Genome sequencing and analysis of the emerging pathogen *Corynebacterium ulcerans*

Eva Trost, Andreas Tauch

Institut für Genomforschung und Systembiologie, Centrum für Biotechnologie, Universität Bielefeld, Bielefeld, Germany

Abstract

Since the first isolation of *Corynebacterium ulcerans* from a throat lesion, described in 1927, the human and animal pathogen is becoming the major cause of diphtheriae in industrialized countries. As *C. diphtheriae, C. ulcerans* can be infected by corynephage β, that codes for the diphtheria toxin. For a better understanding of the cellular physiology of *C. ulcerans* the complete genome of the isolates 809 and BR-AD22, has been sequenced by 454 technology. The strain 809 was isolated from a patient with an abnormal infection of the lung, while BR-AD22 was isolated from a healthy dog.

Introduction

Corynebacterium ulcerans has been detected as a commensal in domestic and wild animals that may serve as reservoirs for zoonotic infections. During the last decade, the frequency and severity of human infections associated with *C. ulcerans* appear to be increasing in various countries. As the knowledge of genes contributing to the virulence of this bacterium was very limited, the complete genome sequences of two *C. ulcerans* strains detected were determined and characterized by comparative genomics: The molecular data provides considerable knowledge of virulence factors in *C. ulcerans*.This bacterium is apparently equipped with a broad and varying set of virulence factors, including a novel type of a shiga-like toxin. Whether the respective protein contributes to the severity of human infections (and a fatal outcome) remains to be elucidated by targeted genetic experiments with defined bacterial mutants.

Materials and methods

The genome sequences of *C. ulcerans* and *C. ulcerans* BR-AD22 were determined by pyrosequencing using a quarter of a sequencing run with the Genome Sequencer FLX Instrument for each strain. The resulting reads were assembled, and the remaining gaps were closed by PCR strategies that were supported by the related reference contig arrangement tool r2cat [Husemann and Stoye, 2010] using the genome sequence of *C. pseudotuberculosis* FRC41 as a reference sequences were subsequently analyzed with the microbial genome annotation system GenDB.

General features of the genomes

Feature	809	BR-AD22
Genome size [bp]	2,502,095	2,606,374
Mean G+C content of DNA	53.3%	53.4%
No. of ribosomal RNAs	4 x	3 x (16S-23S-5S)
No. of transfer RNAs	51	51
No of protein-coding sequences	2,183	2,339
Coding density	87.7%	87.8%
Average gene length [bp]	1,006	979

