16S Pipeline: Software Requirements Specification

Author: Sarah Burns
Date: 14/10/2019

1. Sign offs

I agree that this document represents our best understanding of the requirements for this project today and the system described will satisfy our needs.

Meera Chand		
Name	Signature	Date
Penny Cliff		
Name	Signature	Date
Rossella Baldan		
Name	Signature	Date

2. Introduction

2.1. Purpose

To develop an analytical pipeline to detect bacterial species from 16S nanopore sequencing data.

2.2. Scope

The pipeline will be developed within the Centre for Clinical Infection and Diagnostics Research (CIDR) with the eventual aim of inclusion within a UKAS-accredited clinical service. The analysis will be automated and performed in real-time, and be managed by a bioinformatician in addition to selected personnel.

2.3. Definitions, acronyms and abbreviations

ATCC	American Type Culture Collection (reference standards)
DNAnexus	Cloud-based data analysis and management platform
WDL	Workflow development language; a framework for writing pipelines.
Nextflow	A framework for writing pipelines.

3. Overall description

3.1. Product perspective

This will form part of a new biological diagnostic assay with improved speed and accuracy compared to current gold-standard methods. The pipeline will evolve over the course of a pilot study prior to implementation as a routine clinical service.

3.2. Product features

The raw output of the nanopore sequencing is FAST5 files which are produced and base-called in real-time, producing FASTQ files. The pipeline will monitor the generation of the FASTQ files and, at select time intervals, merge these into a single FASTQ file, classify the reads against a reference database and produce a clinically useful report alongside quality control information (e.g. read length, yield, base-call quality). See Figure 1 in Supporting information.

3.3. Operating environment

- Pipeline framework: DNAnexus-compatible (e.g. WDL or Nextflow)
- Operating system: Linux

3.4. Assumptions and dependencies

- A maximum of five samples will be sequenced in parallel.
- All latest versions of the necessary software will be installed in the run environment.
- Input data will be in the correct format.
- There will be a continuous internet connection.

4. System features

4.1. Functional requirements

Requirement	Priority
Database to include sequences for all relevant clinical species (see Table 1 in Supporting information).	*Essential*
Classifier can detect to species level.	*Essential*
Pipeline output to report hits in >1% classified reads (subject to testing).	*Essential*
Quality report to be produced for every sample.	*Essential*
Pipeline executable both within DNAnexus and on local workstation.	*Desirable*
Data to be stored within DNAnexus.	*Desirable*

4.2. Non-functional requirements

Requirement		Priority
Capacity	Able to process five samples in parallel.	*Essential*
Performance	Complete within 30 minutes (larger files >1GB permitted to exceed this providing the duration remains within an acceptable timeframe for rapid clinical service).	*Desirable*
Accessibility	Continuous service.	*Desirable*

Maintainability and extensibility	Code will be modular, easy to understand and written in line with conventions, with detailed commenting and effective error handling.	*Desirable*

4.3. Software releases

Initial release (v1.0.0) containing all essential requirements as a minimum will be available prior to the start of the pilot study.

4.4. Validation

- 167 ATCC isolates to be used as “truth” data to validate classifier.
- Clinical samples containing known species to be used to determine sensitivity and specificity of test.

4.5. Documentation

- Standard operating procedure
- Testing specification
- Validation specification

