D660–D665 *Nucleic Acids Research, 2019, Vol. 47, Database issue* *Published online 8 November 2018*
doi: 10.1093/nar/gky1123

ICEberg 2.0: an updated database of bacterial integrative and conjugative elements

Meng Liu[1,2], Xiaobin Li[1], Yingzhou Xie[1], Dexi Bi[3], Jingyong Sun[4], Jun Li[5], Cui Tai[1], Zixin Deng[1] and Hong-Yu Ou[1,2,*]

[1]State Key Laboratory of Microbial Metabolism, Joint International Laboratory on Metabolic & Developmental Sciences, School of Life Sciences & Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, China, [2]Shanghai Key Laboratory of Veterinary Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China, [3]Department of Pathology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai 200072, China, [4]Department of Clinical Microbiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China and [5]College of Biotechnology and Bioengineering, Zhejiang University of Technology, Hangzhou 310014, China

Received September 15, 2018; Revised October 16, 2018; Editorial Decision October 22, 2018; Accepted October 23, 2018

ABSTRACT

ICEberg 2.0 (http://db-mml.sjtu.edu.cn/ICEberg/) is an updated database that provides comprehensive information about bacterial integrative and conjugative elements (ICEs). Compared with the previous version, three major improvements were made. First, with the aid of text mining and manual curation, it now recorded the details of 1032 ICEs, including 270 with experimental supports and 762 from bioinformatics prediction. Second, as increasing evidence has shown that ICEs frequently mobilize the so-called 'hitchhikers', such as integrative and mobilizable elements (IMEs) and *cis*-mobilizable elements (CIMEs), 83 known transfer interactions between 49 IMEs and 7 CIMEs with 19 ICEs taken from the literature were included and illustrated with visually intuitive directed graphs. An expanded collection of 260 chromosome-borne IMEs and 235 CIMEs was also added. At last, ICEberg 2.0 provides an online tool ICEfinder to predict ICEs or IMEs in bacterial genome sequences. It combines a similarity search for the integrase, relaxase and/or type IV secretion system and the co-localization of these corresponding homologous genes. With the recent updates, ICEberg 2.0 might provide better support for understanding the biological traits of ICEs, especially as their interaction with cognate mobilizable elements may further promote horizontal gene flow.

INTRODUCTION

Integrative and conjugative elements (ICEs) are important members of the bacterial mobile genetic elements and are integrative to the bacterial chromosome, encoding fully functioning conjugation machinery and, thus, are self-transmissible between bacterial cells (1). The cargo genes of ICEs, such as those encoding virulence factors and acquired antibiotic resistances, can confer the hosts with selective advantages, making ICEs a vital driving force for bacterial adaptation and evolution (2,3). ICEs typically contain the recombination and conjugation modules, and their features have been recently elucidated. The integrases (Int) responsible for ICE recombination are usually confined to the tyrosine, serine or DDE family. ICEs have two conjugal manners. In all Gram-negative and most of the Gram-positive bacteria, ICEs are delivered as linear single-stranded DNA (ssDNA) involving a relaxase, a conjugative type IV secretion system (T4SS) (2). However, in Actinobacteria, some ICEs are transferred as double-stranded DNA (dsDNA) involving a RepSA- or RepAM-type replication initiator protein (Rep) and an FtsK/SpoIIIE domain-containing protein (Tra) for translocation (4). Thus, ICEs have been categorized as T4SS-type ICEs and actinomycete ICEs (AICEs).

ICEs have been identified in increasing numbers from the exponentially expanding pool of bacterial genome sequences (5–11). For example, 200 T4SS-type ICEs were identified from 2484 complete bacterial chromosomes based on comparative genomics, in which the conjugation modules were scanned by CONJScan (11). In addition, 144 FtsK/SpoIIIE-type ICEs (AICE) were predicted in 275 Actinobacteria genomes (6). Recently, the online tool VRprofile detected ICEs based on the identification of recombination modules (Int) and their adjacent conjuga-

*To whom correspondence should be addressed. Tel: +86 216 293 2943; Fax: +86 216 293 2418; Email: hyou@sjtu.edu.cn

tion modules (12), but it could not distinguish the T4SS-type ICEs and AICEs. In 2012, we reported the web-based open-access database ICEberg 1.0 (13), which archived both experimentally validated and computationally predicted ICEs. Since then, a vast number of new ICEs and new ICE types, such as Tripartite ICEs (14), have been identified experimentally. Thus, the demand for a database update and a tool to help predict new ICEs became urgent.

