In Silico Identification of Three Types of Integrative and Conjugative Elements (ICEs) in *Elizabethkingia anophelis* Strains Isolated from Around the World

Jiannong Xu[a#] Dong Pei[a*], Ainsley Nicholson[b], Yuhao Lan[a], Qing Xia[a]

[a]Biology Department, New Mexico State University, Las Cruces, NM, USA
[b]Special Bacteriology Reference Laboratory, Bacterial Special Pathogens Branch, Division of High-Consequence Pathogens and Pathology, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Running Head: Three types of ICE in various strains of *E. anophelis*

#Address correspondence to Jiannong Xu, jxu@nmsu.edu.

*Present address: Dong Pei, Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, USA

ABSTRACT

Elizabethkingia anophelis is an emerging global multidrug-resistant opportunistic pathogen. We assessed the diversity among 13 complete genomes and 23 draft genomes of *E. anophelis* derived from various environmental settings and human infections from different geographic regions around the world over past decades from 1950s. Thirty-one of these 36 strains harbor integrative and conjugative elements (ICEs). A total of 52 ICEs were identified, and categorized into three ICE types based on the architecture of signature genes in the conjugation module. The type II and III ICEs were found to integrate into regions adjacent to tRNA genes, while type I ICEs used a variety of integration sites, inserting into intergenic regions or even directly into a gene, sometimes disrupting gene function. Integrases such as tyrosine recombinases, serine recombinases and DDE transposases were found in most ICEs. The ICEs carry various cargo genes including transcription regulators and those involved in antibiotic resistance. The CRISPR-Cas system was found in nine strains, including four strains in which CRISPR-Cas machinery and ICEs co-exist. ICE distribution in the strains showed no geographic or temporal patterns. The ICEs in *E. anophelis* differ in gene structure and sequence from CTnDOT, a well-studied ICE prevalent in *Bacteroides* spp. This is the first set of ICEs identified in the family Flavobacteriaceae. As a prevalent type of mobile genetic elements in various strains of *E. anophelis* around the world, the categorization of ICEs will facilitate further investigations such as virulence, genome epidemiology and adaptation genomics of *E. anophelis.*

Importance

Elizabethkingia anophelis is an opportunistic human pathogen, and the genetic diversity between strains from around the world becomes apparent as more genomes are sequenced. The Integrative Conjugative Element (ICE), found in many bacterial species, contains genes

for transfer via conjugation and integration into the chromosome, along with various cargo genes. ICEs are identified in 31 of 36 strains and categorized into three types based on architecture of modular genes, integrases, and integration sites. ICE distribution in different strains displays no spatial and temporal patterns. Several ICE-containing strains also possessed CRISPR-Cas units, considered to be the bacterial adaptive immune system providing protection against phage and predatory mobile genetic elements. This co-existence suggests that ICEs are beneficial or at least not harmful to the bacterial cells they inhabit. ICEs as a component of the mobile genetic repertoire enable recipients to resist antibiotics, survive disinfecting agents, and adapt to various ecological niches.

Introduction

The genus *Elizabethkingia* belongs to family *Flavobacteriaceae*, and was recognized as distinct from the genus *Chryseobacterium* in 2005 (1). The two species initially recognized were *E. meningoseptica*, named based on its initial isolation as the causative agent for neonatal meningitis (2), and *E. miricola*, an isolate obtained from condensation water of Space Station Mir (3). A third species, *E. anophelis*, was proposed in 2011(4) based on the description of the type strain R26^T that was originally isolated from the midgut of *Anopheles gambiae* mosquitoes maintained in Stockholm University in Sweden (5). In 2015, a new species, *E. endophytica* was proposed (6) but whole genome sequence (WGS) based genome comparison revealed it to be a homotypic synonym of *E. anophelis* (7, 8). Additional species have since been added to the genus (8).

In 2011, the first human infection attributed to *E. anophelis* was documented, a neonatal meningitis that occurred in Central Africa Republic (9). Later, an *E. anophelis* outbreak in an intensive-care unit in Singapore was reported (10), followed by the worrying account of *E. anophelis* transmission from a mother to her infant (11). In 2016, a large *E. anophelis* outbreak occurring in Wisconsin, USA, was unusual in that a substantial proportion of patients were not already hospitalized, and were instead admitted directly from their homes. No indication of human-to-human transmission was found (12-14). Human cases of *E. anophelis* have been reported with increasing frequency around the world (13, 15-17), aided in part by improved identification methods, and several strains previously described as *E. meningoseptica* were determined to actually belong to the *E. anophelis* species (18). In addition to human-derived strains, multiple strains of the species have been isolated from its namesake, the mosquito, and their genomes have been sequenced. Both the type strain R26^T and strain Ag1 were derived from *An. gambiae* mosquitoes (19, 20). Strain EaAs1 was isolated from *An. stephensi* mosquitoes (21), and strains AR4-6 and AR6-8 were obtained

from wild caught specimens of the mosquito *An. sinensis* in China (20). There has been no epidemiological link between human cases and mosquitoes, and the circumstances of certain outbreaks preclude any possibility of mosquitos being a vector for *E. anophelis.*

Genome comparison of pathogenic bacteria has greatly increased our understanding of the evolution, pathogenesis and epidemiology in many pathogen outbreak investigations (22-26). Horizontal transfer of mobile genetic elements between bacterial strains occurs continually, and is a major source of genetic variation in bacteria (27). One group of modular mobile genetic elements, known as integrative and conjugative elements (ICEs), are capable of transferring between bacteria horizontally via conjugation (28-30). ICEs integrate into a host chromosome and replicate along with the genome (28, 31). The cargo genes that are brought in by the ICEs may endow the recipient bacteria with new phenotypes (32). For example, in *Pseudomonas syringae* pv. *actinidiae*, resistance to copper, arsenic and cadmium was attributed to the resistant genes that were carried in ICEs (33). In *Helicobacter pylori*, ICE *tfs4* has been identified in certain strains mediating gastric disease. The presence of *tfs4* ICE has been associated with virulence, although mechanisms behind the virulence phenotype remain unknown (34).

Genome analysis of various *E. anophelis* strains has revealed substantial genetic diversity (11, 16, 35-39). The ICE named ICE*Ea*1 has been identified in the strains associated with the outbreak in Wisconsin, and unrelated strains from the outbreak in Singapore (40). In this study, we searched for ICEs in 13 complete and 23 draft genomes of *E. anophelis* strains around the world by locating clusters of transfer (*tra*) genes. Based on the architecture of conjugation modules and associated signature genes, three types of ICEs were recognized. As *E. anophelis* is recognized as an environmental bacterium, the high prevalence and diversity of ICEs as well as the mosaic integration pattern in different strains suggest that ICEs play a

significant role in shaping its genome to adapt to different environmental niches. In recent years, whole-genome sequencing has been utilized in epidemiological studies to type outbreak strains, acquire genomic determinants of virulence and antibiotic resistance in outbreaks, and estimate pan-genome diversity (41-44). The identification and classification of ICEs is expected to facilitate the genomic epidemiological studies of the pathogenic strains of *E. anophelis* in the future.

Results

Origin and geographic distribution of the strains

In this study, 13 complete genomes and 23 draft genomes of *E. anophelis* were compared. Mosquito-derived strains included R26^T and Ag1 from *An. gambiae* (19, 20), EaAs1 from *An. stephensi* (21) and AR4-6 and AR6-8 from wild caught *An. sinensis* (20). Strains that were associated with human infections and recent outbreaks were from the Central Africa Republic (9, 39), Singapore (35), Hong Kong (11), China (45), Wisconsin, USA (13, 14), and Taiwan (16). Geographic distribution of included isolates was spread across Asia, Europe, Africa and North America (Figure 1). The available metadata for the strains is noted in Table 1, along with the WGS accession numbers.

Core genome based phylogeny of the strains

The pan-genome of a bacterial species is comprised of a core and an accessory genome. The core genome is a set of genes that are conserved and shared by all strains while the accessory gene set is shared by only some strains (46, 47). Because the core genome contains essential genes and their inheritance is necessitated, they carry more reliable evolutionary information for inferring phylogeny than do accessory genes. Hence, single nucleotide polymorphism (SNP) typing based on core genome comparison has become a

recognized method for accurate evolutionary reconstructions (48, 49). We used the core genome alignment to generate a SNP tree to infer phylogeny of the strains, which was implemented by Parsnp in the Harvest suit (49). Parsnp recognized 37781 anchors and 985 maximal unique matches (MUMs) in the 36 genomes, a total of 37764 anchors and MUMs were used to anchor the genome alignment and make a SNP tree (49). The tree topology appeared similar when each of the complete genomes R26^{T}, 0422, CSID3015183678 and NUHP1 was used as reference for core genome alignment (data not shown). The strains were sorted into three clusters (Figure 1, Table 1). Cluster I contained 14 strains with the genome of 12012-2_PRCM at the basal position. Cluster II consisted of two clades, each included six strains. The strains in this cluster were globally distributed and collected over decades between the 1950s and the present. Cluster III had two clades as well. One was comprised of all five mosquito-derived strains, the other contained strain 37-75 obtained from a human infection in France in 1979 and strain CSID3015183678, which was a representative of the strains associated with the Wisconsin outbreak in 2016 (40). There was no indication of geography-based clustering. In each cluster, there were strains that were very closely related, which were marked as B, C, D and E in the tree (Figure 1). The clustering of these strains was corroborated by the higher percent identity in the protein sequences between the genomes. The overall identity was greater than 99.9% in the clades B, C, D and E (Figure 1). In the clade A, the strain isolated from a carp in France was closely related to the two isolates that were derived from human patients in Central Africa in 2006 and 2011, respectively (9, 39). In the clade C, 3375 was isolated from South Carolina, US, in 1957, and F3543 was isolated from Florida, US, in 1982, while FDAARGOS_198 was obtained in Sweden. In the clade D, 0422 was collected in Florida, US, in 1950, and PW2806 and PW2809 were isolated in Hong Kong in 2012. The clade E grouped the strains from three different mosquito species, *An. gambiae, An. stephensi* and *An. sinensis,* and these strains were collected in different

geographic locations in separate times. The clade A presented a group of strains with certain distance, in which the identity of protein sequences was between 95-100% in a pairwise comparison. Overall, it appears that closely related strains can be widely separated both temporally and spatially.

Identification of ICEs in the strains

Structurally, ICEs consist of the modules for genome integration and excision, conjugative transfer, and maintenance (31). We define a genome neighborhood as an ICE if it harbors genes encoding a relaxase, a coupling protein VirD4 ATPase (T4CP) and several *Tra* proteins. Among the 36 strains, 31 contained at least one ICE, and a total of 52 ICEs were detected (Table 2). According to the architecture of the modular genes, these ICEs were classified into three types (Figure 2).

