Six Nucleotide Palindrome Occurrence in Mycobacteriophages

Catherine Halpern

Introduction:

Bacteria produce restriction enzymes, capable of cutting DNA at a specific site and stopping the replication of viruses that infect the bacteria. Most of these restriction enzymes recognize short palindromic sequences, between four and eight nucleotides long, as the sites near or at which they bind [1]. In order to protect themselves from these enzymes, some viruses have evolved to avoid these short palindromes, and they do not display large numbers of them in their genome [2]

Mycobacteriophages are viruses that infect bacteria of the family Mycobacteriaceae. Two members of the Mycobacteriacae family, Mycobacterium tuberculosis and Mycobacterium leprae, were analyzed and shown not to avoid palindromes of any length [3]. According to the analysis done by Rocha's team of scientists, phages that infect bacteria that do not avoid palindromes also do not avoid palindromes. However, they found an exception to this in two Mycobacteriophages, L5 and D29, both of which avoided palindromes of length six [3]. Sequencing of M. tuberculosis and M. leprae did not reveal restriction modification systems that would explain this [4] but some other Mycobacteria infected by L5 and D29 do have these systems that could explain this avoidance [5].

It was also of note that only two Mycobacteriophages were examined. At this point, several hundred have been sequenced and are available in GenBank for analysis [6]. More can be learned about these phages by examining if this trend continues in all Mycobacteriophages, or if was merely an anomaly in the two genomes looked at by Rocha et. al [3].

Methods:

As other researchers have used before when coming up with values for expected occurrences of a nucleotide sequence [7]. a Markov model was used to calculate the number of times all six nucleotide palindromic sequences should occur in each Mycobacteriophage genome in BioBIKE [8]. This model counts the number of times the first five nucleotides of each six nucleotide palindrome occur in sequence in a given genome, counts the number of times the last five nucleotides of each palindrome occur in sequence, multiplies these numbers and divides by the number of times the middle four nucleotides of the palindrome occur in sequence.

$$Expected = \frac{Occurrence_{nt1\ thru\ nt5} \times Occurrence_{nt2\ thru\ nt6}}{Occurrence_{nt2\ thru\ nt5}}$$

This gets an expected value for the number of occurrences for six nucleotide palindromes within the sequence, taking into account how often subsequences occur, as it is not remarkable if a six nucleotide sequence occurs often and the five nucleotide sequence it extends occurs almost equally as often.

In order to calculate the actual number of 6 nucleotide palindromes in a given genome, BioBIKE's "matches of pattern" function was used, with each genome being searched for patterns matching each of the 64 possible 6 nucleotide palindromes. The number of occurrences of each possible palindrome was recorded, and the 64 values were summed for each Mycobacteriophage genome.

The loop used to accomplish this was written as (SUM-OF

  (COUNTS-OF

    (APPLY-FUNCTION

      (MATCHES-OF-PATTERN X IN *MYCOBACTERIOPHAGE*

        :ONE-STRAND)

      REPLACING (X)

      WITH 6NTPALINDROMES)

)

)

where 6ntpalindromes is a list containing all possible six nucleotide palindromes.

The expected value was then compared to the actual number of all six nucleotide palindromes found in the genome of each Mycobacteriophage (Table 1).

Results and Discussion:

