OPEN ACCESS Freely available online

Journal of Proteomics & Bioinformatics

Editorial

An Editorial on ActDES: A Curated Actinobacterial Database for Evolutionary Studies

Paul Anderson*

Department of Proteomics, Edge Hill University, St Helens Road, Ormskirk, Lancashire L39 4QP, UK

Actinobacteria is a large and diverse phylum of bacteria that contains medically and ecologically relevant organisms. Many members are valuable sources of bioactive natural products and chemical precursors that are exploited in the clinic and made using the enzyme pathways encoded in their complex genomes. Whilst the number of sequenced genomes has increased rapidly in the last 20 years, the large size, complexity and high G+C content of many *actinobacterial* genomes means that the sequences remain incomplete and consist of large numbers of contigs with poor annotation, which hinders large-scale comparative genomic and evolutionary studies. To enable greater understanding and exploitation of *actinobacterial* genomes, specialized genomic databases must be linked to high-quality genome sequences. Here, we provide a curated database of 612 high-quality *actinobacterial* genomes from 80 genera, chosen to represent a broad phylogenetic group with equivalent genome re-annotation. Utilizing this database will provide researchers with a framework for evolutionary and metabolic studies, to enable a foundation for genome and metabolic engineering, to facilitate discovery of novel bioactive therapeutics and studies on gene family evolution. This article contains data hosted by Microreact. The *Actinobacteria* is a large diverse phylum of bacteria, often with large, complex genomes with a high G+C content. Sequence databases have great variation in the quality of sequences, equivalence of annotation and phylogenetic representation, which makes it challenging to undertake evolutionary and phylogenetic studies. To address this, we have assembled a curated, taxa-specific, non-redundant database to aid detailed comparative analysis of *Actinobacteria*. The increase in availability of bacterial whole-genome sequencing provides large amounts of data for evolutionary and phylogenetic analysis. However, there is great variation in the quality, annotation and phylogenetic skew of the data available in large universal databases, meaning that evolutionary and phylogenetic studies can be challenging.

To address this variation, curated, high-level, taxa-specific, non-redundant sub-databases need to be assembled to aid detailed analysis. Given that there is a direct correlation between phylogenetic distance and the discovery of novel function, it is imperative that any derived databases must be phylogenetically representative and non-redundant to enable insight into the evolution of genes, proteins and pathways within a given group of taxa. The phylum *Actinobacteria* is a major taxon amongst the Bacteria, which includes phenotypically and morphologically diverse organisms found on every continent and in virtually every ecological niche. They are particularly common in soils, yet within their ranks are potential human and animal pathogens such as *Corynebacterium*, *Mycobacterium*, *Nocardia* and *Tropheryma*, inhabitants of the gastrointestinal tract (Bifidobacterium and Scardovia), as well as plant commensals and pathogens such as *Frankia*, *Leifsonia* and *Clavibacter*. Perhaps the most notable trait of the phylum is the renowned ability to produce bioactive natural products such as antibiotics, anti-cancer agents and immuno-suppressive agents, with genera such as *Amycolatopsis*, *Micromonospora* and *Streptomyces* being particularly prominent. As a result, computational 'mining' of *actinobacterial* genomes has become an important part of the drug-discovery pipeline, with increasing numbers of online resources and software devoted to identification of natural-product biosynthetic gene clusters (BGCs). It is important to move beyond approaches that rely on similarity searches of known BGCs and to expand searches to identify hidden chemical diversity within the genomes. A recent study of 830 *actinobacterial* genomes found >11 000 BGCs comprising 4122 chemical families, indicating that there is a vast diversity of strains and chemistry to exploit, yet within each of these strains there will be hidden diversity in the form of cryptic BGCs. To exploit this undiscovered diversity as the technology develops and databases expand, new biosynthetic logic will emerge, yet we know little of how natural selection shapes the evolution of BGCs and how biosynthetic precursors are supplied to gene products of BGCs from primary metabolism and to identify targets for metabolic engineering of industrially relevant strains.

Correspondence to: Paul Anderson, Department of Proteomics, Edge Hill University, St Helens Road, Ormskirk, Lancashire L39 4QP, UK; E-mail: andersonp@ehu.uk

Received: June 02, 2021; **Accepted:** June 16, 2021; **Published:** June 23, 2021

Citation: Anderson P. (2021) An Editorial on ActDES: A Curated Actinobacterial Database for Evolutionary Studies. J Proteomics Bioinform. 14:e125.