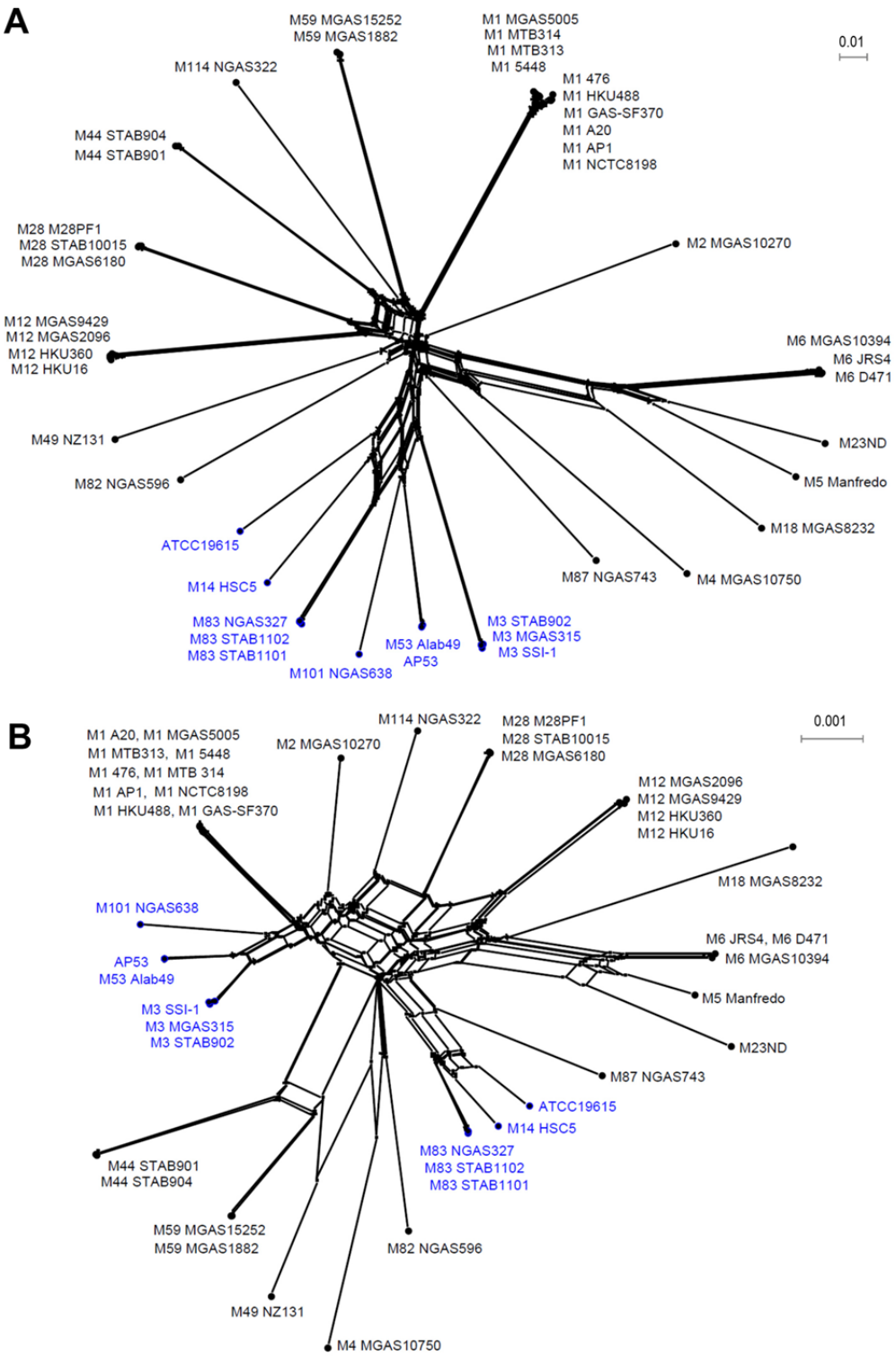


FIG S1 The phylogenetic network structure of GAS strains constructed from pair-wise distances of all GAS genomes based on the SNPs. The networks were determined for: (A) the core genome of GAS, containing 71,558 SNP loci; (B) seven housekeeping genes, *i.e.*, *gki*, *gtr*, *murI*, *mutS*, *recP*, *xpt*, and *yqiL*, generating 434 SNP loci. Skin-tropic strains, *i.e.*, AP53, Alab49, M101 NGAS638, M83 strains, M3 strains, ATCC19615, and M14 HSC5, were clustered together in (A), whereas re-distributed in distinct clusters in (B). Those strains were highlighted with blue. The pair-wise distance was estimated based on the model of Maximum Composite Likelihood and the network reconstruction was implemented by Neighbornet in SplitsTree.

