MICROBIAL GENOMICS

RESEARCH ARTICLE

Nepal *et al.*, *Microbial Genomics* 2021;7:000726

DOI 10.1099/mgen.0.000726

Prophages encoding human immune evasion cluster genes are enriched in *Staphylococcus aureus* isolated from chronic rhinosinusitis patients with nasal polyps

Roshan Nepal[1,2], Ghais Houtak[1,2], Gohar Shaghayegh[1,2], George Bouras[1,2], Keith Shearwin[3], Alkis James Psaltis[1,2], Peter-John Wormald[1,2] and Sarah Vreugde[1,2,*]

Abstract

Prophages affect bacterial fitness on multiple levels. These include bacterial infectivity, toxin secretion, virulence regulation, surface modification, immune stimulation and evasion and microbiome competition. Lysogenic conversion arms bacteria with novel accessory functions thereby increasing bacterial fitness, host adaptation and persistence, and antibiotic resistance. These properties allow the bacteria to occupy a niche long term and can contribute to chronic infections and inflammation such as chronic rhinosinusitis (CRS). In this study, we aimed to identify and characterize prophages present in *Staphylococcus aureus* from patients suffering from CRS in relation to CRS disease phenotype and severity. Prophage regions were identified using PHASTER. Various *in silico* tools like ResFinder and VF Analyzer were used to detect virulence genes and antibiotic resistance genes respectively. Progressive MAUVE and maximum likelihood were used for multiple sequence alignment and phylogenetics of prophages respectively. Disease severity of CRS patients was measured using computed tomography Lund–Mackay scores. Fifty-eight *S. aureus* clinical isolates (CIs) were obtained from 28 CRS patients without nasal polyp (CRSsNP) and 30 CRS patients with nasal polyp (CRSwNP). All CIs carried at least one prophage (average=3.6) and prophages contributed up to 7.7% of the bacterial genome. Phage integrase genes were found in 55/58 (~95%) *S. aureus* strains and 97/211 (~46%) prophages. Prophages belonging to Sa3int integrase group (phiNM3, JS01, phiN315) (39/97, 40%) and Sa2int (phi2958PVL) (14/97, 14%) were the most prevalent prophages and harboured multiple virulence genes such as *sak, scn, chp, luk*E/D, *sea*. Intact prophages were more frequently identified in CRSwNP than in CRSsNP (P=0.0021). Intact prophages belonging to the Sa3int group were more frequent in CRSwNP than in CRSsNP (P=0.0008) and intact phiNM3 were exclusively found in CRSwNP patients (P=0.007). Our results expand the knowledge of prophages in *S. aureus* isolated from CRS patients and their possible role in disease development. These findings provide a platform for future investigations into potential tripartite associations between bacteria-prophage-human immune system, *S. aureus* evolution and CRS disease pathophysiology.

Received 17 July 2021; Accepted 21 October 2021; Published 15 December 2021

Author affiliations: [1]Adelaide Medical School, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide, Australia; [2]The Department of Surgery – Otolaryngology Head and Neck Surgery, University of Adelaide and the Basil Hetzel Institute for Translational Health Research, Central Adelaide Local Health Network, Adelaide, South Australia, Australia; [3]School of Biological Sciences, Faculty of Sciences, The University of Adelaide, Adelaide, Australia.

***Correspondence:** Sarah Vreugde, sarah.vreugde@adelaide.edu.au

Keywords: phage; bacteriophage; *S. aureus*; phage-encoded virulence factors (PEVF); CRS; prophage.

Abbreviations: ANOVA, analysis of variance; ARG, antimicrobial resistance gene; BLAST, Basic Local Alignment Search Tool; CC, clonal complex; CHIPS, chemotaxis inhibitory protein; CI, clinical isolate; CRS, chronic rhinosinusitis; CRSsNP, CRS without nasal polyposis; CRSwNP, CRS with nasal polyposis; DB, database; EPOS, European position paper on rhinosinusitis and nasal polyps; GO, gene ontology; HGT, horizontal gene transfer; IEC, immune evasion cluster; IS, insertion sequence; LMK, Lund–Mackay score; MAFFT, multiple alignment using fast Fourier transform; MGE, mobile genetic element; MLST, multilocus sequence typing; MRSA, methicillin-resistant *Staphylococcus aureus*; NCBI, National Center for Biotechnology Information; ORF, open reading frame; PHASTER, PHAge search tool – enhanced release; PVL, Panton–Valentine leukocidin; SaPIs, *Staphylococcus aureus* Pathogenicity Islands; SCC, staphylococcal cassette chromosome; SCIN, staphylococcal complement inhibitor; SCV, small colony variant; ST, sequence type; Tn, transposons; VF, virulence factor; VGT, vertical gene transfer; VRSA, vancomycin-resistant *Staphylococcus aureus*.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. One supplementary table and six supplementary figures are available with the online version of this article.

HIGHLIGHTS

- In total, 211 prophage regions were identified in 58 *S. aureus* genomes isolated from chronic rhinosinusitis (CRS) patients suggesting widespread distribution of prophage elements in clinical strains colonizing nasal niche.
- Sa2int and Sa3int group prophages belonging to family *Siphoviridae* and genus *Biseptimavirus* were most frequently found in *S. aureus* from CRS patients.
- *S. aureus* isolated from CRS patients with nasal polyps predominantly harboured intact Sa3int group prophages encoding human immune evasion cluster (IEC) genes.
- Prophages in *S. aureus* did not encode any antibiotic resistant genes (ARGs).

SUMMARY

Prophages of *Staphylococcus aureus* modulate bacterial fitness on multiple levels like infectivity, toxicity, virulence regulation, immune evasion and microbiome competition as they arm bacteria with accessory genes. These properties allow *S. aureus* to persist in a nasal niche, possibly contributing to the severity and phenotype of infections like chronic rhinosinusitis (CRS). Here, we report that *S. aureus* isolated from CRS patients carried at least one prophage and contributed up to 7.7% of the total bacterial genome. Intact prophages were more frequently identified in CRS patients with nasal polyp (CRSwNP) compared to patients without nasal polyp (CRSsNP). Prophages belonging to Sa3int and Sa2int group were the most prevalent. Further, *S. aureus* isolates from CRSwNP patients often harboured Sa3int prophages encoding human immune evasion cluster genes. In summary, prophage encoded accessory genes may play a significant role in the pathogenicity of *S. aureus* and impact CRS disease phenotype as well as severity.

DATA SUMMARY

Genomes of previously sequenced *S. aureus* (*n*=58) were retrieved from the local database. The sequences are also publicly available in NCBI Genome Depository under BioProject Accession Number: PRJNA436815. The additional sequences from control group are included as Data S1 (available in the online version of this article) and complete information on prophage (analysed: November 2020) is available as Data S2. All supporting data are publicly available for download at figshare (https://doi.org/10.6084/m9.figshare.16590359).

Impact Statement

Mobile genetic elements like prophages alter the genetic make-up and profoundly impact the virulence of *S. aureus* including but not limited to toxin secretion, biofilm formation, niche adaptation. As chronic rhinosinusitis (CRS) is often associated with persistence of *S. aureus*, it is crucial to identify prophages predominantly circulating in CRS patients and associated virulence factors. Those prophage-associated virulence factors could be predictive of CRS disease progression and severity and assist in identifying appropriate therapeutic interventions to quench clonal expansion and survivability of pathogenic lysogens. We report that *S. aureus* clinical isolates carrying Sa3int group prophages encoding human immune evasion factors like *scn, chp, sak, sea* were predominantly found in CRS patients with nasal polyps. These findings provide a platform for investigation into the contribution of those factors in the pathophysiology of CRS and their potential use as diagnostic, prognostic and therapeutic targets. Our findings will not only be of interest to clinicians but also will equally be important in disease epidemiology particularly inflammatory diseases, including but not limited to CRS as prophage associated toxins have been known to cause deadly outbreaks in the past.

INTRODUCTION

Chronic rhinosinusitis (CRS) is a multifactorial inflammatory disease of the sinonasal mucosa associated with relapsing infections [1]. Phenotypically, CRS is broadly differentiated into CRS with nasal polyps (CRSwNP) and CRS without nasal polyps (CRSsNP). Development of polyp tissue results in reduced nasal airflow and anatomical obstruction of the sinus drainage pathways which exacerbates CRS symptoms and is often mirrored by high levels of inflammation seen on computed tomography (CT) [2]. The pathophysiology of CRS remains unclear and no single genetic and/or environmental factor has been solely linked to the development of this disorder. In the last decade, there has been increasing evidence that bacterial virulence, the presence of microbial mucosal biofilms and microbiome dysbiosis can affect the persistence of symptoms, disease severity and post-operative recovery [3–6]. Although *Staphylococcus aureus* is considered a commensal capable of colonizing diverse ecological niches within human and animals and is carried by ~30% of the human population asymptomatically [7, 8], it is also one of the most invasive, highly pathoadaptive, opportunistic pathogens and etiological agent of diverse human and animal maladies including CRS. An increased colonization of *S. aureus* was demonstrated in patients with CRSwNP (64%) but not in patients with CRSsNP (33%) versus control (20%) patients suggesting contribution of *S. aureus* in CRS [9, 10]. Of further concern is the emergence and spread of methicillin resistant *S. aureus* (MRSA) and vancomycin resistant *S. aureus* (VRSA). The successful pathoadaptive evolution of virulent *S. aureus* is largely due to acquisition of large mobile genetic elements (MGEs) carrying virulence, toxin and resistance genes [11]. Such MGEs include plasmids, transposons (Tn), insertion sequences (IS), *S. aureus* pathogenicity islands (SaPIs), staphylococcal cassette chromosomes (SCCs) and (pro)phages. They can be exchanged between strains by horizontal gene transfer (HGT) and/or transferred to progeny through vertical gene transfer (VGT) [12–14]. Among multiple MGEs contributing

to virulence and pathogenicity of *S. aureus*, active prophages are one of the most efficient elements, that can mobilize 'clusters' of genes between genetically related clones [15–17].

In contrast to virulent (lytic) phages that are unable to insert their DNA into the bacterial host genome, temperate (lysogenic) phages can integrate their DNA into the bacterial host genome or occasionally exist as extrachromosomal DNA. Once stably integrated, the phage DNA is named 'prophage' and the host bacteria becomes 'lysogenic'. By doing so, temperate phages can introduce and mobilize resistance genes, toxins and phage-associated virulence factors (VFs) via phage mediated transduction [18], thereby altering bacterial genomic information and phenotype [19]. Such prophages can switch to the lytic cycle through a variety of mechanisms, producing infectious phage particles provided they have all the functional and structural genes required for genome excision, replication and phage particle assembly. One mechanism by which this lysogenic to lytic switch can occur is because of biotic and/or abiotic stresses which gives rise to DNA damage (UV exposure, antibiotics, chlorine, H_2O_2) [20–22]. In other phage, the switch to lytic development can be a stochastic decision, influenced by the density of phages in the environment [23, 24].

As more genomic sequences of clinical isolates become available, a considerable number of prophages are discovered recently that account for as much as 20% of the host genome [25]. Lysogens can release phages as weapons against other invading bacterial strains, accelerate clonal expansion of virulent bacteria through lateral transduction and/or trigger the immune system to produce specific antibodies that may worsen inflammatory disease [26]. Further, Li, Wang [27] demonstrated that integration of specific prophage φSA169 in methicillin-resistant *S. aureus* increased biofilm formation, enhanced δ-hemolysin activity and reduced vancomycin sensitivity.

