

# Phylogenetic and genomic analysis reveals high genomic openness and genetic diversity of *Clostridium perfringens*

Yuqing Feng<sup>1†</sup>, Xuezheng Fan<sup>2†</sup>, Liangquan Zhu<sup>2†</sup>, Xinyue Yang<sup>1</sup>, Yan Liu<sup>1</sup>, Shiguang Gao<sup>3</sup>, Xiaolu Jin<sup>1</sup>, Dan Liu<sup>1</sup>, Jiabo Ding<sup>2</sup>, Yuming Guo<sup>1</sup> and Yongfei Hu<sup>1,\*</sup>

## Abstract

*Clostridium perfringens* is associated with a variety of diseases in both humans and animals. Recent advances in genomic sequencing make it timely to re-visit this important pathogen. Although the genome sequence of *C. perfringens* was first determined in 2002, large-scale comparative genomics with isolates of different origins is still lacking. In this study, we used whole-genome sequencing of 45 *C. perfringens* isolates with isolation time spanning an 80-year period and performed comparative analysis of 173 genomes from worldwide strains. We also conducted phylogenetic lineage analysis and introduced an openness index (OI) to evaluate the openness of bacterial genomes. We classified all these genomes into five lineages and hypothesized that the origin of *C. perfringens* dates back to ~80000 years ago. We showed that the pangenome of the 173 *C. perfringens* strains contained a total of 26954 genes, while the core genome comprised 1020 genes, accounting for about a third of the genome of each isolate. We demonstrated that *C. perfringens* had the highest OI compared with 51 other bacterial species. Intact prophage sequences were found in nearly 70.0% of *C. perfringens* genomes, while CRISPR sequences were found only in ~40.0%. Plasmids were prevalent in *C. perfringens* isolates, and half of the virulence genes and antibiotic resistance genes (ARGs) identified in all the isolates could be found in plasmids. ARG-sharing network analysis showed that *C. perfringens* shared its 11 ARGs with 55 different bacterial species, and a high frequency of ARG transfer may have occurred between *C. perfringens* and species in the genera *Streptococcus* and *Staphylococcus*. Correlation analysis showed that the ARG number in *C. perfringens* strains increased with time, while the virulence gene number was relative stable. Our results, taken together with previous studies, revealed the high genome openness and genetic diversity of *C. perfringens* and provide a comprehensive view of the phylogeny, genomic features, virulence gene and ARG profiles of worldwide strains.

## DATA SUMMARY

This study generated sequencing data for 45 *C. perfringens* isolates and used published data for 128 isolates. All sequence data have been upload to the National Center for Biotechnology Information under accession number PRJNA597712. Isolate metadata and accession numbers can be found in Table S1 (available in the online version of this article).

## INTRODUCTION

*Clostridium perfringens* is a Gram-positive pathogen which inhabits various natural environments as well as the gastrointestinal tracts of humans and animals [1]. This spore-forming bacterium is known to cause diverse human and animal systemic and intestinal diseases. In humans, *C. perfringens* causes diseases including gas gangrene [2], food poisoning [3], gastrointestinal disorders [4], and liver and kidney damage [5, 6]. As a commensal,

Received 09 April 2020; Accepted 08 September 2020; Published 25 September 2020

**Author affiliations:** <sup>1</sup>State Key Laboratory of Animal Nutrition, College of Animal Science and Technology, China Agricultural University, Beijing 100193, PR China; <sup>2</sup>China Institute of Veterinary Drug Control, Beijing 100081, PR China; <sup>3</sup>Dandong Customs, Liaoning 118000, PR China.

\***Correspondence:** Yongfei Hu, huyongfei@cau.edu.cn

**Keywords:** Antibiotic Resistance Gene; *Clostridium perfringens*; Openness Index; Phylogenetic Tree; Toxins.

**Abbreviations:** AGR, antimicrobial resistance gene; ANI, average Nucleotide Identity; BHI, brain heart infusion; CARD, Comprehensive Antibiotic Resistance Database; CASC, CASC Ain't Simply CRT; CRISPR, clustered regularly interspaced short palindromic repeats; ESS, effective sample size; FTG, fluid thioglycollate; GTR, general time-reversible; NCBI, National Center for Biotechnology Information; NE, necrotic enteritis; OI, openness index; PCA, principal-component analysis; PHASTER, PHAge Search Tool Enhanced Release; TSC, tryptose-sulfite-cycloserine; VF, virulence factor; VFDB, Virulence Factor Database.

Repositories: the *C. perfringens* sequencing project has been registered in BioProject database at NCBI under accession number PRJNA597712.

†These authors contributed equally to this work

**Data statement:** Six supplementary figures and 19 supplementary tables are available with the online version of this article.

000441 © 2020 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution NonCommercial License.

*C. perfringens* has potentially co-evolved with humans for thousands of years, as evidenced by the identification of this bacterium in the gut of a >5000-year-old mummified Neolithic ‘Tyrolean Iceman’ [7]. In animals, especially in pigs and poultry, *C. perfringens* is the causative agent of gangrenous dermatitis [8], enterotoxaemia [9] and necrotic enteritis (NE) [10].

The pathogenicity of *C. perfringens* depends largely on the various types of extracellular toxins and enzymes they produce [11]. The toxin-based *C. perfringens* typing method indicated that different toxinotype strains had different host preference and were responsible for specific diseases. For example, type B strains are predominantly isolated from sheep suffering dysentery; type F (previously CPE-positive type A) strains are associated with human food poisoning; and type G (previously NetB-positive type A) strains comprise isolates causing necrotic enteritis in chickens [12]. Genes of these typing toxins have different genetic locations with the  $\alpha$ -toxin encoding genes carried by chromosomes while the other typing toxin genes are carried by mobile genetic elements, commonly large conjugative plasmids [13].

The first genome of *C. perfringens* was sequenced in 2002, highlighting the genetic features of this bacterium with pronounced low G+C content, diverse toxin and degradative enzyme-encoding genes, the presence of anaerobic fermentation pathways and the lack of enzymes for amino acid biosynthesis [14]. Since then, several comparative genomics analyses for *C. perfringens* of either human or animal origins have been performed [15–19]. By analysing 56 strains, Kiu *et al.* [20] showed that *C. perfringens* has a large pangenome with only 12.6% core genes present in the pangenome. The study also indicated that *C. perfringens* is relatively lacking in the CRISPR system compared with other bacteria [21], which may be a reason for its high genomic variation caused by horizontal gene transfer. Recently, a comparative genomic study that included 88 chicken-associated *C. perfringens* genomes revealed that virulence genes *netB*, *pfoA*, *cpb2*, *tpeL* and *cna* variants are closely associated with chicken NE-linked strains [22]. These studies provide useful information on the diversity and evolution of *C. perfringens*. However, large-scale genomic analysis concerning isolates from a range of hosts and conditions is still needed to understand the genomic features, toxinotype and antimicrobial resistance gene profile of *C. perfringens* [12].

In this study, we performed whole-genome sequencing of 45 *C. perfringens* isolates preserved in our culture collection centre. These isolates, collected over a period of ~80 years, were originally isolated from at least 14 different animal hosts with the majority associated with chicken NE. We then downloaded 128 additional publicly available genomes in GenBank (as of December 2018), including isolates of human gas gangrene and food poisoning, and performed comparative analysis of the 173 *C. perfringens* genomes.

### Impact Statement

*Clostridium perfringens* is associated with a variety of human and animal enteric diseases, but large-scale genomic understating of *C. perfringens* strains of different origin is still lacking. In this study, we sequenced 45 *C. perfringens* isolates with isolation times spanning an 80-year period and performed comparative analysis of 173 genomes from different sources. We showed that *C. perfringens* is an ancient bacterium, the origin of which dates back ~80000 years ago. We introduced an openness index (OI) to evaluate the openness of bacterial genomes and demonstrated that *C. perfringens* had the highest OI compared with other randomly selected bacterial species. We showed that the various gene contents in phages and plasmids contribute to the genetic diversity of this bacterium. We also revealed that the *C. perfringens* typing toxin genes are conserved among strains, while the number of antibiotic resistance genes in isolates increased over time, which may suggest that the bacterium has been subjected to antibiotic selection pressures.

## METHODS

### Isolate collection and metadata

By considering isolation time and host range, 45 *C. perfringens* isolates conserved in our culture collection centre were selected for genome sequencing (Table S1). All these isolates were of animal origin and of long isolation age; the oldest strain, CP-40, was originally isolated from the former Soviet Union in 1937, while the most recent, CP-45, was isolated from Hebei province, China, in 2018. Before DNA extraction, the freeze-dried bacteria were inoculated in fluid thioglycollate medium broth (FTG; Beijing Land Bridge Technology) and then plated on tryptose-sulfite-cycloserine agar (TSC; Beijing Land Bridge Technology). A single colony was picked and incubated in brain heart infusion broth (BHI; Beijing Land Bridge Technology) at 37°C for 24 h under anaerobic conditions.

For comparative genomics, all available *C. perfringens* genomes (as of December 2018) deposited in the National Center for Biotechnology Information (NCBI) ( $n=128$ ) were retrieved. The year of isolation and the geographical origin of each of these isolates were taken from the published research papers or from the information provided by the submitters (Table S1).

