Virus Genes
DOI 10.1007/s11262-015-1223-8

Complete genome analysis of two new bacteriophages isolated from impetigo strains of *Staphylococcus aureus*

Tibor Botka[1] · Vladislava Růžičková[1] · Hana Konečná[2] · Roman Pantůček[1] · Ivan Rychlík[3] · Zbyněk Zdráhal[2] · Petr Petráš[4] · Jiří Doškař[1]

Received: 1 April 2015 / Accepted: 23 June 2015

Abstract Exfoliative toxin A (ETA)-coding temperate bacteriophages are leading contributors to the toxic phenotype of impetigo strains of *Staphylococcus aureus*. Two distinct *eta* gene-positive bacteriophages isolated from *S. aureus* strains which recently caused massive outbreaks of pemphigus neonatorum in Czech maternity hospitals were characterized. The phages, designated ϕB166 and ϕB236, were able to transfer the *eta* gene into a prophageless *S. aureus* strain which afterwards converted into an ETA producer. Complete phage genome sequences were determined, and a comparative analysis of five designed genomic regions revealed major variances between them. They differed in the genome size, number of open reading frames, genome architecture, and virion protein patterns. Their high mutual sequence similarity was detected only in the terminal regions of the genome. When compared with the so far described *eta* phage genomes, noticeable differences were found. Thus, both phages represent two new lineages of as yet not characterized bacteriophages of the *Siphoviridae* family having impact on pathogenicity of impetigo strains of *S. aureus*.

Keywords ETA-converting bacteriophages · *Staphylococcus aureus* · Complete genome sequences · Virion protein patterns

Edited by Joachim Jakob Bugert.

Electronic supplementary material The online version of this article (doi:10.1007/s11262-015-1223-8) contains supplementary material, which is available to authorized users.

✉ Vladislava Růžičková
vladkar@sci.muni.cz

[1] Department of Experimental Biology, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic

[2] Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

[3] Veterinary Research Institute, Hudcova 70, 621 00 Brno, Czech Republic

[4] National Reference Laboratory for Staphylococci, National Institute of Public Health, Šrobárova 48, 100 42 Prague, Czech Republic

Introduction

Bacteriophages are the most abundant microorganisms in the biosphere that play an important role in bacterial biology, pathogenicity, and evolution. While many phages simply infect bacteria, lyse them, and spread, others take up long-term residence in their host. As far as *Staphylococcus aureus* is concerned, most strains harbour at least one prophage that replicates along with the host chromosome [1–6]. More than 1880 complete genomes of bacteriophages are reported; hereof, 96 belong to staphylococcal phages (EMBL-EBI Genomes, 2015). Recent sequence analyses indicate that many strains carry a large spectrum of prophages encoding additional factors which enhance bacterial virulence as well as survival ability in various hosts. Staphylophages participate in the mediation of horizontal transfer of pathogenicity islands and raise intra-strain and inter-strain exchange frequency of toxic or resistance genes [7–11].

Some human and animal strains of *S. aureus* produce an exfoliative toxin (ET) [12, 13]. Of the four known ETs-causing epidermal splitting in infected neonatal mice [14, 15], two (ETA and ETB) have been identified as the major causative agents of the blistering skin disease pemphigus neonatorum and/or generalised staphylococcal scalded skin

Published online: 02 July 2015

syndrome (SSSS) [16]. Only the *eta* gene encoding ETA is located on a *S. aureus* prophage [17, 18], while other ET genes were localized on other accessory elements such as a plasmid (*etb*) [19] and/or a pathogenicity island (*etd*) [15].

The staphylococcal bacteriophages belonging to the *Siphoviridae* family, the genus "Phietalikevirus" commonly have a linear double-stranded DNA with a size ranging from 42 to 44 kb, the modular structure of genome, and the B1 morphotype showing an isometric head with non-contractile tail [13, 17, 20, 21]. Most of the impetigo strains of *S. aureus* isolated in 23 maternity hospitals in the Czech Republic contained *eta* gene-positive prophages [12, 22], which can be classified into six ETA-B groups [23].

In this study, the two *eta* gene-positive phages (hereinafter termed *eta* phages) isolated from *S. aureus* strains implicated in massive pemphigus neonatorum outbreaks in two Czech geographically distant maternity hospitals were characterized. Their whole-genome sequences were analysed and compared with the published genomes of three *eta* phages and 24 staphylococcal *eta* gene-negative phages of the *Siphoviridae* family to determine their relatedness and elucidate their impact on the pathogenicity of the host bacteria.

Materials and methods

Bacterial strains

Two ETA-producing *S. aureus* strains (SA 166 and SA 236) were recovered from the skin blisters of child patients in 2008 and 2009 [12]. Both the strains carried prophage B along with prophage F. Strain SA 236 of clonal lineage CC121 harboured the *eta* and *etb* genes, while SA 166 of clonal lineage CC9 carried only *eta*. Prophageless and non-toxigenic *S. aureus* strain CCM 4890, originally designated 1039, was kindly provided by Dr Y. Yoshizawa (Jikei University School of Medicine, Tokyo, Japan).

Isolation of phages

The particles of phages, designated ϕB166 and ϕB236, were isolated from bacterial cells as described previously [20, 24]. The phage stock lysates, prepared by propagation of phages on strain CCM 4890, were used for the isolation of DNA, virion protein analysis, and lysogenization experiments.

Isolation of phage DNA

High-titre lysates of 10^9 pfu/ml were prepared on the propagation strain CCM 4890 growing in BHI broth (Oxoid, UK). Phage lysates were centrifuged at $5000\times g$ for 15 min and filtered by 0.45 μm pore size filter to remove cell debris. The phage DNA was purified as described previously [25].

