Phage genome annotation: is this phage actually safe for therapeutic use ?

Antoine Culot. Rime Bioinformatics SAS, 104 avenue du Général Leclerc, 91120, Palaiseau, FRANCE.

INTRODUCTION : Phage's genomes peculiar structure

The study of phages genomic properties is one of the main steps to assess their safety for therapeutic use [1]. This process enables detection of genes that make phages potentially harmful for the patient or the environment, such as antibiotic resistance, lysogeny, and pathogenicity genes. Phage genome annotation is traditionally performed using bioinformatics tools that are designed for bacterial genomes [2]. However, the different genome structure of phages and lack of phage gene entries in annotation databases cause poor gene calling performance and poor gene function annotation when bacteria-focused tools are used for phages.

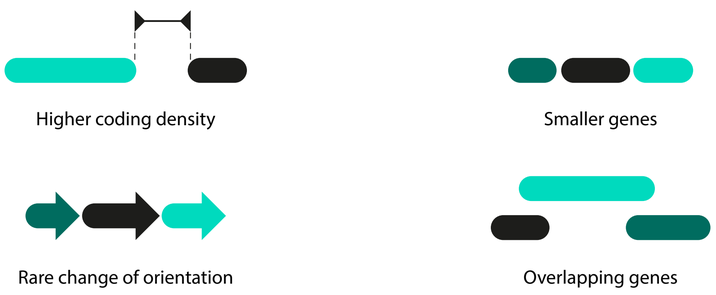
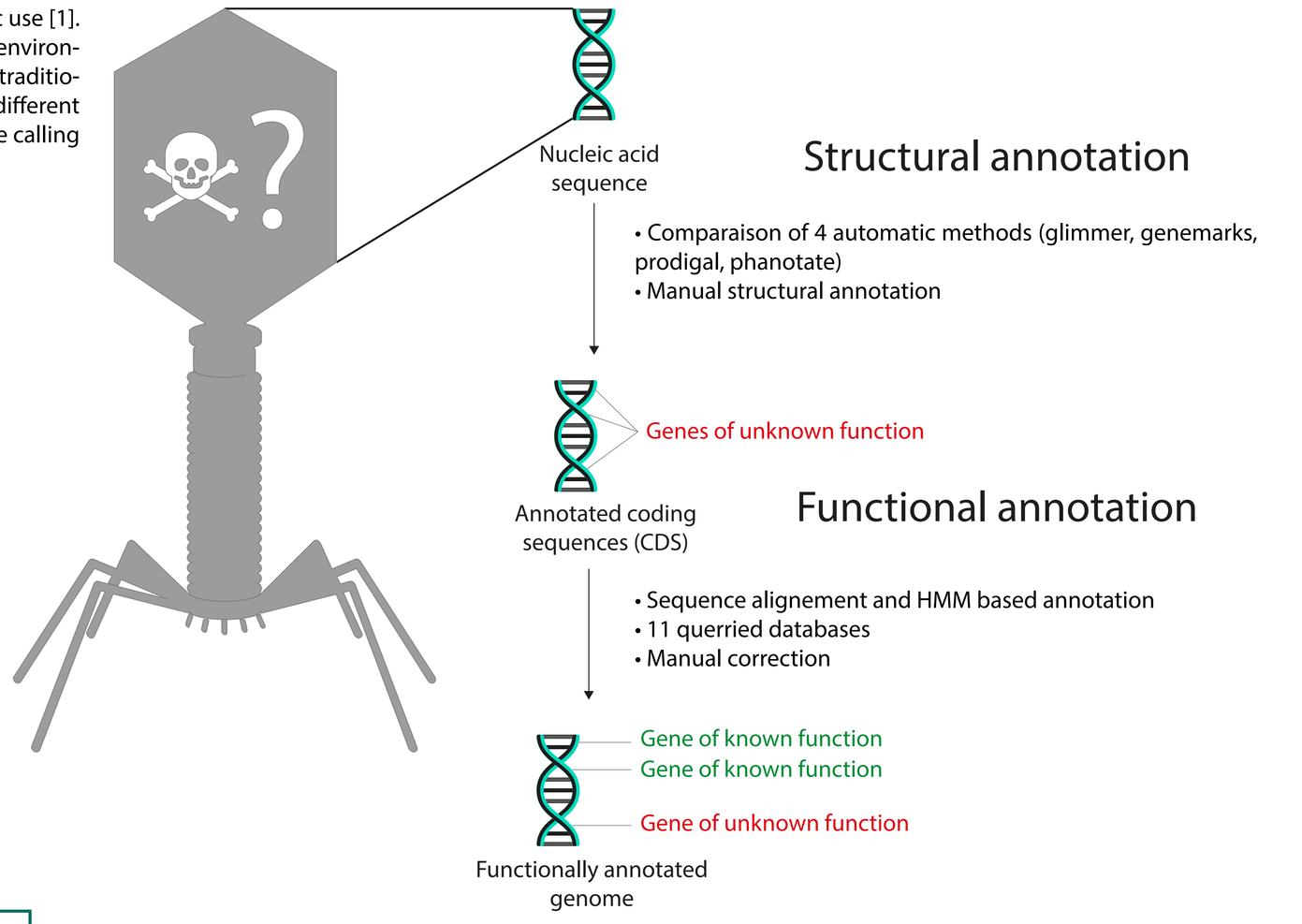


