Current Microbiology (2020) 77:762–768
https://doi.org/10.1007/s00284-019-01867-6

# The Complete Genome of *Emcibacter congregatus* ZYL$^T$, a Marine Bacterium Encoding a CRISPR-Cas 9 Immune System

**Zhe Zhao[1] · Rui-an Zhang[3] · Ge-yi Fu[4] · Ran Zhang[1] · Yan-fang Nie[1] · Cong Sun[2] · Min Wu[1]**

Received: 3 July 2019 / Accepted: 30 December 2019 / Published online: 9 January 2020


## Abstract

*Emcibacter congregatus* ZYL$^T$ was isolated from a sediment sample cultured in situ in a coast located in the East China Sea. The genome of *E. congregatus* ZYL$^T$ was sequenced and assembled into one single circular chromosome with the size of 4,189,011 bp and G+C content of 52.6%. Genomic annotation showed that *E. congregatus* ZYL$^T$ had an intact Type II-C CRISPR-Cas system consists of three cas genes (*cas 9*, *cas 1*, and *cas 2*), 34 direct repeat sequences with the length of 36 bp, and 33 spacers. The predicted Cas 9 protein was smaller than most of existing genome editing tools. This structure might have potential in developing new gene editing system and uncovering the regulatory mechanisms of CRISPR-Cas system. Besides, the comparison between *E. congregatus* ZYL$^T$ and its relative species living in neritic environments unraveled some common traits of the defective strategies of these bacteria to face inshore challenges including the motility, multidrug resistance, and universal efflux pumps.

## Introduction

The CRISPR-Cas immune system is one of the most significant discoveries in modern microbiology. The CRISPR-Cas modules are present in most archaea (nearly 90%) and many bacteria (50%) [1] and provide sequence-specific protection against invading genetic elements including foreign DNA and RNA [2].

CRISPR-Cas systems have been assigned to two classes, which are further subdivided into six types and several subtypes that each possess signature *cas* genes. Class 1 CRISPR-Cas systems, including types I, III, and IV, possess multi-subunit effector complexes comprised of multiple Cas proteins. For example, all type I CRISPR–Cas systems contain the signature gene *cas3*, among which the I–E system of *Escherichia coli* is most thoroughly characterized. The cascade of the type I-E CRISPR-Cas system has a molecular weight of 405 kDa and displays the following composition: $(Cas5e)_1$–$(Cas6e)_1$–$(Cas7e)_6$–$(Cas8e)_1$–$(Cas11e)_2$ [1]. However, class 2 systems, including types II, V, and VI, accomplish interference by a single, large effector protein [3]. Type I systems are most commonly seen in archaeal genomes (64%), and type III systems are the second commonly seen in archaeal genomes (34%) [4]. However, the most widely used and best studied CRISPR-Cas 9 system

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s00284-019-01867-6) contains supplementary material, which is available to authorized users.

✉ Cong Sun
michael_sc@sina.com

✉ Min Wu
wumin@zju.edu.cn

Zhe Zhao
zhaozhe@zju.edu.cn

Rui-an Zhang
zra@zju.edu.cn

Ge-yi Fu
fugeyi@163.com

Ran Zhang
zhangran@zju.edu.cn

Yan-fang Nie
yanfangnie18@163.com

[1] College of Life Sciences, Zhejiang University, Hangzhou 310058, People's Republic of China

[2] College of Life Sciences and Medicine, Zhejiang Sci-Tech University, Hangzhou 310018, People's Republic of China

[3] School of Materials Science and Engineering, Zhejiang University, Hangzhou 310058, People's Republic of China

[4] Ocean College, Zhejiang University, Hangzhou 310058, People's Republic of China

belongs to types II and is only a small part of bacterial CRISPR-Cas systems.