### DNA sequencing, assembly and annotation

For whole genome sequencing, the genomic DNA of 45 *C. perfringens* isolates was extracted by using a QIAamp DNA Mini Kit (Qiagen). We performed paired-end genome sequencing on the Illumina HiSeq 2500 with an genome coverage of 300× [coverage was calculated from the length of the genome (G), the number of reads (N) and the read

length (L) as  $N \times L / G$ . Illumina HiSeq-generated sequencing reads were clipped for adaptors with Trimmomatic [23]. The sequence reads were assembled using Unicycler (v0.4.7) with default parameters [24], which employs SPAdes (v3.13.0) for genome assembly [25]. The genomes were annotated using Prokka (v1.13.3) [26].

### Phylogenetic analysis and lineage assignment

A maximum-likelihood phylogenetic tree was generated based on the core genomes, using RAxML-NG (v0.9.0) with 500 bootstrap replicates [27]. Bayesian phylogenetic analysis was performed on the core genome using BEAST 2 (v2.5) [28] to infer the timescale of *C. perfringens* evolution. All unknown dates were set to year 0 [29]. Evolutionary rates and tree topologies were analysed using the general time-reversible (GTR) substitution models with gamma substitution models with gamma distributed among-site rate variation with four rate categories. We tested a strict molecular clock, which assumes the same evolutionary rate for all branches in the tree. A Yule tree prior was selected as the tree prior, which assumes a constant lineage birth rate for each branch in the tree. Convergence was assessed using Tracer (v1.7.1) [30], ensuring all relevant parameters reached an effective sample size (ESS) of over 200. A run of  $1 \times 10^8$  generations was sampled every 10000 generations to ensure independent convergence of the chains. Lineage assignment for the 173 strains was made using RhierBAPS based on the core genome [31] with default parameters. For verification of lineage assignment, we applied three additional methods to confirm the current definition: principal-component analysis (PCA) was carried out with Jalview [32]; pairwise SNP distances (core genome SNPs) were calculated using the ape package in R (v3.5.3) [33]; pairwise fixation index ( $F_{ST}$  values, estimation of population separation) were calculated using Arlequin (v3.5) [34].

### Openness index definition and comparison

We introduced an openness index (OI) to evaluate the openness of a bacterial strain, which was calculated according to equation 1:

$$\text{Openness index (OI)} = \frac{n_{\text{strain}} - n_{\text{core}}}{n_{\text{strain}}} \quad (1)$$

Here,  $n_{\text{strain}}$  is the total number of genes in a specific strain;  $n_{\text{core}}$  is the number of core genes for the corresponding bacterial species; and  $n_{\text{strain}} - n_{\text{core}}$  represents the number of accessory genes in that strain. Therefore, OI reflects the ratio of accessory genes in a specific strain. From the equation, we can see that the  $n_{\text{core}}$  value determines the value of OI for a single strain, but the  $n_{\text{core}}$  in a bacterial species largely depends on the strain number used for building the core gene catalogue. So, for comparison of the OI among different bacterial species, it is essential to make sure the core gene analysis is based on using the same number of genomes in different species. To achieve this, we first selected all the species having a sufficient genome number (>200) in NCBI. Then, to ensure a high-quality genome was included in the following study, we randomly (using the 'shuf' command in Linux) selected 100 genomes with N50 over 50 kb and an average nucleotide identity (ANI)

greater than 0.95 [35] calculated by pyani (v0.2.3) [36] among genomes for each species. Finally, 51 bacterial species (5100 genomes in total) met our filter demands and were compared with *C. perfringens* (100 genomes selected from 173 isolates using the same genome selection criteria as for other species). The core genome and pangenome prediction was carried out using the Roary software (v3.11.2) [37], and the core genome was defined by including both the hard-core and soft-core genes, i.e. genes present in more than 95.0% of strains [38]. For each species, the rRNA operon copy number was retrieved from *rrnDB* (v5.5) [39]. The core gene and accessory gene functional annotations of *C. perfringens* were carried out using eggNOG (v4.5) (evolutional genealogy of genes: Non-supervised Orthologous Groups) platform [40].

### In silico analysis of virulence factors, ARGs, prophages, CRISPR and plasmids

Virulence genes were identified by BLAST (The Basic Local Alignment Search Tool, v2.9.0+) [41] against the latest version of the Virulence Factor Database (VFDB) [42] with a cutoff value of  $\geq 80\%$  identity and  $\geq 50\%$  coverage. We added the nucleotide sequences of 11 genes to the core dataset of VFDB, as these genes are not included in the core dataset of VFDB but are important for *C. perfringens* (Table S2). To test the conservation of the toxins used for strain toxinotyping [12], we retrieved the *plc*, *cpb*, *etx*, *iap*, *ibp*, *cpe* and *netB* gene sequences from the 173 genomes, and expanded the dataset by adding the nucleotide sequences of these seven toxins from the UniProt database. Then, we curated the toxin nucleotide sequences dataset by CD-HIT (v4.6) [43] at 100.0% identity to remove redundant sequences. A total of 158, 8, 21, 5, 6, 8 and 6 nucleotide sequences for these seven genes, respectively, were compared. KaKs\_Calculator (v2.0) [44] was applied to calculate non-synonymous ( $K_a$ ) and synonymous ( $K_s$ ) substitution rates of nucleotide sequences for all the seven toxins.

Antibiotic resistance gene (ARG) analysis was performed by using the Comprehensive Antibiotic Resistance Database (CARD, v4.2.2), which is a manually curated resource containing high-quality reference data on the molecular basis of antibiotic resistance [45]. We also added the nucleotide sequences of four genes (*bcrA*, *bcrB*, *bcrD* and *bcrR*) encoding resistance to bacitracin [46] to the database (Table S3). ARG-sharing network analysis was performed according to our previous procedure, and if an ARG was shared between two bacterial species with  $\geq 99.0\%$  nucleotide identity, the ARG was considered to have the potential to transfer between species [47]. For this analysis, the 18 ARGs (34 sequences) identified in *C. perfringens* were analysed by BLAST against 11395 complete genomes downloaded from NCBI on 26 September 2018.

Prophage predictions were carried out using the PHAge Search Tool Enhanced Release (PHASTER) web server [48] to explore the 'intact', 'questionable' and 'incomplete' prophages. CRISPR (clustered regularly interspaced short Palindromic Repeats) predictions were performed with CASC (CASC Ain't Simply CRT, v2.6) [49]. To explore the source of the CRISPR

spacer sequence, we performed a BLASTn analysis against the NCBI Virus database ([ncbi.nlm.nih.gov/labs/virus/vssi/#/](https://ncbi.nlm.nih.gov/labs/virus/vssi/#/)) with a cutoff value of  $\geq 90\%$  identity and  $\geq 90\%$  coverage.

Complete plasmid sequences were retrieved from the complete genomes, and partial plasmid sequences from the draft genomes were predicted using PlasFlow (v1.1) [50] and ABRicate (v1.0.1, <https://github.com/tseemann/abricate>). For plasmid identification, the predicted plasmid sequences were analysed by BLAST against a custom plasmid database which was constructed based on a previous study [51], using best-hit approach with query coverage threshold  $\geq 70\%$  and nucleotide identity  $\geq 90\%$ . Virulence and antibiotic resistance genes carried by the plasmid sequences were analysed using the same procedure and cutoff for the whole genome analysis.

### Statistical analysis

Spearman's rank correlation coefficient was applied to test the correlation between two variables. A Wilcoxon rank-sum test was used to identify the difference. Fisher's exact test was used to determine if there are non-random associations between two categorical variables. All statistical analyses were performed using R software.

## RESULTS

### Strain information and sequence statistics

In the present study, we sequenced 45 *C. perfringens* and downloaded additional 128 genomes from NCBI. These strains, covering all the seven toxinotypes, and were from at least 14 species of animals as well as from humans and the environment. In addition, the strains we sequenced were isolated earlier (median year of isolation: 1985) than the 128 genomes downloaded from the NCBI (median year of isolation: 2008). Sequencing statistics can be found in Table S4. The estimated size of the genomes, as predicted from total contig size, ranged from 2696148 to 4127102 bp, with an average length of 3382459 bp.

### Phylogenetic analysis reveals five lineages of *C. perfringens*

To investigate the phylogenetic relationship of *C. perfringens* strains, we performed core-genome phylogenetic analysis of 173 genomes (including 45 genomes we sequenced in this study and 128 genomes downloaded from NCBI) and estimated their divergence dates using the BEAST 2 software package [28]. The 173 isolates were from at least 19 different hosts and 19 geographical locations on five continents (Table S1). The majority of the strains were from animal sources ( $n=133$ ), followed by human sources ( $n=21$ ) and environmental sources ( $n=15$ ).

