Supplementary Information for

Rates of mutation and recombination in *Siphoviridae* phage genome evolution over three decades

Anne Kupczok, Horst Neve, Kun D. Huang, Marc P. Hoeppner, Knut J. Heller, Charles M. A. P. Franz, Tal Dagan

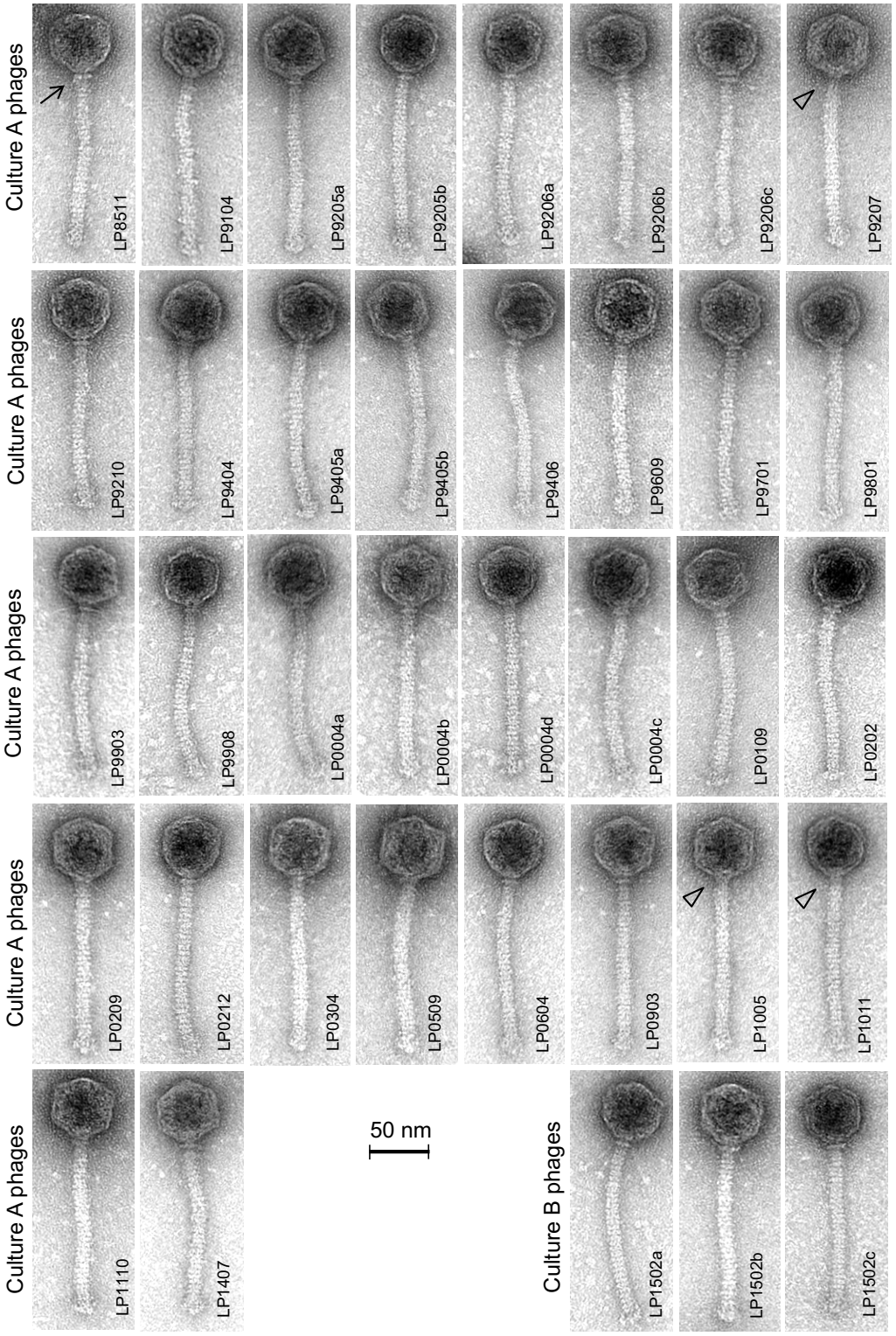


Figure S1: Electron micrographs of phages sequenced in this study. Phage names indicate isolation time (in YYMM format). All 34 culture A and all three phages from culture B are members of the lactococcal 936 group of *Siphoviridae* phages with isometric heads (ca. 55 nm diameter) and flexible, non-contractile tails (ca. 160 nm length). Except for few phages,

all phages possess a neck passage structure (NPS) with long thin whiskers (see arrow on LP8511 micrograph). The *nps* genes are truncated in four culture A phage genomes (LP9207, LP9404, LP1005, and LP1011). As expected, the neck is not visible in LP9207, LP1005, and LP1011 (see open triangles). The assembly of LP9404 shows a polymorphism at the nonsense mutation. Thus, the presence of phages with and without the nonsense mutation explains the discrepancy between the genome sequencing and the phenotype.

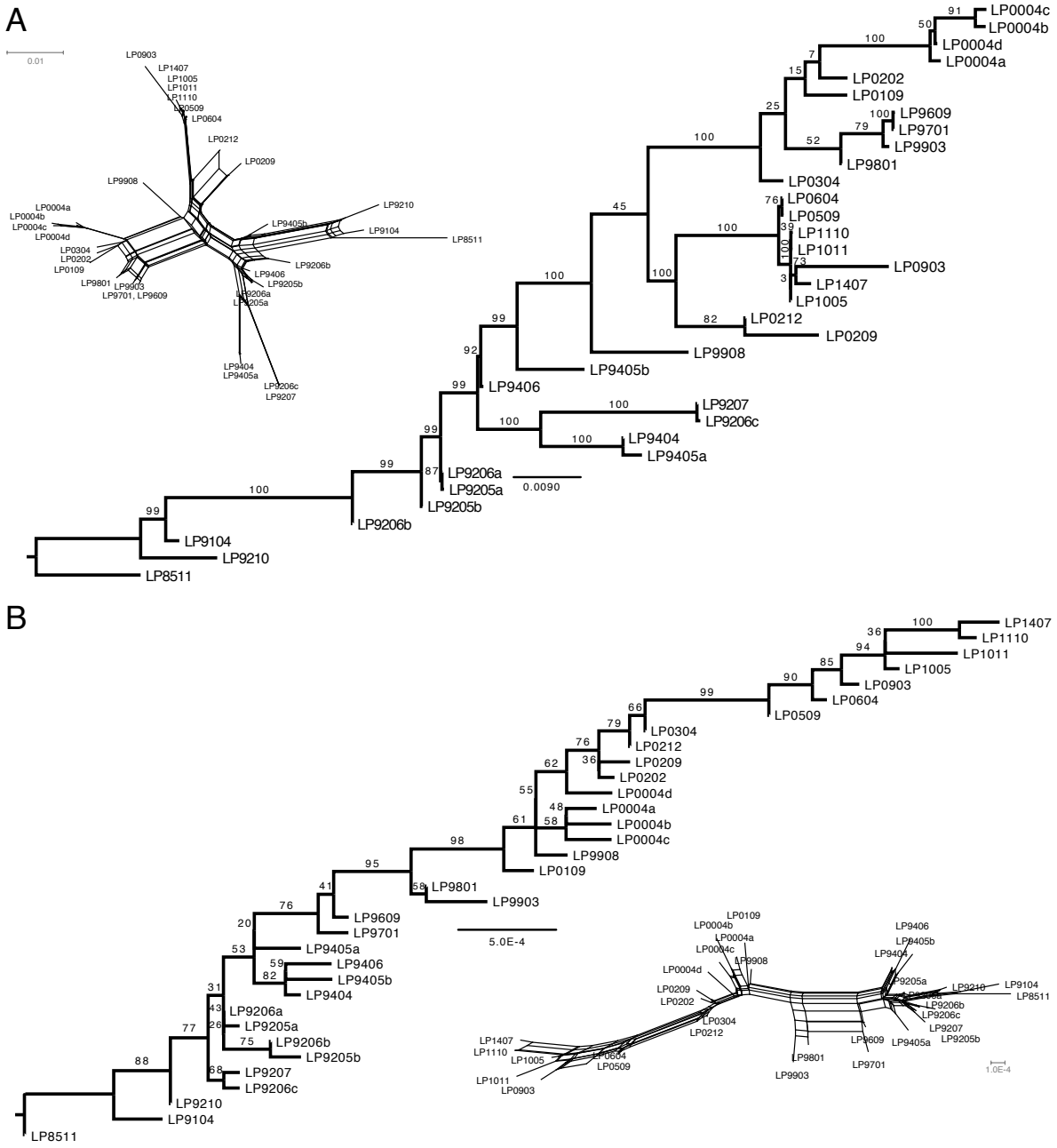


Figure S2: ML Phylogenies of the whole-genome alignment. Maximum likelihood phylogeny based on (A) whole alignment and (B) masked alignment. Insert figures: Neighbor-net based on ordinary least squares distances calculated with SplitsTree (Huson and Bryant 2006).

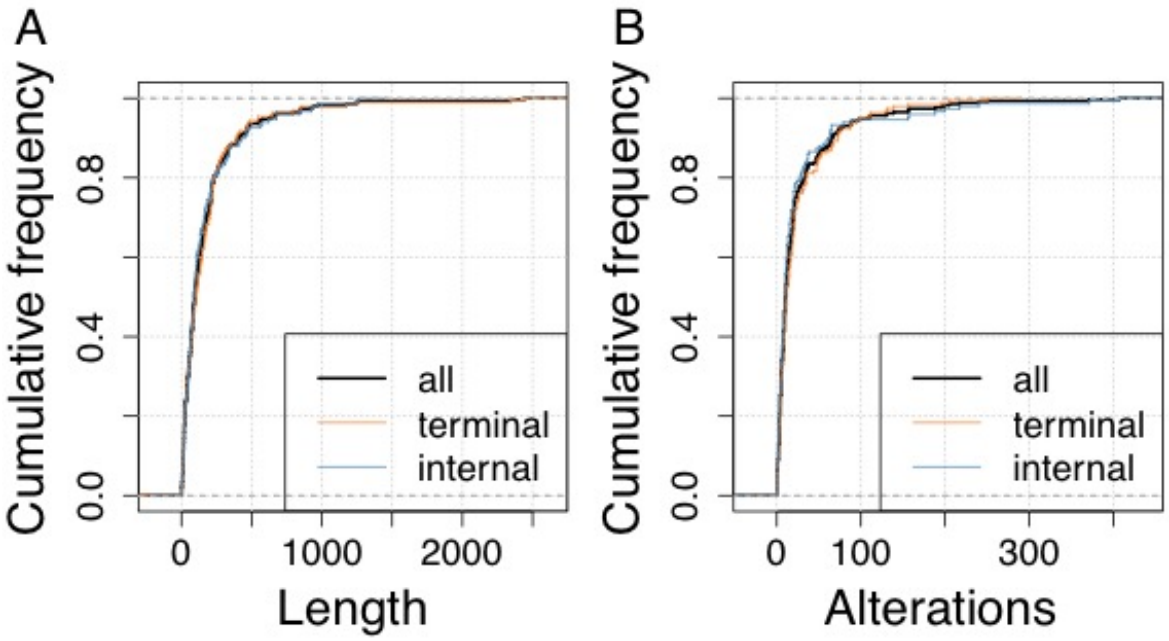


Figure S3: Recombination statistics. (A) Length distribution of 345 detected recombined segments, mean=176.9, median=96, no difference between terminal and internal branches ($p>0.05$, Anderson-Darling Test). (B) Distribution of number of nucleotide alterations introduced by one recombination event, mean=26.35, median=11, no difference between terminal and internal nodes ($p>0.05$, Anderson-Darling Test). Sequencing errors in single strains would lead to an overrepresentation of recombination events along terminal branches. Since the distribution of recombination length and altered nucleotides is not significantly different between terminal and internal branches, we can exclude the presence of strain-related artefacts in the data. Posterior mean values of parameters estimated by ClonalFrameML are: ratio of recombination to mutation rate $R/\theta=1.247$, import length $\delta=111.2$nt, nucleotide distance of imports $\nu=0.1694$ differences per nucleotide. Thus, the relative effect of recombination to mutation is $r/m=(R/\theta)\times\delta\times\nu=23.50$.

