A Novel Genome-Wide, qPCR-Based Approach to Identify Genetic Determinants of Phage Host Range

Steven Gayder [1,2,3*], Lars Fieseler [1], Alan Castle [2], Antonet Svircev [3]

1. Zurich University of Applied Sciences, Wädenswil, Switzerland; 2. Brock University, St. Catharines, Ontario, Canada; 3. Agriculture and Agri-Food Canada, Vineland Station, Ontario, Canada. *steven.gayder@zhaw.ch

Introduction

Phages have enormous potential as tools to make targeted modifications to microbial populations, especially for the reduction of pathogenic bacteria in clinical, biotechnological, and agricultural settings. To maximize this potential, we need a far deeper understanding of how different phages interact with their target hosts. Identifying host genes that can affect phage infection can grant powerful insight into the mechanisms that phages use to infect their hosts. This information will be critical for designing smarter and more effective phage-based treatments and products.

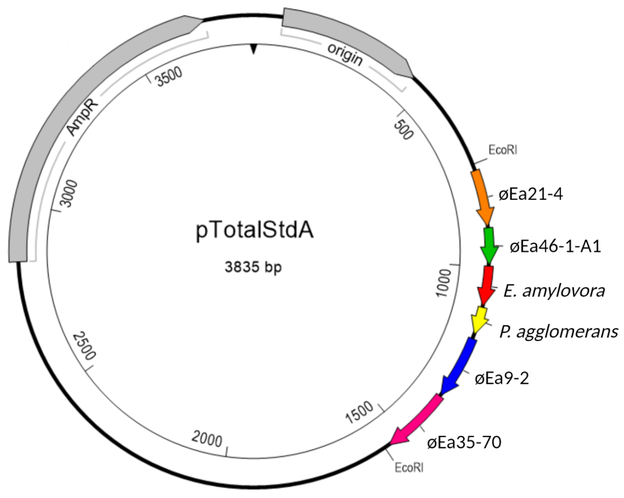
Erwinia amylovora is a pathogenic bacterium of apples, pears and other Rosaceous plants, which causes the disease fire blight. Phages targeting this pathogen are being investigated for their application to apple and pear orchards for the control of fire blight. Phages infecting *E. amylovora* tend to have broad host ranges and are also able to infect other epiphytic species in the *Erwinia* and *Pantoea* genera.

We designed a plasmid DNA standard for qPCR quantification of phages in solution. Using this we performed a quantitative host range analysis of 4 different *Erwinia* phages against 80 sequenced strains of *Erwinia* and *Pantoea* species by measuring the amount of each phage produced on each host after 8 h. A genome-wide association study (GWAS) was performed for each phage to identify host genetic variants associated with variations in phage replication. These genetic variants exist in environmental populations and may represent mutations which evolved naturally in response to phage predation. Therefore, the genes identified from this work may be indicative of novel host mechanisms of phage resistance and phage infection methods.

Methods

qPCR for Phage Quantification

The plasmid standard pTotalStdA (Gayder et al., 2019) contains qPCR amplicons targeting the four phages ϕEa21-4, ϕEa46-1-A1, ϕEa9-2, and ϕEa35-70, as well as the pathogen *E. amylovora* and the epiphytic *P. agglomerans*. A dilution series of pTotalStdA at known copy numbers from 10^{11} to 10^{5} copies/mL was used to quantify phage genomes using qPCR.



Quantitative Phage Host Range

A quantitative host range analysis was performed using 4 phages each against a sequenced collection of 50 *E. amylovora* and 30 *Erwinia* and *Pantoea* spp. epiphytes. In 1 mL cultures in nutrient broth, each host at 10^{8} CFU/mL was infected with 10^{4} genomes/mL (*Erwinia*) or 10^{6} genomes/mL (*Pantoea*) and grown 8 h at 27°C, 200 rpm. qPCR was used to quantify the amount of phages produced.

GWAS Identification of Host Genes Affecting Phage Production

Genomic sequences were acquired for all 80 strains tested, including the reference genome of *E. amylovora* strain Ea273. This reference genome was split into 19 bp k-mers and each of these k-mers were compared against all 79 other genomes as being present or absent, using kSNP3 (Gardner et al., 2015). This binary data was used to test every k-mer for significant association with production of each of the phages, using TreeWas (Collins and Didelot, 2018). The genes containing the significant k-mers in the Ea273 reference genome were then determined and identified as significantly associated with phage production.

