Identification of Prophages and Prophage Remnants within the Genome of *Avibacterium paragallinarum* Bacterium

Y. Roodt, R. R. Bragg, and J. Alberyn

Department of Microbial, Biochemical and Food Biotechnology, University of the Free State, Bloemfontein 9300, South Africa

Correspondence should be addressed to R. R. Bragg, braggr@ufs.ac.za

Received 10 September 2012; Revised 21 November 2012; Accepted 28 November 2012

Academic Editor: Alexei Sorokin

Copyright © 2012 Y. Roodt et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Bacterial whole genome sequencing has delivered an abundance of prophage sequences as a by-product and the analysis of these sequences revealed ways in which phages have affected the genome of their host bacteria in various bacterial species. The aim of this study was to identify the phage-related sequences in the draft assembly of the *Avibacterium paragallinarum* genome, the causative agent of infectious coryza in poultry. Whole genome assembly was not possible due to the presence of gaps and/or repeats existent on the ends of contigs. However, genome annotation revealed prophage and prophage remnant sequences present in this genome. From the results obtained, a complete Mu-like bacteriophage could be identified that was termed AvpmuC-2M. A complete sequence of HP2-like bacteriophage, named AvpC-2M-HP2, was also identified.

1. Introduction

Bacteriophages were first discovered in England, 1915, by Frederick W. Twort and independently thereof in 1917 by Felix d’Herelle at the Pasteur Institute in Paris [1, 2]. Bacteriophages are viruses which infect bacteria and can be divided into two groups according to their means of interaction with the bacterial cell, namely, lytic (virulent) and temperate (lysogenic) phages. Lytic phages proceed typically with replication the moment after infecting the host cell, where large numbers of new viruses are released through the lysis of the host cell. Temperate phages do not necessarily start replication immediately after infection, and depending on a number of conditions, these phages may integrate their chromosome into the genome of the host cell thus remaining silent until induced [1]. Once a bacteriophage genome is integrated into the host cell genome it is referred to as a prophage. Prophages are the primary suspects in the adaptation event of existing pathogens to new hosts or the emergence of new pathogens or epidemic clones [3]. Bacteria have no sexual life cycle and the exchange of alleles within a population is fulfilled via horizontal gene transfer, and the source of this DNA may be phages, amongst others [4]. Unless restricted by the species barrier, entire functional units can be imported from these sources and the transferred DNA can range from 1 kb to more than 100 kb in size and can encode complex surface structures or even entire metabolic pathways [4]. It has long been established that bacteriophages contribute to the pathogenicity of their bacterial hosts as many genes have been shown to undergo transfer among bacteria through phages [4]. These genes can code for a diverse subset of virulence factors such as toxin, regulatory factors (which upregulate the expression of host virulence genes), and enzymes, which can alter the bacterial virulence components [4].

Bacterial whole genome sequencing has delivered an abundance of prophage sequences [6]. Prophages can constitute as much as between 10 and 20% of a bacterium’s genome though many of these prophages are cryptic and in a state of mutational decay [7]. Analysis of these sequences revealed numerous ways in which prophages shape the genome of their host bacteria [6]. Prophages have the ability to affect their host genome architecture by changing the content of the genome, by modifying the organization of the genome, or by altering the location and the order of genes [3, 8]. Temperate bacteriophages play an intricate role in the generation of microbial diversity and in the evolution of bacterial genomes through the mediation of rearrangement within the bacterial
chromosome and as a result they contribute significantly towards interstrain differences within the same bacterial species [3, 7, 9]. By comparison, two *Streptococcus pyogenes* strains belong to different M serotypes and they are unconnectedly associated with different diseases—but when these strains were compared on DNA level the key differences correlated to prophage sequences [3]. Another example can be observed within *Campylobacter jejuni*, where interstrain differences may be attributed towards intragenomic inversions of Mu-like prophage DNA sequences [10]. Comparisons between colinear genomes showed that gaps between the relevant genomes were attributed to the presence of prophage sequences or as a result of rearranged bacterial genomes [11]. In many cases prophages with limited DNA sequence identity or duplicated prophages serve as anchoring points for homologous recombination [3]. In addition prophages can recombine with other prophages contributing to their mosaic structure [11].

*Avibacterium paragallinarum* is a pathogen targeting chickens and is causing the disease infectious coryza [12, 13]. Phylogenetically this bacterium belongs to the *Pasteurellaceae* family and was formerly known as *Haemophilus paragallinarum* renamed in 2005 [14]. The family *Pasteurellaceae* is made up of closely related bacteria [14] and within this family, various bacteriophages have been reported in *Haemophilus influenzae*, *Actinobacillus actinomycetemcomitans*, *Pasteurella multocida*, and *Mannheimia haemolytica*. Bacteriophage-like sequences have been reported in *H. somni* [2, 15]. In this paper, we describe prophage and prophage remnants detected in *Av. paragallinarum*.

## 2. Materials and Methods

### 2.1. Bacterial Strains and Growth Conditions. *Av. paragallinarum* strain C-2 (Modesto) was obtained from Onderstepoort Biological Products, Onderstepoort, South Africa. Strain Modesto is NAD⁺ dependent and was cultivated in TM/SN supplemented with 1% (v/v) chicken serum, 1% (v/v) NAD⁺, and 0.0005% (m/v) thiamine solution and cultivated under oxygen limiting conditions in a candle jar at 37°C for 16 h [16].