*Emcibacter congregatus* ZYL$^T$, isolated from sediment cultured in situ in the East China Sea, is an aerobic, rod-shaped, motile, and non-spore-forming bacterium belonging to the order *Emcibacterales*. To date, the order contains only one family *Emcibacteraceae*, and this family consists of only one genus *Emcibacter* [5]. There are only two species in this genus, *E. nanhaiensis* and *E. congregatus*. They form a special branch on the phylogenetic tree [6]. Both of the two species and most of their relative genera were isolated from neritic zones, including sediment and sea water. Though the draft genome of *E. congregatus* ZYL$^T$ had been sequenced [6], no complete genome has hitherto been reported in this special taxon. In this study, the genome of *E. congregatus* ZYL$^T$, which is the first complete genome within the genus *Emcibacter*, was sequenced and assembled into one circular chromosome, providing a basis for exploring ecological roles, genetic mechanisms, and application potentials of strains in genus *Emcibacter*.

According to the phylogenetic tree of *E. congregatus* ZYL$^T$ [6], the most related strain of *E. congregatus* ZYL$^T$ was *E. nanhaiensis* MCCC 1A06723$^T$. To understand the features of the genome better, we also sequenced the draft genome of *E. nanhaiensis* MCCC 1A06723$^T$. The other three representative species were chosen as references, including *Kordiimonas lacus* S3-22$^T$, *K. lipolytica* M41$^T$, and *Sneathiella glossodoripedis* MKT133$^T$. These three strains had close relationships with *E. congregatus* ZYL$^T$ on the phylogenetic tree [6]. Besides, they were all isolated from neritic environments [7–10] just like *E. congregatus* ZYL$^T$. We compared the genomes of these species and showed their universal mechanisms to prevent injuries from antibiotics and viruses in neritic environments.

## Materials and Methods

*E. congregatus* ZYL$^T$ was cultivated in Marine Broth 2216 (MB; BD), at 25 °C for 48 h, while *E. nanhaiensis* MCCC 1A06723$^T$ was cultivated in the same medium at 37 °C for 48 h. The cells were centrifuged at 12,000 rpm for 3 min at 4 °C. The genomic DNA of *E. congregatus* ZYL$^T$ was extracted using the high-salt cetyltrimethyl ammonium bromide (CTAB) method [11], while the genomic DNA of *E. nanhaiensis* MCCC 1A06723$^T$ was extracted using the SDS-Proteinase K method described by Marmur [12]. Total DNA of them was quantified by a Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), and the quality of DNAs was controlled by 1% gel electrophoresis.

For *E. congregatus* ZYL$^T$, the genome was sequenced using a PacBio RS II platform and Illumina HiSeq 4000 platform with insert length of 270 bp at the Beijing Genomics Institute (BGI, Shenzhen, China) (Table 1). PacBio libraries and Illumina paired-end libraries were constructed and sequenced yielding 170 × and 270 × average coverages, respectively. Draft genomic unitigs were assembled using the Celera Assembler 8.3 [13]. One circular chromosome with 0 gap was constructed finally.

For *E. nanhaiensis* MCCC 1A06723$^T$, paired-end sequencing was applied and two insert-size libraries (350 bp, 6000 bp) were constructed. A total of 1743 Mb clean data were generated, representing approximately 436-fold genomic coverage. De novo assembly was performed using ABySS v2.0.2 [14], and the contigs shorter than 2 k bp were removed. The quality of the assembled genome sequence was checked by CheckM v1.0.7 [15].

Genomes of *K. lacus* S3-22$^T$, *K. lipolytica* M41$^T$, and *S. glossodoripedis* MKT133$^T$ were obtained from DDBJ/EMBL/GenBank database under accession numbers of LRUA00000000, LRUB00000000, and BAMS00000000, and their completeness scores were assessed using CheckM 1.0.7 [15].