To analyse the phylogeny of the 173 isolates, we compared the topology of two trees: a maximum-likelihood phylogenetic tree and a Bayesian-based tree. These two trees shared almost identical topologies, except regarding strains An68 and An185 (Figs S1 and S2). According to divergence dates of the phylogenetic tree from BEAST 2 and the assignment

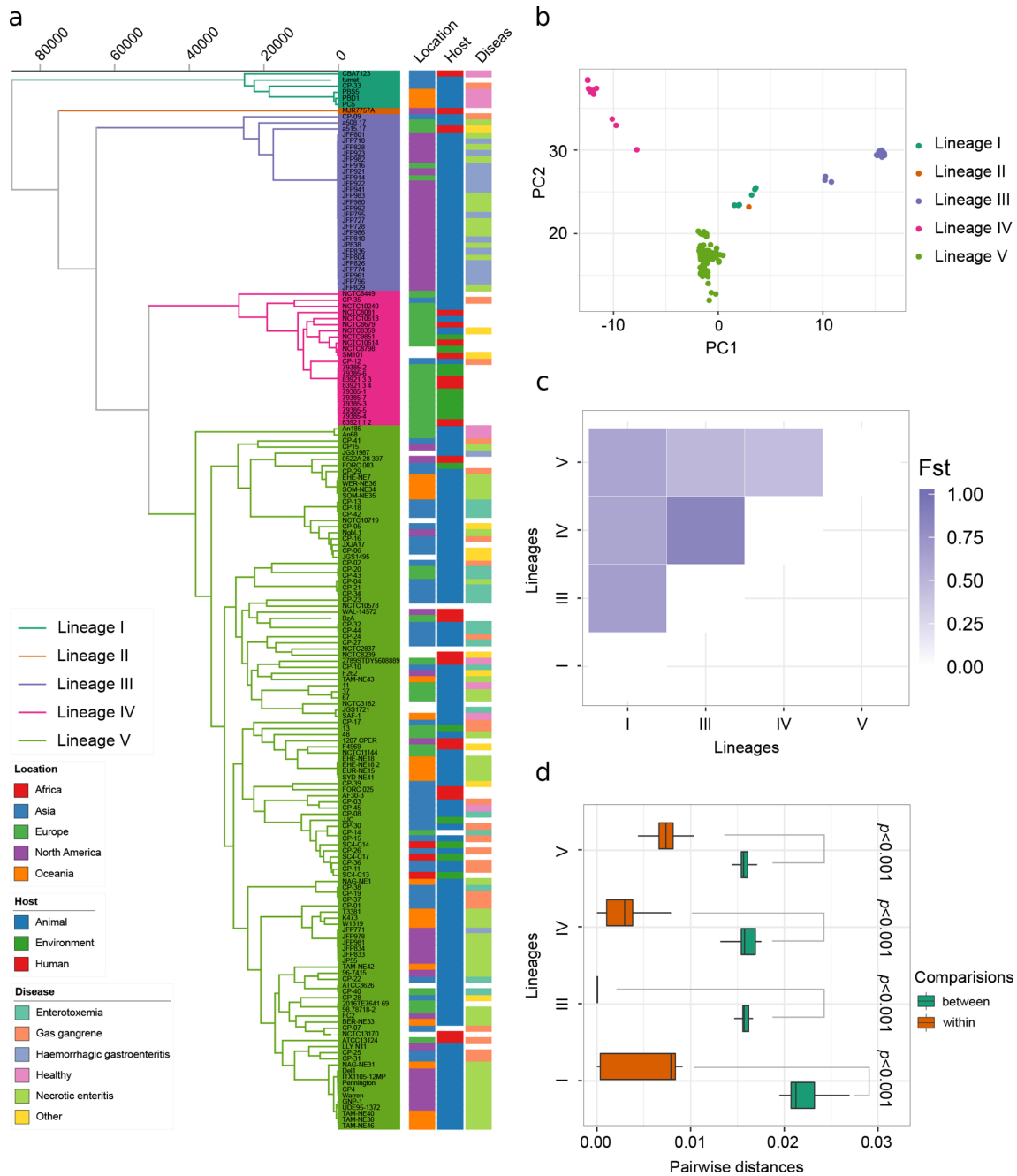
from hierBAPS, we assigned all the isolates into five major lineages (Lineage I to Lineage V) and two sub-lineages within Lineage V (Fig. 1a, Table S5). We combined the two sub-lineages within Lineage V into a major lineage considering the complexity and the divergence dates inferred from the Bayesian-based tree (Table S5). PCA showed clear separation of the four lineages (Lineages I, III, IV and V). As there is only one strain in Lineage II, which is blurred with Lineage I, no clear boundary can be found between Lineage I and Lineage II (Fig. 1b). All  $F_{ST}$  values between the lineages were well above 0.25, indicating that these lineages are clearly separated (Fig. 1c). The pairwise genetic distance between pairs of isolates within the lineages was significantly lower than that between lineages (Wilcoxon rank-sum test,  $P < 0.001$ , Fig. 1d). Taken together, these results suggest that the current lineage definition is reasonable and reflects the true genetic relatedness of the *C. perfringens* strains.

Lineage I encompassed six strains, representing the oldest *C. perfringens* clade as shown in the phylogenetic tree. Coincidentally, the *C. perfringens* strain Tumat, isolated from the mummified remains of an ancient puppy found in Siberian permafrost, was placed in this lineage. Three isolates, CBA7123, Tumat and CP-33 in this lineage were from Asia, and the other three, PBS5, PBD1 and PC5, were from Oceania. Lineage II contained only one strain, MJR7757A, which was isolated from a human in the USA. Strains from Lineages III and IV showed the clearest eco-epidemiological relationships. Lineage III included 29 strains, the majority of which were from North America and Europe and were genetically closely related to each other. Lineage IV included 22 strains, and interestingly, except for CP-12 and CP-35, the remaining 20 strains were isolated from European countries, and may have expanded from the same clonal population. Lineage V was the largest one, comprising 115 strains of various geographical and host origins, which can be regarded as the modern *C. perfringens* lineage (Table S6). Eco-epidemiological relationships of some strains can be found in the sub-clades in Lineage V. For example, strains JFP771, JFP978, JFP981, JFP834, JFP833 and JP55 isolated from dogs or horses in 2000–2011 in Canada and the USA were probably from the same clonal population.

Based on the Bayesian phylogenetic inference, we tentatively hypothesized that the *C. perfringens* species originally emerged in Asia and Oceania ~80000 years ago, and subsequently spread to America and Europe ~60000 years ago. Thereafter, *C. perfringens* spread back to Asia ~50000 years ago, and it became worldwide in the last 40000 years (Fig. S2).

### *C. perfringens* has an extremely open genome

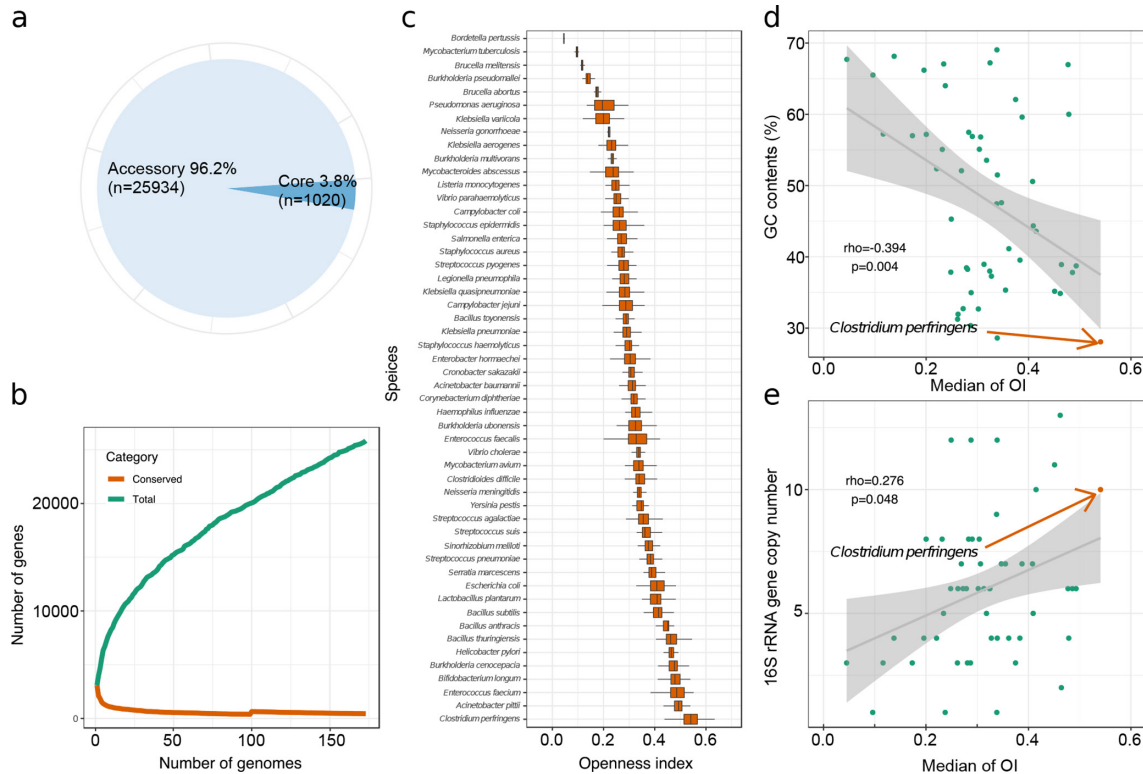
Previous result suggested that *C. perfringens* has a high pangenome variation, suggesting a high plasticity of the *C. perfringens* genome [20]; however, the plasticity/openness of *C. perfringens* was not compared with other bacterial species [52]. We showed that the pangenome of the 173 *C. perfringens* strains contained a total of 26954 genes, while the core genome (1020 genes, 3.8%) makes up about one-third



**Fig. 1.** Phylogenetic assignment of *C. perfringens* isolates and validation. (a) Bayesian-dated phylogeny of *C. perfringens*. We classified the 173 *C. perfringens* genomes into five lineages according to their phylogenetic relationships. The five lineages are labelled with different colours. Information about location, host and diseases are shown to the right of the phylogenetic tree. (b) PCA of core genome sequences in 173 strains. (c) Pairwise  $F_{ST}$  values between the four lineages. (d) Pairwise SNP distances within and between isolates of the four lineages.

of the genome of each isolate and the rest is of variable gene content (Fig. 2a, b). The accessory genome harboured ten-fold more genes in 15 eggNOG function categories than the core genome (Fig. S3a). Genes associated with ‘[V] Defence mechanisms’ and ‘[N] Cell motility’ were clearly more enriched in the accessory genome than in the core genome (603 vs six genes and 241 vs four genes, respectively). We then compared

the accessory genome and core genome sizes of *C. perfringens* with 51 other species (Table S7) and introduced an OI to evaluate the genome openness (see Materials and Methods). We showed that *C. perfringens* displayed the highest value of OI among the studied bacterial species, ranging from 0.43 to 0.66 (median 0.54), suggesting the accessory genes accounted for a large proportion of each *C. perfringens* genome (Fig. 2a–c).



