Starts and Stops: Identification of conserved sequences in subcluster A11 and cluster K mycobacteriophages associated with gene expression.

Bryan A. Pasqualucci, Elena M. Haury, Maurice Newton, Kara Jones, Dathan Stone, Liam J. McNally

Nicholas P. Edgington, Ph.D., Department of Biology
Southern Connecticut State University

Introduction

- SCSU Genomics I students isolated 7 mycobacteriophages that infect *M. smegmatis* mc^2 155 at 42°C.
- Three of these genomes were sequenced which revealed their identity in the A11 (Lucivia and Flaverint, joining Snape which was sequenced at SCSU in 2014) and K4 (Wintermute) subclusters.
- Wintermute, the sixth phage in the K4 subcluster, is part of a cluster of phages that is capable of also infecting *M. tuberculosis* (*Mtb*). Fionnbharth is Wintermute's closest relative and can infect *Mtb* 30 times more effectively than it can *M. smeg*.
- Bioinformatic analyses of Lucivia demonstrate how subcluster A11 phages are capable of forming lysogens without the phage and bacteria *attP/B* cassette and an integrase. Analyses may also show the presence of stoperators which are unique to cluster A phages.

The Issues

- The investigation of the unique genomes of cluster K and A phages allows us to gain insight on how they can infect their hosts. Each cluster has characteristics that are unique to that cluster, and investigating them using bioinformatics may show us how different clusters are capable of infecting bacteria.
- Understanding how certain clusters of mycobacteriophages can also infect *Mtb* has proven to be difficult to investigate bioinformatically, but these analyses have the potential to lead to wet lab experiments that will hopefully provide definitive answers.

Results/Discussion

- **Phage Hunting at SCSU**
  - 3 phages sequenced at SCSU in 2015
    - Lucivia (A11)
    - Flaverint (A11)
    - Wintermute (K4)

- **The K Cluster**
  - Now includes newly sequenced Wintermute (K4)
  - 75 Members in 6 subclusters
  - Several phage shown to have *M.smeg/Mtb* infectivity

- **Start-Associated Sequences (SAS)**
  - 13 bp motif found in all and only K cluster phages (doi:10.1371/journal.pone.0026750)
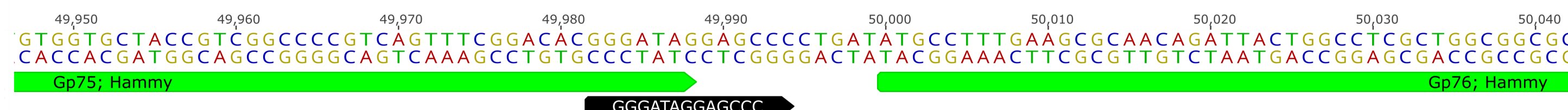 - Thus far, SAS's are located in the more variable "right"-side of the genomes.

