RESISTOME ANALYSIS OF STAPHYLOCOCCUS AUREUS IN PUBLICLY AVAILABLE GENOMES

Main author: Vincenzo Pennone (University of Leòn - ULE)

Co-authors: José F. Cobo-Dìaz, Miguel Prieto Maradona, Avelino Alvarez-Ordoñez

INTRODUCTION

Antimicrobial resistance (AMR) is a key issue worldwide, affecting both human and animal health and welfare. The occurrence and persistence of pathogens in foods and environments, together with the widespread use of antibiotics, can lead to the development of resistance in these microbes of concern. Many genetic mechanisms that can lead to AMR have been identified, including point mutations, insertions/deletions and the exchange of mobile genetic elements (MGEs) via horizontal gene transfer (HGT). The analysis of the data of thousands of bacterial whole genomes that are available in public databases can provide invaluable information on the development and spread of AMR, allowing us to understand when, where and how AMR appears and develops. Staphylococcus aureus is a pathogen of major concern for its AMR, with some resistant strains that are difficult to eliminate and which can lead to severe disease or death. The aim of this work was to study in-silico the resistome of publicly available S. aureus genomes using bioinformatics tools, in order to elucidate the occurrence of the main AMR genes (ARGs) in different hosts and Clonal Complexes, plus their temporal and geographical distribution.

METHODOLOGY

The assembled genomes and metadata were downloaded from the NCBI, PATRIC and PubMLST databases. The corresponding metadata were manually curated where necessary. The assembled genomes from each of the three databases were analysed in silico with the staramr pipeline, which was automated by using a modified version of the Ruby script 08.staramr_auto.rb, as described in a previous study. The Clonal Complexes (CCs) were assigned using the list of CCs associated with STs in the S. aureus PubMLST scheme. PlasFlow was used for the identification of plasmidic contigs. Two datasets including both resistome and metadata were generated: one for chromosomic data and one for plasmidic data. The chromosomic and plasmidic datasets were both explored and analysed in RStudio. ARG occurrence and distribution patternswere compared among CCs, source categories, isolation years and geographical distributions. Heatmaps and boxplots were generated using the pheatmap and ggplot2 packages, respectively. Statistical analysis was performed in RStudio, through the Shapiro-Wilk normality test, and followed by Student’s T-Test or Wilcoxon test.

RESULTS

A total of 29,679 genomes were downloaded, with 24,765 chromosomes (Chrs) and 21,006 plasmidic contigs (PCs) carrying 1,041 and 100 ARG patterns, respectively. The most abundant ARG in Chrs was mecA, while in PCs it was blaZ, both of which were highly prevalent in almost all CCs, while mecC was only found in CC1, CC130 and unassigned CCs. CC398 showed multiple ARGs, while CC8 harboured ARGs for β-lactams, tetracyclines and aminoglycosides. CC5 showed ARGs for β-lactams, aminoglycosides and macrolides. More than half of the genomes lacked metadata, however, in the Chrs, mecA and blaZ were higher among human and environmental isolates. In PCs, tet(K) was more abundant in the environment, while blaZ was spread throughout all of the categories. In the Chrs, mecA, erm(A), aph(3')-III, spc and erm(C) showed a decreasing trend, while tetracycline resistance genes, erm(B), fexA and lnuB increased. In PCs, blaZincreased over time, as did tet(L), mphC, ant(6)-Ia and msrA, while ermA, ermC, tet(K) and spc showed a decreasing trend. In the Chrs, differences were observed between continents as regards the occurrence of 13 ARGs. In PCs, ermC was more abundant in Asia, while mphC, msrA and spc were more abundant in America.

DISCUSSION

More than half of the genomes downloaded exhibited a significant lack of metadata, highlighting the necessity of compulsory minimum metadata requirements for genome data submission. In the literature, most of the works focus on ARGs found in a limited number of genomes, often isolated from specific sources, in a limited period or from a specific location, while a global view is presented in this work.

The most dominant ARGs found are associated with resistance to those antimicrobials that are more frequently used in humans and cattle. Furthermore, we were able to confirm the relevance of some mobile genetic elements. For example, we located the majority of the ARGs found in CC130 on its plasmidic contigs. These findings give an idea of which ARGs can be easily mobilised between different strains and different bacterial species. The temporal and geographical analysis showed differences in the occurrence of ARGs that are in line with general data on the usage of antibiotics worldwide. Here, the differentiation between plasmidic and chromosomic ARGs is important for understanding the potential spread of AMR.