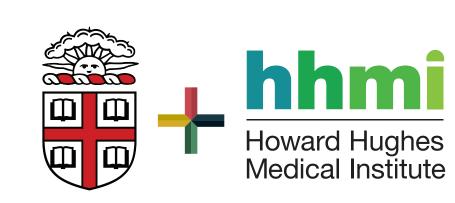
Characterization of palindrome usage in mycobacteriophage genomes

Julia Gross, Kenyon Alexander, Brandon Le, Zachary Ricca, Hwai-Ray Tung, Benjamin Siranosian, Yang Zhou, Sarah Taylor, Christopher de Graffenried Brown University, Providence RI



BACKGROUND

Palindromes: DNA

sequences with identical reverse complements

GATGATCCGT

The reverse complement of GATC is GATC!

In double stranded DNA, palindromic sequences are often recognized and cleaved by restriction enzymes.



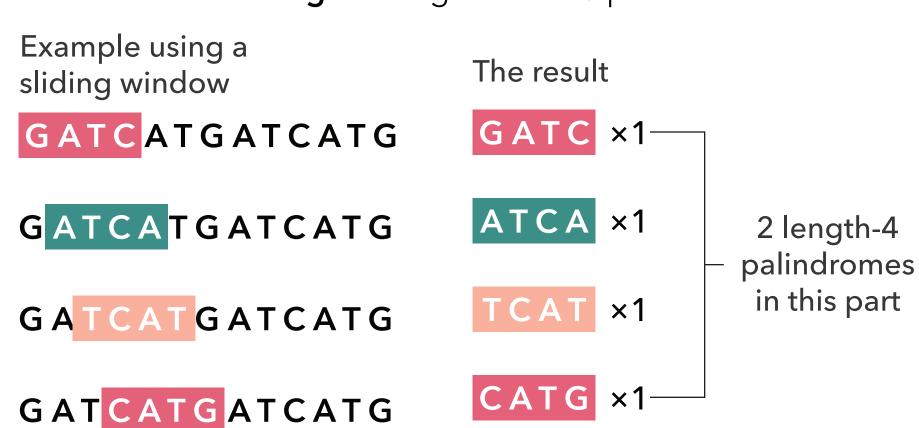
To prevent cleavage, bacteria and phage genomes typically avoid using palindromic sequences.



We characterized the presence of palindromic sequences of length 4 and 6 in mycobacteriophage genomes to look for global trends and specific interesting examples.

METHODS

Palindrome counting: all length-4 and 6 palindromes



Normalization of counts

Palindrome occurrence depends on the underlying nucleotide counts in the genome. To normalize, we computed the expected number of a palindrome w in a genome with identical nucleotide composition but randomized sequence order:

$$Exp(w) = [(A^a * C^c * G^g * T^t) * N]$$
 A, C, G, T : genomic frequency of respective nucleotides a, c, g, t : tetranucleotide frequency of nucleotides N : length of genome

The normalized count is the observed count divided by the expected count:

