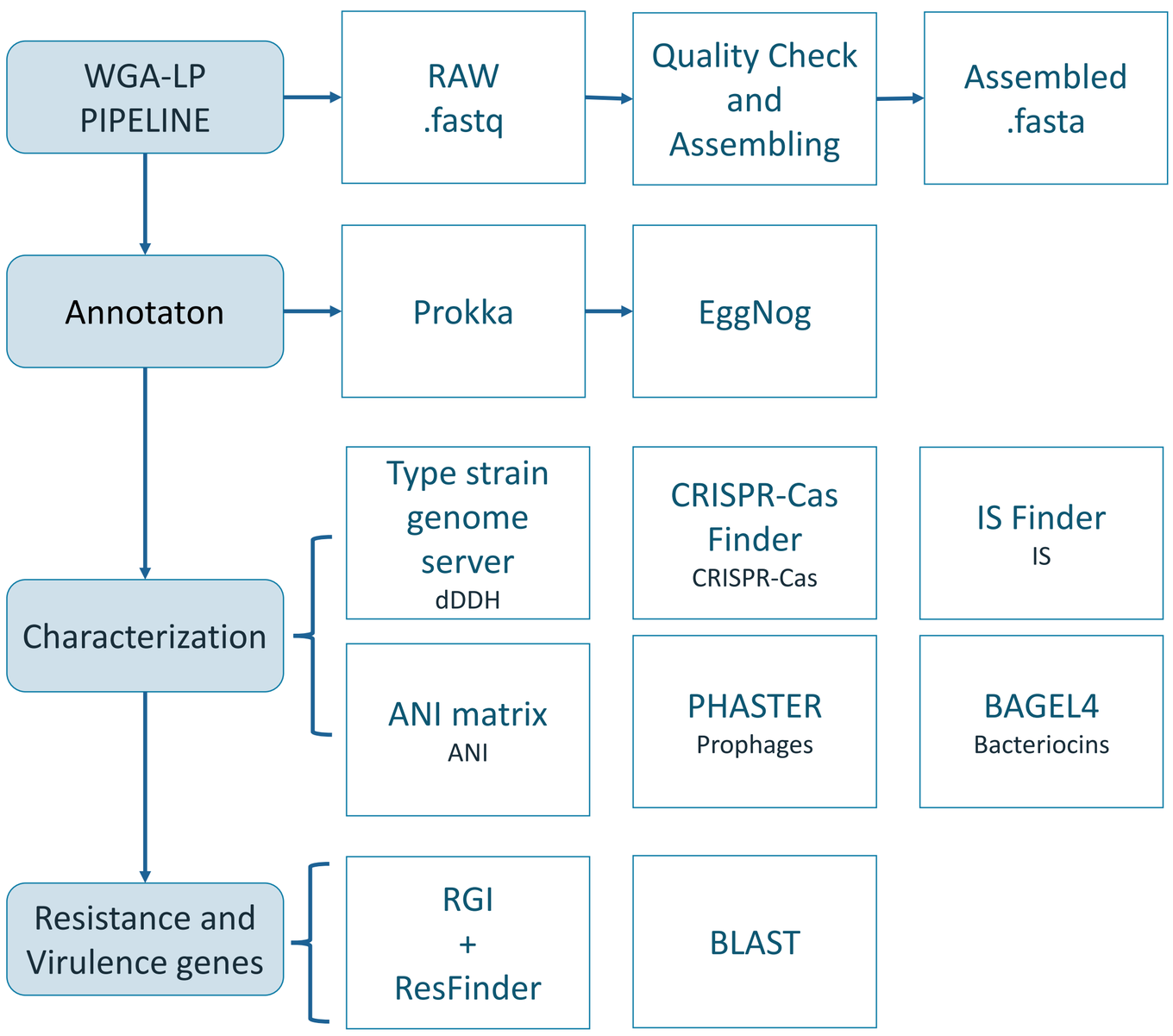
Probiotics for genetic and rare diseases