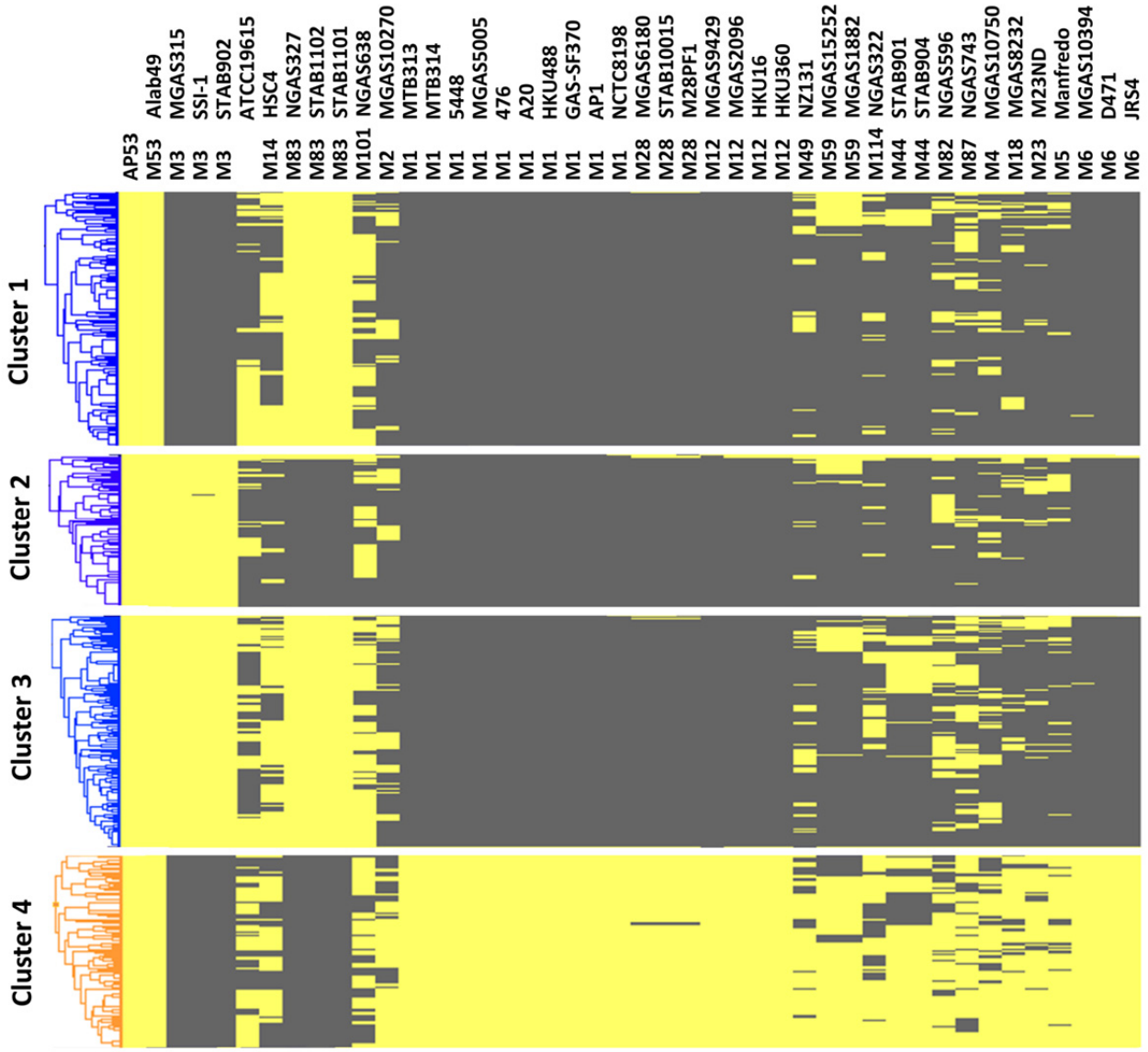
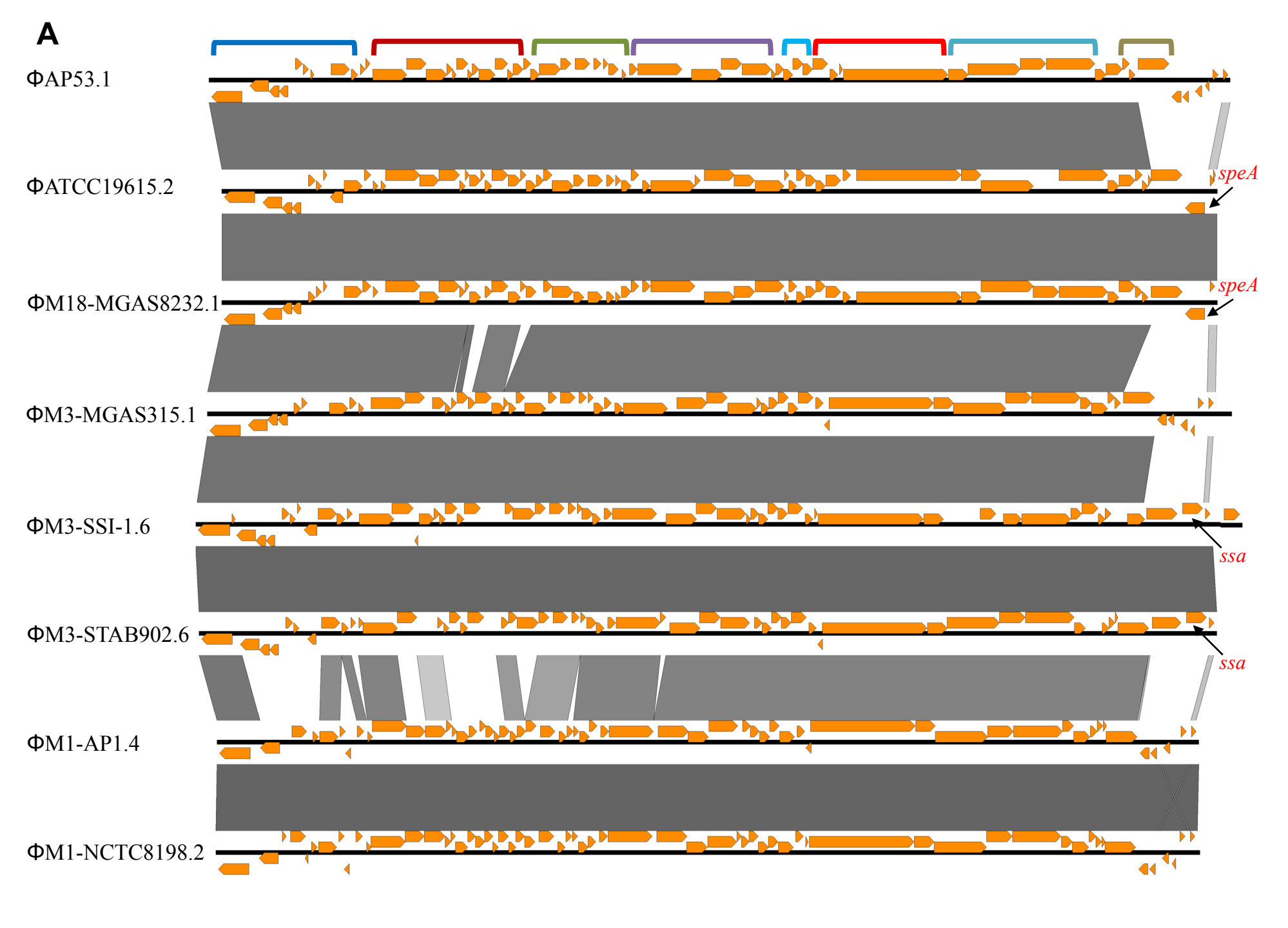
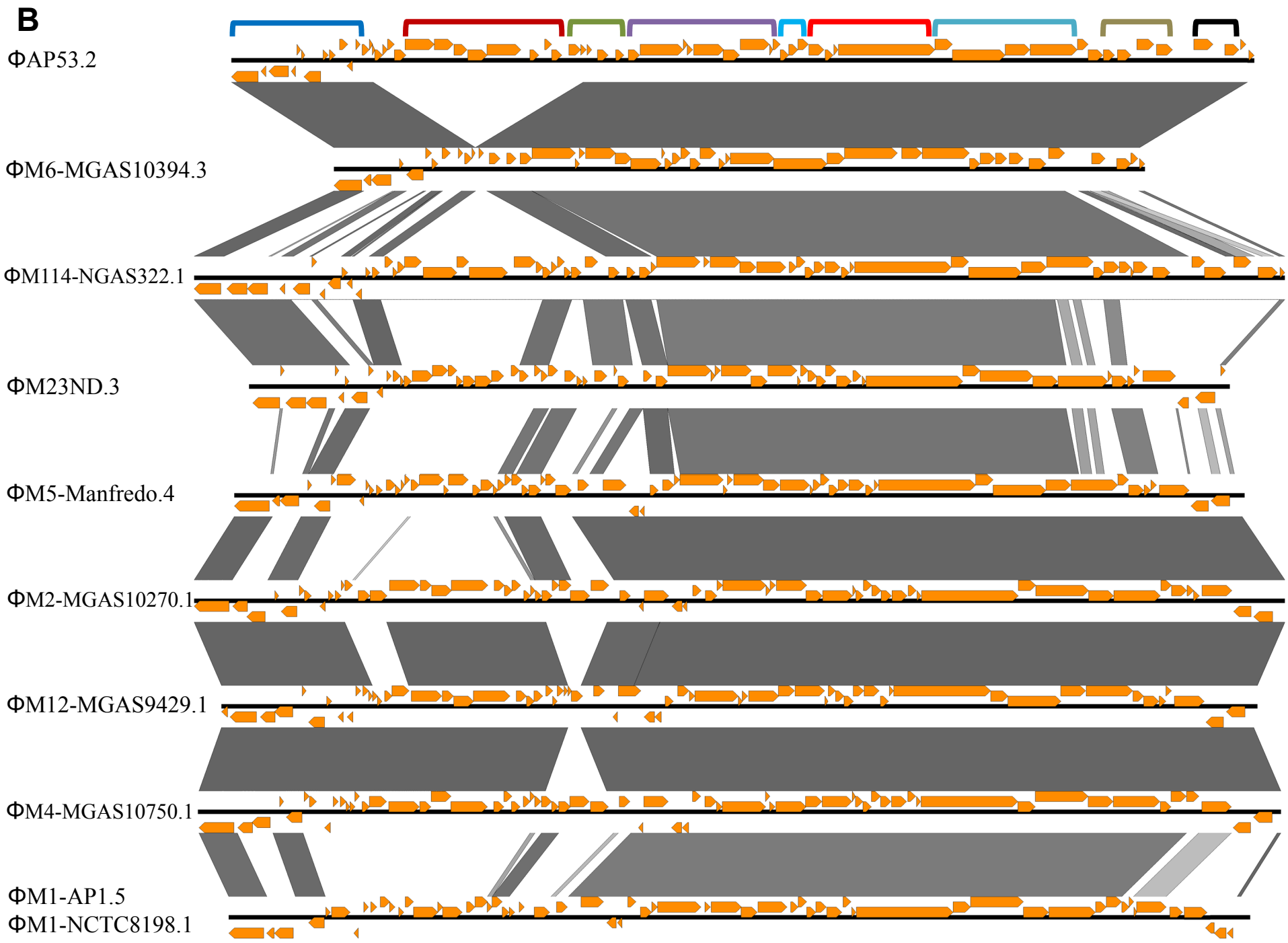
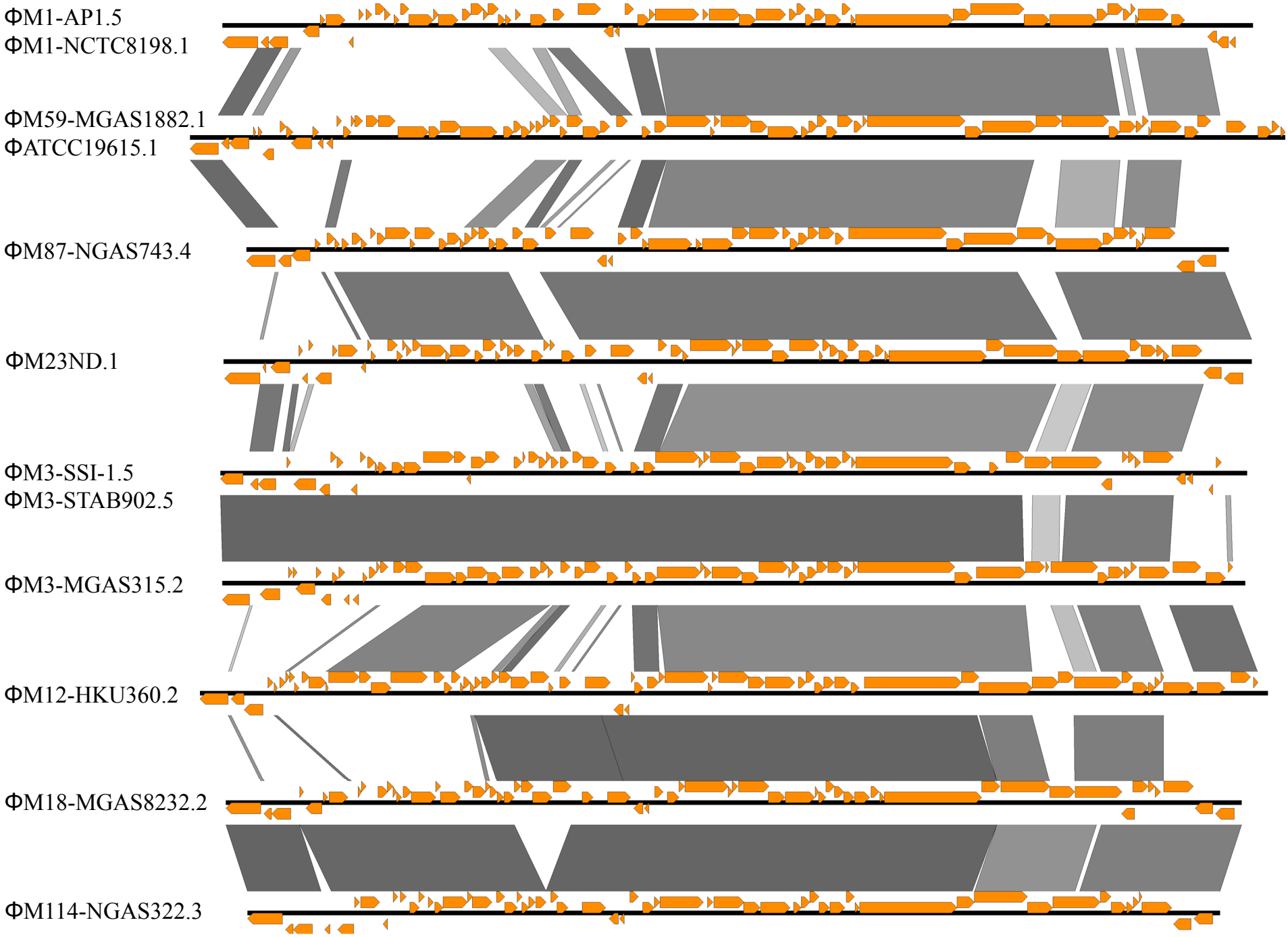
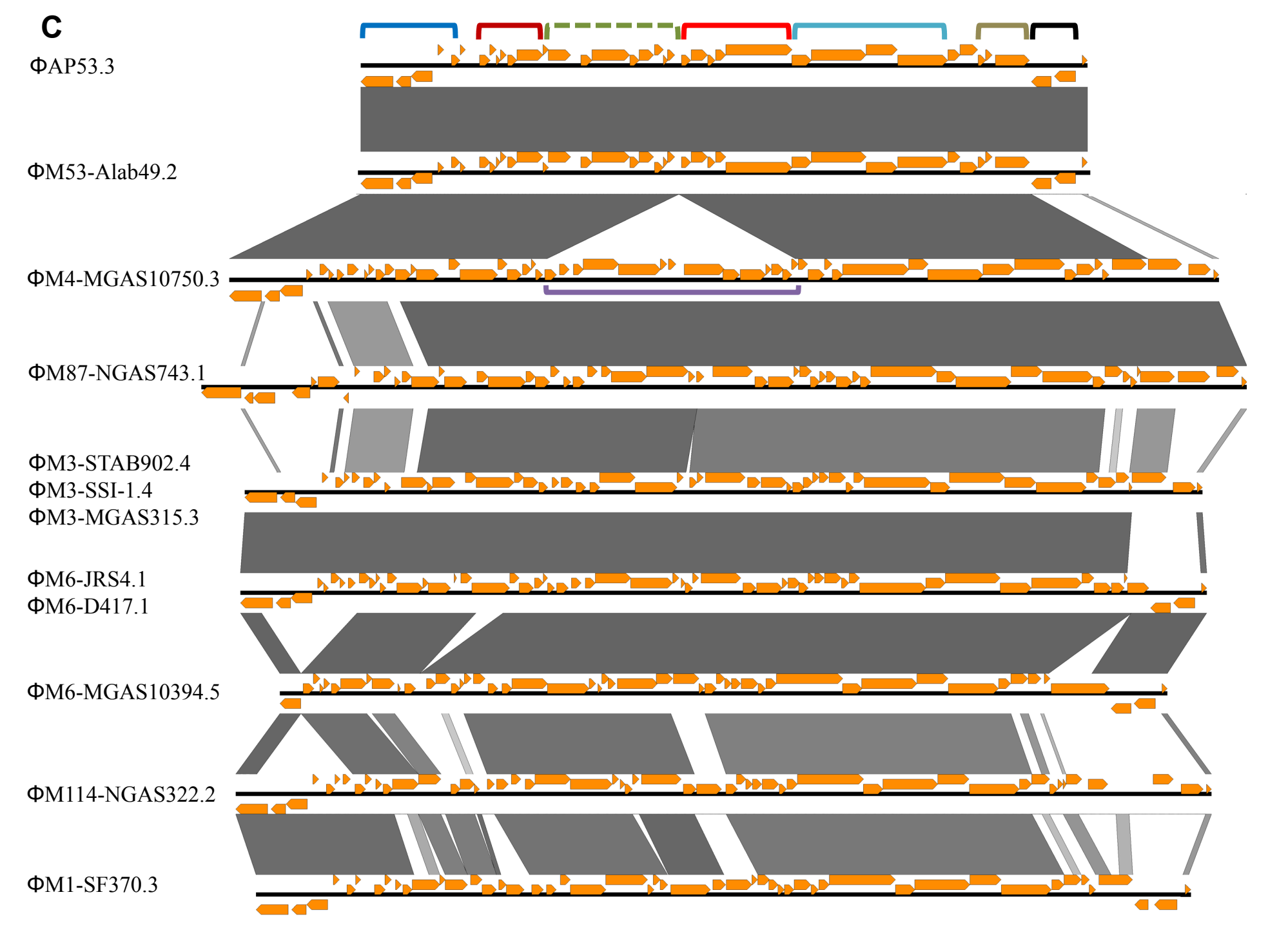