**Fig. 2.** Openness of the *C. perfringens* genome. (a) Proportions of accessory genes and core genes in the pangenome. (b) Accumulation of the pan genes and core genes with increasing genome number. (c) Comparison of the *C. perfringens* genome OI with 51 other bacterial species. (d) Spearman's correlation between bacterial genome OI and GC contents of each species. (e) Spearman's correlation between median values of the bacterial genome OI and median values of 16S rRNA gene copy number.

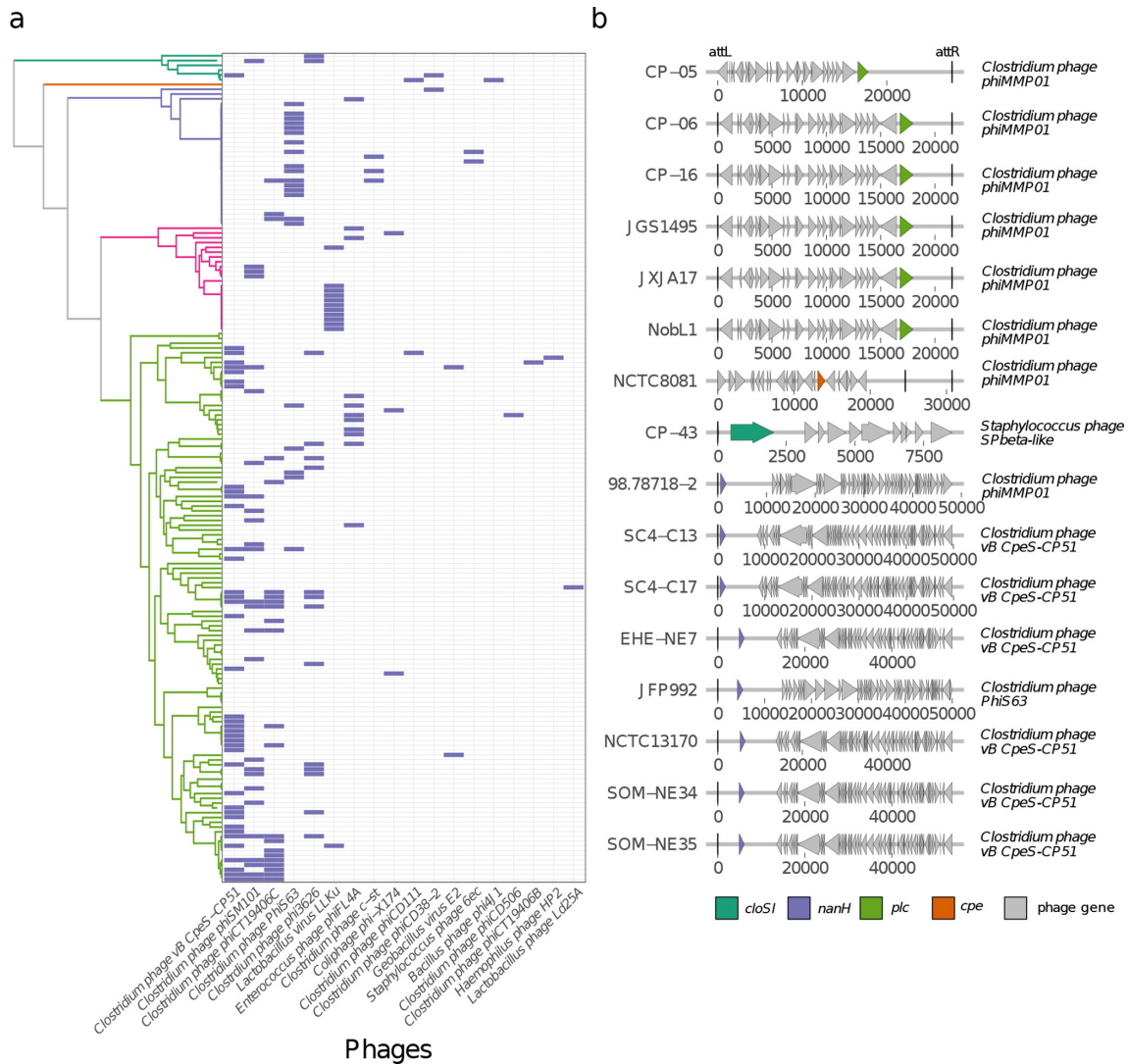
The human pathogens *Bordetella pertussis* and *Mycobacterium tuberculosis* showed the lowest OI; these two bacteria have relatively conserved genomes which mainly involve small genetic changes, such as SNPs, but rarely acquired new genes [53, 54].

To understand factors associated with the openness of the *C. perfringens* genome, we explored the potential connection of OI with bacterial genome GC content and 16S rRNA gene copy number, which positively correlates with bacterial reproductive rate [55]. Interestingly, using the same dataset above, we found that the OI values were negatively correlated with the bacterial genome GC content (Spearman's correlation,  $P=0.004$ ) (Fig. 2d), and were positively with the median value of 16S rRNA gene copy number ( $P=0.048$ ) (Fig. 2e). Collectively, these results indicated that *C. perfringens* has a highly open genome with diverse functions encoded in its accessory genes; the low GC content and rapid growth speed (high 16S rRNA gene copy number) may be closely related to its high genome plasticity.

### The interaction between prophages/phages and *C. perfringens* contributes to genetic variations

We next investigated the prevalence of prophages and CRISPR systems in the *C. perfringens* genomes. A total of 527 phages including 'incomplete', 'questionable' and 'intact'

were found in the 173 genomes, and the average length of the phage sequences was ~26.7 kb (Fig. S4a, b). Phage sequences were found in each *C. perfringens* isolate; as many as eight phage fragments were detected in a single genome (Fig. S4c). Sequences belonging to *Clostridium* phage *vB CpeS-CP51* were the most abundant in *C. perfringens* (Fig. S4d). A total of 18 intact prophages were found among the 173 genomes, and nearly 70.0% of isolates (119 genomes) contained at least one intact prophage (Table S8). The most common prophage in *C. perfringens* was *Clostridium* phage *vB CpeS-CP51* (40/173, 23.1%), followed by *Clostridium* phage *phiSM101* (26/173, 15.0%) and *Clostridium* *phiCT19406C* (22/173, 12.7%). The different *C. perfringens* lineages have different prophage profiles (Fig. 3a, Table S9): Isolates from Lineage III had a higher prevalence of *Clostridium* phage *PhiS63* (55.2%, 16/29), while Lineage IV had *Lactobacillus* phage *LLKu* (50.0%, 11/22) and Lineage V had *Clostridium* phage *vB CpeS-CP51* (33.9%, 39/115). The prophages in *C. perfringens* contained genes encoding various functions that were classified into 19 of the 25 eggNOG functional categories; these gene functions accounted for 14.1% of the total gene functions in the *C. perfringens* accessory genome (Fig. S3b). Interestingly, four toxin genes, *cloSI*, *cpe*, *nanH* and *plc*, were found to be carried by at least four intact prophages in 16 isolates (Fig. 3b): the *cloSI* and *cpe* genes each in one



