Role of mobile genetic elements in the global dissemination of the carbapenem resistance gene \( \text{bla}_{\text{NDM}} \)

Mislav Acman\(^1\text{*}\), Ruobing Wang\(^2\), Lucy van Dorp\(^1\), Liam P. Shaw\(^3\), Qi Wang\(^2\), Nina Luhmann\(^4\), Yuyao Yin\(^2\), Shijun Sun\(^2\), Hongbin Chen\(^2\), Hui Wang\(^2\) & Francois Balloux\(^1\)

The mobile resistance gene \( \text{bla}_{\text{NDM}} \) encodes the NDM enzyme which hydrolyses carbapenems, a class of antibiotics used to treat some of the most severe bacterial infections. The \( \text{bla}_{\text{NDM}} \) gene is globally distributed across a variety of Gram-negative bacteria on multiple plasmids, typically located within highly recombining and transposon-rich genomic regions, which leads to the dynamics underlying the global dissemination of \( \text{bla}_{\text{NDM}} \) to remain poorly resolved. Here, we compile a dataset of over 6000 bacterial genomes harbouring the \( \text{bla}_{\text{NDM}} \) gene, including 104 newly generated PacBio hybrid assemblies from clinical and livestock-associated isolates across China. We develop a computational approach to track structural variants surrounding \( \text{bla}_{\text{NDM}} \), which allows us to identify prevalent genomic contexts, mobile genetic elements, and likely events in the gene’s global spread. We estimate that \( \text{bla}_{\text{NDM}} \) emerged on a Tn125 transposon before 1985, but only reached global prevalence around a decade after its first recorded observation in 2005. The Tn125 transposon seems to have played an important role in early plasmid-mediated jumps of \( \text{bla}_{\text{NDM}} \), but was overtaken in recent years by other elements including IS26-flanked pseudo-composite transposons and Tn3000. We found a strong association between \( \text{bla}_{\text{NDM}} \)-carrying plasmid backbones and the sampling location of isolates. This observation suggests that the global dissemination of the \( \text{bla}_{\text{NDM}} \) gene was primarily driven by successive between-plasmid transposon jumps, with far more restricted subsequent plasmid exchange, possibly due to adaptation of plasmids to their specific bacterial hosts.
Antimicrobial resistance (AMR) poses a major challenge to human and veterinary health. AMR can be conferred by vertically inherited point mutations or via the acquisition of horizontally transmitted ‘accessory’ genes, often located in transposons and plasmids. The blaNDM gene represents a typical example of a mobile AMR gene. blaNDM encodes a metallo-β-lactamase capable of hydrolyzing most β-lactam antibiotics. These antibiotics are used as a first-line treatment for severe infections and to treat multidrug-resistant Gram-negative bacterial infections. As such, the global prevalence of bacteria carrying blaNDM represents a major public health concern.

blaNDM was first described in 2008 from a Klebsiella pneumoniae isolate from a urinary tract infection in a Swedish patient returning from New Delhi, India. While blaNDM now has a worldwide distribution, most of the earliest cases have been linked to the Indian subcontinent, leading to this region being suggested as the likely location for the initial mobilization event. NDM-positive Acinetobacter baumannii isolates have been retrospectively identified from an Indian hospital in 2005, which remain the earliest observations to date. Nevertheless, an NDM-positive A. pittii isolate was also collected in 2006 from a Turkish patient with no history of travel outside Turkey.

Although no complete genome sequences are publicly available from these earliest observations, the first NDM-positive isolates from 2005 were shown to carry blaNDM on multiple non-conjugative, but potentially mobilizable plasmid backbones. In addition, blaNDM in these early isolates was positioned within a complete Tn125 transposon with IS26 insertion sequences (IISs) as well as ISCR27 (IS-containing common region 27), suggesting the possibility of complex patterns of mobility since the gene’s initial integration. Subsequent NDM-positive isolates across multiple species consistently harbour either a complete or fragmented ISAba125 (an IS constituting Tn125), immediately upstream of blaNDM, which provides a promoter region, leading to the initial mobilization event. The presence of ISAba125 in some form in all NDM-positive isolates to date and the early observations in A. baumannii have led to Tn125 being proposed as the ancestral transposon responsible for the mobilization of blaNDM, and A. baumannii as the ancestral host.

The NDM enzyme itself is of possible chimeric origin, with the first six amino acids in NDM matching to those in aphA6, a gene providing aminoglycoside resistance and also flanked by ISAba125. It is hypothesized that ISCR27, which uses a rolling-circle (RC) transposition mechanism, initially mobilized a progenitor of blaNDM in Xanthomonas sp. and placed it downstream of ISAba125. At least 29 distinct sequence variants of the NDM enzyme have been described to date. The most prevalent of these variants is the first to have been characterized, denoted NDM-1. Different NDM variants are mostly distinguished by a single amino-acid substitution, apart from NDM-18 that carries a tandem repeat of five amino acids. None of the observed substitutions occur in the active site and their functional impact remains under debate.

blaNDM is found in at least 11 bacterial families and NDM-positive isolates have heterogeneous clonal backgrounds, supporting multiple independent acquisitions of blaNDM. Although blaNDM has been observed on bacterial chromosomes, it is most commonly found on plasmids, comprising multiple different backbones or types. Thus far, blaNDM has been associated with at least 20 different plasmid types, predominantly IncFIB, IncFII, IncAC (IncC), IncX3, IncH, and IncL/M, and also in untyped plasmids. Furthermore, even within the same plasmid type, blaNDM can be found in a variety of genomic contexts, often interspersed by multiple ISs and composite transposons. The immediate environment of blaNDM has been reported to vary even in isolates from the same patient. Many mobile elements are thought to play important roles in dissemination, including ISAba125, IS3000, IS26, IS5, ISCR1, Tn3, Tn125, and Tn3000. It is therefore clear that the spread of blaNDM was, and is, a multi-layer process involving multiple mobile genetic elements—the mobileome. blaNDM mobility involves diverse processes, including genetic recombination, transposition, conjugation and transformation of plasmids, transduction, and transfer through outer-membrane vesicles.