Figure 1. . Subcluster K4 is most closely related to subclusters K2 & K5. *Phylogenetic tree of cluster K phages created using Splitstree of aligned whole genome sequences from ClustalO.*
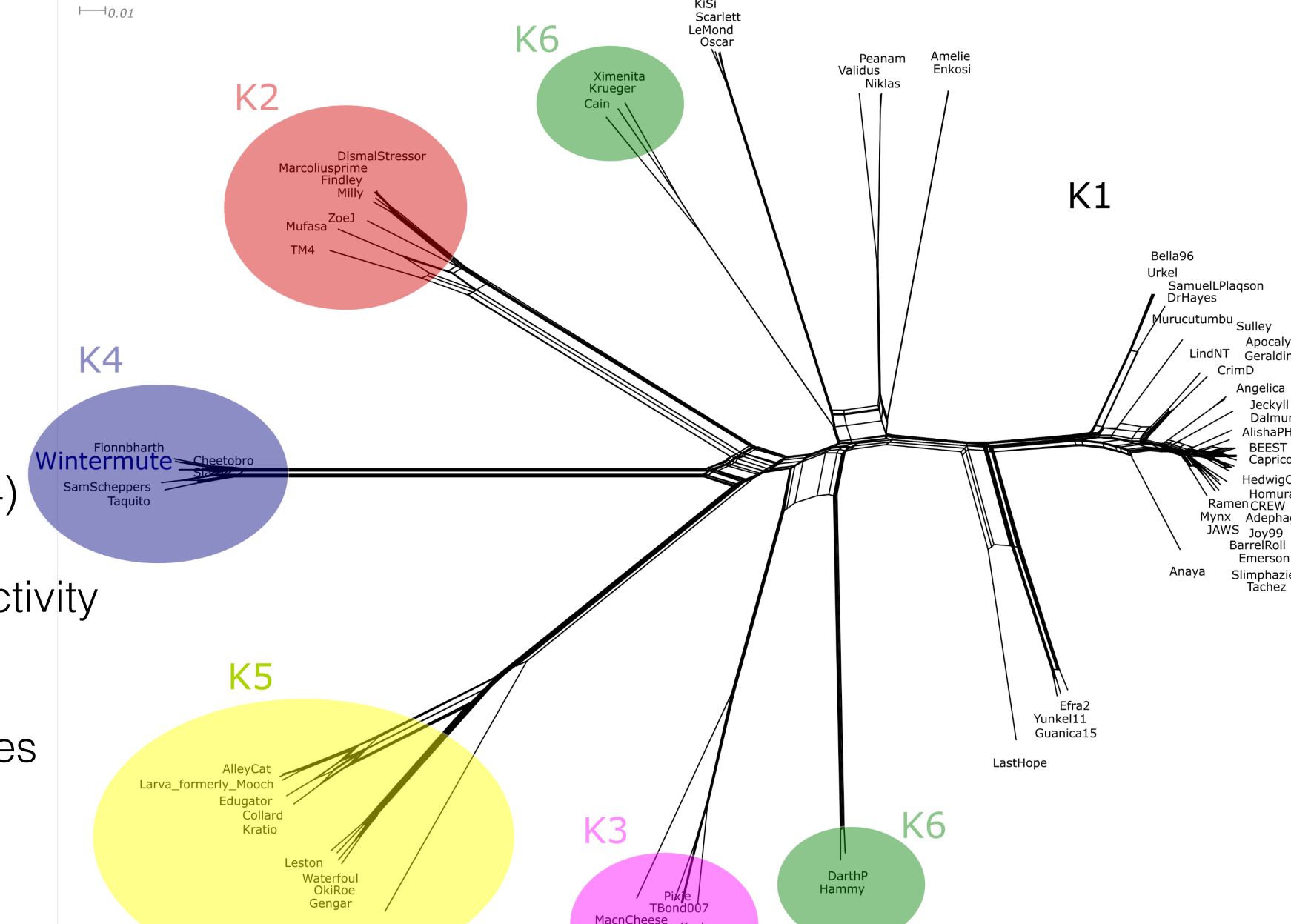
Wintermute (K4): General Intergenic SAS; 692 instances
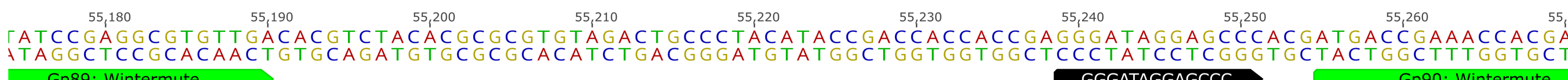
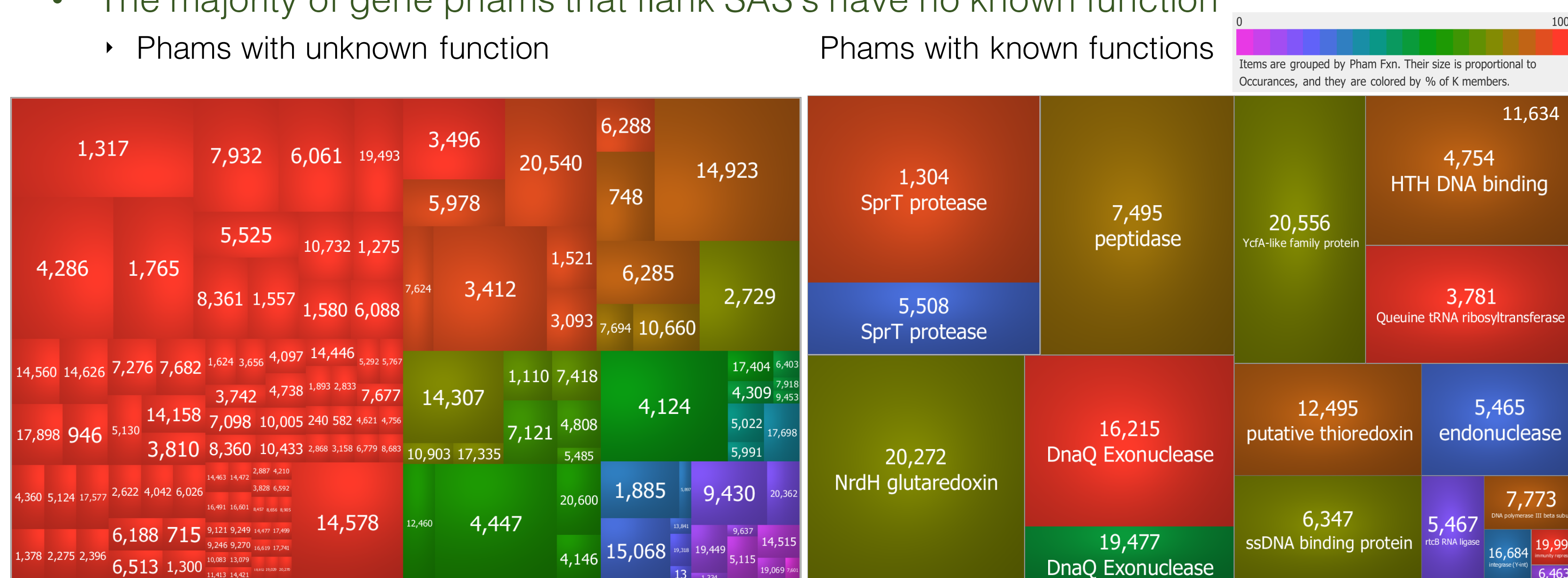Hammy (K6): SAS exhibits 7bp overlap; 143 instances

