# Genome sequence of p. acnes ed1 strain

**Abstract**

*Propionibacterium acnes* is a Gram-positive bacterium that forms part of the normal human microbiota. *P. acnes* is pervasive on human skin and under specific conditions can be an opportunistic pathogen. The draft genome sequence of *P. acnes* ED1 strain, which was isolated from a hip implant, was examined in order to investigate the genomic biology of these pathogens. Various methods were used in order to examine the genome sequence of the *P. acnes* strain ED1 as well as its assembly and annotation. From these findings, the length of the genome was 2. 8 Mb (200 contigs, N50= 86 kb) and the GC content was 60%. The coding sequences of P. acnes genome (2, 788) in 318 subsystems was found as well as the allocation of the genes in the genome. The histidine metabolism and antibiotic resistance genes of P. acnes are analysed further in comparison with the P. acnes strain KPA171202 reference genome.

## The Genome Announcement

*Propionibacterium acnes* is a bacterium of normal human microbiota that mainly colonises the skin. It is associated with many infections such as acne vulgaris and other skin diseases (1). For that reason, sequencing this genome is really important.

A sample of P. acnes ED1 strain, isolated from a hip implant at the time of revision operation, was examined. The DNA concentration was measured using the Qubit protocol. A transposome-based sequencing library was generated using Illumina NextERA XT Tagmentation kit (2) and it was amplified by PCR. This genomic library was subsequently submitted to the *Edinburgh Genomics Facility* in order to be normalised and then sequenced

on the *Illumina* MiSeq instrument (2). The raw DNA sequence data were trimmed for adapters and low quality sequences.

The quality of the data was assessed using *FASTQC* (3) which indicated the GC content over all reads which resulted in the generation of two peaks (Figure 1). One of them showed the expected for the P. acnes (4) 60% GC and the other one a much lower GC content (roughly 34%). The presence of the second peak of GC content indicated that the ED1 sample was contaminated with DNA from another bacterium. This bacterium is most likely to be *Staphylococcus lugdunensis* as it also colonises the human skin and its GC content is 33. 87% (5, 6). The FASTQC analysis displayed the sequence duplication levels that were really high (>= 40. 52%) and possibly to incorrect library generation. Many duplication events contributed to the reduction of the coverage of the P. acnes ED1 genome. The fold coverage of the P. acnes ED1 genome was 49. Contamination, library errors and tagmentation bias might be the reason why data is of such low quality.

*De novo* genome assembly was performed using *CLC Genomics Workbench* (7). The quality of the genome assembly was assessed using QUAST (8). The total length of the *Propionibacterium* genomes is between 2. 2 and 2. 5 Mb. QUAST showed that the total length of the P. acnes ED1 was about 2. 8 Mb (200 contigs, N50= 86 kb) and that might be due to contamination with *Staphylococcus* . Thereafter, the genome assembly was submitted to the RAST annotation server, (9) which using the SEED system, identify the genes. The SEED system (10) identified the genes that are present in the P. acnes ED1 genome but absent from the P. acnes strain KPA171202 reference. A SEED comparison revealed that 766 of the total 2, 298 protein

sequences have similarity > 99%. Most proteins/genes are not identical with the reference genome due to mistakes in PCR amplification or duplications events. The SEED annotated the P. acnes ED1 genome and defined that the coding sequences were 2, 788 and the number of subsystems was 318. A histogram showed that the proportion of features in the P. acnes ED1 genome that were annotated by SEED was 39%. Additionally, the annotated functions in the genome were 1, 306 (Figure 3). From all these findings, it can be concluded that there are genes being repeated in many other subsystems. In order to examine the allocation of the genes and whether the genome was complete, it is necessary to further identify more specific processes.

## Two Analysis Summaries

The P. acnes ED1 genome with the reference genome is now compared in terms of amino acid metabolism and potential virulence (antibiotic resistance).

The examination of histidine metabolism is really interesting in order to assess whether the P. acnes ED1 genome assembly is complete related to the reference genome and is sufficient in a complete biochemical pathway. Using the SEED system, the identification of the presence of histidine biosynthesis and degradation genes in both genomes was accomplished. The KEGG map in SEED system shows that the P. acnes ED1 genome contains exactly the same set of 10 histidine biosynthesis genes and 5 histidine degradation genes as the reference genome. For example, ATP phosphoribosyltransferase is an enzyme that catalyses the first step in the histidine biosynthesis in P. acnes and has a crucial role in the pathway

because the rate of histidine biosynthesis appears to be controlled mainly by regulation of HisG enzymatic activity. Another enzyme is the Imidazoleglycerol-phosphate dehydratase that catalyses the seventh step in the biosynthesis of histidine. Furthermore, the Histidinol dehydrogenase (HDH) catalyses the terminal step in the biosynthesis of histidine. Moreover, there are some proteins that catalyse the histidine degradation as the imidazolonepropionase catalyses the third step in degradation of histidine and the histidine ammonia-lyase catalyses the first reaction in histidine degradation (11). Therefore, all these histidine biosynthesis and degradation proteins play an important role in histidine metabolism. Histidine biosynthesis is really important for the survival of P. acnes ED1 genome (Figure 4).

The examination of P. acnes ED1 genome in terms of antibiotic resistance and sensitivity is also an important aspect, as for any clinical infection it is vital to choose the correct drugs for treatment. Using SEED system in RAST server, the identification of the genes resistant to antibiotics and toxic compounds was achieved. The P. acnes ED1 genome contains exactly the same set of 10 antibiotic and toxic resistance genes as the reference genome. More specifically, the P. acnes ED1 genome has the cobalt/zinc/cadmium resistance protein and the mercuric reductase protein that confer resistance to cobalt, zinc, cadmium and mercury respectively. So in an environment with high concentration of cobalt, zinc, cadmium and mercury, the P. acnes ED1 would not be poisoned. Thereafter, the P. acnes ED1 genome shows resistance against fluoroquinolone antibiotics. Fluoroquinolone antibiotics disrupt the Propionibacterium DNA replication by

degenerating DNA gyrase and topoisomerase in a concentration-dependent manner. It could be mentioned that, P. aeruginosa shows increased fluoroquinolone resistance, which could be evidence for intra-generic horizontal gene transfer and evolution in action in Propionibacterium genomes. All these remarks are consistent with there being fluoroquinolones antibiotic resistance in the P. acnes ED1 genome (12). It would be of interest to study this HGT event by a Protein BLAST analysis (13); yet, further examination would be outside the scope of this analysis. However, there are beta-lactamase class C and other penicillin binding proteins in P. acnes ED1 genome but not in the reference one. Beta-lactamase is an enzyme that provides resistance in beta-lactam antibiotics like penicillin. The lack of beta-lactamase from the reference genome might be due to wrong annotation. Yet, a cytoplasmic copper homeostasis protein (CutC) is present in the reference genome but absent from the P. acnes ED1. This protein is really important as it confers resistance to copper. It could be mentioned that mutation of this protein leads to an increased copper sensitivity. The absence of CutC from the P. acnes ED1 genome might be caused by incorrect annotation (11).