Results

Quantitative Phage Host Range

The quantitative host range data of the four phages show that each have broad host ranges and are able to infect strains in more than one genus (Fig. 1). However, the distinct differences in host preference among related species is likely due to host genetic variations.

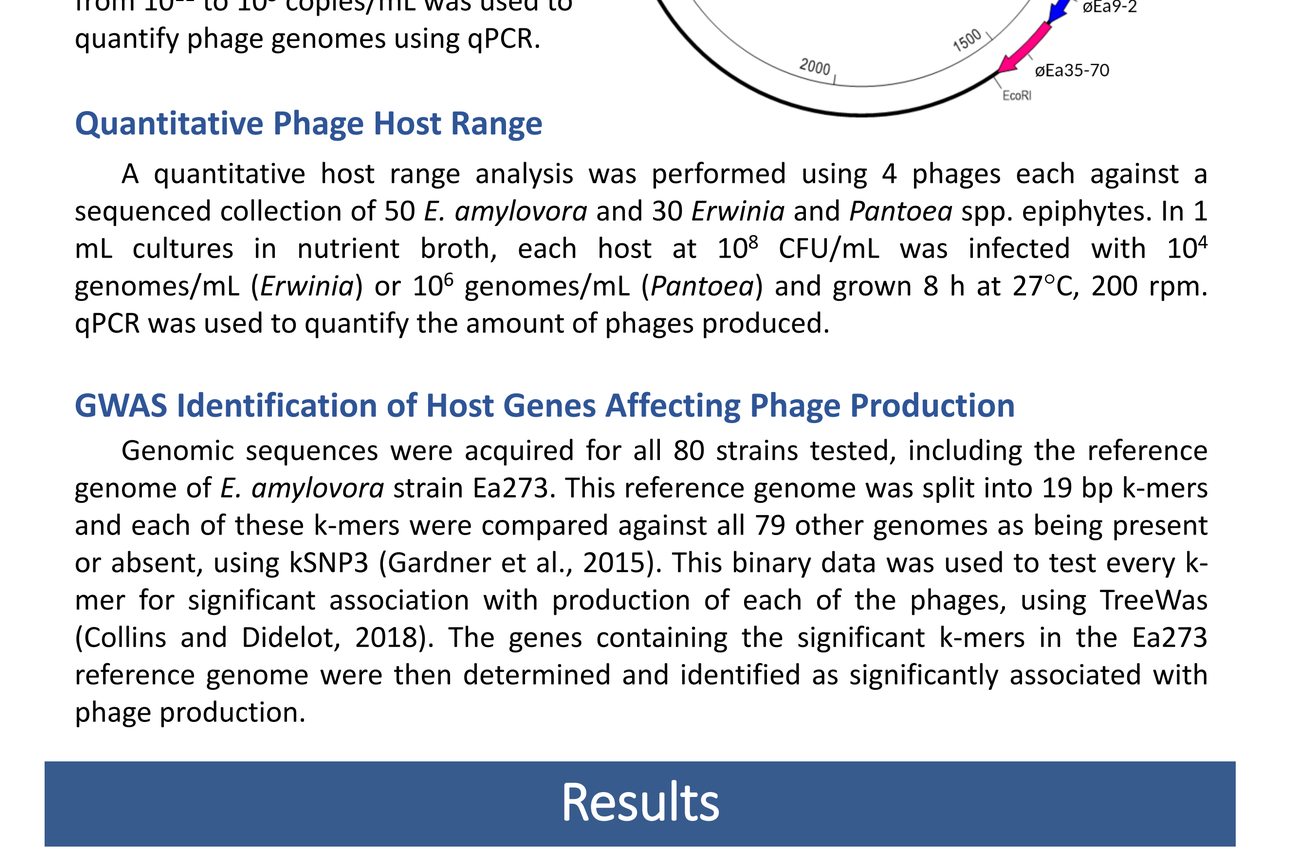


Figure 1. Heatmap of normalized phage production from no phage growth (blue) to maximal (yellow; 7.5x10^{11} genomes/mL) on all 80 tested strains. Phylogeny was produced by kSNP3 (Gardner et al., 2015).