Previous surveys of blaNDM-positive genomes have led to a better understanding of its evolution. However, a major difficulty, as for other AMR genes, is relating the diverse genomic contexts to temporal evolution. Here, we outline an alignment-based method to identify flanking structural variants and use it to build a history of the insertion and mobilization events. We compile a global dataset of more than 6000 NDM-positive isolates. In line with previous studies, we identify Tn125, IS26 and Tn3000 as the main contributors to blaNDM mobility but go further and estimate the timing of the initial emergence of blaNDM to pre-1990, around two decades prior to its first detection and rapid dissemination. Our findings suggest that this global spread was driven primarily by transposons, with plasmids playing more of a role in local transmission.

### Results

**A global dataset of blaNDM carriers.** We compiled a dataset of 6155 bacterial genomes (7148 contigs) carrying at least one copy of blaNDM (Fig. 1). These include: published assemblies from NCBI RefSeq (n = 2632), NCBI GenBank (n = 1158) and Enterobase (n = 1379); bacterial genomes assembled using short-read de novo assembly from NCBI’s Sequence Read Archive (SRA) (n = 882); and newly generated bacterial genomes isolated from 79 hospitalized patients across China and 25 live-stock farms assembled using hybrid PacBio-Illumina de novo assembly (n = 104) (Supplementary Table 1 and Supplementary Fig. 1). While public genomes have inherent sampling biases, taking advantage of them offers the most comprehensive approach available. Data were included from 251 different Bioprojects, with more than half the samples linked to two large-scale database refinement efforts. Quality assessment of blaNDM-positive contigs obtained from SRA de novo assemblies showed good contig mapping coverage and did not reveal any problems that could potentially compromise downstream analyses (see Methods and Supplementary Fig. 2).

The dataset included blaNDM-positive isolates from 88 states (Fig. 1a) mostly collected in Asia, particularly mainland China (n = 1270), European countries (941), USA (461), Thailand (419) and India (361). At least 27 bacterial genera were represented, with a large fraction of Klebsiella and Escherichia isolates (2664 and 2154 genomes respectively; Fig. 1b and Supplementary Data 1). Collection dates were recorded for 4816 samples (78.25%). Of these, the majority were collected between 2014 and 2019 (71.05%, Fig. 1c). The dataset also includes 55 genomes collected in 2010 or earlier. These include the 2008 K. pneumoniae isolate from Sweden in which blaNDM was first characterized; one 2008 Enterobacter hormaechei isolate from India; one 2008 S. enterica isolate from London, UK; one A. baumannii isolate from an individual of Balkan origin collected in Germany in 2007; and nine assembled E. coli genomes from urine samples collected in Greece in 2007 (Supplementary Data 1).

The dataset contained 17 known variants of NDM. NDM-1 was the most abundant (n = 4127; Supplementary Fig. 3a) with NDM-5 (n = 2394) increasing in prevalence after 2012 (Supplementary Fig. 2b, c). Variants showed different associations with plasmid types (Supplementary Fig. 2d) and genera.

ARTICLE NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-022-28819-2

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-022-28819-2

2 NATURE COMMUNICATIONS | (2022) 13:1131 | https://doi.org/10.1038/s41467-022-28819-2 | www.nature.com/naturecommunications
Supplementary Fig. 2e) but were fairly evenly distributed across the world except for \(\text{blaNDM-4}\)-carrying isolates largely collected in Southeast Asia and \(\text{blaNDM-9}\) predominantly found in East Asia (Supplementary Fig. 2f).

Plasmid backbones carrying \(\text{blaNDM}\). We identified 33 different replicon types on 1222 contigs using PlasmidFinder42 (Fig. 1d). The most prevalent replicon type was IncX3 (444 contigs), and abundant types exhibited geographic structure (Supplementary Fig. 4). To further identify uncharacterized plasmid sequences, we mapped 3599 contigs to a set of complete plasmid reference sequences after discarding short contigs (see Methods). This revealed 181 clusters of similar putative plasmid sequences (Fig. 2 and Supplementary Data 2). Most clusters (\(n = 105\)) grouped contigs of the same replicon type and contained a small number of contigs (only 27 clusters included >10 contigs), in line with a diverse and dynamic population of plasmid backbones for \(\text{blaNDM}\).

The majority (\(n = 2427; 68.4\%\)) of \(\text{blaNDM}\)-carrying contigs were associated with small putative plasmids (<10 Kb; Fig. 2). While this could suggest small plasmids play a key role as \(\text{blaNDM}\) carriers, this pattern could also result from consistently fragmented de novo assemblies due to duplicated ISs and transposons. Consistent with this latter hypothesis, 610 contigs mapped to pKP-YQ12450 that is likely a 7.8-Kb fragment of a larger plasmid21. Conversely, Roach et al. provide evidence that other small \(\text{blaNDM}\)-carrying plasmids (Peruvian pKP-NDM-1, isoforms 1–5) are inherited by descent and are a result of transposon-mediated plasmid fusion43.

Resolving structural variants in the \(\text{blaNDM}\) flanking regions. To go beyond a static reference-based view of variation around \(\text{blaNDM}\) and gain a detailed overview of the possible events in its evolution, we developed an alignment-based approach to progressively resolve genomic variation moving upstream or downstream from the gene (see Methods, Fig. 3). In brief, a pairwise contiguous Mega BLAST search (v2.10.1+)44,45 is applied to all \(\text{blaNDM}\)-carrying contigs to identify all possible homologous regions between each contig pair. Only BLAST hits covering the complete \(\text{blaNDM}\) gene are retained (Fig. 3a). Next, starting from \(\text{blaNDM}\), a gradually increasing ‘splitting threshold’ is used to monitor structural variants as they appeared upstream or downstream of the gene. At each step, a network of contigs (nodes) that share a BLAST hit with a minimum length as given by the ‘splitting threshold’ is assessed (Fig. 3b). As we move upstream or downstream and further away from the gene, the network starts to split into smaller clusters, each carrying contigs that share an uninterrupted stretch of homologous DNA, which
can be represented as a tree (Fig. 3c). This approach treats the upstream and downstream flanking regions separately rather than simultaneously and is agnostic to whether splitting into ‘sequence clusters’ is caused by structural variants of the same genomic background or different genomic backgrounds.