Restriction endonuclease profiling and PCR genotyping

The phage DNA was digested with *Hin*dIII restriction endonuclease (Roche Diagnostics, Germany) and electrophoresed as described previously [26]. Phage ETA-B type, serogroup, and the *eta* gene were identified by PCR as described previously [23, 27, 28].

DNA sequencing

Phage genome sequences were obtained by the 454 sequencing method. A GS Junior System (Roche 454 Life Science, USA) was used. Samples of DNA were prepared as described previously [25]. Fully covered genome sequences of ϕB166 and ϕB236, with more than 28,300 numreads each, were obtained and assembled using the GS De Novo Assembler, version 2.6. The sequences have been deposited in the GenBank database under accession nos. KP893289 and KP893290.

Genome analysis software

To handle and visualize sequence data, the CLC Genomics Workbench software (CLC, Inc., Aarhus, Denmark) was used. ORFs were identified by GeneMarkS, optimized for phage sequences (http://exon.gatech.edu), and annotated by BLAST (http://blast.ncbi.nlm.nih.gov/) and InterPro (http://www.ebi.ac.uk/interpro/). Nucleotide sequence similarity comparisons are based on megablast algorithm calculations. BPROM [29] was applied to promoter prediction. Only the promoters with an overall score above 3.5 and located in close vicinity to ORF were considered. ARNold [30] was applied to Rho-independent terminator identification. Only the predicted terminators with a dG value of less than −11 kcal/mol and with functional position were considered. The GC content and its variation in the phage genomes were analysed by GC-Profile [31].

Lysogenic conversion and detection of ETA production

Staphylococcus aureus strain CCM 4890 was lysogenized with induced phages as described previously [20]. Lysogenic strains were incubated in BHI medium at 37 °C for 24 h. Single colonies were cultivated three times in medium with 10^{-3} M Na citrate to obtain pure culture of lysogens. PCR detection of prophages and the *eta* gene as well as PFGE of *Sma*I macrorestriction fragments

confirmed the presence of integrated prophages. ETA produced by the lysogens was detected using the Reverse Passive Latex Agglutination Kit (RPLA, Denka Seiken for Unipath, Japan).

PFGE analysis

Genomic DNA of lysogenized strains was prepared as described previously [32]. *Sma*I macrorestriction fragments were separated with a CHEF Mapper (Bio-Rad Laboratories) under the following conditions: 1.2 % agarose in 1× TAE electrophoresis buffer (0.04 M Tris/acetate, 0.001 M EDTA, pH 8.2) at 14 °C, 6 V cm^{-1}, pulse times of 1–55 s for 24 h.

Structure proteins analysis

Virion proteins were isolated from phage lysate according to Eyer et al. [33]. Vertical one-dimensional electrophoresis (1-DE), using the Bio-Rad equipment and reagents, was performed in Protean II xi Cell (discontinuous 12 % T SDS-PAGE). Precision Plus Protein Standard (catalogue no. 161-0363) was applied as the molecular weight marker. Proteins were stained with Bio-Safe Coomassie Stain. The GS-900 calibrated densitometer and Image Lab 5.1 software (cubic spline regression) were used for image analysis. Protein bands were excised and subjected to tryptic digestion (40 °C, 2 h). Tryptic peptides were analysed by LC–MS/MS system consisting of RSLCnano chromatograph (Thermo Fisher Scientific) connected to Impact II Ultra-High Resolution Qq-Time-Of-Flight mass spectrometer (Bruker). After desalting and concentration on trap column, the peptides were separated using Acclaim Pepmap100 C18 column (2 µm particles, 75 µm × 500 mm; Thermo Fisher Scientific, 300 nl/min) and 0.1 % FA/acetonitrile gradient. MS data were acquired in a data-dependent strategy with 3 s long cycle time. Mass range was set to 150–2200 m/z, and precursors were selected from 300 to 2000 m/z. Mascot search engine was used in combination of phage translated sequence data for protein identification. Mass tolerance for peptides and MS/MS fragments were 15 ppm and 0.05 Da, respectively.

Results and discussion

The present study focused on the characterization of two as yet undescribed bacteriophages, ϕB166 and ϕB236, originating from staphylococcal impetigo strains isolated from the skin blisters of two neonatal patients. The phages, identified by PCR as B phages carrying the *eta* gene, were investigated by whole-genome sequencing and assayed for *eta* gene transfer activity and lysogenic conversion of ETA production. The phage ϕB236 was assigned to type ETA-B5 and ϕB166 to ETA-B6. Prophages of the ETA-B5 type were found in 24 *S. aureus* strains isolated in eight maternity hospitals, and prophages of the ETA-B6 type have been reported in 17 impetigo strains from two distant Czech maternity hospitals [12].

In accordance with previous study [20], the ϕB166 as well as ϕB236 genome integrated into the 365-kb *Sma*I fragment of the recipient strain (Fig. 1). Selected *eta* gene-positive lysogens turned out to be producers of exfoliative toxin A. These results proved the induced *eta* phages to be efficient ETA-converting phages. Two lysogenized strains designated CCM 8515 (1039/ϕ166+) and CCM 8512 (1039/ϕ236+), and both phages, ϕB166 (CCM 8513) and ϕB236 (CCM 8514), were deposited into the Czech Collection of Microorganisms.