In addition to self-transfer, ICEs have also been reported to be capable of mobilizing other genetic elements *in trans* or *in cis*, such as the chromosome-borne integrative and mobilizable elements (IMEs), *cis*-mobilizable elements (CIMEs) and plasmids. IMEs are genomic islands that encode their own excision and integration but lack an intact conjugative apparatus for autonomously conjugative transfer, whereas CIMEs have lost genes for both integration and transfer but retain intact *attL* and *attR* recombination sites (15). These elements are devoid of conjugal apparatus and can function as hitchhikers frequently picked up by ICEs and other conjugative elements (15,16). For example, in *Streptococcus thermophilus*, CIMEL_3catR_3could be mobilized by ICE*St3 in cis* via the integration of ICE*St3* and subsequently excise and transfer as a whole composite element (17). Moreover, the ICEs, IMEs and CIMEs have shown profound diversity in *Streptococcus* (5,8,10). Indeed, IMEs and CIMEs, as the important vehicles for the dissemination of virulence factor genes and acquired antibiotic resistance genes, are thought to be more widespread than ICEs (7,9,16). To date, a well-organized IME- and CIME-specific database has not been reported. The complex interactions between ICEs and IMEs or CIMEs await exploration (3,15,16).

Here, we report the release of ICEberg version 2.0, which reflects a large expansion of the dataset of curated ICEs and their crosstalk with the IMEs. A newly developed online prediction tool for ICEs and IMEs was also integrated. We expect that ICEberg will provide a better support for researchers interested in bacterial horizontal gene transfer.

MATERIALS AND METHODS

ICE data update by text mining and manual curations

After manual curation of the search results, 284 PubMed-archived papers published since 2012 were collected and added into ICEberg 2.0, resulting in a total of 694 papers in the database. In ICEberg 2.0, 604 ICEs (of which 84 derived from experimental data) have been newly added, for a total collection of 1032 ICEs (of which 270 were experimentally validated) (Supplementary Table S1). Notably, the number of the collected AICEs has been expanded from 27 to 181. The existing data accuracy and reliability, including some mistakes in ICEberg 1.0 highlighted by peers (8), have also been manually revisited and cured as best as possible. Similar to the previous version, ICEberg 2.0 organized the ICE data using the PostgreSQL relational database, the PHP data pipeline and HTML web interfaces. However, the ICE graphical display was achieved with Perl scripts to highlight the re-annotated genes (cluster) encoding both the recombination module and the conjugation module. The modules were shown with different colors to facilitate users to compare the ICE genetic structures. In addition, the putative *oriT* (origin of transfer) regions in the ICEs with the entire nucleotide sequences were detected by oriTfinder (18) and listed in ICEberg 2.0. The length and the GC content of the ICEs with the entire nucleotide sequences were also calculated (Supplementary Figures S1 and S2).

ICE hitchhikers, IMEs and CIMEs

ICEs or conjugative plasmids can mobilize IMEs *in trans.* Furthermore, IMEs and/or CIMEs can integrate with ICEs to form 'tandem accretion' composite elements, which then can be mobilized with ICE *in cis.* Using a visually intuitive directed graph, ICEberg 2.0 illustrated the 83 interactions of 19 ICEs and 18 conjugative plasmids with their hitchhiker, 49 IMEs and 7 CIMEs. For example, the Supplementary Figure S3 shows the main known interactions among conjugative elements and mobilizable elements in the five genera. In these directed networks, each node represents one type of mobilizable genetic element (ICEs, IMEs, CIMEs or plasmids); each edge, with direction from A to B node, indicates that element A could mobilize element B *in trans* or *in cis*. Below the graph, a table lists the detailed information about their transfer interactions, including the specific donors and recipients and the corresponding experimental literature (see the example in Figure 1). The detailed interaction information for each IME and CIME is accessible via the 'Browse' page of ICEberg 2.0. In addition, ICEberg 2.0 now contains the information of 260 IMEs from 160 bacterial strains and 235 CIMEs from 165 bacterial strains, including 32 IMEs from 31 bacterial strains and 5 CIMEs from 5 bacterial strains with their experimental supports.