Upstream of \( \text{bla}_{\text{NDM}} \), >98% of sufficiently long contigs included a ~75 bp fraction of IS\textsuperscript{Aba125}, supporting Tn\textsubscript{125} as an ancestral transposon of the \( \text{bla}_{\text{NDM}} \) gene in agreement with previous work\textsuperscript{1,5,9,10} (Supplementary Figs. 5 and 6). However, the homology of the region upstream of \( \text{bla}_{\text{NDM}} \) falls quickly: within a few hundred base pairs of the \( \text{bla}_{\text{NDM}} \) start codon the region splits into multiple structural variants, none of which dominate the considered pool of contigs (Supplementary Figs. 5 and 6). We identified 141 different structural variants within 1200 bp upstream of \( \text{bla}_{\text{NDM}} \). This upstream region contained a high number of ISs (e.g., IS\textsuperscript{Aba125} \( n = 243 \), IS\textsuperscript{5} \( n = 426 \), IS\textsuperscript{3000} \( n = 60 \), IS\textsuperscript{Kpn14} \( n = 55 \), and IS\textsuperscript{Eco33} \( n = 147 \)). This transposition hotspot probably contributes to fragmented assemblies: 2269 contigs were excluded from further analysis for being too short (Supplementary Fig. 5).

The downstream flanking region exhibits more gradual structural diversification than the upstream region, with one dominant putative ancestral background (Fig. 4). As illustrated by the stem of the tree of structural variants, many of the 7014 contigs analyzed contained complete sequences of the same set of genes: \( \text{ble} \) (6863 contigs), \( \text{trpF} \) (6038), \( \text{dsbD} \) (5551), \( \text{cutA} \) (2731), \( \text{groS} \) (2175), \( \text{groL} \) (1631). When restricted to \( \text{bla}_{\text{NDM}} \)-positive contigs of sufficient length to possibly harbour the full repertoire of these genes \( n = 3786 \), almost half carry all of them \( n = 1631 \); 43.1%. In addition, we find dominant structural variants associated with various source databases and sequence lengths hence diminishing the impact of the sampling bias (Supplementary Fig. 7).

**Early events in the spread of \( \text{bla}_{\text{NDM}} \).** While we did not observe any strong overall signal in the distribution of associated plasmid backbones, bacterial genera, or sampling locations, closer examination of mobilome features common to sufficiently large numbers of isolates indicated early events in the spread of \( \text{bla}_{\text{NDM}} \). The putative ancestral Tn\textsubscript{125} background, with an uninterrupted downstream IS\textsuperscript{Aba125} element, was seen in contigs mainly from \textit{Acinetobacter} and \textit{Klebsiella} (Fig. 4 top). Conversely, the diversity of bacterial genera carrying IS\textsuperscript{Aba125} upstream is more substantial (Supplementary Fig. 5 top). Only 203 contigs carried a complete IS\textsuperscript{Aba125} downstream of \( \text{bla}_{\text{NDM}} \) of which 147 carried an IS\textsuperscript{Aba125} sequence in proximity (<9 Kb) to the \( \text{bla}_{\text{NDM}} \) start codon. These account for a minority (7%; 147/2097) of isolates when sufficiently long contigs are considered.

This supports the initial dissemination of \( \text{bla}_{\text{NDM}} \) by Tn\textsubscript{125} to other plasmid backbones predominately being mediated by \textit{Acinetobacter} and \textit{Klebsiella}, after which the transposon was disrupted by other rearrangements. IS\textsuperscript{3000}, both upstream and downstream, was almost exclusively associated with samples from \textit{Klebsiella} (Fig. 4 and Supplementary Fig. 5). Thus, as suggested by Campos et al.\textsuperscript{24}, Tn\textsubscript{3000}—a
composite transposon made of two copies of IS3000—likely remobilized blaNDM, following the ‘fossilisation’ of Tn125; our findings suggest this secondary mobilization primarily happened in Klebsiella species. Tn5403 was found extensively associated with IncN2 plasmids (Fig. 4) that could have placed blaNDM in this background via cointegrate intermediate as previously suggested by Poirel et al.9. Some elements of the mobiome were geographically linked, e.g., IS5 that was predominantly found upstream of blaNDM on IncX3 plasmids in Escherichia from East Asia (Supplementary Fig. 5). IS5 is known to enhance transcription of nearby promoters in E. coli46 and its abundance and positioning just upstream of blaNDM suggests a similar role in this case.

One of the most commonly identified transposable elements in the downstream flanking region (~30% prevalence) was ISCR1 (IS91 family transposase) (Fig. 4) always accompanied by sull and occasionally in configuration with ant1 or pspF, ampR, and dap genes. In some cases, a small and possibly fragmented putative IS, which we refer to as ‘IS-?', is found further downstream. IS-? bears little similarity to known ISs and it is unclear what role it plays in the mobility of blaNDM. ISCR1 is found at various positions downstream of blaNDM and often in Escherichia and Klebsiella species. We note that, in most cases, the orientation of ISCR1 should prevent this element from mobilizing blaNDM (Fig. 4)13. Nevertheless, the prevalence of this element could be due to the several AMR genes it can mobilize, such as sull or ampR. ISCR1s are mainly found in complex class 1 integrons13; however, not many annotated integrase genes are located within the vicinity of blaNDM. In fact, only 15 contigs were found to have an integrase <50 Kb away from blaNDM and none showed any consistency in integrase placement with respect to blaNDM. This suggests integrase plays a minor role in the dissemination of blaNDM.

Another notable ISCR element is ISCR27 that is consistently found immediately downstream of the groL gene at high prevalence (33.1% of sufficiently long contigs; Fig. 4). Contrary to its ISCR1 relative, ISCR27 is correctly oriented to mobilize blaNDM as is presumed to have happened during the initial mobilization of the progenitor of blaNDM10. However, we find no evidence of subsequent ISCR27 mobility. The origin of RC replication of ISCR27 (oriIS; GCGGTGAACCTCTATACC) is located 236 bp downstream of the ISCR27 transposase stop codon. The region downstream of this stop codon in all structural variants bearing a complete ISCR27 is highly conserved for at least 750 bp (Fig. 4).