There is growing evidence that accessory genes carried by prophages of *S. aureus* significantly modulate bacterial fitness as they carry multiple VFs. These VFs include human immune evasion cluster (IEC) comprising the genes *sak, chp, scn* and *sea/sep* which encodes staphylokinase, chemotaxis inhibitory protein of *S. aureus* (CHIPS), staphylococcal complement inhibitor (SCIN) and enterotoxin A/P (SEA or SEP) respectively in different combinations [28]. In addition, they also comprise a bi-component cytotoxin Panton-Valentine leukocidin (PVL, *luk* F/S) and related leukocidins (*luk* M/F) involved in necrotic infections; and exfoliative toxin A (*eta*) involved in skin infections [29, 30]. Furthermore, phage-associated virulence is strongly associated with the phage 'integrase' (*int* types) type in *S. aureus,* Sa3int type being the most abundant among nasal colonisers [31]. Further, expression of prophage-associated VFs varies according to the infection site and external stimulus. Despite the widespread presence of prophages in *S. aureus* clinical isolates and their role in pathoadaptive gene acquisition, mobility, virulence and pathogenicity, prophages are one of the most understudied elements. Knowledge of prophage presence and organisation in *S. aureus* clinical isolates and their potential role in CRS disease pathophysiology is not known. Previous research by our team on *S. aureus* core genome (*n*=58) found even distribution of virulence genes in CRS sub-groups (CRSsNP versus CRSwNP) and their origin, status and/or evolutionary association was elusive. Further, no significant difference in pathogenic gene abundance was observed between CRSsNP and CRSwNP [32].

Here, we implement an *in silico* approach to re-analyse the data focussing primarily on accessory genes (particularly prophages) in the genomes of 58 *S. aureus* clinical isolates from CRS patients. We report the discovery of 211 prophage-like regions and provide detailed insight into prophage types, genomics and their phylogenetics. We further explore the contribution of these prophages to the bacterial genome, major VFs they encode and investigate a possible contribution of prophage-rich lysogens in CRS disease status and severity.

METHODS

Bacterial isolates and measure of disease severity

S. aureus clinical isolates (CIs) were obtained from patients with CRS and non-CRS patients at the time of endoscopic sinus surgery, isolated by an independent laboratory (Adelaide Pathology Partners, South Australia) and stored at −80 °C in glycerol stocks (20%). CRS patients fulfilled the CRS diagnostic criteria according to the European Position Paper on Rhinosinusitis and Nasal Polyps (EPOS2020) [1]. Control patients did not have symptoms of CRS with no evidence of mucosal inflammation on endoscopic evaluation of the nasal and paranasal sinuses. CRS type (CRSsNP or CRSwNP) was determined based on presence/absence of nasal polyp tissue and disease severity was scored based on Lund–Mackay (LMK) staging system [33] by the surgeon (PJW and AJP) at the time of clinical isolate collection.

Prophage prediction and characterization within *S. aureus* genomes

Integrated prophage regions were predicted using PHASTER (Phage Search Tool – Enhanced Release) (https://phaster.ca/) with default settings (Text S1) and the regions were classified as intact, questionable and incomplete [34] which roughly translates to active (intact) and inactive (questionable and incomplete). Further, prophage sequences, putative prophage attachment sites (*attL*/*attR*), the GC percentage, size, protein hit (total ORFs), most similar phage and details of protein family were manually identified and extracted from the output. Most similar prophage was further queried against viral nr/nt NCBI database (taxid:10239) and Virus-Host DB (https://genome.jp/virushostdb/) to predict the prophage family and genus based on their maximum homology. All visualisations were performed using GraphPad Prism 9 (Ver 9.1), R (Ver 4.0.0) in RStudio (Ver 1.3.1093) using the R package 'ggplot2' (Ver 3.3.2) unless stated otherwise.

In silico detection of virulent and antimicrobial resistance genes within prophages

A concatenated DNA sequence file (FASTA) of prophage sequences was created. Antimicrobial resistance genes (ARGs) and virulence factors (VFs) associated with *S. aureus* were scanned within the prophage sequences using ResFinder 4.1 [35] and VFanalyzer [36] respectively. The biological (pathogenesis) and/or molecular function for major VFs associated with prophage was assigned according to gene ontology (GO) knowledgebase through UniProtKB (https://uniprot.org/).

Multiple sequence alignment (MSA) and prediction of major phage-associated VF clusters

The complete sequences of predicted prophages were extracted and concatenated in a separate file (FASTA) with most similar phage hit as a reference. Groups with more than four intact prophage hits were considered. The prophage sequences of all intact and questionable prophages were aligned with the reference sequence (extracted from NCBI) using progressive Mauve in R package 'genoPlotR' [37]. Only major pathogenic genes were visualised in MSA analysis. Incomplete prophages (scores <70) were excluded. To determine the IEC clusters, a customised BLAST database was created with the amino acid sequences of the five possible genes (*sea, sep, chp, sak, scn*). Each intact Sa3int group prophage genome was then compared against the database using BLAST, specially using the blastx algorithm. A threshold of 95% identity was chosen as the cut-off for presence of the genes. The prophages were then assigned into the clusters based on the classification by van Wamel *et al.* (2006) [28].

Genome assembly and phylogenetics

For *S. aureus*, genomes were assembled using Unicycler (v 0.4.8) and annotated with Prokka (v 1.14.6). Assemblies were quality controlled using QUAST (v 5.0.2) [38–40]. CIs were grouped into clonal complexes (CC) by assigning Multi-Locus Sequence Typing using the programme MLST [41]. The core genome of *S. aureus* isolates was inferred with Roary (v 3.7.0) with the Prokka annotations as input [42]. This core genome alignment was used to create a maximum likelihood phylogenetic tree using IQtree (v 2.0.3) [43]. Specifically, the resulting maximum likelihood tree was created using 1000 ultrafast bootstrap replicates, applying the SH-like approximate likelihood ratio test (Guindon *et al.*, 2010). For prophage phylogenetics, DNA sequences of all putative prophage were aligned using MAFFT7 (Multiple Alignment using Fast Fourier Transform, ver 7) [44] and a maximum likelihood tree was created with FastTree 2.1 [45]. Further, amino acid (aa) sequences for integrase genes were extracted from PHASTER annotations. Representative integrase sequences (Sa1int-Sa12int) were retrieved from NCBI [31, 46, 47], aligned along with query sequences using MAFFT and phylogenetic diversity was inferred using FastTree 2.1 in Geneious Prime 11.09 (ver 21.1, Biomatters Ltd. Auckland, New Zealand). All trees, unless specified, were visualized using iTOL V5 (https://itol.embl.de) [48]. The percentage identity heat-map matrix was also exported from Geneious Prime 11.09. Further, integrase sequences of unassigned phages were retrieved from Virus-Host DB (https://genome.jp/virushostdb/) and homology was inferred using similar approach as mentioned above.

Statistical analysis

Descriptive statistical methods were used to determine the frequency, percentage, and means while one-way ANOVA was used to compare between groups. Fisher's exact test (two-tailed) was used to determine significance of each prophage (intact) between CRSsNP/CRSwNP and lower/higher LMK severity groups. Unless mentioned, all statistical analyses were performed using GraphPad Prism 9 (ver 9.1) and $P<0.05$ was considered statistically significant. No statistical methods were used for predetermination of sample size and experiments were not randomized.

RESULTS

Prophages are significant components of *S. aureus* clinical isolates

Although *S. aureus* has often been associated with CRS, phylogenetics analysis has failed to correlate any specific sequence type (ST) or clonal complex (CC) with CRS disease severity and/or phenotype including methicillin resistance (Figs 1 and S1). We analysed genomes of *S. aureus* clinical strains isolated from CRS patients (n=58) and control (n=9). All CIs were predicted to be lysogenic as they carried at least one recognisable prophage (range=1–10, average=3.63 prophages/strain) (Figs 1a and S2a). All *S. aureus* from control patients had at least one intact prophage (Fig. S2a). Among 58 strains isolated from CRS patients, 53 (91%) were poly-lysogenic (Fig. 1a), 47 (81%) harboured at least one 'intact' prophage, four (7%) had only 'incomplete' prophages whereas seven (12%) had a combination of questionable and incomplete prophages (Figs 1b and S2b). Altogether, 211 prophage-like sequences were predicted from 58 *S. aureus* genomes (Fig. 1a, c). Out of those, 64 (30%, average=1.1/strain) were intact, 33 (16%, average=0.57/strain) were questionable and, 114 (54%, average=1.96/strain) were incomplete. The mean genome size of intact, questionable and incomplete prophage was 44.30, 27.83 and 17.83 kb respectively (Fig. 1c). Prophages accounted for a maximum of 220.8 kb which amounts to 7.7% (average=3.57%) of the total bacterial genome (Fig. 1d). Although there was no significant difference in average prophage percentage between CRSsNP and CRSwNP groups, density of intact prophages were significantly higher in CRSwNP group (Fig. 1e) and most of them belonged to size range of 20–70 kb (Figs 1f and S2c, d) The average GC% of the bacterial genome was 32.7% (range=32.6–32.8), whereas the average GC% of intact and incomplete prophages was 33.54% (range=31.93–36.31) and 30.92 (range=25.56–34.93) respectively (Fig. 1g).

From the 58 CRS patients, 28 were classified as CRSsNP and 30 as CRSwNP. Although the average number of prophage regions was similar between CRSsNP (3.64/strain) and CRSwNP (3.63/strain), intact prophages were more frequently