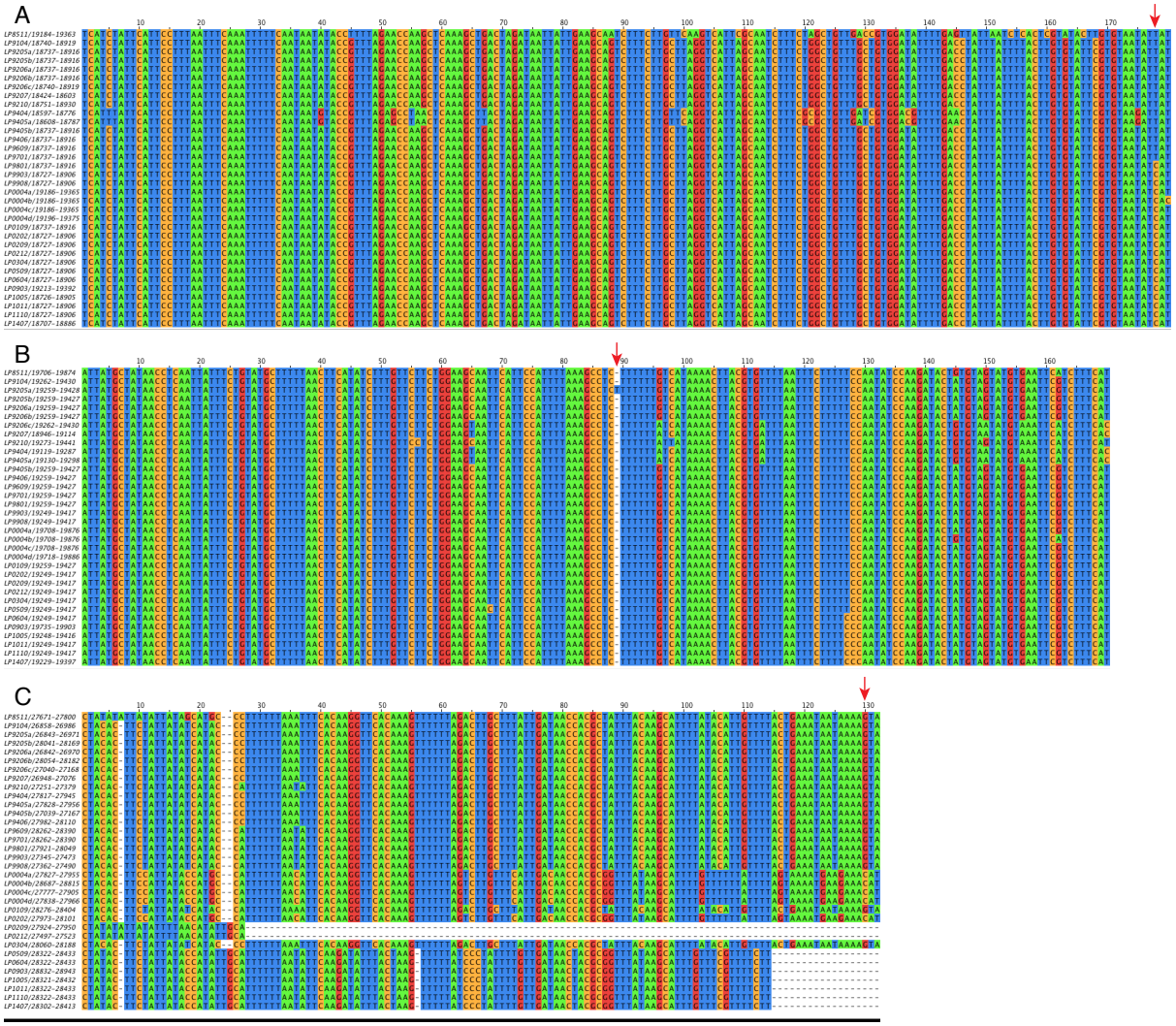


Figure S4: Alignments involving nonsense mutations. (A) Family 48 (alignment position 20179-20358, located on reverse strand) appeared first in isolate LP9801 due to a substitution in the start codon and persists until the end of the sampling. (B) Family 49 (alignment position 20701-20870, located on reverse strand) is absent in LP9205a due to a 1bp insertion resulting in a frameshift. (C) Family 70 (alignment position 32053-32184, located on reverse strand) is present in 5 strains due to substitutions in start codon. Nonsense mutations are marked by red arrows. Note that the nonsense mutations that involve start codons might also involve functional alternative start codons (AUA in A and UAC in B) (Hecht et al. 2017).

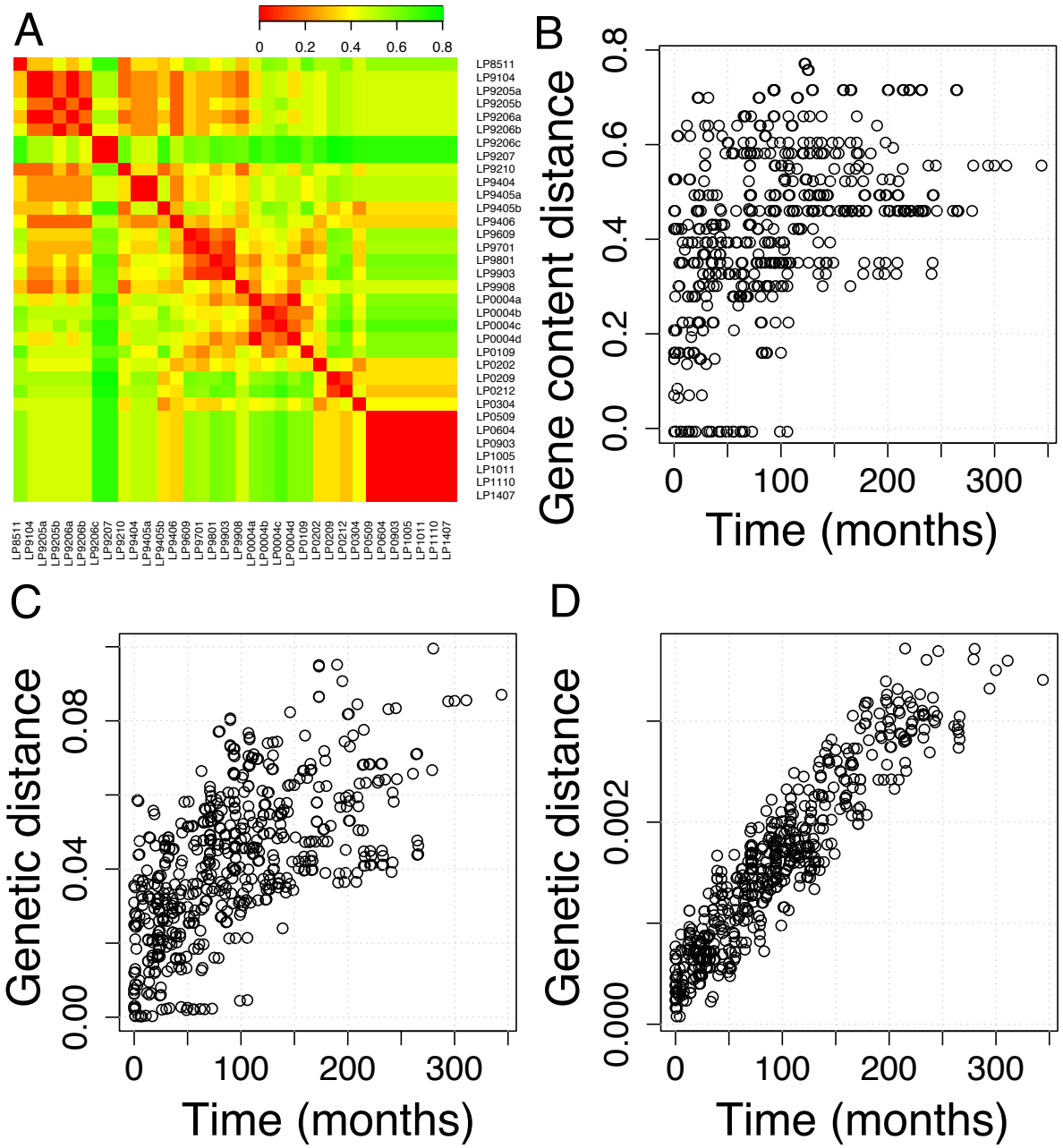


Figure S5: Gene indel family evolution over time. (A) Heatmap of pairwise gene content distance based on 25 gene indel families. We observe that the gene content distance is lower between isolates close in time, although there appears to be no gradual change of the gene content distance with time. (B) Pairwise gene content distance against pairwise isolation time differences (r^2=0.19). In accordance with a, we observe a weak correlation of pairwise gene content distance with time distance. (C) Pairwise genetic distances based on whole-genome alignment against pairwise isolation time differences (r^2=0.39). (D) Genetic distances based on whole-genome alignment after recombination masking against pairwise isolation time differences (r^2=0.87). In comparison to the gene content distances, the genetic distances based on the whole-genome alignment have a stronger correlation with time, and this correlation even increased when recombinations were masked.

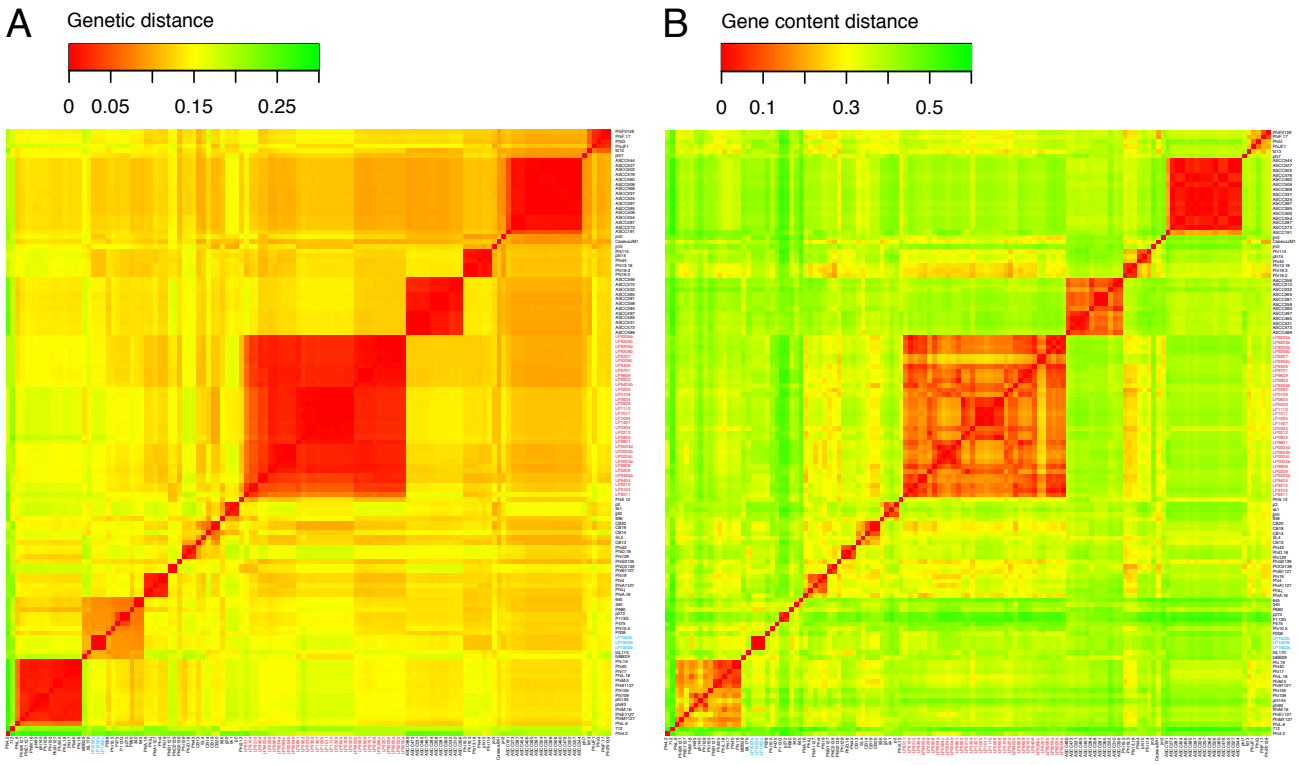


Figure S6: Heatmap of (A) genetic distances, and (B) gene content distances based on Jaccard index, for all phages of the 936 group. Genome order is the same in both matrices, based on clustering of genetic distances. Culture A phages are highlighted in red and culture B phages are highlighted in blue.