The genomes and virion protein patterns of ϕB166 and ϕB236

Restriction endonuclease analysis of the phage genomes gave the first evidence that ϕB166 and ϕB236 are not identical (Supplementary material, Fig. S1). Description of the genome sequences, their characteristics, and predicted gene products are summarized in Table 1 and Fig. 2. The sequences of these

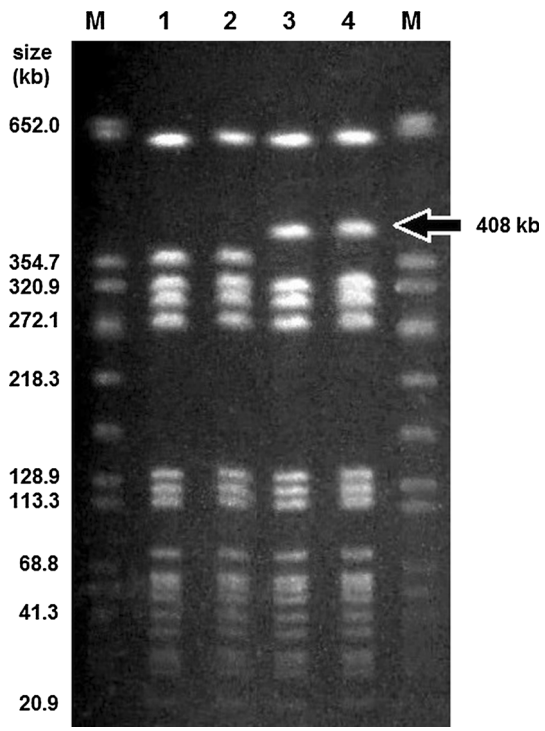

Fig. 1 A pulsed-field electrophoresis agarose gel of DNAs from prophageless *S. aureus* strain CCM 4890 (*lines 1* and *2*) and lysogens CCM 8515 [+ϕB166] (*line 3*) and CCM 8512 [+ϕB236] (*line 4*). M, size standard (DNA of *S. aureus* NCTC 8325); the *Sma*I restriction fragment with integrated *eta* prophage is indicated by a *full arrow*

Table 1 Predicted phage ϕB166 and ϕB236 products categorized into five genomic regions

ORF	ϕB166[a]		ORF	ϕB236[b]	
	Predicted function Region I (1–9097)	aa		Predicted function Region I (1–8404)	aa
01	Integrase	349	01	Integrase	349
02	Hypothetical protein	156	02	PemK-like phage protein	226
03	Hypothetical protein	173	03	cI-like repressor	239
04	Putative bacterial membrane protein	46	04	cro-like repressor	72
05	cI-like repressor	212	05	Antirepressor	263
06	cro-like repressor	84	06	Hypothetical protein	74
07	Putative antirepressor	262	07	Hypothetical protein	59
08	Hypothetical protein	67	08	Putative bacterial membrane protein	160
09	Hypothetical protein	46	09	XRE transcriptional regulator	61
10	Hypothetical protein	66	10	Putative antirepressor	250
11	Hypothetical protein	73	11	Hypothetical protein	71
12	DUF 1270 hypothetical protein	53	12	Hypothetical protein	84
13	DUF 2482 hypothetical protein	100	13	Hypothetical protein	73
14	DUF1108 hypothetical protein	85	14	DUF1270 hypothetical protein	53
15	Hypothetical protein	92	15	DUF1108 hypothetical protein	98
16	ATPase	651	16	DUF2483 hypothetical protein	73
17	ssDNA annealing protein RecT	306	17	SAK3-like protein	207
ORF	ϕB166[a]		ORF	ϕB236[b]	
	Predicted function Region II (9185–16,920)	aa		Predicted function Region II (8405–17,114)	aa
18	Metallo-hydrolase	161	18	ssDNA-binding protein	141
19	ssDNA-binding protein	156	19	DUF968 hypothetical protein	231
20	Replication protein	297	20	Replication protein	268
21	Hypothetical protein	72	21	Helicase inhibitor G39P-like protein	118
22	RusA-like protein	138	22	Replicative DNA helicase	413
23	ϕPVL_ORF50-like protein	125	23	Hypothetical protein	71
24	Phage_ORF51 superfamily protein	80	24	DUF3269 hypothetical protein	76
25	Hypothetical protein	68	25	DUF1064 hypothetical protein	134
26	Hypothetical protein	134	26	DUF3113 hypothetical protein	61
27	Hypothetical protein	64	27	ϕPVL_ORF50-like protein	119
28	Hypothetical protein	150	28	Phage_ORF51 superfamily protein	82
29	Hypothetical protein	167	29	DUF1024 hypothetical protein	87
30	Hypothetical protein	65	30	Hypothetical protein	66
31	DUF1024 hypothetical protein	82	31	Hypothetical protein	79
32	dUTPase	184	32	HNH endonuclease domain protein	184
33	Putative bacterial membrane protein	81	33	dUTPase	186
34	DUF1523 hypothetical protein	131	34	Hypothetical protein	57
35	Transcriptional activator RinB	57	35	Hypothetical protein	88
36	Hypothetical protein	133	36	DUF1381 hypothetical protein	96
			37	Hypothetical protein	78
			38	Transcriptional activator RinB	58
			39	Hypothetical protein	133