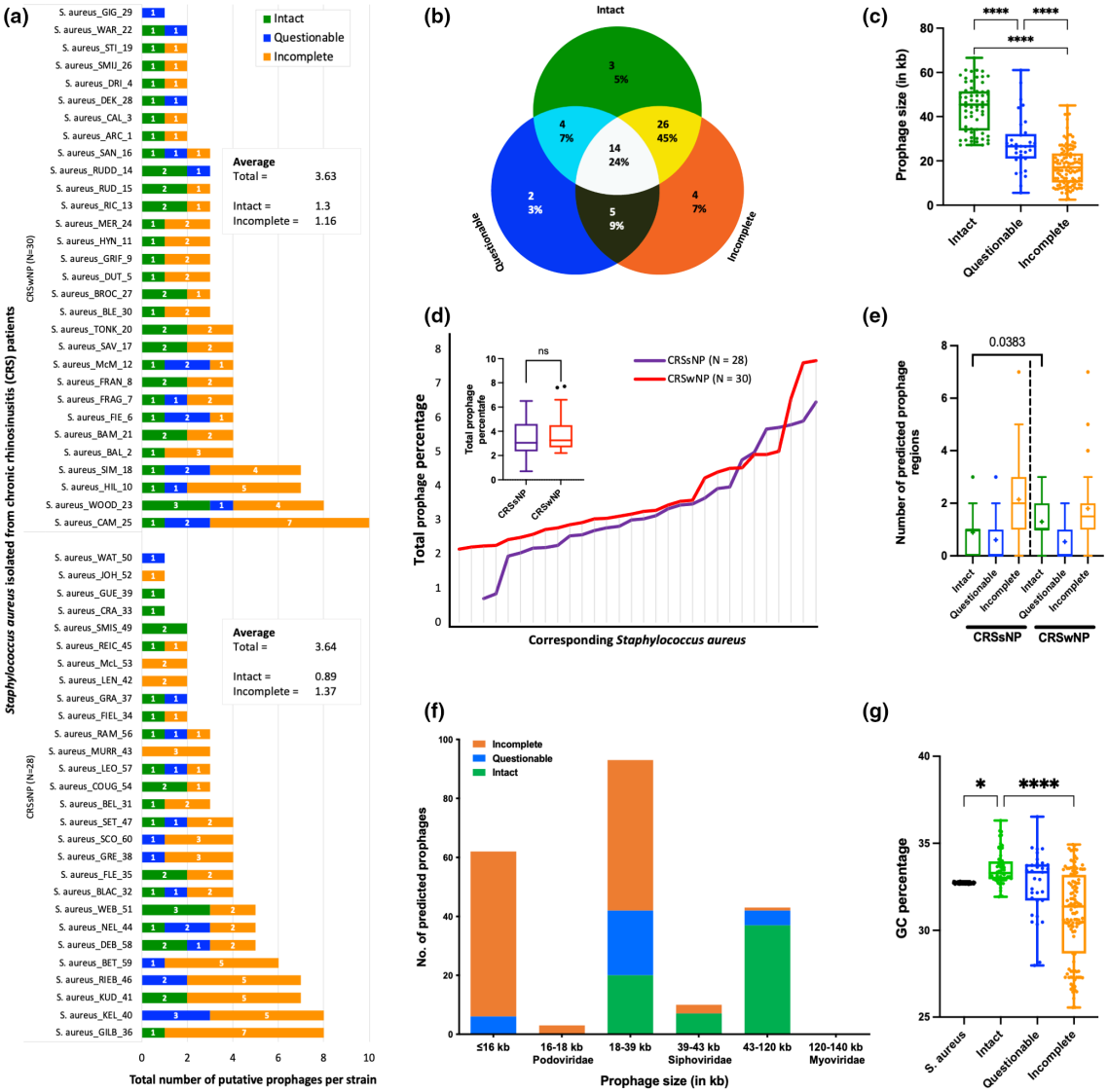

Fig. 1. Prediction and distribution of prophages from *S. aureus* genome. (a) Among 58 clinical strains, 53 (~91%) were poly-lysogenic, while only five strains had single prophage. Out of total 211 (3.6 prophages/strain) predicted prophages, 64 (30.33%) were intact, 33 (15.64%) were questionable and 114 (54.03%) were incomplete. The numbers inside the bar represents number of prophages. (b) Venn-diagram representing distribution of prophages. Out of 58 strains, 47 harboured at least one intact prophage, four had only incomplete prophage while seven had mix of questionable and incomplete prophages but lacked intact prophages. (c) Distribution of predicted prophages according to their size. The average size of prophages decreased from intact to incomplete. The solid red line represents median. (d) The genome shares of prophages on the host genome ranged from 0.7–7.7% (average=3.6%). The box plot on the inset shows difference in prophage genome between CRSsNP and CRSwNP. Although prophage content in CRSwNP was relatively higher, the difference was not statistically significant. (e) Distribution of prophages between CRSsNP and CRSwNP. The number of intact prophages was significantly higher in CRSwNP (P=0.038, Welch's t-test). (f) Distribution of candidate prophage regions based on their predicted size and reference genome size. All intact prophages fell in size range closer to *Siphoviridae* (39–43 kb). (g) Comparison of GC% across host genome, combined prophage, and different types of prophages. The average GC% of the host (*S. aureus*) was 32.72% compared to 31.98% of the combined prophages. Further, the average GC% of intact, questionable and incomplete prophages were 33.5%, 32.7 and 30.9% respectively.

identified in CRSwNP (29/30, 96.6%, average=1.3/strain) than in CRSsNP 18/28, 64.28%, average=0.89/strain) (P=0.0021, Fisher's exact test) (Table 1). Similarly, intact prophages were more frequent (29/32, 91%, average=1.21/strain) in *S. aureus* strains isolated from CRS patients with more severe disease (LMK score >12) compared to those with less severe disease (LMK score <12) (16/22, 72%, average=1.04/strain) even though statistical significance was not reached (P=0.1363, Fisher's exact test, Table 1). Similar analysis of *S. aureus* isolated from 'control' group (n=9) revealed that at least

Table 1. Correlation between CRS disease status/severity and presence of prophages* in *S. aureus* recovered from CRS patients

CRS disease type/severity		Intact prophage			*P*-value (Fisher's exact test)
		Average density	Present	Absent	
Disease phenotype	CRSsNP (*N*=28)	0.89	18	10	0.0021 (significant)
	CRSwNP (*N*=30)	1.30	29	1	
Disease severity (LMK)†	LMK ≤12 (*N*=22)	1.04	16	6	0.1363
	LMK >12 (*N*=32)	1.21	29	3	
Control (*N*=9)		1.33	9	0	

*Only intact prophages considered as they are likely functional and comprise complete sets of genes (including virulence genes), have ability to switch between lytic-lysogenic cycle and pass virulence to other strains.
†LMK scores only available for 54 patients. Refer to Fig. S4/data.
LMK, Lund–Mackay score; NA, Not available because 'control' groups are not scored for LMK.

one intact prophage was present in all strains (9/9, 100%, average=1.33/strain), indicating prophage associated adaptation is common in human nasal colonization and prophage retention and/or gain may occur at the later stage.

Prophage genomes significantly contribute to *S. aureus* strain variability

We then compared the distribution and abundance of phage-hit genes across intact, questionable and incomplete prophages through a heat-map according to their corresponding structural and/or functional gene families assigned by PHASTER. Among 211 predicted prophages, only 118 (56%) were flanked by at least one pair of attachment sites *(attL/attR)* (Data S2). Similarly, head-like protein genes were found in 125/211 (59%) followed by tail in 92/211 (44%) and capsid in 51/211 (24%). Integrase genes were found in 97/211 (46%) followed by portal in 86/211 (41%), terminase in 75/211 (36%) prophages. Lysin, protease, transposase and recombinase were less frequent and found only in 28/211 (13%), 25/211 (12%), 19/211 (9%) and 2/211 (1%) prophages, respectively (Fig. 2a, b). Compared to intact prophages, incomplete prophages often lacked tail, capsid, portal, terminase, lysin and protease genes. Further, transposases were relatively more frequent in incomplete prophages (15/114, 13%) than in intact prophages (2/64, 3%) whilst recombinase genes were found exclusively in incomplete prophages (Fig. 2c). However, as genomes are spliced at these regions during short-read sequencing, this may be underestimated and thus carefully reported. Altogether, 7523 open reading frame hits (ORF-hits) (average=35.65 ORFs/prophage, including hypothetical proteins) were predicted from 211 prophage regions (Data S2). Out of those, 3655 (48%), 1177 (16%) and 2691 (36%) were in intact, questionable and incomplete prophages respectively (Table S1). Further, 6693 (89%) had known functions, mainly involved in phage structure, transcription, replication, and lytic/lysogenic regulation, while 830 (11%) were 'hypothetical' with unknown function. The total number of phage-hit proteins (including hypothetical) and the total prophage genome size significantly correlated with the size of *S. aureus* genome (P<0.0001, linear regression fit, Fig. 2d).

Gene density in a prophage is inversely proportional to its genome size

The number of phage-hit proteins in prophage genomes positively correlated with the size of the prophage (r^2=0.86, P<0.0001) (Fig. 3a) and the GC% was higher in larger prophage genomes (Fig. 3a). In addition, prophage sequences had a high gene density (average=1.43 genes/kb) (Fig. 3b) which was highest in smaller prophage sequences (genome size <10 kb) and those had relatively low GC% (Fig. 3b).

Most prevalent prophages were similar to *S. aureus* phages from the genus *Biseptimavirus*

Based on nucleotide homology, among 211 prophages, 196 (93%) were Staphylococcus prophage whereas 15 (7%) were non-Staphylococcus prophage (Fig. S3a, b). Altogether 44 different phage strains were found, mostly belonging to the *Siphoviridae* family (41/44, 93%) (Fig. 3c, d), out of which 36 were Staphylococcus phages while eight resembled non-Staphylococcus phages (Fig. 3c, indicated by star). Among the 44 prophage strains, five (PT1028, phiNM3, JS01, phiN315 and phi2958PVL) accounted for almost 51% (108/211) of the prophages and at least one of those was present in 54/58 (93%) isolates.

Further, 22/44 prophage strains (50%) (a total of 64 prophages) were found in intact form and none of those were non-Staphylococcal phages. The most abundant intact prophage was similar to Staphylococcus phage JS01 (14/64, 21.8%) followed by Staphylococcus phage phiNM3 (10/64, 15.6%), Staphylococcus phage phi2958PVL (9/64, 14.0%) and Staphylococcus phage phiN315 (4/64, 6.25%) (Fig. 3c). Further, among 196 Staphylococcal like prophages, most of them belonged to the genus *Biseptimavirus* (75/196, 38%) followed by *Phietavirus* (44/196, 22.4%) and *Triavirus* (22/196, 11.22%) (Table 2). Based on the most similar phage-hit, among 197 *S. aureus* prophages, most of the prophages