The open reading frames (ORFs) were predicted and annotated with a fully automated annotating service of Rapid Annotation using Subsystem Technology (RAST) server online [16]. Classification of predicted genes was analyzed by blastp with COGs database [17], and the functional annotation of genes and pathways was analyzed by KEGG databases [18] with KAAS web server (https://www.genome.jp/tools/kaas/). CDSs with signal peptides and transmembrane helices were identified through Signal 5.0 web server (https://www.cbs.dtu.dk/services/SignalP/) with set organism group to "Gram-negative bacteria" and TMHMM web server v.2.0 (https://www.cbs.dtu.dk/services/TMHMM/), respectively. Transfer RNA (tRNA) genes were predicted with tRNAscan-SE 2.0 On-line server (https://lowelab.ucsc.edu/tRNAscan-SE/). Ribosome RNAs (rRNA) were predicted with rRNAmmer 1.2 [19]. We predicted biosynthetic gene clusters of secondary metabolites by using antiSMASH 3.0 web server (https://antismash.secondarymetabolites.org). And the CRISPR array was predicted with CRISPRFinder (https://crispr.i2bc.paris-saclay.fr/Server/). Insertion sequences (IS) was predicted using ISfinder (https://www-is.biotoul.fr/) with e-value 0.001. Circular figure of *E. congregatus* ZYL$^T$ was visualized by using CGView (https://wishart.biology.ualberta.ca/cgview/).

## Results and Discussion

The genome of *E. congregatus* ZYL$^T$ contained one circular chromosome with size of 4,189,011 bp as well as DNA G+C contents of 52.6%. It had 49 tRNA genes and 9 operons of 16-23S-5S rRNA genes (Fig. 1 and supplementary material). The total number of predicted coding sequences (CDS) was

**Table 1** General features of *E. ongregates* ZYL$^T$ and MIGS mandatory information

| Items | Descriptions |
|---|---|
| Classification | |
| Domain | *Bacteria* |
| Phylum | *Proteobacteria* |
| Class | *Alphaproteobacteria* |
| Order | *Emcibacterales* |
| Family | *Emcibacteraceae* |
| Genus | *Emcibacter* |
| Species | *Emcibacter congregatus* |
| General features | |
| Gram strain | Negative |
| Cell shape | Rod |
| Color of colonies | Pale yellow |
| Temperature range (°C) (optimum) | 4–30 (25) |
| NaCl concentration (w/v) (Optimal) | 0–7% (2%) |
| pH range (optimal) | 6.0–8.5 (7.0) |
| Observed biotic relationship | Free-living |
| Trophic level | Chemoorganotroph |
| Relationship to oxygen | Aerobic |
| MIGS data | |
| Investigation type | Bacteria_archaea |
| Project name | Genome sequence of E. congregatus ZYLT |
| Latitude and longitude | 29° 56′ N, 122° 05′ E |
| Depth | Not reported |
| Geographical location | China: East China Sea |
| Collection date | 2017-11-01 |
| Environment (biome) | Neritic sediment cultured in situ |
| Number of replicons | 1 |
| Estimated size | 4,189,011 bp |
| Source material identifiers | ZYL$^T$ = KCTC 62328$^T$ = JCM 32378$^T$ = MCCC 1K03526$^T$ |
| Sequencing method | PacBio RS II, Illumina Hiseq |
| Assembly data | |
| Assembly method | Celera Assembler v.8.3, Pbdagcon v. MAY-2018, GATK v.4.0.2.0, SOAPindel v.2 |
| Genome coverage | 170×(PacBio RS II), 270×(Illumina Hiseq) |

3862, only 1664 from which could be classified into COG categories. The major categories of *E. congregatus* ZYL$^T$ were inorganic ion transport and metabolism (P), amino acid transport and metabolism (E), and transcription (K) (Fig. S1). Only 1841 CDS were classified into KEGG categories, the major pathways were two-component system, ribosome, purine metabolism, and oxidative phosphorylation. Compared to other references, the quantities and categories of COGs classified were quite similar (Fig. S1). The summary of general features and statistics of the genomes of *E. congregatus* ZYL$^T$ and references are listed in Tables 1 and 2.