ZoeJ (K2): SAS 54 bp from ORF start; 8 instances

Figure 2. Different types of SAS found in K4, K6, and K2 phages found in different regions in each genome.

- **The majority of gene phams that flank SAS's have no known function**
  - Phams with unknown function
 - Phams with known functions

Figure 3. Size denotes number of pham genes associated with SAS. Color denotes the number of SAS associated genes in a pham divided by the total number of cluster K genes in that pham as a percentage. *Generated with TreeMap by Macrofocus.*

- **Are there phams that correlate and predict *Mtb* infectivity?**
  - There are no phams that only have membership in K, G, A2, and A3 sub/clusters.
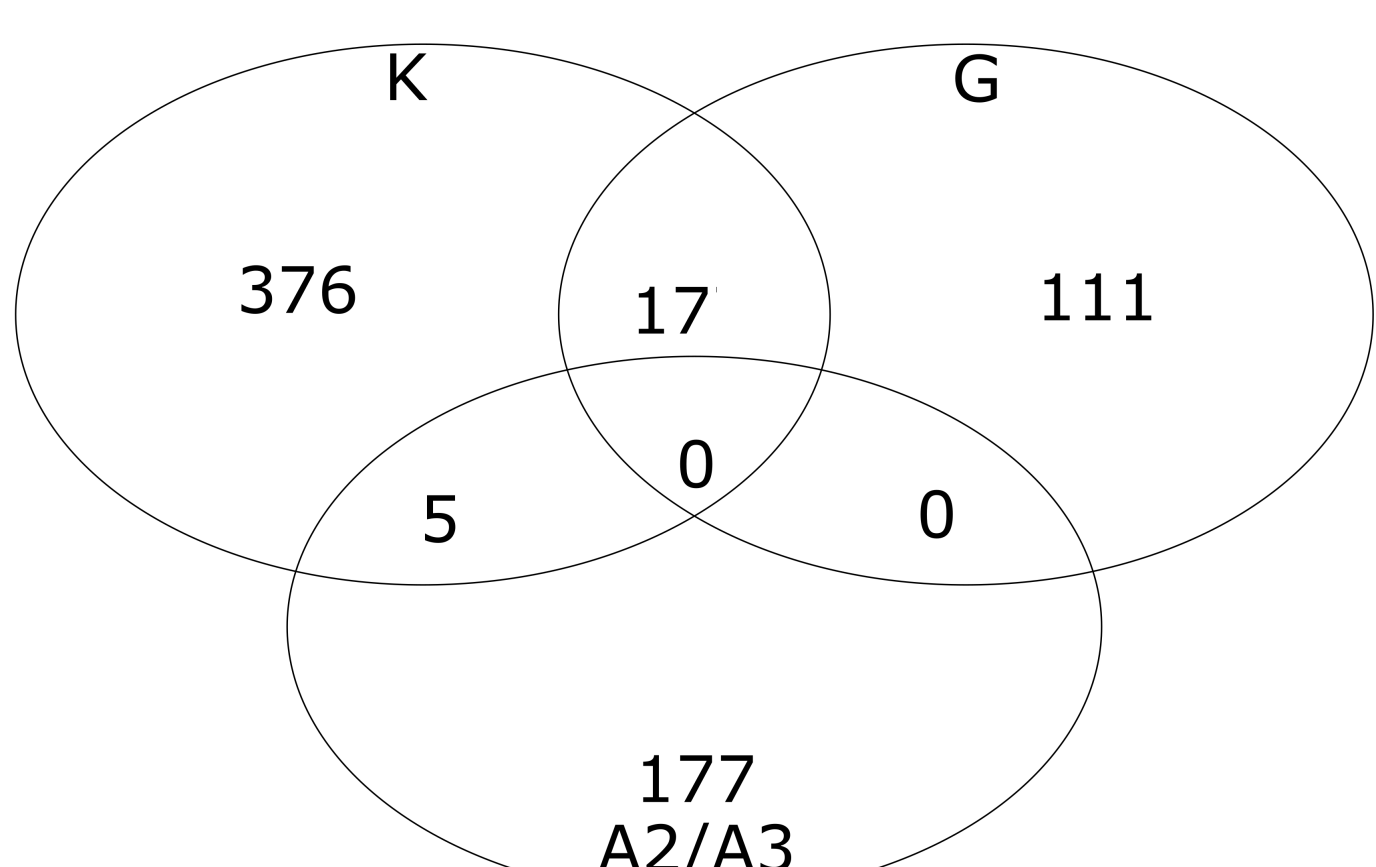
Figure 4. Venn plot of Pham Uniqueness in K, G cluster and A2,A3 subclusters. *Generated in the program "R".*
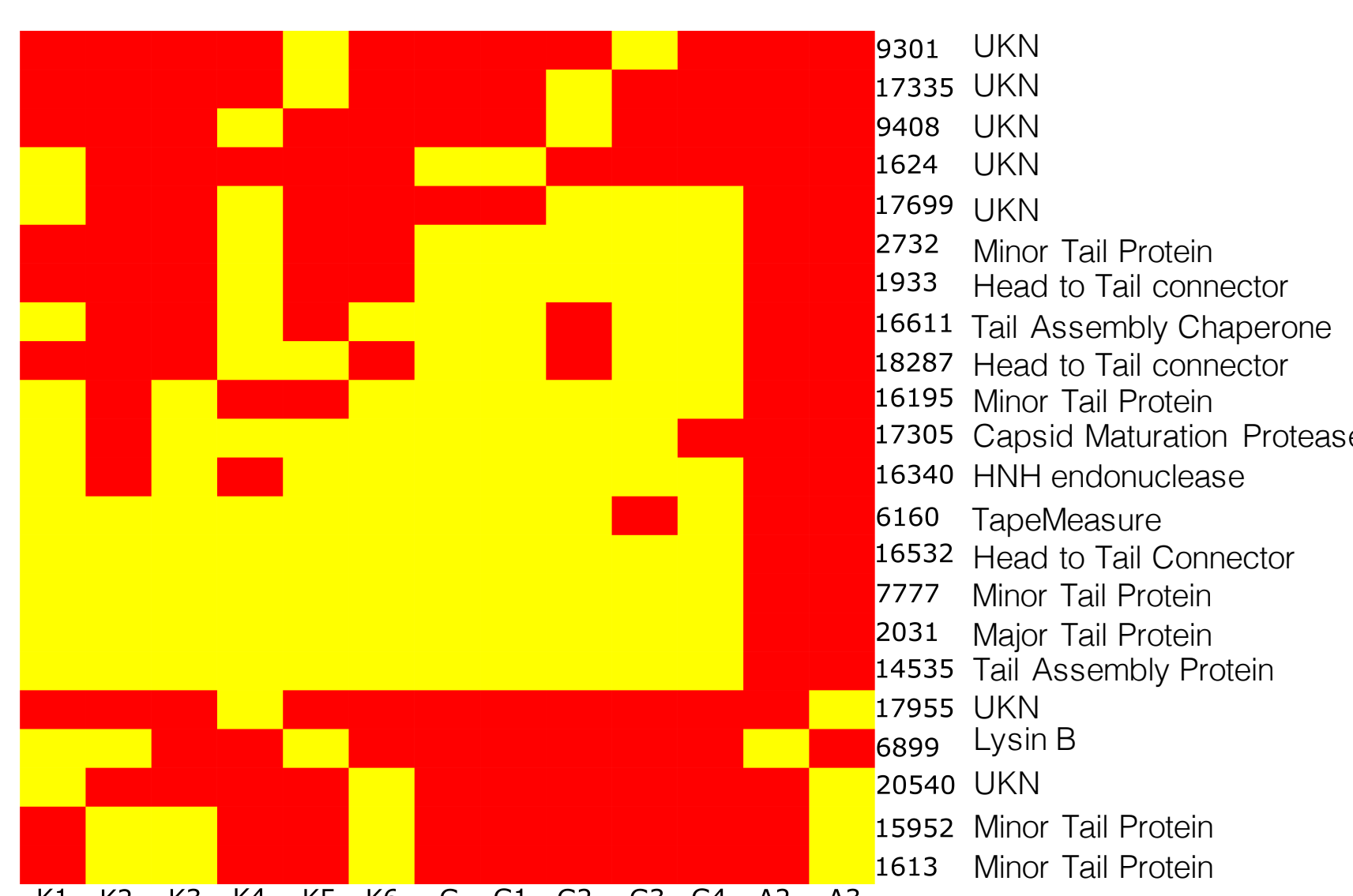
Figure 5. Distribution of shared unique phams by subcluster. Yellow denotes pham presence, and red denotes absence. *Generated in the program "R".*