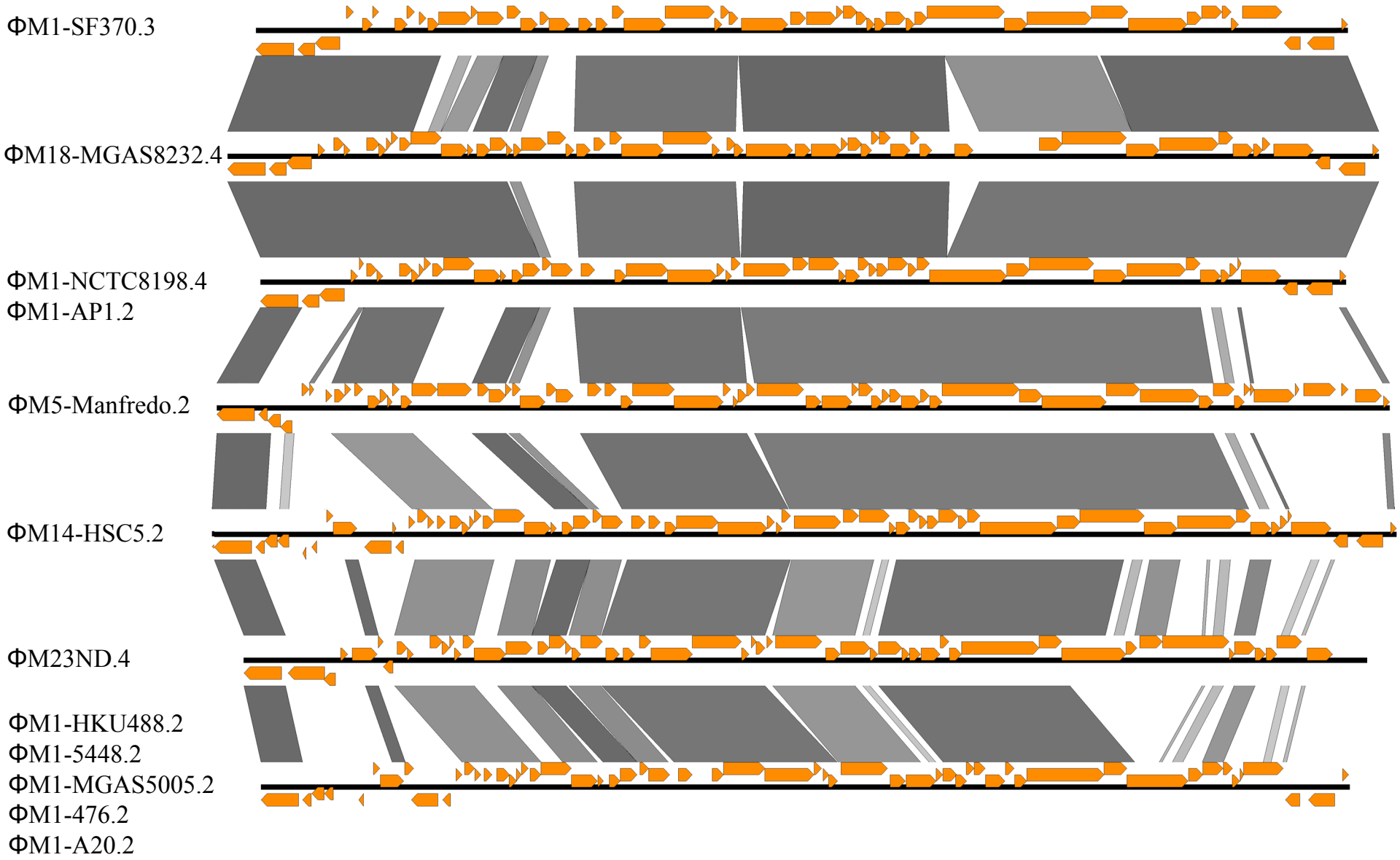
FIG S2 The candidate skin-specific SNPs assessed by clustering of the non-synonymous SNP allele patterns for all the compared genomes. Four clusters were inferred: (1) the SNP alleles present in strains of M53 and M83; (2) the SNP alleles present in strains of M53 and M3; (3) the SNP alleles present in strains of M53, M83 and M3; (4) The SNP alleles present in strains of M83 and M3. The simultaneous occurrence of the alleles in different strains is indicated by the same color. The detailed description of SNP alleles at each locus can be found in Supplementary Dataset 1.