Table 1 continued

ORF	ϕB166[a]		ORF	ϕB236[b]	
	Predicted function Region III (16,903–24,920)	aa		Predicted function Region III (17,115–25,245)	aa
37	Terminase small subunit	164	40	Terminase small subunit	164
38	Terminase large subunit	407	41	Terminase large subunit	403
39	Portal protein	474	42	Portal protein	492
40	Head morphogenesis protein	316	43	Head morphogenesis protein	326
41	Hypothetical protein	194	44	Minor structural protein GP20	198
42	Major capsid protein	304	45	Capsid protein	274
43	Hypothetical protein	48	46	Rho domain protein	108
44	Head–tail connector	116	47	Head–tail connector	104
45	Head–tail adaptor	111	48	Head–tail adaptor	111
46	Putative tail component	135	49	Putative tail component	137
47	DUF3168 hypothetical protein	141	50	DUF3168 hypothetical protein	145
48	Tail protein	185	51	Putative major tail protein	186
ORF	ϕB166[a]		ORF	ϕB236[b]	
	Predicted function Region IV (24,886–36,266)	aa		Predicted function Region IV (25,224–36,614)	aa
49	DUF3647 hypothetical protein	168	52	DUF3647 hypothetical protein	164
50	Hypothetical protein	107	53	Hypothetical protein	105
51	Tape measure protein	961	54	Tape measure protein	961
52	Siphovirus-type tail component	313	55	Siphovirus-type tail component	311
53	Tail endopeptidase	628	56	Tail endopeptidase	628
54	Putative minor structural protein	632	57	Putative minor structural protein	632
55	DUF2479 hypothetical protein	607	58	DUF2479 hypothetical protein	607
56	DUF2977 hypothetical protein	125	59	DUF2977 hypothetical protein	125
57	XkdX superfamily protein	58	60	XkdX superfamily protein	60
58	DUF2951 hypothetical protein	99	61	DUF2951 hypothetical protein	77
ORF	ϕB166[a]		ORF	ϕB166[a]	
	Predicted function Region V (36,241–42,881)	aa		Predicted function Region V (36,590–43,228)	aa
59	Tail tip cell wall hydrolase	624	62	Tail tip cell wall hydrolase	624
60	Tail fibre protein	412	63	Tail fibre protein	412
61	phiETA ORF63-like protein	131	64	phiETA ORF63-like protein	131
62	Holin	91	65	Holin	91
63	Amidase	473	66	Amidase	473
64	Exfoliative toxin A	280	67	Exfoliative toxin A	280

Position of each region in the phage genome is indicated in brackets

[a] ORFs 01–05 and 09 were localized on the negative strand

[b] ORFs 01–03, 08, and 12 were localized on the negative strand

phages were divided into five genomic regions, which were designed with respect to the functional genomic architecture, based on the predicted regulatory elements and modular arrangement. Each of the five regions carries specific genes as follows: Region I: genes for lysogeny and regulation of transcription, region II: genes for products involved in DNA replication, region III: genes for terminase and capsid structure proteins, region IV: genes for tail proteins, and the smallest region V: genes for host lysis and the *eta* gene encoding exfoliative toxin A (Table 1; Fig. 3).

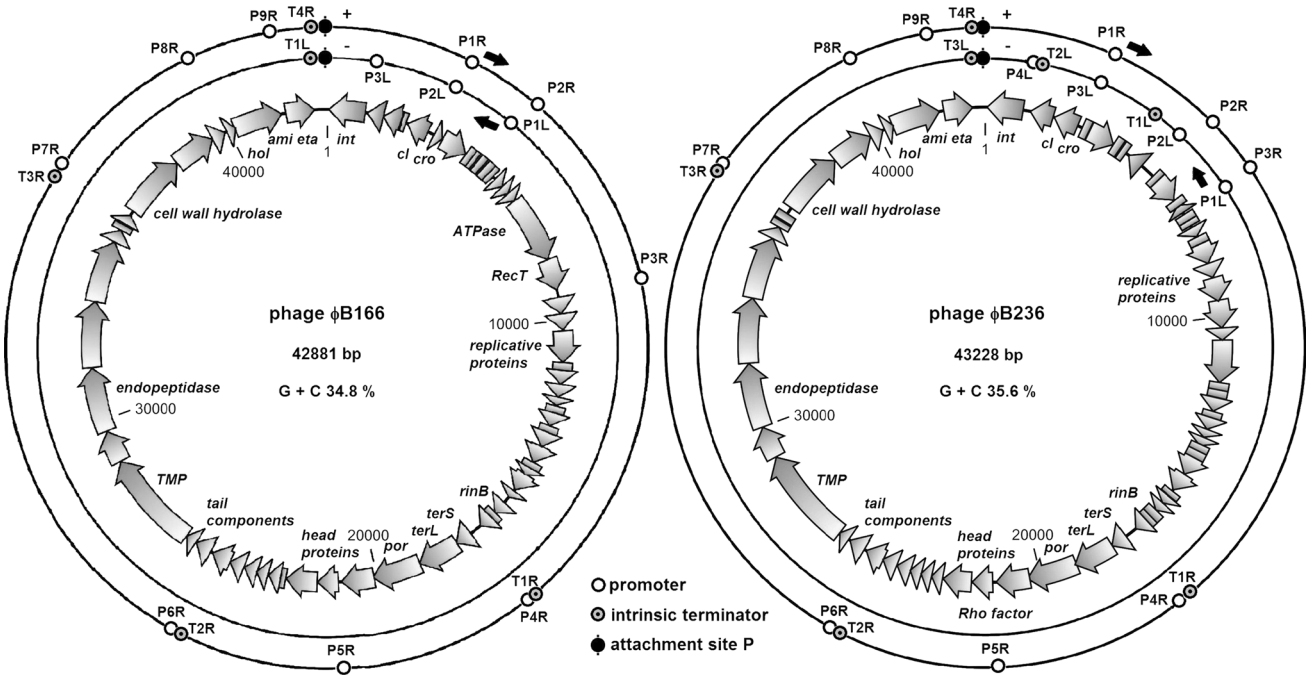

Fig. 2 Circular maps of the ϕB166 and ϕB236 genomes. ORFs are represented by *dark arrows* on *inner circles*. Promoters and terminators are marked on *outer circles*. Directions of transcription are indicated by *black arrows*. The main (+) and complementary (−) strands are distinguished. The attachment site P is indicated at the *top* of each map. The whole-genome size and GC content are provided. Positions are shown in bp. Detailed information is in Table 1 and Supplementary material in Table S1