Among all of the six species, *E. congregatus* ZYL$^T$ was the only one annotated to encode a CRISPR-Cas system. Based on the Cas 9 protein and gene structure of *E. congregatus* ZYL$^T$, we classify its CRISPR-Cas system as the Type II-C system. The CRISPR-Cas9 system contained three cas genes: *cas 9* (GenBank accession number MN064751), *cas 1* (GenBank accession number MN064749), and *cas 2* (GenBank accession number MN064750). The 3135 bp *cas 9* gene, 905 bp *cas 1* gene, and 305 bp *cas 2* gene interlock on genome and transcript in the same direction. The downstream CRISPR array consisted of 33 spacers. Each spacer was 30 bp (Table 3). The direct repeat sequence was 5′-GCC ATAGCTTCCATTCCGACACGGTTATGCTATGAT-3′. In *Streptococcus pyogenes* SF370, whose type II CRISPR-Cas 9 adaptive immune system has been repurposed for mammalian genome engineering as artificial plasmids: pX330 (U6-chimeric guide RNA + CBh-SpCas9 backbone, Addgene ID 42230) [20], the *cas 9* gene is 4101 bp, encoding 1367 amino acids. However, predicted Cas 9 protein

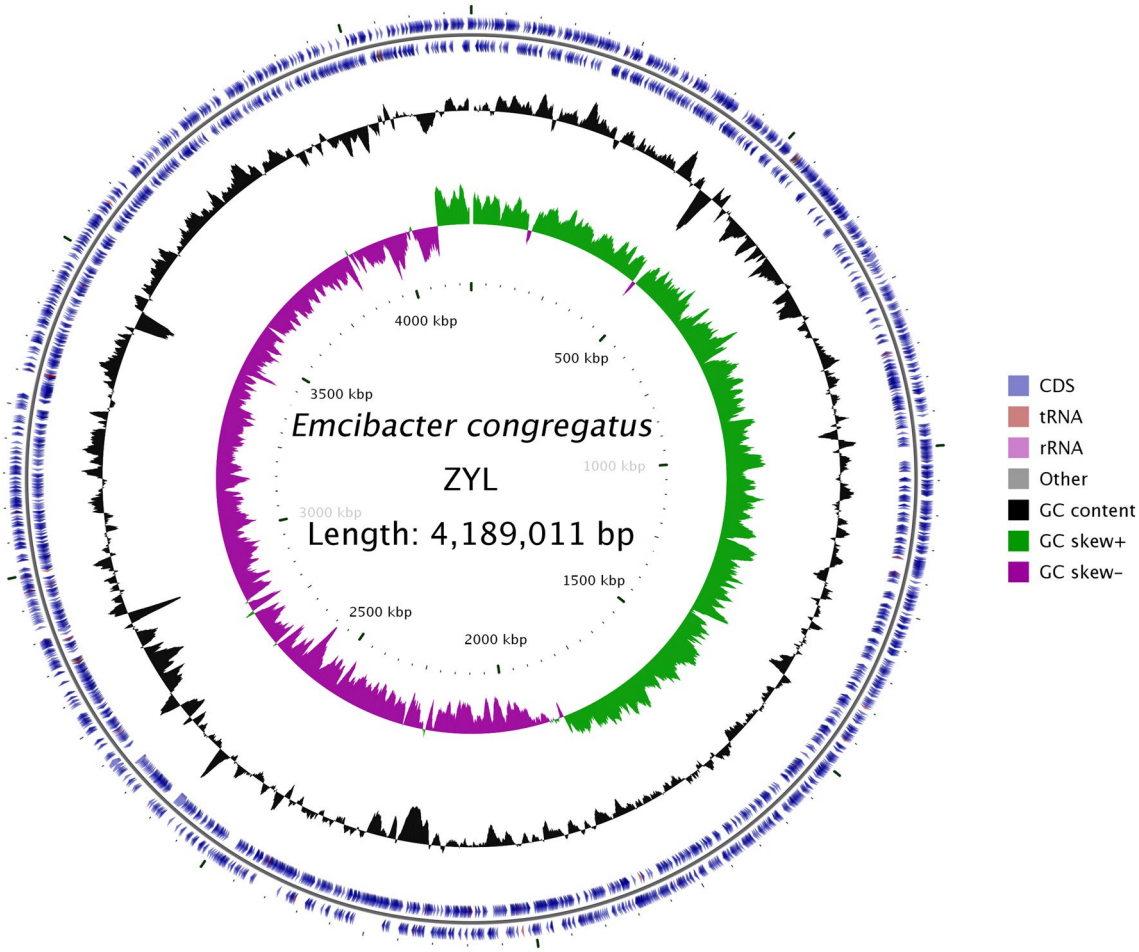