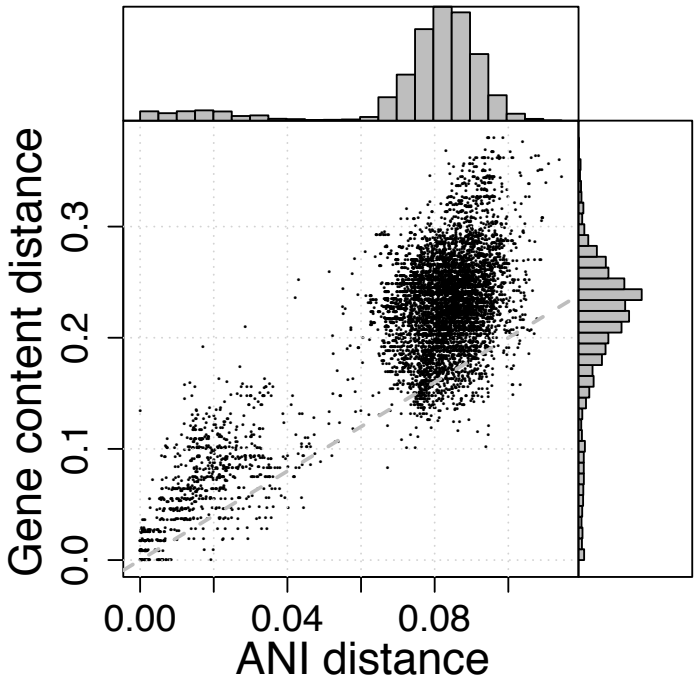


Figure S7: Association between pairwise gene content distance and ANI distance for all *L. lactis* phages of the 936 group. Gene content distances are based on the average proportion of shared genes. These statistics are similar to the ones used by (Mavrich and Hatfull 2017). The clustering pattern observed in Fig. 6 is also observed with this statistic. For an ANI distance cutoff of 0.05, low distance pairs have an average gene content distance of 0.06947 and r^2 of 0.3925 (n=881, $p<10^{-6}$), and high distance pairs have an average gene content distance of 0.2331 and r^2 of 0.07739 (n=7120, $p<10^{-6}$). We cannot distinguish between phages of high or low gene flux modes (Mavrich and Hatfull 2017) due to sampling density.

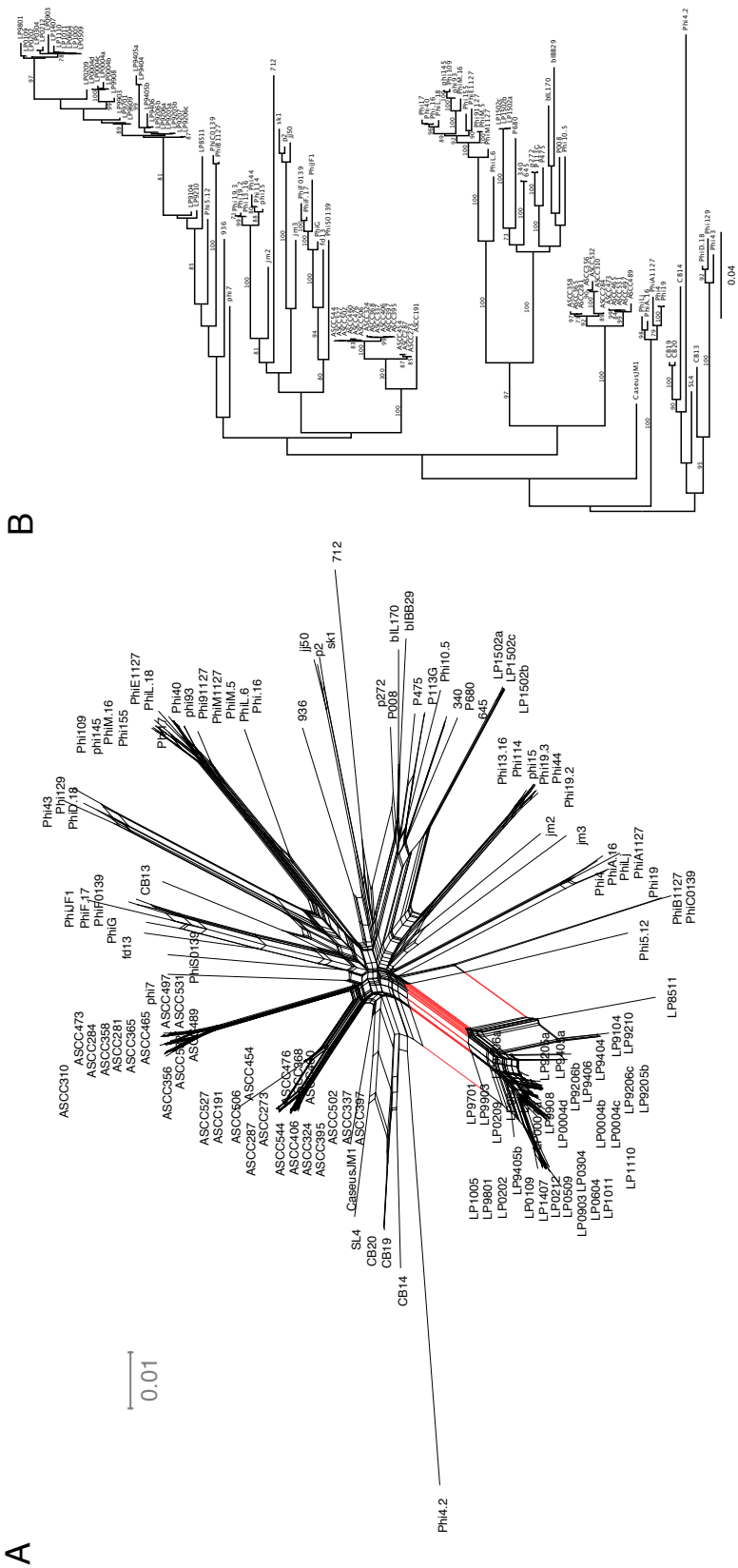
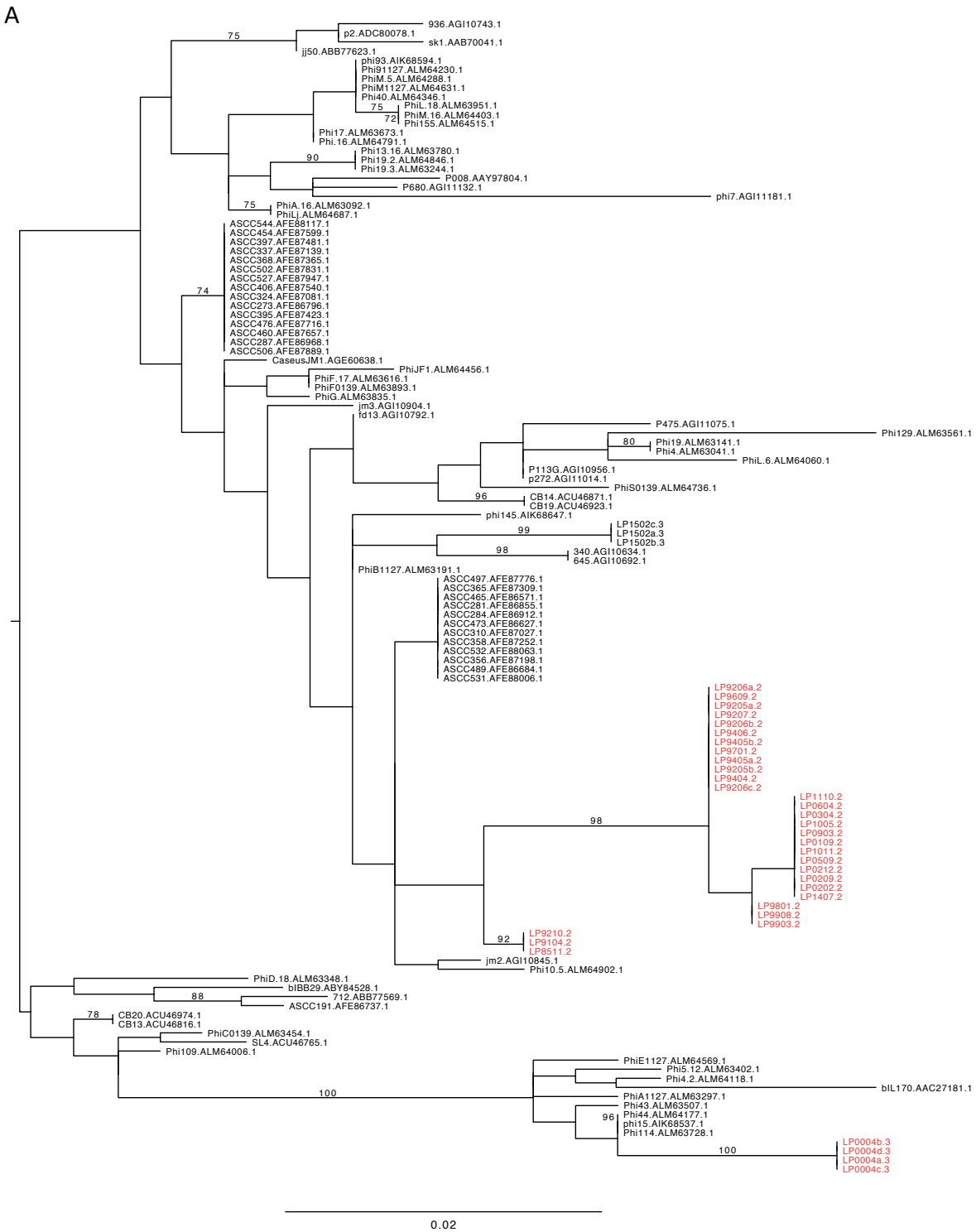