Integration of the ICE and IME detection tool

We have developed a tool, called ICEfinder, available online and as a standalone version for the rapid detection of ICEs and IMEs in bacterial genome sequences. ICEfinder employs a method we called 'Pattern-based hit co-localization' (see the Supplementary Methods) that detects the signature sequences of the recombination modules and conjugation modules based on their profile HMMs (19) (Supplementary Table S2, S3 and Figure S4). It also searches for the *oriT* region using the approach proposed by oriTfinder (18). It then co-localizes, filters and groups the corresponding genes. At last, those elements carrying an integrase gene, a relaxase gene and T4SS gene clusters (12,20) are considered as T4SS-type ICEs, while those without T4SS but with integrase, replication and the AICE translocation-related proteins are thought to be putative AICEs. Those without T4SS but with integrase and relaxase are tagged as putative IMEs. ICEfinder also tries to detect some particular IMEs with integrase and an *oriT* but no relaxase. ICEfinder employs ARAGORN (21) with the default parameters to identify the 3′ termini of the tRNA/tmRNA genes as the putative ICE insertion sites. It also uses Vmatch (http://vmatch.de/) with the default options to detect the directed repeats as the tRNA-distal boundaries. The acquired antibiotic resistance genes and virulence factors are also identified by NCBI BLASTp (22) with the cut-off of *Ha*-value of 0.64 (12).

The ICEfinder online tool allows users to submit a GenBank file containing a nucleotide sequence and its annotation as a query. A FASTA format file of a raw nucleotide