Type I ICEs are featured by 13 *tra* genes (*traABBDEGIJKMNOQ*) in addition to genes coding for T4CP, relaxsase and integrase. In most cases, genes *t4cp* and *relaxase* were located tandemly at upstream of the *tra* gene cassette, and *integrase* was located at downstream of the *tra* cassette (Figure 2). A total of 12 ICE*Ea*I elements were found in 11 strains, with strain FDAARGOS-198 harboring two ICE*Ea*Is (Table 2). The integration site varied. Each of the 12 elements had a distinct insertion site (Figure 3). The type I ICE in CSID3015183678 has been characterized previously, named as ICE*Ea*1 (40). This element inserted into and disrupted the gene encoding MutY, which is an adenine DNA glycosylase that is required for fixing G-A mis-pairs, making the strain more prone to mutation (40). The type I ICE in strain 37-75 inserted into the gene encoding an alkyl hydroperoxide reductase (Ahp) between codons K207 and I208. This protein is a primary scavenger of H_2O_2 in *E.coli* (50). Its disruption by the ICE may result in a loss of function and make the strain vulnerable to oxidative stress. The integration sites of the remaining nine type I ICEs were in the

intergenic regions. The locations of these elements were listed in Table 2 and marked in Figure 3. Interestingly, The ICE*Ea*I(1) was shared by four strains, CSID3015183678 (Wisconsin, US, 2015) , NUPH1 (Singapore, 2012), 3375 (South Carolina, US, 1957) and FDAARGOS-198 (Sweden, collection time unknown). The sequences of the element were nearly identical, but inwertion sites were different, indicating that the integrations occurred independently.

Type II ICEs were identified in 13 strains. Unlike in type I, genes *t4cp* and *relaxase* were separated by one or more CDS, followed by 13 *tra* genes, *traABDEFGIJKMNOQ*. A gene is located between *traD* and *traE*, encoding a RadC-domain-containing protein (Figure 2). All 13 ICE*Ea*IIs integrated next to a tRNA gene, 12 elements reside at the 3' end of the tRNA-Leu-CAA. Only the ICE*Ea*II in NUHP1 was inserted after the tRNA-Ser-GGA (Table 2, Figure 3). Integrases were found in 4 of the 13 elements (see below).

Type III ICEs were recognized in 16 strains. The structure was quite different from that in type I and II ICEs, with only seven *tra* genes present, *traAEGJKMN*. The *relaxase* and *t4cp* flank the *tra* genes. In 11 out of 16 elements, a gene is present after the *traN*, encoding a large protein (791-1177 aa) with a RadC domain (Figure 2). Like ICE*Ea*II, the type III ICEs after a tRNA gene. The tRNA-Arg-ACG, tRNA-Gln-TTG, tRNA-Asp-TGA, tRNA-Ser-TGA, and tRNA-Glu-TTC were targeted by ICE*Ea*IIIs (Figure 3). In NUHP1, two type III elements, ICE*Ea*III(7) and ICE*Ea*II(8), are co-located between two tRNA-Asp-GTC, and the two elements are separated by a tRNA-Asp-GTC. In NUH11, ICE*Ea*I(8) and ICE*Ea*III(12) combined as one segment, co-localized between the tRNA-Asp-GTC and the gene encoding a beta-lactamase. In most type III elements, at least one integrase gene is present.

In sum, a total of 52 ICEs were identified in 31 genomes, and 15 elements in 11 draft genomes were partially assembled. The size of ICEs is up to 116.3 kb in complete genomes.

Annotated ICEs were listed in Table S1, which included predicted CDS and gene functions. Some strains had more than one ICE. No ICEs were found in five strains: PW2810, B2D, 8707, CIP11067 and GTC_10754 (Table 2).

Phylogenetic relationship of the ICEs

To track the evolutionary history of these ICEs, the nucleotide sequences of the genes encoding relaxase, T4CP, TraG and TraJ from each ICE were compared. As shown in Figure 4, in all four gene trees, the ICEs were separated based on their types. The type I and II elements are more closely related, forming a clade distinct from the type III clade. This pattern was consistent with the type classification based on the structure of the conjugation module (Figure 2). The distribution of the ICE types did not imply geographic or temporal patterns.

Integrases

Integrases, required for ICE integration and excision, could be tyrosine or serine recombinases, or DDE transposases containing an acidic amino acid triad (DDE)(32). Tyrosine recombinases were found in all type I and III ICEs, and DDE transposases were identified in four type III ICEs. In the 12 type II ICEs, only four elements carried integrases. Two serine recombinases were detected in NCTC10588 and one was identified in JM-87. A DDE transposase was found in FMS-007, and a tyrosine recombinase was detected in NUHP1. The remaining 9 type II ICEs lack co-localized integrases (Table S1).

Cargo genes

Restriction and modification (R-M) systems provide an innate defense against invading DNA to protect the genome stability. Mobile genetic elements (including ICEs) usually carry an R-M system, which enables an evolutionary interplay between the mobile

genetic elements (MGEs) and their hosts (51). The DNA modification methylase, type I R-M system, and type II restriction enzymes were prevalent in the ICEs of all three types. In addition, the anti-restriction protein ArdA was detected in three type II ICEs and one type III ICE. Some ICEs also carried DNA topoisomerase, DNA helicase, DNA primase, DNA polymerase. These enzymes may be involved in the integration process, and perhaps in DNA replication while the ICE is in its plasmid state. Genes encoding proteins in the resistance-nodulation-division (RND) family (52), a multidrug efflux pump system found in Gram-negative bacteria, are prevalent in the ICEs. The genes coding for three key component proteins of the tripartite RND complex (the outer membrane protein, membrane fusion protein and inner membrane protein), are carried in the type I ICEs in six strains, one of the three type III ICEs in NUHP1 and the type II ICE in F3543 (Table S1). ABC transporters for various substrates, such as manganese, potassium and oligopeptides were found in all ICEs. TonB-dependent receptors for siderophore import or carbohydrates uptake (SusC) are present in some ICEs. There are various transcriptional regulators in the ICEs, such as AraC family, ArsR family, MarR family, HxlR family, and TetR family. The AraC family transcriptional regulators (AFTRs) are most prevalent, a total of 36 these were found in 19 strains. In addition, two-component regulatory system are present in the type I ICEs in three strains.

CRISPR-Cas loci

The CRISPR-Cas system is a prokaryotic adaptive defense machinery against invading nucleic acids (53, 54). The type II-C CRISPR-Cas system was detected in nine strains (Table 1), which is featured by a CRISPR array and the genes encoding Cas9, Cas1 and Cas2 proteins. In these nine strains, the CRISPR-Cas system is located downstream of the gene encoding a cobalt-zinc-cadmium resistance protein CzcD. In a CRISPR array, the direct repeat is 47 nt, and the spacer is 30 nt in length. The number of spacers varies between

6 in strain CIP111046 to 47 in strain CIP111067. No CRISPR locus was detected in the draft assembly of strain 8707, but one complete direct repeat was identified in the assembly. Likely the CRISPR locus was not assembled well in the draft genome. The CRISPR loci of strains Po0527107 and V0378064 have been described previously (39). Po0527107 has 21 spacers and V0378064 has the same 21 plus two more spacers. Both strains were isolated in Bangui, the Central Africa Republic, but five years apart (9, 55). Strain LDVH-AR107 has two CRISPR loci, each with 21 spacers. LDVH-AR107 shares 12 spacers with both Po0527107 and V0378064 despite being isolated on a different continent, from the internal organ of a common carp *Cyprinus carpio* collected in 2004 in Montpellier, France (Table 1).

Discussion

Many genomes have been sequenced for *E. anophelis* strains derived from different sources including mosquitoes, human infections, hospital environments, fish and corn stems. This genome availability enabled a comparative genomics approach to investigate the genetic architecture and repertoire of the *E. anophelis* population.

Core genome based phylogeny

By definition, the core genome is shared by all strains and consists of essential genes that are vertically transmitted while genes in the accessory genome are present only in a subset of strains. The sum of the core and accessory genomes in all strains constitutes a pan-genome for the species (47, 56, 57). Due to its inheritability, the core genome is intrinsically suitable for inferring phylogeny (49). The core genome phylogenetic analysis separated the strains into three clusters, however, genetically similar strains do not show a spatial or temporal aggregation pattern. As an example, clade C (Figure 1) contains closely related strains F3543, 3375 and FDAARGOS-198, the locations where they were isolated were Florida in 1982, South Carolina in 1957 and Sweden (collection time unknown), respectively

(Table1, Figure 1). On the other hand, the mosquito-associated strains were very closely related, and distinct from the strains derived from human infections (Figure 1), despite having been isolated from three different mosquito species, and from both lab reared (19, 21) and wild caught mosquitoes (20). The possibility that they might be lab contaminants was carefully evaluated, and rejected based on the timeline of sample collection, processing and sequencing. Furthermore, the presence of *E. anophelis* in the gut microbiota of wild populations of *An. gambiae* in Kenya and *An. sinensis* mosquitos in Shandong, China, has been confirmed from the shotgun metagenomic sequencing data (Xu, unpublished data). Association of *E. anophelis* has been documented in several different mosquito species, including *An. gambiae* (58), *Culex quinquefasicatus* (59) and *Aedes aegypti* (60), and a role for the bacteria in larval development has been hypothesized. Taken together, these results indicate that there is selection in the mosquito gut environment for these specific strains.

ICEs in the strains

Genome diversification promotes bacterial adaptation and evolution. ICE, as a type of mobile genetic elements, contributes significantly to the pan-genome reservoir with potentially adaptive genes (28, 31). Several ICEs have been found in the taxa of *Bacteroides* (29, 61-64), which belong to the family Bacteroidaceae in phylum Bacteroidetes. Of these elements, CTnDOT in *B. thetaiotaomicron* has been studied extensively (65-68). The ICEs identified in *E. anophelis* are structurally distinct from CTnDOT.

The conjugation machinery is a necessary component of ICEs. Relaxase, coupling protein and *tra* proteins are required for element excision from the genome and transfer to a recipient (69). Conserved features indicative of conjugative elements have been used for identifying ICEs in a wide variety of genomes (29). Using this approach, we categorized the ICEs in *E. anophelis* strains into three types based on the architecture of the genes in the

conjugation module and the phylogeny of sequences from four genes in the module (Figure 2, 4). Type II and III ICEs tended to integrate adjacent to a tRNA gene, while the type I ICEs used a non-tRNA gene region for integration. In some strains, the type I ICE insertion into a gene resulted in the loss of function. For example, in CSID3015183678 the integration of ICE disrupted the gene *mutY*, which is required for excision of G-A mismatch (70, 71). The disruption of *mutY* by ICE*Ea*I resulted in a higher rate of nonsynonymous to synonymous substitution in the strain and subsequent sublineages, which may fuel the adaptive capacity of an outbreak strain (40). In strain 37-75, the insertion of a type I ICE inactivates the gene encoding alkyl hydroperoxide reductase (Ahp). As the Ahp protein is a scavenger of hydrogen peroxide in *E. coli* (50), the disruption of the *ahp* gene by ICE may affect host defense against elevated oxidative stress. Twelve type II ICEs used tRNA-Leu-CAA as the integration locus, however, these elements lacked an integrase. This warrants further investigation to elucidate whether or not the type II ICEs are mobile without an integrase in the element. The presence of diverse integration loci, including tRNA gene and non-tRNA loci (Figure 3), suggests a high degree of genome plasticity of the *E. anophelis* in favor of a dynamic genetic repertoire of the pan-genome accommodating mobile elements like ICEs (28, 63).