Subsequent rearrangements dominated by IS26. Three sharp drops in the number of considered contigs at particular distances downstream of blaNDM (see Fig. 4, e.g., region 3000–3300 bp) prompted us to investigate these distinct cut-offs. We mapped 781 raw Illumina paired-end sequencing reads from our dataset back to their matching blaNDM contigs. The read overhangs (≥50 bp) that mapped to the downstream end of the contigs were screened against the ISFinder database47. The ≥50 bp overhangs associated with 3000–3300 long flanks downstream of blaNDM corresponding to the largest observed drop almost exclusively match the left inverted repeat of the IS26 sequence (Supplementary Fig. 8). Another hotspot associated with IS26 was found around 7500 bp, while at around 7800 bp, a number of overhanging reads mapped to ISAba125. These positions roughly match the third drop in the number of contigs observed 7500–8000 bp downstream of blaNDM. No ISs were found to match the second drop in number of contigs (5000–5250 bp).

IS26, although often found in two adjacent copies forming a seemingly composite transposon, is a so-called pseudo-composite (or pseudo-compound) transposon48. In contrast to composite transposons, a fraction of DNA flanked by the two IS26 is mobilized either via cointegrate formation or in the form of a circular translocatable unit (TU), which consists of a single IS26 element and a mobilized fraction of DNA, and inserts preferentially next to another IS2648,49. Taken together, the presented results, including Supplementary Fig. 8, suggest three possible explanations for the
Fig. 4 Splitting of structural variants downstream of blaNDM. The ‘splitting’ tree for the most common (i.e., ≥10 contigs) structural variants is shown on the left-hand side. The labels on the nodes indicate the number of contigs remaining on each branch. Labels of (yellow) branches with <20 contigs are not shown. The other contigs either belong to other structural variants or were removed due to being too short in length. The number of contigs cutting short is indicated by the area chart at the bottom. Similarly, the number of less common structural variants is indicated by the upper area chart. Genome annotations provided by the Prokka and Roary pipelines of the most common structural variants are shown in the middle of the figure. The homologous regions among structural variants are indicated by the grey shading. Some of the structural variants and branches were intentionally cut short even though their contigs were of sufficient size or longer. This was done to prevent excessive bifurcation and to make the tree easier to interpret. In particular, branches with more than 75% of contigs lost due to variation and short length were truncated. The distribution of genera, plasmid backbones and geographical regions of samples that belong to each of the common structural variants is shown on the right-hand side.

presence of short blaNDM carrying contigs in the dataset: (i) the presence of IS26 TUs in the host cell; (ii) other circular DNA formations mediated by plasmid recombination, transposons or ISCR elements; or (iii) misassembly of contigs due to the presence of multiple copies of the same Is64.

To further investigate the mobility of blaNDM, we characterized the most common (pseudo-)composite transposons theoretically capable of mobilizing blaNDM (Fig. 5). These were defined as stretches of DNA flanked by two matching complete or partial Is5 <30 Kb apart and enclosing blaNDM. In total, we identified 640 composite transposons in 468 contigs that comprised 31 different types with the most frequent being: IS26 (231 instances), IS3000 (forming Tn3000; 168), ISAba125 (forming Tn125; 138 instances), and IS15 (28) (Fig. 5b). Interestingly, we observe 80 cases where >2 of the same IS flank blaNDM. These are mostly IS26 (59) that could indicate the presence of cointegrate formation and showcase increased activity of this particular insertion element. Only 431 of the 640 putative composite transposons identified contained both complete flanking Is5, while others had at least one Is partially truncated. In addition, 1681 ISCR27, and 150 ISCR1 were found in similar proximity and appropriate orientation to mobilize blaNDM (Fig. 5b). However, as mentioned earlier, their role in the transposition of blaNDM appears minor.

In the majority of cases, composite transposons Tn125 and Tn3000 were found to have a consistent length ranging from 7 to 10 Kb (Fig. 5a). Similarly, ISCR1 and ISCR27 are found at fixed positions downstream of blaNDM. However, the lengths of transposons formed by IS15, a known variant of IS26, and especially IS26 were found to be more variable. Pairs of IS26 are found to be 2.5-30 Kb apart again consistent with increased activity and multiple independent insertions. We note that IS15 and IS26 occur at increased presence in samples collected in East and Southeast Asia (Fig. 5c). These occur roughly equally in Escherichia and Klebsiella genera (Fig. 5d) and are associated with multiple plasmid backbones, but predominantly on IncF plasmids (Fig. 5e). Tn125 and Tn3000 have a notable predominance in the Indian subcontinent (Fig. 5c) and largely in the Acinetobacter and Klebsiella genera, respectively (Fig. 5d).

Molecular dating of key events. We estimated the relative timing of the formation of the Tn125 and Tn3000 transposons (see Methods). After selecting only contigs with conserved transposon configurations we aligned each transposon region and identified
the likely root (i.e., ancestral) sequence by assessing temporal patterns (Supplementary Figs. 9 and 10; see Methods). Overall, we observed fewer SNPs, mostly located within the transposase gene, in the alignment of Tn\(_{3000}\) compared to Tn\(_{125}\), but observed a notable temporal signal for both (Supplementary Figs. 11 and 12). We also assessed temporal signal for three other prevalent insertion events (Fig. 4), namely: \(\text{bla}_{\text{NDM}}\) with down-stream ISCR27, \(\text{bla}_{\text{NDM}}\) with correctly oriented downstream \(\text{folP}\)-ISCR1 (+ strand), and \(\text{bla}_{\text{NDM}}\) – dsbD with downstream ISCR1 (− strand) ending with an unknown putative IS (labelled IS-?). However, no significant temporal signal was recovered for these events.