After examining the occurrences of six nucleotide palindromes in the Mycobacteriophages in BioBIKE, it was determined that based on the expected values computed using the Markov model explained above, not all Mycobacteriophages avoid six nucleotide palindromes. Table one displays a full list of the phages, the expected number of palindromes of length six, the observed number, and the ratio of observed to expected. These were broken down into six categories: those with a ratio less than .5, which were said to significantly avoid palindromes, those with a ratio between .5 and .75, which were said to avoid them, those with a ration between .75 and .8, which were said to slightly avoid them, those with a ratio between .8 and 1.0, which were said to not avoid them, and those with a ratio greater than 1.0, in which six nucleotide palindromes were said to be overrepresented.
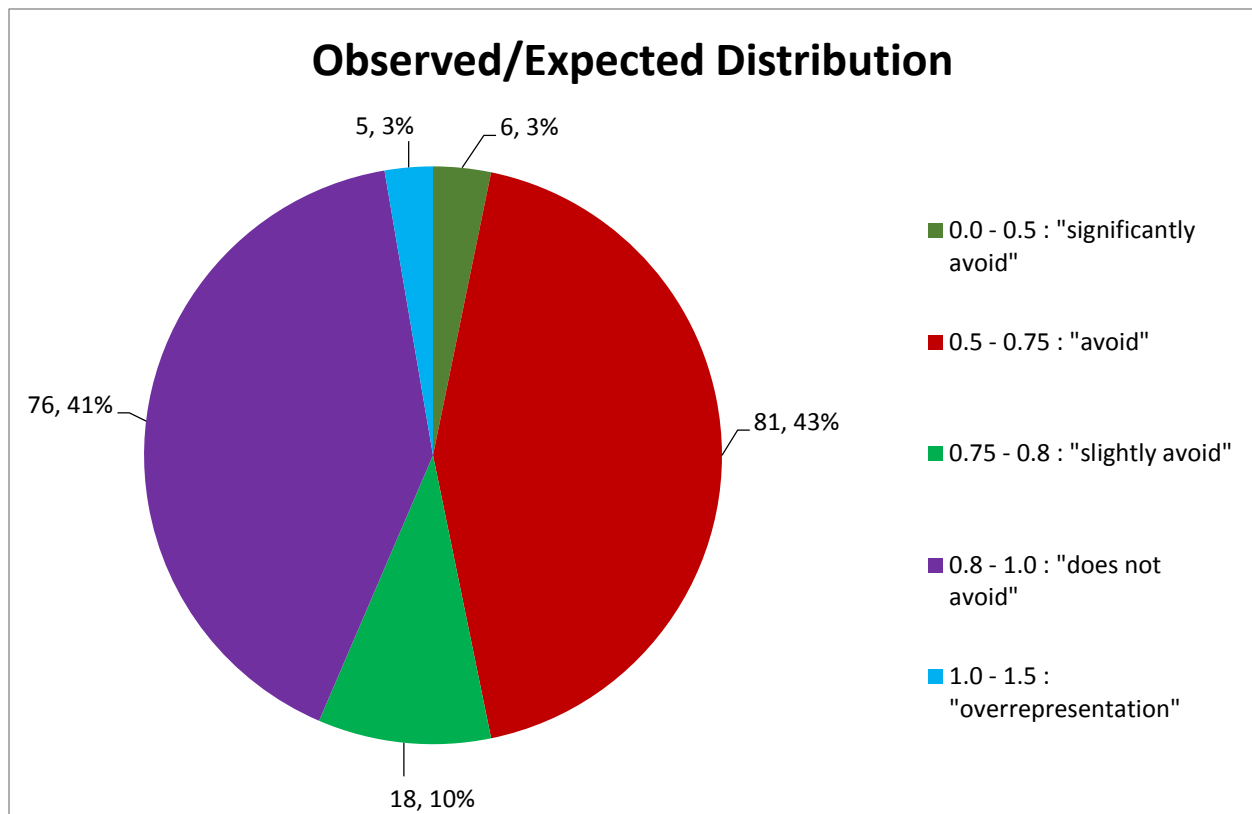
Observed/Expected Distribution

Figure 1: Frequency of each category of observed/expected 6 nucleotide palindrome ratio. Genomes with a ratio less than .5 were said to significantly avoid 6nt palindromes. Those with a ratio between .5 and .75 were said to avoid 6nt palindromes. Those between .75 and .8 were said to slightly avoid 6nt palindromes. Those between .8 and 1.0 were said to not avoid 6nt palindromes. Those with a ratio over 1.0 were said to have an overrepresentation of 6nt palindromes.

Those that avoid palindromes the most significantly (have the lowest observed to expected ratio) are all in the same cluster, which means they have similar genomic characteristics [6]. All those with a less than .5 ratio are in the A cluster. 81 of those examined had an observed to expected ratio of .8 or higher, and could not be said to avoid six nucleotide palindromes. Five have more six nucleotide palindromes than would be expected to occur, and the four of these with the highest observed to expected ratio are in the G cluster. As a whole, the statement that Mycobacteriophages avoid six nucleotide palindromes can be said to be false. Only 56% of those examined fell into the "avoid" range, with 44% showing no avoidance or an overrepresentation.