Among the cargo genes carried by the ICEs for which a function could be identified, most are involved with host defense, nutrient acquisition, or transcription regulation. Innate immune mechanisms like R-M systems provide host defense by protecting against invading DNA (51, 72), and the RND type multidrug efflux pumps enable the chemical defense by recognizing and expelling various structurally diverse compounds and toxins (73), including antibiotics, as shown previously in R26^{T} (37), and the strains associated with the outbreak in Singapore (35, 36). Regarding nutrient acquisition, a variety of ABC transporters and TonB dependent receptors are present in the ICEs, which greatly enhances the host

capability to satisfy the nutrient needs in the environment they thrive. The ICEs also equip the host with a variety of transcriptional regulators, with those in the AraC family being particularly prevalent (Table S1). These transcription regulators can sense various chemical signals involved in carbon metabolism, quorum-sensing signaling, virulence and stress response, which leads to the expression of a global gene repertoire (74-77). Such genetic capability brought in by ICEs enables the host strains a broader adaptive flexibility to eco-niche changes, which may contribute to establishment of an infection in humans as well.

CRISPR loci

CRISPR-Cas system is an adaptive immune mechanism against invading nucleic acids. Spacer sequences are acquired and integrated into the CRISPR locus. Upon encountering the same non-self invaders, the spacers will be transcribed into guide RNAs to direct Cas nuclease to cleave specific target sequences (53, 78). In this study, the class II CRISPR-Cas system was identified in nine strains of *E. anophelis.* A total of 209 spacers were recognized in eight strains. These spacers reflected previous events of the horizontal transfer of mobile elements or phage infections these strains have experienced in the past. Indeed, it has been shown by Breurec *et al.* (39) that strain V0378064 (isolate E18064 in (39)) obtained in 2011 contained two newly acquired spacers, S22 and S23, which are a perfect match to two protospacers from the strain Po0527107 (isolate E27107) isolated in 2006. In fact, the two protospacers reside in a phage-derived mobile genetic element integrated at the 3' end of tRNA-Arg-ACG in Po0527107. The phage element is absent from V0378064, likely due to the action of the CRISPR-Cas mediated defense.

Of nine strains that carry a CRISPR-Cas system, five have ICEs as well. The co-existence of mobile elements and CRISPR defense system suggests an evolutionary balance between the genome stability and plasticity. In the family *Flavobacteriaceae*, the CRISPR-

Cas systems in *Flavobacterium columnare* were shown to play a role in host-phage interactions, which drives long-term genome coevolution through arms-race-like mechanisms between the host and phages (79), serving as an example of genome stability. On the other hand, the CRISPR-Cas machinery appeared not to play a role as a resistance mechanism against phages in fish pathogenic *Flavobacterium psychrophilum* (80). Availability of the *E. anophelis* strains that possess CRISPR-Cas with and without ICEs enables further studies on co-evolution of mobile elements and host defense systems in the bacteria.

Conclusion

In this study, we identified ICEs in the genomes of 36 *E. anophelis* strains isolated from a wide array of hosts, collected at diverse geographic locations over several decades. The ICEs can be categorized into three types based on the distinctive architecture of conjugation module and integration sites. The identification of the ICEs enables further studies on the genetic diversity of the pan-genome and its impact on the virulence of this global opportunistic pathogen.

Materials and Methods

Bacterial strains

In this study, complete genomes of 13 strains and draft genomes of 23 strains were analyzed (Table 1). Mosquito-derived strains were isolated from *An. gambiae* (R26^T, Ag1), *An. stephensi* (As1) and *An. sinensis* (AR4-6, AR6-8). Strain CSID3015183678 was one of the strains associated with the Wisconsin outbreak 2015-2016, which has been characterized in (40), and strain NUHP1 was a representative of the strains isolated in the Singapore outbreak, the genomes of these strains have been described (35). Strain LDVH-AR107 was derived from a common carp *Cyprinus carpio*. Strain JM-87 was isolated from maize *Zea mays*,

389 which was originally described as *E. endophytica*. The genome comparison put it as a
390 synonym of *E. anophelis* (7). The other strains were all derived from human patients or
391 hospital environments, collected from 1950 to 2016.

392 The draft genomes of strains LDVH-AR107, 8707, NCTC10588, AmMS250 and 37-75 were
393 reassembled using Illumina reads downloaded from SRA database. The reads were *de novo*
394 assembled by CLC genomics workbench v 10.1.1. The genomes were annotated using the
395 SEED and Rapid Annotations using Subsystems Technology (RAST) at the RAST server
396 (81).

397 **Phylogenetic relationship of the strains based on core genome comparison**

398 The phylogenomic relationships of the strains were estimated by the Harvest suite (49). The
399 core genome that is shared by all strains was identified by multiple genome alignment, and
400 SNPs in the core genome were typed to infer phylogenies between the strains implemented
401 by Parsnp module in the Harvest suite (49). Four complete genomes R26^T, 0422,
402 CSID3015183678 and NUHP1, each was used iteratively as the reference for tree
403 construction, and tree topology of all four core genome trees was unaffected by choice of
404 reference genome. The tree with 0422 as the reference was depicted in Figure 1.

405 **Identification of ICEs**

406 Pairwise genome comparison based on protein identity was used to identify variable regions.
407 To identify an ICE, the genome was searched for a cluster of genes coding for a relaxase, a
408 coupling ATPase (T4CP) and transfer (Tra) proteins including a VirB4 ATPase (TraG) in the
409 conjugation module. These proteins are the key components of an ICE (28, 69). The
410 boundary of an element was delimited as between the two open reading frames (ORFs) that
411 flank the ICE. The genome of the type strain R26^T was used as a reference to mark the

integration sites. Each ICE was categorized into one of the three Types based on its gene structure (Figure 2). The ICEs were named based mainly on the nomenclature proposed by Burrus *et al.* (82): the acronym ICE was followed by the initials of the name of the bacterium (*Ea*), a Roman numeral as type, a strain name and a ordinal number in brackets to identify the same sequence if it is encountered in a different strain. For example, the type I ICE found in both NUHP1 and CSID3015183678 would be named ICE*Ea*I(1)_CSID3015183678 and ICE*Ea*I(1)_NUHP1, respectively. Each type has its set of numbers; for example, the four type III ICEs in NUHP1 are designated as ICE*Ea*III(1)_NUHP1 thru ICE*Ea*III(4)_NUHP1.

GenBank accession numbers

Accession numbers of the genomes used in this study were listed in Table 1. GenBank accession numbers for ICE sequences were listed in Table 2.

Supplemental Material

Table S1. Annotated ICEs in the strains. (excel file)

Author contributions

JX conceived and designed study. JX, DP, YL,QX collected genome data and performed data analysis. JX and AN wrote the manuscript.

Acknowledgements

This work was supported by the National Institutes of Health [SC1AI112786 to J.X.] and the National Science Foundation [No. 1633330 to J.X.] and CDC program funds designated for the study of emerging infectious agents. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of

Health and the National Science Foundation, and the Centers for Disease Control and Prevention. Mention of company names or products does not constitute endorsement.

Funding for open access charge: National Institutes of Health [SC1AI112786].

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Kim KK, Kim MK, Lim JH, Park HY, Lee ST. 2005. Transfer of Chryseobacterium meningosepticum and Chryseobacterium miricola to Elizabethkingia gen. nov. as Elizabethkingia meningoseptica comb. nov. and Elizabethkingia miricola comb. nov. Int J Syst Evol Microbiol 55:1287-93.
2. King EO. 1959. Studies on a group of previously unclassified bacteria associated with meningitis in infants. Am J Clin Pathol 31:241-7.
3. Li Y, Kawamura Y, Fujiwara N, Naka T, Liu H, Huang X, Kobayashi K, Ezaki T. 2003. Chryseobacterium miricola sp. nov., a novel species isolated from condensation water of space station Mir. Syst Appl Microbiol 26:523-8.
4. Kampfer P, Matthews H, Glaeser SP, Martin K, Lodders N, Faye I. 2011. Elizabethkingia anophelis sp. nov., isolated from the midgut of the mosquito Anopheles gambiae. Int J Syst Evol Microbiol 61:2670-5.
5. Lindh JM, Borg-Karlson AK, Faye I. 2008. Transstadial and horizontal transfer of bacteria within a colony of Anopheles gambiae (Diptera: Culicidae) and oviposition response to bacteria-containing water. Acta Trop 107:242-50.

460 6. Kampfer P, Busse HJ, McInroy JA, Glaeser SP. 2015. Elizabethkingia endophytica sp.
461 nov., isolated from Zea mays and emended description of Elizabethkingia
462 anophelisKampfer et al. 2011. Int J Syst Evol Microbiol 65:2187-93.
463 7. Doijad S, Ghosh H, Glaeser S, Kampfer P, Chakraborty T. 2016. Taxonomic
464 reassessment of the genus Elizabethkingia using whole-genome sequencing:
465 Elizabethkingia endophytica Kampfer et al. 2015 is a later subjective synonym of
466 Elizabethkingia anophelis Kampfer et al. 2011. Int J Syst Evol Microbiol 66:4555-
467 4559.
468 8. Nicholson AC, Gulvik CA, Whitney AM, Humrighouse BW, Graziano J, Emery B,
469 Bell M, Loparev V, Juieng P, Gartin J, Bizet C, Clermont D, Criscuolo A, Brisse S,
470 McQuiston JR. 2017. Revisiting the taxonomy of the genus Elizabethkingia using
471 whole-genome sequencing, optical mapping, and MALDI-TOF, along with proposal
472 of three novel Elizabethkingia species: Elizabethkingia bruuniana sp. nov.,
473 Elizabethkingia ursingii sp. nov., and Elizabethkingia occulta sp. nov. Antonie Van
474 Leeuwenhoek doi:10.1007/s10482-017-0926-3.
475 9. Frank T, Gody JC, Nguyen LB, Berthet N, Le Fleche-Mateos A, Bata P, Rafai C,
476 Kazanji M, Breurec S. 2013. First case of Elizabethkingia anophelis meningitis in the
477 Central African Republic. Lancet 381:1876.
478 10. Teo J, Tan SY, Tay M, Ding Y, Kjelleberg S, Givskov M, Lin RT, Yang L. 2013.
479 First case of E anophelis outbreak in an intensive-care unit. Lancet 382:855-6.
480 11. Lau SK, Wu AK, Teng JL, Tse H, Curreem SO, Tsui SK, Huang Y, Chen JH, Lee RA,
481 Yuen KY, Woo PC. 2015. Evidence for Elizabethkingia anophelis transmission from
482 mother to infant, Hong Kong. Emerg Infect Dis 21:232-41.
483 12. Figueroa Castro CE, Johnson C, Williams M, VanDerSlik A, Graham MB, Letzer D,
484 Ledeboer N, Buchan BW, Block T, Borlaug G, Munoz-Price LS. 2017.