**Fig. 1** Circular map of the chromosome of *E. congregatus* ZYL$^T$. From the outside to the center: rRNA and tRNA genes on forward strand, CDS genes assigned to the comparative genomic result on forward strand, CDS genes assigned to the comparative genomic result on reverse strand, rRNA and tRNA genes on reverse strand, DNA G+C content and GC skew

**Table 2** Comparative genomic features of *E. congregatus* ZYL$^T$ and related strains

| Strains | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Contigs | 1 | 19 | 10 | 36 | 12 |
| Size(bp) | 4,189,011 | 3,963,826 | 4,024,630 | 4,557,767 | 3,629,390 |
| DNA G+C content (%) | 52.6 | 56.2 | 57.2 | 56.3 | 46.9 |
| CDSs | 3862 | 3554 | 3694 | 4238 | 3800 |
| CDSs assigned to COGs | 1664 | 1561 | 1566 | 1741 | 1434 |
| CDSs assigned to KEGG | 1841 | 1804 | 1733 | 1975 | 1832 |
| CDSs with signal peptides | 757 | 573 | 752 | 821 | 504 |
| CDSs with trans-membrane helices | 897 | 743 | 811 | 932 | 785 |
| Biosynthetic gene clusters of secondary metabolites | 5 | 5 | 5 | 5 | 4 |
| rRNA opron (16-5S-23S rRNA) | 9 | 12 | 12 | 9 | 3 |
| tRNAs | 49 | 43 | 56 | 53 | 43 |
| Insertion sequences (IS) | 22 | 8 | 71 | 51 | 5 |
| Completeness (%) | 100 | 99.07 | 99.57 | 100 | 99.12 |

Strains: 1. *E. congregatus* ZYL$^T$; 2. *E. nanhaiensis* MCCC 1A06723$^T$; 3. *K. lacus* S3-22$^T$; 4. *K. lipolytica* M41$^T$; 5. *S. glossodoripedis* MKT133$^T$