### 2.2. DNA Extraction. Genomic DNA extraction was performed with the use of a QIAamp DNA Mini kit from Qiagen. The manufacturer's recommendations were meticulously followed except for the elution procedure where genomic DNA was eluted in 2 × 80 μL 10 mM Tris-HCl, pH 8.

### 2.3. Identification of *Av. paragallinarum*. *Av. paragallinarum* was identified by using a PCR protocol developed specifically for this bacterium [17]. Additionally, the 16S ribosomal DNA (rDNA) region was amplified in order to identify *Av. paragallinarum* [18]. The PCR product obtained was excised from the agarose gel, purified, and directly sequenced to avoid any possible foreign genomic DNA contamination.

### 2.4. Genome Assembly and Annotation. Genomic DNA samples of *Av. paragallinarum* strain C-2 (Modesto) were sent to Agowa Genomics (now known as LGC Genomics), Germany, for GS-FLX Titanium pyrosequencing. Sequences/reads obtained from the pyrosequencing were assembled using Newbler 2.0.01.14 from 454 Life Sciences Corporation with an overlap minimum match identity of 90% and an overlap minimum match length of 40 bases. The sequences of obtained contigs were submitted to the J. Craig Venter Institute (JCVI) as a pseudomolecule for annotation. The pipeline includes gene finding with Glimmer [19, 20], BLAST-Extend-Repraze (BER) searches [21, 22], Hidden Markov Models (HMM) searches [23], Trans Membrane prediction [24], SignalP predictions [25], and automatic annotations from AutoAnnotate. All of this information is stored in a MySQL database and associated files which was downloaded to our site. The manual annotation tool Manatee was downloaded from SourceForge (http://manatee.sourceforge.net/) and used to manually review the output from the pipeline. pDraw32 from ACACLONE software was used to draw genetic maps for prophage region.

### 2.5. Accession Numbers. AvpmuC-2M (accession no.: JN627905); HP-2 like prophages contig A (JN627906), B (JN627907), and C (JN627908); lamboid prophage
**Figure 2:** Lambdoid genetic representation. Various lambda genes have been mapped to the contigs from which they were identified and grouped according to associated functions [5]. Beginnings and ends of the contigs are indicated by //.

**Figure 3:** HP2-like prophage genetic representation. The various contigs (a), (b), and (c) carrying HP2-like open reading frames are illustrated within this diagram. A complete HP2-like prophage AvpC-2M-HP2 was identified within the genome of *Av. paragallinarum* Modesto represented by (c) in this diagram.
fragments (accession nos.: JN627909-JN627919); prophage integrase genes (accession nos.: JN627920-JN627928).

3. Results and Discussion

The whole genome sequencing project of *Av. paragallinarum* yielded a total number of 160 743 reads/sequences through pyrosequencing chemistry. The total number of contigs assembled from these sequences was 300 with the largest contig comprising 78 465 bp and the average contig size longer than 500 bp were 10 727 bp. A closed genome could not be assembled for *Av. paragallinarum* from the sequencing results obtained due to repeat sequences that existed on the end(s) of contigs or as a result of sequence gaps between contigs. A pseudogenome molecule was assembled from the 300 contigs obtained and was sent to the JCVI for genome annotation. Annotation results indicated that a total of 7% or 141 open reading frames were assigned to phage protein functions. The 141 open reading frames were identified on 60 of the 300 contigs where either one or a series of prophage genes were mapped. Nine prophage integrase genes were identified and mapped to seven of the 60 contigs. Investigations into open reading frames adjacent to the identified integrases revealed no additional prophage-related genes, but rather genes encoding membrane proteins; metabolic proteins; ribosomal proteins; cell membrane function proteins, or membrane export proteins. No significant homology was observed between the integrases.

Mu-like prophage genes were identified and mapped to 11 different contigs. The 11 contigs were compared to the genome map of Mu-like phages reported by [26]. Here we suggest a putative Mu-like prophage for *Av. paragallinarum*, AvpMuC-2M (Figure 1). A large portion of the Mu gene (25) was identified on a single contig comprising 27, 107 bp. Two additional contigs were identified completing the map of a Mu-like prophage for *Av. paragallinarum*.

Forty-one lamboid genes were mapped to 11 different contigs (Figure 2). With the data available from the whole genome sequencing project it could not be concluded if the lamboid prophage is complete or if it has undergone massive loss of functional DNA resulting in bacteriophage genome fragmentation and ultimately leading to its disappearance [3].

A complete lysogen of a temperate phage was identified on a single contig within the genome. The succession of the genes identified is similar to that of *H. influenzae* HP2 prophage. A nucleotide-to-nucleotide comparison revealed a weak similarity between the *H. influenzae* prophage HP2 and the identified prophage. The identified prophage is therefore unique to *Av. paragallinarum*. We suggest calling it AvpC-2M-HP2.

Two additional contigs also contained HP2-like genes (Figure 3). Further investigations of similarity between the identified contigs (A, B, and C of AcpC-2M-HP2) indicated that the three contigs share the same genomic organization but they are not identical.

An abundance of other prophage genes have also been identified but could not be assigned to any specific phage or phage family.

Thus the mapping of all of the prophage genes uncovered allowed us to suggest two prophages in *Av. paragallinarum*. One prophage, AvpMyC-2M, resembles a Mu-like prophage, whereas the other prophage, AvpC-2M-HP2, resembles the HP2 prophage of *H. influenzae*.

Conflict of Interests

The authors declare that no conflict of interests exists with any of the commercial identities mentioned within the paper.

Acknowledgment

The authors would like to thank JCVI for providing the JCVI Annotation Service which provided them with automatic annotation data and the manual annotation tool Manatee.

References