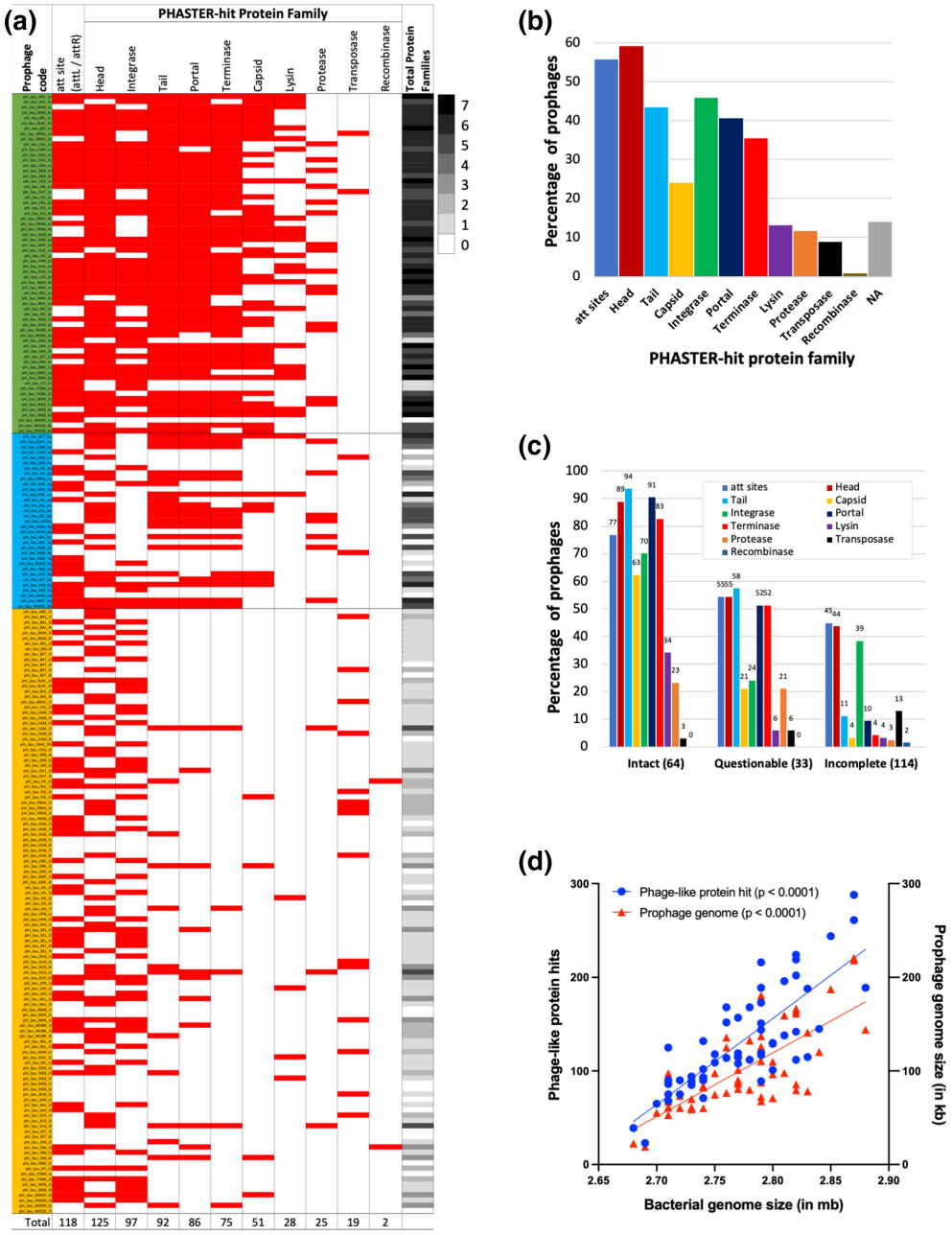

Fig. 2. Distribution of phage-like proteins (PLP) across different types of prophages. (a) Heat map of prophages and phage associated proteins in all *S. aureus* strains. Prophages (y-axis) are plotted in alphabetical order grouped according to their status (green=intact, blue=questionable, yellow=incomplete) against each protein hit (x-axis). Red boxes indicate the presence of the indicated protein. White spaces indicate the lack thereof. The numbers in the last column indicate total number of PHASTER-hit protein families and is also represented by gradient of black-colour. The number in last row indicates total number of prophages with corresponding protein-family hit. Please refer to the PDF of the figure and use the zoom function to identify names of prophages and proteins. (b) Among 211 prophages, at least one attachment site (attL/attR) was present in 118 (56%), while the most abundant structural protein was associated with head (125) followed by tail (92) and capsid (51). Similarly, the most abundant functional protein was integrase (97) followed by portal (86) and terminase (75). Lysin, protease, transposase and recombinase were found only in 28, 25, 19 and two prophages, respectively. (c) Comparison of phage-associated protein distribution between intact, questionable and incomplete prophages revealed that intact and questionable prophages completely lacked recombinase genes, and transposases were significantly enriched in incomplete prophage (compared to present only in two each in intact (3%) and questionable (6%) prophages. Arrows represent enriched proteins in incomplete prophages compared to the complete ones. (d) Correlation between host genome (*S. aureus*) vs number of phage-like proteins (PLPs) (P<0.0001, linear regression) and prophage genome (P<0.0001, linear regression). The gain of genome size is significantly contributed by prophage as the prophage content increases with increase in genome of the host.

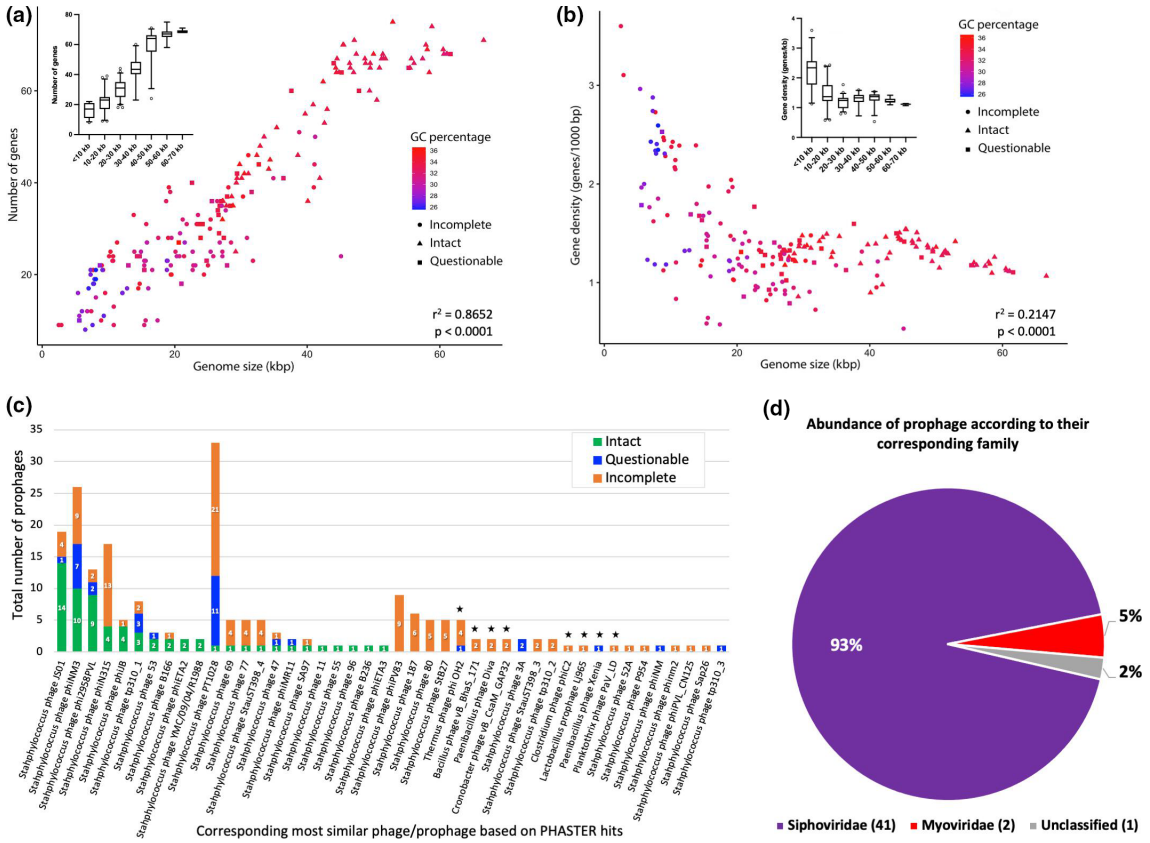

Fig. 3. Identification and characteristics of predicted prophages. (a) Correlation between number of genes, prophage genome size and GC%. The number of genes and GC% increases with increase in size of prophage genome indicating bigger prophages have more coding sites and high GC. (b) Correlation between gene density, prophage genome size and GC%. The gene density (genes/kb) is relatively high in smaller prophages accompanied by lower GC (higher AT), suggesting that they efficiently pack more genes within their small genome as compared to intact prophages. (c) Distribution of predicted prophages based on their most similar hit. Although 211 prophages were predicted by PHASTER, they all were most similar to 44 different phages available in the PHASTER database. Among 211, 108 (~51%) belonged to five most common temperate phages (Staphylococcus phage PT1028, Staphylococcus phage phiNM3, Staphylococcus phage JS01, Staphylococcus phage phiN315, Staphylococcus phage phi2958PVL) and almost 83% (175/211) of prophages were represented by 18 different strains of prophages. (Stars represent non-Staphylococcus phage-hits, and numbers inside bar represents total prophages of that type). (d) Among 44 (pro)phage hits, most of them (41, 93%) belonged to Siphoviridae family, two were from Myoviridae family (non-Staphylococcal) whereas one phage (PT1028) was unclassified till date.

were similar to Sa3int group (68, 35%) phages, followed by Sa2int (27,14%) and Sa1int (11, 6%) (Table 2).

S. aureus isolated from CRS patients with nasal polyps often carried Sa3int group prophages

We then performed phylogenetics analysis to identify integrase groups based on previously characterized representative sequences based on Goerke's classification [31]. Amino acid (aa) sequences of all 97 integrase genes identified in prophage regions were considered (intact=45/64, 70%, questionable=8/33, 24%, incomplete=44/114, 38%). Phage integrases were found in 55/58 (~95%) *S. aureus* strains and were always accompanied by the presence of attachment sites (data not shown). The most prevalent prophage type based on integrase gene polymorphism was Sa3int followed by Sa2int and Sa1int (Table 2, Fig. 4a, b). We further report an unassigned integrase group (~390 aa) in 16 incomplete prophages that did not relate with any of the major Sa1int-Sa12int groups but had 100% identity with tyrosine-type recombinase/integrase (NCBI Ref. Seq: WP_048667711.1, non-redundant protein sequences (nr) database) in *S. aureus* (Fig. S3c). Limiting the BLAST search within NCBI virus database (taxid:10239) showed 88.24% identity (query coverage=100%) with putative integrase from uncultured Caudovirales phage (GenBank: ASN72555.1) (Fig. S3d). Similar dot-matrix and phylogenetic analysis of lysin and tail-fibre genes showed limited polymorphism in *S. aureus* prophages (Fig. S5a–d). Details and amino acid sequences of representative integrase proteins are available as Data S2.

Further, 'intact' Sa3int prophages were significantly more prevalent in clinical isolates fromCRSwNPpatients than CRSsNP patients (Table 3, Figs 4a and S4). Specific Staphylococcus prophage phiNM3 (also belonging to Sa3int prophages) was

Table 2. Predicted Staphylococcal prophage*, associated integrase group, major virulence factors (VFs), corresponding phage genus and family based on maximum homology (as assigned by PHASTER)

Most similar phage hit	Integrase group^	Associated VFs†	No. of prophages			Prophage genus Total (IN, Q, IC)	Predicted family
			IN	Q	IC		
Staphylococcus phage PT1028	NA	NA	1	11	21	NA 33 (1, 11, 21)	Unclassified
Staphylococcus phage StB27	NA	NA	0	0	5	NA 22 (4, 0, 18)	*Siphoviridae*
Staphylococcus prophage phiN315	Sa3int	*sak, chp, scn, sep*	4	0	13		
Staphylococcus phage JS01	Sa3int‡	*sak, chp, scn, sep*[c]	14	1	4		
Staphylococcus phage phiNM3	Sa3int	*sak, chp, scn, sea*	10	7	9		
Staphylococcus phage StauST398-4	Sa3int‡	–	1	0	4		
Staphylococcus phage tp310-3	Sa3int	*sak, chp, scn*	0	1	0		
Staphylococcus phage tp310-1	Sa2int	*luk* S/F-PV	3	3	2	*Biseptimavirus* 75 (29, 12, 34)	
Staphylococcus phage phiPVL-CN125	Sa2int	*luk* S/F-PV	0	0	1		
Staphylococcus phage 77	Sa6int	–	1	0	4		
Staphylococcus prophage phiPV83	Sa5int	*luk* M, *luk* F-PV	0	0	9		
Staphylococcus phage P954	Sa7int	–	0	0	1		
Staphylococcus phage phi2958PVL	Sa2int	*luk* S/F-PV	9	2	2	*Triavirus* 22 (12, 5, 5)	
Staphylococcus phage YMC/09/04/R1988	Sa2int‡	–	2	0	0		
Staphylococcus phage 47	Sa2int	–	1	1	1		
Staphylococcus phage 3A	NT	–	0	2	0		
Staphylococcus phage tp310-2	Sa6int	–	0	0	2		
Staphylococcus phage phiJB	Sa6int	–	4	0	1	*Phietavirus* 44 (18, 3, 23)	
Staphylococcus phage B166	Sa1int‡	–	2	0	1		
Staphylococcus phage phiETA2	Sa1int	*eta*	2	0	0		
Staphylococcus phage SA97	Sa1int‡	–	1	0	1		
Staphylococcus phage 55	Sa1int	–	1	0	0		
Staphylococcus phage B236	Sa1int‡	–	1	0	0		
Staphylococcus phage phiETA3	Sa1int	*eta*	1	0	0		
Staphylococcus phage Sap26	Sa1int‡	–	0	0	1		
Staphylococcus phage 69	Sa5int	–	1	0	4		
Staphylococcus phage 11	Sa5int	–	1	0	0		
Staphylococcus phage 187	Sa5int	–	0	0	6		
Staphylococcus phage phiNM1	Sa5int	–	0	1	0		
Staphylococcus phage 53	Sa7int	–	2	1	0		
Staphylococcus phage phiNM2	Sa7int	–	0	0	1		
Staphylococcus phage 96	Sa9int	–	1	0	0		
Staphylococcus phage StauST398-3	Sa9int‡	–	0	0	2		
Staphylococcus phage 80	Sa6int	–	0	0	5		
Staphylococcus phage 52A	Sa6int	–	0	0	1		
Staphylococcus phage phiMR11	Sa12int	–	1	1	0		
	Total (Staphylococcal prophages)		**64**	**31**	**101**		

Continued

Table 2. Continued

Most similar phage hit	Integrase group^	Associated VFs†	No. of prophages			Prophage genus Total (IN, Q, IC)	Predicted family
			IN	Q	IC		

*Based on most-similar phage-hit in PHASTER. Non-Staphylococcus prophage hits excluded.
†Reference: Goerke *et al.* 2009 [31], Kahánková *et al.* 2010 [46], Varga *et al.* 2016. Colour coded according to *S. aureus* phage integrase group.
‡Predicted from this study based on integrase gene homology and phylogeny of the reference sequence with reference sequences of integrase gene (Fig. S4a, b).
Different colours represent different group of integrases found based on most similar hit.
IN, Intact (or complete); Q, Questionable; IC, Incomplete; NA, Not assigned, NT, Non-typeable.