Further examination of why some do and some do not avoid six nucleotide palindromes may yield interesting results about the reasons for the presence of these restriction enzyme recognition sites in those phages that have a higher than expected number of them. It may also provide insight into the ability to tell purely through genome analysis which phages will be able to infect some strains of Mycobacteria, and which will not. For instance, G and K cluster phages are known to be able to infect and lyse Mycobacterium tuberculosis [9], a useful medical application, and the 5 phages with an overrepresentation of 6nt palindromes are G and K cluster. Consideration of the location

of these palindromes and examination of which palindromes occur most and least frequently could be useful in learning more about Mycobacteriophages. Also of interest could be the examination of palindrome avoidance in the different strains of Mycobacteria infected by these phages, and comparison of the strains able to be infected by each phage with the over or underrepresentation of six nucleotide palindromes in both the host bacteria and the infecting virus.

Table 1:

Phage Name	Expected	Observed	Observed/Expected
Mycobacterium-Phage-Rockstar	826.5075	320	0.38717132
Mycobacterium-Phage-Bxz2	876.0103	352	0.401821759
Mycobacterium-Phage-Vix	880.50714	365	0.414533833
Mycobacterium-Phage-Microwolf	876.0072	367	0.418946328
Mycobacterium-Phage-Jhc117	879.7546	373	0.423981869
Mycobacterium-Phage-George	800.0028	397	0.496248263
Mycobacterium-Phage-Backyardigan	828.5	427	0.515389258
Mycobacterium-Phage-Zemanar	1448.3317	760	0.524741673
Mycobacterium-Phage-Joedirt	1280.335	672	0.524862634
Mycobacterium-Phage-Frederick	1459.0752	770	0.527731538
Mycobacterium-Phage-Lebron	1259.1168	672	0.533707437
Mycobacterium-Phage-Eagle	848.5	453	0.533883324
Mycobacterium-Phage-Peaches	847.75	453	0.534355647
Mycobacterium-Phage-D29	874.25	468	0.535315985
Mycobacterium-Phage-Marvin	1135.991	610	0.536976085
Mycobacterium-Phage-Maverick	847.75	456	0.537894426
Mycobacterium-Phage-Upie	1266.3824	685	0.540910865
Mycobacterium-Phage-Airmid	783.25885	424	0.541328068
Mycobacterium-Phage-Lhtscc	859.75	467	0.543181157
Mycobacterium-Phage-Chrisnmich	1444.6351	788	0.545466464
Mycobacterium-Phage-Stinger	1414.6256	772	0.545727435
Mycobacterium-Phage-Benedict	781.2545	427	0.546556852
Mycobacterium-Phage-Shaka	842.75	461	0.547018689
Mycobacterium-Phage-Acadian	1522.4598	846	0.555679697
Mycobacterium-Phage-Nigel	1427.1075	800	0.560574449
Mycobacterium-Phage-Barnyard	1153.9502	655	0.567615483
Mycobacterium-Phage-L5	883.50116	502	0.568193934
Mycobacterium-Phage-Cooper	1466.992	856	0.583506931
Mycobacterium-Phage-Pukovnik	944.251	565	0.598357852