Results

GWAS Identification of Host Genes Affecting Phage Production

There were 15 k-mers identified as significantly associated with production of one of the four phages (Fig. 2), and these k-mers were located within 10 genes in the Ea273 genome (Fig. 3).

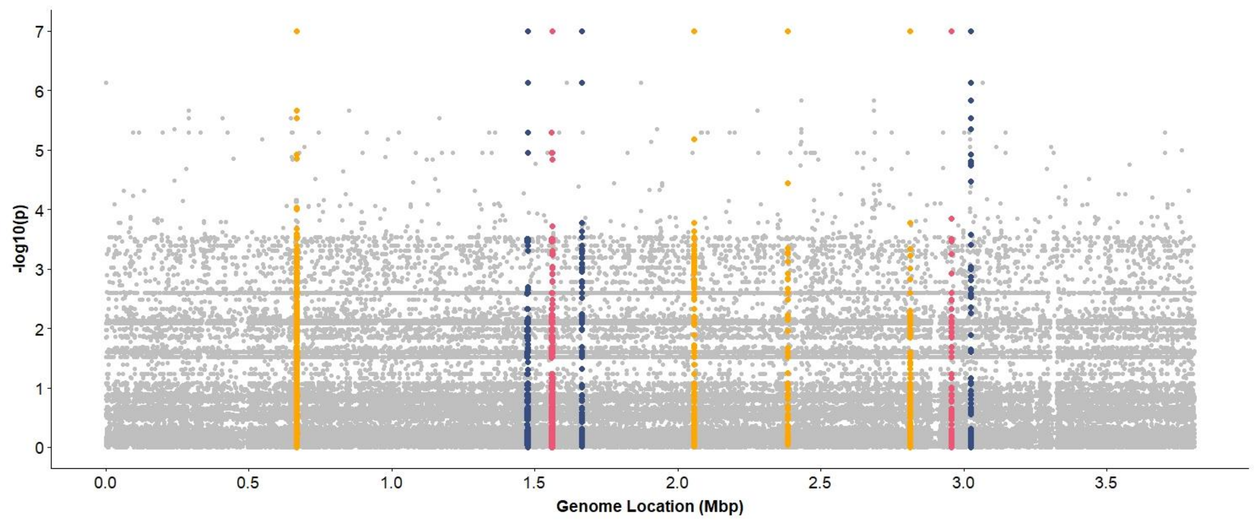


Figure 2. Manhattan plot of k-mers associated with phage infection. The –log10 p-values from these tests are plotted at their corresponding location in the Ea273 reference genome. The coloured dots represent k-mers, or loci, within the identified genes, where orange, blue, and magenta represent phages ϕEa21-4, ϕEa9-2, and ϕEa35-70 respectively. Grey dots represent k-mers that were not significantly associated with phage production, and only a random 5% of the total are shown to prevent oversaturation of the figure.

Conclusions

Using this approach, we identified 10 host genes which may have evolved naturally to avoid phage predation (Fig. 3), and novel hypotheses can be formed about the interaction of these phages with their hosts based on their potential functions (Table 1). Most notably, VgrG and CdiA have similar roles within different secretion systems and are responsible for 'loading' toxic effectors and delivering them to neighboring cells, suggesting these secretion systems (Type VI and Type V respectively) may also function as phage resistance mechanisms. Additionally, ϕEa21-4 may initially bind to the flagella and reach the cell surface through the direction-dependent rotation of the flagella. Several other host processes, such as peptidoglycan structure, protein secretion, and cell metabolism are also possibly used by the host to affect phage infection.

Further work is required to confirm these hypotheses, and currently efforts are being made to generate and study knockouts of these genes.