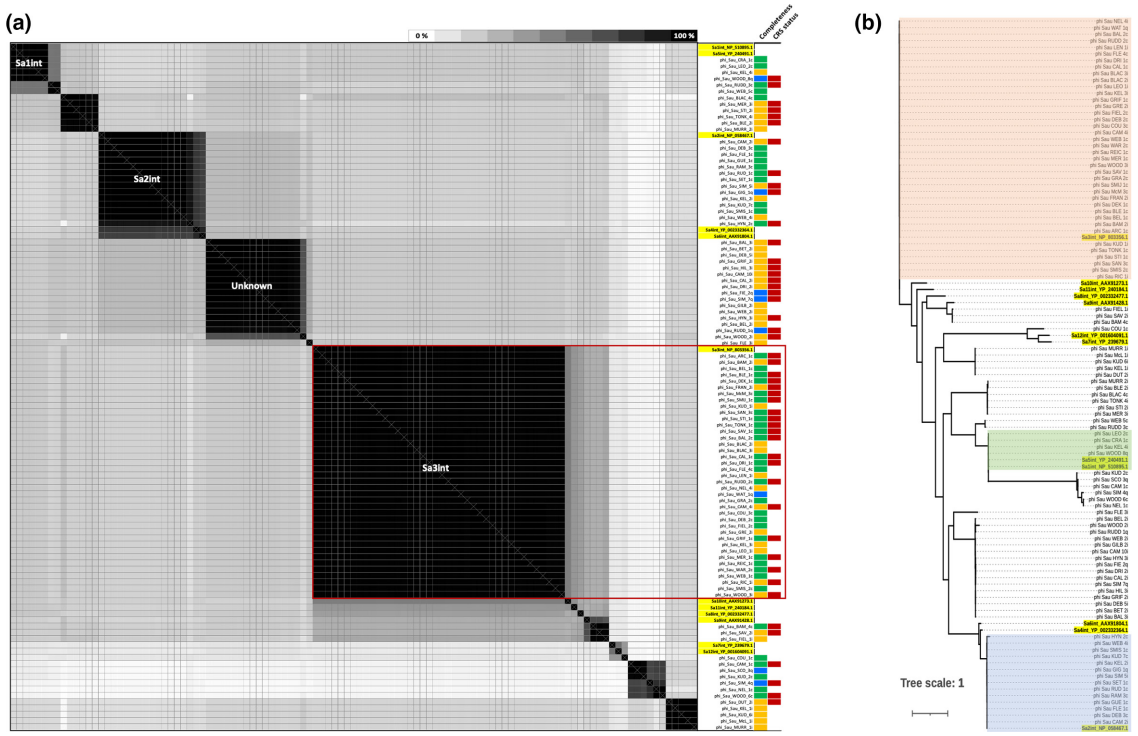

Fig. 4. Percentage identity dot-matrix and phylogenetics of integrase. (a) Percentage identity dot-matrix of integrase (*N*=97) gene. The gradient bar at the top-right represents percentage identity, darkest being 100%. The green, blue and orange bar represents completeness (intact, questionable and incomplete respectively) of the corresponding prophage. The red bar represents positive polyp status (CRSwNP) of the corresponding *S. aureus*. (b) Phylogenetics of integrase (*N*=97) gene. Together, these findings reveal that Sa3int group of phage infection (as prophage) is the most widely distributed in *S. aureus* clinical isolates isolated from chronic rhinosinusitis patients followed by Sa2int and Sa1int.

significantly more prevalent in patients within high disease severity compared to those with low disease severity (LMK ≥12 vs LMK <12, $P = 0.0073$, Fisher's exact test) (Table 3).

Prophages of *S. aureus* carry virulence factors but not antimicrobial resistance genes

Prophages carried multiple phage-associated virulence factors. These included *sak, scn, chp, hlb, luk*G/H, *seg, seln, selu, sei, selm, selo, spl*C, *eap/map, sea* (Table 4). The most abundant phage associated VFs were *sak, scn, hlb, entA,* and *chp* found in 45, 40, 37, 36 and 22 prophages respectively. All seven types of serine protease-like proteins (*slp*A/B/C/D/E/G/H) were found within prophage sequences suggesting them to be phage associated. VFs that are known to be human immune evasion factors such as *scn, chp* and *sak* were mostly present in prophages belonging to Sa3int or Sa3int homologues (JS01, phiNM3, phiN315) while prophages similar to Sa2int group (phi2958PVL) lacked those genes (Fig. 5a–d). IEC typing of all *S. aureus* strains and intact Sa3int prophages did not correlate with any specific type with CRS disease presentation (Table 5). Further, antimicrobial resistance genes (ARGs) were not identified within any of the prophage genomes in any of the *S. aureus* strains although 15/67 (22% including control group) were MRSA. A complete list of common VFs and other phage associated accessory genes is shown in Table 4 and IEC type of *S. aureus* and intact Sa3int prophages is elaborated in Table 5. Further, multiple sequence alignment (MSA) of prophages with the most similar phage-hit as a reference sequence confirmed that Sa3int group prophages (JS01, phiNM3, phiN315) consistently carried pathogenic

Table 3. Distribution of prophage, integrase typing among various groups of patients based on polyp status and Lund–Mackay severity score (LMK)

Disease status	Prophage groups/strains	No. of strains having intact prophages*			*P*-value (Fisher's exact test between CRSsNP and CRSwNP)
	Prophage strains based on integrase group	**Control (*N*=9)**	**CRSsNP (*N*=28)**	**CRSwNP (*N*=30)**	
Control / CRSsNP / CRSwNP	Sa3int	3 (33%)	7 (25%)	21 (70%)	0.0008 (significant)
	Sa2int	4 (44%)	10 (36%)	4 (13%)	0.0667
	Sa1int	2 (22%)	3 (11%)	5 (17%)	0.7073
	Individual phage strain				
	Staphylococcus phage JS01 (Sa3int)†	0	3 (11%)	11 (36%)	0.0331 (significant)
	Staphylococcus phage phiNM3 (Sa3int)	0	4 (14%)	6 (20%)	0.7316
	Staphylococcus phage phi2958PVL (Sa2int)	1 (11%)	6 (21%)	3 (10%)	0.2904
	Prophage strains based on integrase group	**Control (*N*=9)**	**LMK ≤12 (*N*=22)**	**LMK >12 (*N*=32)**	
LMK≤12/ LMK>12	Sa3int	NA	9 (41%)	17 (53%)	0.4180
	Sa2int	NA	6 (27%)	8 (25%)	1.0000
	Sa1int	NA	6 (27%)	2 (6%)	0.0512
	Individual phage strain				
	Staphylococcus phage JS01 (Sa3int)†	NA	8 (25%)	5 (16%)	0.1092
	Staphylococcus phage phiNM3 (Sa3int)	NA	0	9 (28%)	0.0073 (significant)
	Staphylococcus phage phi2958PVL (Sa2int)	NA	3 (14%)	6 (19%)	0.7230

*Only intact prophages considered. The integrase group is based on corresponding integrase group of phage identified as most similar hit by PHASTER through maximum homology.
†Identified from this study (Fig. S4a, b).
Please refer to Fig. S5 for complete list of prophage distribution.
LMK, Lund–Mackay score; na, Not available because 'control' groups are not scored for LMK.

IEC genes (*sak, chp, scn*) which were more conserved and uniformly distributed across intact prophages (Fig. 5a–c). In contrast, Sa2int group prophage (phi2958PVL) lacked IEC genes (Fig. 5d).

Prophage phylogenetics

Phylogenetic analysis based on maximum likelihood revealed three distinct evolutionary lineages of prophages with more diversified sub-clusters (Fig. 6). There was a heterogeneous distribution of intact, questionable and incomplete prophages across the three major clusters. Further, within clusters, there were many highly unrelated sub-clusters and singletons representing both intact and incomplete prophages. No intact prophages found in a same strain were found to be phylogenetically related (clustered) (Fig. 6).

DISCUSSION

This study demonstrated that all 58 *S. aureus* CIs from CRS patients carried at least one recognisable prophage, with a total of 211 prophage-like regions identified from the cohort. The majority of those were similar to temperate phages belonging to the *Siphoviridae* family, more specifically, the *Biseptimavirus* genus. The ubiquitous presence of prophages in *S. aureus* clinical isolates and strong positive correlation of prophage size and phage-hit proteins with the bacterial genome size indicate that the acquisition of prophage-encoded genetic

Table 4. Major virulence factors and their GO* annotation encoded by *S. aureus* prophages

VF class	Virulence factor	Related genes	No. of prophages	GO* annotation (biological process)
Immune evasion cluster (IEC)	Staphylococcal complement inhibitor (SCIN)	*scn*	40	pathogenesis
	Chemotaxis inhibitory protein (CHIPS)	*chp*	22	pathogenesis
	Staphylokinase	*sak*	45	pathogenesis
Enzyme	Serine protease	*sspA*	0	
	Serine protease-like proteins	*splA*	8	hydrolase and protease†
		splB	15	
		splC	7	
		splD	1	
		splE	9	
		splF	3	
Toxins	Delta-hemolysin	*hld*	7	pathogenesis
	Leukocidin	*luk* E/D	10	pathogenesis
	Enterotoxins (SEs)	*entA (sea)*	36	pathogenesis
		entB (seb)	8	pathogenesis
		entC	17	biosynthetic process
		entD (sed)	20	pathogenesis
		entE (see)	7	pathogenesis
		entG (seg)	13	pathogenesis
		entH (seh)	1	pathogenesis
		seln, selu, selu2	16 each	NA
		yent2	16	pathogenesis
		sei	15	pathogenesis
		selm, selo	15 each	NA
	Exfoliative toxin A	*eta*	3	pathogenesis
	Toxic shock syndrome toxin	*tst (tsst)*	5	pathogenesis
Other (non-virulent, prophage associated, responsible for successful prophage excision and induction)	Cell wall hydrolase	*lytN*	63	cell wall organization
	Tyrosine recombinase	*xerC*	39	cell division, transposition
	ssDNA-binding protein A	*ssbA*	38	DNA repair, replication, recombination
	Chromosome partition protein	*smc*	35	chromosome condensation, DNA replication, sister chromatid cohesion
	ATP-dependent *clp* protease proteolytic subunit	*clpP*	33	serine-type endopeptidase activity[b]
	DNA recombination protein	*recT*	33	NA

Continued

Table 4. Continued

VF class	Virulence factor	Related genes	No. of prophages	GO* annotation (biological process)
	60 kDa chaperonin	*groL*	22	protein refolding
	DNA replication protein	*dnaC*	21	DNA replication, synthesis of RNA primer
	ten kDa chaperonin	*groS*	21	protein folding

*Gene ontology.
†GO Molecular function
Pease refer to supplementary data for complete list of virulent and non-virulent gene hits in prophage sequence.
The gene name in parenthesis indicates the alternative name.
NA, Not categorized according to GO knowledgebase.