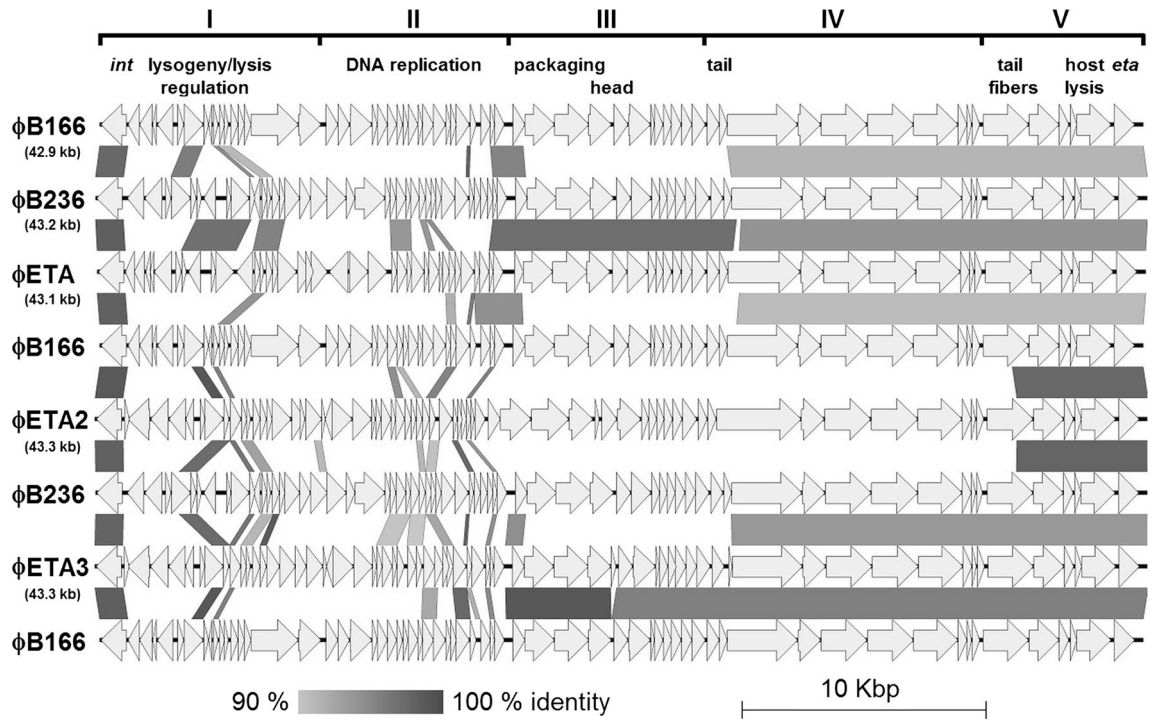

Fig. 3 Alignment of the genome maps of ϕB166 and ϕB236 and ϕETA, ϕETA2, and ϕETA3. Regions I−V and modules are depicted on the *top line*. ORFs are represented by *full arrows* and intergenic sequences by abscissa. *Shaded boxes* express the sequence identity levels between genomes. The phage genome sizes are noted in *brackets*

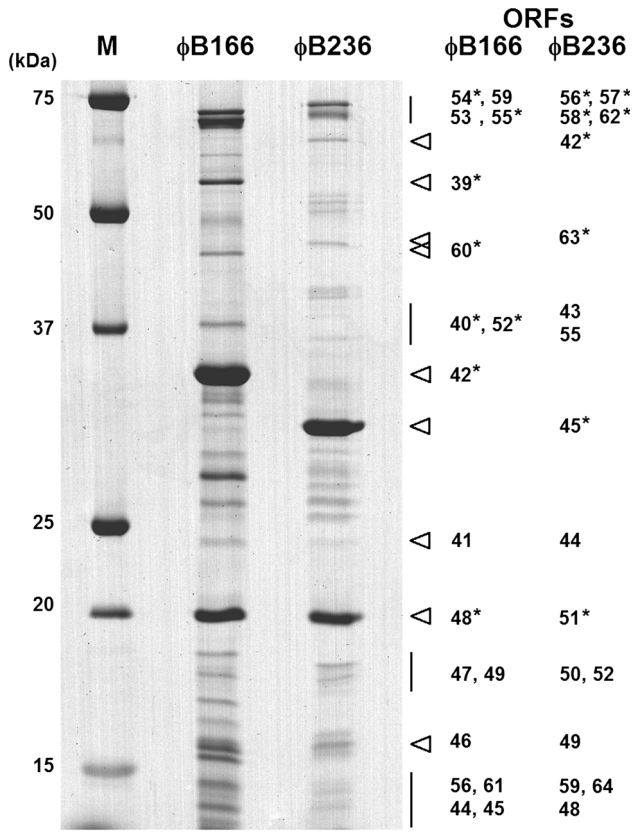

Fig. 4 SDS-PAGE patterns of virion proteins. M = Precision Plus Protein Standard. ORFs corresponding to protein bands are noted on the right. Described positions are marked by empty *arrow heads* or *vertical lines*. Protein bands used for MS analysis are marked by *asterisk*. Fragments of tape measure protein were detected only by MS. TerL and amidase fragments were identified only in the ϕB166 pattern. Details of ORFs are given in Table 1

The determination of the complete genome sequences revealed that ϕB166 and ϕB236 have genome sizes of 42.9 and 43.2 kb, respectively. The coding sequences (ORFs) take over 94.4 % of the genome of ϕB166 and 93.7 % of the genome of ϕB236. In addition, 12 promoters and five Rho-independent terminators were predicted in the genome of ϕB166, and 13 promoters and seven Rho-independent terminators were predicted in the ϕB236 genome (Fig. 2; Supplementary Material, Table S1).