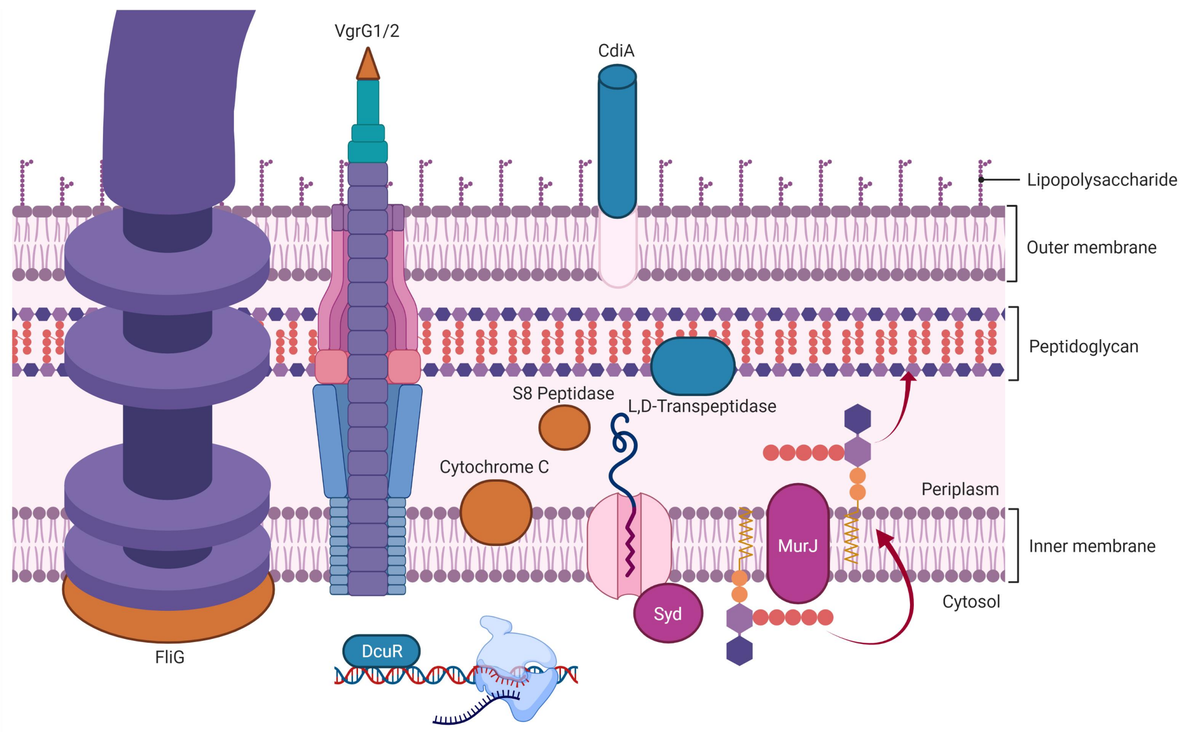


Figure 3. Diagram of cellular locations of the proteins associated with phage host range. The proteins are labelled and colour coded by phage where orange, blue, and purple are the genes associated with ϕEa21-4, ϕEa9-2, and ϕEa35-70 respectively.

Table 1. Host genes identified as significantly associated with phage infection, and their potential functions

Gene	Function
FliG	Part of the flagellar motor switch, determines rotation direction
VgrG	Spike protein complex in the T6SS. Binds with toxic effector to inject in nearby cells
Cytochrome C	Transports electrons across the membrane as part of the electron transport chain
S8 Peptidase	Subtilisin-like peptidase
DcuR	Transcriptional regulator that controls the expression of C_4-dicarboxylate metabolism
CdiA	Toxic effector component of the T5SS, contact-dependent inhibition system
L,D-Transpeptidase	Creates D-Ala^4→DAP^3 cross-linkages in the peptidoglycan
Syd	Associates with the SecY protein transport channel and is contributes to its stability
MurJ	Flippase, translocates the peptidoglycan precursor lipid II across the inner membrane.

References

- Collins, C., and Didelot, X. (2018). *PLoS Computational Biology* 14, e1005958.
- Gardner, S.N., Slezak, T., and Hall, B.G. (2015). *Bioinformatics* 31, 2877-2878.
- Gayder, S., Parcey, M., Castle, A.J., and Svircev, A.M. (2019). *Viruses* 11, 910.
- Gayder, S., Parcey, M., Nesbitt, D., Castle, A.S., and Svircev, A.M. (2020). *Microorganisms* 8, 1449.
- Parcey, M., Gayder, S., Morley-Senkler, V., Bakkeren, G., Urbez-Torres, J.R., Ali, S., Castle, A.J., and Svircev, A.M. (2020). *Genomics* 112, 3762-3772.