$$TUD(w) = \frac{observed}{expected}$$

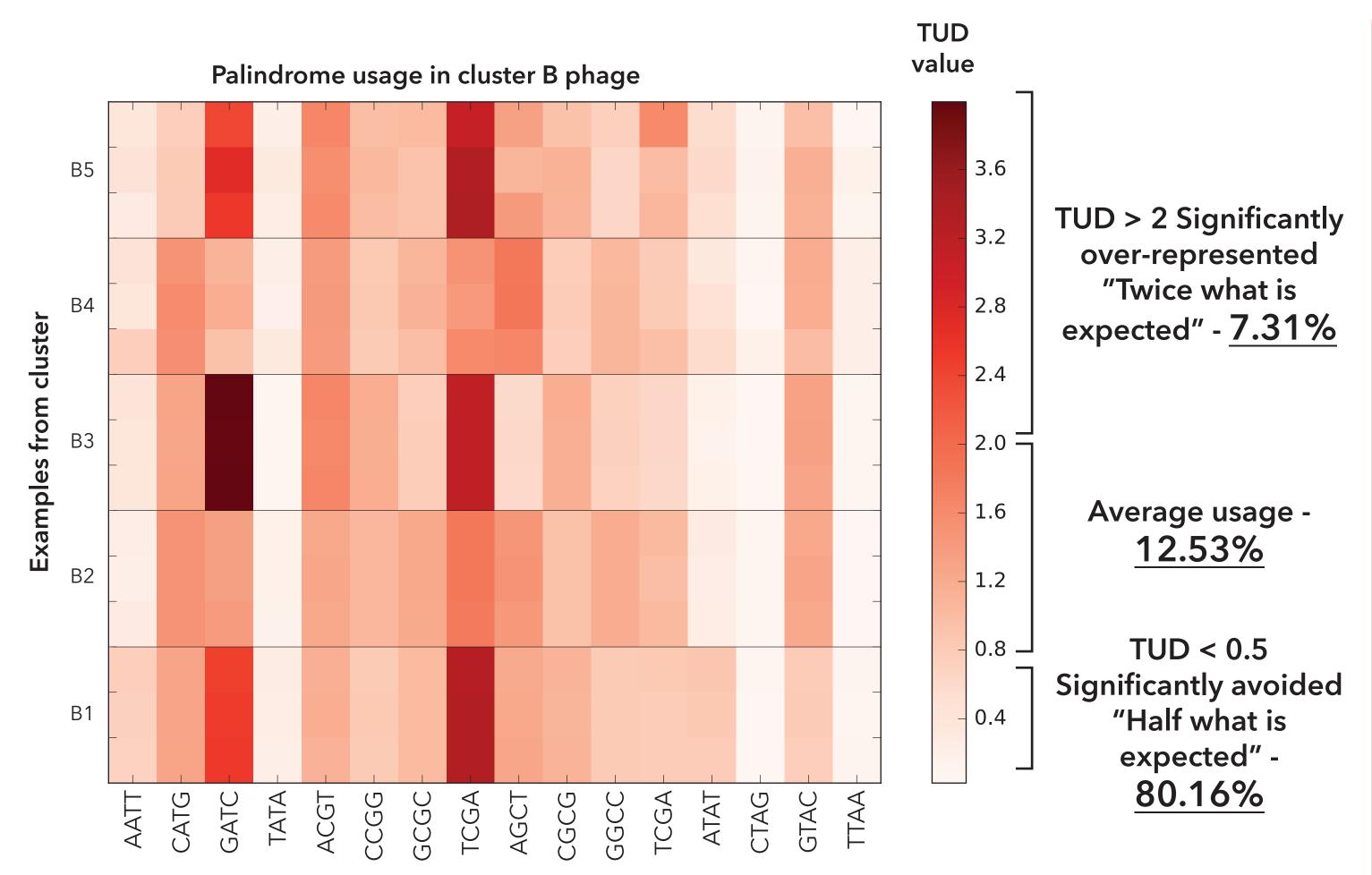
Tetranucleotide

TUD = Usage
Deviation

This removes biases from GC count and highlights interesting patterns of palindrome occurrences. One TUD value for every possible palindrome. We calculated this in all mycobacteriophage genomes.

PALINDROME Q&A

Q1: ARE PALINDROMES GLOBALLY AVOIDED IN MYCOBACTERIOPHAGE GENOMES?



R1: Heterogeneous usage between clusters

The vast majority of palindromes in most phages are underrepresented. However, there are a small number of striking deviations from this pattern -- both palindromes that do not occur at all in certain phages, and ones that occur at many times the expected rate. Overall, palindrome usage is heterogeneous, but similar within phage clusters.

Q2: WHAT INFLUENCES PALINDROME OCCURRENCE IN PHAGES?



Restriction enzymes can cut palindromes in double-stranded DNA. Are there systems that recognize palindromes in Mycobacteria?

