Searching Algorithm for Type IV Effector proteins (S4TE) 2.0: tool for Type IV effectors prediction

Christophe Noroy[1,2,3], Chávez AS. [1,2], Nathalie Vachiéry[1,2], Thierry Lefrançois[1,2] and Damien F. Meyer[1,2*]

[1]CIRAD, UMR CMAEE, F-97170 Petit-Bourg, Guadeloupe, France
[2]INRA, UMR1309 CMAEE, F-34398, Montpellier, France
[3]Université des Antilles, 97159 Pointe-à-Pitre Cedex, Guadeloupe, France

Bacterial pathogens have evolved numerous strategies to corrupt, hijack or mimic cellular processes in order to survive and proliferate. Among those strategies, Type IV effectors (T4Es) are proteins secreted by pathogenic bacteria to manipulate host cell processes during infection. They are delivered into eukaryotic cells in an ATP-dependent manner by a specialized multiprotein complex, the type IV secretion system. T4Es contain a wide spectrum of features such as eukaryotic-like domains, localization signals or a C-terminal translocation signal. A combination of these 14 features enables prediction of T4Es in a given bacterial genome. In this research, we implemented a workflow, called Searching Algorithm for Type IV Effector proteins 2.0 (S4TE 2.0), to provide a comprehensive computational tool for accurate prediction and comparison of T4Es with a web-based graphical user interface. Applications range from characterizing effector features and identifying potential T4Es to analyzing effectors localization among the genome, according to G+C composition and local gene density. Following upload of Genbank files, bacterial genomes can be analyzed with default or user parameters. Furthermore, each feature can be searched independently making S4TE2.0 a useful tool to analyze a genome. Finally, S4TE 2.0 allows the comparison of putative T4Es repertoires among up to four bacterial strains. The software identifies T4Es orthologs between compared strains and returns the Venn diagram and lists of genes for each intersection. Last, we added interactive new features to offer the best visualization experience of the localization of identified candidate T4Es, including hyperlinks to NCBI and Pfam databases. S4TE 2.0 has been conceived to evolve rapidly with the publication of new experimentally validated T4Es, which will reinforce the predictive power of the algorithm. Our computational methodology is general and can be applied to the identification of a wide spectrum of bacterial effectors that lack sequence conservation but have similar amino acid characteristics. This approach will provide highly valuable information about bacterial host-specificity, virulence factors and to identify host targets for the development of new anti-bacterial molecules.