Andrea Colautti (colautti.andrea@spes.uniud.it)
Dipartimento di Scienze agroalimentari, ambientali e animali – DI4A, University of Udine, Italy
Tutor: Prof. Lucilla Iacumin

Introduction

This work consist of a preliminary screening of bacterial stains to research possible risk factors in their use in the food sector and use as probiotics. As representatives of a collection consisting of 200 bacteria belonging to the *Lacticaseibacillus* genus, based on the source of isolation and electrophoretic profiles obtained by RAPD, REP, and SAU, the genome of 14 strains (4 identified as *L. casei*, 6 identified as *L. paracasei*, and 4 identified as *L. rhamnosus*) were sequenced and assembled with a newly built pipeline. Following a literature review, genes related to antibiotic resistance and associated with characteristics that confer virulence within *Lactobacillus* spp. were identified. The reported genes were then searched within the 14 assembled genomes using BLAST. As further verification, the genomes were analyzed using databases of resistance and virulence genes

Materials and methods



To assembly the paired-end reads obtained with MiSeq platform (Illumina), a new assembly pipeline that pays particular attention to the quality of the assemblies, WGA-LP [1], was used. The quality assessment of raw reads was based on the usage of FastQC [2], Kraken [3], and Bracken [4], the evaluation of coverage of the assembled contigs with a new tool, with a final verification of the assemblies thus obtained with CheckM [5], Merqury [6], and QUAST [7]. After trimming Illumina adapters with Trimmomatic [8], SPAdes [9] coupled with the plasmid search function was used for the assembly of the genomes. The genomes thus assembled were then aligned and ordered to a reference sequence specific for the strain (*L. casei* NZ_CP006690.1, *L. paracasei* NC_008526.1, *L. rhamnosus* NC_017482.1) using Mauve [10]. The reference sequences of the various virulence and antibiotic resistance genes were obtained from the NCBI database, and using BLAST's tblastn function, were aligned to the 14 genomes to verify their eventual presence. For further verification, the amino acid sequences that showed correspondence with reference genes were translated into amino acid sequences and aligned with Protein BLAST to the reference sequence. As further verification, the genomes were aligned with RGI [11] and ResFinder 4.1 [12] databases.

Results

After the alignments carried out through BLAST, the presence of any gene linked to these determinants was not detected. Not even the verification through specific databases showed any match with any determinant linked to resistance or virulence factors. Therefore, the 14 lactobacilli examined in this experimentation can be considered valid candidates for the experimentation as they do not involve the presence of factors that could lead to potential damage to the host.

Antibiotic Resistance	**Protein synthesis inhibitor**	**Cell wall synthesis inhibitors**	**DNA inhibitor**	**Antimetabolite**	**RNA inhibitor**
	Tetracycline MLS Chloramphenicol	Aminoglycosides Vancomycin β-lactam Bacitracin	Ciprofloxacin	Sulfamethoxazole	Rifampicin
	tet, otr, tcr, erm, mef, msr, lnu, vat, cat	*aac, ant aph, aad, van, bla, bcr*	*gyr, par*		*rpo*
Virulence	**Adhesion and Aggregation Factors**	**Sex pheromones**	**Cytolysin toxin**	**Gelatinase**	**Hyaluronidase**
	ace, asa1, esp, efaA	ccf, cob, cpd	cyl	gelE, fsr	hyl

References: [1] Rossi N, Colautti A, Iacumin L, Piazza C (2021) WGA-LP: a pipeline for Whole Genome Assembly of contaminated reads; [2] Andrews, S (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data; [3] Wood DE, Salzberg SL (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments; [4] Lu J, Bretwieser FP, Thielen P, Salzberg, SL (2017) Bracken: estimating species abundance in metagenomics data; [5] Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes; [6] Rhie A, Walenz BP, Koren S et al. (2020) Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies; [7] Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies; [8] Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data; [9] Bankevich A, Nurk S, Antipov D, et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing; [10] Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. [11] Alcock B (2020) CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database; [12] Bortolaia V (2020) ResFinder 4.0 for predictions of phenotypes from genotypes.