485 Elizabethkingia anophelis: Clinical Experience of an Academic Health System in
486 Southeastern Wisconsin. Open Forum Infect Dis 4:ofx251.
487 13. Navon L, Clegg WJ, Morgan J, Austin C, McQuiston JR, Blaney DD, Walters MS,
488 Moulton-Meissner H, Nicholson A. 2016. Notes from the Field: Investigation of
489 Elizabethkingia anophelis Cluster - Illinois, 2014-2016. MMWR Morb Mortal Wkly
490 Rep 65:1380-1381.
491 14. Nicholson AC, Whitney AM, Emery BD, Bell ME, Gartin JT, Humrighouse BW,
492 Loparev VN, Batra D, Sheth M, Rowe LA, Juieng P, Knipe K, Gulvik C, McQuiston
493 JR. 2016. Complete Genome Sequences of Four Strains from the 2015-2016
494 Elizabethkingia anophelis Outbreak. Genome Announc 4.
495 15. Hu S, Jiang T, Zhang X, Zhou Y, Yi Z, Wang Y, Zhao S, Wang M, Ming D, Chen S.
496 2017. Elizabethkingia anophelis Isolated from Patients with Multiple Organ
497 Dysfunction Syndrome and Lower Respiratory Tract Infection: Report of Two Cases
498 and Literature Review. Front Microbiol 8:382.
499 16. Lin JN, Lai CH, Yang CH, Huang YH, Lin HH. 2017. Genomic features,
500 phylogenetic relationships, and comparative genomics of Elizabethkingia anophelis
501 strain EM361-97 isolated in Taiwan. Sci Rep 7:14317.
502 17. Lau SK, Chow WN, Foo CH, Curreem SO, Lo GC, Teng JL, Chen JH, Ng RH, Wu
503 AK, Cheung IY, Chau SK, Lung DC, Lee RA, Tse CW, Fung KS, Que TL, Woo PC.
504 2016. Elizabethkingia anophelis bacteremia is associated with clinically significant
505 infections and high mortality. Sci Rep 6:26045.
506 18. Janda JM, Lopez DL. 2017. Mini review: New pathogen profiles: Elizabethkingia
507 anophelis. Diagn Microbiol Infect Dis 88:201-205.

19. Kukutla P, Lindberg BG, Pei D, Rayl M, Yu W, Steritz M, Faye I, Xu J. 2013. Draft Genome Sequences of Elizabethkingia anophelis Strains R26T and Ag1 from the Midgut of the Malaria Mosquito Anopheles gambiae. Genome Announc 1.

20. Pei D, Nicholson AC, Jiang J, Chen H, Whitney AM, Villarma A, Bell M, Humrighouse B, Rowe LA, Sheth M, Batra D, Juieng P, Loparev VN, McQuiston JR, Lan Y, Ma Y, Xu J. 2017. Complete Circularized Genome Sequences of Four Strains of Elizabethkingia anophelis, Including Two Novel Strains Isolated from Wild-Caught Anopheles sinensis. Genome Announc 5.

21. Raygoza Garay JA, Hughes GL, Koundal V, Rasgon JL, Mwangi MM. 2016. Genome Sequence of Elizabethkingia anophelis Strain EaAs1, Isolated from the Asian Malaria Mosquito Anopheles stephensi. Genome Announc 4.

22. Georgiades K. 2012. Genomics of epidemic pathogens. Clin Microbiol Infect 18:213-7.

23. Raskin DM, Seshadri R, Pukatzki SU, Mekalanos JJ. 2006. Bacterial genomics and pathogen evolution. Cell 124:703-14.

24. Worby CJ, Lipsitch M, Hanage WP. 2017. Shared Genomic Variants: Identification of Transmission Routes Using Pathogen Deep-Sequence Data. Am J Epidemiol 186:1209-1216.

25. Hu R, Yuan J, Meng Y, Wang Z, Gu Z. 2017. Pathogenic Elizabethkingia miricola Infection in Cultured Black-Spotted Frogs, China, 2016. Emerg Infect Dis 23:2055-2059.

26. Peter S, Oberhettinger P, Schuele L, Dinkelacker A, Vogel W, Dorfel D, Bezdan D, Ossowski S, Marschal M, Liese J, Willmann M. 2017. Genomic characterisation of clinical and environmental Pseudomonas putida group strains and determination of

their role in the transfer of antimicrobial resistance genes to Pseudomonas aeruginosa. BMC Genomics 18:859.

27. Bellanger X, Payot S, Leblond-Bourget N, Guedon G. 2014. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. FEMS Microbiol Rev 38:720-60.

28. Cury J, Touchon M, Rocha EPC. 2017. Integrative and conjugative elements and their hosts: composition, distribution and organization. Nucleic Acids Res 45:8943-8956.

29. Guglielmini J, Quintais L, Garcillan-Barcia MP, de la Cruz F, Rocha EP. 2011. The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. PLoS Genet 7:e1002222.

30. Wozniak RA, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, Garriss G, Dery C, Burrus V, Waldor MK. 2009. Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. PLoS Genet 5:e1000786.

31. Wozniak RA, Waldor MK. 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol 8:552-63.

32. Johnson CM, Grossman AD. 2015. Integrative and Conjugative Elements (ICEs): What They Do and How They Work. Annu Rev Genet 49:577-601.

33. Colombi E, Straub C, Kunzel S, Templeton MD, McCann HC, Rainey PB. 2017. Evolution of copper resistance in the kiwifruit pathogen Pseudomonas syringae pv. actinidiae through acquisition of integrative conjugative elements and plasmids. Environ Microbiol 19:819-832.

34. Sugimoto M, Watada M, Jung SW, Graham DY, Yamaoka Y. 2012. Role of Helicobacter pylori plasticity region genes in development of gastroduodenal diseases. J Clin Microbiol 50:441-8.

557 35. Teo J, Tan SY, Liu Y, Tay M, Ding Y, Li Y, Kjelleberg S, Givskov M, Lin RT, Yang
558 L. 2014. Comparative genomic analysis of malaria mosquito vector-associated novel
559 pathogen Elizabethkingia anophelis. Genome Biol Evol 6:1158-65.
560 36. Li Y, Liu Y, Chew SC, Tay M, Salido MM, Teo J, Lauro FM, Givskov M, Yang L.
561 2015. Complete Genome Sequence and Transcriptomic Analysis of the Novel
562 Pathogen Elizabethkingia anophelis in Response to Oxidative Stress. Genome Biol
563 Evol 7:1676-85.
564 37. Kukutla P, Lindberg BG, Pei D, Rayl M, Yu W, Steritz M, Faye I, Xu J. 2014.
565 Insights from the genome annotation of Elizabethkingia anophelis from the malaria
566 vector Anopheles gambiae. PLoS One 9:e97715.
567 38. Chew KL, Cheng B, Lin RTP, Teo JWP. 2018. Elizabethkingia anophelis Is the
568 Dominant Elizabethkingia Species Found in Blood Cultures in Singapore. J Clin
569 Microbiol 56.
570 39. Breurec S, Criscuolo A, Diancourt L, Rendueles O, Vandenbogaert M, Passet V, Caro
571 V, Rocha EP, Touchon M, Brisse S. 2016. Genomic epidemiology and global
572 diversity of the emerging bacterial pathogen Elizabethkingia anophelis. Sci Rep
573 6:30379.
574 40. Perrin A, Larsonneur E, Nicholson AC, Edwards DJ, Gundlach KM, Whitney AM,
575 Gulvik CA, Bell ME, Rendueles O, Cury J, Hugon P, Clermont D, Enouf V, Loparev
576 V, Juieng P, Monson T, Warshauer D, Elbadawi LI, Walters MS, Crist MB, Noble-
577 Wang J, Borlaug G, Rocha EPC, Criscuolo A, Touchon M, Davis JP, Holt KE,
578 McQuiston JR, Brisse S. 2017. Evolutionary dynamics and genomic features of the
579 Elizabethkingia anophelis 2015 to 2016 Wisconsin outbreak strain. Nat Commun
580 8:15483.

581 41. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, Garcia-Cobos S,
582 Kooistra-Smid AM, Raangs EC, Rosema S, Veloo AC, Zhou K, Friedrich AW,
583 Rossen JW. 2017. Application of next generation sequencing in clinical microbiology
584 and infection prevention. J Biotechnol 243:16-24.
585 42. Tagini F, Greub G. 2017. Bacterial genome sequencing in clinical microbiology: a
586 pathogen-oriented review. Eur J Clin Microbiol Infect Dis 36:2007-2020.
587 43. Vernikos G, Medini D, Riley DR, Tettelin H. 2015. Ten years of pan-genome
588 analyses. Curr Opin Microbiol 23:148-54.
589 44. Robinson ER, Walker TM, Pallen MJ. 2013. Genomics and outbreak investigation:
590 from sequence to consequence. Genome Med 5:36.
591 45. Sun G, Wang L, Bao C, Li T, Ma L, Chen L. 2015. Complete Genome Sequence of
592 Elizabethkingia meningoseptica, Isolated from a T-Cell Non-Hodgkin's Lymphoma
593 Patient. Genome Announc 3.
594 46. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. 2005. The microbial pan-
595 genome. Curr Opin Genet Dev 15:589-94.
596 47. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV,
597 Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M,
598 Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R,
599 Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH,
600 Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins
601 K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC,
602 Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. 2005. Genome analysis
603 of multiple pathogenic isolates of Streptococcus agalactiae: implications for the
604 microbial "pan-genome". Proc Natl Acad Sci U S A 102:13950-5.

48. Leekitcharoenphon P, Kaas RS, Thomsen MC, Friis C, Rasmussen S, Aarestrup FM. 2012. snpTree--a web-server to identify and construct SNP trees from whole genome sequence data. BMC Genomics 13 Suppl 7:S6.

49. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol 15:524.