The whole GC content corresponds to those in other B phages and is slightly superior to those in *S. aureus* (approximately 32.7 %). Based on the low percentage of G+C, genome location, and InterPro prediction, some ORFs seem to have a diverse origin. A good example is ϕB236-ORF08 and its promoter P2L (27.5 % G+C).

SDS-PAGE analysis of ϕB166 and ϕB236 virion proteins revealed major differences between the molecular weights of the capsid proteins which correlated with gene sequence variability detected in the respective ORFs (Fig. 4). The differences between SDS-PAGE patterns were also observed in the previously characterized *eta* phages ϕB557 and ϕB122 [20]. On the other hand, a resemblance was found between the tail and tail fibre components.

Comparison of ϕB166 and ϕB236 with *eta*-positive and *eta*-negative phages

Sequence analyses revealed that ϕB166 and ϕB236 are not closely related to each other or to other previously reported *eta* phages (Fig. 3; Table 2). The phage ϕB166 showed the following levels of total sequence similarity: 51 % to ϕETA, 36 % to ϕETA2, 67 % to ϕETA3, and 54 % to ϕB236. In respect of non-*eta* phages, the highest similarities were found between ϕB166 and the ϕNM4 (62 %), ϕ96 (63 %), ϕ71, and ϕ55 (51 % each). The region III showed 98 % homology to ϕNM4 and ϕ96; therefore, it may originate from a common ancestor or could have been acquired as a result of recombination exchange between phages during co-replication.

Whole-genome comparison of ϕB236 with other *eta* phages revealed the following identity levels: 78 % with ϕETA, 37 % with ϕETA2, 58 % with ϕETA3, and 53 % with ϕB166. As far as non-*eta* phages are concerned, the highest similarities were found between ϕB236 and ϕ71 (72 %), ϕ29 (70 %), ϕ55 (69 %), and StauST398-1 (67 %). The regions I and II of ϕB166 and ϕB236 are the most variable and distinct from those of other *eta* phages and non-*eta* phages.

Despite genome variability, all *eta* phages share highly similar sequences of the region V harbouring the major virulence gene encoding ETA. The phages ϕB166 and ϕB236 demonstrated the same type of integrase locus as ϕETA, ϕETA2, and ϕETA3. Much like ϕETA, these phages did not share the *xis* gene with ϕETA2 and ϕETA3, which indicates that the excisionase does not play a crucial role in releasing an *eta* prophage from the host chromosome.

The gene for Rho-termination factor domain protein was identified in ϕB236, but not in ϕB166. The same and/or very similar gene was found in ϕETA and in the non-*eta* phages ϕ88, ϕ92, StauST398-1, StauST398-5, ϕ55, ϕ71, ϕ29, and ϕ52A. The promoter and terminator sequences of phages showed reciprocal similarities almost analogously as the nearby ORFs. Differences between regulatory sequences are caused by variation in region II and the presence of presumptive morons (e.g. ϕB166 ORF09 or ϕB236 ORFs 08 and 12) flanked by their own promoters and/or terminators.

Table 2 Percentages of similarities between the genomic regions of ϕB166, ϕB236, and selected phages

Phage[a]		ϕB166 region[b]					ϕB236 region[b]				
Name	ID	I	II	III	IV	V	I	II	III	IV	V
eta-phages											
ϕB166 (Sa1, B)	KP893289	100	100	100	100	100	31	16	18	79	**98**
ϕB236 (Sa1, B)	KP893290	**28**	19	18	79	98	100	100	100	100	100
ϕETA (Sa1, B)	NC_003288	18	**32**	15	77	98	**50**	32	**97**	**90**	**98**
ϕETA2 (Sa1, B)	NC_008798	**28**	17	–	39	80	33	25	–	41	80
ϕETA3 (Sa1, B)	NC_008799	27	30	**98**	**93**	**99**	33	**34**	24	83	**98**
non eta-phages											
55 (Sa1, B)	NC_007060	17	42	15	75	56	38	34	**97**	90	55
71 (Sa1, B)	NC_007059	18	30	10	77	**81**	42	**62**	53	**93**	**80**
SAP-26 (Sa1, B)	NC_014460	40	19	–	41	38	40	25	–	43	38
StauST398-1 (Sa3, B)	NC_021326	4	36	25	76	77	13	34	**97**	89	77
TEM126 (Sa3, B)	HQ127381	31	35	–	39	34	21	10	–	42	35
ϕ11 (Sa5, B)	NC_004615	17	10	–	41	34	34	54	–	43	36
29 (Sa5, B)	NC_007061	7	17	15	77	55	30	56	84	**93**	54
69 (Sa5, B)	NC_007048	10	18	–	43	36	33	31	–	42	37
88 (Sa5, B)	NC_007063	14	34	10	77	55	27	47	62	92	55
ϕMR25 (Sa5, B)	NC_010808	15	24	–	42	35	**45**	37	–	42	36
52A (Sa6, B)	NC_007062	7	27	15	77	53	5	33	84	**93**	53
80 (Sa6, B)	DQ908929	8	25	–	77	53	10	33	84	**93**	53
ϕNM4 (Sa6, B)	DQ530362	3	49	**98**	90	34	27	34	24	80	33
92 (Sa7, B)	NC_007064	7	34	10	77	54	30	47	62	92	53
TEM123 (Sa7, B)	NC_017968	**60**	60	–	39	34	21	10	–	42	35
96 (Sa9, B)	NC_007057	7	31	**98**	**91**	53	3	56	24	79	54
StauST398-5 (-, B)	NC_023500	3	36	25	76	77	3	34	**97**	89	77
ϕPVL (Sa2, F)	NC_002321	35	50	–	–	<1	17	8	–	–	<1
ϕBU01 (Sa3, F)	KF831354	34	61	–	–	<1	11	12	–	–	<1
JS01 (Sa3, F)	NC_021773	49	**66**	15	–	–	16	13	–	–	–
ϕ13 (Sa3, F)	NC_004617	4	16	–	2	–	22	27	–	–	–
ϕNM3 (Sa3, F)	NC_008617	42	46	–	–	<1	12	9	–	–	<1
StauST398-4 (Sa3, F)	NC_023499	42	46	–	–	<1	12	9	–	–	–
P954 (Sa7, F)	NC_013195	2	24	–	2	<1	26	50	–	–	<1