- **The A11 subcluster**
  - 9 A11 subcluster phages: 3 found at SCSU
    - Lucivia (2015)
    - Flaverint (2015)
    - Snape (2014)
  - Discovery date range = 2012-2015
  - Ave. length = 52,278 bp
  - Ave. GC% = 63.7%

- **There are at least three different ways in which a phage can infect a bacterial cell**
  - Lytic cycle or lysogeny, or pseudolysogeny
  - Lysogeny requires an integration cassette so that the phage genome can integrate into the host genome. Most cluster A phages have this integration cassette, but a small number do not.

- **Subcluster A11 phages do not have an integrase gene!**
  - *ParABS* function
 - Phages with parABS systems are replicated extrachromosomally instead of integrating into the genome.
  - *ParABS* in A11s
    - In A11 genomes, *ParA* genes belong to pham 7133(66) and *ParB* genes belong to pham 814(38). Both of these phams only exist in cluster A phages.
    - Not all cluster A phages with a *ParA* gene also have a *ParB* gene, but all phages with a *ParB* gene also have a *ParA* gene.
    - *ParABS* systems may coevolve due to evolutionary pressure (doi: 10.1111/mmi.13414).
    - To analyze this in the A11 phams, we created phylogenetic trees from protein multiple sequence alignments to visualize phage relationships and how they differ.