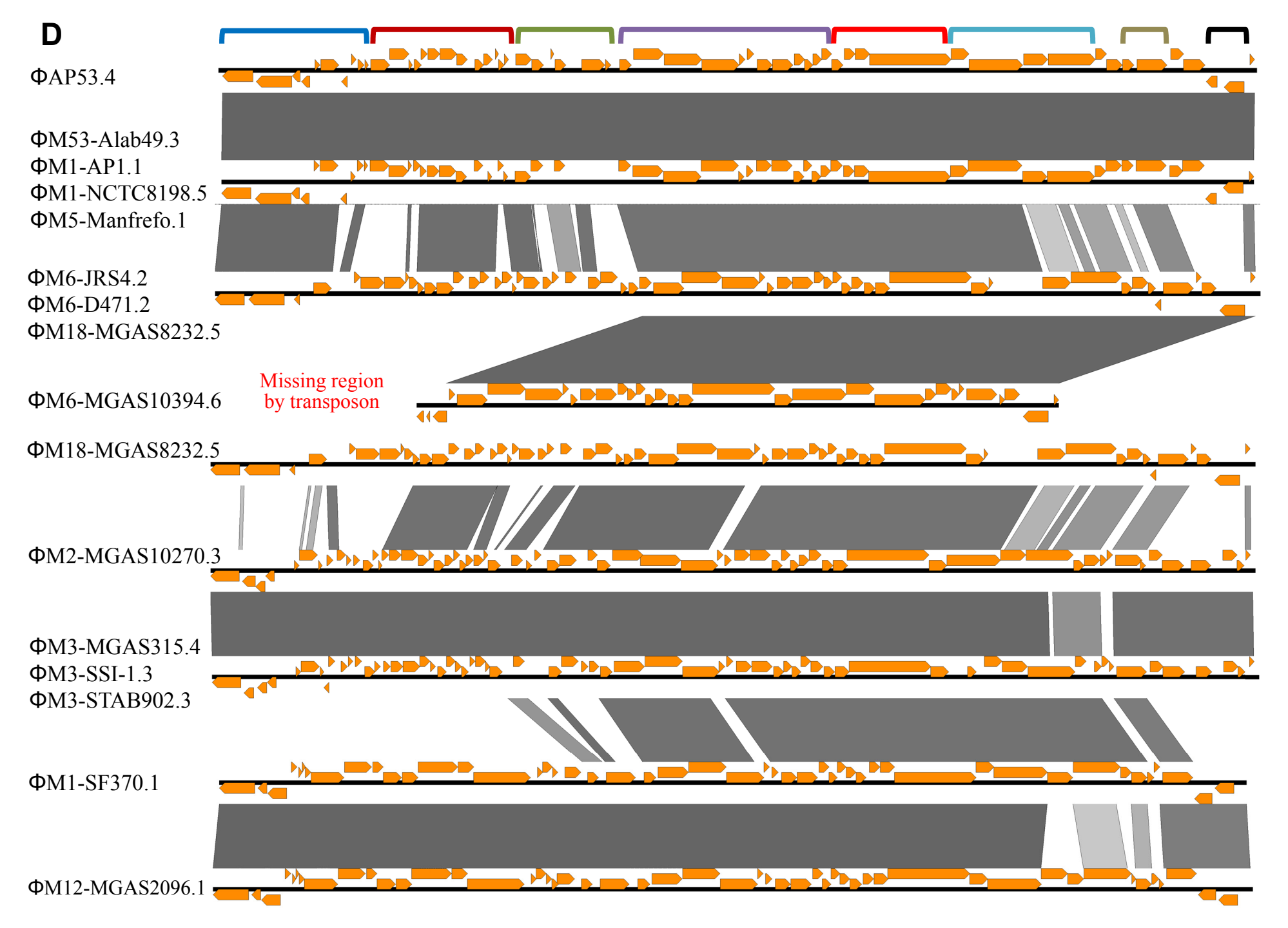

A
ΦAP53.1
ΦATCC19615.2
speA
ΦM18-MGAS8232.1
speA
ΦM3-MGAS315.1
ΦM3-SSI-1.6
ssa
ΦM3-STAB902.6
ssa
ΦM1-AP1.4
ΦM1-NCTC8198.2