R2: Lack of restriction enzymes in M. smegmatis



We searched the restriction enzyme database REBASE for information on restriction/modification and methylase systems in Mycobacteria.

We found:

- M. Smegmatis does not encode for any restriction enzymes
- Mgol in M. Gordonae recognizes GATC —
- A methylase in M. Smegmatis recognizes GATC —

Do these explain the high abundance of GATC in B3 phages?

Q3: What palindromes are underrepresented?

Some palindromes are used infrequently or not at all in some genomes.

R3: Certain adenine-thymine palindromes in A-cluster phages





Not in any A4 genomeNot in half of A2 genomes



- Not in 5 A2 genomes

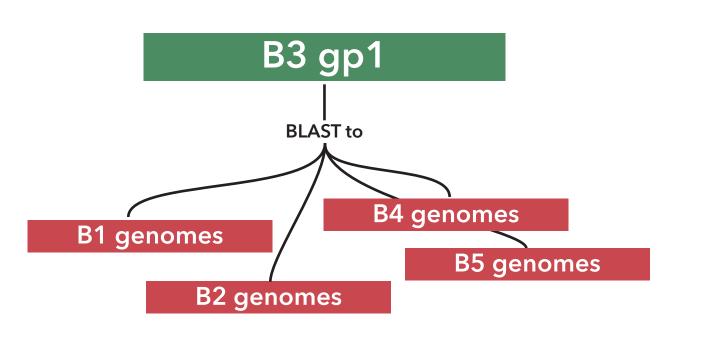
Probability of this happening by random chance alone

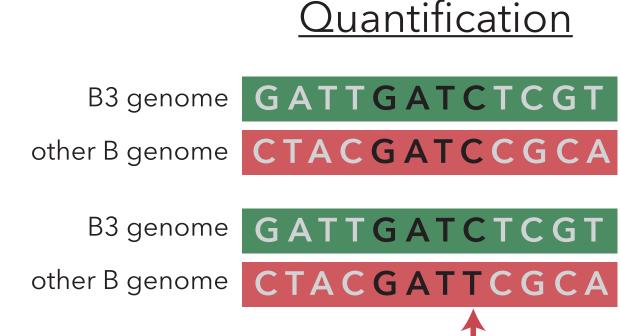
10⁻²³ -or- 0.00000000000000000000001

Conclusion: Something is suppressing ATAT and AATT usage in cluster A phages. Maybe they are targets for currently unknown restriction enzymes in the native host bacteria of cluster A phages?

Q4: How did GATC BECOME SO FREQUENT IN B3 GENOMES?

GATC is used 4 times more frequently than expected in B3 genomes. We investigated if consistent single nucleotide polymorphisms (SNPs) lead to an abundance of this palindrome.