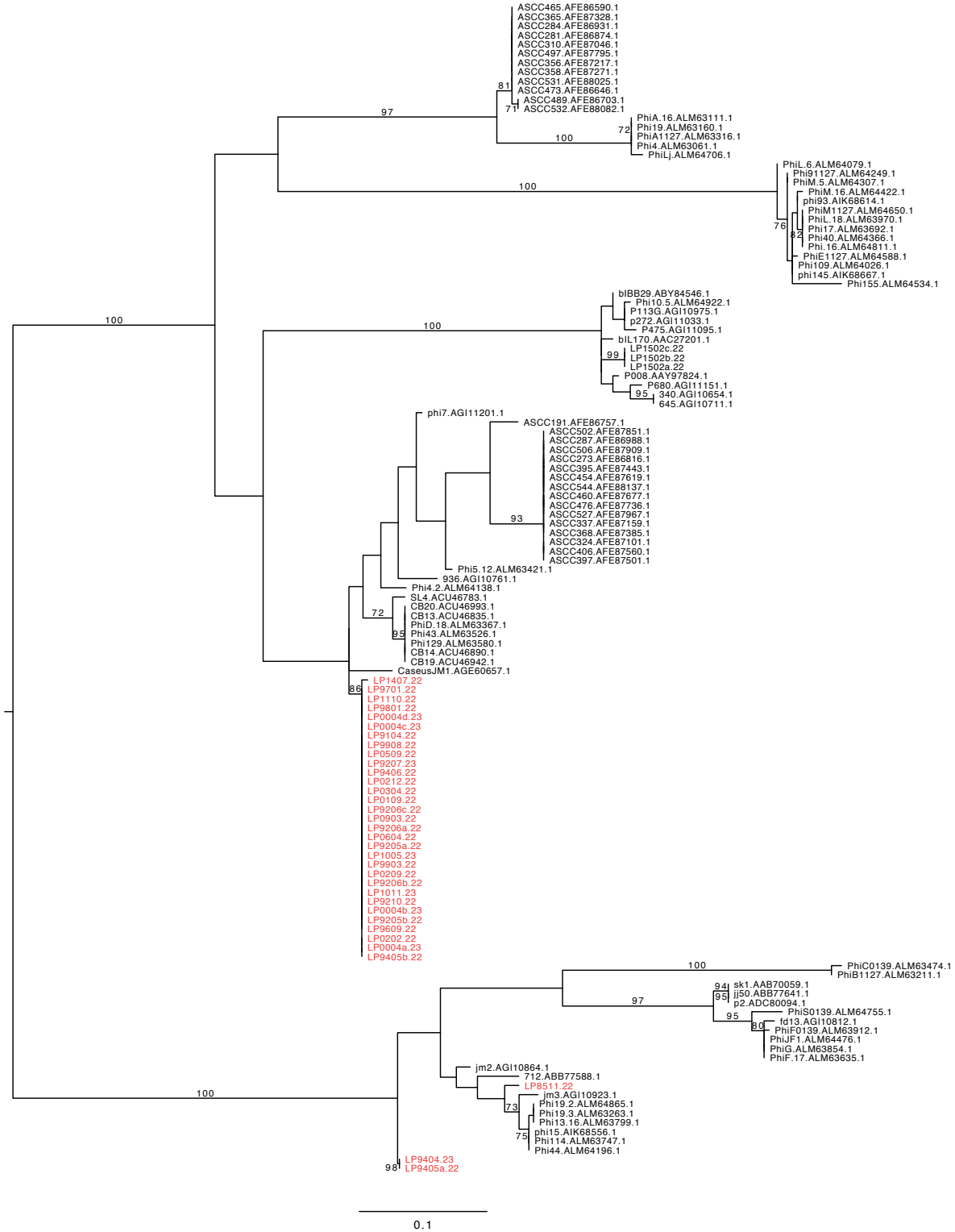
Figure S8: Relationships of *L. lactis* phages of the 936 group based on core gene alignment. (A) Neighbor-net based on ordinary least squares distances calculated with SplitsTree (Huson and Bryant 2006). The split that supports the monophyly of culture A phages is marked in red. (B) ML phylogeny, only bootstrap values above 70 are displayed. Note that this is an unrooted tree. Culture A phages are monophyletic in the tree whereas culture B phages group in a notably different region.


A
936.AGI10743.1
p2.ADC80078.1
sk1.AAB70041.1
75
jj50.ABB77623.1
phi93.AIK68594.1
Phi91127.ALM64230.1
PhiM.5.ALM64288.1
PhiM1127.ALM64631.1
Phi40.ALM64346.1
75
PhiL.18.ALM63951.1
72
PhiM.16.ALM64403.1
Phi155.ALM64515.1
Phi17.ALM63673.1
Phi.16.ALM64791.1
90
Phi13.16.ALM63780.1
Phi19.2.ALM64846.1
Phi19.3.ALM63244.1
P008.AAY97804.1
P680.AGI11132.1
phi7.AGI11181.1
75
PhiA.16.ALM63092.1
PhiLj.ALM64687.1
ASCC544.AFE88117.1
ASCC454.AFE87599.1
ASCC397.AFE87481.1
ASCC337.AFE87139.1
ASCC368.AFE87365.1
ASCC502.AFE87831.1
ASCC527.AFE87947.1
ASCC406.AFE87540.1
74
ASCC324.AFE87081.1
ASCC273.AFE86796.1
ASCC395.AFE87423.1
ASCC476.AFE87716.1
ASCC460.AFE87657.1
ASCC287.AFE86968.1
ASCC506.AFE87889.1
CaseusJM1.AGE60638.1
PhiJF1.ALM64456.1
PhiF.17.ALM63616.1
PhiF0139.ALM63893.1
PhiG.ALM63835.1
jm3.AGI10904.1
fd13.AGI10792.1
P475.AGI11075.1
Phi129.ALM63561.1
80
Phi19.ALM63141.1
Phi4.ALM63041.1
PhiL.6.ALM64060.1
P113G.AGI10956.1
p272.AGI11014.1
PhiS0139.ALM64736.1
96
CB14.ACU46871.1
CB19.ACU46923.1
phi145.AIK68647.1
99
LP1502c.3
LP1502a.3
LP1502b.3
98
340.AGI10634.1
645.AGI10692.1
PhiB1127.ALM63191.1
ASCC497.AFE87776.1
ASCC365.AFE87309.1
ASCC465.AFE86571.1
ASCC281.AFE86855.1
ASCC284.AFE86912.1
ASCC473.AFE86627.1
ASCC310.AFE87027.1
ASCC358.AFE87252.1
ASCC532.AFE88063.1
ASCC356.AFE87198.1
ASCC489.AFE86684.1
ASCC531.AFE88006.1
LP9206a.2
LP9609.2
LP9205a.2
LP9207.2
LP9206b.2
LP9406.2
LP9405b.2
LP9701.2
LP9405a.2
LP9205b.2
LP9404.2
LP9206c.2
LP1110.2
LP0604.2
LP0304.2
LP1005.2
LP0903.2
LP0109.2
LP1011.2
LP0509.2
LP0212.2
LP0209.2
LP0202.2
LP1407.2
98
LP9801.2
LP9908.2
LP9903.2
92
LP9210.2
LP9104.2
LP8511.2
jm2.AGI10845.1
Phi10.5.ALM64902.1
PhiD.18.ALM63348.1
bIBB29.ABY84528.1
88
712.ABB77569.1
ASCC191.AFE86737.1
78
CB20.ACU46974.1
CB13.ACU46816.1
PhiC0139.ALM63454.1
SL4.ACU46765.1
Phi109.ALM64006.1
PhiE1127.ALM64569.1
Phi5.12.ALM63402.1
Phi4.2.ALM64118.1
bIL170.AAC27181.1
100
PhiA1127.ALM63297.1
Phi43.ALM63507.1
96
Phi44.ALM64177.1
phi15.AIK68537.1
Phi114.ALM63728.1
100
LP0004b.3
LP0004d.3
LP0004a.3
LP0004c.3
0.02