The Bayesian analysis indicated that the most recent common ancestor (MRCA) of the Tn\(_{125}\) transposon carrying the \(\text{bla}_{\text{NDM}}\) gene dated to before 1990 (Fig. 6a). While the time intervals are uncertain, the results are consistent with an MRCA in the mid-twentieth century— strikingly half a century prior to the first reported Tn\(_{125}\)-\(\text{bla}_{\text{NDM}}\)-positive isolates\(^7\). Conversely, the mobilization of \(\text{bla}_{\text{NDM}}\) by Tn\(_{3000}\) is estimated to have happened later at the turn of the millennium (Fig. 6b). These findings are consistent with a wider narrative whereby the spread of \(\text{bla}_{\text{NDM}}\) was initially driven by Tn\(_{125}\) mobilization before subsequent transposition by Tn\(_{3000}\), IS\(_{26}\) and others.

**Temporal diversity in \(\text{bla}_{\text{NDM}}\) isolates suggests the role of plasmids.** The earliest samples in our dataset are from 2007 to 2010 and comprise 21 \(\text{bla}_{\text{NDM}}\)-positive isolates. These already encompass seven bacterial species, collected in eight countries spanning four geographic regions (17 clinical samples and four of unknown origin from South Asia, Middle East, Oceania, and Europe). Such a wide host and geographic distribution, even in the earliest available genomes, illustrates the extraordinarily high mobility of \(\text{bla}_{\text{NDM}}\) at this stage and is consistent with our molecular dating estimates.

In order to trace the progress of \(\text{bla}_{\text{NDM}}\)’s rapid spread after 2005 (coinciding with the first published observations), we measured

---

**Fig. 5 Global prevalence and genetic context of most frequent putative (pseudo-) composite transposons and insertion sequences containing common regions (ISCRs) capable of mobilizing \(\text{bla}_{\text{NDM}}\) gene.** Transposons were defined as stretches of DNA flanked by two matching complete or partial ISs <30 Kb apart and enclosing \(\text{bla}_{\text{NDM}}\). Putative pseudo-composite transposons were labelled according to their constituent ISs (IS15 and IS26). a Marginal distributions of transposon lengths or distances of ISCRs from \(\text{bla}_{\text{NDM}}\) start codon. b Overall counts of the frequent transposable elements (i.e., >25 representatives). c–e Bar plots, respectively, indicating proportions of plasmid backbones, bacterial genera and sampling location associated with most frequent transposable elements.
Fig. 6 Molecular dating of blaNDM mobilization by Tn125 and Tn3000. Posterior density distributions of ancestral sequence age (i.e., root height) for the Tn125 (a) and Tn3000 (b) transposons. The ancestral sequence emergence was estimated using two Bayesian tip-dating approaches implemented in BactDating and BEAST2. Three different population growth priors were used in case of BEAST2: Coalescent Constant Population, Coalescent Exponential Population, and Coalescent Bayesian Skyline as given by the colour scheme and legend at right. Median estimates with 95% highest density interval (HDI) are provided in the panel legends.

Discussion

In this study, we have characterized the extent structural variation around blaNDM in a large global dataset in order to reconstruct its evolutionary history and the main actors underlying its spread. Our results highlight an ancestral background of blaNDM as well as several insertion events and a myriad of other genetic reshufflings, together pointing to an early emergence of blaNDM followed by more recent and rapid dissemination globally. Genetic reshuffling and mobilization of blaNDM by multiple transposons aided its rapid dissemination via a multitude of plasmid backbones.

We go beyond previous smaller studies by dating the MRCA of the hypothesized ancestral form—the transposon Tn125, together with blaNDM in its chimeric form—to pre-1990, and possibly well back into the mid-twentieth century. A likely scenario is an origin in Acinetobacter in the Indian subcontinent. We note that Tn125 is mostly present in Acinetobacter and Klebsiella species and it is likely this transposon played an important role in early plasmid jumps of blaNDM, given it is the dominant transposon in our comprehensive dataset that encompasses the ancestral genetic background of blaNDM—groS/groL genes and ISCR27 sequence. We also estimate the formation of a secondary transposon, involving Tn3000, which remobilized the region likely in Klebsiella species sometime between the 1980s and early 2000s. However, we suggest Tn3000 likely played a lesser role in the
early spread of blaNDM as it does not include the ISCR27 found at least partially preserved in many samples.

In total, 31 different putative transposons were identified within our dataset. Their role, together with integrons and other transposable elements, is likely mostly minor or disruptive, as suggested for ISCR1. However, we do identify IS26 as of interest, given it frequently forms putative transposons in our dataset, especially in IncF plasmids. IS26 is known for its increased activity and rearrangement of plasmids in clinical isolates and has been observed to drive within-plasmid heterogeneity even in a single E. coli isolate. Thus, IS26-flanked pseudo-composite transposons likely represent the most important contributor to the genetic reshuffling of blaNDM in recent times.

Our assessment of temporal diversity of countries of origin of blaNDM positive isolates supports a globalization peak in 2013–2015. Such a rapid 8–10-year worldwide spread has been suggested for other important mobile resistance genes such as the mcr-1 gene, mediating colistin resistance. The extent to which this model of ‘rapid global spread’ applies to other transposon-borne resistance elements remains to be determined.

We found 33 different plasmid types carrying blaNDM and a positive correlation between genetic distance calculated for differing lengths of plasmid backbones and geographic distances of sampling locations. This observation is consistent with the existence of a constraint on plasmid spread, i.e., plasmid niches, that may exist as a result of local ecological and evolutionary pressures acting on particular plasmid backbones. Such forces may include country boundaries limiting population movement, region-specific patterns in antibiotic usage, influence of co-resistance, plasmid fitness costs, conjugation rates and copy numbers, the narrow host range of the majority of bacterial plasmids, or plasmids being associated with particular locations or environments. All may contribute to restricting plasmid geographical range. Thus, an introduction of another plasmid into a foreign plasmid niche may lead to plasmid loss or fast adaptation by, for instance, acquisition of resistance and other accessory elements. This hypothetical scenario also provides an opportunity for resistance to spread by transposition or recombination, by which a new resistance gene could establish itself into another plasmid niche. In the case of blaNDM, this would also imply that after the initial introduction of blaNDM to a geographic region, dissemination and persistence of the gene could proceed idiosyncratically—selection for carbapenem resistance being just one of many selective pressures acting on plasmid diversity.