50. Seaver LC, Imlay JA. 2001. Alkyl hydroperoxide reductase is the primary scavenger of endogenous hydrogen peroxide in Escherichia coli. J Bacteriol 183:7173-81.

51. Oliveira PH, Touchon M, Rocha EP. 2014. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res 42:10618-31.

52. Venter H, Mowla R, Ohene-Agyei T, Ma S. 2015. RND-type drug e ffl ux pumps from Gram-negative bacteria: molecular mechanism and inhibition. Front Microbiol 6:377.

53. Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J. 2016. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science 353:aad5147.

54. Koonin EV, Makarova KS, Zhang F. 2017. Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol 37:67-78.

55. Bobossi-Serengbe G, Gody JC, Beyam NE, Bercion R. 2006. [First documented case of Chryseobacterium meningosepticum meningitis in Central African Republic]. Med Trop (Mars) 66:182-4.

56. Bentley SD, Parkhill J. 2015. Genomic perspectives on the evolution and spread of bacterial pathogens. Proc Biol Sci 282:20150488.

629 57. Mira A, Martin-Cuadrado AB, D'Auria G, Rodriguez-Valera F. 2010. The bacterial
630 pan-genome:a new paradigm in microbiology. Int Microbiol 13:45-57.
631 58. Wang Y, Gilbreath TM, 3rd, Kukutla P, Yan G, Xu J. 2011. Dynamic gut microbiome
632 across life history of the malaria mosquito Anopheles gambiae in Kenya. PLoS One
633 6:e24767.
634 59. Telang A, Skinner J, Nemitz RZ, McClure AM. 2018. Metagenome and Culture-
635 Based Methods Reveal Candidate Bacterial Mutualists in the Southern House
636 Mosquito (Diptera: Culicidae). J Med Entomol doi:10.1093/jme/tjy056.
637 60. Terenius O, Lindh JM, Eriksson-Gonzales K, Bussiere L, Laugen AT, Bergquist H,
638 Titanji K, Faye I. 2012. Midgut bacterial dynamics in Aedes aegypti. FEMS
639 Microbiol Ecol 80:556-65.
640 61. Burrus V, Waldor MK. 2004. Shaping bacterial genomes with integrative and
641 conjugative elements. Res Microbiol 155:376-86.
642 62. Malanowska K, Salyers AA, Gardner JF. 2006. Characterization of a conjugative
643 transposon integrase, IntDOT. Mol Microbiol 60:1228-40.
644 63. Nguyen M, Vedantam G. 2011. Mobile genetic elements in the genus Bacteroides,
645 and their mechanism(s) of dissemination. Mob Genet Elements 1:187-196.
646 64. Wang Y, Wang GR, Shelby A, Shoemaker NB, Salyers AA. 2003. A newly
647 discovered Bacteroides conjugative transposon, CTnGERM1, contains genes also
648 found in gram-positive bacteria. Appl Environ Microbiol 69:4595-603.
649 65. Laprise J, Yoneji S, Gardner JF. 2013. IntDOT interactions with core sites during
650 integrative recombination. J Bacteriol 195:1883-91.
651 66. Whittle G, Shoemaker NB, Salyers AA. 2002. Characterization of genes involved in
652 modulation of conjugal transfer of the Bacteroides conjugative transposon CTnDOT.
653 J Bacteriol 184:3839-47.

654 67. Waters JL, Salyers AA. 2013. Regulation of CTnDOT conjugative transfer is a
655 complex and highly coordinated series of events. MBio 4:e00569-13.
656 68. Wood MM, Gardner JF. 2015. The Integration and Excision of CTnDOT. Microbiol
657 Spectr 3:MDNA3-0020-2014.
658 69. Guglielmini J, Neron B, Abby SS, Garcillan-Barcia MP, de la Cruz F, Rocha EP.
659 2014. Key components of the eight classes of type IV secretion systems involved in
660 bacterial conjugation or protein secretion. Nucleic Acids Res 42:5715-27.
661 70. Au KG, Clark S, Miller JH, Modrich P. 1989. Escherichia coli mutY gene encodes an
662 adenine glycosylase active on G-A mispairs. Proc Natl Acad Sci U S A 86:8877-81.
663 71. Michaels ML, Pham L, Nghiem Y, Cruz C, Miller JH. 1990. MutY, an adenine
664 glycosylase active on G-A mispairs, has homology to endonuclease III. Nucleic Acids
665 Res 18:3841-5.
666 72. Murray NE. 2000. Type I restriction systems: sophisticated molecular machines (a
667 legacy of Bertani and Weigle). Microbiol Mol Biol Rev 64:412-34.
668 73. Nikaido H, Pages JM. 2012. Broad-specificity efflux pumps and their role in
669 multidrug resistance of Gram-negative bacteria. FEMS Microbiol Rev 36:340-63.
670 74. Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL. 1997. Arac/XylS family
671 of transcriptional regulators. Microbiol Mol Biol Rev 61:393-410.
672 75. Yang J, Tauschek M, Robins-Browne RM. 2011. Control of bacterial virulence by
673 AraC-like regulators that respond to chemical signals. Trends Microbiol 19:128-35.
674 76. Tan A, Petty NK, Hocking D, Bennett-Wood V, Wakefield M, Praszkier J, Tauschek
675 M, Yang J, Robins-Browne R. 2015. Evolutionary adaptation of an AraC-like
676 regulatory protein in Citrobacter rodentium and Escherichia species. Infect Immun
677 83:1384-95.

77. Tobes R, Ramos JL. 2002. AraC-XylS database: a family of positive transcriptional regulators in bacteria. Nucleic Acids Res 30:318-21.

78. Koonin EV, Makarova KS. 2017. Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back. Genome Biol Evol 9:2812-2825.

79. Laanto E, Hoikkala V, Ravantti J, Sundberg LR. 2017. Long-term genomic coevolution of host-parasite interaction in the natural environment. Nat Commun 8:111.

80. Castillo D, Christiansen RH, Dalsgaard I, Madsen L, Middelboe M. 2015. Bacteriophage resistance mechanisms in the fish pathogen Flavobacterium psychrophilum: linking genomic mutations to changes in bacterial virulence factors. Appl Environ Microbiol 81:1157-67.

81. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206-14.

82. Burrus V, Pavlovic G, Decaris B, Guedon G. 2002. Conjugative transposons: the tip of the iceberg. Mol Microbiol 46:601-10.

TABLE AND FIGURES LEGENDS

Table 1. ***Elizabethkingia anophelis*** **strains in the study**

Table 2. Types of ICEs identified in the strains

Figure 1. Evolutionary relationship and geographic locations of the strains. (A) The phylogenetic was tree derived from the core genome SNP comparison. Circles A-E demonstrate the protein identity between the genomes in the corresponding clades. The

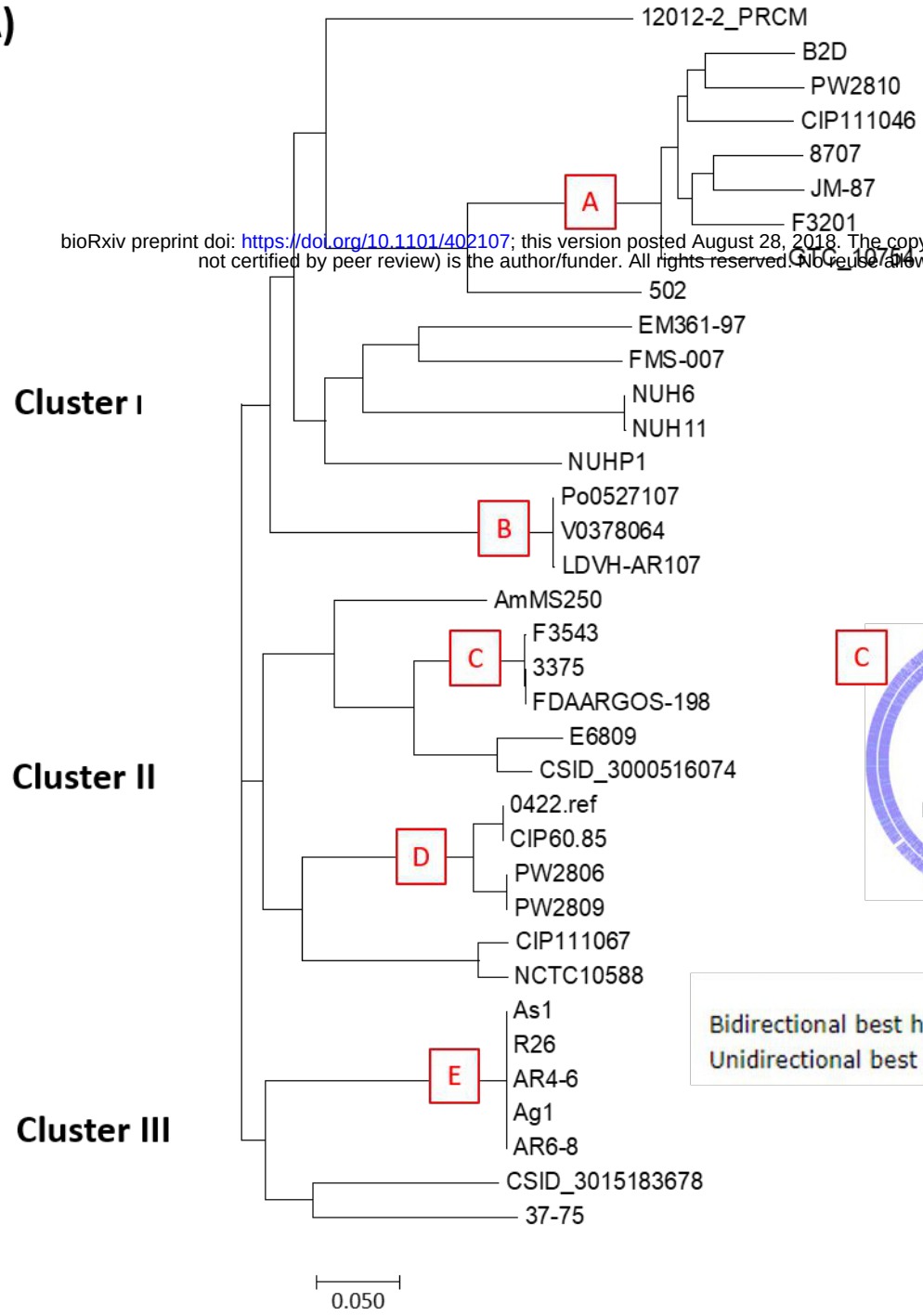
colour represents the percent identity when a genome was compared to the reference genome. (B) Geographic distribution of the strains. The letters correspond to the clades in part A of this figure.