Fig 1. Phage genome structure peculiarities that should be taken into account for annotation

The peculiarities of phages genomes (Fig. 1) are rarely accounted for. This leads to insufficient characterization of phages that are considered for phage therapy. This problem is illustrated by phage VVP001, which infects *Vibrio vulnificus*, a marine bacteria responsible for heavy loss in aquaculture farms as well as food poisoning [3]. The very promising anti-bacterial properties of this virus have been studied *in vitro, in vivo, in silico* and published by Kim *et al* in 2021 [3]. The genome of VVP001 was studied using state of the art -but bacteria focused- analysis methods and is poorly described : only 17% of the genes are functionally annotated. **How can we be sure no dangerous genes are present in the remaining 83% of the genome ?** Similar cases to VVP001 (about 20% of functionnally annotated genes) are common in recent scientific literature, which underlines the difficulty of phage genome annotation.

Our aim is to demonstrate how combining multiple manual and automatic approaches for structural and functional annotation can improve the analysis result to reach higher confidence in the phage's harmlessness for patients and the environment.

Method: New annotation of a published genome



RESULTS & DISCUSSION: 5x more functions annotated

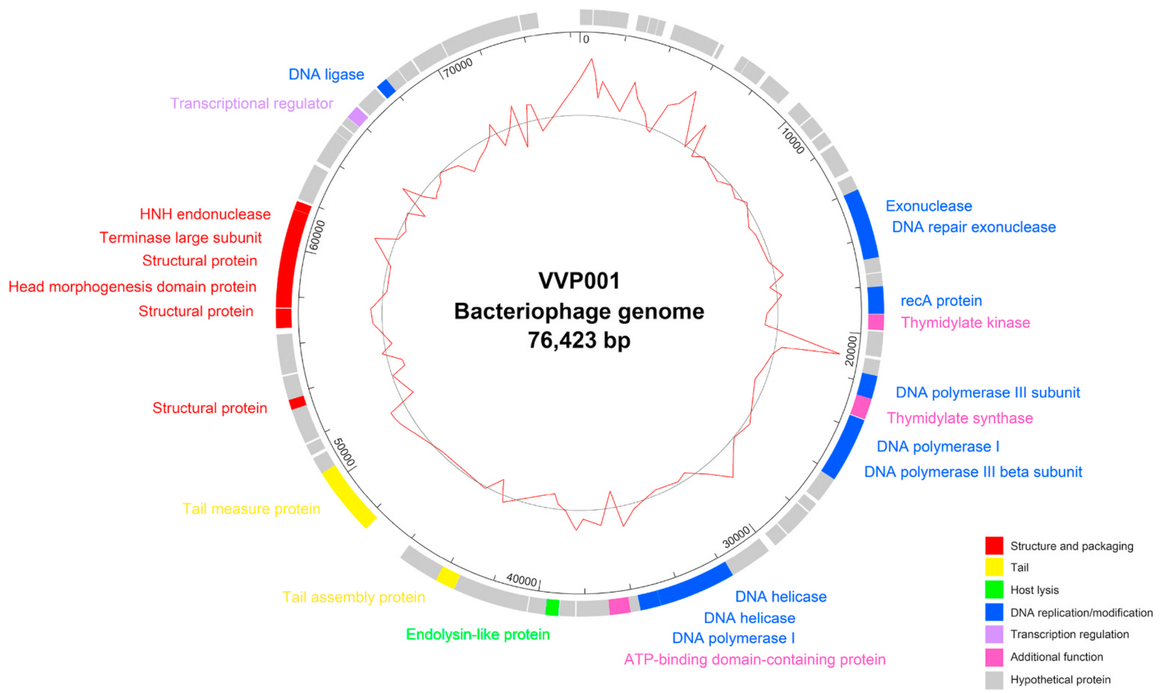


Fig 2. Genome map of phage VVP001, as published by Kim et al in 2021 [3]