5. Supporting information

Figure 1 – General pipeline workflow

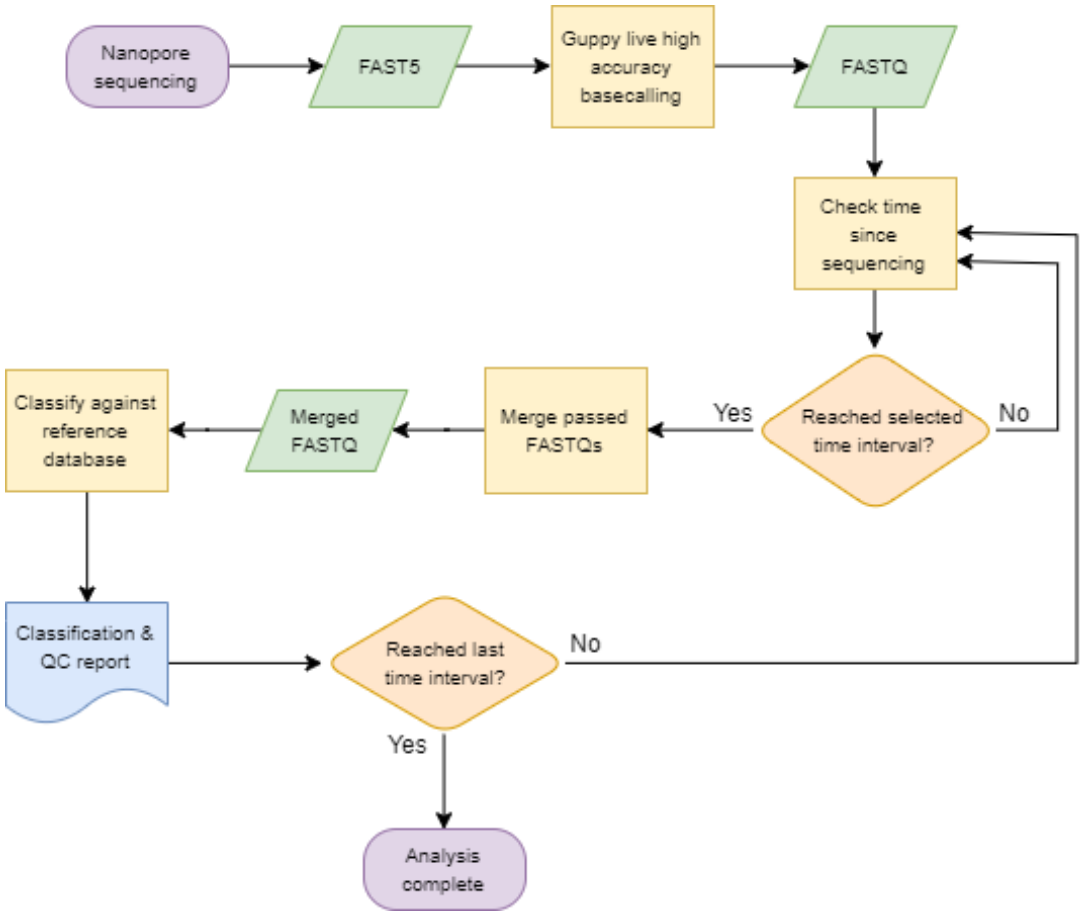

Table 1 – Clinical species desirable for pipeline to detect

Abiotrophia defectiva	Coxiella burnetii	Peptoniphilus harei
Achromobacter xylosoxidans	Cutibacterium acnes	Peptostreptococcus anaerobius
Acinetobacter baumannii	Cutibacterium avidum	Plesiomonas shigelloides
Acinetobacter baylyi	Delftia acidovorans	Pluralibacter gergoviae
Acinetobacter haemolyticus	Dermabacter hominis	Prevotella denticola
Acinetobacter johnsonii	Eikenella corrodens	Proteus mirabilis
Acinetobacter junii	Elizabethkingia miricola	Proteus vulgaris
Acinetobacter lwoffii	Enterobacter asburiae	Providencia rettgeri
Acinetobacter nosocomialis	Enterobacter cloacae	Providencia stuartii
Acinetobacter pittii	Enterobacter kobei	Pseudomonas aeruginosa
Acinetobacter radioresistens	Enterobacter ludwigii	Pseudomonas koreensis
Acinetobacter tjernbergiae	Enterococcus avium	Pseudomonas luteola
Acinetobacter ursingii	Enterococcus casseliflavus	Pseudomonas mosselii
Actinomyces naeslundii	Enterococcus durans	Pseudomonas oryzihabitans
Actinomyces neuii	Enterococcus faecalis	Pseudomonas putida
Actinomyces odontolyticus	Enterococcus faecium	Pseudomonas stutzeri
Actinomyces radingae	Enterococcus raffinosus	Pseudomonas taiwanensis
Actinomyces turicensis	Erysipelotrix rhusiopathiae	Raoultella ornithinolytica
Aerococcus urinae	Escherichia coli	Raoultella planticola
Aeromonas caviae	Finegoldia magna	Rhodococcus hoagii
Aeromonas media	Fusobacterium necrophorum	Salmonella enterica
Aeromonas veronii	Fusobacterium nucleatum	Serratia liquefaciens
Aggregatibacter actinomycecomitans	Gemella haemolysans	Serratia marcescens
Aggregatibacter aphrophilus	Gemella morbillorum	Serratia ureilytica
Aggregatibacter segnis	Gemella sanguinis	Staphylococcus aureus
Agrobacterium tumefaciens	Granulicatella adiacens	Staphylococcus capitis
Alcaligenes faecalis	Haemophilus haemolyticus	Staphylococcus caprae
Anaerococcus lactolyticus	Haemophilus influenzae	Staphylococcus epidermidis
Anaerococcus octavius	Haemophilus parainfluenzae	Staphylococcus haemolyticus
Arcanobacterium haemolyticum	Hafnia alvei	Staphylococcus hominis
Bacillus cereus	Helcococcus kunzil	Staphylococcus lugdunensis
Bacillus licheniformis	Kingella kingae	Staphylococcus pasteuri