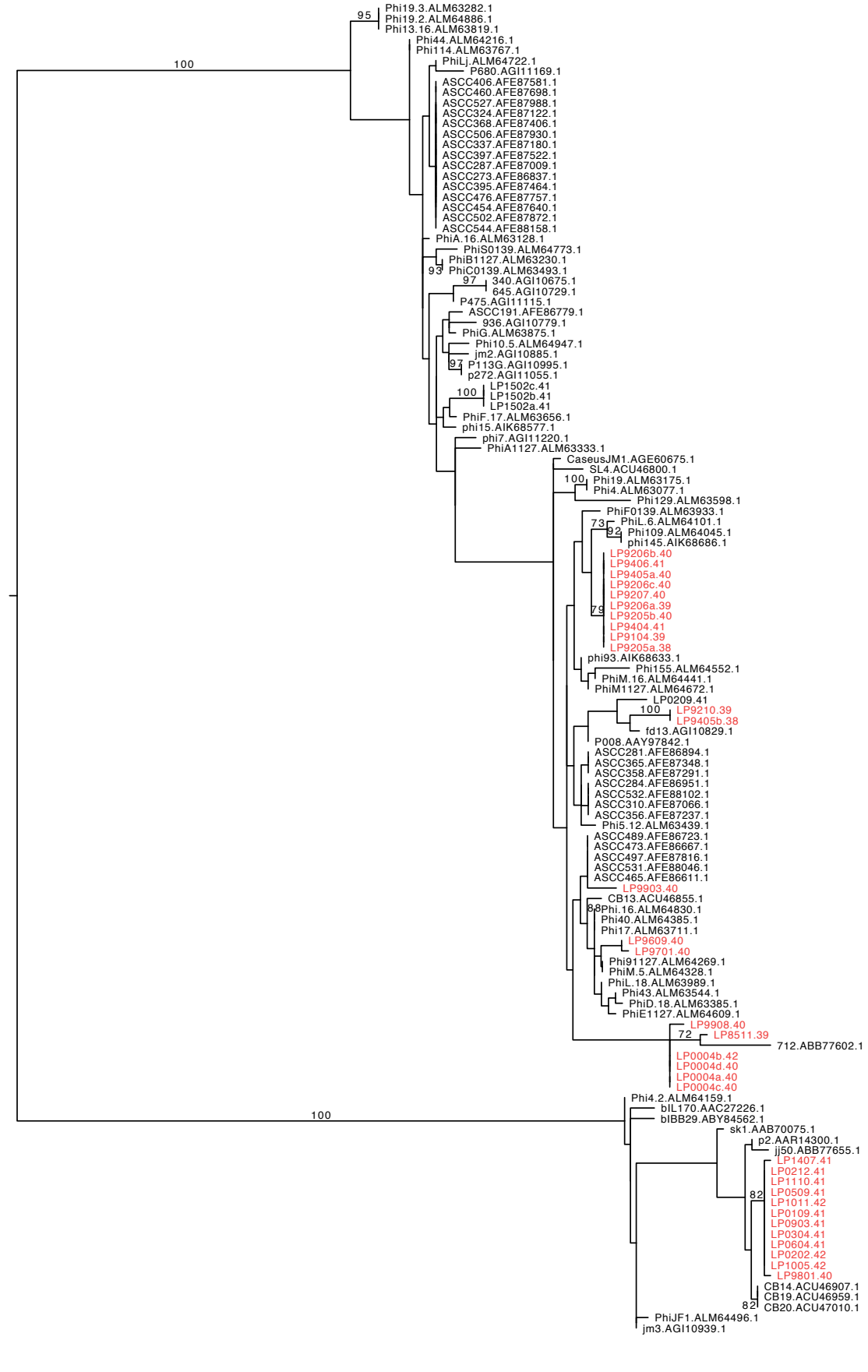
B


CaseusJM1.AGE60651.1
CB13.ACU46829.1
Phi4.2.ALM64131.1
PhiJF1.ALM64470.1
PhiF0139.ALM63906.1
PhiF.17.ALM63629.1
PhiG.ALM63848.1
phi7.AGI11195.1
fd13.AGI10806.1
SL4.ACU46778.1
CB20.ACU46987.1
CB19.ACU46936.1
jm3.AGI10917.1
Phi19.2.ALM64859.1
phi15.AIK68550.1
Phi19.3.ALM63257.1
Phi44.ALM64190.1
Phi114.ALM63741.1
Phi13.16.ALM63793.1
Phi5.12.ALM63415.1
LP9210.16
LP9104.16
LP8511.16
PhiC0139.ALM63468.1
PhiB1127.ALM63205.1
936.AGI10755.1
CB14.ACU46884.1
sk1.AAB70053.1
p2.ADC80089.1
lj50.ABB77635.1
PhiA1127.ALM63310.1
PhiLj.ALM64700.1
PhiA.16.ALM63105.1
Phi4.ALM63055.1
Phi19.ALM63154.1
LP0202.16
LP0212.16
LP0903.16
LP1011.17
LP0209.16
LP0604.16
LP1005.17
LP0304.16
LP0509.16
LP1407.16
LP1110.16
LP0004d.17
LP9207.17
LP9205a.16
LP9206b.16
LP9609.16
LP9206a.16
LP9405b.16
LP9206c.16
LP9205b.16
LP0004b.17
LP9701.16
LP0004c.17
LP9406.16
LP9801.16
LP9903.16
LP9404.17
LP0004a.17
LP9908.16
LP0109.16
LP9405a.16
ASCC532.AFE88076.1
ASCC356.AFE87211.1
ASCC310.AFE87040.1
ASCC531.AFE88019.1
ASCC358.AFE87265.1
ASCC473.AFE86640.1
ASCC497.AFE87789.1
ASCC365.AFE87322.1
ASCC281.AFE86868.1
ASCC465.AFE86584.1
ASCC489.AFE86697.1
ASCC284.AFE86925.1
ASCC324.AFE87095.1
ASCC395.AFE87437.1
ASCC406.AFE87554.1
ASCC397.AFE87495.1
ASCC337.AFE87153.1
ASCC368.AFE87379.1
ASCC544.AFE88131.1
ASCC460.AFE87671.1
ASCC476.AFE87730.1
ASCC506.AFE87903.1
ASCC527.AFE87961.1
ASCC502.AFE87845.1
ASCC191.AFE86751.1
ASCC287.AFE86982.1
ASCC273.AFE86810.1
ASCC454.AFE87613.1
jm2.AGI10858.1
PhiS0139.ALM64749.1
Phi10.5.ALM64915.1
P680.AGI11145.1
bIBB29.ABY84540.1
340.AGI10648.1
645.AGI10705.1
P008.AAY97818.1
p272.AGI11027.1
P113G.AGI10969.1
P475.AGI11089.1
LP1502a.16
LP1502b.16
LP1502c.16
bIL170.AAC27195.1
712.ABB77582.1
Phi43.ALM63520.1
PhiD.18.ALM63361.1
Phi129.ALM63574.1
PhiL.6.ALM64073.1
PhiM.5.ALM64301.1
Phi91127.ALM64243.1
Phi109.ALM64020.1
phi145.AIK68661.1
phi93.AIK68608.1
PhiM.16.ALM64416.1
PhiM1127.ALM64644.1
Phi.16.ALM64805.1
PhiL.18.ALM63964.1
Phi155.ALM64528.1
PhiE1127.ALM64582.1
Phi40.ALM64360.1
Phi17.ALM63686.1
0.06

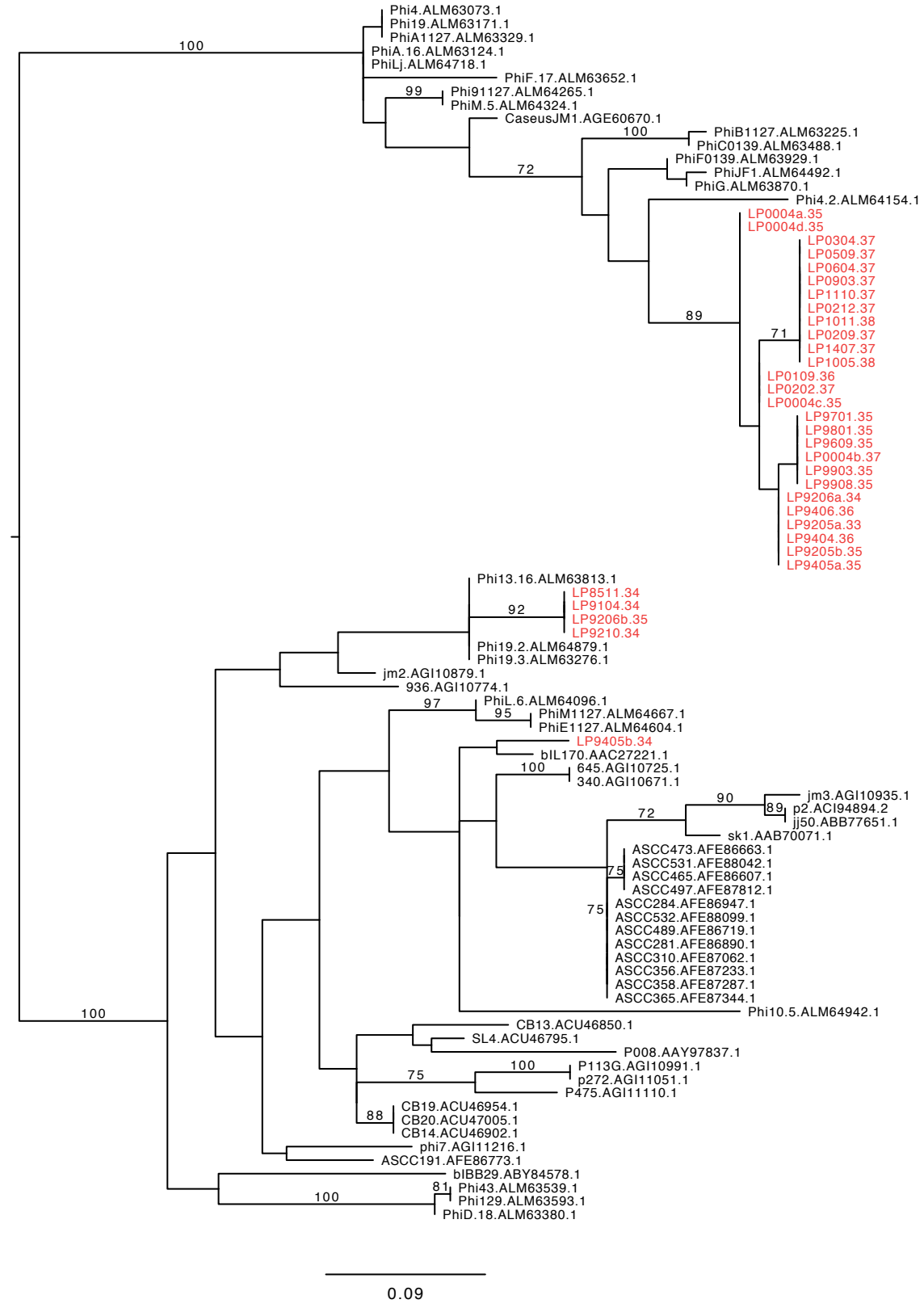
C


ASCC465.AFE86590.1
ASCC365.AFE87328.1
ASCC284.AFE86931.1
ASCC281.AFE86874.1
ASCC310.AFE87046.1
ASCC497.AFE87795.1
ASCC356.AFE87217.1
ASCC358.AFE87271.1
ASCC531.AFE88025.1
ASCC473.AFE86646.1
ASCC489.AFE86703.1
ASCC532.AFE88082.1
PhiA.16.ALM63111.1
Phi19.ALM63160.1
PhiA1127.ALM63316.1
Phi4.ALM63061.1
PhiLj.ALM64706.1
PhiL.6.ALM64079.1
Phi91127.ALM64249.1
PhiM.5.ALM64307.1
PhiM.16.ALM64422.1
phi93.AIK68614.1
PhiM1127.ALM64650.1
PhiL.18.ALM63970.1
Phi17.ALM63692.1
Phi40.ALM64366.1
Phi.16.ALM64811.1
PhiE1127.ALM64588.1
Phi109.ALM64026.1
phi145.AIK68667.1
Phi155.ALM64534.1
bIBB29.ABY84546.1
Phi10.5.ALM64922.1
P113G.AGI10975.1
p272.AGI11033.1
P475.AGI11095.1
bIL170.AAC27201.1
LP1502c.22
LP1502b.22
LP1502a.22
P008.AAY97824.1
P680.AGI11151.1
340.AGI10654.1
645.AGI10711.1
phi7.AGI11201.1
ASCC191.AFE86757.1
ASCC502.AFE87851.1
ASCC287.AFE86988.1
ASCC506.AFE87909.1
ASCC273.AFE86816.1
ASCC395.AFE87443.1
ASCC454.AFE87619.1
ASCC544.AFE88137.1
ASCC460.AFE87677.1
ASCC476.AFE87736.1
ASCC527.AFE87967.1
ASCC337.AFE87159.1
ASCC368.AFE87385.1
ASCC324.AFE87101.1
ASCC406.AFE87560.1
ASCC397.AFE87501.1
Phi5.12.ALM63421.1
936.AGI10761.1
Phi4.2.ALM64138.1
SL4.ACU46783.1
CB20.ACU46993.1
CB13.ACU46835.1
PhiD.18.ALM63367.1
Phi43.ALM63526.1
Phi129.ALM63580.1
CB14.ACU46890.1
CB19.ACU46942.1
CaseusJM1.AGE60657.1
LP1407.22
LP9701.22
LP1110.22
LP9801.22
LP0004d.23
LP0004c.23
LP9104.22
LP9908.22
LP0509.22
LP9207.23
LP9406.22
LP0212.22
LP0304.22
LP0109.22
LP9206c.22
LP0903.22
LP9206a.22
LP0604.22
LP9205a.22
LP1005.23
LP9903.22
LP0209.22
LP9206b.22
LP1011.23
LP9210.22
LP0004b.23
LP9205b.22
LP9609.22
LP0202.22
LP0004a.23
LP9405b.22
PhiC0139.ALM63474.1
PhiB1127.ALM63211.1
sk1.AAB70059.1
jj50.ABB77641.1
p2.ADC80094.1
PhiS0139.ALM64755.1
fd13.AGI10812.1
PhiF0139.ALM63912.1
PhiJF1.ALM64476.1
PhiG.ALM63854.1
PhiF.17.ALM63635.1
jm2.AGI10864.1
712.ABB77588.1
LP8511.22
jm3.AGI10923.1
Phi19.2.ALM64865.1
Phi19.3.ALM63263.1
Phi13.16.ALM63799.1
phi15.AIK68556.1
Phi114.ALM63747.1
Phi44.ALM64196.1
LP9404.23
LP9405a.22
0.1