The methodological framework we developed reconstructs the immediate up- and downstream backgrounds of the blaNDM gene separately, which is a limitation inherent to our network-based algorithm. In the case of blaNDM, we deem this approach to be satisfactory since the downstream region is dominated by a single putative ancestral background while the alignment rapidly breaks down upstream. For future applications to other mobile genetic elements, it may be worthwhile to jointly resolve genomic variation in both directions. The solution to this problem using the presented methodology does not appear straightforward or computationally efficient. A simple satisfactory alternative would be to select an anchorpoint falling upstream or downstream of the genetic element of interest, rather than starting from the gene of interest as we did in the present study.

The importance of transposon movement has been previously demonstrated by work on plasmid networks, as well as several papers promoting a Russian-doll model of resistance mobility. Considering our results, we suggest a conceptual framework of AMR gene dissemination across genera where plasmid mobility is for the most part restricted. Although plasmids can facilitate rapid spread within species and geographical regions, plasmid transfer is the main driver of widespread dissemination. Instead, most plasmid horizontal transfers are likely only transient, with plasmids generally failing to establish themselves in the new bacterial host, though such aborted plasmid exchanges still provide a crucial opportunity for between-plasmid transposon jumps and genetic recombination to spread AMR genes across bacterial species.

**Methods**

**Compiling the curated dataset of blaNDM sequences.** We compiled an extensive dataset of 6155 bacterial genomes carrying the blaNDM gene from several publicly available databases. A total of 2632, 1156 and 1379 fully assembled genomes were downloaded from NCBI Reference Sequence Database (v1.60), ReSeq (accessed on 15 April 2021), NCBI’s GenBank (accessed on 15 April 2021), and Enterobase (accessed on 27 April 2021), respectively. The Enterobase repository was screened for blaNDM using BlastBlast (v1.0.0) for allowing one mismatch. In addition, we used the Bacterial Genomic Signature Index (BIOSG) tool (v0.3.1) to identify all SRA unassembled reads that carry the blaNDM gene. At the time of writing, a publicly available BIOSG demo did not include sequencing datasets from after December 2016. Therefore, we manually indexed and screened an additional 335,375 SRA bacterial sequencing datasets starting from January 2017 to January 2019. We required the presence of 95% of blaNDM-k-mers to identify NDM-positive samples from raw SRA reads. This led to the inclusion of a further 882 isolates. The dataset also included 104 NDM-positive genomes from 79 hospitalized patients across China and 25 livestock farms selected from two previous studies. These were sequenced using paired-end Illumina (Illumina HiSeq) and paired-end PacBio (PacBio RS II) sequencing reads. Each read was assigned to the Sequence Read Archive (SRA) under accession number PRJNA761884. All reads were de novo assembled using Unicycler (v0.4.8) with default parameters while also specifying hybrid mode for those isolates for which we had both Illumina short-read and PacBio long-read sequencing data. Spades (v3.11.1) was applied, without additional polishing, for cases where Unicycler assemblies failed to resolve.

Assembled genomes were retained when they were derived from a single BioSample identifier. Contigs carrying the blaNDM gene were identified using BLAST (v2.10.1+) with the 40–49-mer hits. Forty-eight contigs were found to carry more than one copy of blaNDM, and were not included in our analyses and 88 contigs were excluded due to having partial (>90%) blaNDM hits. Fourteen assemblies had a single blaNDM gene split into two contigs; these 28 contigs were also excluded. Several contigs were also removed due to poor assembly quality. The filtering resulted in a dataset of 7148 contigs (6155 samples) that were used in all subsequent analyses. Of these, 938 assembled genomes were found to contain blaNDM, on multiple (mostly two) contigs, each harbouring a single and complete copy of blaNDM. Even though the information about sequencing platform or assembly methods of most samples from ReSeq, GenBank and Enterobase databases could not be determined, the distribution of blaNDM-positive contig lengths (Supplementary Fig. 1) reveals they are likely to be based on short reads with the minority of contigs, mostly from ReSeq, reaching the quality of a hybrid de novo assembly. The full table of contigs and metadata considered is available as Supplementary Data 1.

**Quality assessment of SRA blaNDM-positive contigs.** Quality assessment was performed on all blaNDM-positive contigs assembled from samples obtained from SRA database. The phred quality scores and mapping coverage was obtained with BBMap tools. Overall, the blaNDM-positive contigs showed good mapping coverage predominantly exceeding per base coverage of 30x (Supplementary Fig. 3a). Similar coverage was obtained for SRA contigs used in molecular dating (Supplementary Fig. 3b). Phred score was used to assess the quality of SNP calls found in SRA contigs used in molecular dating analysis. A total of three variable positions (i.e., SNPs) were found within 30 SRA contigs compared to the inferred ancestral sequence (Supplementary Fig. 2c, d). In ten contigs, two of these three variable positions were found, to have slightly lower however still acceptable (~20), Phred quality score.

**Annotating the dataset.** Full metadata for each genome was collected from its respective database and the R package ‘taxize’ was used to retrieve taxonomic information for each sample. In the case of samples for which exact sampling coordinates were not provided, GeoCoding API from Google cloud computing services was used to retrieve coordinates based on location names.

Coding sequences of all NDM-positive contigs were annotated using the annotation tool Prokka (v1.14.6) and Rosary (v3.13.0) with minimum blastp similarity percentage of 90% (±0.9) and disabled paralog splitting. To identify plasmid replicon types, contigs were screened against the PlasmidFinder replicon database, reaching 2020-02-23) using BlastBlast (v2.10.1+) with only BLAST hits with a minimum coverage of 80% and percentage identity of ≥95% were retained. In cases where two or more replication hits were found at overlapping positions on a contig, the one with the higher percentage identity was retained. Identified plasmid types were used to cluster contigs into broader plasmid groups: IncX3, Inc, IncC, IncC2, IncF, IncFII, IncHI2, IncHI1B and other (Fig. 1d).