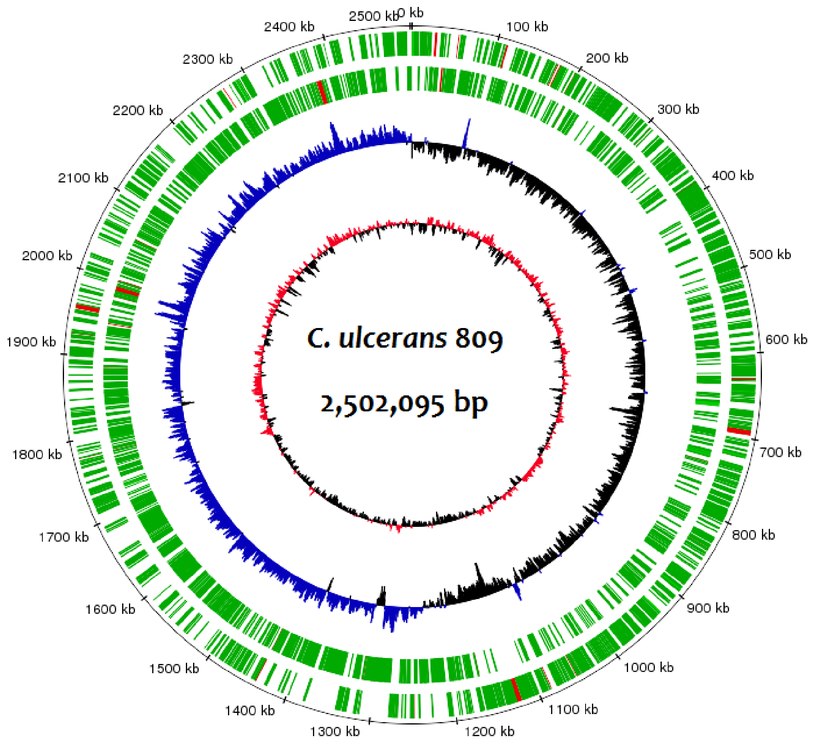

Fig. 1.Circular representation of the chromosomes from *C. ulcerans* 809. The circles represent the following features: circle 1, DNA base position; circles 2 and 3, predicted coding sequences transcribed clockwise and anticlockwise, respectively; circle 4, G/C skew [(G+C)/(G+C)] plotted using a 10-kb window; circle 5, G+C content plotted using a 10-kb window. Color code in circles 2 and 3: green, predicted protein-coding regions; red, rRNA or tRNA genes.

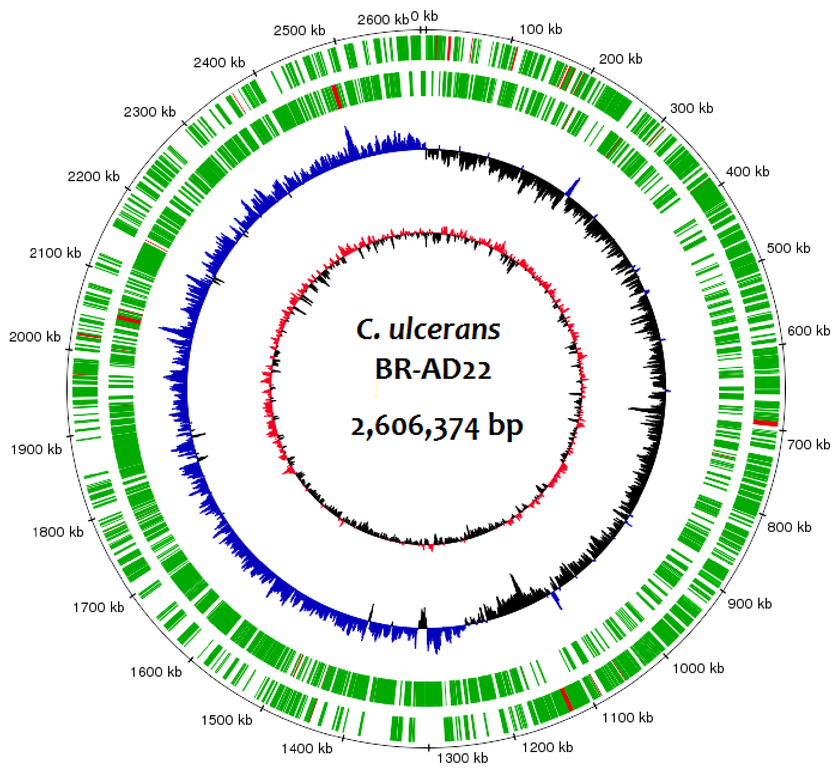

Fig. 2. Circular representation of the chromosomes from *C. ulcerans* BR-AD22. The circles represent the following features: circle 1, DNA base position; circles 2 and 3, predicted coding sequences transcribed clockwise and anticlockwise, respectively; circle 4, G/C skew [(G+C)/(G+C)] plotted using a 10-kb window; circle 5, G+C content plotted using a 10-kb window. Color code in circles 2 and 3: green, predicted protein-coding regions; red, rRNA or tRNA genes.