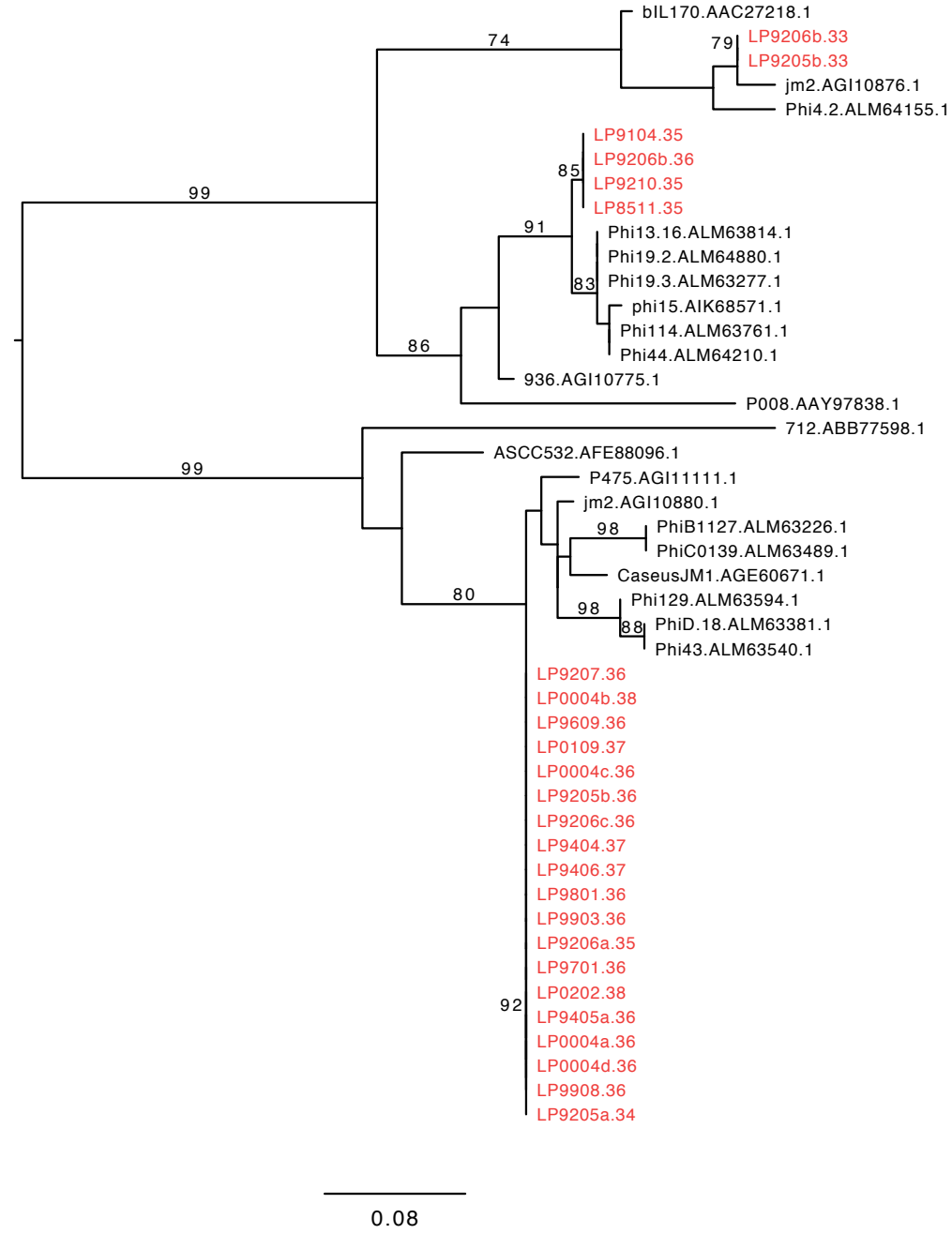
D


Phi19.3.ALM63282.1
Phi19.2.ALM64886.1
Phi13.16.ALM63819.1
Phi44.ALM64216.1
Phi114.ALM63767.1
PhiLj.ALM64722.1
P680.AGI11169.1
ASCC406.AFE87581.1
ASCC460.AFE87698.1
ASCC527.AFE87988.1
ASCC324.AFE87122.1
ASCC368.AFE87406.1
ASCC506.AFE87930.1
ASCC337.AFE87180.1
ASCC397.AFE87522.1
ASCC287.AFE87009.1
ASCC273.AFE86837.1
ASCC395.AFE87464.1
ASCC476.AFE87757.1
ASCC454.AFE87640.1
ASCC502.AFE87872.1
ASCC544.AFE88158.1
PhiA.16.ALM63128.1
PhiS0139.ALM64773.1
PhiB1127.ALM63230.1
PhiC0139.ALM63493.1
340.AGI10675.1
645.AGI10729.1
P475.AGI11115.1
ASCC191.AFE86779.1
936.AGI10779.1
PhiG.ALM63875.1
Phi10.5.ALM64947.1
jm2.AGI10885.1
P113G.AGI10995.1
p272.AGI11055.1
LP1502c.41
LP1502b.41
LP1502a.41
PhiF.17.ALM63656.1
phi15.AIK68577.1
phi7.AGI11220.1
PhiA1127.ALM63333.1
CaseusJM1.AGE60675.1
SL4.ACU46800.1
Phi19.ALM63175.1
Phi4.ALM63077.1
Phi129.ALM63598.1
PhiF0139.ALM63933.1
PhiL.6.ALM64101.1
Phi109.ALM64045.1
phi145.AIK68686.1
LP9206b.40
LP9406.41
LP9405a.40
LP9206c.40
LP9207.40
LP9206a.39
LP9205b.40
LP9404.41
LP9104.39
LP9205a.38
phi93.AIK68633.1
Phi155.ALM64552.1
PhiM.16.ALM64441.1
PhiM1127.ALM64672.1
LP0209.41
LP9210.39
LP9405b.38
fd13.AGI10829.1
P008.AAY97842.1
ASCC281.AFE86894.1
ASCC365.AFE87348.1
ASCC358.AFE87291.1
ASCC284.AFE86951.1
ASCC532.AFE88102.1
ASCC310.AFE87066.1
ASCC356.AFE87237.1
Phi5.12.ALM63439.1
ASCC489.AFE86723.1
ASCC473.AFE86667.1
ASCC497.AFE87816.1
ASCC531.AFE88046.1
ASCC465.AFE86611.1
LP9903.40
CB13.ACU46855.1
Phi.16.ALM64830.1
Phi40.ALM64385.1
Phi17.ALM63711.1
LP9609.40
LP9701.40
Phi91127.ALM64269.1
PhiM.5.ALM64328.1
PhiL.18.ALM63989.1
Phi43.ALM63544.1
PhiD.18.ALM63385.1
PhiE1127.ALM64609.1
LP9908.40
LP8511.39
712.ABB77602.1
LP0004b.42
LP0004d.40
LP0004a.40
LP0004c.40
Phi4.2.ALM64159.1
bIL170.AAC27226.1
bIBB29.ABY84562.1
sk1.AAB70075.1
p2.AAR14300.1
jj50.ABB77655.1
LP1407.41
LP0212.41
LP1110.41
LP0509.41
LP1011.42
LP0109.41
LP0903.41
LP0304.41
LP0604.41
LP0202.42
LP1005.42
LP9801.40
CB14.ACU46907.1
CB19.ACU46959.1
CB20.ACU47010.1
PhiJF1.ALM64496.1
jm3.AGI10939.1
0.2

E


Phi4.ALM63073.1
Phi19.ALM63171.1
PhiA1127.ALM63329.1
PhiA.16.ALM63124.1
PhiLj.ALM64718.1
PhiF.17.ALM63652.1
Phi91127.ALM64265.1
PhiM.5.ALM64324.1
CaseusJM1.AGE60670.1
PhiB1127.ALM63225.1
PhiC0139.ALM63488.1
PhiF0139.ALM63929.1
PhiJF1.ALM64492.1
PhiG.ALM63870.1
Phi4.2.ALM64154.1
LP0004a.35
LP0004d.35
LP0304.37
LP0509.37
LP0604.37
LP0903.37
LP1110.37
LP0212.37
LP1011.38
LP0209.37
LP1407.37
LP1005.38
LP0109.36
LP0202.37
LP0004c.35
LP9701.35
LP9801.35
LP9609.35
LP0004b.37
LP9903.35
LP9908.35
LP9206a.34
LP9406.36
LP9205a.33
LP9404.36
LP9205b.35
LP9405a.35
Phi13.16.ALM63813.1
LP8511.34
LP9104.34
LP9206b.35
LP9210.34
Phi19.2.ALM64879.1
Phi19.3.ALM63276.1
jm2.AGI10879.1
936.AGI10774.1
PhiL.6.ALM64096.1
PhiM1127.ALM64667.1
PhiE1127.ALM64604.1
LP9405b.34
bIL170.AAC27221.1
645.AGI10725.1
340.AGI10671.1
jm3.AGI10935.1
p2.ACI94894.2
jj50.ABB77651.1
sk1.AAB70071.1
ASCC473.AFE86663.1
ASCC531.AFE88042.1
ASCC465.AFE86607.1
ASCC497.AFE87812.1
ASCC284.AFE86947.1
ASCC532.AFE88099.1
ASCC489.AFE86719.1
ASCC281.AFE86890.1
ASCC310.AFE87062.1
ASCC356.AFE87233.1
ASCC358.AFE87287.1
ASCC365.AFE87344.1
Phi10.5.ALM64942.1
CB13.ACU46850.1
SL4.ACU46795.1
P008.AAY97837.1
P113G.AGI10991.1
p272.AGI11051.1
P475.AGI11110.1
CB19.ACU46954.1
CB20.ACU47005.1
CB14.ACU46902.1
phi7.AGI11216.1
ASCC191.AFE86773.1
bIBB29.ABY84578.1
Phi43.ALM63539.1
Phi129.ALM63593.1
PhiD.18.ALM63380.1
0.09

F


bIL170.AAC27218.1
74
79
LP9206b.33
LP9205b.33
jm2.AGI10876.1
Phi4.2.ALM64155.1
LP9104.35
LP9206b.36
85
LP9210.35
LP8511.35
99
91
Phi13.16.ALM63814.1
Phi19.2.ALM64880.1
83
Phi19.3.ALM63277.1
phi15.AIK68571.1
Phi114.ALM63761.1
86
Phi44.ALM64210.1
936.AGI10775.1
P008.AAY97838.1
712.ABB77598.1
ASCC532.AFE88096.1
99
P475.AGI11111.1
jm2.AGI10880.1
98
PhiB1127.ALM63226.1
PhiC0139.ALM63489.1
CaseusJM1.AGE60671.1
80
Phi129.ALM63594.1
98
88
PhiD.18.ALM63381.1
Phi43.ALM63540.1
LP9207.36
LP0004b.38
LP9609.36
LP0109.37
LP0004c.36
LP9205b.36
LP9206c.36
LP9404.37
LP9406.37
LP9801.36
LP9903.36
LP9206a.35
LP9701.36
92
LP0202.38
LP9405a.36
LP0004a.36
LP0004d.36
LP9908.36
LP9205a.34
0.08