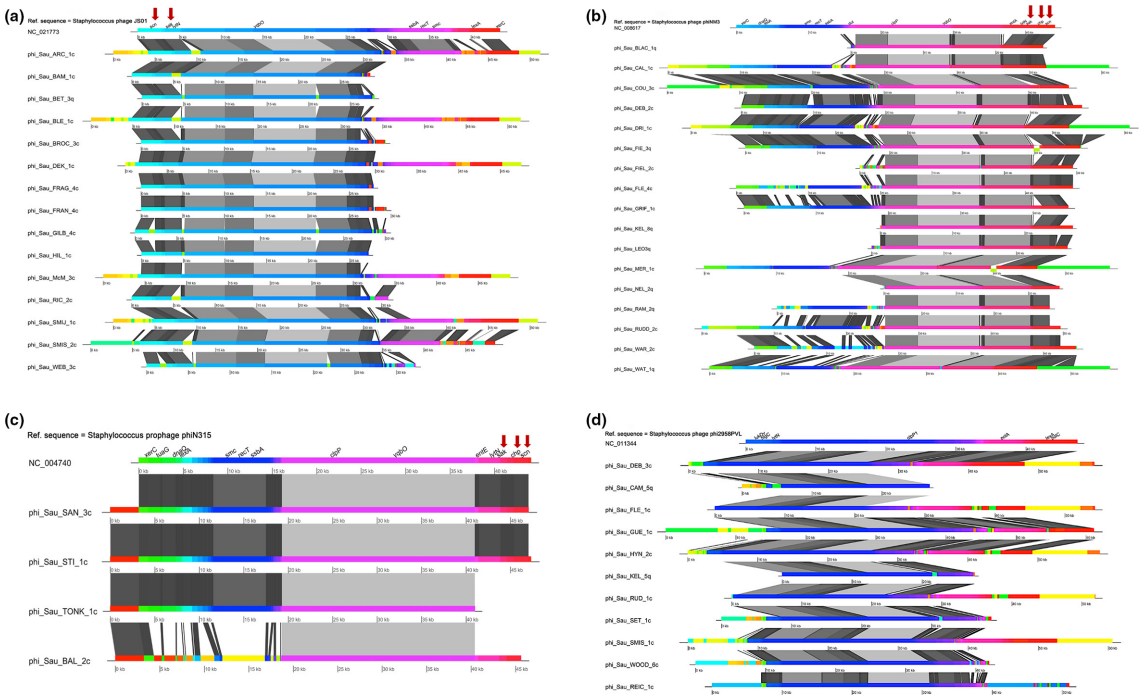

Fig. 5. Multiple sequence alignment (MSA) of predicted prophages (intact and questionable) using progressive MAUVE against most similar phage-hit as a reference sequence. (a) Sequence alignment of prophages with reference sequence Staphylococcus phage JS01 (Sa3int). (b) Sequence alignment of prophages with reference sequence Staphylococcus phage phiNM3 (Sa3int). (c) Sequence alignment of prophages with reference sequence Staphylococcus phage phiN315 (Sa3int). (d) Sequence alignment of prophages with reference sequence Staphylococcus phage phi2958PVL (Sa2int). The downward pointing red-arrow represents the immune evasion cluster (IEC) genes, the same colour between different prophage sequence indicates homology between prophages and the dark-grey band below every sequence represents percentage identity with the previous sequence. Please use zoom function from the PDF image for other individual genes.

material in *S. aureus* is common and likely an important driver of *S. aureus* evolution and host adaptation. This further implies that genome plasticity between *S. aureus* strains is likely to be driven in part by variability in temperate phage infection and integration. This process may improve the bacterial fitness and adaptation to the host environment potentially long term as these integrated phage DNA can pass to progeny. Further, a significant correlation between the prevalence of intact Sa3int group prophages carrying IEC genes including enterotoxins in *S. aureus* CIs from CRSwNP indicate that prophage associated VFs may contribute to the CRS disease severity and phenotype. Also, *S. aureus* prophages lacked AMR genes indicating phage-mediated spread of AMR genes is unlikely to be a major driver of antimicrobial resistance in the *S. aureus* population in this region (South Australia).

Clinical strains are usually laden with prophages [49] and multiple prophage encoded genes impacting the ability of

Table 5. Prevalence of different immune evasion cluster (IEC) types* in *S. aureus* and intact Sa3int (IEC) prophages

Sample	Immune Evasion Cluster (IEC) Type							
	–	A	B	C	D	E	F	G
		sea, sak, chp, scn	*sak, chp, scn*	*chp, scn*	*sea, sak, scn*	*sak, scn*	*sep, sak, chp, scn*	*sep, sak, scn*
Control (*N*=9)	0 (0%)	0 (0%)	3 (33%)	1 (11%)	1 (11%)	4 (44%)	0 (0%)	0 (0%)
CRSsNP (*N*=28)	2 (7%)	3 (11%)	13 (46%)	2 (7%)	3 (11%)	3 (11%)	0 (0%)	2 (7%)
CRSwNP (*N*=30)	1 (3%)	2 (7%)	8 (27%)	1 (3%)	6 (20%)	8 (27%)	3 (10%)	1 (3%)
Intact Sa3int prophages (*N*=28)	7 (25%)	1 (4%)	5 (18%)	1 (4%)	5 (18%)	4 (14%)	2 (7%)	3 (11%)

*IEC typing is based on presence/absence of IEC genes (*sak, chp, scn, sea/sep*) based on van Wamel *et al.* (2006) [28].

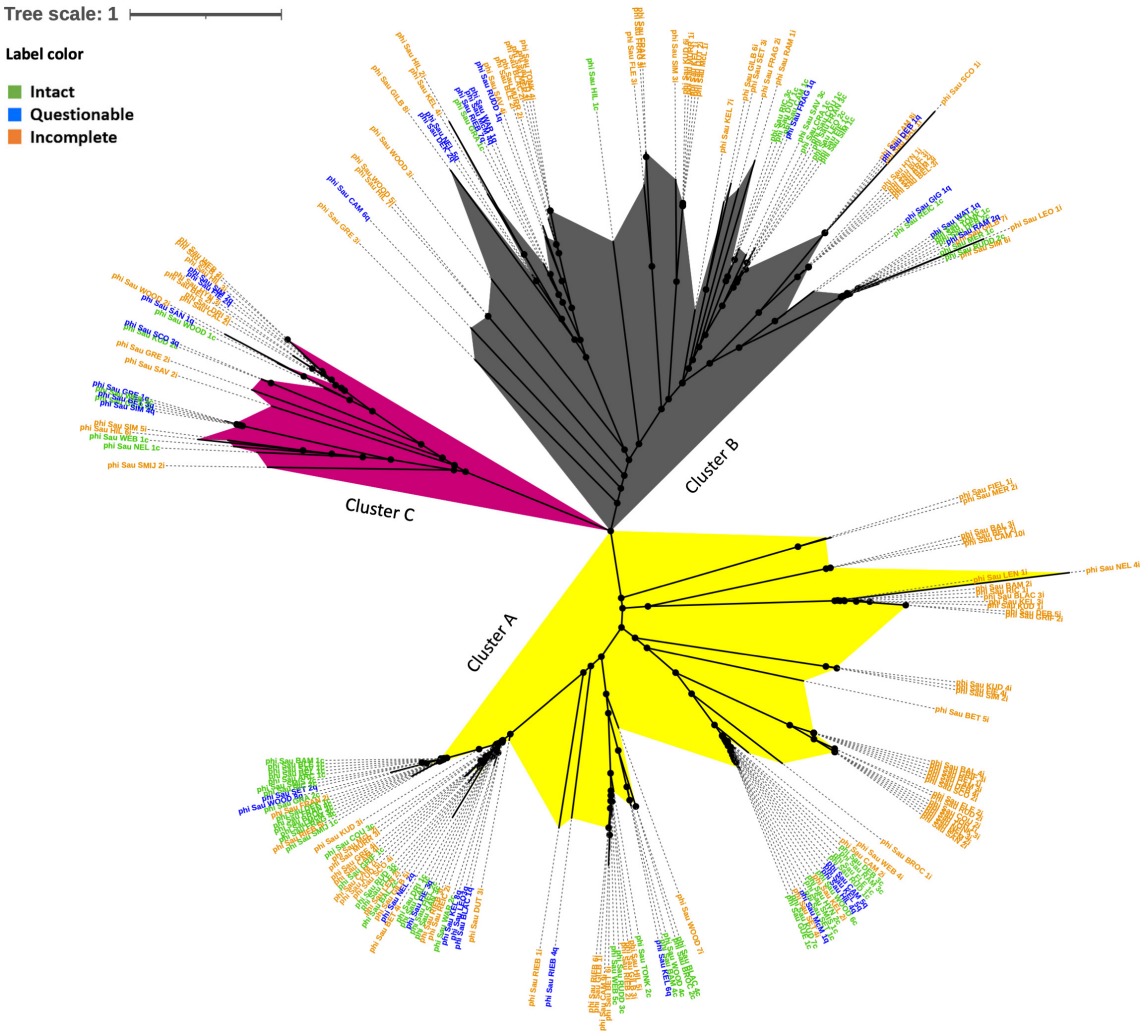

Fig. 6. Phylogenetic tree of 211 prophages from 58 *S. aureus* genomes isolated from patients having chronic rhinosinusitis. Multiple sequence alignment of the prophage sequences was created with MAFFT 7 and maximum likelihood tree was created with FastTree 2.1 through Geneious Prime 2021.1. The tree was further edited using iTol (ver 6). The tree signified that intact (green label), questionable (blue label) and incomplete (red label) prophages are not separate entities but related to each other in mosaic distribution. Please refer to the PDF of the figure and use the zoom function to identify label names of prophages.

S. aureus to colonize and persist in the human nasal niche have been reported [50]. Our results are in line with those observations and indicate that all *S. aureus* clinical isolates from CRS patients and non-control patients carried at least one prophage and prophages could contribute up to 7.7% of accessory genomic data to the core *S. aureus* genome. As different prophages are known to carry different VFs, polylysogeny, that is the presence of various prophages within an individual strain, significantly contributes to pathoadaptive genome variation in clinical strains. Lysogeny furthermore provides a selective advantage to the bacterial strain as the prophage provides immunity against secondary phage attack [51]. This is supported by our findings where no two intact prophages found in the same strain were phylogenetically related or clustered.