NDM-positive contigs were also screened against a dataset of complete bacterial plasmids. Bacterial plasmid references were obtained from ReSeq and curated to

**NATURE COMMUNICATIONS** | https://doi.org/10.1038/s41467-022-28819-2 | www.nature.com/naturecommunications

(2022)13:131 | https://doi.org/10.1038/s41467-022-28819-2 | www.nature.com/naturecommunications
include plasmids from a bacterial host and with a sequence description, which implies a complete plasmid sequence (regular expression term used: \( \text{plasmid}^* \text{complete} \text{sequence}^* \)). MinHash based genome distance estimator\cite{indexing} was applied with default settings to evaluate pairwise genetic distances between contig sequences and plasmid references. Contig-reference hits with less than 0.05 Mash distance and less than 20% difference in length were retained. A table of pairwise genetic distances between contigs and references was represented as a network that was then analyzed with the informap\cite{informap} community detection algorithm implemented in igraph\cite{igraph} R package. Contigs were annotated according to their community membership and the network was visualized using Cytoscape\cite{cytoscape} (Fig. 2).

### Resolving structural variants of NDM-positive contigs

A novel alignment-based approach was used to identify stretches of homology (i.e., maximal alignable regions) as well as structural variations upstream and downstream of \( \text{bla} \text{NDM} \) gene. A conceptual illustration of the method is presented in Fig. 3. First, contigs carrying \( \text{bla} \text{NDM} \) were reoriented such that the \( \text{bla} \text{NDM} \) gene was located on the positive-sense DNA strand (i.e., facing 5' to 3' direction).

A discontinuous Mega BLAST (v2.10.1)\cite{MegaBLAST} search with default settings was then applied against all pairs of retained contigs. This method was selected over the regular Mega BLAST implementation as it is comparably faster, but more permissive towards dissimilar sequences with frequent gaps and mismatches. BLAST hits including a complete \( \text{bla} \text{NDM} \) gene represent maximal stretches of homology around the gene for every pair of contigs. The analysis continues by considering only the position (i) the start of the \( \text{bla} \text{NDM} \) gene and the \( \text{bla} \text{NDM} \) upstream sequence or (ii) the end of the \( \text{bla} \text{NDM} \) gene and the upstream sequence depending on the analysis at hand: the downstream or the upstream analysis, respectively. This trimming of BLAST hits establishes \( \text{bla} \text{NDM} \) as an anchor and enables comparisons to be made across all contigs. A table of BLAST hits can be considered as a network (graph), where each pair of contigs (i.e., nodes) are connected by the edge weighted by the length of the BLAST hit. The algorithm proceeds with a stepwise network analysis of BLAST hits. For this purpose, a ‘splitting threshold’ was introduced. Starting from zero that represents the start/end position of \( \text{bla} \text{NDM} \) gene, the threshold is gradually increased by 10 bp at each step. BLAST hits with a length shorter than the value given by the ‘splitting threshold’ are excluded. Thus, as the ‘splitting threshold’ increases, a network of BLAST hits is also pruned and broken down into components—groups of interconnected nodes (contigs). It is expected that contigs within each component share a homologous region downstream (or upstream) of \( \text{bla} \text{NDM} \) gene given by the threshold. It is therefore not possible for a single contig to be assigned to multiple components. Components of size <10 are labelled as ‘Other Structural Variants’. Also, contigs that are shorter than the ‘splitting threshold’ and share no edge with any other contig are considered as ‘cutting short’. Similar clustering algorithms have been previously described, e.g., CAST\cite{CAST}, but were used for a different application.

By tracking the splitting of the network as the ‘splitting threshold’ is increased, one can determine clusters of homologous contigs at any given position downstream or upstream from the anchor gene (here \( \text{bla} \text{NDM} \)), as well as the homology breakpoint. The precision of the algorithm is directly influenced by the step size of the threshold and the alignment algorithm, in this case the discontinuous Mega BLAST. We assessed the precision of the algorithm on the tree of structural variations downstream of \( \text{bla} \text{NDM} \) (Fig. 4). To this end, we compared extended 30 bp sequence fragments of each branching point in the tree checking for missed homologies and comparing Mash distances between pairs of branched-out contigs.

We found no similarities among 50 bp fragments of any split branches. The described algorithm is available at https://github.com/macman123/track_structural_variants.

### Analyzing the contig overhanging reads

To investigate the reasons behind a number of distinctively short \( \text{bla} \text{NDM} \)-carrying contigs, we mapped 781 raw Illumina paired-end sequencing reads (originally downloaded from SRA) back to their matching contigs. The mapping was done using BWA\cite{BWA} (v0.7.17) and minindel = 0 and memindel = 0. Within the output file, only the overhanging reads with the CIGAR string matching the [0-9]*[M|I|D|S]* regular expression were selected. All overhangs of reads ≥50 bp were screened against ISFinder database\cite{ISFinder}.

### Molecular tip-dating analysis

The 112 complete \( \text{Tn}125 \) and 73 complete \( \text{Tn}3000 \) contigs with a known collection date and harbouring \( \text{bla} \text{NDM} \) were sequentially aligned (‘pilup flag) using Clustal Omega (v1.2.3)\cite{ClustalOmega} specifying the \( \text{bla} \text{NDM} \) sequence (FN396876.1) as a profile. Each alignment was manually inspected using UGENE\cite{UGENE} (v0.8.25) QualiCut and the alignment algorithm, in this case the Smith-Waterman algorithm, was used for the final inspection of the alignments did not reveal any error-prone site; thus, no site masking was implemented in molecular tip-dating analysis. The ancestral (i.e., root) sequence was determined by evaluating SNP frequencies over time (Supplementary Figs. 9 and 10). Due to a short sampling time span and relatively few mutations observed, it is unlikely that any non-ancestral SNP has become dominant in the population. Therefore, we expect the ancestral sequence to have a higher SNP frequency in earlier years.