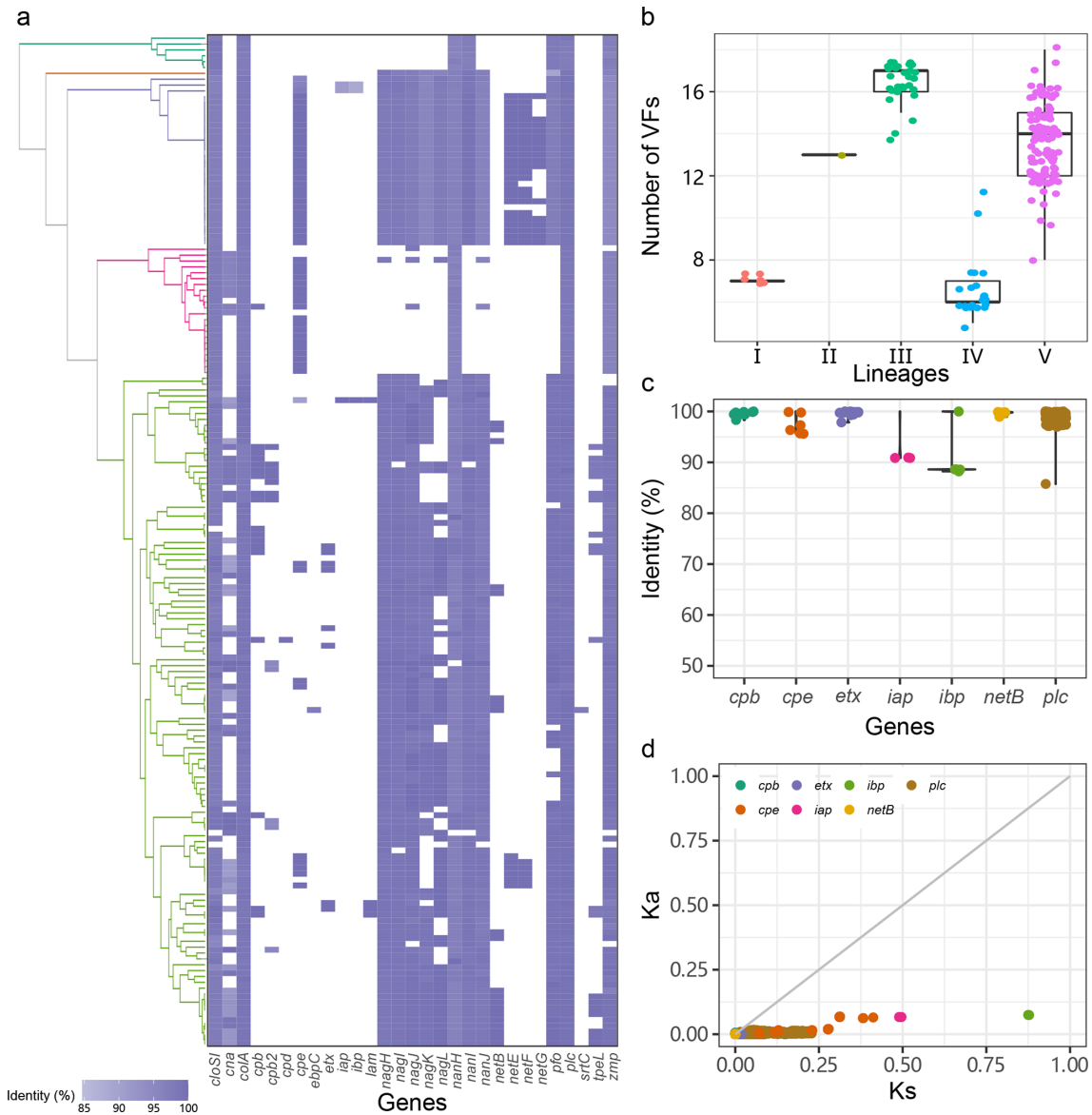
**Fig. 3.** Prophages in the *C. perfringens* genome and the toxins carried by the prophages. (a) Heatmap visualizing the prophage profile of 173 *C. perfringens* genomes. Cells in purple and white indicate the presence and absence of prophages, respectively. (b) Toxin genes carried by prophages in 16 isolates.

isolate, the *plc* gene in six isolates, and the *nanH* gene in eight isolates.

Further analysis showed that 73 *C. perfringens* genomes (Fig. S5), 42.2% of the total, harboured CRISPR systems. CRISPR systems were less common in strains from animals (35.3%) compared with those from humans and the environment (57.1 and 73.3%, respectively) (Table S10), and strains from Lineage III showed the lowest prevalence of CRISPR systems (3.4%, 1/29) (Table S11). The spacer sequences in these CRISPR systems of 43 isolates matched 12 specific phage genomes. The most frequently encountered phage genome sequence in the CRISPR spacers was *Clostridium* phage 39-O (36 isolates), followed by *Clostridium* phage *phi*CP130 (30 isolates), *Clostridium* phage *Phi*S63 (15 isolates) and *Clostridium* phage *phi*SM101 (15 isolates), suggesting

that these four phages are more common invaders of *C. perfringens*. Among the 43 isolates with CRISPR spacers matched to specific phage sequences, 40 isolates contained no corresponding intact phages in their genomes, suggesting that the CRISPR systems in *C. perfringens* worked efficiently to protect against phage infection. The exception was the three isolates CP-40, CP-43 and ATCC 3626 from sheep, which had their CRISPR spacers matched to *Clostridium* phage *phi*SM101 genomes, although the complete *Clostridium* phage *phi*SM101 sequences were still found in their genomes.

Taken together, these results suggested that the frequent interaction between prophages/phages and *C. perfringens* is an important factor impacting the genetic variations within isolates.



**Fig. 4.** Virulence genes in the *C. perfringens* genome. (a) Heatmap visualizing the virulence gene profile of 173 *C. perfringens* genomes. Cells in purple and white indicate the presence and absence of genes, respectively. (b) The virulence gene number in each lineage. (c) The identity of the nucleotide sequences of the seven toxin genes. (d)  $K_a/K_s$  values of all the seven toxins.

### Virulence gene distribution and conservation of the typing toxin genes

To provide a comprehensive view on *C. perfringens* virulence, we first analysed the virulence gene profile of the 173 isolates (Fig. 4a). A total of 29 genes displaying high nucleotide identity with known virulence genes were found in all these strains. In general, all the strains were positive for the genes encoding the  $\alpha$ -toxin (*plc*) and  $\kappa$ -toxin (*colA*) [56], and more than 50.0% of the isolates harboured genes encoding  $\alpha$ -clostripain (*cloSI*),  $\beta$ -toxin (*cpb2*),  $\theta$ -toxin (*pfo*), hyaluronidase (*nagH*, *nagI*, *nagJ*, *nagK*, *nagL*), *Zmp* and sialidase (*nanH*, *nanI* and *nanJ*). Sixteen isolates, mainly from dogs and horses, possessing as

many as 18 virulence genes, encoded the highest number of virulence genes among the 173 isolates.

We then calculated the virulence gene number in each lineage (Fig. 4b). We showed that Lineage III harboured more virulence genes than the other lineages ( $P < 0.05$ , Wilcoxon rank-sum test), while Lineage IV contained the least number of virulence genes ( $P < 0.05$ , Wilcoxon rank-sum test). Isolates from Lineage IV, mainly from Europe (Fig. 1), were lacking  $\mu$ -toxin and sialidases, which may decrease bacterial attachment to host cells and their spread into deeper tissues [57]. Lineages III and IV showed a higher prevalence of the *cpe* gene, which is responsible for human food poisoning [58].



However, the *cpe* gene was not host-associated, as the *cpe*-positive strains in Lineages III and IV were not only of human, but also of animal and environmental origins. Interestingly, the *netB* gene was found in Lineage V strains, and all *netB*-positive isolates were associated with chicken NE (Tables S1 and S12), suggesting this gene's essential role in causing chicken NE [59]. The rarely reported *ebpC* and *srtC* genes were solely contained in a Lineage V strain (SYD-NE41) isolated from chickens having chicken NE in Australia, which were probably acquired from *Enterococcus faecalis* by horizontal gene transfer [42].

According to a recent update, *C. perfringens* is classified into seven toxinotypes (types A to G) based on the combination of seven major toxins they produce [12]. We investigated the nucleotide sequence conservation of these typing toxins (Table S13). The results showed that the nucleotide sequences of the *cpb*, *cpe*, *etx* and *netB* genes all displayed an identity over 95.0% (Fig. 4c). The *iap* and *ibp* genes were relatively less conserved, with an identity ~90.0%. The overwhelming majority of *plc* genes showed more than 97.0% sequence identity [60] with only one exception (85.8%) which was from a swan *C. perfringens* isolate [61]. In addition, we showed that the  $K_a/K_s$  values of all the seven toxins were much lower than 1 (Fig. 4d), indicating purifying selection during evolution. These results suggested that the typing toxin genes in *C. perfringens* are highly conserved, which warrants using their genomic sequences for strain toxinotyping.

### ARG profile of *C. perfringens*

To reveal the ARG distribution in *C. perfringens* genomes, we searched the 173 genomes for known ARGs collected in the CARD database and genes encoding resistance to bacitracin. Unlike its proficiency in producing toxins and extracellular enzymes, *C. perfringens* possessed fewer ARGs, with only 18 ARGs found in all these genomes (Fig. 5a). The *mprF* gene encoding multiple peptide resistance factor F was the most prevalent ARG in *C. perfringens* genomes (169/173, 97.7%), probably indicating that it is an intrinsic ARG in this bacterium. The tetracycline resistance genes *tetA*(P) and *tetB*(P) also showed high prevalence, present in 65.3 and 35.2% of the total strains, respectively. Macrolide resistance genes, bacitracin resistance genes and aminoglycoside resistance genes were less frequently found. Interestingly, strains from the modern lineage, Lineage V, harboured more ARGs than strains from other lineages: all the 18 ARGs can be found in Lineage V while only three ARGs were contained in the other four lineages (Fig. 5b). Geographically, strains from Asia and Africa contained more ARGs than those from Europe, North America and Oceania (Fig. 5c).

We then determined the potential recent ARG exchange between *C. perfringens* and other bacteria (3069 species in total, 11395 genomes) using an ARG-sharing network analysis [47] (Fig. 5d, Table S14). In general, *C. perfringens* shared ARGs (99.0% nucleotide identity) with a total of 55 different bacterial species, and 16 and six species in the genera *Streptococcus* and *Staphylococcus*, respectively, had at