G

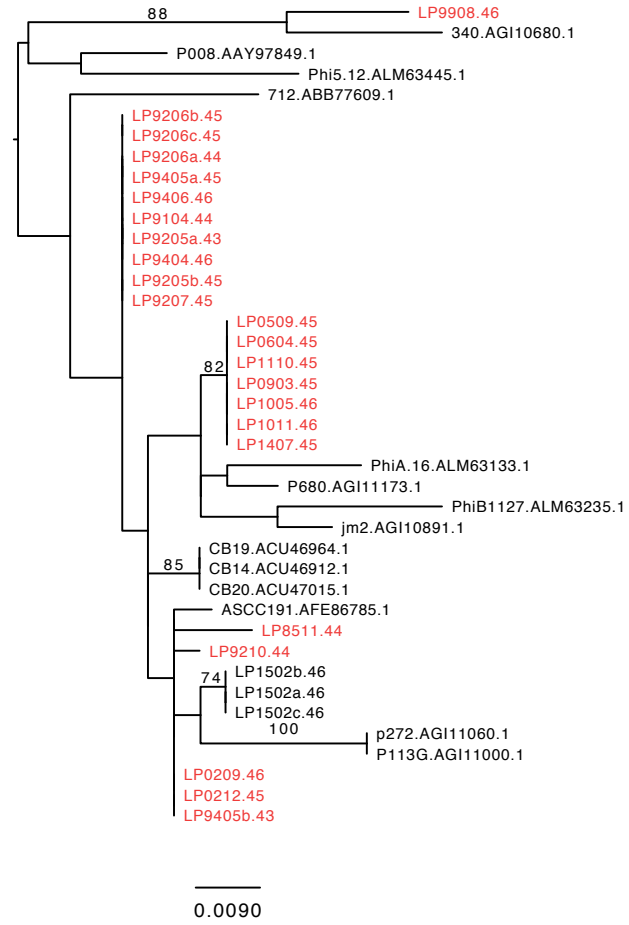


Figure S9: Gene phylogenies that conflict the monophyly of culture A lactococcal phages. All trees are unrooted. Proteins from culture A and culture B phages are labelled by the name and CDS number. Remaining proteins are labelled by name and protein ID. Culture A phages are highlighted in red. Only bootstrap support values above 70 are displayed. All phylogenies are shown where the monophyly of culture A phages is contradicted by a branch with bootstrap support of more than 70. (A) Family 2 – Phage terminase, large subunit. (B) Family 16 – Phage tail length tape measure protein. (C) Family 22 – Phage lysin. (D) Family 31 – Sak3. Family 31 shows similarity to ORF35 from phage p2 (blastp e-value≤2×10^{-64}) that was described as Sak3 (Bouchard and Moineau 2004). Sak3 is a DNA-single strand annealing protein involved in homologous recombination (Bouchard and Moineau 2004; Scaltriti et al. 2011). Sequences from family 31 were also detected as homologous to the eukaryotic recombination protein Rad52 using virfam (Lopes et al. 2010). Proteins from the Rad52 family promote recombination between phages and prophages by facilitating homologous recombination between divergent sequences (De Paepe et al. 2014). The alignment exhibits a high amount of conflict as displayed by the split network (Fig. S10A). The conflict can be traced back to different regions of the protein (Fig. S10). (E) Family 60 – conserved hypothetical protein. (F) Family 61 – HNH homing endonuclease. (G) Family 66 – DNA polymerase. Two protein families (64,66) are predicted to function as DNA polymerases, but they show less than 50% global protein identity. Exactly one gene from each family is present in each isolate of culture A phages at a conserved genomic location. In addition, family 66 shows a conflicting phylogeny. This implies the frequent exchange of multiple versions of DNA polymerase between phages of the 936 group.

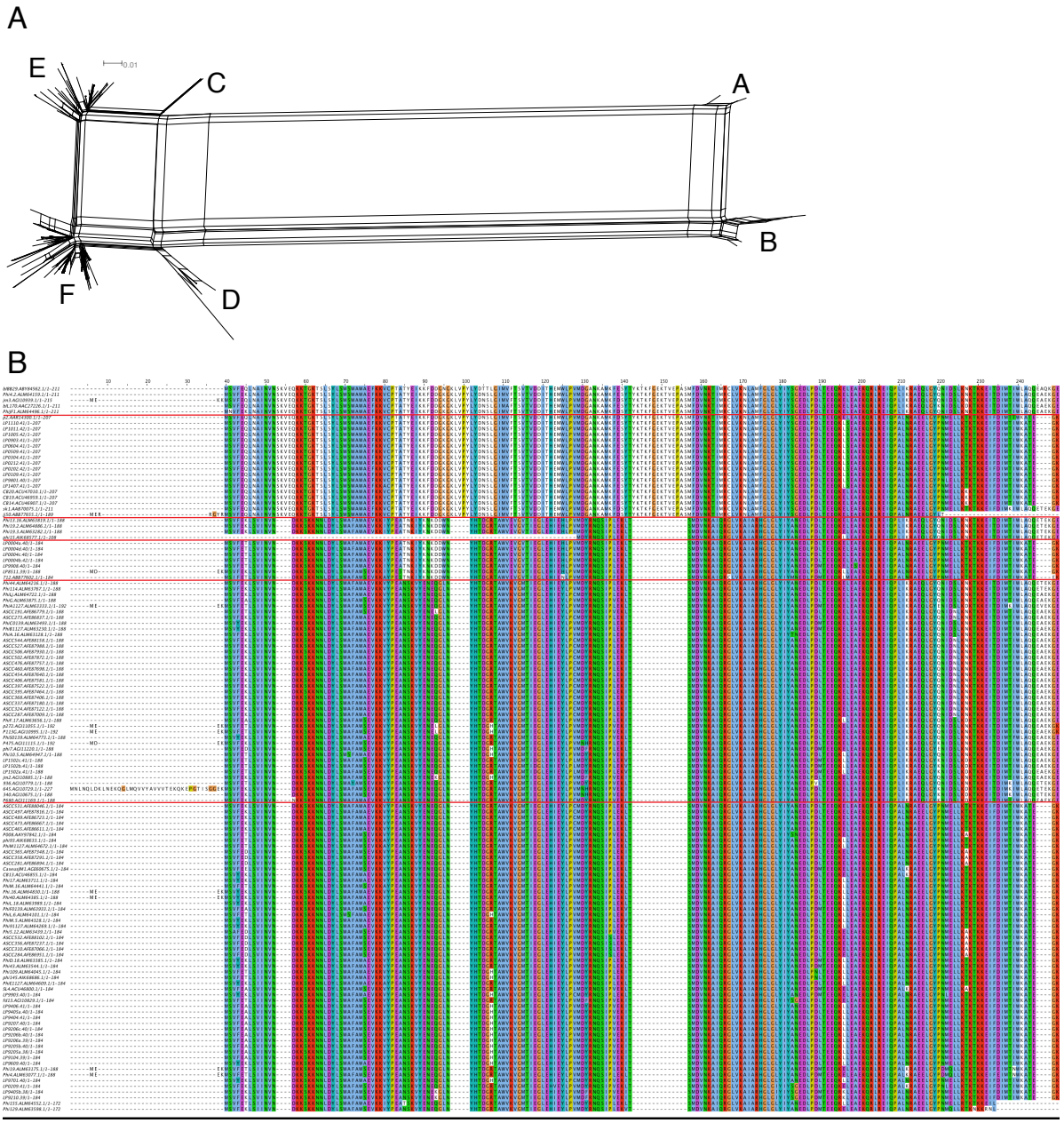


Figure S10: Alignment of Sak3 proteins from phages of the 936 group. (A) Neighbor-net based on ordinary least squares distances calculated with SplitsTree (Huson and Bryant 2006). (B) Alignment delimited into the taxon groups A, B, C, D, E, and F (top to bottom) as marked in the neighbor-net. Differences that split taxa A and B from the rest can be found in the N-terminal region (alignment positions 40-184), whereas differences that split taxa A, C, and E from B, D, and F can be found in the C-terminal region (alignment positions 212-250). ORF35 from phage p2 belongs to group B. Its N-terminal region contains the domains for oligomerization, DNA binding, and the motifs for ATPase activity that are important for the stimulation of RecA, whereas its C-terminal region might constitute the interaction domain with RecA (Scaltriti et al. 2011).

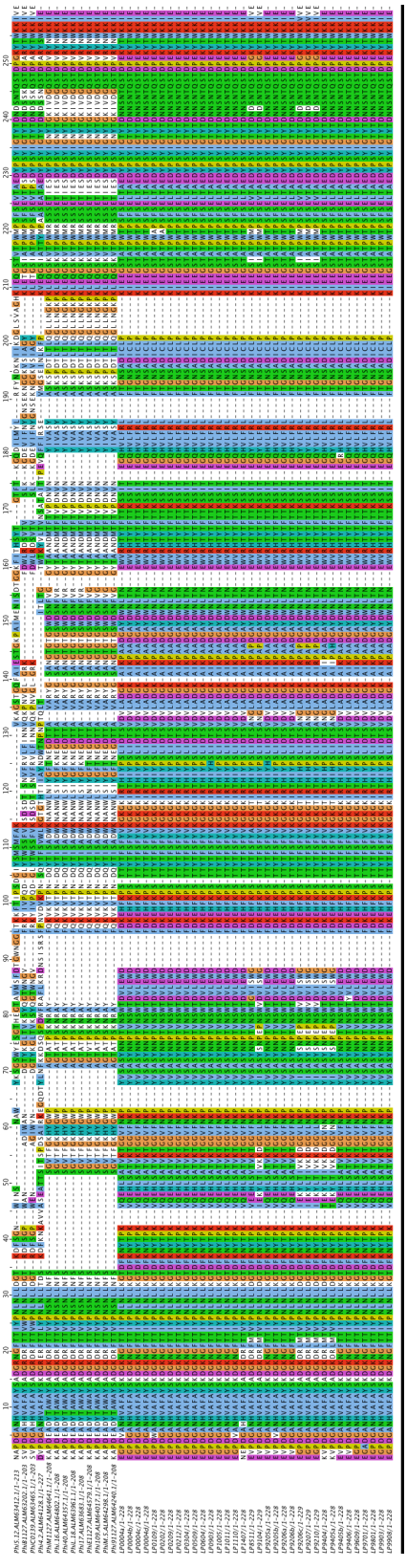


Figure S11: Protein alignment of TpeX region in phages of the 936 group. In the initial annotation, the tail protein extension TpeX (family 13) did not show a conserved start codon. The C-terminal region of the major tail protein is similar to other phages of the 936 group that contain a tail protein extension (Murphy et al. 2016). We thus conclude that family 13 is also an extension protein generated by a +1 frameshift. Notably, our version of TpeX only shows sequence similarity to previously described TpeX in the termini.

Table S1: Protein families.