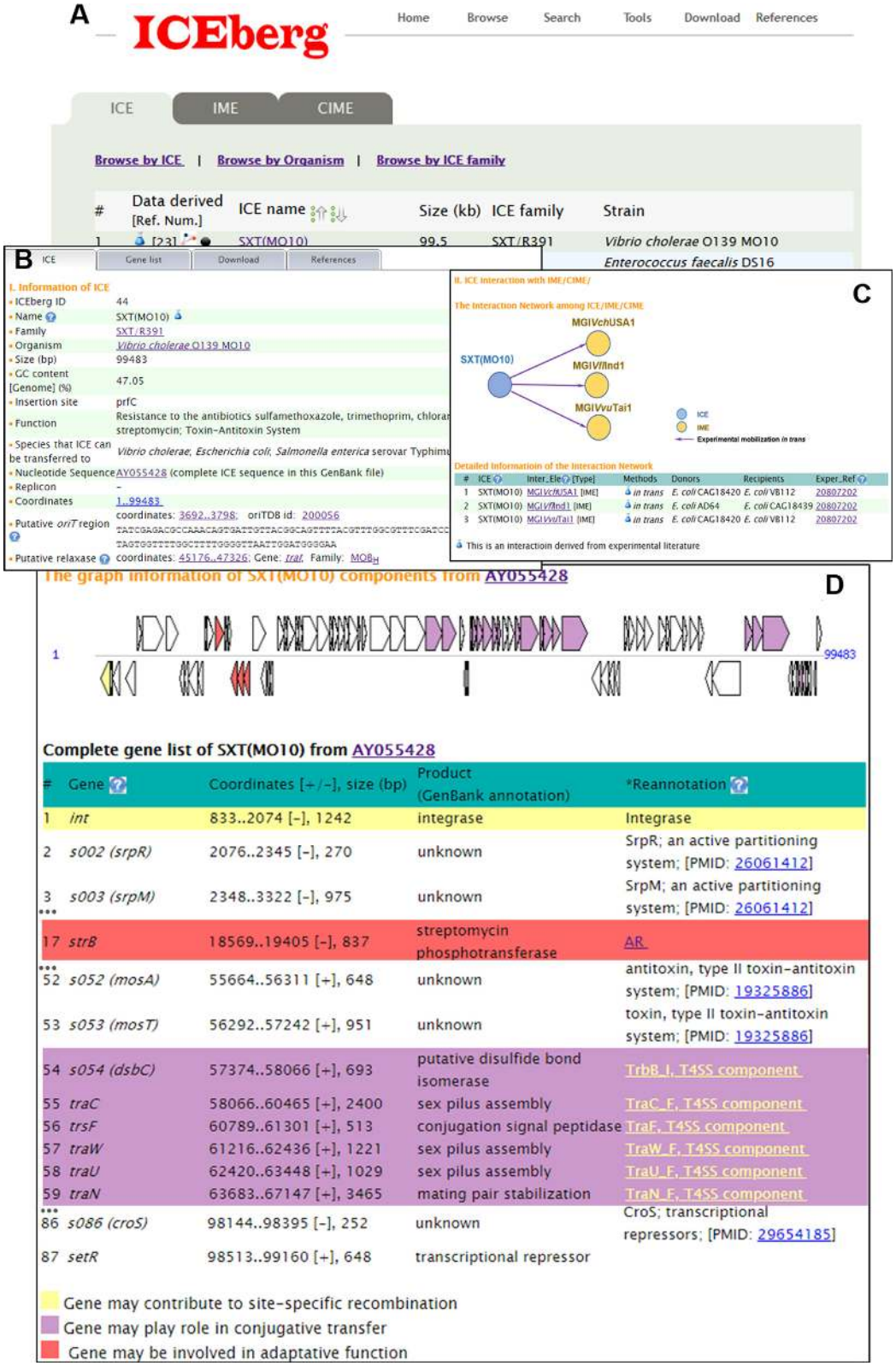

Figure 1. The updated web interface of ICEberg 2.0 using the SXT(MO10) in *Vibrio cholerae* O139 MO10 as an example. (**A**) The browse modules of ICEberg on the home page, consisting of three sections (ICEs, IMEs and CIMEs). (**B**) An overview of the feature of SXT(MO10) with the newly added putative *oriT* region (*oriT* coordinates, the corresponding link of most related *oriT* in oriTDB and the detailed sequences) and relaxase gene (coordinates, locus tag and MOB family). (**C**) Detailed information on the mobile interactions between ICEs and IMEs or CIMEs in both graph and table format. (**D**) A graphical display of ICE genetic modules was constructed using a local Perl script. AR: acquired antibiotic resistance genes; T4SS, type IV secretion system.

sequence is also accepted, which is annotated using our gene annotation tool CDSeasy (12) and is then used as the input for the following ICE detection. ICEfinder uses the CGView circular genome visualization tool (23) to display the distribution of the predicted T4SS-type ICEs, IMEs and AICEs in the query bacterial genome. In addition, the ICEfinder has a comparison module (Supplementary Figure S5) that allows performing the alignment between the identified ICE loci against the ICEberg-archived ICEs using MultiGeneBlast (24).

RESULTS AND DISCUSSION

Recent developments present in this study have further improved the data quality of ICEberg, such as the ICE curation based on experimental evidence. Compared with ICEberg 1.0, the updated 2.0 version offers three major improvements: (i) new T4SS-type ICE and AICE data via manual curation; (ii) transfer interactions with IMEs and CIMEs and a supplement of IME and CIME data; and (iii) a prediction tool for ICEs and IMEs. ICEberg 2.0 currently recorded the details of 1032 ICEs, including 270 ICEs with experimental supports (Supplementary Table S1). It also collected 260 IMEs and 235 CIMEs.

ICEberg browse module

ICEberg provides a flexible and biologist-friendly web interface. The browse module contains detailed information on all archived ICEs that are tagged with thumbnail icons corresponding to the availability of experimental support, full nucleotide sequences, and information on the transfer interactions with IMEs. Notably, the 83 transfer interactions between 49 IMEs and 7 CIMEs with 19 ICEs and 18 conjugative plasmids were also illustrated using intuitive directed graphs to visualize the complex and diverse interactions of IMEs with ICEs (Supplementary Figure S3), including *trans*- or *cis*-mobilization after a tandem accretion. As an example, the ICE SXT in *Vibrio* is shown in Figure 1. The SXT has been reported to mobilize three IMEs: MGI*Vvu*Tai1, MGI*Vch*USA1 and MGI*Vfl*Ind1 (25) (Figure 1). These IMEs only contain *oriT* sites similar to the *oriT* of SXT (63% identities for the 282 bp matching regions) (25) for the conjugation and lack relaxase and coupling protein genes. In addition, CTnDOT, the most widely studied ICE in *Bacteroides*, can *in trans* mobilize co-existing *ermF* region (26), NBU1 (27), NBU2 (27) or Tn*4399* (27) from *Bacteroides* donors to *Bacteroides* recipients. The four IMEs all encode MOB_P relaxases and cognate *oriT* regions. The first three IMEs also encode tyrosine recombinases. Similarly, another *Bacteroides* ICE, CTnERL, which is highly identical to CTnDOT but lacks the *ermF* portion, can also mobilize NBU1 (27,28), NBU2 (27,28) and the *ermF* region (26) *in trans*. However, the available experimental data of the mobilizing interactions between ICEs and IMEs (Supplementary Table S4) are still inadequate for the extensive prediction of the putative mobilized elements of ICEs or the mobilizing elements of IMEs inversely.