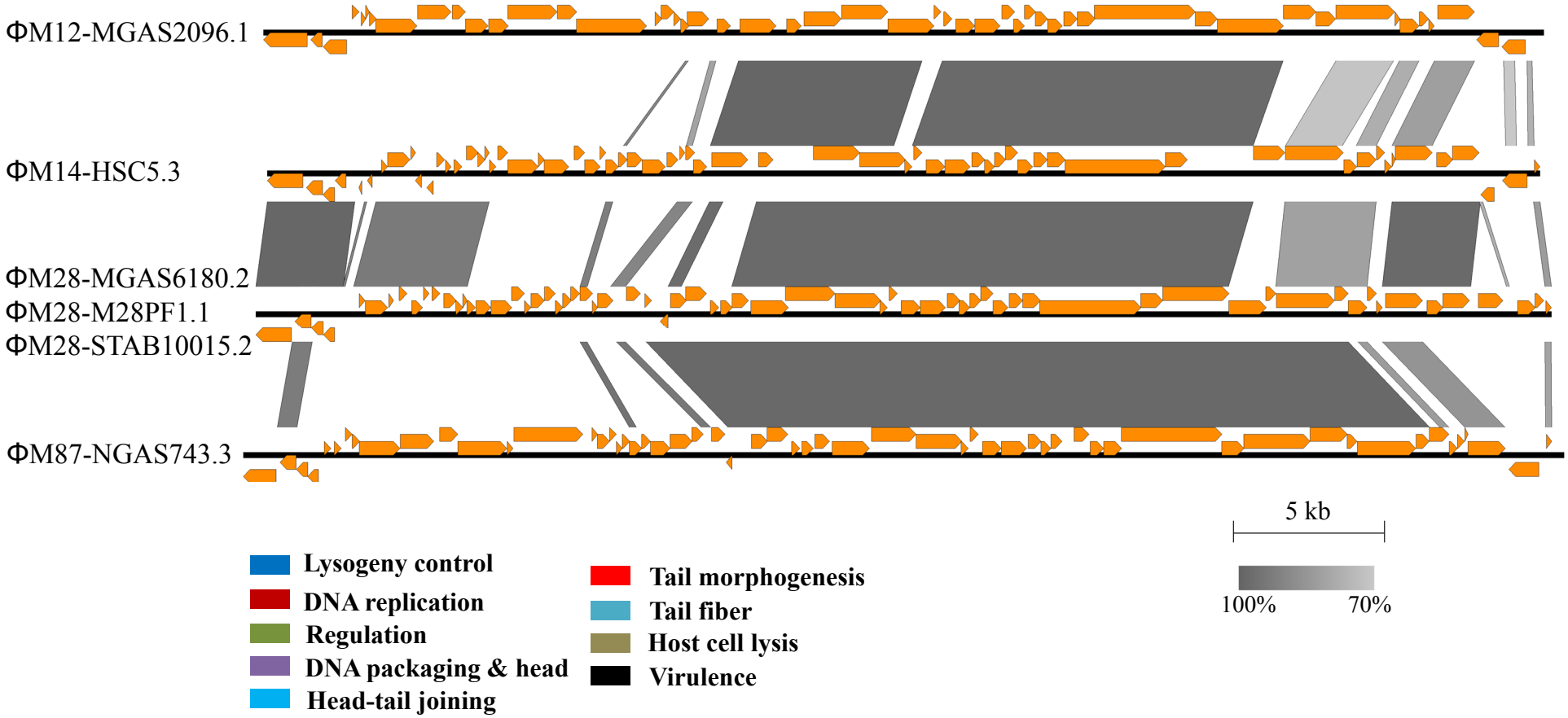

B
ΦAP53.2
ΦM6-MGAS10394.3
ΦM114-NGAS322.1
ΦM23ND.3
ΦM5-Manfredo.4
ΦM2-MGAS10270.1
ΦM12-MGAS9429.1
ΦM4-MGAS10750.1
ΦM1-AP1.5
ΦM1-NCTC8198.1

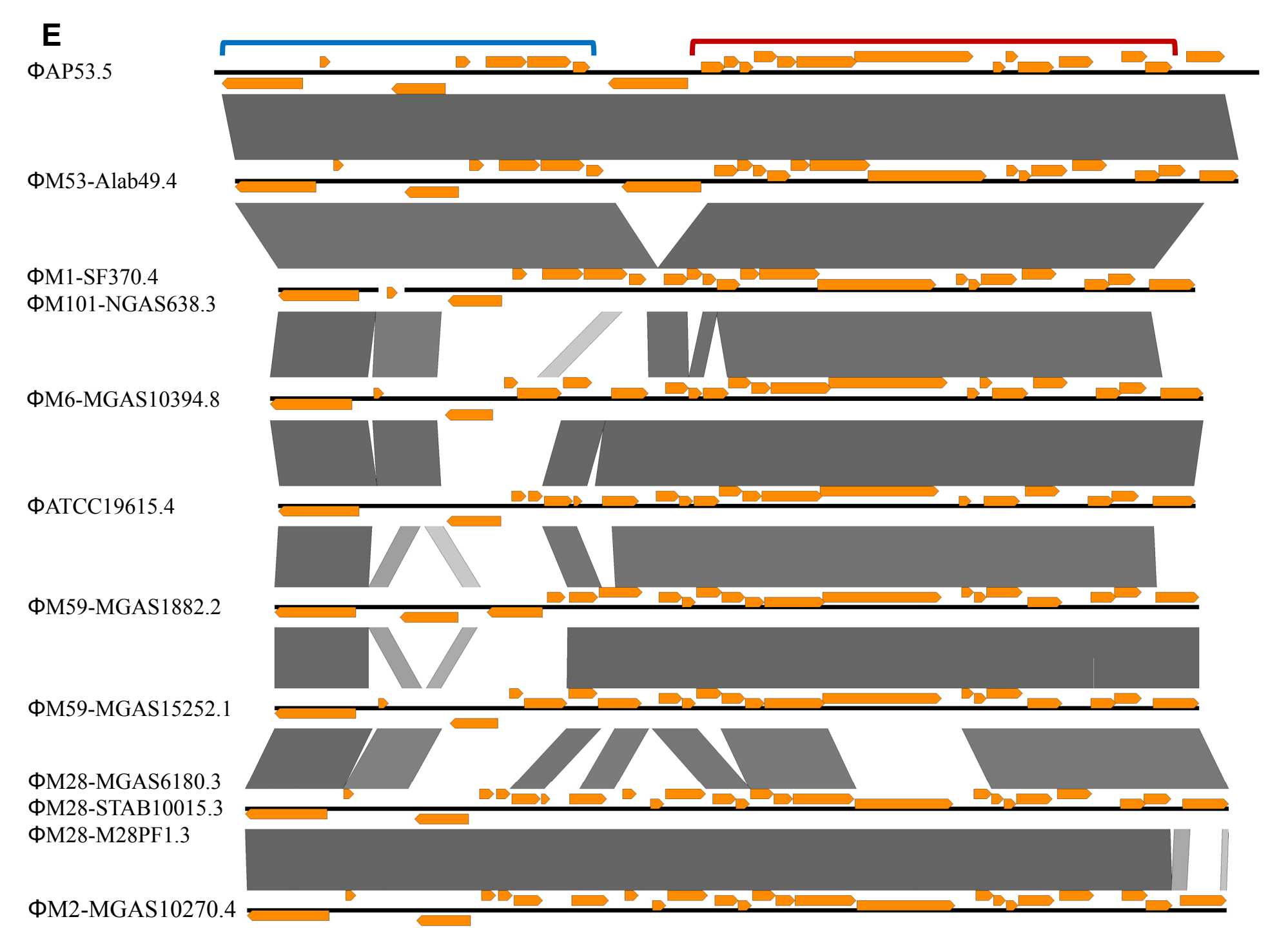

ΦM1-AP1.5
ΦM1-NCTC8198.1
ΦM59-MGAS1882.1
ΦATCC19615.1
ΦM87-NGAS743.4
ΦM23ND.1
ΦM3-SSI-1.5
ΦM3-STAB902.5
ΦM3-MGAS315.2
ΦM12-HKU360.2
ΦM18-MGAS8232.2
ΦM114-NGAS322.3


