Microsynth

Resequencing of Prokaryotic Organisms

Search for virulence and resistance genes and check for mutations
Validate your bacterial production stem
Characterize microbial isolates by MLST

Introduction

Due to the emergence of powerful next generation sequencing (NGS) technologies, it has never been easier to resequence whole genomes. In case a fully annotated reference genome exists, resequencing may be used to compare differences between genomes of individuals from the same species (please see our *de novo* service if no reference exists). Using results from such measurements, researchers may search for virulance, resistance and toxin genes, study inherited or acquired mutations and characterize microbial isolates by multi locus sequence typing. Besides offering our customers state-of-the-art NGS platforms, Microsynth has also made significant investments in the bioinformatics analysis area. This application note will give you an overview of Microsynth's various services in the field of microbial resequencing as well as its possible impact on your research.

Microsynth's Competences and Services

With more than 10 years of experience in the field of next generation sequencing, one of Microsynth's core competences is to provide high quality one-stop services from experimental design to bioinformatics analysis. You may either outsource the entire analysis or only single steps to us as illustrated in ***Figure 1***.

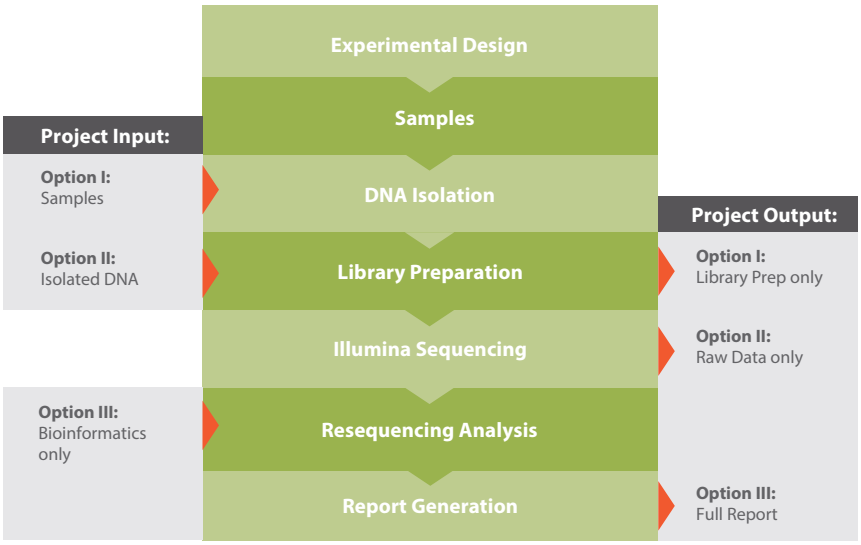

Figure 1. *Microsynth's workflow for resequencing projects. The workflow can be entered and exited at various steps depending on the requirements of the customer.*

Experimental Design

Microsynth's NGS specialists will help you define suitable experimental setups for your resequencing projects and discuss possible strategies to address your research questions.

DNA Isolation

You may either perform the DNA extraction yourself or outsource this critical step to Microsynth. We have long-standing experience in processing various sample matrices and DNA/RNA sources.

Library Preparation and Sequencing

Following a quality check of your samples, Microsynth will construct Illumina libraries including specific adaptors with barcodes. Depending on the experimental design, the libraries are pooled and sequenced on the Illumina NextSeq 500/550 platform. By offering datapackages for this service, flexible sequencing of any sample number is possible.

Microsynth AG, Switzerland
Schützenstrasse 15 · P.O. Box · CH-9436 Balgach · Phone +41 71 722 83 33 · Fax +41 71 722 87 58 · info@microsynth.ch · www.microsynth.com

Bioinformatics Analysis

Sequencing reads are quality filtered and mapped against the reference genome of the organism to be studied. After further refinement based on best practices, possible single nucleotide variations (SNVs) and small insertions and deletions (InDels) are detected and annotated (see ***Table 1***). One or multiple variant callers may be used at once, thus searching in a sensitive or specific way depending on the goals of the study. Whole genome sequencing (WGS) data may also be used for Multi Locus Sequence Type (MLST) microbial isolates using predefined loci and reference databases (see ***Figure 2***). If there is no prior knowledge, sequenced reads may be *de novo* assembled and predicted genes are screened against community accepted database containing resistance, virulence and toxin genes (see ***Table 2***).

Beside the sequencing raw data and the output of the bioinformatics analysis, a user-friendly summary report is provided.

Provided Output Files:

Raw data: Fastq

Mapping: BAM/BAI files

Variant calling, protein consequences:

VCF (for each sample separately) and HTML (includes all samples).

Example Results

Sample_ID	Ref_ID	Gene	Tag	Start	End	Strand	Protein_Id	Product	Var_Position	Reference	Alternative	DP	AF	SB	Mutation_Type	Ref_Protein/Alt_Protein
Sample_1	NC_002695	stx2A	CDS	1267107	1268066	sense	NP_309232.1	Shiga toxin 2 subunit A	1268061	A	G	290	1.000000	0	Missense (319)	MKC...TTGK* MKC...TTGE*
Sample_1	NC_002695	mdtH	CDS	1483403	1484611	anti-sense	NP_309470.2	multidrug resistance protein MdtH	1483524	A	G	278	0.996403	0	Missense (363)	MSR...GKSVHQP...RDA* MSR...GKSAHQP...RDA*
Sample_1	NC_002695	insI1	CDS	2612114	2613145	anti-sense	NP_305380.1	IS30 transposase	2612212	C	CTAG	291	0.941581	0	Insertion (312)	MLR...QHEDLV...LTD* MLR...QHELDLV...LTD*

Table 1. *SNV und InDels - Typical output overview file (HTML) resulting from the variant calling step for SNVs and small InDels. For each sample and chromosome/contig, the SNV and small InDels are reported separately and the effect of SNVs and small InDels for all annotated features of the reference genome are shown. Besides the html format the data are also given in tab separated format (to import into Excel) and as vcf.*

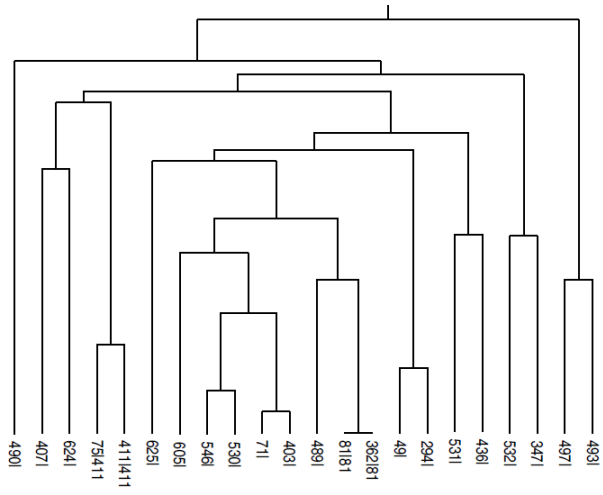

Figure 2. *MLST subtree: As an illustrative example, a small subtree of all phylogenetic profiles stored on the PubMLST database [3] of the Cronobacter sequence type is depicted. A customer sample would be placed into its appropriate phylogenetic context to determine its exact sequence type or its next relative.*

Sample_ID	Resistance genes	Virulence genes	Toxin genes
Sample_1	fosB1	-	-
Sample_2	-	astA1	-
Sample_3	-	-	TRI1

Table 2. *Exemplary cut-out of a table with screening results for resistance, virulence and toxin genes.*