Figure 6. Phylogeny of the subcluster A11 phages. Lucivia and Flaverint are most closely related to each other. *Heatmap created using 'pyani'.*
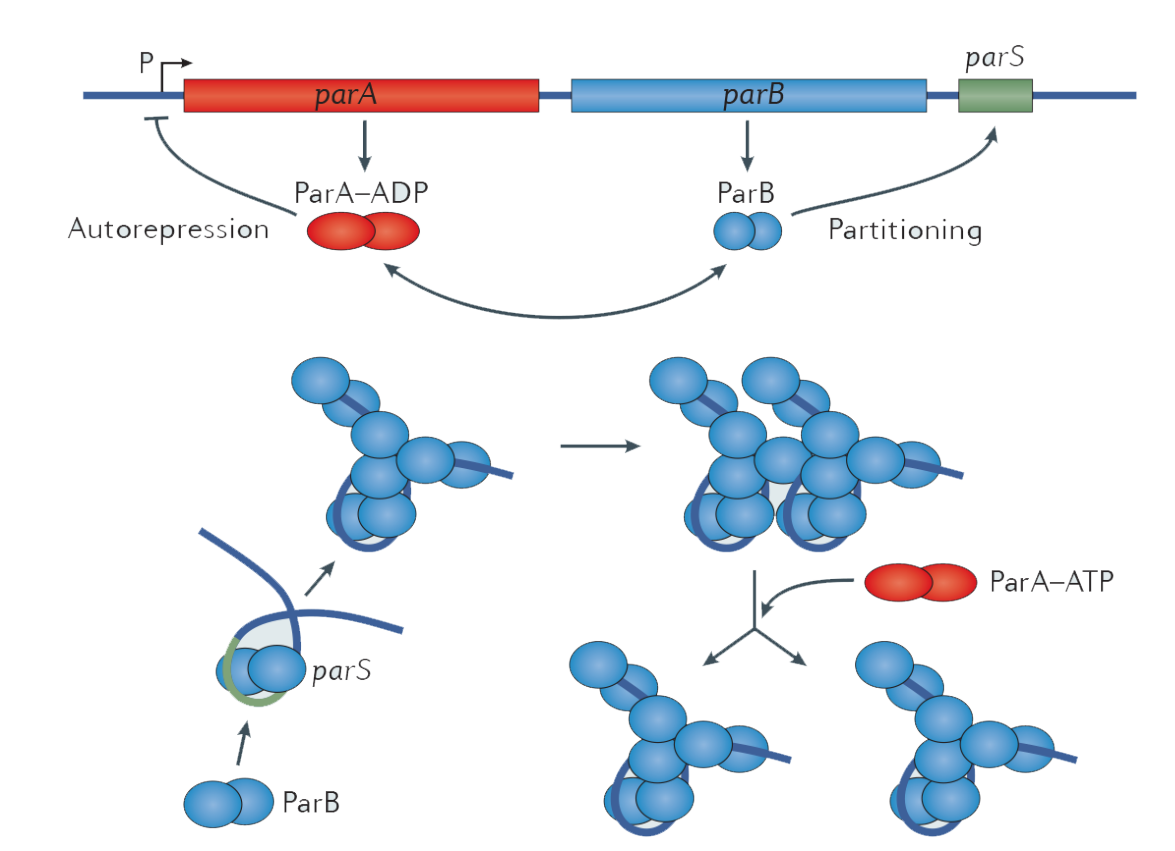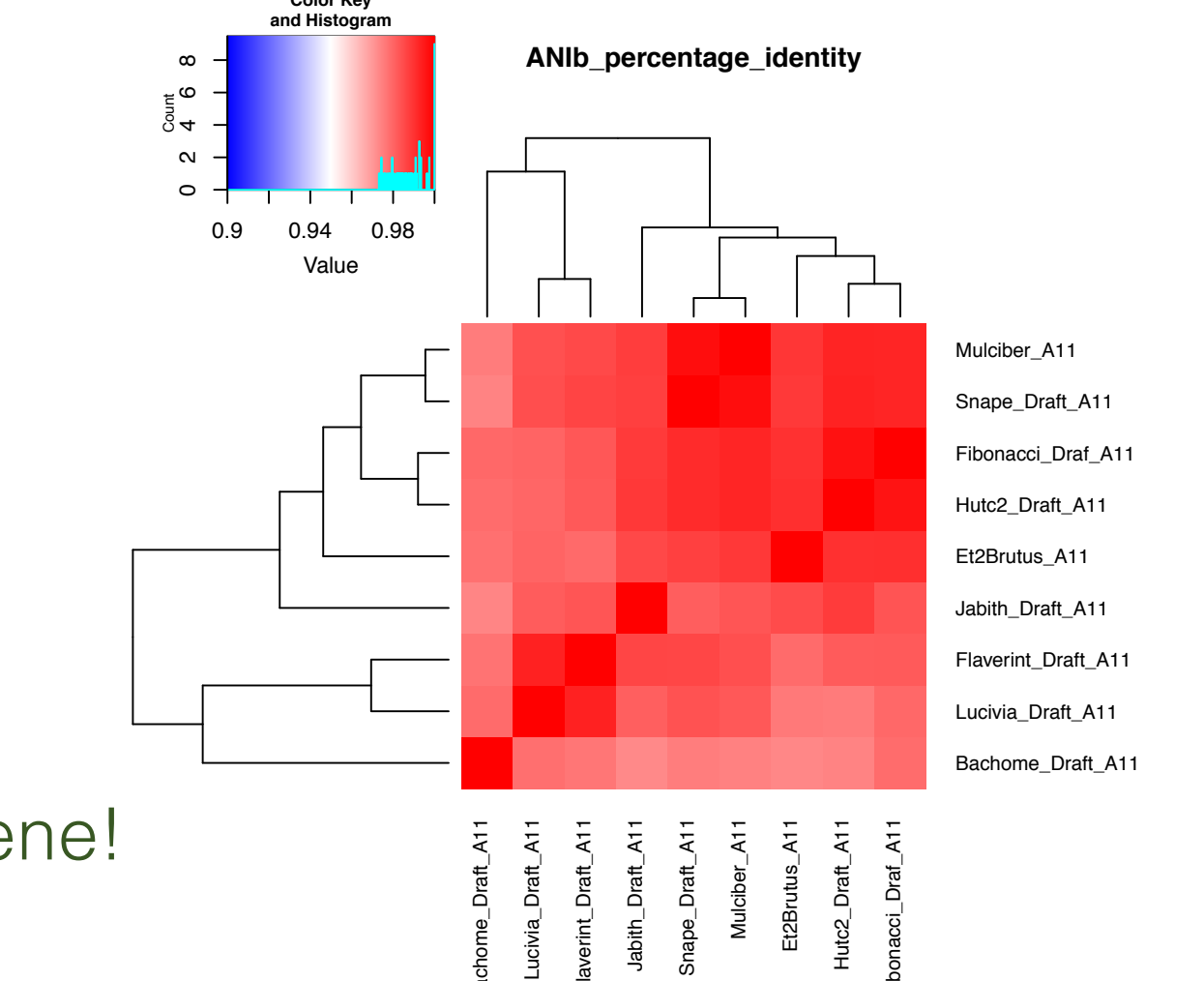