C
ΦAP53.3
ΦM53-Alab49.2
ΦM4-MGAS10750.3
ΦM87-NGAS743.1
ΦM3-STAB902.4
ΦM3-SSI-1.4
ΦM3-MGAS315.3
ΦM6-JRS4.1
ΦM6-D417.1
ΦM6-MGAS10394.5
ΦM114-NGAS322.2
ΦM1-SF370.3


ΦM1-SF370.3
ΦM18-MGAS8232.4
ΦM1-NCTC8198.4
ΦM1-AP1.2
ΦM5-Manfredo.2
ΦM14-HSC5.2
ΦM23ND.4
ΦM1-HKU488.2
ΦM1-5448.2
ΦM1-MGAS5005.2
ΦM1-476.2
ΦM1-A20.2


D
ΦAP53.4
ΦM53-Alab49.3
ΦM1-AP1.1
ΦM1-NCTC8198.5
ΦM5-Manfrefo.1
ΦM6-JRS4.2
ΦM6-D471.2
ΦM18-MGAS8232.5
ΦM6-MGAS10394.6
Missing region
by transposon
ΦM18-MGAS8232.5
ΦM2-MGAS10270.3
ΦM3-MGAS315.4
ΦM3-SSI-1.3
ΦM3-STAB902.3
ΦM1-SF370.1
ΦM12-MGAS2096.1


ΦM12-MGAS2096.1
ΦM14-HSC5.3
ΦM28-MGAS6180.2
ΦM28-M28PF1.1
ΦM28-STAB10015.2
ΦM87-NGAS743.3
5 kb
100%
70%
Lysogeny control
DNA replication
Regulation
DNA packaging & head
Head-tail joining
Tail morphogenesis
Tail fiber
Host cell lysis
Virulence


E
ΦAP53.5
ΦM53-Alab49.4
ΦM1-SF370.4
ΦM101-NGAS638.3
ΦM6-MGAS10394.8
ΦATCC19615.4
ΦM59-MGAS1882.2
ΦM59-MGAS15252.1
ΦM28-MGAS6180.3
ΦM28-STAB10015.3
ΦM28-M28PF1.3
ΦM2-MGAS10270.4

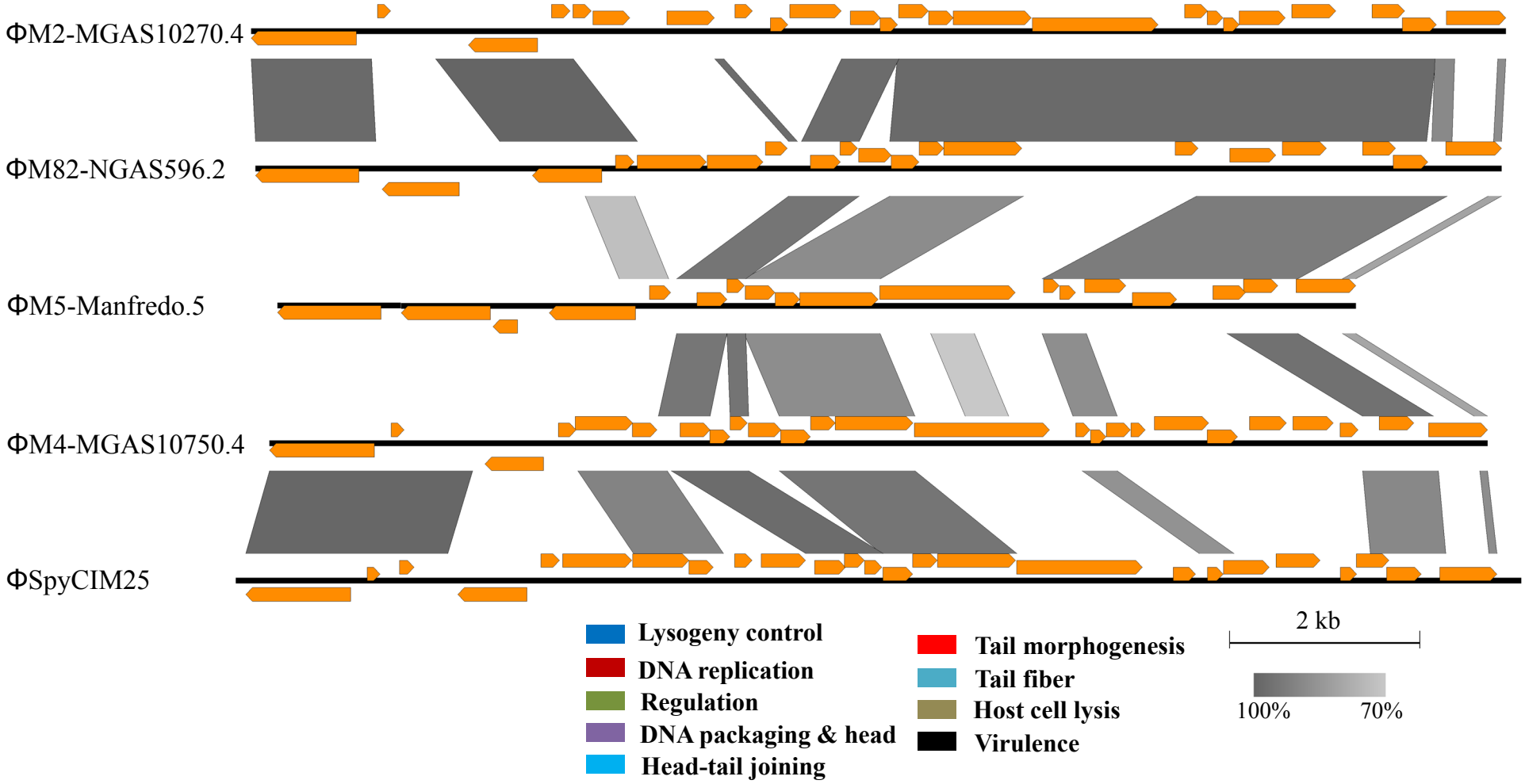


FIG S3 Multiple sequence alignments of homologous phages of ΦAP53.1 (A), ΦAP53.2 (B), ΦAP53.3 (C), ΦAP53.4 (D), and ΦAP53.5 (E). The sequence similarities are shown in gradient greyscale from 70%-100%. The modular functions of the aligned genes are indicated by brackets and color-coding. The phages within the same homologous groups contain conserved regions in the module of head and tail morphogenesis, although divergent in other regions, most significantly in DNA replication, host lysis, and virulence. Specifically, the group of ΦAP53.1-like phages are more conserved across the whole span of the phage region, except in the module of lysis and virulence, where the sequences are highly mosaic and the homology terminates abruptly. This indicates that the recombination events occurred in this region, resulting in the deletion of toxin genes in some of the host strains, including AP53.