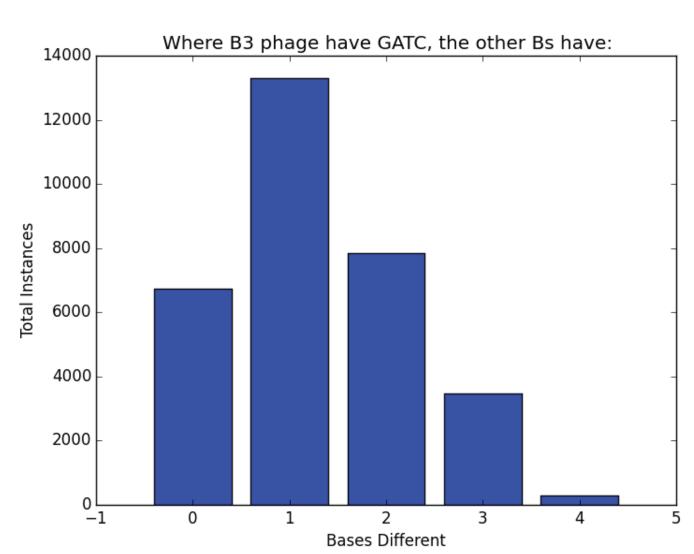


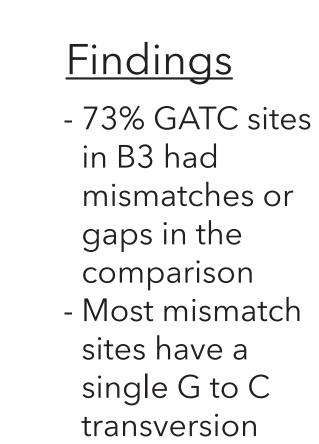


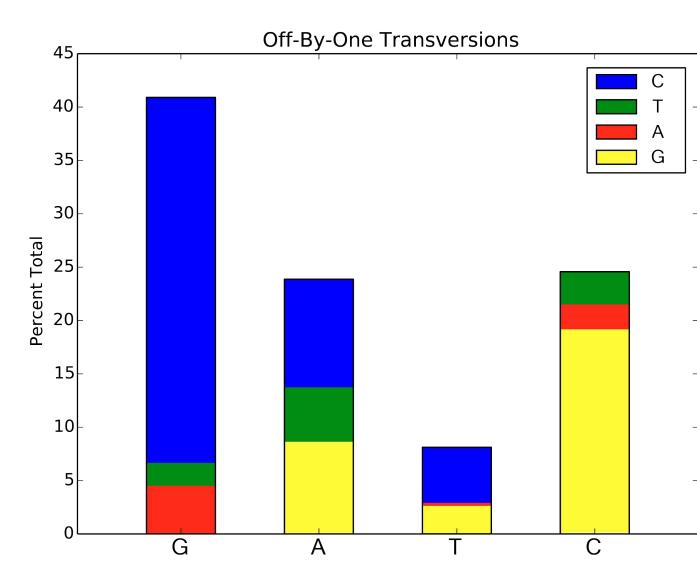
GCA GATC matches

CGT Mismatches, by type

R4: INTERESTING MUTATION PATTERNS







Conclusions

- B3 phages maintain palindromic GATC sites - Related phages commonly have NATC

Acknowledgments

We would like to thank

- The other instructor of the Brown University Phage

Hunters, Peter Shank

- The MDL lab staff at Brown, and HHMI SEA-PHAGES for providing us with the opportunity to conduct this research - Chen Ye for designing some of the graphics.

References

1. Gelfand, M. S. & Koonin, E. V. Avoidance of palindromic words in bacterial and archaeal genomes: a close connection with restriction enzymes. Nucl. Acids Res. 25, 2430-2439 (1997).

2. Hemavathy, K. C. & Nagaraja, V. DNA methylation in mycobacteria: absence of methylation at GATC (Dam) and CCA/TGG (Dcm) sequences. FEMS Immunol. Med. Microbiol. 11, 291-296 (1995).

of restriction sites in DNA sequences. Nucleic Acids Res 20, 1363–1370 (1992).

4. Rocha, E. P. C., Danchin, A. & Viari, A. Evolutionary Role of Restriction/Modification Systems as Revealed by Comparative Genome Analysis. Genome Res. 11, 946–958 (2001). 5. Rocha, E. P. C., Viari, A. & Danchin, A. Oligonucleotide bias in Bacillus subtilis: General trends and taxonomic comparisons. Nucl. Acids Res. 26, 2971–2980 (1998).

6. Shankar, S. & Tyagi, A. K. Purification and characterization of restriction endonuclease Mgol from Mycobacterium gordonae. Gene 131, 153–154 (1993).

7. Shankar, S. & Tyagi, A. K. MchAl and MchAll, two class-II restriction endonucleases

3. Karlin, S., Burge, C. & Campbell, A. M. Statistical analyses of counts and distributions

from Mycobacterium chelonei. Gene 132, 119-123 (1993). 8. Sharp, P. M. Molecular evolution of bacteriophages: evidence of selection against the recognition sites of host restriction enzymes. Mol Biol Evol 3, 75-83 (1986).