ICE reference dataset

The downloadable dataset of ICEberg has been popularly used, such as the analysis of the interaction between bacterial mobile genetic elements and the CRISPR (clustered regularly interspaced short palindromic repeats) immune system (29). The new download module of ICEberg 2.0 now provides diverse types of reference data of ICEs for offline analysis. ICEs datasets are more clearly categorized according to the availability of experimental evidence and/or full sequences. Here, we show an example of using the downloadable ICE data to analyze the targets of CRISPR spacers. Like phages or conjugative plasmids, the transferred ICEs in recipients might be recognized and cleaved by CRISPR systems. In turn, some ICEs might evolve mechanisms to prevent the inhibition. Here, we search for the targets of CRISPR spacers in the ICEberg-collected ICE sequences. Using CRISPRTarget (30) with the cut-off score of 20 by default and the 125 497 CRISPRdb-collected non-redundant spacers (31), a total number of 831 targets of CRISPR spacers are found in 340 ICEs, in which 272 spacers are distributed on 104 experimentally validated ICEs (Supplementary Table S5). Notably, 34 ICEs with detectable CRISPR targets also code for the putative anti-CRISPR (Acr) proteins (BLASTp *E*-value < 0.001; Supplementary Table S5) that evade that CRISPR immunity (32), such as the PAGI-5 of *Pseudomonas aeruginosa* PSE9 (33), ICE*Ml*$Sym^{NZP2037}$ of *Mesorhizobium loti* NZP2037 and ICE*Ml*Sym^{R7A} of *M. loti* R7A. These findings indicate that these ICEs might process the complex interactions with the CRISPR immunity system.

ICE and IME prediction tool

To facilitate the rapid identification of ICEs and their potentially associated IMEs in bacterial chromosome sequences, we developed a user-friendly tool called ICEfinder that is publicly available both online and as a standalone version. It detects both T4SS-type ICEs and AICE. The performance of ICEfinder has been evaluated by three frequently used metrics, recall, precision and F1 (F-measure) (Supplementary Methods and Table S6). As an example of the application of ICEfinder, we scanned the 9434 bacterial genomes downloaded from the NCBI RefSeq database in April 2018. ICEfinder detected 1109 T4SS-type ICEs in 928 bacterial chromosomes and 375 AICEs in 198 Actinobacteria genomes (Supplementary Table S7). In addition, 1777 IMEs were identified in 1470 bacterial genomes (Supplementary Table S7). For example, 216 putative ICE regions with the size from 37 to 143 kb were discovered from the 279 completely sequence chromosomes of *Klebsiella pneumoniae*, which is an increasingly important human pathogen. Notably, 90 ICE*Kp1*-like regions carried the biosynthetic gene clusters for yersiniabactin, 14 of which harbored an additional colibactin biosynthetic gene cluster (Supplementary Figure S6). Only few putative IMEs of *K. pneumoniae* have been reported before, such as GIE492 (34), which was 22.3 kb and was predicted to be mobilizable by ICE*Kp1*; however, in this study, 92 putative IMEs with sizes ranging from 7 to 40 kb were identified in 86 *K. pneumoniae* chromosomes using ICEfinder. These findings indicate that the

prevalent and diverse ICEs and IMEs might play an important role in the genome evolution of *K. pneumoniae.* However, it should be noted that for various reasons, ICEfinder may not provide the precise boundaries of ICEs and IMEs. For example, some ICEs and IMEs are also reported to integrate into the 3′ or 5′ ends of protein-coding genes besides the most common tRNA gene sites. Cury *et al.* used comparative genomics to delimit ICEs at the gene level (11), but the accurate delimitation of ICEs and IMEs still has a long way to go.