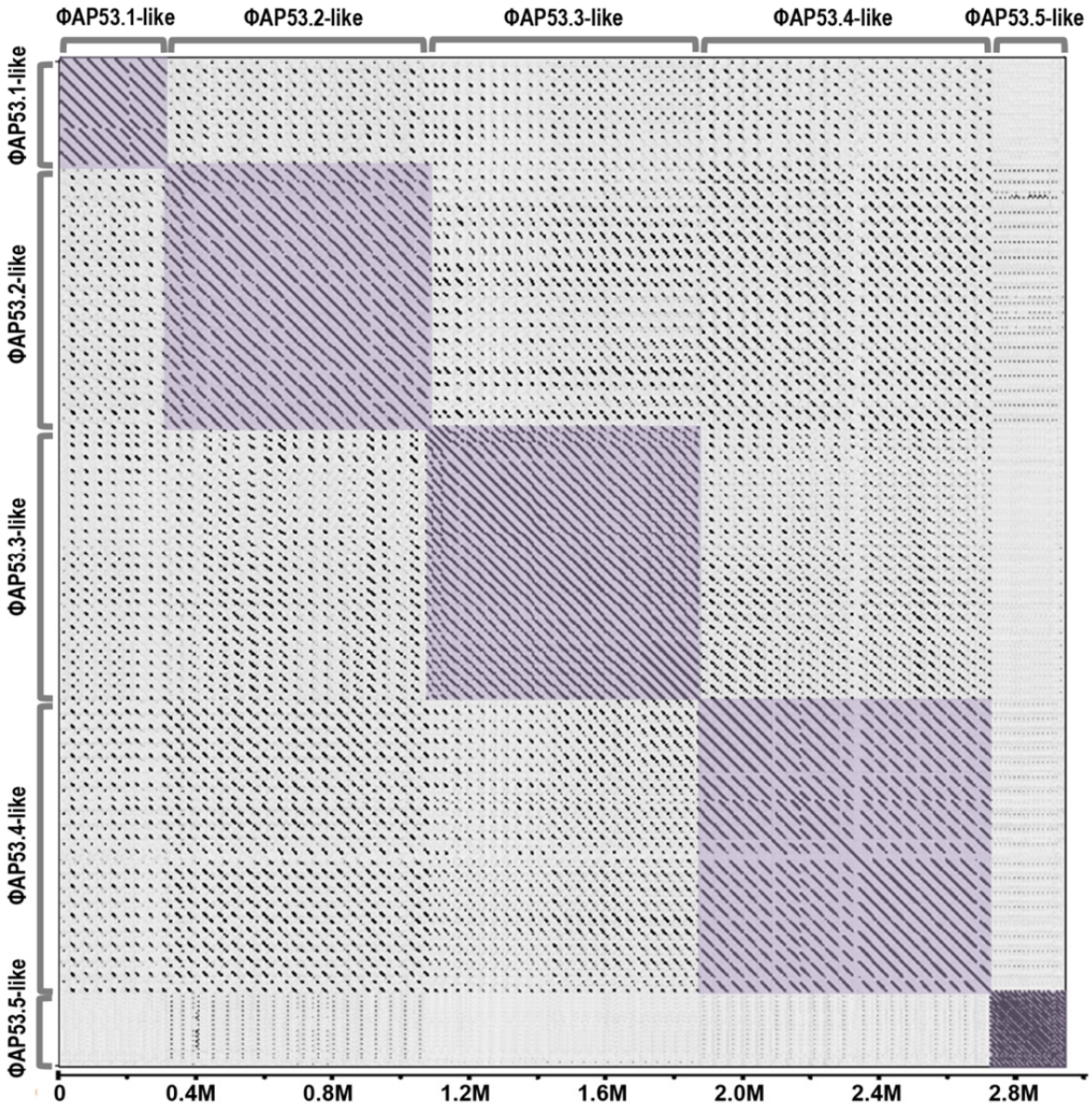


FIG S4 Dot-plot representation of multiple sequence comparison of phages from AP53 and their homologues. The diagonal lines indicate high levels of sequence identity and the discontinuous dots represent sporadic similarities. Five phages encoded by AP53 fall into distinct homology clusters (in purple shading), and only sporadic similarities were detected between different clusters in the segments of tail fiber genes (or hyaluronidase) (represented by the dots).

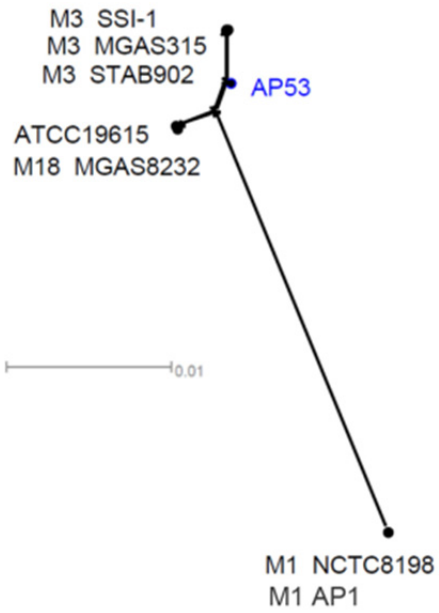


FIG S5 Phylogenetic network representation of phages from the ΦAP53.1-like group. The phylogeny structures were built based on multiple sequence alignment of conserved regions in the gene module of head and tail morphogenesis. A two-branch structure was formed, similar to that for the groups of ΦAP53.5, ΦAP53.3 and ΦAP53.4-like phages.

TABLE S1 Primers used for qRT-PCR experiments.

Primer	Sequence (5'−>3')	Sequence position[#] (Length)
spd3-F	ACAAGTACTGTTACGGCAGCCAG	73-379 (307)
spd3-R	GGTAACCAACTAAATGGCCACGG	
speC-F	CGGTGAGTACATCTATGGAGGAAT	327-562 (236)
speC-R	CAATTTCGATTCTGCCGCTTACA	
spd1-F	GAATGTATCCCATGCAAACACAC	108-345 (238)
spd1-R	CCATGAAGATTTTGAGCCATCTC	
speK-F	GCCAAACAAGGAACGCAATTGAT	220-677 (458)
speK-R	GTGTGTCTAATGCCACCGTCTT	
slaA-F	AGCTTCCAAAATGGTGAACTCCT	127-473 (347)
slaA-R	ACTAGCTGACGGTTACATTCACA	
plr-F	GACGGTACTGAAACAGTTATCTC	241-655 (234)
plr-R	GATAGCTTTAGCAGCACCAGTTG	

The numbering is from the position of start codon of the specific gene. Length refers to the amplicon length.

TABLE S2 The list of SNPs detected by the genomic comparison of two M53 strains, *i.e.*, AP53 and Alab49.

GeneID	AA change[1]	Function
SNPs unique to Alab49		
98	Leu122Ser	tRNA-binding protein
127	Leu429Ile	V-type ATP synthase subunit B
209	Gln89Arg	Putative deoxyribonuclease YcfH
415	Asn33Ser	Membrane protein
574	His60Tyr	Phenylalanyl-tRNA synthetase beta chain
920	Lys248Glu	Tyrosine recombinase XerC
940	Arg198Ser	Luciferase-like monooxygenase
973	Arg863His	ABC transporter permease protein
1110	Ala242Thr	5-Enolpyruvylshikimate-3-phosphate synthase
1163	Glu22Lys	Cold-shock DEAD-box protein A
1291	Trp643Stp	Beta-galactosidase
1400	Ile64Thr	NAD synthetase
1429	Met523Ile	Glycyl-tRNA synthetase beta chain
1430	Phe7Leu	Glycyl-tRNA synthetase alpha chain
1510	Ser229Tyr	ABC transporter permease protein
1564	Ser143Pro	Aquaporin (Major Intrinsic Protein Family)
1579	Ile7Thr	SSU ribosomal protein S14p (S29e)
1716	Lys22Glu	Glycerol dehydrogenase
1764	Ser176Arg	UPF0246 protein YaaA
1834	Ala48Thr	Phosphoesterase, DHH family protein
SNPs unique to AP53		
130	Lys142Asn	Hypothetical protein
151	His401Asn	DNA polymerase I
261	Gly67Glu	Trk system potassium uptake protein TrkA
511	Phe43Leu	Unsaturated glucuronyl hydrolase
686	Tyr171Asp	Uracil phosphoribosyltransferase
925	Val210Ala	Hypothetical protein
926	Cys291Arg	GMP synthase amidotransferase subunit
979	Ile125Thr	General stress protein, Gls24 family
1226	Val178Ala	Exodeoxyribonuclease VII large subunit
1492	Val172Met	Heat shock protein GrpE
1679[2]	Met221Ile	Chromosome segregation helicase

1. The column "AA change" indicates the amino acid change caused by the mutation. The numbers represent the codon position in the translated proteins.
2. The SNP in this gene is not unique to Alab49 or AP53.