Mycobacterium-Phage-Hammer	824.9999	507	0.614545529
Mycobacterium-Phage-Phipps	1411.0684	869	0.615845412
Mycobacterium-Phage-Fang	1410.7926	872	0.61809227
Mycobacterium-Phage-Myrna	3127.9697	1936	0.618931827
Mycobacterium-Phage-Harvey	1402.0616	868	0.619088348
Mycobacterium-Phage-Kikipoo	1419.607	880	0.619889871
Mycobacterium-Phage-Isaaceli	1419.607	880	0.619889871
Mycobacterium-Phage-Orion	1413.8383	877	0.620297243
Mycobacterium-Phage-Scoot17c	1416.0873	879	0.620724443
Mycobacterium-Phage-Chah	1410.5685	878	0.622444071
Mycobacterium-Phage-Pg1	1426.8433	889	0.623053702
Mycobacterium-Phage-Threeoh3d2	1421.0746	886	0.623471843
Mycobacterium-Phage-Puhltonio	1409.7981	880	0.624202856
Mycobacterium-Phage-Yoshand	1413.078	884	0.625584716
Mycobacterium-Phage-Unclehowie	1410.633	883	0.625960119
Mycobacterium-Phage-Serendipity	1418.8256	890	0.62727935
Mycobacterium-Phage-Suffolk	1410.5844	885	0.627399537
Mycobacterium-Phage-Packman	879.7501	552	0.627450909
Mycobacterium-Phage-Abu	1422.8665	894	0.628309121
Mycobacterium-Phage-Klucky39	1400.0165	880	0.62856402
Mycobacterium-Phage-Hertubise	1412.3282	889	0.629457091
Mycobacterium-Phage-Oline	1407.591	889	0.631575507
Mycobacterium-Phage-Vista	1423.1383	899	0.631702485
Mycobacterium-Phage-Jacattac	1407.0596	890	0.632524735
Mycobacterium-Phage-Ericb	831.2512	526	0.632781041
Mycobacterium-Phage-Vortex	1410.5896	895	0.634486459
Mycobacterium-Phage-Jeffabunny	785.50287	499	0.635261842
Mycobacterium-Phage-Colbert	1393.3801	886	0.635863825
Mycobacterium-Phage-Oosterbaan	1416.8264	901	0.635928297
Mycobacterium-Phage-Wildcat	1082.3589	689	0.636572582
Mycobacterium-Phage-Gladiator	828.7501	528	0.637103996
Mycobacterium-Phage-Osmaximus	1425.1263	910	0.638539896
Mycobacterium-Phage-Tallgrassmm	1406.8848	899	0.639000436
Mycobacterium-Phage-Blue7	826.49994	529	0.640048443
Mycobacterium-Phage-Ares	1548.9417	995	0.64237408
Mycobacterium-Phage-Davinci	823.75275	530	0.643396942
Mycobacterium-Phage-Qyrzula	1545.6176	995	0.643755609

Mycobacterium-Phage-Rosebush	1545.1055	996	0.644616177
Mycobacterium-Phage-Arbiter	1531.7783	990	0.646307628
Mycobacterium-Phage-Hedgerow	1542.1447	997	0.646502238
Mycobacterium-Phage-Predator	1248.9448	829	0.66376032
Mycobacterium-Phage-Che12	907.2502	605	0.666850225
Mycobacterium-Phage-Akoma	1517.8437	1032	0.679911904
Mycobacterium-Phage-Pipefish	1520.1497	1038	0.682827487
Mycobacterium-Phage-Phlyer	1531.3407	1047	0.683714604
Mycobacterium-Phage-Daisy	1516.3138	1037	0.683895378
Mycobacterium-Phage-Phaedrus	1509.8226	1035	0.685511
Mycobacterium-Phage-Athena	1540.6119	1065	0.691283768
Mycobacterium-Phage-Yahalom	1510.5651	1047	0.693118092
Mycobacterium-Phage-Gadjet	1522.1165	1059	0.695741752
Mycobacterium-Phage-Kamiyu	1527.857	1064	0.696400252
Mycobacterium-Phage-Konstantine	1232.577	862	0.699347789
Mycobacterium-Phage-Redrock	990.5114	699	0.705696068
Mycobacterium-Phage-Rey	1358.0016	977	0.719439506
Mycobacterium-Phage-Corndog	1343.2139	969	0.721404089
Mycobacterium-Phage-Ibhubesi	858.9189	634	0.738137209
Mycobacterium-Phage-Tweety	937.6478	693	0.739083481
Mycobacterium-Phage-Che8	932.82654	696	0.746119423
Mycobacterium-Phage-Sg4	970.2353	739	0.761670906
Mycobacterium-Phage-Drago	849.9029	650	0.764793249
Mycobacterium-Phage-Ardmore	823.89746	631	0.765872005
Mycobacterium-Phage-Llij	899.18964	690	0.767357595
Mycobacterium-Phage-Gumbie	901.3692	692	0.767720929
Mycobacterium-Phage-Wee	969.2276	746	0.769685056
Mycobacterium-Phage-Pmc	887.1334	683	0.769895486
Mycobacterium-Phage-Dlane	956.01404	737	0.77090918
Mycobacterium-Phage-Fruitloop	951.72003	741	0.778590317
Mycobacterium-Phage-Anaya	1325.9509	1033	0.779063538
Mycobacterium-Phage-Mozy	908.9232	713	0.784444714
Mycobacterium-Phage-Dotproduct	893.3888	702	0.785772107
Mycobacterium-Phage-Pacc40	918.38495	724	0.788340445
Mycobacterium-Phage-Ramsey	908.1366	716	0.788427644
Mycobacterium-Phage-Boomer	905.5742	714	0.788450024
Mycobacterium-Phage-Tm4	1267.8118	1001	0.789549364