**Table 3** Spacer sequences of *E. congregatus* ZYL$^T$ predicted by CRISPRFinder

| Spacer numbers | Spacer sequences |
|---|---|
| Spacer 1 | TCAAGGGTGGTGACCGGTAGGGCCTCTGTG |
| Spacer 2 | CTGGCCGTTGAAGCAGAACAGCGTGAAGCC |
| Spacer 3 | GGATATGCAATTCTCCGCTTGGAATATCGG |
| Spacer 4 | GATTTTTCATTCCCTTTAAAGCCTTTTTGG |
| Spacer 5 | TGTTCAAGAGTTAAGTTTTGCTCTAACGAA |
| Spacer 6 | TTTAACTCTCTGGATTGGAAAGTCGTTTAT |
| Spacer 7 | TTAATGTTTTCAATGGCAGGGTGTCCCACG |
| Spacer 8 | GTCGATGTGCTTCAGGACAGAATTAAAGAC |
| Spacer 9 | GGCCAATATGCGGATCTGATGCCGAAAGAA |
| Spacer 10 | TGATCATATGCGATTGTTTGAATATGTTCA |
| Spacer 11 | TTTTTTCTTGACAGTATGCCTTATACATTA |
| Spacer 12 | GGTATTACGTCGCGGCTATATCTGGGCCGT |
| Spacer 13 | CGCGTAATCGATAAGGCTTATCGCAAGAGC |
| Spacer 14 | AAAGGAAATATTCGCCGTGCTGGGCTATTA |
| Spacer 15 | ATACAGCAGGGCACAAGTTATAACCGTGGG |
| Spacer 16 | TCCGACCCCTTGAAACCGCAACCGTAAACA |
| Spacer 17 | GATTGCCCGCCCATTTTGACCATCTCGTAG |
| Spacer 18 | TCGGGGTCGTCGCATAACAATCACCATAAT |
| Spacer 19 | TCACGCGGGCGGCGGGGTAATGGACGGCGA |
| Spacer 20 | TGAAAAACTTTGAATATATCTACTATCAAG |
| Spacer 21 | TAAGTGCATATTGTCATGAAGCGTGCGGCC |
| Spacer 22 | GCGGTTTTGATCCGCAGTGCGCGCCGCCGA |
| Spacer 23 | TGTGTTGTTCTTTACAACACACCCTTCACT |
| Spacer 24 | GAAGAAAAGCATGTTTCACATGATCGGCGT |
| Spacer 25 | AGCAATTTATTATATGAGGGGATAAGTAAC |
| Spacer 26 | TATTGACATAATCATTATCATGATATTCTT |
| Spacer 27 | AAAAATAATACATCATTTGATATGTAAAGC |
| Spacer 28 | TTCGCTTTCATACATTTCATAACCAAAACA |
| Spacer 29 | TACACCGGTAAACGAATGTTGCGCCGTCGA |
| Spacer 30 | TGTCTGTAGCATCTGTGACCTTATCCACAA |
| Spacer 31 | CACGAAACGGGAACCCATTTAAACCTCGAC |
| Spacer 32 | AGCCCCACCAGAGAGGTAAACGGTGGACGT |
| Spacer 33 | TGTTATACCGTCCGGGGATGTCATAACTCT |

of strain ZYL$^T$ only consisted of 1044 amino acids, even smaller than the new genome editing tool, type-V CRISPR effector Cas12b (1108 amino acids) [21]. This strain might have big potential in developing new gene editing systems, because that smaller Cas proteins will be more efficient during the process of intracellular delivery via viral vectors [21]. Furthermore, there were ten predicted genes encoding hypothetical proteins on the upstream of *cas 9* gene (peg. 2549–peg. 2558, location: 2,760,706–2,767,137). Since the knowledge of the regulation mechanism of CRISPR-Cas9 system in their native hosts is limited [22], we supposed that these genes could be associated with the regulation of CRISPR-Cas9 system.

After BLASTp in the nr database, we found that the genes included in the CRISPR-Cas 9 system of *E. congregatus* ZYL$^T$ all had low similarities to other sequences. The Cas 9 of *E. congregatus* ZYL$^T$ (peg. 2548) was most closely related to type II CRISPR RNA-guided endonuclease Cas 9 (*Alphaproteobacteria bacterium*) with 53.5% identity. The Cas 1 (peg. 2547) was most closely related to type II CRISPR-associated endonuclease Cas1 (*Alphaproteobacteria bacterium* CG_4_9_14_3_um_filter_47_13) with 62.2% identity. And the Cas 2 (peg. 2546) was most closely related to CRISPR-associated endonuclease Cas2 (*Alphaproteobacteria bacterium* CG_4_9_14_3_um_filter_47_13) with 80.2% identity. None of the identities was higher than 85%, including the most highly conserved Cas protein, Cas1, which was generally used as a suitable marker to trace the evolutionary history of CRISPR-Cas [23]. And the high matches were all uncertain species. Hence, it was pretty hard to speculate the source of these genes with multiple sequence alignment. Among the 33 spacers, 6 could be matched in CRISPR databases [24], whereas the other 27 spacers could not be matched. In addition, the direct repeat sequence could neither be matched in the database. Although we could not determine the origin of this CRISPR-Cas 9 system, we inferred that it was a result of horizontal gene transfer (HGT), since we cannot find any CRISPR-Cas 9 system in the related species. And the short sequences might go through fast evolution process under evolutionary pressure [25] after they was transferred into the genome of *E. congregatus* ZYL$^T$.