Figure 7. Function of the *ParABS* system (doi:10.1038/nrmicro2882).
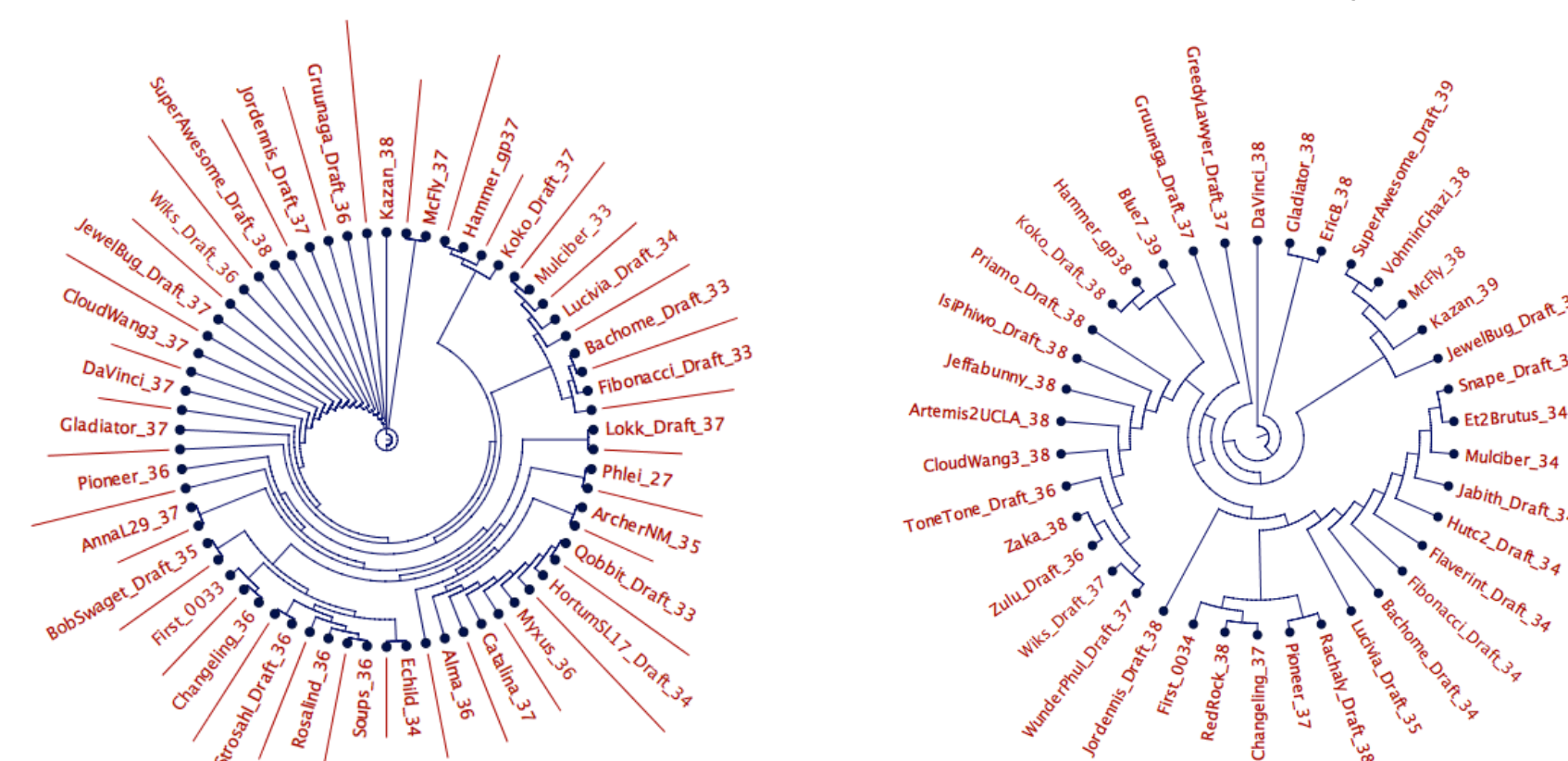
Figure 8. Co-evolution of cluster A *parA/B* gene 'phamilies' *Created in CLC Sequence Viewer.*

