# Abstract

This SOP describes the workflow for annotating prokaryote genome sequences. It covers topics such as relevant tools and downstream quality control and manual curation options.

# Introduction

Annotation of genomes includes the identification of gene structures as well as attaching meaningful biological information to the predicted protein products of these genes. This SOP uses the Prokka package for this purpose as an all-in-one solution. Prokka automatically identifies both coding and many classes of non-coding genes and infers function from highly similar, previously characterized proteins. Prokka’s output format is furthermore compatible with all relevant tools and downstream repositories.

# Requirements

## Genome assembly

The input to Prokka is a genome assembly in FASTA format.

## Reference proteomes

Prokka includes reference data sets for commonly studied bacteria. If your species does not fall within those groups, you should consider supplying an additional protein FASTA file that adheres to the Uniprot naming schema to improve the annotation.

## Software tools and versions

Prokka (1.11)

Barrnap (0.6)

Aragorn (1.2.36)

Prodigal (2.6.2)

SignalP (4.1)

Blast+ (2.2.30)

Hmmer3 (3.0)

Infernal (1.0.2)

# Procedure

## Running Prokka

Invoking Prokka (assuming all dependencies are loaded)

## Quality control

## Curation

# Implementation

# Related Documents

# Revision history