Protein Family	Samples	Alignment length	% Recombination	Type	Annotation	Annotation evidence
1	34	174	7.5%	core (all)	Phage terminase, small subunit	RAST
2	34	540	72.0%	core (all)	Phage terminase, large subunit	RAST
3	34	94	0.0%	core (all)	Phage-associated homing endonuclease	RAST
4	34	378	0.0%	core (all)	Phage portal protein	RAST
5	34	178	100.0%	core (all)	Putative capsid maturation protein	Similarity to YP_762516.1
6	34	393	100.0%	core (all)	Major capsid protein	Similarity to ALM64064.1
7	34	87	0.0%	core (all)	Conserved hypothetical protein	
8	34	104	0.0%	core (all)	Probable head completion protein 1	Similarity to HCP1_BPLSK
9	34	113	0.0%	core (all)	Probable head completion protein 2	Similarity to HCP2_BPLSK
10	34	121	0.0%	core (all)	Probable tail terminator protein	Similarity to TTTP_BPLSK
11	34	656	33.8%	core	Neck passage structure	RAST, Similarity to ANS02693.1
12	34	301	72.1%	core (all)	Major tail protein	Similarity to ALM63411.1
13	34	228	71.5%	core	Tail protein extension	Similarity to ALM63412.1
14	34	91	0.0%	core (all)	Probable chaperone for the tape measure protein	Similarity to GP12_BPLSK
15	34	170	100.0%	core (all)	Conserved hypothetical protein	
16	34	919	44.0%	core (all)	Phage tail length tape measure protein	RAST
17	34	460	0.0%	core (all)	Distal tail protein	Similarity to ALM64750.1
18	34	379	0.0%	core	Tail associated lysin	Similarity to ALM63688.1
19	34	97	0.0%	core	Conserved hypothetical protein	
20	34	264	0.0%	core	Putative receptor binding protein	Similarity to AAT81500.1
21	34	117	0.0%	core (all)	Phage holin	RAST
22	34	263	3.8%	core (all)	Phage lysin	RAST
23	34	89	1.1%	core (all)	Putative transglycolase	Phagonaute based on similarity to sk1p21
24	34	115	58.3%	core (all)	Conserved hypothetical protein	
25	34	62	100.0%	core	Conserved hypothetical protein	
26	34	102	100.0%	core	SaV protein	Similarity to SAV_BPLSK
27	34	39	100.0%	core	Conserved hypothetical protein	
28	34	172	100.0%	core	Conserved hypothetical protein	
29	34	71	100.0%	core (all)	Conserved hypothetical protein	
30	34	119	100.0%	core (all)	Single stranded DNA-binding protein, phage associated	RAST
31	34	211	77.7%	core (all)	Sak3	Similarity to ALM64385.1
32	34	94	100.0%	core	Conserved hypothetical protein	
33	34	96	100.0%	core	Conserved hypothetical protein	
34	34	90	100.0%	core	Conserved hypothetical protein	
35	34	73	100.0%	core	Conserved hypothetical protein	

Protein Family	Samples	Alignment length	% Recombination	Type	Annotation	Annotation evidence
36	34	56	100.0%	core	Conserved hypothetical protein	
37	34	75	100.0%	core	Conserved hypothetical protein	
38	34	43	0.0%	core (all)	Conserved hypothetical protein	
39	34	160	16.3%	core (all)	Putative holliday junction endonuclease	Similarity to YP_762566.1
40	34	55	0.0%	core	Conserved hypothetical protein	
41	4	153	0.0%	gene indel	HNH homing endonuclease	RAST
42	1	-	-	nonsense	Neck passage structure	RAST
43	1	-	-	nonsense	Neck passage structure	RAST
44	1	-	-	nonsense	Neck passage structure	RAST
45	1	-	-	nonsense	Neck passage structure	RAST
46	1	-	-	gene indel	Putative lytic transglycosylase	Similarity to YP_002004010.1
47	31	53	0.0%	gene indel	Putative lytic transglycosylase	Similarity to B3TJQ7_9CAUD
48	19	59	0.0%	nonsense	Conserved hypothetical protein	
49	33	56	100.0%	nonsense	Conserved hypothetical protein	
50	7	82	100.0%	gene indel	Conserved hypothetical protein	
51	31	86	100.0%	gene indel	Conserved hypothetical protein	
52	13	81	0.0%	gene indel	Conserved hypothetical protein	
53	32	220	0.0%	gene indel	Methylase	Similarity to WP_062806946.1
54	2	107	0.0%	gene indel	Conserved hypothetical protein	
55	2	116	0.0%	gene indel	Putative excisionase	Similarity to A5GZ54_9CAUD
56	2	45	0.0%	nonsense	Methylase	Similarity to ALM64768.1
57	22	325	70.2%	gene indel	Methylase	Similarity to ALM64939.1
58	30	82	37.8%	gene indel	Conserved hypothetical protein	
59	32	57	0.0%	gene indel	Conserved hypothetical protein	
60	32	109	100.0%	gene indel	Conserved hypothetical protein	
61	23 (25)*	147	100.0%	gene indel	HNH homing endonuclease	RAST
62	11	169	100.0%	gene indel	HNH homing endonuclease	RAST
63	20	60	33.3%	gene indel	Conserved hypothetical protein	
64	11	316	2.8%	gene indel	DNA Polymerase	Similarity to AIK68583.1
65	5	145	0.0%	gene indel	HNH homing endonuclease	RAST
66	23	298	100.0%	gene indel	DNA polymerase	Similarity to YP_008320131.1
67	7	153	0.0%	gene indel	Phage anti-repressor protein	RAST
68	4	54	0.0%	nonsense	Phage anti-repressor protein	RAST
69	7	81	0.0%	gene indel	Conserved hypothetical protein	
70	5	42	0.0%	nonsense	Conserved hypothetical protein	
71	18	48	0.0%	gene indel	Conserved hypothetical protein	
72	1	-	-	gene indel	HNH homing endonuclease	RAST
73	32	62	100.0%	gene indel	Conserved hypothetical protein	
74	2	163	0.0%	gene indel	HNH homing endonuclease	RAST

Alignment length is the length of the protein alignment. Type marks core families of culture A phages and core families of all phages of the 936 group are denoted as “core (all)”. One family (marked *) occurs two times in two strains, thus the total number of sequences in the family is given in brackets.

Table S2: Substitution rate estimation with BEAST.

Data set	Model	Clock model	Population model	Substitution rate estimate [95% HPD interval] in substitutions/site/year	Marginal likelihood	Bayes factor
Masked whole-genome alignment	1	strict	constant	1.888 [1.470-2.317] x 10^{-4}	-22798.77	0
Masked whole-genome alignment	2	strict	exponential	1.890 [1.489-2.330] x 10^{-4}	-22800.48	-1.715
Masked whole-genome alignment	3	strict	Bayesian Skyline	1.900 [1.464-2.317] x 10^{-4}	-22798.89	-0.1258
Masked whole-genome alignment	4	strict	Extended Bayesian Skyline	1.889 [1.460-2.331] x 10^{-4}	-22798.29	0.4780
Masked whole-genome alignment	5	Lognormal	constant	1.915 [1.486-2.371] x 10^{-4}	-22803.45	-4.683
Masked whole-genome alignment	6	Random local clock	constant	1.889 [1.479-2.342] x 10^{-4}	-22821.26	-22.49
Core codon alignment	1	strict	constant	1.737 [1.329-2.143] x 10^{-4}	-	-
Core codon alignment, position 1	1	strict	constant	1.494 [0.9759-2.087] x 10^{-4}	-	-
Core codon alignment, position 2	1	strict	constant	0.8751 [0.5001-1.264] x 10^{-4}	-	-
Core codon alignment, position 3	1	strict	constant	2.887 [2.107 - 3.770] x 10^{-4}	-	-
Core codon alignment, position 1	6	Random local clock	constant	1.525 [0.9690-2.105] x 10^{-4}	-	-
Core codon alignment, position 2	6	Random local clock	constant	0.8892 [0.5160-1.301] x 10^{-4}	-	-
Core codon alignment, position 3	6	Random local clock	constant	2.927 [2.148 - 3.823] x 10^{-4}	-	-

Substitution rate estimate is mean rate for lognormal and random local clock models. Bayes factors compare H1 to H0 with H0 being the null model of a strict clock and a constant population size. Complex population size models have an absolute Bayes factor difference of less than two to the basic model. For the exponential growth model (model 2), the growth rate is estimated as -0.032 [-0.11,0.040]. This interval includes 0, thus the data is compatible with constant population size. Non-strict clock models are clearly rejected in comparison to the strict clock model (Bayes Factor > 4). For the random local clock model (model 6), the rate indicator estimate is 0.48 [0,2] and the highest posterior probability is at zero rate changes, thus the data is compatible with a strict molecular clock. The random local clock model based on a partition of the core codon alignment supports a strict molecular clock for each codon position, since the rate indicator estimates are 0.88 [0,2] for position 1, 0.78 [0,2] for position 2 and 0.50 [0,2] for position 3.

Table S3: Publicly available genomes for *L. lactis* phages of the 936 group.

Accession	Phage
AF009630	bIL170
AF011378	sk1
DQ054536	P008
DQ227763	712
DQ227764	jj50
EU221285	bIBB29
FJ848881	SL4
FJ848882	CB13
FJ848883	CB14
FJ848884	CB19
FJ848885	CB20
GQ979703	p2
JQ740787	ASCC191
JQ740788	ASCC273
JQ740789	ASCC281
JQ740790	ASCC284
JQ740791	ASCC287
JQ740792	ASCC310
JQ740793	ASCC324
JQ740794	ASCC337
JQ740795	ASCC356
JQ740796	ASCC358
JQ740797	ASCC365
JQ740798	ASCC368
JQ740799	ASCC395
JQ740800	ASCC397
JQ740801	ASCC406
JQ740802	ASCC454
JQ740803	ASCC460
JQ740804	ASCC465
JQ740805	ASCC473
JQ740806	ASCC476
JQ740807	ASCC489
JQ740808	ASCC497
JQ740809	ASCC502
JQ740810	ASCC506
JQ740811	ASCC527
JQ740812	ASCC531
JQ740813	ASCC532
JQ740814	ASCC544
KC182542	340
KC182543	645
KC182544	936
KC182545	fd13
KC182546	jm2
KC182547	jm3
KC182548	P113G

Accession	Phage
KC182549	p272
KC182550	P475
KC182551	P680
KC182552	phi7
KC522412	CaseusJM1
KM091442	phi15
KM091443	phi93
KM091444	phi145
KP793101	Phi4
KP793102	PhiA.16
KP793103	Phi19
KP793104	PhiB1127
KP793105	Phi19.3
KP793106	PhiA1127
KP793107	PhiD.18
KP793108	Phi5.12
KP793109	PhiC0139
KP793110	Phi43
KP793111	Phi19.2
KP793112	Phi129
KP793113	PhiF.17
KP793114	Phi17
KP793115	Phi114
KP793116	Phi13.16
KP793117	PhiG
KP793118	PhiF0139
KP793119	Phi10.5
KP793120	PhiL.18
KP793121	Phi109
KP793122	PhiL.6
KP793123	Phi4.2
KP793124	Phi44
KP793125	Phi91127
KP793126	PhiM.5
KP793127	Phi40
KP793128	PhiM.16
KP793129	PhiJF1
KP793130	Phi155
KP793131	PhiE1127
KP793132	PhiM1127
KP793133	PhiLj
KP793134	PhiS0139
KP793135	Phi.16

References

Bouchard JD, Moineau S. 2004. Lactococcal phage genes involved in sensitivity to AbiK and their relation to single-strand annealing proteins. J. Bacteriol. 186:3649–3652.

De Paepe M, Hutinet G, Son O, Amarir-Bouhram J, Schbath S, Petit M-A. 2014. Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. PLOS Genet. 10:e1004181.

Hecht A, Glasgow J, Jaschke PR, Bawazer LA, Munson MS, Cochran JR, Endy D, Salit M. 2017. Measurements of translation initiation from all 64 codons in *E. coli*. Nucleic Acids Res. 45:3615–3626.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254–67.

Lopes A, Amarir-Bouhram J, Faure G, Petit M-A, Guerois R. 2010. Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. Nucleic Acids Res. 38:3952–62.

Mavrich TN, Hatfull GF. 2017. Bacteriophage evolution differs by host, lifestyle and genome. Nat. Microbiol. 2:17112.

Murphy J, Bottacini F, Mahony J, Kelleher P, Neve H, Zomer A, Nauta A, van Sinderen D. 2016. Comparative genomics and functional analysis of the 936 group of lactococcal *Siphoviridae* phages. Sci. Rep. 6:21345.

Scaltriti E, Launay H, Genois M-M, Bron P, Rivetti C, Grolli S, Ploquin M, Campanacci V, Tegoni M, Cambillau C, et al. 2011. Lactococcal phage p2 ORF35-Sak3 is an ATPase involved in DNA recombination and AbiK mechanism. Mol. Microbiol. 80:102–116.