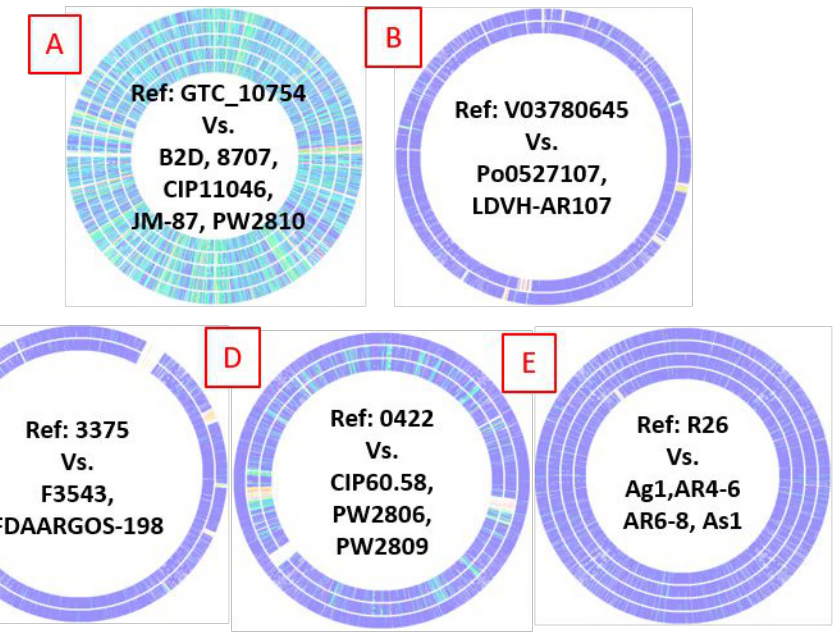
Figure 2. Schematic view of the architecture of conjugation modular genes in the three types of ICEs.

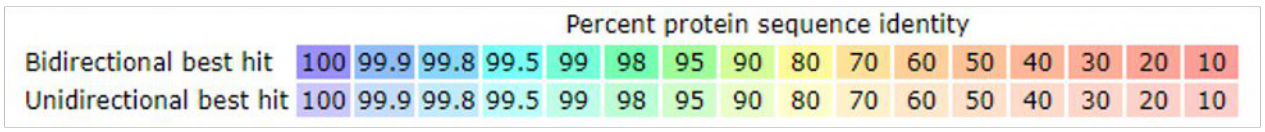
Figure 3. Integration sites of ICEs in different strains. (A) Location of the ICE*Ea*I integration sites. (B) Location of the tRNA genes where ICE*Ea*II and ICE*Ea*III integrated. The ICE types were color-coded. Refer to Table 2 for strain information.

Figure 4. Phylogenetic relationship of the genes *T4CP, relaxase, TraG* and *TraJ*. The nucleotide sequences from different ICEs were aligned and the evolutionary history was inferred using the Neighbor-Joining method. The evolutionary distances were computed using different models to reconstruct the phylogenetic trees with the bootstrap test using 1000 replicates, which generated similar tree topology. The consensus trees generated using Kimura 2-parameter model were presented. The bootstrap values were shown on the node.

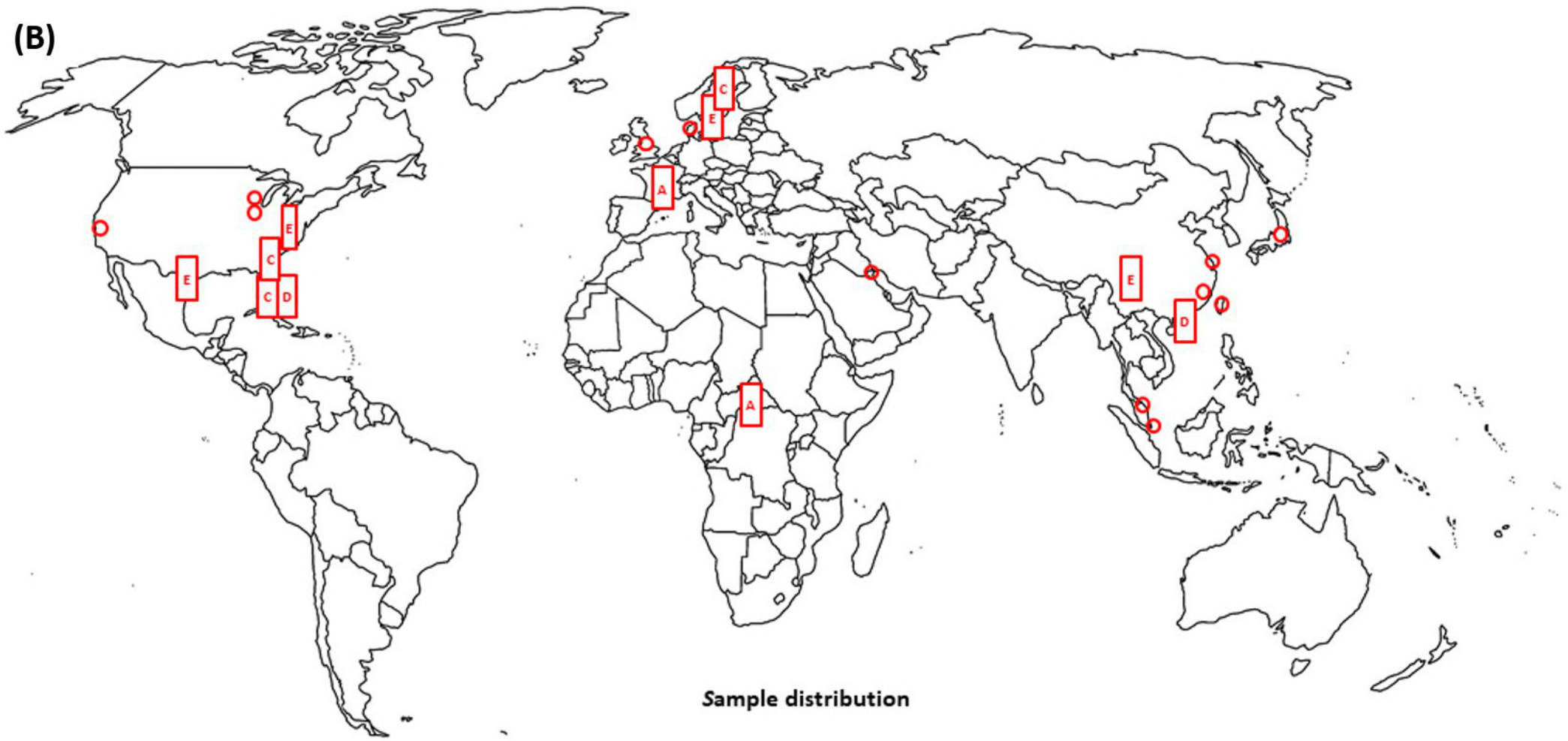
(A)

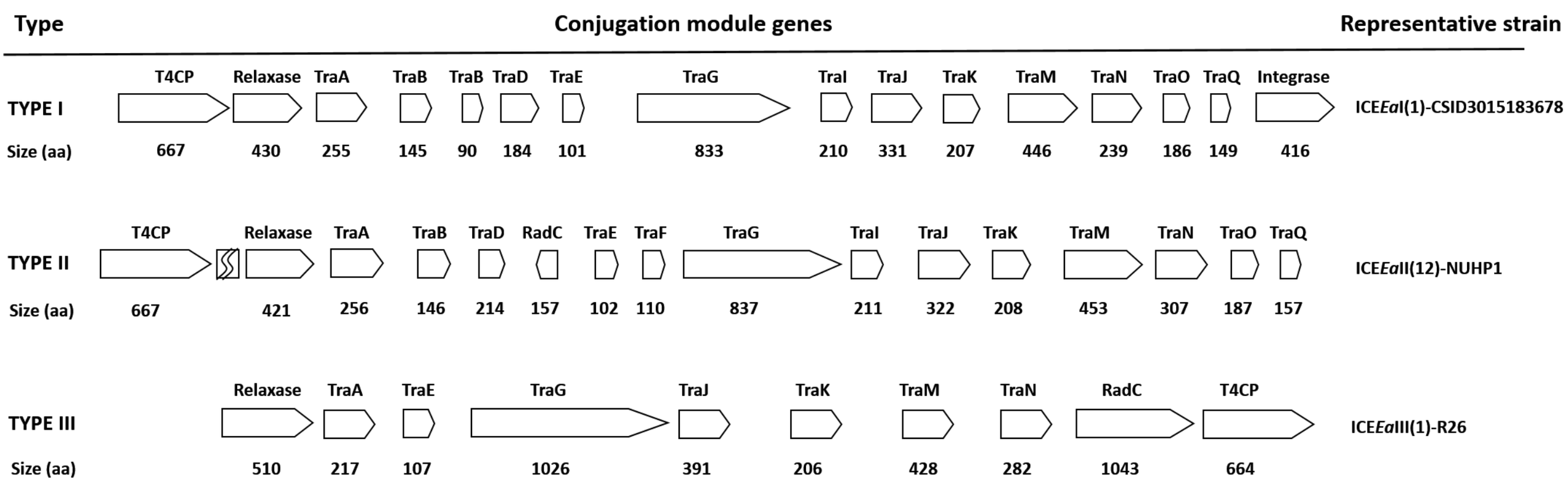


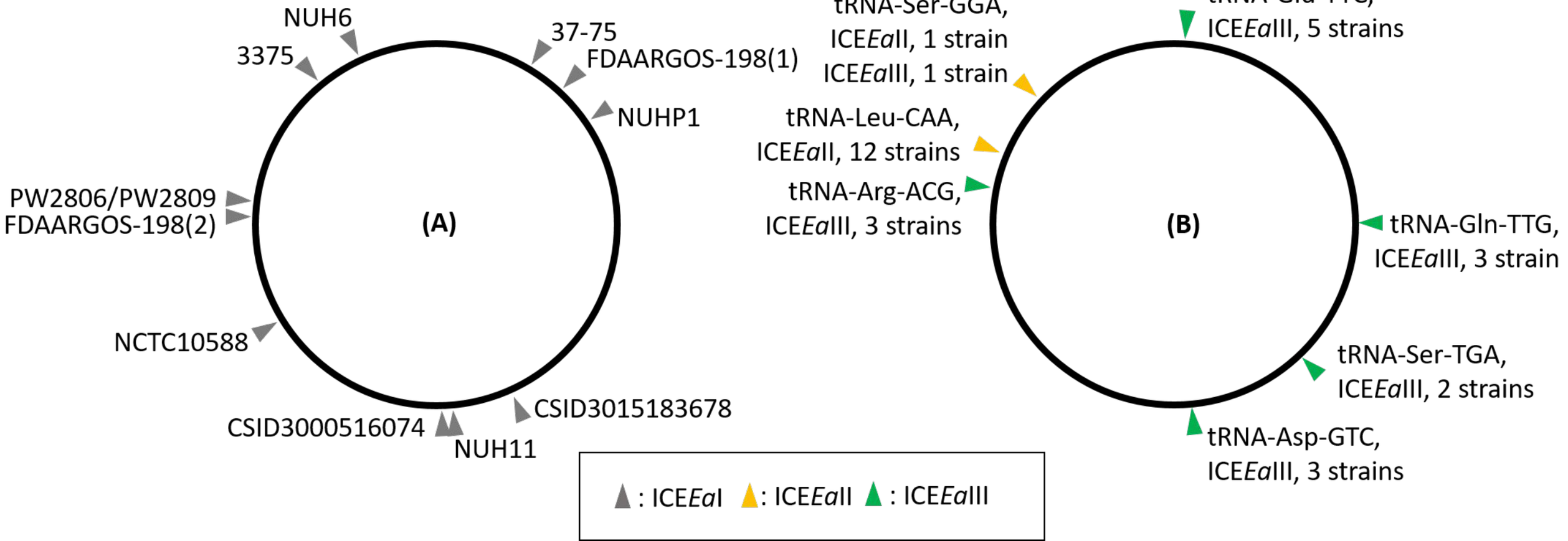




(B)




