
A1D182	Isolate 2 (pump)	barcode05	4.68	5	4.35	99.99	0.44	12.50
A1D183	Isolate 3 (South loading dock)	barcode06	4.79	8	4.41	99.69	0.32	0.00
A1D184	Isolate 4 (Door knob)	barcode07	4.52	3	4.26	99.99	0.44	12.50
A1D185	Isolate 5 (Lab coat)	barcode08	6.07	69	4.12	97.93	33.15	0.00
A1D186	Isolate 6 (Membrane)	barcode09	4.68	7	4.26	99.99	0.44	12.50
A1D187	Isolate 7 (Floor)	barcode10	7.82	78	4.12	100.00	90.36	0.00
A1D188	Isolate 8 (Gloves)	barcode11	9.59	6	4.35	98.96	98.96	98.18
A1D189	Isolate 9 (East loading dock)	barcode12	4.62	4	4.35	99.99	0.44	12.50
A1D190	Isolate 10 (Pump station)	barcode13	4.66	5	4.42	99.69	0.32	0.00
A1D191	Isolate 11 (Lunch room)	barcode14	4.61	3	4.35	99.99	0.44	12.50
A1D192	Isolate 12 (Light switch)	barcode15	4.52	3	4.26	99.99	0.44	12.50
A1D193	Isolate 13 (Soap dispenser)	barcode16	10.93	83	0.32	98.79	84.18	4.55
A1D194	Isolate 14 (Towel)	barcode17	7.53	24	4.12	100.00	99.54	0.00

Highly contaminated samples cannot be subjected to a cgSNP analysis and efforts should be made to minimize (during culturing) or remove contaminants. It should be noted that genome completion and contamination levels are estimated to provide a quantitative measure of the quality of the sample. A specific prokaryotic lineage can be associated with a specific set or number of single-copy marker genes (i.e. lineage-specific marker genes). If the entire set can be located, the genome is said to be 100 % complete (e.g. 110 located marker genes out of 110 expected marker genes for that lineage). If no additional marker-genes are located (i.e. more than the 110 present in the lineage), the sample is said to have 0 % contamination. If multiple single-copy marker genes are found, the sample is said to be contaminated in a proportion proportional to the number of additional located marker genes. For substantially complete genomes (≥70%–90%) with medium contamination (5% to ≤10%), completeness and contamination estimates generally have an absolute error of ≤6%, and the error in the quality estimates tends to decrease as the quality of a genome improves (Parks et al., 2015). The contamination estimates should therefore be interpreted with some caution. Extreme values or values above the 6 % threshold may indicate samples requiring further inspection. The concept of strain heterogeneity is used to indicate the relatedness of an isolate contaminant, i.e. based on the identity of the duplicate marker genes. High strain heterogeneity suggests that the majority of the reported

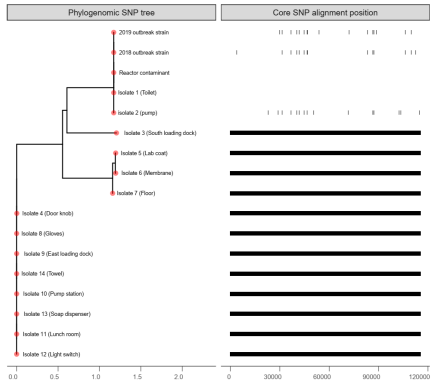


Figure 6: ***SNP signature****. Left panel shows a reference-based phylogenomic tree based on the single-nucleotide polymorphisms (SNPs) in the core genome SNP alignment. Branch lengths are proportional to the amount of inferred evolutionary change (the axis units are arbitrary). Right panel shows the positions of the individual SNPs relative to the reference genome. Samples are aligned to their corresponding tree tip label (left panel). The x-axis represents the SNP position in the the core genome SNP alignment.*

Price example*

Service	Analysis	Sample fee (pr. isolate)	Fast-track fee	Turn-around-time**	24 isolate price example
Normal	2000 EUR	200 EUR	0 EUR	≤ 15 days	6800 EUR
Fast-track	2000 EUR	200 EUR	1200 EUR	≤ 5 days	8000 EUR
Ultra fast-track	2000 EUR	200 EUR	2000 EUR	≤ 3 days	8800 EUR

*Prices assume that isolates are pure culture isolates (~ 500 Mbp/sample). ** Working days

Contact us today at
info@DNASense.com +45 7199 2020

Founded in 2014, DNASense's mission is to make cutting-edge DNA sequencing and bioinformatics readily available to life science research and industry