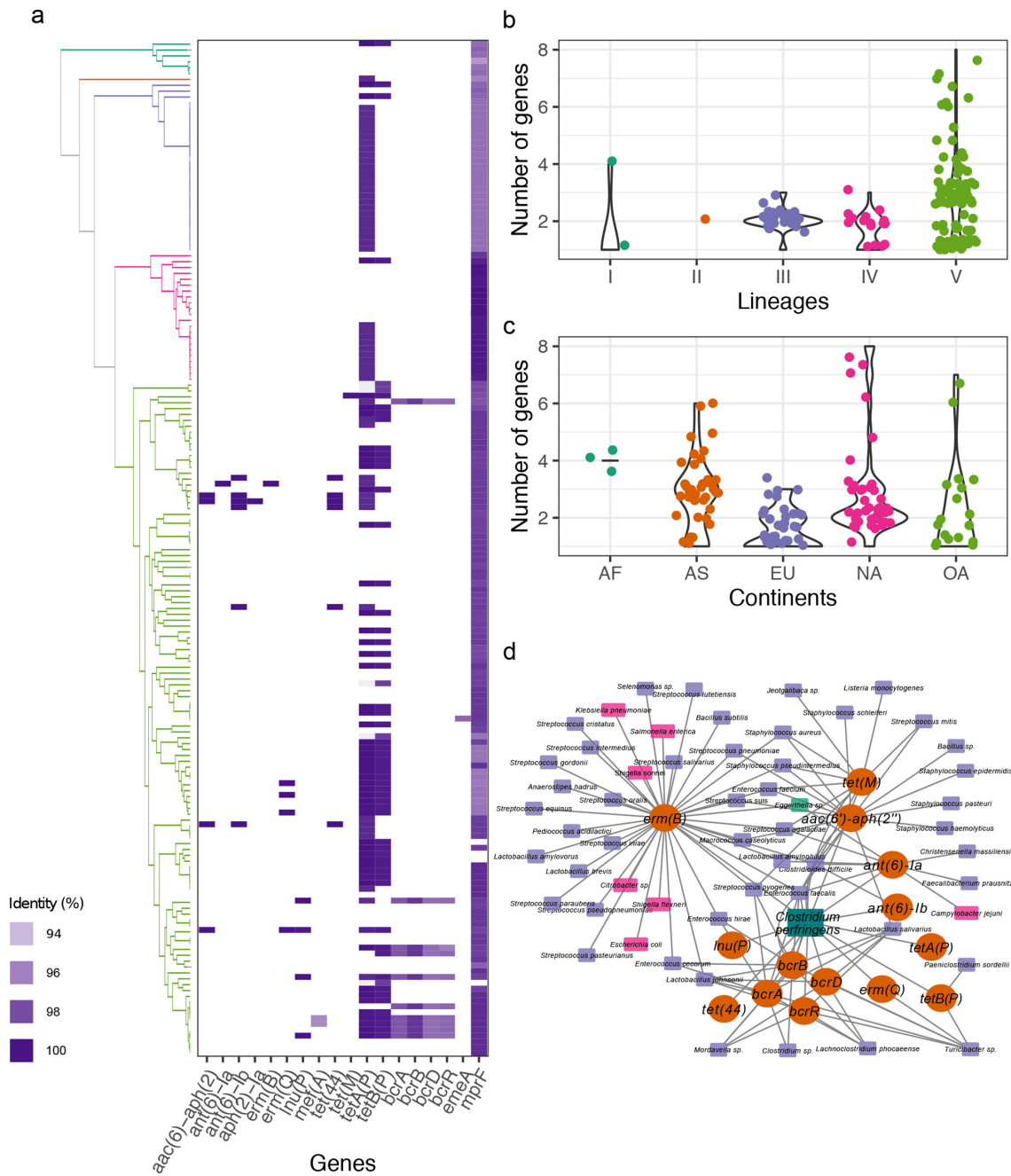
least one ARG shared with *C. perfringens* (Table S15). The *erm*(B) gene was the most frequently transferred ARG among *C. perfringens* and other bacteria. Highly similar genes to *tetA*(P), *tet*(44) and *lnu*(P) were not found in other bacterial species, which means that potential horizontal gene transfer events for these three ARGs have not been identified. These results indicate that although *C. perfringens* isolates contained a limited number of ARGs, it can readily exchange ARGs with other bacterial species, even phylogenetically distant species.

### *In silico* analysis of virulence genes and ARGs carried by *C. perfringens* plasmids

Plasmids are important for the biology of *C. perfringens*, and interspecies plasmid transfer may enhance the pathogenicity and the tolerance to antibiotics of the recipients [13]. We then explored plasmids harboured in the 173 isolates and found that more than 90.0% (157 out of 173) of the genomes contained plasmid sequences (Table S16). Annotation of these plasmid sequences showed that 14 out of 29 virulence genes and nine out of 18 ARGs identified in all the isolates could be found in the plasmids. Additionally, 81.0, 20.0, 93.5 and 93.1% of isolates harbouring virulence genes *cpe*, *netB*, *netE* and *netF*, respectively, were found located in plasmids. For ARGs, 23.9 and 37.7% of isolates containing *tetA*(P) and *tetB*(P), respectively, were located in plasmids (Fig. 6, Table S17). By using the custom plasmid database (Table S18), we identified 34 different known plasmids presented in 26 isolates (Table S16 and Fig. S6). Among these plasmids, pCP15\_1 ( $n=5$ ) was the most common, followed by pCP15\_3 ( $n=3$ ) (Table S19). Only two ARGs [*tetA*(P) and *tetB*(P)] and seven virulence genes (*cna*, *cpb2*, *cpe*, *etx*, *netB*, *netE* and *netF*) were present in these known plasmids, and the ARGs and virulence genes were not found to coexist in the same plasmid (Fig. S6).

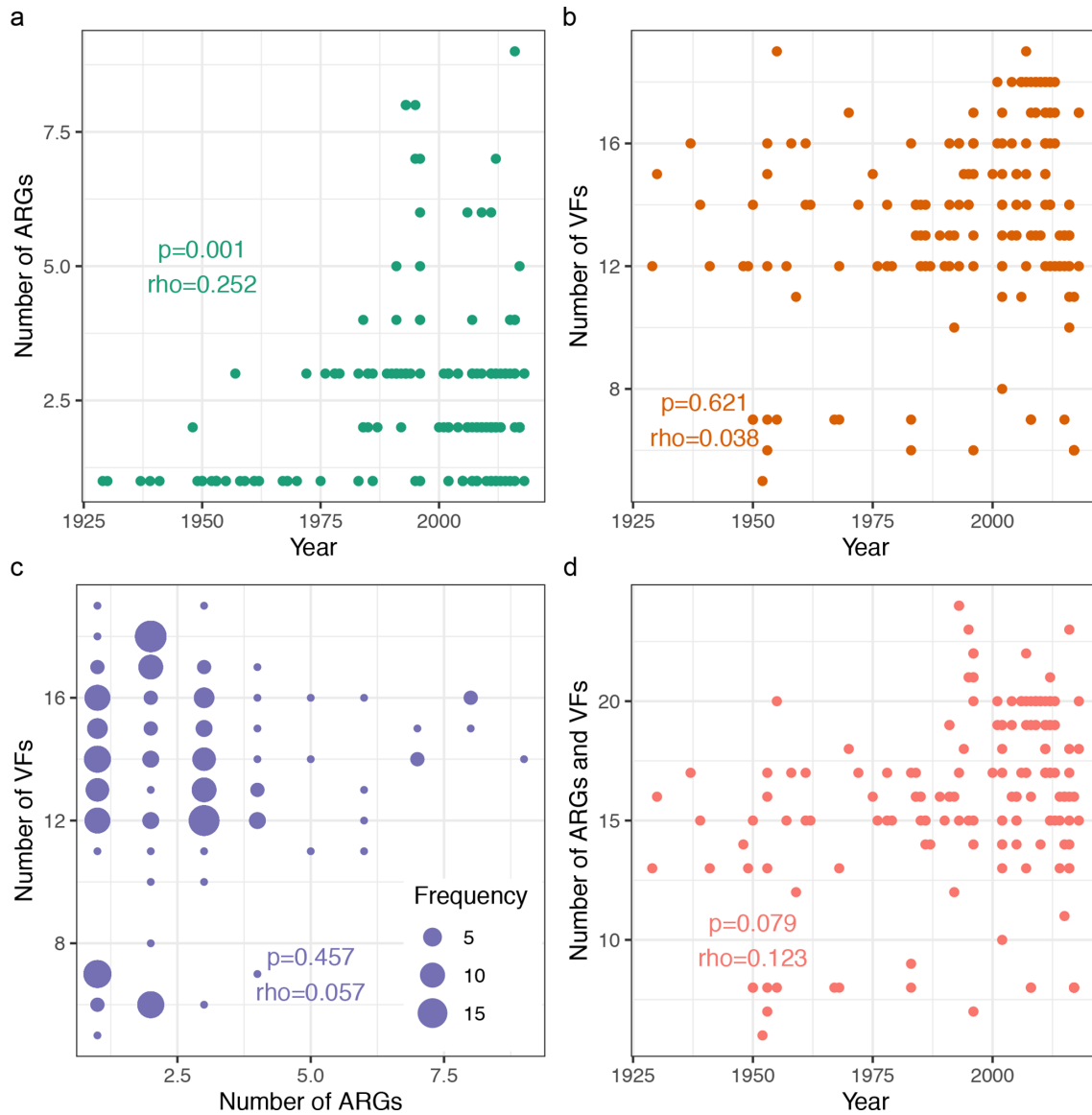
### Antibiotic resistance but not virulence of *C. perfringens* increased over time

To investigate the changes of antibiotic resistance and virulence of *C. perfringens* with time, we performed correlation analysis of ARG and virulence gene number with the isolation time of the strain (Fig. 7). The *C. perfringens* genomes used in this study allow us to detect such changes over an 80-year time span. The results showed that the number of ARGs in *C. perfringens* has increased significantly with the strains' isolation time (Spearman correlation,  $P=0.001$ ), while no significant change was found for the virulence genes ( $P=0.621$ ), and no correlation between the ARG number and virulence gene number in specific isolates ( $P=0.457$ ). When both ARG and virulence were taken into account, an increasing trend (marginally significant) of their total number over time was observed ( $P=0.079$ ). These results reflect that although *C. perfringens* virulence has remained relative stable, its antibiotic resistance has increased steadily, which might be due to the selective pressures of antibiotic use in both animals and humans, especially over the last ~30 years (Fig. 1b).