Mycobacterium-Phage-Crimd	1309.6504	1041	0.794868615
Mycobacterium-Phage-Hamulus	920.6539	733	0.796173242
Mycobacterium-Phage-Angelica	1292.3444	1035	0.800870109
Mycobacterium-Phage-Barrelroll	1312.6677	1053	0.802183218
Mycobacterium-Phage-Rockyhorror	880.1192	709	0.8055727
Mycobacterium-Phage-Adephagia	1300.6812	1060	0.814957578
Mycobacterium-Phage-Lrrhood	2958.8184	2445	0.826343381
Mycobacterium-Phage-Alice	2939.767	2437	0.828977262
Mycobacterium-Phage-Et08	2964.2844	2466	0.831903983
Mycobacterium-Phage-Mrgordo	966.1661	804	0.832155051
Mycobacterium-Phage-Rizal	2954.8162	2465	0.834231246
Mycobacterium-Phage-Catera	2955.3108	2466	0.834430003
Mycobacterium-Phage-Spud	2978.0637	2485	0.834434804
Mycobacterium-Phage-Scottmcg	2967.5542	2480	0.83570504
Mycobacterium-Phage-Pio	2996.3384	2505	0.836020391
Mycobacterium-Phage-Nappy	3023.6382	2528	0.836078867
Mycobacterium-Phage-Bxz1	2991.0889	2501	0.83615034
Mycobacterium-Phage-Cali	2969.844	2485	0.836744287
Mycobacterium-Phage-Babsiella	1089.9365	914	0.838580963
Mycobacterium-Phage-Linstu	2954.5876	2478	0.838695729
Mycobacterium-Phage-Drazdys	2986.845	2506	0.839012403
Mycobacterium-Phage-Lockley	968.1002	813	0.839789104
Mycobacterium-Phage-Bpbiebs31	997.6448	841	0.842985399
Mycobacterium-Phage-Sebata	2988.1057	2519	0.843009001
Mycobacterium-Phage-Dd5	966.6004	815	0.843161248
Mycobacterium-Phage-Museum	957.3462	810	0.846088907
Mycobacterium-Phage-Bxb1	959.3336	813	0.847463281
Mycobacterium-Phage-Kssjeb	974.6021	826	0.847525364
Mycobacterium-Phage-Pari	964.87524	819	0.848814402
Mycobacterium-Phage-Doom	978.8794	831	0.848929909
Mycobacterium-Phage-Giles	1284.9414	1091	0.849065957
Mycobacterium-Phage-Ghost	2969.8413	2524	0.849877062
Mycobacterium-Phage-Brujita	1066.9844	911	0.853808172
Mycobacterium-Phage-Jc27	976.10455	835	0.855441151
Mycobacterium-Phage-Skipole	1002.3972	860	0.857943338
Mycobacterium-Phage-Kbg	992.8924	852	0.858099025
Mycobacterium-Phage-Island3	1073.7094	922	0.858705344