[a] Phages are listed in order of assignment to serogroup and integrase family according to Kahánková et al. [27]

[b] The highest similarities determined in each region are highlighted in bold; (–) no significant similarity was found

Phage genomic mosaicism

In agreement with the known extensive mosaic structure of prophages resulting from the recombination among members of a phage species and horizontal transfer of sequences [34–36], many non-homologous ORFs and/or sequences located in the genomes of ϕB166 and ϕB236 were identified. In ϕB166, these were ORFs 2–4 and ORFs 29–30. In ϕB236, ORFs 31, 32, and 35 were unique. Both phages use the ϕETA-like *pac*-site located within the *terS* sequence as reported by Yamaguchi et al. [17]. In accordance with the recently proposed division of the *Siphoviridae* family [21], both phages studied have a tail protein architecture and catalytic domain distribution corresponding to the genus "Phietalikevirus", subgroup I. In contrast, there is a HNH domain in the ϕB236-ORF32 product, which is considered a typical characteristic of the proposed genus "3alikevirus".

Except the *eta* gene and its promoter, conserved region V is highly similar to that of the prophage of *S. aureus* strain 04-02981 (CP001844). Its genome contains the *ami* gene and a nearby sequence, highly similar to the terminal part of *eta* phages. It indicates that the *eta* gene and its promoter could be incorporated between the *ami* gene and the *eta*-terminator by recombination, which is inconsistent with the previously assumed aberrant excision [17].

Recombination events in *eta* phages are supported by the finding that a large part of the ϕB166 genome (ORF 14–22) is very similar to that of F phages (ϕPVL and ϕJS01). ORFs 13 and 23 in ϕB166 can be considered as the most probable recombination sites. They resemble those of some F phages (i.e. JS01, BU01, and ϕNM3), prophage StauST398-4 [37], and B phage ϕETA3. The genes involved in recombination were found in both phages studied. The ϕB166 genome contains the genes for RecT and RusA-like proteins, while ϕB236 contains Sak3-like and HNH endonuclease domain proteins. These proteins can mediate new rearrangements in the phage genomes, especially in polylysogenic strains or strains with defective prophages [36].

Conclusions

The emerging data strongly suggest that ETA-converting phages are crucial in the pathogenesis of staphylococcal skin blistering infections because of the extensive prevalence of ETA-producing impetigo strains of *S. aureus*. This work reports detailed genomic characteristics of two *eta* gene-positive bacteriophages as new members of a large group of diverse temperate phages encoding exfoliative toxin A. Based on the genomic comparisons between phages isolated in this work and those formerly characterized as *S. aureus* ETA-like B bacteriophages, it can be concluded that ϕB166 and ϕB236 represent the recent new lineages of as yet undescribed ETA-converting bacteriophages. Polymorphisms detected in their genomic sequences could be used as an identification tool for detailed classification of the *eta* phages. Their genomic mosaicism reveals them as a fluid gene pool to confer new properties to co-replicating phages and also as mediators of the *S. aureus* pathogenicity.

Acknowledgments The authors would like to thank their colleagues from the participating hospitals for providing the strains of *S. aureus*. Thanks are also due to Ms Eva Kodytková for valuable help. This work was supported by Grant No. NT12395-5 from the Internal Grant Agency of the Ministry of Health of the Czech Republic and project AdmireVet CZ.1.05/2.1.00/01.0006-ED0006/01/01 from the Czech Ministry of Education, Youth and Sports. The proteomic part was carried out with the support of the CEITEC (Central European Institute of Technology) Proteomics Core Facility, ID No. CZ.1.05/1.1.00/02.0068, financed from the European Regional Development Fund and Czech Science Foundation (Project No. P206-12-G151).

