Microsynth

Resequencing of Prokaryotic and Eukaryotic Organisms

Detect inherited and acquired mutations
Characterize microbial isolates by MLST

Introduction

Due to the emergence of powerful next generation sequencing (NGS) technologies, it has never been easier to *de novo* sequence or 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 study inherited or acquired mutations or may characterize microbial isolates. 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 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.

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.

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.*

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 either on the Illumina MiSeq or NextSeq 500 platform. These flexible sequencing platforms allow for optimal sequencing depending on the number of samples and on the required read length.
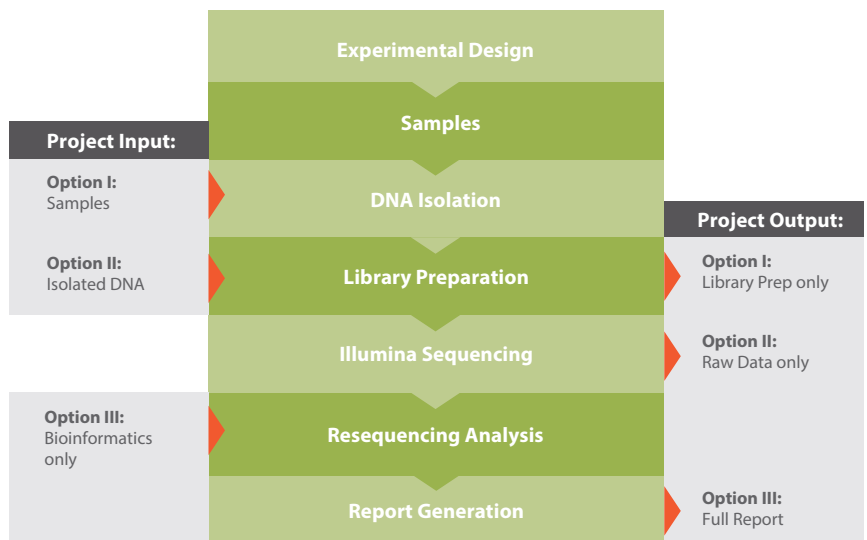
Bioinformatics Analysis

Sequencing reads are quality filtered and mapped against the reference genome of the organism to be studied. After further refinement based on GATK Best Practices [1], possible single nucleotide variations (SNVs) and small insertions and deletions (InDels) are detected and annotated (see *Figure 2*). 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. For eukaryotic organisms, large structural variations are also called by default. If the study provides tumor and control sample pairs, somatic variations may be identified. For human samples, detected variations are compared against public databases like the ClinVar - NCBI - NIH [2] to separate already known variations from novel ones. 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 3*). Beside the sequencing raw data and the output of the bioinformatics analysis, a user-friendly summary report is provided.

Example Results

Primary_ID	Gene_Id	Primary_Tag	Start	End	Strand	Transcript_Id	Gene_Name	Var_Position	Reference	Alternative	DP	AF	MQ	Mutation_Type	Ref_Protein/Alt_Protein
chr20	DEFB125	CDS	96005	96417	sense	NM_153325.2	DEFB125	96321	T	C	146	1.00	60.00	Silent	MNILMLTFIICGLLTRVTKGSFEPQKCWKNNVGH MNILMLTFIICGLLTRVTKGSFEPQKCWKNNVGH
chr20	DEFB128	CDS	187886	188118	anti-sense	NM_001037732.1	DEFB128	188087	T	A	202	1.00	60.00	Missense (27;27)	MKLFLVLIILLFEVLTDGARLKKCFN**K**VTGYCRK MKLFLVLIILLFEVLTDGARLKKCFN**N**VTGYCRK
chr20	DEFB132	CDS	257779	257836	sense	NM_207469.2	DEFB132	257794	CTGGTCT	C	44	1.00	60.00	Deletion (7;7)	MKFLLL**VL**AALGFLTQVIPASAGGSKCVSNTPGY MKFLLLAALGFLTQVIPASAGGSKCVSNTPGYCR

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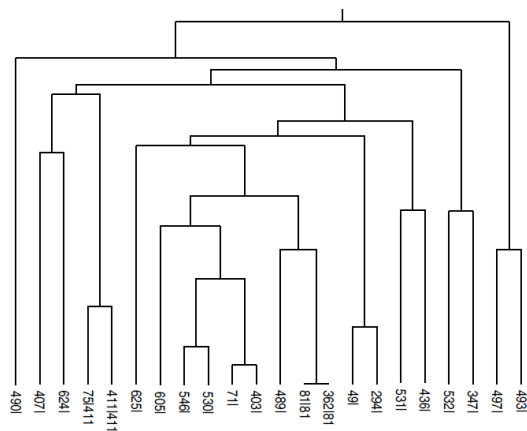
Figure 3: MLST subtree: As an illustrative example, a small subtree of all phylogenetic profiles stored on the PubMLST database [3] of the Cronobacter sequence type is depicted. A customer sample would be placed into its appropriate phylogenetic context to determine its exact sequence type or its next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

Related Topics

Microsynth also provides whole exome sequencing (WES) services which focus on sequencing selected genomic features. The data analysis however, is similar to the one described in this application note.

References

[1] GATK Best Practices online at: https://software.broadinstitute.org/gatk/best-practices/ (accessed 01.05.2018)
[2] Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res . 2016 Jan 4;44(D1):D862-8. doi: 10.1093/nar/gkv1222. PubMed PMID:26582918
[3] The primary PubMLST site is hosted at The Department of Zoology, University of Oxford, UK and is funded by The Wellcome Trust. online at: http://pubmlst.org/