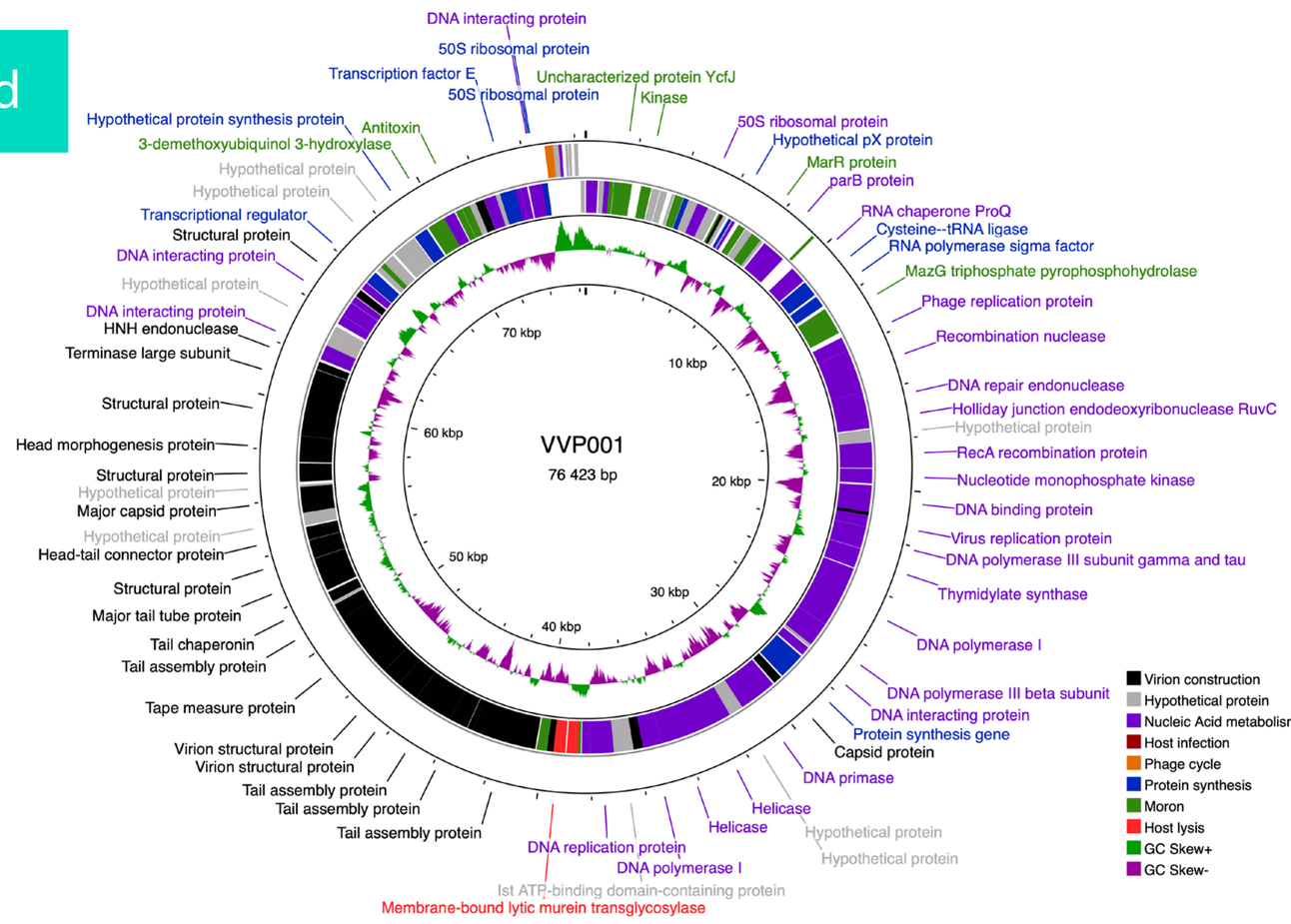


Fig 3. Genome map of newly annotated phage VVP001

The original annotation of VVP001's genome described **102** CDS, among which **17 (16,7%)** were functionally annotated (Fig 2). This annotation leaves surprisingly large non-coding zones (in white, at 45 906 bp or 74 683 bp for example) which are especially rare in the usually densely packed genomes of phages. **The new functionnal annotation** (Fig 3) enabled the detection of **126 (+24%)** genes. **93 of them (74%)** were functionally annotated.
At the end of their study of phage VVP001's genome, the authors underline the absence of toxin-coding or virulence factor coding genes, thus suggesting the harmlessness of this phage for phage therapy. This exemplifies the widely spread difficulty of phage genome annotation in the literature: authors are often left betting on the absence of unwanted genes when only 20-30% of the genome is annotated. Such conclusions are hazardous : using phages of which 70-80% of the genome is unknown could help spreading antibiotic resistance or virulence genes.

The analysis presented in this poster determined that the presence of unwanted genes is unlikely, thus making VVP001's usage for phage therapy safe, from the gene content point of view. Still, it is not impossible that new, never described and unwanted genes are present in this genome among the grey regions (Hypothetical proteins, Fig. 3). Despite the high quality of annotations currently achievable, bioinformatics tools still need to be improved to reach a full confidence on the harmlessness of a given phage.

Finally, the greater precision of the multi-method pipeline comes at the expense of time and computing power. Manual annotation is indeed a lengthy process, and so are the automatic functionnal annotation steps, which also require a powerful computing cluster to run. Such detailed analysis are therefore not reasonably doable for all phages.