Our results on GC content of the whole *S. aureus* genome (32.7%) that is lower than intact prophage (33.5%) is in line with the GC content observed by Kwan, Liu [52] [*S. aureus* (32.9%) and *S. aureus* phage (33.7%)]. This is contrary to the tendency of a higher GC content in core genomes, compared to the corresponding accessory genomes in the majority of pathogens [53]. The retention of such relatively stable but energetically expensive GC nucleotides of intact prophages within the bacterial core genomes suggests that selective pressures are at work and that those intact prophages are likely important components of host adaptation with a potential involvement in the disease process. This is further supported by our finding that the presence of intact prophages (particularly Sa3int group) significantly correlated with the CRSwNP phenotype, suggesting a role of intact prophages and/or associated accessory VFs in CRS disease pathophysiology. Similar correlations may be observed in other diseases associated with persistence of *S. aureus* as these active prophage elements are proven to increase bacterial fitness and mobilize VFs among competing populations. Unlike incomplete prophage regions that are considered non-inducible because they lack genes essential for production of new phage particles, intact prophages may be induced into infectious phage particles. Prophage induction can occur spontaneously or can be promoted in the context of bacterial stress such as antibiotic pressure [54]. This can in turn facilitate horizontal gene transfer (HGT) and support the distribution of prophage-encoded virulence factors within the community promoting host adaptation and colonization of the niche. In this study, most of these intact prophages belonged to Sa3int group phages which encode the immune evasion cluster (IEC) genes (*sak, scn, chp*, *sea/sep*). Furthermore, intact phiNM3 prophages (belonging to the Sa3int group, also carrying IEC) were more abundant in CRS patients that had high severity scores compared to those that had low disease severity scores. *S. aureus* is well known to deploy an arsenal of immune evasive strategies and the IEC genes are well known factors that interfere with host complement and immunoglobulins (*sak* and *scn*) and neutrophil and monocyte chemotaxis (CHIPS) [55]. *Sak* also neutralizes host antimicrobial peptides [56] and promotes *S. aureus* invasion [57]. Interestingly, *S. aureus* invasion within sinonasal mucosa is also seen in the context of CRSwNP [58, 59] and the potential involvement of Sa3int prophages and *sak* in that process requires further investigation. Comparison of prophage abundance and prophage type in CRS with control group also revealed that prophage acquisition in CIs is common and the higher prevalence of Sa3int prophage in CRSwNP compared to CRSsNP could be due to the gain of Sa3int prophage in CRSwNP or the loss of Sa3int prophage in CRSsNP. The gain or loss of specific prophage and associated VFs may impact the persistence of given bacteria, their role in chronic infections and development of nasal polyps. As CRS is known to be associated with dysbiosis with an increased prevalence of *S. aureus*, we speculate that the gain of Sa3int group prophage in CRSwNP may contribute to CRS severity and chronicity as CIs carrying IEC genes are better equipped to persist. Activation and mobilization of those genes would therefore likely assist *S. aureus* in escaping immune surveillance in those patients. Interestingly, the Sa3int prophages also encode enterotoxins that can cross-link the T-cell receptor (TCR) and class-II major histocompatibility complex non-specifically and trigger a massive polyclonal T-cell activation and cytokine release. Through the production of cytokines and chemokines, a type-2 immune response is favoured which is common in the context of CRSwNP [60]. This type-2-biassed immune response promotes the differentiation of immunotolerant M2 macrophages which demonstrate decreased phagocytosis of *S. aureus* and may contribute to its persistence in CRSwNP [61]. Despite strong immune activation, *S. aureus* superantigen driven inflammation can skew adaptive immune responses of the host away from a protective response against *S. aureus* to the benefit of its own survival [62]. Further, as it has been established that Sa3int prophages insert themselves into the beta-haemolysis (*hlb*) gene locus rendering it inactive, we postulate that beta-hemolysin activity is not required for nasal colonization by *S. aureus*. However, more research is required to evaluate the role of prophage-encoded VFs and the relevance of active prophages in *S. aureus* persistence in nasal microenvironment. Also, further studies are required to establish the potential causal relationships between the integrity of prophage in *S. aureus* and the formation or presence of nasal polyps.

In contrast to intact prophages, incomplete prophages had lower GC% (30.92%) than *S. aureus* core genome. It is well known that endosymbionts like prophages are often AT biassed, as AT rich regions are metabolically cheaper to maintain [63]. Such relatively high AT contents can also result from increased levels of genetic drift and mutational bias and it has been shown that increased AT content increases the bacterial fitness of the host [63]. Furthermore, prophage regions showed higher gene density compared to its host *S. aureus* genome (1.43 vs 0.97 genes/kb) [64]. This result is similar to that of temperate *S. aureus* phages (1.67 genes/kb) reported by Kwan, Liu [52] which implies that intact prophage regions have similar gene densities as temperate *S. aureus* phages and they are most likely recently integrated phage regions and are inducible. Gene density in prophage regions is expected to be higher as non-coding DNA segments (introns and intergenic

regions) are continuously under selection pressure to manage the metabolic burden imposed by the addition of genomic material and by the limitations imposed by the amount of DNA able to be packaged into phage heads. Although gene density and prophage size are inversely correlated, phage associated genes (including those necessary for viral replication) were less frequent in incomplete prophages. Loss of phage associated genes like portal (critical roles in head assembly, genome packaging, neck/tail attachment, genome ejection), terminase (catalyse site-specific endo-nucleolytic cleavage of DNA and its packaging into phage proheads), lysin (cleave host's cell wall), proteases (encapsulation of viral DNA into capsid) leads to permanent domestication of a prophage and yet still confers a selective advantage [65–68]. As most of the virulent phages of *S. aureus* belong to *Myoviridae* and almost all temperate phages to *Siphoviridae* family and, to our best knowledge, there are no known *Siphoviridae* phage <20 kb (16–18 kb phages belong to *Podoviridae*) [29, 69], we can infer that prophage regions smaller than 20 kb in *S. aureus* may represent gene remnants of a *S. aureus* temperate phage that still confers evolutionary benefits to the progeny through vertical gene transfer or are remnants that are still in the process of being lost and confer relatively less fitness as they cannot get induced and offer competitive advantage to the host.

In this study, multiple clusters of phage integrases that do not belong to any of the reference (Sa1int-Sa12int) groups within prophage regions were identified. Protein blasting of one of the most prevalent 'unknown' integrase (a 390 aa long) against the NCBI database showed 100% homology with an integrase present in *S. aureus* which was also reported by Bui and Kidd [70] in small colony variants (SCVs) of *S. aureus*. As this integrase type has been reported in unculturable phage and SCVs can underly chronic infections, it may be interesting to see if such association is clinically important and integrase typing can further predict transformation into SCVs. However, supporting experimental evidence is required to associate prophage with the SCV and its association with disease.

Prophage encoded ARGs are sparsely reported in clinically important bacteria like *Acinetobacter*, *Pseudomonas*, *Escherichia* [16–18, 67, 71]. Also ARGs have been occasionally reported in *S. aureus* prophages [47], especially in regions where inappropriate use of antibiotics is highly prevalent. Our results could not identify complete ARGs in any of the prophage regions although ARGs like *tet-38, norA, blaZ, fosB* were highly prevalent in these isolates [32]. This indicates phage-mediated spread of AMR genes may not be a major driver of antimicrobial resistance in the *S. aureus* population in South Australia.

Although Rezaei Javan, Ramos-Sevillano [72] suggests that complete and incomplete (satellite) prophages have separate evolutionary lineage and must be considered a separate entity, our results contradict those findings. Despite incomplete prophages having significantly lower GC%, higher gene density and lower prevalence of phage-hit genes compared to intact (complete) prophages, the heterogenous distribution of intact, questionable and incomplete prophages across major clusters in the phylogenetic tree indicate that incomplete prophages do not belong to separate evolutionary lineages. Rather, they may be truncated remnants of past infection suggesting an AT-biassed endosymbiont-like co-evolution in *S. aureus* prophages that may have important roles in co-evolution of bacteria [51, 73, 74]. This is further supported by the MSA with the reference sequence and the fact that such cryptic entities encode multiple phage-associated structural as well as functional genes. Further, comparison between sub-clusters representing intact and incomplete prophages within a cluster indicate that, evolution of intact prophages into incomplete is possibly non-specific resulting in highly unrelated sub-clusters and singletons. However, this may be because of the different programme used as the authors use their own algorithm to categorize prophage.

CONCLUSION

In summary, our findings expand the knowledge of prophages in *S. aureus* isolated from CRS patients, and their possible role in disease development. Discovery of 22 diverse strains of intact prophages in *S. aureus* within a restricted geographic region and from a well-defined population (CRS disease) reveals circulation of diverse temperate phages contributing to genotypic and phenotypic plasticity as well as virulence. Of further concern is poly-lysogeny which aids in accumulation of additional phage encoded VFs. We also report prophages belonging to Sa3int (phiNM3, JS01, phiN315) and Sa2int (phi2958PVL) group most dominant in *S. aureus* from CRS patients that consistently harboured multiple pathogenic genes such as *sak, scn, chp, sea/sep, luk*E/D. We further speculate that *S. aureus* carrying Sa3int type prophage might impact CRS disease severity and phenotype as they are better equipped to evade the immune system as well as increase the pathogenicity of the strain. However, the potential role of Sa3int prophage in CRS severity and the development of nasal polyps requires further study.

We believe that our findings reveal a novel area for future investigations which will not only increase our understanding of prophage biology, but also uncover undiscovered tripartite associations between prophage-bacteria-human immune system, *S. aureus* evolution and CRS disease epidemiology.

Future directions

Our study was designed to understand the distribution of prophages in *S. aureus*, potential prophage encoded virulence factors and its possible correlation with disease phenotype and severity in a very defined population, CRS. As our results showed significant correlation between the presence of Sa3int group prophages in *S. aureus* and the presence of nasal polyps in CRS disease, it may be important to see if these prophages release any protein(s) that impacts disease development and severity. Also, as phage released from

lysogens are known to directly stimulate/induce/worsen the mammalian immune response, thus impacting inflammation and disease outcomes, it will be important to see if these intact prophages can be induced either spontaneously and/or under stress conditions.

Limitations of the study

We acknowledge that experimental verification of prophage induction is required in addition to *in silico* population genomics to claim that intact prophages are inducible and specific prophage impact CRS disease phenotype, progression, and severity. We also acknowledge that genetic makeup and prior environmental predisposition has a profound impact on the inflammatory response to any external stimulus, and overall CRS pathogenesis and prophage is unlikely to be the sole factor affecting CRS disease pathogenesis. We further note that the sample size is not large enough for robust statistical correlation and similar sized control (non-CRS) group must be included in future research.

Funding information
R.N. was supported by THRF Postgraduate Research Scholarship and The University of Adelaide Scholarship. G.H. was supported by The University of Adelaide International Scholarships and a THRF Postgraduate Top-up Scholarship. The research was funded by AusHealth Research in a grant awarded to S.V. and P.J.W.

Acknowledgements
We sincerely thank Dr Anna Megow for her kind assistance in retrieving CTs for LMK scoring from the database. Further, we'd also like to extend our sincere thanks to all the members of ENT Surgery Group, Basil Hetzel Institute (BHI) for their constructive suggestions during the research. We'd also like to extend our sincere thanks to past/present members of the group who directly and indirectly assisted in collecting clinical isolates from the patients.

Author contributions
R.N. and S.V. conceived the idea and designed the study. G.S. identified and sequenced the control samples. P.J.W., A.J.P. and G.H. assisted in LMK scoring. R.N. and G.B. analysed the data. R.N. and S.V. drafted the manuscript. S.V., K.S., A.J.P. and P.J.W. supervised the study. All authors revised the manuscript and approved for the submitted version.

Conflicts of interest
The authors declare that the research was conducted in the absence of any commercial and/or financial relationships that could be construed as a potential conflict of interest.

Ethical statement
Ethics approval and written informed consent was obtained from each patient prior to the study for the use of *S. aureus* clinical isolates (HREC/18/CALHN/69).