**Fig. 5.** ARGs in the *C. perfringens* genome and their transfer network among different bacterial genomes. (a) Heatmap visualizing the ARG profile of 173 *C. perfringens* genomes. Cells in purple and white indicate the presence and absence of genes, respectively. (b) ARG number in each lineage. (c) ARG number in strains from different continents (AF: Africa; AS: Asia; EU: Europe; NA: North America; OA: Oceania). (d) The ARG transfer network between *C. perfringens* and other species. The cyan octagon represents *C. perfringens*; nodes with a brown circle represent ARGs found in the genomes of *C. perfringens*; nodes with squares represent bacterial species sharing (>99.0% nucleotide identity) ARGs with *C. perfringens* (purple: species from Firmicutes; pink: species from Proteobacteria; green: species from Actinobacteria).





**Fig. 7.** The changes of ARGs and virulence factors (VFs) in *C. perfringens* over time. The correlation of ARG (a) and VF (b) number with strain isolation times. (c) The correlation between ARG number and VF number in different strains. Dot size is proportional to the number of strains. (d) Changes in the total number of ARGs and VFs in *C. perfringens* over time.

## DISCUSSION

As a human and animal pathogen that can cause various diseases, *C. perfringens* has gained increasing attention globally in recent years [12]. Revealing the evolutionary and genomic features of this bacterium will undoubtedly expand our understanding of its basic biological properties. In this study, we performed phylogenetic and genomic analysis of *C. perfringens* and provided a further view on the bacterium's phylogeny, genome plasticity, virulence and antibiotic resistance. The newly sequenced strains in this study, especially those isolated longest ago, have enriched the genome collection in the databases, allowing us to address the evolution of *C. perfringens* from an historical perspective.

Previously, 56 *C. perfringens* genomes had been clustered into four major clades [20]. In the present study, based on the phylogenetic relationship of strains, we assigned the 173 isolates to five lineages. The major difference with previous studies was that isolates from the reported Clades 2 and 3 were assigned to Lineage V here, which we defined as the modern lineage of *C. perfringens*. Analysis of ARGs and virulence from the available isolates prompted us to suggest that Lineage V may be a more successful clade during the evolutionary process; the acquisition of higher antibiotic resistance and unique toxins has given this lineage an improved survival advantage. Based on a Bayesian phylogenetic analysis, we tentatively inferred that *C. perfringens*

originated ~80000 years ago and has prospered over the last 40000 years. This long evolutionary history is comparable to the evolution of *Mycobacterium tuberculosis* which was inferred to have originated ~70000 years ago [62], suggesting *C. perfringens* may be as old as this famous human pathogen. We should stress that this hypothesis may be biased by the sample size, and global isolates with different isolation times, locations and sources.

It has been reported that the core genome of *C. perfringens* is small, suggesting a high plasticity of the *C. perfringens* genome, but the openness of the *C. perfringens* genome has never been compared with other bacterial species. In fact, a few studies have been performed to investigate the openness of bacterial genomes among different species [63–65]. Here we were surprised to find that *C. perfringens* has the most open genome compared with 51 other bacterial species. We demonstrated that the OI of a bacterial genome was clearly associated with the GC content and 16S rRNA gene copy number in the genome. The rRNA operon copy number has been verified to correlate positively with the growth rate of bacteria. Therefore, the extremely low genome GC content (~28.0%) [20] and relatively high growth speed (a generation time as short as ~7 min) [66] of *C. perfringens* may be major contributors to the high level openness of the *C. perfringens* genome. A highly open genome implies this bacterium's good ability to acquire new genes from external environments and thus to cope with and adapt to hostile survival conditions easily.

Compared with the presence of intact prophages in *C. perfringens* genomes (nearly 70.0% of isolates), we found that the prevalence of CRISPR systems in *C. perfringens* genomes is low (42.2% isolates). As an acquired immune system in bacteria, the CRISPR system plays critical roles in defence against phage infection [49]. The low prevalence of CRISPR systems may indicate a low barrier to horizontal gene transfer events and probably a frequent interaction of phage/prophage with *C. perfringens*. Phage/prophage has been regarded as one of the most important factors driving the evolution of bacterial hosts [67]. Through horizontal gene transfer, phage/prophage can mediate the alteration of host physiological functions by changing their antibiotic resistance and virulence. It is noteworthy in our findings that four toxin genes were predicted to be carried by intact prophages in 16 *C. perfringens* isolates, especially the *cpe* gene which has been shown to exist in both genome and plasmid [68], but never in phage/prophage.

*C. perfringens* does not invade healthy cells but produces various toxins and enzymes that are responsible for the associated lesions and symptoms [1]. Remarkably, this bacterium is reported to secrete >20 degradative toxins which constitute its primary arsenal to initiate histotoxic pathogenesis in both humans and animals [13]. We have shown that although the genome of *C. perfringens* is highly plastic, most of the toxin genes, especially those used for strain toxinotyping, are conserved, which facilitates using PCR or genome sequencing for strain typing. The conservation of these toxin genes and the purifying selection ( $K_a/K_s < 1$ ) on them during

evolution may imply their critical roles for the better survival of *C. perfringens* in different ecological niches. The majority of the virulence genes have no lineage preference, probably due to either their carriage by conjugative plasmids, providing them with the means to be readily transferred between isolates or their chromosomal carriage as intrinsic genes. However, the high prevalence of the food poisoning-related *cpe* gene in Lineage III and IV, and the presence of the NE-associated *netB* gene in only Lineage V warrants further investigation on the interaction between these two important toxin genes and their specific host genetic background.

Our analysis suggested that *C. perfringens* has the potential to exchange ARGs with various bacterial species, especially those from the genera *Staphylococcus* and *Streptococcus*. As species in these two genera have plentiful ARGs [69], it is possible that *C. perfringens* will acquire new ARGs from these bacteria in the future. We should note that the ARG-sharing network presented here does not mean direct ARG transfer events between *C. perfringens* and other bacterial species, as the strains we analysed have no direct contact with each other. However, the presence of the same ARG (more than 99% sequence identity) in different species suggests these genes have the ability to transfer.

Plasmids are important mobile genetic elements which play critical roles in disseminating both virulence genes and ARGs. We showed that plasmid sequences were present in more than 90% of *C. perfringens* isolates, suggesting the high prevalence or even frequent transfer of plasmids within this species. Virulence genes *cpe*, *netB*, *netE* and *netF* were frequently found carried by plasmids, which may indicate the important role of plasmids in spreading these important virulence factors. It should be noted that plasmids pJFP838C and pJFP55F both contained two virulence genes *netE* and *netF*, which deserves further investigation. The *tetA(P)* and *tetB(P)* genes were found in seven known plasmids, while they were still found in many strains having no predicted plasmid sequences, suggesting they were located in chromosomes or in novel plasmids in these strains. As we cannot accurately predict all the plasmid sequences from the draft genomes (only 17 complete genomes included in this study), the virulence gene and ARG contents in *C. perfringens* plasmids may have been underestimated, which will be further unravelled with more complete genomes sequenced.

In conclusion, we analysed 173 *C. perfringens* genomes from strains of worldwide origin. We assigned these genomes to five different lineages having different genomic features and demonstrated that *C. perfringens* has an extremely open genome, probably due to its low genome GC content and rapid growth speed. The gene contents in phages and plasmids contribute to the species' genetic diversity. The limitation of this study is that the isolates/genomes included are not a good global representation of the geographical, animal or environmental distribution of the organism because the availability of sequences is skewed by the specific studies and their different aims. As a result, the lineage features and distributions presented here may be biased by the limitation

of currently available genomes. Therefore, larger scale sampling and sequencing efforts are continuously needed to fully understand this important human and animal enteric pathogen.

#### Funding information

This work was supported by National Key Research and Development Programme of China under grant number 2018YFD0500505. Y.H. was funded by the Starting Grants programme for young talents in China Agricultural University.

#### Acknowledgements

We thank Na Lyu and Fei Liu at the CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, for assistance with whole genome sequencing and data processing.

#### Author contributions

Y.H. conceived and designed the project. Y.F. analysed the data. X.F., L.Z., X.Y., Y.L., S.G. and J.D. collected the isolates and did the experiment. X.J., D.L., Y.G. and J.D. provided invaluable feedback and insights into the analysis. Y.F. and Y.H. wrote the initial manuscript. All authors approved the final version of the manuscript.

#### Conflicts of interest

The authors declare that there are no conflicts of interest.