CONCLUSION

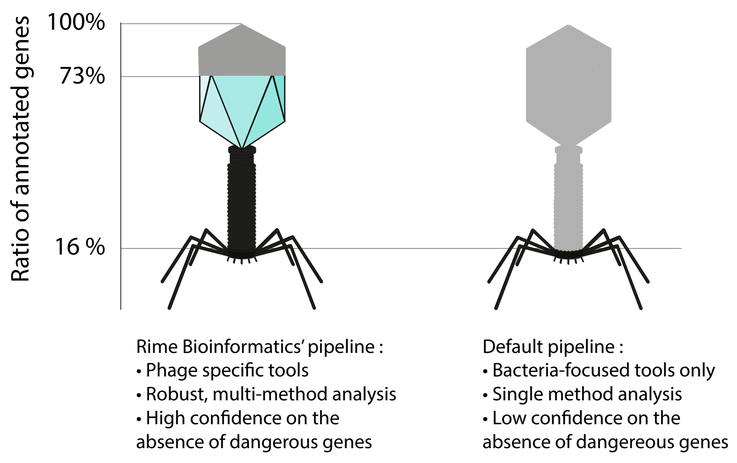


Fig 4. Comparison of the annotation results discussed in this study

Taking the peculiar genome structure of phages into account allowed to detect 33 (24%) more genes in phage VVP001's genome. As this detection is the result of the comparison of several gene calling methods, it is more robust. The usage of numerous large databases, querried using different strategies, allowed the functionnal annotation of 93 (73%) genes, when only 17 (16%) genes were previously associated with a function.

Knowing the function of 73% of the genome of phage VVP001 enables its use for phage therapy with greater confidence. It is now indeed less likely to discover lysogeny, virulence or antibiotic genes in this genome, which would forbid any therapeutic use for this phage. **Phage VVP001 seems therefore actually safe for therapeutic use.**

In order to promote the safest possible use of phage therapy for patients and for the environment, it is highly recommended to use multi-method, phage focused pipelines, which are more sensitive and robust (Fig 4).

[1] A. Culot, N. Grosset, M. Gautier, Overcoming the challenges of phage therapy for industrial aquaculture: A review, Aquaculture, 513 (2019) 734423.
[2] K. McNair, C. Zhou, E.A. Dinsdale, B. Souza, R.A. Edwards, PHANOTATE: a novel approach to gene identification in phage genomes, Bioinformatics, (2019)
[3] Hye-Jin Kim, You-Tae Kim, Hyeun Bum Kim, Sang Ho Choi, et Ju-Hoon Lee. Characterization of bacteriophage VVP001 and its application for the inhibition of Vibrio vulnificus causing seafood-borne diseases. Food Microbiology, 94:103630, April 2021. ISSN 07400020.