Mycobacterium-Phage-Jasper	954.0794	821	0.860515383
Mycobacterium-Phage-Kugel	993.583	858	0.863541345
Mycobacterium-Phage-Switzer	1002.6178	866	0.863738904
Mycobacterium-Phage-Bethlehem	964.3784	835	0.865842702
Mycobacterium-Phage-Solon	936.6064	812	0.866959696
Mycobacterium-Phage-Violet	990.3454	862	0.870403397
Mycobacterium-Phage-Lesedi	958.36206	836	0.872321678
Mycobacterium-Phage-U2	986.61145	866	0.877751824
Mycobacterium-Phage-Billknuckles	981.8593	866	0.882000099
Mycobacterium-Phage-Troll4	1058.4268	934	0.882441752
Mycobacterium-Phage-Send513	1029.521	911	0.88487753
Mycobacterium-Phage-Plot	1063.1927	941	0.885070035
Mycobacterium-Phage-Sirharley	1053.6843	934	0.886413511
Mycobacterium-Phage-Gumball	1053.4343	935	0.88757315
Mycobacterium-Phage-Toto	1280.7897	1137	0.88773356
Mycobacterium-Phage-Pixie	1395.1572	1241	0.889505498
Mycobacterium-Phage-Butterscotch	1058.9402	945	0.892401667
Mycobacterium-Phage-Adjutor	1058.9402	947	0.894290348
Mycobacterium-Phage-Pbi1	1057.4404	947	0.895558747
Mycobacterium-Phage-Dori	1325.668	1188	0.896151978
Mycobacterium-Phage-Cjw1	1269.942	1139	0.896891354
Mycobacterium-Phage-Nova	1066.2172	957	0.897565712
Mycobacterium-Phage-Lilac	1280.2687	1150	0.898248938
Mycobacterium-Phage-Porky	1296.033	1167	0.900440035
Mycobacterium-Phage-Pumpkin	1242.404	1119	0.900673211
Mycobacterium-Phage-Elph10	1254.4476	1134	0.903983554
Mycobacterium-Phage-Sirduracell	1283.2993	1161	0.904699317
Mycobacterium-Phage-Bask21	1266.3124	1153	0.910517815
Mycobacterium-Phage-244	1248.4114	1137	0.910757463
Mycobacterium-Phage-Rakim	1289.1188	1175	0.911475343
Mycobacterium-Phage-Henry	1279.8008	1170	0.914204773
Mycobacterium-Phage-Yoshi	873.5246	801	0.916974748
Mycobacterium-Phage-Kostya	1289.8457	1184	0.917939254
Mycobacterium-Phage-Baka	1872.1436	1720	0.918732943
Mycobacterium-Phage-Optimus	1838.0822	1698	0.923788936
Mycobacterium-Phage-Che9d	843.51654	782	0.927071329
Mycobacterium-Phage-Littlee	1924.6324	1805	0.937841429

Mycobacterium-Phage-Omega	1932.4293	1821	0.942337192
Mycobacterium-Phage-Che9c	1178.1769	1129	0.958260173
Mycobacterium-Phage-Thibault	1804.4525	1730	0.958739562
Mycobacterium-Phage-Ds6a	1329.5009	1275	0.959006496
Mycobacterium-Phage-Larva	1465.9172	1840	1.255186855
Mycobacterium-Phage-Angel	1076.7394	1522	1.413526801
Mycobacterium-Phage-Hope	1079.9922	1530	1.416676898
Mycobacterium-Phage-Bps	1080.2422	1531	1.417274756
Mycobacterium-Phage-Halo	1099.5356	1566	1.424237651

References:

1. Jeltsch, A. and Pingouda, A. (2001). Structure and function of type II restriction endonucleases. Nucleic Acids Res. Sep 15, 2001, 29(18): 3705–3727.
2. Sharp, P. M. (1986). Molecular evolution of bacteriophages: Evidence of selection against the recognition sites of host restriction enzymes. Molecular Biology and Evolution, 3(1), 75-83.
3. Rocha, E., Danchin, A., & Viari, A. (2001). Evolutionary role of Restriction/Modification systems as revealed by comparative genome analysis. Genome Research, (11), 946-958. doi:10.1101/gr.153101
4. Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., et al. 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393: 537–544
5. Roberts, Richard J. and Vincze, Tamas and Posfai, Janos and Macelis, Dana. (2003). REBASE: restriction enzymes and methyltransferases. Nucl. Acids Res. 31 (1): 418-420. doi:10.1093/nar/gkg069
6. phagesdb.org
7. Gelfand, M. S., & Koonin, E. V. (1997). Avoidance of palindromic words in bacterial and archaeal genomes: A close connection with restriction enzymes. Nucleic Acids Research, 25(12), 2430-2439. doi:10.1093/nar/25.12.2430
8. Elhai J, Taton A, Massar JP, Myers JK, Travers M, Casey J, Slupesky M, Shrager J (2009). BioBIKE: A web-based, programmable, integrated biological knowledge base. Nucl Acids Res 37:W28-W32
9. Deborah Jacobs-Sera, Laura J. Marinelli, Charles Bowman, Gregory W. Broussard, Carlos Guerrero Bustamante, Michelle M. Boyle, Zaritza O. Petrova, Rebekah M. Dedrick, Welkin H. Pope, Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program, Robert L. Modlin, Roger W. Hendrix, Graham F. Hatfull, On the nature of mycobacteriophage diversity and host preference, Virology, Volume 434, Issue