#### References

- Petit L, Gibert M, Popoff MR. *Clostridium perfringens*: toxinotype and genotype. *Trends Microbiol* 1999;7:104–110.
- Flores-Díaz M, Alape-Girón A. Role of *Clostridium perfringens* phospholipase C in the pathogenesis of gas gangrene. *Toxicon* 2003;42:979–986.
- Kiu R, Caim S, Painset A, Pickard D, Swift C et al. Phylogenomic analysis of gastroenteritis-associated *Clostridium perfringens* in England and Wales over a 7-year period indicates distribution of clonal toxigenic strains in multiple outbreaks and extensive involvement of enterotoxin-encoding (CPE) plasmids. *Microb Genomics* 2019;5:e000297.
- Mpamugo O, Donovan T, Brett MM. Enterotoxigenic *Clostridium perfringens* as a cause of sporadic cases of diarrhoea. *J Med Microbiol* 1995;43:442–445.
- Fernandez Miyakawa ME, Zabal O, Silberstein C. *Clostridium perfringens* epsilon toxin is cytotoxic for human renal tubular epithelial cells. *Hum Exp Toxicol* 2011;30:275–282.
- Law ST, Lee MK. A middle-aged lady with a pyogenic liver abscess caused by *Clostridium perfringens*. *World J Hepatol* 2012;4:252–255.
- Lugli GA, Milani C, Mancabelli L, Turroni F, Ferrario C et al. Ancient bacteria of the ötzi's microbiome: a genomic tale from the copper age. *Microbiome* 2017;5:5.
- Li G, Lillehoj HS, Lee KW, Jang SI, Marc P et al. An outbreak of gangrenous dermatitis in commercial broiler chickens. *Avian Pathol* 2010;39:247–253.
- Okewole PA, Itodo AE, Oyetunde IL, Chima JC, Irokanulo EA et al. *Clostridium perfringens* type A enterotoxaemia in pigs: a report of five cases. *Br Vet J* 1991;147:484–485.
- Uzal FA, Senties-Cué CG, Rimoldi G, Shivaprasad HL. Non-*Clostridium perfringens* infectious agents producing necrotic enteritis-like lesions in poultry. *Avian Pathol* 2016;45:326–333.
- Canard B, Saint-Joanis B, Cole ST. Genomic diversity and organization of virulence genes in the pathogenic anaerobe *Clostridium perfringens*. *Mol Microbiol* 1992;6:1421–1429.
- Kiu R, Hall LJ. An update on the human and animal enteric pathogen *Clostridium perfringens*. *Emerg Microbes Infect* 2018;7:1–15.
- Li J, Adams V, Bannam TL, Miyamoto K, Garcia JP et al. Toxin plasmids of *Clostridium perfringens*. *Microbiol Mol Biol Rev* 2013;77:208–233.
- Shimizu T, Ohtani K, Hirakawa H, Ohshima K, Yamashita A et al. Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc Natl Acad Sci U S A* 2002;99:996–1001.
- Hassan KA, Elbourne LDH, Tetu SG, Melville SB, Rood JI et al. Genomic analyses of *Clostridium perfringens* isolates from five toxinotypes. *Res Microbiol* 2015;166:255–263.
- Lacey JA, Allnutt TR, Vezina B, Van TTH, Stent T et al. Whole genome analysis reveals the diversity and evolutionary relationships between necrotic enteritis-causing strains of *Clostridium perfringens*. *BMC Genomics* 2018;19:379.
- Lacey JA, Johanesen PA, Lyras D, Moore RJ. Genomic diversity of necrotic enteritis-associated strains of *Clostridium perfringens*: a review. *Avian Pathol* 2016;45:302–307.
- Myers GSA, Rasko DA, Cheung JK, Ravel J, Seshadri R et al. Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*. *Genome Res* 2006;16:1031–1040.
- Ronco T, Stegger M, Ng KL, Lilje B, Lyhs U et al. Genome analysis of *Clostridium perfringens* isolates from healthy and necrotic enteritis infected chickens and turkeys. *BMC Res Notes* 2017;10:270.
- Kiu R, Caim S, Alexander S, Pachori P, Hall LJ. Probing genomic aspects of the multi-host pathogen *Clostridium perfringens* reveals significant pangenome diversity, and a diverse array of virulence factors. *Front Microbiol* 2017;8:2485.
- McAllister KN, Sorg JA. CRISPR genome editing systems in the genus *Clostridium*: a timely advancement. *J Bacteriol* 2019;201:e00219–19.
- Kiu R, Brown J, Bedwell H, Leclaire C, Caim S et al. Genomic analysis on broiler-associated *Clostridium perfringens* strains and exploratory caecal microbiome investigation reveals key factors linked to poultry necrotic enteritis. *Anim Microbiome* 2019;1:12.
- Bolger AM, Lohse M, Usadel B, Bolger Anthony M, Marc L, Bjoern U. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 2014;30:2114.
- Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
- Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A et al. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019;35:4453–4455.
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H et al. Beast 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 2014;10:e1003537.
- Rasmussen S, Allentoft ME, Nielsen K, Orlando L, Sikora M et al. Early divergent strains of *Yersinia pestis* in Eurasia 5000 years ago. *Cell* 2015;163:571–582.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst Biol* 2018;67:901–904.
- Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. RhierBAPS: an R implementation of the population clustering algorithm hierbaps. *Wellcome Open Res* 2018;3:93.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25:1189–1191.
- Paradis E, Claude J, Strimmer K. Ape: analyses of phylogenetics and evolution in R language. *Bioinformatics* 2004;20:289–290.
- Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinforma* 2005;1:117693430500100.
- Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P et al. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007;57:81–91.

36. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Anal Methods* 2016;8:12–24.
37. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.
38. Livingstone PG, Morphey RM, Whitworth DE. Genome sequencing and pan-genome analysis of 23 *Corallococcus* spp. strains reveal unexpected diversity, with particular plasticity of predatory gene sets. *Front Microbiol* 2018;9:3187.
39. Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res* 2015;43:D593–D598.
40. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D et al. EGGNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 2016;44:D286–D293.
41. UniProt Consortium, Bateman A. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;47:D506–D515.
42. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res* 2016;44:D694–D697.
43. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006;22:1658–1659.
44. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs\_Calculator 2.0: a toolkit incorporating Gamma-Series methods and sliding window strategies. *Genomics, Proteomics Bioinforma* 2010;8:77–80.
45. Jia B, Raphenya AR, Alcock B, Wagtechner N, Guo P et al. Card 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2017;45:D566–D573.
46. Charlebois A, Jalbert L-A, Harel J, Masson L, Archambault M. Characterization of genes encoding for acquired bacitracin resistance in *Clostridium perfringens*. *PLoS One* 2012;7:e44449.
47. Hu Y, Yang X, Li J, Lv N, Liu F et al. The bacterial mobile resistome transfer network connecting the animal and human microbiomes. *Appl Environ Microbiol* 2016;82:6672–6681.
48. Arndt D, Grant JR, Marcu A, Sajed T, Pon A et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44:W16–W21.
49. Nasko DJ, Ferrell BD, Moore RM, Bhavsar JD, Polson SW et al. Crispr spacers indicate preferential matching of specific viroplankton genes. *mBio* 2019;10:e02651–18.
50. Krawczyk PS, Lipinski L, Dziembowski A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res* 2018;46:e35.
51. Brooks L, Kaze M, Sstrom M. A curated, comprehensive database of plasmid sequences. *Microbiol Resour Announc* 2019;8:e01325–18.
52. Gohari IM, Prescott JF. Commentary: Probing genomic aspects of the multi-host pathogen *Clostridium perfringens* reveals significant pangenome diversity, and a diverse array of virulence factors. *Front Microbiol* 2018;9.
53. Wan X, Koster K, Qian L, Desmond E, Brostrom R et al. Genomic analyses of the ancestral Manila family of *Mycobacterium tuberculosis*. *PLoS One* 2017;12:e0175330.
54. Mooi FR. *Bordetella pertussis* and vaccination: the persistence of a genetically monomorphic pathogen. *Infect Genet Evol* 2010;10:36–49.
55. Roller BRK, Stoddard SF, Schmidt TM. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nat Microbiol* 2016;1:16160.
56. Rood JI. Virulence genes of *Clostridium perfringens*. *Annu Rev Microbiol* 1998;52:333–360.
57. Li J, Sayeed S, Robertson S, Chen J, McClane BA. Sialidases affect the host cell adherence and epsilon toxin-induced cytotoxicity of *Clostridium perfringens* type D strain CN3718. *PLoS Pathog* 2011;7:e1002429.
58. Lindström M, Heikinheimo A, Lahti P, Korkeala H. Novel insights into the epidemiology of *Clostridium perfringens* type A food poisoning. *Food Microbiol* 2011;28:192–198.
59. Keyburn AL, Boyce JD, Vaz P, Bannam TL, Ford ME et al. NetB, a new toxin that is associated with avian necrotic enteritis caused by *Clostridium perfringens*. *PLoS Pathog* 2008;4:e26.
60. Sheedy SA, Ingham AB, Rood JI, Moore RJ. Highly conserved alpha-toxin sequences of avian isolates of *Clostridium perfringens*. *J Clin Microbiol* 2004;42:1345–1347.
61. Justin N, Walker N, Bullifent HL, Songer G, Bueschel DM et al. The first strain of *Clostridium perfringens* isolated from an avian source has an alpha-toxin with divergent structural and kinetic properties. *Biochemistry* 2002;41:6253–6262.
62. Comas I, Coscolla M, Luo T, Borrell S, Holt KE et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* 2013;45:1176–1182.
63. Chan AP, Sutton G, DePew J, Krishnakumar R, Choi Y et al. A novel method of consensus pan-chromosome assembly and large-scale comparative analysis reveal the highly flexible pan-genome of *Acinetobacter baumannii*. *Genome Biol* 2015;16:143.
64. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ et al. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biol* 2010;11:R107.
65. Park SC, Lee K, Kim YO, Won S, Chun J. Large-scale genomics reveals the genetic characteristics of seven species and importance of phylogenetic distance for estimating pan-genome size. *Front Microbiol* 2019;10:834.
66. Smith S, Schaffner DW. Evaluation of a *Clostridium perfringens* predictive model, developed under isothermal conditions in broth, to predict growth in ground beef during cooling. *Appl Environ Microbiol* 2004;70:2728–2733.
67. Fortier LC, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* 2013;4:354–365.
68. Freedman JC, Shrestha A, McClane BA. *Clostridium perfringens* enterotoxin: action, genetics, and translational applications. *Toxins* 2016;8:73.
69. Frieden T. Antibiotic resistance threats in the United States. *Centers Dis Control Prev* 2013;114.

### Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at [microbiologyresearch.org](http://microbiologyresearch.org).