

**PROGETTI DI “RICERCA CORRENTE 2020”**  
**RELAZIONE FINALE**

**N. identificativo progetto: IZS LT 02/20 RC**

**Progetto presentato da:**

**ISTITUTO ZOOPROFILATTICO SPERIMENTALE**

**LAZIO E TOSCANA “M. ALEANDRI”**

**Area tematica:** Sanità animale

**Titolo del progetto:** Valutazione delle performance della metodica Whole Genome Sequencing (WGS) in Escherichia coli oggetto di Piano di Monitoraggio Nazionale AMR, per la determinazione dell'antibioticoresistenza “in silico” e predizione di fenotipi di resistenza tramite elaborazione di modelli di “machine learning”

Ricerca finanziata dal Ministero della Salute

**Responsabile Scientifico:** Virginia Carfora

## SINTESI

**Titolo: Valutazione delle performance della metodica Whole Genome Sequencing (WGS) in *Escherichia coli* oggetto di Piano di Monitoraggio Nazionale AMR, per la determinazione dell'antibioticoresistenza "in silico" e predizione di fenotipi di resistenza tramite elaborazione di modelli di "machine learning"**

Parole chiave: WGS, AMR, analisi bioinformatiche, machine learning

Il Progetto ha analizzato le caratteristiche di performance di HTS-WGS ed appropriata analisi bioinformatica nel determinare i pattern di sensibilità/resistenza agli antibiotici "*in silico*", in relazione ai risultati fenotipici "tradizionali" (es. microdiluizione in brodo e determinazione della MIC). Per far sì che il dato di WGS prodotto possa essere ritenuto affidabile e idoneo per le successive analisi bioinformatiche per la determinazione delle basi genetiche AMR, è stato necessario innanzitutto stabilire e ottimizzare protocolli di wet-lab e dry-lab. E' stata poi standardizzata una pipeline bioinformatica (dry-lab) *ad hoc* per l'analisi dei dati (dal QC alla caratterizzazione *in silico* degli isolati), focalizzata in particolare sull'analisi delle basi genetiche AMR. Sono state poi valutate le performance dei metodi stabiliti e implementati a seguito dei risultati ottenuti ed esplorate le potenzialità e i limiti di modelli di machine learning per la predizione dei fenotipi AMR sulla base di soli dati genomici.

## SUMMARY

**Title: Evaluation of the performance of Whole Genome Sequencing (WGS) for *Escherichia coli* in the frame of the National AMR Monitoring Program, for "in silico" analysis of AMR and prediction of resistance phenotypes through the development of "machine learning" models**

Key words: WGS, AMR, bioinformatics analyses, machine learning

The Project analysed the performance characteristics of HTS-WGS and appropriate bioinformatics analysis in determining "in silico" antimicrobial susceptibility/resistance patterns, in relation to "traditional" phenotypic methods (e.g. broth microdilution and MIC determination). With the aim to determine whether WGS data can be considered reliable and suitable for subsequent bioinformatics analyses for the detection of the genetic basis of AMR, wet-lab and dry-lab protocols were first established and optimized. An *ad hoc* bioinformatics pipeline (dry-lab) was then standardized for data analysis (from QC to *in silico* characterization of the isolates), focused mainly on the analysis of the genetic basis of AMR. Following the results obtained the performance of the wet-lab, dry-lab methods were then evaluated, and the potential and limitations of machine learning models for AMR prediction based on genomic data were explored.