TABLE S3. The homologous phages of each of the five phages encoded by AP53 and the corresponding integration sequences

GAS strain	Phage Order number	Integration sequences
Homologous phages of ΦAP53.1		
AP53	1	Non-identifiable[1]
ATCC19615	2	Non-identifiable
M18 MGAS8232	1	Non-identifiable
M3 MGAS315	1	Non-identifiable
M3 SSI-1	6	Non-identifiable
M3 STAB902	6	Non-identifiable
M1 AP1	4	Non-identifiable
M1 NCTC8198	2	Non-identifiable
Homologous phages of ΦAP53.2		
AP53	2	ACTCCCACCAGCTCCATCAATGCTTACCGTAAGTAATCATAACTTACTA AAACCTTGTTACATCAAGGTTTTTTCTTTTTGTCTTGTTCATGAGTT
M6 MGAS10394	3	ACTCCCACCAGCTCCATCAATGCTTACCGTAAGTAATCATAACTTACTA AAACCTTGTTACATCAAGGTTTTTTCTTTTTGTCTTGTTCATGAGTT
M114 NGAS322	1	ACTCCCACCAGCTCCATCAATGCTTACCGTAAGTAATCATAACTTACTA AAACCTTGTTACATCAAGGTTTTTTCTTTTTGTCTTGTTCATGAGTT
M23_M23ND	3	ACTCCCACCAGCTCCATCAATGCTTACCGTAAGTAATCATAACTTACTA AAACCTTGTTACATCAAGGTTTTTTTCTTTTTGTCTTGTTCATGAGTTTCC
M5 Manfredo	4	CATGTACAACTATACTCGT
M2 MGAS10270	1	CATGTACAACTATACTCGT
M12 MGAS9429	1	CATGTACAACTATACTCGT
M4 MGAS10750	1	CATGTACAACTATACTCGT
M114 NGAS322	3	CATGTACAACTATACTCGT
M1 AP1	5	CATGTACAACTATACT[CGT/---][2]
M1 NCTC8198	1	CATGTACAACTATACT[---/CGT][2]
ATCC19615	1	AATTATTTAACAGCGTCTTT
M59 MGAS1882	1	AATTATTTAACAGCGTCTTT
M87 NGAS743	4	AATTATTTAACAGCGTCTTT[3]
M23 M23ND	1	CATGTACAACTATACTC
M3 SSI-1	5	ACTCCCACCGGCTCCATCAATGCTTACCGTAAGTAATCATAACTTACTA AAACCTTGTTACATCAAGGTTTTTTCTTTTTGTCTTGTTCATGAGTT[3]
M3 STAB902	5	ACTCCCACCGGCTCCATCAATGCTTACCGTAAGTAATCATAACTTACTA AAACCTTGTTACATCAAGGTTTTTTCTTTTTGTCTTGTTCATGAGTT[3]
M3 MGAS315	2	ACTCCCACCAGCTCCATCAATGCTTACCGTAAGTAATCATAACTTACTA AAACCTTGTTACATCAAGGTTTTTTCTTTTTGTCTTGTTCATGAGTT
M12 HKU360	2	AATTATTTAACAGCGTCTTT
M18 MGAS8232	2	CATGTACAACTATACTCGT

Homologous phages of ΦAP53.3

AP53	3	AATTATTTAACAGCGTCTTT
M53 Alab49	2	AATTATTTAACAGCGTCTTT
M4 MGAS10750	3	AATTATTTAACAGCGTCTTT
M3 STAB902	4	AATTATTTAACAGCGTCTTT
M3 SSI-1	4	AATTATTTAACAGCGTCTTT
M3 MGAS315	3	AATTATTTAACAGCGTCTTT
M6 JRS4	1	AATTATTTAACAGCGTCTTT
M6 D471	1	AATTATTTAACAGCGTCTTT
M114 NGAS322	2	AATTATTTAACAGCGTCTTT
M1 GAS-SF370	3	AATTATTTAACAGCGTCTTT
M18 MGAS8232	4	AATTATTTAACAGCGTCTTT
M1 NCTC8198	4	AATTATTTAACAGCGTCTTT
M1 AP1	2	AATTATTTAACAGCGTCTTT
M5 Manfredo	2	AATTATTTAACAGCGTCTTT
M14 HSC5	2	AATTATTTAACAGCGTCTTT
M23 M23ND	4	AATTATTTAACAGCGTCTTT
M1 HKU488	2	AATTATTTAACAGCGTCTTT
M1 5448	2	AATTATTTAACAGCGTCTTT
M1 MGAS5005	2	AATTATTTAACAGCGTCTTT
M1 476	2	AATTATTTAACAGCGTCTTT
M1 A20	2	AATTATTTAACAGCGTCTTT

Homologous phages of ΦAP53.4

AP53	4	TATAATGAACATAT
M53 Alab49	3	TATAATGAACATAT
M1 NCTC8198	5	TATAATGAACATAT
M1 AP1	1	TATAATGAACATAT
M5 Manfredo	1	TATAATGAACATAT
M18 MGAS8232	5	TATAATGAACATAT
M6 D471	2	TATAATGAACATAT
M6 JRS4	2	TATAATGAACATAT
M6 MGAS10394	6	TATAATGAACATAT
M2 MGAS10270	3	TCTT[C/A]TATTATATCAGA^{3}
M3 SSI-1	3	TCTT[A/T]TATTATATCAGA^{3}
M3 MGAS315	4	TCTT[T/A]TATTATATCAGA^{3}
M3 STAB902	3	TCTT[T/A]TATTATATCAGA^{3}
M1 GAS-SF370	1	CATGTACAACTATACT
M12 MGAS2096	1	CATGTACAACTATACT
M14 HSC5	3	CTTATATTATAACAAAAA
M28 M28PF1	1	TCTT[A/C]TATTATATCAGA^{3}
M28 MGAS6180	2	TCTT[C/A]TATTATATCAGA^{3}
M28 STAB10015	2	TCTT[C/A]TATTATATCAGA
M87 NGAS743	3	CAATAATGACCCCTGCCGGAATC

Homologous phages of ΦAP53.5		
AP53	5	CAATAATGTTTGTCATAATTT
M53 Alab49	4	CAATAATGTTTGTCATAATTT
M1 GAS-SF370	4	CAATAATGTTTGTCATAATTT
M101 NGAS638	3	CAATAATGTTTGTCATAATTT
M6 MGAS10394	8	CAATAATGTTTGTCATAATTT
ATCC19615	4	CAATAATGTTTGTCATAATTT
M59 MGAS1882	2	CAATAATGTTTGTCATAATTT
M59 MGAS15252	1	CAATAATGTTTGTCATAATTT
M28 STAB10015	3	CAATAATGTTTGTCATAATTT
M28 M28PF1	3	CAATAATGTTTGTCATAATTT
M28 MGAS6180	3	CAATAATGTTTGTCATAATTT
M2 MGAS10270	4	CAATAATGTTTGTCATAATTT
M82 NGAS596	2	CAATAATGTTTGTCATAATTT
M5 Manfredo	5	CAATAATGTTTGTCATAATTT
M4 MGAS10750	4	CAATAATGTTTGTCATAATTT
T25-3__SypCIM25	1	CAATAATGTTTGTCATAATTT

1. “Non-identifiable”: the integration sequences for the group of ΦAP53.1-like phages cannot be precisely identified due to the short and divergent *att*-site sequences on both sides of the phages.
2. There are mismatches between the two *att*-site sequences on the two sides of the phages, shown in brackets.
3. The indicated integration sequences were only identified on either side of the phages due to a large inversion, which interrupted the paired relationship of the integration sequences on both sides of the phages.