CRISPR-Cas 9 system is a kind of prokaryotic defense system, which performs its function through editing the genome, some researches supposed it might have played a role in the evolutionary process of genomic composition [1]. In the comparison of *E. congregatus* ZYL$^T$ and four relative species, we found that it only had one gene which encodes the predicted restriction endonuclease (COG 3183), whereas other four species had more genes encode restriction endonuclease: 2 in *E. nanhaiensis* MCCC 1A06723$^T$ (COG 0732, COG 4096), 3 in *K. lacus* S3-22$^T$ (COG 1002, COG 1403, COG 3587), 4 in *K. lipolytica* M41$^T$ (COG 1403), and 2 in *S. glossodoripedis* MKT133$^T$ (COG 0732, COG 0496). We speculated that this might be an effect of keeping a CRISPR-Cas 9 system in genome. With the great power of CRISPR-Cas 9 system, strain *E. congregatus* ZYL$^T$ needs fewer varieties and quantities of restriction endonuclease to prevent infection by other mobile genetic elements (MGEs). We hope to verify this hypothesis in subsequent studies.

Based on its special systematic status, *E. congregatus* ZYL$^T$ might have the potential to help us to understand the spread and evolution of CRISPR-Cas system in marine bacterial strains and also might be a valuable material for developing new gene editing systems.

## Nucleotide Sequence Accession Number

The complete genome sequence of *E. congregatus* ZYL$^T$ and *E. nanhaiensis* MCCC 1A06723$^T$ was deposited in DDBJ/EMBJ/GenBank under the accession number CP041025 and VFIY00000000, respectively.

The genes *cas 1*, *cas 2*, and *cas 9* in *E. congregatus* ZYL$^T$ were deposited in DDBJ/EMBJ/GenBank under the accession number MN064749, MN064750, and MN064751, respectively.

**Acknowledgements** This work was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LQ19C010005, China Postdoctoral Science Foundation (Grant No. 2018M642382), and the Science & Technology Basic Resources Investigation Program of China (Grant No. 2017FY100300).