CONCLUSION

Here, we reported a major update of ICEberg and focused on ICEs and their interactions with the genetic elements of other types. ICEberg 2.0 collected and integrated the systematic information of ICEs and their interactions with mobilizable elements from peer-reviewed publications. A tool, called ICEfinder, available in both web server and standalone versions for the prediction of ICEs and IMEs, was also provided, which could facilitate the investigation of the widely distributed conjugative and mobilizable elements from the rapidly growing bacterial genomic data. Newly available information about ICEs and associated IMEs will be updated and improved regularly to keep up with the rapidly expanding microbial genome database. Ultimately, we propose an updated ICE-specific resource to facilitate efficient exploration of large numbers of these elements and an improved understanding of their biological traits.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Key R&D Program of China [2017YFC1600100 to H.Y.O.]; National Natural Science Foundation of China [31670074 to H.Y.O., 21661140002 to Z.D.]; Medicine and Engineering Interdisciplinary Research Fund of Shanghai Jiao Tong University [YG2015MS59 to J.S.]. Funding for open access charge: National Key R&D Program of China [2017YFC1600105].
Conflict of interest statement. None declared.

REFERENCES

1. Burrus,V., Pavlovic,G., Decaris,B. and Guedon,G. (2002) Conjugative transposons: the tip of the iceberg. *Mol. Microbiol.*, **46**, 601–610.
2. Johnson,C.M. and Grossman,A.D. (2015) Integrative and Conjugative Elements (ICEs): what they do and how they work. *Annu. Rev. Genet.*, **49**, 577–601.
3. Delavat,F., Miyazaki,R., Carraro,N., Pradervand,N. and van der Meer,J.R. (2017) The hidden life of integrative and conjugative elements. *FEMS Microbiol. Rev.*, **41**, 512–537.
4. Thoma,L. and Muth,G. (2015) The conjugative DNA-transfer apparatus of *Streptomyces*. *Int. J. Med. Microbiol.*, **305**, 224–229.
5. Brochet,M., Couve,E., Glaser,P., Guedon,G. and Payot,S. (2008) Integrative conjugative elements and related elements are major contributors to the genome diversity of *Streptococcus agalactiae*. *J. Bacteriol.*, **190**, 6913–6917.
6. Ghinet,M.G., Bordeleau,E., Beaudin,J., Brzezinski,R., Roy,S. and Burrus,V. (2011) Uncovering the prevalence and diversity of integrating conjugative elements in Actinobacteria. *PLoS One*, **6**, e27846.
7. Guglielmini,J., Quintais,L., Garcillan-Barcia,M.P., de la Cruz,F. and Rocha,E.P. (2011) The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLos Genet.*, **7**, e1002222.
8. Ambroset,C., Coluzzi,C., Guedon,G., Devignes,M.D., Loux,V., Lacroix,T., Payot,S. and Leblond-Bourget,N. (2015) New insights into the classification and integration specificity of *Streptococcus* integrative conjugative elements through extensive genome exploration. *Front. Microbiol.*, **6**, 1483.
9. Puymege,A., Bertin,S., Guedon,G. and Payot,S. (2015) Analysis of *Streptococcus agalactiae* pan-genome for prevalence, diversity and functionality of integrative and conjugative or mobilizable elements integrated in the tRNA(Lys CTT) gene. *Mol. Genet. Genomics*, **290**, 1727–1740.
10. Coluzzi,C., Guedon,G., Devignes,M.D., Ambroset,C., Loux,V., Lacroix,T., Payot,S. and Leblond-Bourget,N. (2017) A glimpse into the world of integrative and mobilizable elements in *Streptococci* reveals an unexpected diversity and novel families of mobilization proteins. *Front. Microbiol.*, **8**, 443.
11. Cury,J., Touchon,M. and Rocha,E.P.C. (2017) Integrative and conjugative elements and their hosts: composition, distribution and organization. *Nucleic Acids Res.*, **45**, 8943–8956.
12. Li,J., Tai,C., Deng,Z., Zhong,W., He,Y. and Ou,H.Y. (2018) VRprofile: gene-cluster-detection-based profiling of virulence and antibiotic resistance traits encoded within genome sequences of pathogenic bacteria. *Brief. Bioinform.*, **19**, 566–574.
13. Bi,D., Xu,Z., Harrison,E.M., Tai,C., Wei,Y., He,X., Jia,S., Deng,Z., Rajakumar,K. and Ou,H.Y. (2012) ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. *Nucleic Acids Res.*, **40**, D621–D626.
14. Haskett,T.L., Terpolilli,J.J., Bekuma,A., O'Hara,G.W., Sullivan,J.T., Wang,P., Ronson,C.W. and Ramsay,J.P. (2016) Assembly and transfer of tripartite integrative and conjugative genetic elements. *PNAS*, **113**, 12268–12273.
15. Bellanger,X., Payot,S., Leblond-Bourget,N. and Guedon,G. (2014) Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol. Rev.*, **38**, 720–760.
16. Guedon,G., Libante,V., Coluzzi,C., Payot,S. and Leblond-Bourget,N. (2017) The obscure world of integrative and mobilizable elements, highly widespread elements that pirate bacterial conjugative systems. *Genes*, **8**, 337.
17. Bellanger,X., Morel,C., Gonot,F., Puymege,A., Decaris,B. and Guedon,G. (2011) Site-specific accretion of an integrative conjugative element together with a related genomic island leads to *cis* mobilization and gene capture. *Mol. Microbiol.*, **81**, 912–925.
18. Li,X., Xie,Y., Liu,M., Tai,C., Sun,J., Deng,Z. and Ou,H.-Y. (2018) oriTfinder: a web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. *Nucleic Acids Res.*, **46**, W229–W234.
19. Finn,R.D., Clements,J., Arndt,W., Miller,B.L., Wheeler,T.J., Schreiber,F., Bateman,A. and Eddy,S.R. (2015) HMMER web server: 2015 update. *Nucleic Acids Res.*, **43**, W30–W38.
20. Bi,D., Liu,L., Tai,C., Deng,Z., Rajakumar,K. and Ou,H.Y. (2013) SecReT4: a web-based bacterial type IV secretion system resource. *Nucleic Acids Res.*, **41**, D660–D665.
21. Laslett,D. and Canback,B. (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.*, **32**, 11–16.
22. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
23. Grant,J.R. and Stothard,P. (2008) The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res.*, **36**, W181–W184.
24. Medema,M.H., Takano,E. and Breitling,R. (2013) Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol. Biol. Evol.*, **30**, 1218–1223.
25. Daccord,A., Ceccarelli,D. and Burrus,V. (2010) Integrating conjugative elements of the SXT/R391 family trigger the excision and drive the mobilization of a new class of *Vibrio* genomic islands. *Mol. Microbiol.*, **78**, 576–588.
26. Whittle,G., Hund,B.D., Shoemaker,N.B. and Salyers,A.A. (2001) Characterization of the 13-kilobase ermF region of the *Bacteroides*