Bacillus subtilis	Klebsiella aerogenes	Staphylococcus pettenkoferi
Bacteroides caccae	Klebsiella oxytoca	Staphylococcus pseudointermedius
Bacteroides fragilis	Klebsiella pneumoniae	Staphylococcus saprophyticus
Bacteroides thetaiotaomicron	Klebsiella variicola	Staphylococcus schleiferi
Bartonella henselae	Lactobacillus gasseri	Staphylococcus simulans
Bartonella quintana	Lactobacillus rhamnosus	Staphylococcus warneri
Bordetella pertussis	Lautropia mirabilis	Stenotrophomonas maltophilia
Brevibacterium casei	Legionella longbeachae	Streptococcus agalactiae
Brucella abortus	Legionella pneumophila	Streptococcus anginosus
Burkholderia cenocepacia	Listeria monocytogenes	Streptococcus constellatus
Burkholderia cepacia	Massilia timonae	Streptococcus cristatus
Burkholderia gladioli	Micrococcus luteus	Streptococcus dysgalactiae
Cardiobacterium hominis	Moraxella catarrhalis	Streptococcus gallolyticus
Chlamydophila pneumoniae	Moraxella nonliquefaciens	Streptococcus gordonii
Citrobacter amalonaticus	Moraxella osloensis	Streptococcus intermedius
Citrobacter braakii	Morganella morganii	Streptococcus lutetiensis
Citrobacter freundii	Mycobacterium abscessus	Streptococcus mitis
Citrobacter koseri	Mycobacterium avium	Streptococcus mutans
Clostridium paraputrificum	Mycobacterium chimaerae	Streptococcus oralis
Clostridium perfringens	Mycobacterium fortuitum	Streptococcus parasanguinis
Comamonas testosteroni	Mycobacterium intracellulare	Streptococcus pasterianus
Comoamonas kerstersii	Mycoplasma lipophilum	Streptococcus pneumoniae
Corynebacterium amycolatum	Mycoplasma pneumoniae	Streptococcus pyogenes
Corynebacterium aurimucosum	Neisseria gonorrhoea	Streptococcus salivarius
Corynebacterium confusum	Neisseria meningitidis	Streptococcus sanguinis
Corynebacterium diphtheriae	Neisseria weiveri	Sutterella wadsworthensis
Corynebacterium haemolyticum	Ochrobactrum anthropi	Tropheryma whipplei
Corynebacterium jeikeium	Oligella urethralis	Trueperella bernardiae
Corynebacterium minutissimum	Paenibacillus illinoisensis	Veillonella atypica
Corynebacterium propinquum	Paenibacillus pabuli	Veillonella parvula
Corynebacterium striatum	Pantoea agglomerans	Vibrio fluvialis
Corynebacterium tuberculostearicum	Parvimonas micra	Vibrio parahaemolyticus
Corynebacterium ulcerans	Pasteurella multocida	