## Compliance with Ethical Standards

**Conflicts of interest** The authors declare that they have no conflict of interest.

## References

1. Hille F, Richter H, Wong SP, Bratovic M, Ressel S, Charpentier E (2018) The biology of CRISPR-Cas: backward and forward. Cell 172(6):1239–1259. https://doi.org/10.1016/j.cell.2017.11.032
2. Barrangou R, Marraffini LA (2014) CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol Cell 54(2):234–244. https://doi.org/10.1016/j.molcel.2014.03.011
3. Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, Minakhin L, Joung J, Konermann S, Severinov K, Zhang F, Koonin EV (2015) Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. Mol Cell 60(3):385–397. https://doi.org/10.1016/j.molcel.2015.10.008
4. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJM, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV (2015) An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol 13(11):722–736. https://doi.org/10.1038/nrmicro3569
5. Iino T, Ohkuma M, Kamagata Y, Amachi S (2016) *Iodidimonas muriae* gen. nov., sp nov., an aerobic iodide-oxidizing bacterium isolated from brine of a natural gas and iodine recovery facility, and proposals of Iodidimonadaceae fam. nov., Iodidimonadales ord. nov., Emcibacteraceae fam. nov and Emcibacterales ord. nov. Int J Syst Evol Microbiol 66(12):5016–5022. https://doi.org/10.1099/ijsem.0.001462
6. Zhao Z, Shen X, Chen W, Yu X-Y, Fu G-Y, Sun C, Wu M (2018) *Emcibacter congregatus* sp. nov., isolated from sediment cultured in situ. Int J Syst Evol Microbiol 68(9):2846–2850. https://doi.org/10.1099/ijsem.0.002906
7. Xu XW, Huo YY, Bai XD, Wang CS, Oren A, Li SY, Wu M (2011) *Kordiimonas lacus* sp. Nov., isolated from a ballast water tank, and emended description of the genus Kordiimonas. Int J Syst Evol Microbiol 61(2):422–426. https://doi.org/10.1099/ijs.0.018200-0
8. Wu YH, Jian SL, Meng FX, Maripatay TD, Wang CS, Xu XW (2016) *Kordiimonas lipolytica* sp. nov., isolated from seawater. Int J Syst Evol Microbiol 66(6):2198–2204. https://doi.org/10.1099/ijsem.0.001007
9. Kurahashi M, Fukunaga Y, Harayarna S, Yokota A (2008) *Sneathiella glossodoripedis* sp. nov., a marine alphaproteobacterium isolated from the nudibranch *Glossodoris cincta*, and proposal of Sneathiellales ord. nov. and Sneathiellaceae fam. nov. Int J Syst Evol Microbiol 58(3):548–552. https://doi.org/10.1099/ijs.0.65328-0
10. Liu XP, Li GY, Lai QL, Sun FQ, Du YP, Shao ZZ (2015) Emcibacter nanhaiensis gen. nov. sp. nov., isolated from sediment of the South China Sea. Anton Leeuw Int J Gen 107(4):893–900. https://doi.org/10.1007/s10482-015-0381-y
11. Stefanova P, Taseva M, Georgieva T, Gotcheva V, Angelov AS (2013) A modified CTAB method for DNA extraction from soybean and meat products. Biotechnol Biotechnol Equip 27(3):3803–3810. https://doi.org/10.5504/BBEQ.2013.0026
12. Marmur J (1961) Procedure for isolation of deoxyribonucleic acid from micro-organisms. J Mol Biol 3(2):208. https://doi.org/10.1016/S0022-2836(61)80047-8
13. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KHJ, Remington KA, Anson EL, Bolanos RA, Chou HH, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ, Lai ZW, Liang Y, Nusskern DR, Zhan M, Zhang Q, Zheng XQ, Rubin GM, Adams MD, Venter JC (2000) A whole-genome assembly of Drosophila. Science 287(5461):2196–2204. https://doi.org/10.1126/science.287.5461.2196
14. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19(6):1117–1123. https://doi.org/10.1101/gr.089532.108
15. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25(7):1043–1055. https://doi.org/10.1101/gr.186072.114
16. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O (2008) The RAST server: rapid annotations using subsystems technology. BMC Genom. https://doi.org/10.1186/1471-2164-9-75
17. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA (2003) The COG database: an updated version includes eukaryotes. BMC Bioinform. https://doi.org/10.1186/1471-2105-4-41
18. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 27(1):29–34. https://doi.org/10.1093/nar/27.1.29
19. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35(9):3100–3108. https://doi.org/10.1093/nar/gkm160
20. Canver MC, Bauer DE, Dass A, Yien YY, Chung J, Masuda T, Maeda T, Paw BH, Orkin SH (2014) Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. J Biol Chem 289(31):21312–21324. https://doi.org/10.1074/jbc.M114.564625
21. Strecker J, Jones S, Koopal B, Schmid-Burgk J, Zetsche B, Gao LY, Makarova KS, Koonin EV, Zhang F (2019) Engineering of

CRISPR-Cas12b for human genome editing. Nat Commun. https://doi.org/10.1038/s41467-018-08224-4
22. Patterson AG, Yevstigneyeva MS, Fineran PC (2017) Regulation of CRISPR-Cas adaptive immune systems. Curr Opin Microbiol 37:1–7. https://doi.org/10.1016/j.mib.2017.02.004
23. Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, Abudayyeh OO, Gootenberg JS, Makarova KS, Wolf YI, Severinov K, Zhang F, Koonin EV (2017) Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol 15(3):169–182. https://doi.org/10.1038/nrmicro.2016.184
24. Grissa I, Vergnaud G, Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinform. https://doi.org/10.1186/1471-2105-8-172
25. Westra ER, Dowling AJ, Broniewski JM, van Houte S (2016) Evolution and ecology of CRISPR. Annu Rev Ecol Evol Syst 47:307–331. https://doi.org/10.1146/annurev-ecolsys-121415-032428