Comparative analysis

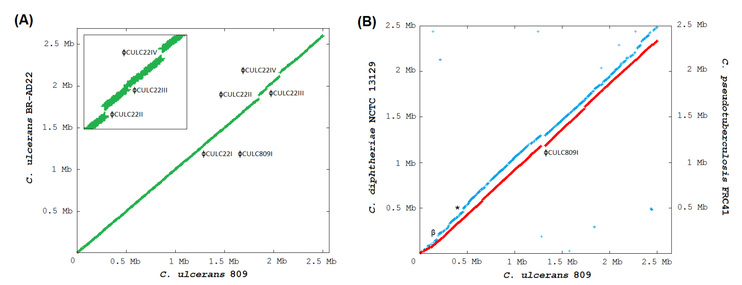

Fig. 3.. **(A),** Synteny between the sequenced chromosomes of *C. ulcerans* 809 and *C. ulcerans* BR-AD22. **(B),** Synteny between the chromosome of *C. ulcerans* 809 and those from *C. diphtheriae* NCTC 13129 (blue) and *C. pseudotuberculosis* FRC41 (red). The graphs represent X-Y plots of dots forming syntenic regions between the selected chromosomes. Each dot represents a predicted protein having an orthologous counterpart in another corynebacterial genome, with co-ordinates corresponding to the position of the respective coding region in each genome. Orthologous proteins were detected by reciprocal best BLASTP matches. The genomic positions of putative prophages detected in *C. ulcerans* are marked in the synteny plots. Symbols: β, β-corynephage of *C. diphtheriae* NCTC 13129; asterisk, nitrate reductase gene region

Comparison of the predicted gene content

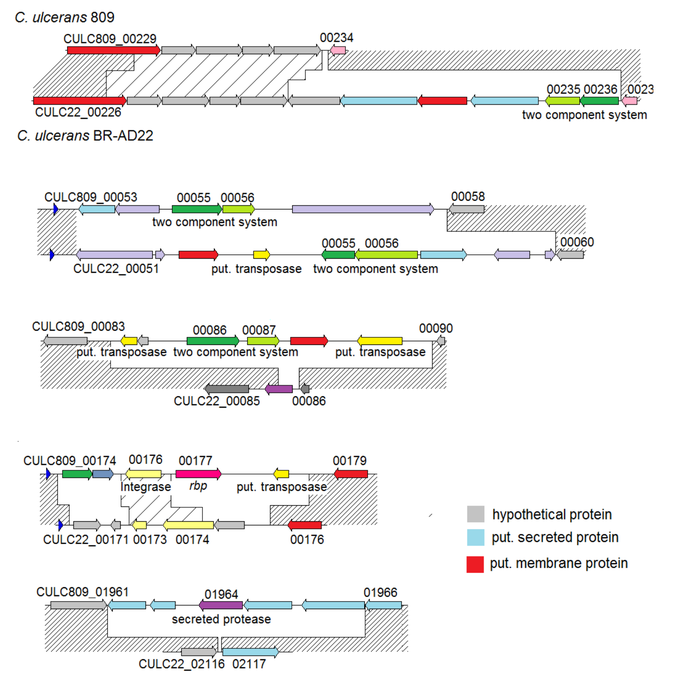

Fig. 4: **(A),** Selected examples of genomic regions comprising strain-specific genes in *C. ulcerans*. Orthologous gene regions are shadded gray. **(B),** Venn diagramm comparing the gene content of *C. ulcerans* 809, *C. diphtheriae* NCTC 13129 and *C. pseudotuberculosis* FRC41. The Venn diagram shows the number of shared and species-specific genes among the three corynebacterial genomes.

References

Tauch et al. 2006. **Ultrafast de novo sequencing of Corynebacterium urealyticum using the Genome Sequencer 20 System**. Biochemica 4, 4-6.
Tauch et al. **Complete genome sequence and analysis of *Corynebacterium urealyticum*, a causative agent of urinary tract infections in hospitalized patients.** Bacteriol. Biotechnol., in press.
Tauch et al.**Ultrafast pyrosequencing of *Corynebacterium kroppenstedtii* DSM44385 revealed insights into the physiology of a lipophilic corynebacterium that lacks mycolic acids**. Bacteriol. Biotechnol., in press.