conjugative transposon CTnDOT. *Appl. Environ. Microbiol.*, **67**, 3488–3495.
27. Shoemaker,N.B., Wang,G.R., Stevens,A.M. and Salyers,A.A. (1993) Excision, transfer, and integration of NBU1, a mobilizable site-selective insertion element. *J. Bacteriol.*, **175**, 6578–6587.
28. Li,L.Y., Shoemaker,N.B., Wang,G.R., Cole,S.P., Hashimoto,M.K., Wang,J. and Salyers,A.A. (1995) The mobilization regions of two integrated Bacteroides elements, NBU1 and NBU2, have only a single mobilization protein and may be on a cassette. *J. Bacteriol.*, **177**, 3940–3945.
29. Zhang,Q., Rho,M., Tang,H., Doak,T.G. and Ye,Y. (2013) CRISPR-Cas systems target a diverse collection of invasive mobile genetic elements in human microbiomes. *Genome Biol.*, **14**, R40.
30. Biswas,A., Gagnon,J.N., Brouns,S.J., Fineran,P.C. and Brown,C.M. (2013) CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA Biol.*, **10**, 817–827.
31. Grissa,I., Vergnaud,G. and Pourcel,C. (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*, **8**, 172.
32. Dong,C., Hao,G.F., Hua,H.L., Liu,S., Labena,A.A., Chai,G., Huang,J., Rao,N. and Guo,F.B. (2018) Anti-CRISPRdb: a comprehensive online resource for anti-CRISPR proteins. *Nucleic Acids Res.*, **46**, D393–D398.
33. Bondy-Denomy,J., Pawluk,A., Maxwell,K.L. and Davidson,A.R. (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, **493**, 429–432.
34. Marcoleta,A.E., Berrios-Pasten,C., Nunez,G., Monasterio,O. and Lagos,R. (2016) *Klebsiella pneumoniae* asparagine tDNAs are integration hotspots for different genomic islands encoding microcin E492 production determinants and other putative virulence factors present in hypervirulent strains. *Front. Microbiol.*, **7**, 849.