References

1. Fokkens WJ, Lund VJ, Hopkins C, Hellings PW, Kern R, *et al*. European position paper on Rhinosinusitis and Nasal Polyps 2020. *Rhin* 2020;0:1–464.
2. Garneau J, Ramirez M, Armato SG 3rd, Sensakovic WF, Ford MK, *et al*. Computer-assisted staging of chronic rhinosinusitis correlates with symptoms. *Int Forum Allergy Rhinol* 2015;5:637–642.
3. Psaltis AJ, Weitzel EK, Ha KR, Wormald P-J. The effect of bacterial biofilms on post-sinus surgical outcomes. *Am J Rhinol* 2008;22:1–6.
4. Foreman A, Psaltis AJ, Tan LW, Wormald P-J. Characterization of bacterial and fungal biofilms in chronic rhinosinusitis. *Am J Rhinol Allergy* 2009;23:556–561.
5. Nayak N, Satpathy G, Prasad S, Thakar A, Chandra M, *et al*. Clinical implications of microbial biofilms in chronic rhinosinusitis and orbital cellulitis. *BMC Ophthalmol* 2016;16:165.
6. Copeland E, Leonard K, Carney R, Kong J, Forer M, *et al*. Chronic rhinosinusitis: potential role of microbial dysbiosis and recommendations for sampling sites. *Front Cell Infect Microbiol* 2018;8:57.
7. Chambers HF, Deleo FR. Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nat Rev Microbiol* 2009;7:629–641.
8. Tong SYC, Davis JS, Eichenberger E, Holland TL, Fowler VG. *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clin Microbiol Rev* 2015;28:603–661.
9. Van Zele T, Gevaert P, Watelet J-B, Claeys G, Holtappels G, *et al*. *Staphylococcus aureus* colonization and IgE antibody formation to enterotoxins is increased in nasal polyposis. *J Allergy Clin Immunol* 2004;114:981–983.
10. Vickery TW, Ramakrishnan VR, Suh JD. The role of *Staphylococcus aureus* in patients with chronic sinusitis and nasal polyposis. *Curr Allergy Asthma Rep* 2019;19:21.
11. Beceiro A, Tomás M, Bou G. Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? *Clin Microbiol Rev* 2013;26:185–230.
12. Malachowa N, DeLeo FR. Mobile genetic elements of *Staphylococcus aureus*. *Cell Mol Life Sci* 2010;67:3057–3071.
13. Hiramatsu K, Ito T, Tsubakishita S, Sasaki T, Takeuchi F, *et al*. Genomic basis for methicillin resistance in *Staphylococcus aureus*. *Infect Chemother* 2013;45:117–136.
14. Lebeurre J, Dahyot S, Diene S, Paulay A, Aubourg M, *et al*. Comparative genome analysis of *Staphylococcus lugdunensis* shows clonal complex-dependent diversity of the putative virulence factor, ess/type VII locus. *Front Microbiol* 2019;10:2479:19.:
15. Davies EV, Winstanley C, Fothergill JL, James CE. The role of temperate bacteriophages in bacterial infection. *FEMS Microbiol Lett* 2016;363:fnw015.
16. Calero-Cáceres W, Ye M, Balcázar JL. Bacteriophages as environmental reservoirs of antibiotic resistance. *Trends Microbiol* 2019;27:570–577.
17. Balcázar JL. Implications of bacteriophages on the acquisition and spread of antibiotic resistance in the environment. *Int Microbiol* 2020;23:475–479.
18. Loh B, Chen J, Manohar P, Yu Y, Hua X, *et al*. A biological inventory of prophages in *A. baumannii* genomes reveal distinct distributions in classes, length, and genomic positions. *Front Microbiol* 2020;11:579802.
19. Harrison E, Brockhurst MA. Ecological and evolutionary benefits of temperate phage: what does or doesn't kill you makes you stronger. *BioEssays* 2017;39:1700112.
20. Kim B, Little JW. LexA and lambda Cl repressors as enzymes: specific cleavage in an intermolecular reaction. *Cell* 1993;73:1165–1173.
21. Shearwin KE, Brumby AM, Egan JB. The tum protein of coliphage 186 is an antirepressor. *J Biol Chem* 1998;273:5708–5715.
22. Jin M, Liu L, Wang D-N, Yang D, Liu W-L, *et al*. Chlorine disinfection promotes the exchange of antibiotic resistance genes across bacterial genera by natural transformation. *ISME J* 2020;14:1847–1856.
23. Selva L, Viana D, Regev-Yochay G, Trzcinski K, Corpa JM, *et al*. Killing niche competitors by remote-control bacteriophage induction. *Proc Natl Acad Sci U S A* 2009;106:1234–1238.
24. de Sousa JAM, Rocha EPC. Environmental structure drives resistance to phages and antibiotics during phage therapy and to invading lysogens during colonisation. *Sci Rep* 2019;9:3149.
25. Khan A, Wahl LM. Quantifying the forces that maintain prophages in bacterial genomes. *Theoretical Population Biology* 2020;133:168–179.
26. Wahida A, Tang F, Barr JJ. Rethinking phage-bacteria-eukaryotic relationships and their influence on human health. *Cell Host & Microbe* 2021;29:681–688.

27. Li L, Wang G, Li Y, Francois P, Bayer AS, *et al*. Impact of the novel prophage φSA169 on persistent methicillin-resistant *Staphylococcus aureus* endovascular infection. *mSystems* 2020;5.
28. van Wamel WJB, Rooijakkers SHM, Ruyken M, van Kessel KPM, van Strijp JAG. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of *Staphylococcus aureus* are located on beta-hemolysin-converting bacteriophages. *J Bacteriol* 2006;188:1310–1315.
29. Deghorain M, Van Melderen L. The *Staphylococci phages* family: an overview. *Viruses* 2012;4:3316–3335.
30. Shearwin KE, Truong JQ. Lysogeny. In: Bamford DH and Zuckerman M (eds). *Encyclopedia of Virology*, 4th edn. Oxford: Academic Press; 2021. pp. 77–87.
31. Goerke C, Pantucek R, Holtfreter S, Schulte B, Zink M, *et al*. Diversity of prophages in dominant *Staphylococcus aureus* clonal lineages. *J Bacteriol* 2009;191:3462–3468.
32. Bardy JJ, Sarovich DS, Price EP, Steinig E, Tong S, *et al*. *Staphylococcus aureus* from patients with chronic rhinosinusitis show minimal genetic association between polyp and non-polyp phenotypes. *BMC Ear Nose Throat Disord* 2018;18:16.
33. Hopkins C, Browne JP, Slack R, Lund V, Brown P. The Lund-Mackay staging system for chronic rhinosinusitis: how is it used and what does it predict? *Otolaryngol Head Neck Surg* 2007;137:555–561.
34. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, *et al*. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44:W16-21.
35. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, *et al*. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020;75:3491–3500.
36. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res* 2019;47:D687–D692.
37. Guy L, Kultima JR, Andersson SGE. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 2010;26:2334–2335.
38. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
39. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
40. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
41. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 2018;3:124.
42. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, *et al*. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.
43. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, *et al*. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;37:1530–1534.
44. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;30:3059–3066.
45. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS ONE* 2010;5:e9490.
46. Kahánková J, Pantůček R, Goerke C, Růžičková V, Holochová P, *et al*. Multilocus PCR typing strategy for differentiation of Staphylococcus aureus siphoviruses reflecting their modular genome structure. *Environ Microbiol* 2010;12:2527–2538.
47. Ene A, Miller-Ensminger T, Mores CR, Giannattasio-Ferraz S, Wolfe AJ, *et al*. Examination of *Staphylococcus aureus* prophages circulating in Egypt. *Viruses* 2021;13:337.
48. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.
49. Tan D, Hansen MF, de Carvalho LN, Røder HL, Burmølle M, *et al*. High cell densities favor lysogeny: induction of an H20 prophage is repressed by quorum sensing and enhances biofilm formation in *Vibrio anguillarum*. *ISME J* 2020;14:1731–1742.
50. McCarthy AJ, Witney AA, Lindsay JA. *Staphylococcus aureus* temperate bacteriophage: carriage and horizontal gene transfer is lineage associated. *Front Cell Infect Microbiol* 2012;2:6.
51. Ramisetty BCM, Sudhakari PA. Bacterial "Grounded" prophages: hotspots for genetic renovation and innovation. *Front Genet* 2019;10:65.
52. Kwan T, Liu J, DuBow M, Gros P, Pelletier J. The complete genomes and proteomes of 27 Staphylococcus aureus bacteriophages. *Proc Natl Acad Sci U S A* 2005;102:5174–5179.
53. Bohlin J, Eldholm V, Pettersson JHO, Brynildsrud O, Snipen L. The nucleotide composition of microbial genomes indicates differential patterns of selection on core and accessory genomes. *BMC Genomics* 2017;18:151.
54. Nanda AM, Thormann K, Frunzke J. Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J Bacteriol* 2015;197:410–419.
55. Canfield GS, Duerkop BA. Molecular mechanisms of enterococcal-bacteriophage interactions and implications for human health. *Curr Opin Microbiol* 2020;56:38–44.
56. Nguyen LT, Vogel HJ. Staphylokinase has distinct modes of interaction with antimicrobial peptides, modulating its plasminogen-activation properties. *Sci Rep* 2016;6:31817.
57. Bergmann S, Hammerschmidt S. Fibrinolysis and host response in bacterial infections. *Thromb Haemost* 2007;98:512–520.
58. Tan NC-W, Cooksley CM, Roscioli E, Drilling AJ, Douglas R, *et al*. Small-colony variants and phenotype switching of intracellular *Staphylococcus aureus* in chronic rhinosinusitis. *Allergy* 2014;69:1364–1371.
59. Tan NC-W, Foreman A, Jardeleza C, Douglas R, Vreugde S, *et al*. Intracellular *Staphylococcus aureus*: the Trojan horse of recalcitrant chronic rhinosinusitis? *Int Forum Allergy Rhinol* 2013;3:261–266.
60. Wise SK, Lin SY, Toskala E, Orlandi RR, Akdis CA, *et al*. International consensus statement on allergy and rhinology: allergic rhinitis. *Int Forum Allergy Rhinol* 2018;8:108–352.
61. Martinez FO, Helming L, Gordon S. Alternative activation of macrophages: an immunologic functional perspective. *Annu Rev Immunol* 2009;27:451–483.
62. Tuffs SW, Haeryfar SMM, McCormick JK. Manipulation of innate and adaptive immunity by staphylococcal superantigens. *Pathogens* 2018;7:E53.
63. Dietel A-K, Merker H, Kaltenpoth M, Kost C. Selective advantages favour high genomic AT-contents in intracellular elements. 2018.
64. Rogozin IB, Makarova KS, Natale DA, Spiridonov AN, Tatusov RL, *et al*. Congruent evolution of different classes of non-coding DNA in prokaryotic genomes. *Nucleic Acids Res* 2002;30:4264–4271.
65. Laumay F, Corvaglia A-R, Diene SM, Girard M, Oechslin F, *et al*. Temperate prophages increase bacterial adhesin expression and virulence in an experimental model of endocarditis due to *Staphylococcus aureus* from the CC398 lineage. *Front Microbiol* 2019;10:742.
66. Dragoš A, Andersen AJC, Lozano-Andrade CN, Kempen PJ, Kovács ÁT, *et al*. Phages weaponize their bacteria with biosynthetic gene clusters. 2020.
67. Kondo K, Kawano M, Sugai M. Prophage elements function as reservoir for antibiotic resistance and virulence genes in nosocomial pathogens. *bioRXiv* 2020.
68. Popescu M, Van Belleghem JD, Khosravi A, Bollyky PL. Bacteriophages and the immune system. *Annu Rev Virol* 2021;8:415–435.
69. Ingmer H, Gerlach D, Wolz C. Temperate phages of *Staphylococcus aureus*. *Microbiol Spectr* 2019;7.
70. Bui LMG, Kidd SP. A full genomic characterization of the development of a stable Small Colony Variant cell-type by a clinical *Staphylococcus aureus* strain. *Infect Genet Evol* 2015;36:345–355.

71. Marti E, Variatza E, Balcázar JL. Bacteriophages as a reservoir of extended-spectrum β -lactamase and fluoroquinolone resistance genes in the environment. *Clin Microbiol Infect* 2014;20:456-O459.
72. Rezaei Javan R, Ramos-Sevillano E, Akter A, Brown J, Brueggemann AB. Prophages and satellite prophages are widespread in Streptococcus and may play a role in pneumococcal pathogenesis. *Nat Commun* 2019;10:4852.
73. Bobay L-M, Touchon M, Rocha EPC. Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci U S A* 2014;111:12127–12132.
74. Wang G-H, Niu L-M, Ma G-C, Xiao J-H, Huang D-W. Large proportion of genes in one cryptic WO prophage genome are actively and sex-specifically transcribed in a fig wasp species. *BMC Genomics* 2014;15:893.