References

1. T. Bae, T. Baba, K. Hiramatsu, O. Schneewind, Mol. Microbiol. **62**, 1035–1047 (2006)
2. C. Goerke, R. Pantůček, S. Holtfreter, B. Schulte, M. Zink, D. Grumann, B.M. Broker, J. Doškař, C. Wolz, J. Bacteriol. **191**, 3462–3468 (2009)
3. R. Hrstka, V. Růžičková, P. Petráš, R. Pantůček, S. Rosypal, J. Doškař, Int. J. Med. Microbiol. **296**, 49–54 (2006)
4. J.J. Iandolo, V. Worrel, K.H. Groicher, Y. Qian, R. Tian, S. Kenton, A. Dorman, H. Ji, S. Lin, P. Loh, S. Qi, H. Zhu, B.A. Roe, Gene **289**, 109–118 (2002)
5. R. Pantůček, J. Doškař, V. Růžičková, P. Kašpárek, E. Oráčová, V. Kvardová, S. Rosypal, Arch. Virol. **149**, 1689–1703 (2004)
6. V. Růžičková, R. Karpíšková, R. Pantůček, M. Pospíšilová, P. Černíková, J. Doškař, Int. J. Food Microbiol. **121**, 60–65 (2008)
7. R.P. Novick, G.E. Christie, J.R. Penadés, Nat. Rev. Microbiol. **8**, 541–551 (2010)
8. A.J. McCarthy, A.A. Witney, J.A. Lindsay, Front. Cell. Infect. Microbiol. **2**, 1–10 (2012)
9. A.D. Dearborn, T. Dokland, Bacteriophage **2**, 70–78 (2012)
10. M. Varga, L. Kuntová, R. Pantůček, I. Mašlaňová, V. Růžičková, J. Doškař, FEMS Microbiol. Lett. **332**, 146–152 (2012)
11. I. Mašlaňová, J. Doškař, M. Varga, L. Kuntová, J. Mužík, D. Malúšková, V. Růžičková, R. Pantůček, Environ. Microbiol. Rep. **5**, 66–73 (2013)
12. V. Růžičková, R. Pantůček, P. Petráš, I. Machová, K. Kostýlková, J. Doškař, Int. J. Med. Microbiol. **302**, 237–241 (2012)
13. Y. Endo, T. Yamada, K. Matsunaga, Y. Hayakawa, T. Kaidoh, S. Takeuchi, Vet. Microbiol. **96**, 81–90 (2003)
14. M. Amagai, N. Matsuyoshi, Z.H. Wang, C. Andl, J.R. Stanley, Nat. Med. **6**, 1275–1277 (2000)
15. T. Yamaguchi, K. Nishifuji, M. Sasaki, Y. Fudaba, M. Aepfelbacher, T. Takata, M. Ohara, H. Komatsuzawa, M. Amagai, M. Sugai, Infect. Immun. **70**, 5835–5845 (2002)
16. S. Ladhani, R.W. Evans, Arch. Dis. Child. **78**, 85–88 (1998)
17. T. Yamaguchi, T. Hayashi, H. Takami, K. Nakasone, M. Ohnishi, K. Nakayama, S. Yamada, H. Komatsuzawa, M. Sugai, Mol. Microbiol. **38**, 694–705 (2000)
18. Y. Yoshizawa, J. Sakurada, S. Sakurai, K. Machida, I. Kondo, S. Masuda, Microbiol. Immunol. **44**, 189–191 (2000)
19. T. Yamaguchi, T. Hayashi, H. Takami, M. Ohnishi, T. Murata, K. Nakayama, K. Asakawa, M. Ohara, H. Komatsuzawa, M. Sugai, Infect. Immun. **69**, 7760–7771 (2001)
20. P. Holochová, V. Růžičková, R. Pantůček, P. Petráš, R. Janisch, J. Doškař, Res. Microbiol. **161**, 260–267 (2010)
21. D. Gutiérrez, E.M. Adriaenssens, B. Martínez, A. Rodríguez, R. Lavigne, A.M. Kropinski, P. García, Arch. Virol. **159**, 389–398 (2014)
22. V. Růžičková, R. Pantůček, P. Petráš, J. Doškař, I. Sedláček, S. Rosypal, Int. J. Med. Microbiol. **292**, 541–545 (2003)
23. P. Holochová, V. Růžičková, L. Dostálová, R. Pantůček, P. Petráš, J. Doškař, Diagn. Microbiol. Infect. Dis. **66**, 248–252 (2010)
24. Y. Duval-Iflah, Can. J. Microbiol. **18**, 1491–1497 (1972)
25. J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular cloning: a laboratory manual*, 2nd edn. (Cold Spring Harbor Laboratory, New York, 1989)
26. J. Doškař, P. Pallová, R. Pantůček, S. Rosypal, V. Růžičková, P. Pantůčková, J. Kailerová, K. Klepárník, Z. Malá, P. Bocek, Can. J. Microbiol. **46**, 1066–1076 (2000)
27. J. Kahánková, R. Pantůček, C. Goerke, V. Růžičková, P. Holochová, J. Doškař, Environ. Microbiol. **12**, 2527–2538 (2010)
28. V. Růžičková, J. Voller, R. Pantůček, P. Petráš, J. Doškař, Folia Microbiol. **50**, 499–502 (2005)
29. V. Solovyev, A. Salamov, in *Metagenomics and its applications in agriculture, biomedicine and environmental studies*, ed. by R.W. Li (Nova Science Publishers, New York, 2011), pp. 61–78
30. E.A. Lesnik, R. Sampath, H.B. Levene, T.J. Henderson, J.A. McNeil, D.J. Ecker, Nucleic Acids Res. **29**, 3583–3594 (2001)
31. F. Gao, C.T. Zhang, Nucleic Acids Res. **34**, 686–691 (2006)
32. R. Pantůček, F. Götz, J. Doškař, S. Rosypal, Int. J. Syst. Bacteriol. **46**, 216–222 (1996)

33. L. Eyer, R. Pantůček, Z. Zdráhal, H. Konečná, P. Kašpárek, V. Růžičková, L. Hernychová, J. Preisler, J. Doškař, Proteomics **7**, 64–72 (2007)
34. R.W. Hendrix, M.C. Smith, R.N. Burns, M.E. Ford, G.F. Hatfull, Proc. Natl. Acad. Sci. U.S.A. **96**, 2192–2197 (1999)
35. T. Kwan, J. Liu, M. DuBow, P. Gros, J. Pelletier, Proc. Natl. Acad. Sci. U.S.A. **102**, 5174–5179 (2005)
36. M. De Paepe, G. Hutinet, O. Son, J. Amarir-Bouhram, S. Schbath, M.-A. Petit, PLoS Genet. (2014). doi:10.1371/journal.pgen.1004181
37. N. Van Der Mee-Marquet, A.-R. Corvaglia, A.-S. Valentin, D. Hernandez, X. Bertrand, M. Girard, J. Kluytmans, P.Y. Donnio, R. Quentin, P. Francois, Infect. Genet. Evol. **18**, 299–308 (2013)