Type
Conjugation module genes
Representative strain
TYPE I
T4CP
Relaxase
TraA
TraB
TraB
TraD
TraE
TraG
TraI
TraJ
TraK
TraM
TraN
TraO
TraQ
Integrase
ICEEaI(1)-CSID3015183678
Size (aa)
667
430
255
145
90
184
101
833
210
331
207
446
239
186
149
416
TYPE II
T4CP
Relaxase
TraA
TraB
TraD
RadC
TraE
TraF
TraG
TraI
TraJ
TraK
TraM
TraN
TraO
TraQ
ICEEaII(12)-NUHP1
Size (aa)
667
421
256
146
214
157
102
110
837
211
322
208
453
307
187
157
TYPE III
Relaxase
TraA
TraE
TraG
TraJ
TraK
TraM
TraN
RadC
T4CP
ICEEaIII(1)-R26
Size (aa)
510
217
107
1026
391
206
428
282
1043
664


NUH6
3375
37-75
FDAARGOS-198(1)
NUHP1
PW2806/PW2809
FDAARGOS-198(2)
(A)
NCTC10588
CSID3000516074
NUH11
CSID3015183678
tRNA-Ser-GGA,
ICEEaII, 1 strain
ICEEaIII, 1 strain
tRNA-Leu-CAA,
ICEEaII, 12 strains
tRNA-Arg-ACG,
ICEEaIII, 3 strains
tRNA-Glu-TTC,
ICEEaIII, 5 strains
(B)
tRNA-Gln-TTG,
ICEEaIII, 3 strain
tRNA-Ser-TGA,
ICEEaIII, 2 strains
tRNA-Asp-GTC,
ICEEaIII, 3 strains
: ICEEaI
: ICEEaII
: ICEEaIII

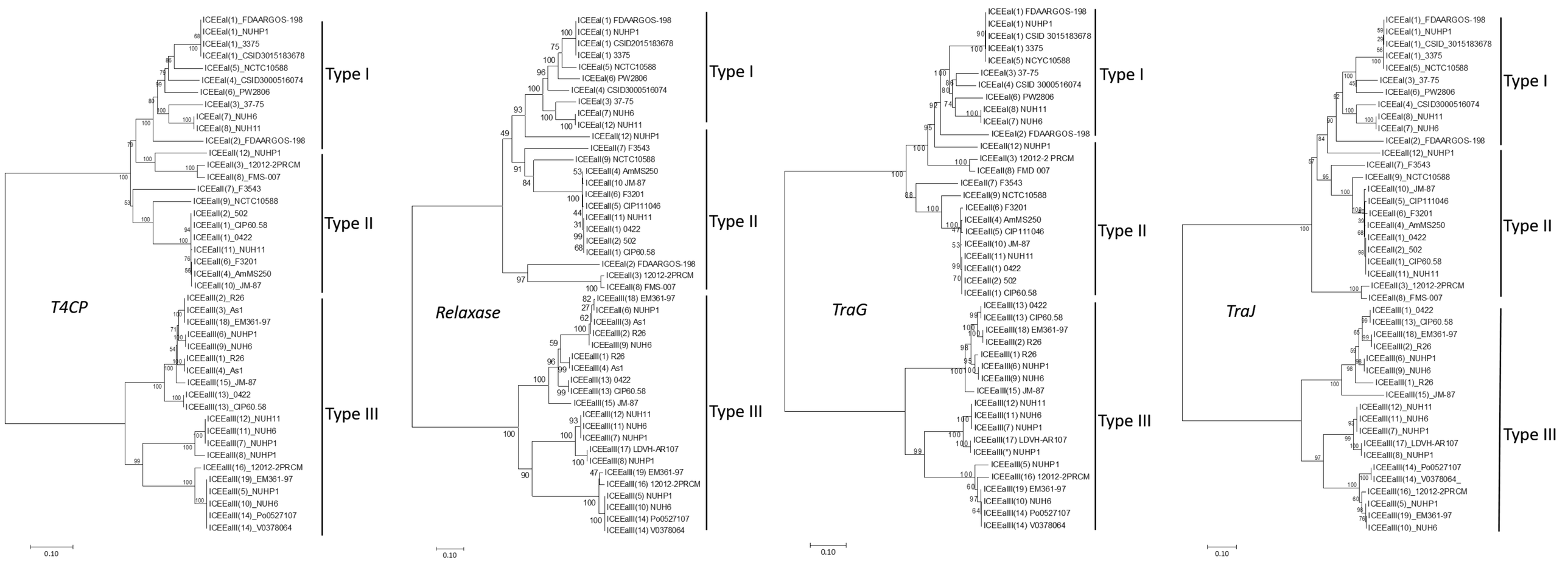

T4CP
ICEEal(1)_FDAARGOS-198
ICEEal(1)_NUHP1
ICEEal(1)_3375
ICEEal(1)_CSID3015183678
ICEEal(5)_NCTC10588
ICEEal(4)_CSID3000516074
ICEEal(6)_PW2806
ICEEal(3)_37-75
ICEEal(7)_NUH6
ICEEal(8)_NUH11
ICEEal(2)_FDAARGOS-198
ICEEall(12)_NUHP1
ICEEall(3)_12012-2PRCM
ICEEall(8)_FMS-007
ICEEall(7)_F3543
ICEEall(9)_NCTC10588
ICEEall(2)_502
ICEEall(1)_CIP60.58
ICEEall(1)_0422
ICEEall(11)_NUH11
ICEEall(6)_F3201
ICEEall(4)_AmMS250
ICEEall(10)_JM-87
ICEEalII(2)_R26
ICEEalII(3)_As1
ICEEalII(18)_EM361-97
ICEEalII(6)_NUHP1
ICEEalII(9)_NUH6
ICEEalII(1)_R26
ICEEalII(4)_As1
ICEEalII(15)_JM-87
ICEEalII(13)_0422
ICEEalII(13)_CIP60.58
ICEEalII(12)_NUH11
ICEEalII(11)_NUH6
ICEEalII(7)_NUHP1
ICEEalII(8)_NUHP1
ICEEalII(16)_12012-2PRCM
ICEEalII(19)_EM361-97
ICEEalII(5)_NUHP1
ICEEalII(10)_NUH6
ICEEalII(14)_Po0527107
ICEEalII(14)_V0378064
Type I
Type II
Type III
0.10
Relaxase
ICEEal(1) FDAARGOS-198
ICEEal(1) NUHP1
ICEEal(1) CSID2015183678
ICEEal(1) 3375
ICEEal(5) NCTC10588
ICEEal(6) PW2806
ICEEal(4) CSID3000516074
ICEEal(3) 37-75
ICEEal(7) NUH6
ICEEal(12) NUH11
ICEEall(12) NUHP1
ICEEall(7) F3543
ICEEall(9) NCTC10588
ICEEall(4) AmMS250
ICEEall(10 JM-87
ICEEall(6) F3201
ICEEall(5) CIP111046
ICEEall(11) NUH11
ICEEall(1) 0422
ICEEall(2) 502
ICEEall(1) CIP60.58
ICEEal(2) FDAARGOS-198
ICEEall(3) 12012-2PRCM
ICEEall(8) FMS-007
ICEEalII(18) EM361-97
ICEEall(6) NUHP1
ICEEalII(3) As1
ICEEalII(2) R26
ICEEalII(9) NUH6
ICEEalII(1) R26
ICEEalII(4) As1
ICEEalII(13) 0422
ICEEalII(13) CIP60.58
ICEEalII(15) JM-87
ICEEalII(12) NUH11
ICEEalII(11) NUH6
ICEEalII(7) NUHP1
ICEEalII(17) LDVH-AR107
ICEEalII(8) NUHP1
ICEEalII(19) EM361-97
ICEEalII(16) 12012-2PRCM
ICEEalII(5) NUHP1
ICEEalII(10) NUH6
ICEEalII(14) Po0527107
ICEEalII(14) V0378064
Type I
Type II
Type III
0.10
TraG
ICEEal(1) FDAARGOS-198
ICEEal(1) NUHP1
ICEEal(1) CSID 3015183678
ICEEal(1) 3375
ICEEal(5) NCYC10588
ICEEal(3) 37-75
ICEEal(4) CSID 3000516074
ICEEal(6) PW2806
ICEEal(8) NUH11
ICEEal(7) NUH6
ICEEal(2) FDAARGOS-198
ICEEall(12) NUHP1
ICEEall(3) 12012-2 PRCM
ICEEall(8) FMD 007
ICEEall(7) F3543
ICEEall(9) NCTC10588
ICEEall(6) F3201
ICEEall(4) AmMS250
ICEEall(5) CIP111046
ICEEall(10) JM-87
ICEEall(11) NUH11
ICEEall(1) 0422
ICEEall(2) 502
ICEEall(1) CIP60.58
ICEEalII(13) 0422
ICEEalII(13) CIP60.58
ICEEalII(18) EM361-97
ICEEalII(2) R26
ICEEalII(1) R26
ICEEalII(6) NUHP1
ICEEalII(9) NUH6
ICEEalII(15) JM-87
ICEEalII(12) NUH11
ICEEalII(11) NUH6
ICEEalII(7) NUHP1
ICEEalII(17) LDVH-AR107
ICEEalII(*) NUHP1
ICEEalII(5) NUHP1
ICEEalII(16) 12012-2PRCM
ICEEalII(19) EM361-97
ICEEalII(10) NUH6
ICEEalII(14) Po0527107
ICEEalII(14) V0378064
Type I
Type II
Type III
0.10
TraJ
ICEEal(1)_FDAARGOS-198
ICEEal(1)_NUHP1
ICEEal(1)_CSID_3015183678
ICEEal(1)_3375
ICEEal(5)_NCTC10588
ICEEal(3)_37-75
ICEEal(6)_PW2806
ICEEal(4)_CSID3000516074
ICEEal(8)_NUH11
ICEEal(7)_NUH6
ICEEal(2)_FDAARGOS-198
ICEEall(12)_NUHP1
ICEEall(7)_F3543
ICEEall(9)_NCTC10588
ICEEall(10)_JM-87
ICEEall(5)_CIP111046
ICEEall(6)_F3201
ICEEall(4)_AmMS250
ICEEall(1)_0422
ICEEall(2)_502
ICEEall(1)_CIP60.58
ICEEall(11)_NUH11
ICEEall(3)_12012-2PRCM
ICEEall(8)_FMS-007
ICEEalII(1)_0422
ICEEalII(13)_CIP60.58
ICEEalII(18)_EM361-97
ICEEalII(2)_R26
ICEEalII(6)_NUHP1
ICEEalII(9)_NUH6
ICEEalII(1)_R26
ICEEalII(15)_JM-87
ICEEalII(12)_NUH11
ICEEalII(11)_NUH6
ICEEalII(7)_NUHP1
ICEEalII(17)_LDVH-AR107
ICEEalII(8)_NUHP1
ICEEalII(14)_Po0527107
ICEEalII(14)_V0378064_
ICEEalII(16)_12012-2PRCM
ICEEalII(5)_NUHP1
ICEEalII(19)_EM361-97
ICEEalII(10)_NUH6
Type I
Type II
Type III
0.10