- **Stoperators: immunity systems in cluster A phages**
  - Short repeating genetic elements that control lysogeny by preventing transcription in genes that are not responsible for integrating the phage into the host genome.
  - First found in L5 (A2) and Bxb1 (A1).
    - 20 copies in Lucivia
    - They are 13 bp in length, and are found mostly in intergenic regions

- **Are stoperator sequences present in A11 phages?**
  - To search for stoperator sequences, we used MEME Suite to search for repeating motifs that occur in Lucivia's genome.
  - The repeating sequence we discovered in Lucivia contains the same consensus sequence as the stoperator sequences in L5.

Figure 9. Consensus stoperator sequences found in A cluster phages L5 and Bxb1, with differences in Bxb1 underlined (doi:10.1093/emboj/16.19.5914).

| L5 (A2) | 5'-GGTGGMTGTCAAG |
| Bxb1 (A1) | 5'-GTTACGWDTCAAG |
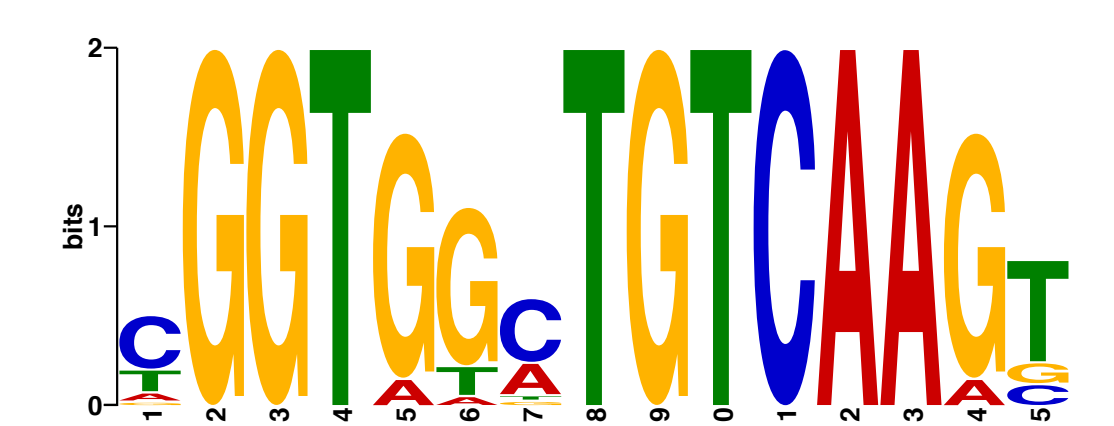
Figure 10. Predicted stoperator concensus sequence of the A11 phage Lucivia. *Created using MEME Suite.*

Conclusions

- We discovered and performed analyses on two new phages in subclusters K4 and A11 and found interesting features exclusive to those subclusters.
- SAS's exhibit three general types, intergenic, 7bp overlap and 54bp from a gene start. The function of these sequences has yet to be determined and their pham association is widely varied.
- There is no clear consensus between the phams of Clusters A2/A3, G and K who have shown the ability to infect *Mtb*.
- Subcluster A11 phages all contain *ParABS* systems and replicate extrachromosomally as opposed to integrase proteins. Cluster A phages also contain stoperator sequences, which have yet to be analyzed in subcluster A11 phages, but were identified in Lucivia with a short identical nucleotide match to the stoperator sequences found in A1 and A2 phages, using bioinformatics software.

Future Studies

- Analysis of the SAS function and determine their position relative to RBS sites.
- Wet lab experiments analyzing the efficiency Wintermute has to infect *Mtb* (done by others).
- Analysis of of the individual proteins in the phams shared by *Mtb* infective phage.
- Further analysis on stoperators in A11.

Acknowledgements