Table 2. Types of ICEs identified in the strains

Element	Strain	Integration site	Size (nt)	ICE TPA GenBank accession
ICE*Ea*I(1)	3375	Siroheme synthase and HP	63,549	BK010586
ICE*Ea*I(1)	CSID3015183678	Inside mutY	62,894	BK010587
ICE*Ea*I(1)	FDAARGOS-198	EngB and ElaA	62,860	BK010588
ICE*Ea*I(1)	NUHP1	SlyD and HP	62,960	BK010589
ICE*Ea*I(2)	FDAARGOS-198	TonB receptor and HP	69,029	BK010590
ICE*Ea*I(3)	37-75	Inside Ahp	65,419	Not available**
ICE*Ea*I(4)	CSID3000516074	HP and LuxR TF	97,160	BK010591
ICE*Ea*I(5)	NCTC10588	Macrolide-efflux protein and LytR/AlgR TR	81,809	Not available**
ICE*Ea*I(6)	PW2806, PW2809	Efflux protein and peptidase	79,842	BK010592
ICE*Ea*I(7)	NUH6	TomB receptor and HP	59,635	BK010593
ICE*Ea*I(8)	NUH11	tRNA-Asp-GTC*	95,803	BK010594
ICE*Ea*II(1)	0422	tRNA-Leu-CAA	63,840	BK010595
ICE*Ea*II(1)	CIP60.58	tRNA-Leu-CAA	63,840	BK010596
ICE*Ea*II(2)	502	tRNA-Leu-CAA	>34270	BK010597
ICE*Ea*II(3)	12012-2 PRCM	tRNA-Leu-CAA	>56734	BK010598
ICE*Ea*II(4)	AmMS250	tRNA-Leu-CAA	>64773	Not available**
ICE*Ea*II(5)	CIP111046	tRNA-Leu-CAA	>22876	BK010599
ICE*Ea*II(6)	F3201	tRNA-Leu-CAA	91,608	BK010600
ICE*Ea*II(7)	F3543	tRNA-Leu-CAA	104,603	BK010601
ICE*Ea*II(8)	FMS-007	tRNA-Leu-CAA	73,167	BK010602
ICE*Ea*II(9)	NCTC10588	tRNA-Leu-CAA	67,087	Not available**
ICE*Ea*II(10)	JM-87	tRNA-Leu-CAA	36,967	BK010603
ICE*Ea*II(11)	NUH11	tRNA-Leu-CAA	>36957	BK010604
ICE*Ea*II(12)	NUHP1	tRNA-Ser-GGA	71,591	BK010605
ICE*Ea*III(1)	R26, Ag1, Ar4-6, AR6-8	tRNA-Ser-TGA	101,692	BK010606
ICE*Ea*III(2)	R26, Ag1, Ar4-6, AR6-8	tRNA-Arg-ACG	77,358	BK010607
ICE*Ea*III(3)	As1	tRNA-Arg-ACG	>44,888	BK010608
ICE*Ea*III(4)	As1	tRNA-Ser-TGA	>31,354	BK010609
ICE*Ea*III(5)	NUHP1	tRNA-Glu-TTC	60,900	BK010610
ICE*Ea*III(6)	NUHP1	tRNA-Glu-TTC	74,499	BK010611
ICE*Ea*III(7)	NUHP1	tRNA-Asp-GTC	116,331	BK010625
ICE*Ea*III(8)	NUHP1	tRNA-Asp-GTC	109,040	BK010626
ICE*Ea*III(9)	NUH6	tRNA-Gln-TTG	>30166	BK010612
ICE*Ea*III(10)	NUH6	tRNA-Glu-TTC	>33929	BK010623
ICE*Ea*III(11)	NUH6	tRNA-Asp-GTC	>98049	BK010613
ICE*Ea*III(12)	NUH11	Beta-lactamase*	84,534	BK010614
ICE*Ea*III(13)	0422	tRNA-Gln-TTG	67,662	BK010615
ICE*Ea*III(13)	CIP60.58	tRNA-Gln-TTG	66,636	BK010616
ICE*Ea*III(14)	Po0527107	tRNA-Glu-TTC	>31721	BK010617
ICE*Ea*III(14)	V0378064	tRNA-Glu-TTC	>37189	BK010618
ICE*Ea*III(15)	JM-87	tRNA-Ser-TGA	73,828	BK010624
ICE*Ea*III(16)	12012-2 PRCM	tRNA-Glu-TTC	>31732	BK010619
ICE*Ea*III(17)	LDVH-AR107	not determined	>24456	BK010620
ICE*Ea*III(18)	EM361-97	tRNA-Arg-ACG	69,393	BK010621

ICE*Ea*III(19)	EM361-97	tRNA-Glu-TTC	>28887	BK010622

*Two elements in NUH11 combine together and integrate between tRNA-Asp-GTC and beta-lactamase. ** There are no assemblies in the NCBI database for these four strains. So we did not submit third party annotation (TPA) for these genomes to the GenBank. >: elements were partially assembled.

Table 1. *Elizabethkingia anophelis* strains in the study

WGS Accession No.	Level	Lineage	Strain	Source	Region	Collection Time	ICE type (n)[a]	CRISPR (n)[b]
ERS1197909[c]	Draft	II	AmMS250/CIP104057	Human patient	US	1994	II (1)	No
CP023010.1	Complete	II	FDAARGOS-198	Human patient	Sweden	Unknown	I (2)	No
MAHA01	Draft	II	CSID 3000516074	Human patient	Illinois, US	2016	I (1)	No
CP016373.1	Complete	II	3375	Human patient	South Carolina, US	1957	I (1)	No
MAHS01	Draft	II	E6809	Human patient	California, US	1979	II (1)	No
CP014340.1	Complete	II	F3543	Human patient	Florida, US	1982	II (1)	No
CP016370.1	Complete	II	0422	Human patient	Florida, US	1950	II (1), III (1)	No
FTQY01	Draft	II	CIP60.58	Unknown	Unknown	Unknown	II (1), III (1)	No
CBYD01	Draft	II	PW2806	Human patient	Hong Kong	2012	I (1)	No
CBYE01	Draft	II	PW2809	Human patient	Hong Kong	2012	I (1)	No
ERS605480[c]	Draft	II	NCTC10588	Human patient	US	1959	I (1), II (1)	No
FTQZ01	Draft	II	CIP111067	Unknown	Unknown	Unknown	No ICE	Yes (37)
CP014805.2	Complete	III	CSID 3015183678	Human patient	Wisconsin, US	2016	I (1)	No
ERS1197911[c]	Draft	III	37-75/CIP79.29	Human patient	St Nazaire, France	1979	I (1)	No
CP023401.1	Complete	III	R26	Mosquito *An.gambiae*	Stockholm, Sweden	2005	III (2)	No
CP023402.1	Complete	III	Ag1	Mosquito *An.gambiae*	New Mexico, US	2012	III (2)	No
CP023404.1	Complete	III	AR4-6	Mosquito *An.sinensis*	Sichuan, China	2015	III (2)	No
CP023403.1	Complete	III	AR6-8	Mosquito *An.sinensis*	Sichuan, China	2015	III (2)	No
LFKT01	Draft	III	As1	Mosquito *An. stephensi*	Pennsylvania, US	2015	III (2)	No
CP007547.1	Complete	I	NUHP1	Human patient	Singapore	2012	I (1), II (1), III (4)	No
ASYJ01	Draft	I	NUH6	Human patient	Singapore	2012	I (1), III (3)	No
ASYK01	Draft	I	NUH11	Human patient	Singapore	2012	I (1), II (1), III (1)	No
CCAC01	Draft	I	Po0527107	Human patient	Central African Republic	2006	III (1)	Yes (21)
CCAB01	Draft	I	V0378064	Human patient	Central African Republic	2011	III (1)	Yes (23)
FTPG01	Draft	I	LDVH-AR107	Common carp *Cyprinus carpio*	Montpellier, France	2004	III (1)	Yes (42)
CP006576.1	Complete	I	FMS-007	Human patient	China	2015	II (1)	Yes (15)
LWDS01	Draft	I	EM361-97	Human patient	Taiwan	2000s	III (1)	No
AVCQ01	Draft	I	502	Human patient	Birmingham, UK	2012	II (1)	No
CP016374.1	Complete	I	F3201	Human patient	Kuwait	1982	II (1)	No
CP016372.1	Complete	I	JM-87	Corn *Zea mays*	Alabama, US	2011	II (1), III (1)	No
ERS1197907[c]	Draft	I	8707/CIP78.9	Human patient	NY, US	1962	No ICE	Yes
JNCG01	Draft	I	B2D	Human patient	Malaysia	2013	No ICE	No

FTRB01	Draft	I	CIP111046	Human patient	Unknown	Unknown	II (1)	Yes (6)
CBYF01	Draft	I	PW2010	Human patient	Hong Kong	2012	No ICE	Yes (27)
DRS013860	Draft	I	GTC_10754	Unknown	Japan	2014	No ICE	Yes (32)
LPXG01	Draft	I	12012-2 PRCM	Human patient	Fujian, China	2009	II (1), III (1)	No

[a]n: number of the ICEs in the type. [b] n: number of the spacers in the CRISPR locus. [c] There are no assemblies in the NCBI database for these four strains. We assembled the genomes from Illumina reads directly.