We find that, in all but two cases, the consensus sequence of an alignment displays this behaviour. The first exception is the consensus sequence allele of \( \text{Tn}125 \) at the variable position 441 (Supplementary Fig. 9). This allele has a low frequency in 2009. However, by inspecting the allele frequency table, we observe the low frequency is based on a single sample. Leaving out this early sample restores the desired frequency pattern; hence, the consensus allele is considered ancestral in this case. The second exception is the variable position 449 in the case of the \( \text{Tn}3000 \) alignment (Supplementary Fig. 10). The consensus allele ‘a’ is not found in the early sample from 2009. Both allele ‘t’, present in the early sample, and allele ‘a’ are found equally frequent in more recent samples. Thus, due to lack of other evidence, allele ‘t’ was considered ancestral. Determined ancestral sequences were used to evaluate temporal signal in the alignment, and in the subsequent rooting of phylogenetic trees.

### Data availability

The accession numbers of bacterial genomes obtained from the RefSeq, Enterobase and SRA databases are given in the Supplementary Data 1. One hundred and four paired-end
References

1. Wu, W. et al. NDM metallo-β-lactamases and their bacterial producers in health care settings. Clin. Microbiol. Rev. 32, e00115–e00118 (2019).

2. Yong, D. et al. Characterization of a new metallo-β-lactamase gene, bla NDM-1, and a novel erythromycin esterase gene carried on a unique genetic structure in Klebsiella pneumoniae sequence type 14 from India. Antimicrob. Agents Chemother. 53, 5046–5054 (2009).

3. Struelens, M. J. et al. New Delhi metallo-beta-lactamase 1-producing Enterobacteriaceae: emergence and response in Europe. Eurosurveillance 15, 197106 (2010).

4. Kumarasamy, K. K. et al. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infect. Dis. 10, 597–602 (2010).

5. Poirel, L., Dortet, L., Bernabéu, S. & Nordmann, P. Genetic features of blaNDM-1-positive Enterobacteriaceae. Antimicrob. Agents Chemother. 55, 5403–5407 (2011).

6. Castanheira, M. et al. Early dissemination of NDM-1- and OXA-181-producing Enterobacteriaceae in Indian hospitals: report from the SENTRY Antimicrobial Surveillance Program, 2006–2007. Antimicrob. Agents Chemother. 56, 1274–1278 (2012).

7. Jones, L. S. et al. Plasmid carriage of blaNDM-1 in clinical Acinetobacter baumannii isolates from India. Antimicrob. Agents Chemother. 58, 4211–4213 (2014).

8. Roca, I. et al. Molecular characterization of NDM-1-producing Acinetobacter pittii isolated from Turkey in 2006. J. Antimicrob. Chemother. 69, 3437–3438 (2014).

9. Poirel, L., Bonnin, R. A. & Nordmann, P. Analysis of the resistome of a multidrug-resistant NDM-1-producing Escherichia coli strain by high-throughput genome sequencing. Antimicrob. Agents Chemother. 55, 4422–4429 (2011).

10. Toleman, M. A., Spencer, J., Jones, L. & Walsh, T. R. blaNDM-1 is a chimera likely constructed in Acinetobacter baumannii. Antimicrob. Agents Chemother. 56, 2773–2776 (2012).

11. Partridge, S. R. & Fredjell, J. R. Genetic contexts of bla NDM-1. Antimicrobial Agents Chemotherapy. 56, 6065–6067 (2012).

12. Toleman, M. A., Bennett, P. M. & Walsh, T. R. ISCR elements: novel gene-capturing systems of the 21st century? Microbiol. Mol. Biol. Rev. 70, 296–316 (2006).

13. Ilyina, T. S. Mobile ISCR elements: structure, functions, and role in emergence, increase, and spread of blocks of bacterial multiple antibiotic resistance genes. Mol. Genet., Microbiol. Virol. 27, 135–146 (2012).

14. Poirel, L. et al. Tn125-related acquisition of blaNDM-like genes in Acinetobacter baumannii. Antimicrob. Agents Chemother. 56, 1087–1089 (2012).

15. Sekizuka, T. et al. Complete sequencing of the blaNDM-1-positive IncAC plasmid from Escherichia coli ST38 isolate suggests a possible origin from plant pathogens. PLoS One 6, e25334 (2011).

16. McArthur, A. G. et al. The comprehensive antibiotic resistance database. Antimicrob. Agents Chemother. 57, 3348–3357 (2013).

17. Basu, S. Variants of the New Delhi metallo-β-lactamase: new kids on the block. Future Microbiol. 15, 465–467 (2020).

18. Baraniak, A. et al. NDm-producing Enterobacteriaceae in Poland, 2012-14: inter-regional outbreak of Klebsiella pneumoniae ST11 and sporadic cases. J. Antimicrob. Chemother. 71, 85–91 (2016).

19. Rahman, M. et al. Prevalence and molecular characterization of New Delhi metallo-beta-lactamases in multidrug-resistant Pseudomonas aeruginosa and Acinetobacter baumannii from India. Microb. Drug Resist. 24, 792–798 (2018).

20. Hu, H. et al. Novel plasmid and its variant harboring both a blaNDM-1 gene and type IV secretion system in clinical isolates of Acinetobacter baumannii. Antimicrob. Agents Chemother. 56, 1698–1702 (2012).

21. Yang, Q. et al. Dissemination of NDM-1-producing Enterobacteriaceae mediated by the IncX3-type plasmid. PLoS One 10, e0129455 (2015).

22. Watan, A. M. et al. Genetic contexts of blaNDM-1 in patients carrying multiple NDM-producing strains. Antimicrob. Agents Chemother. 59, 7405–7410 (2015).

23. Rasheed, J. K. et al. New Delhi metallo-β-lactamase-producing enterobacteriaceae, United States. Emerg. Infect. Dis. 19, 870–878 (2013).

24. Campos, J. C. et al. Characterization of Tri3000, a transposon responsible for blaNDM-1 dissemination among Enterobacteriaceae in Brazil, Nepal, Morocco, and India. Antimicrob. Agents Chemother. 59, 7387–7395 (2015).

25. Feng, Y., Liu, L., McNally, A. & Zong, Z. Coexistence of two blaNDM-5 genes on an IncN plasmid as revealed by nanopore sequencing. Antimicrob. Agents Chemother. 62, e00110–e00118 (2018).

26. Zhao, Q.-Y. et al. IS62 is responsible for the evolution and transmission of blaNDM-harboring plasmids in Escherichia coli of poultry origin in China. mSystems 6, e006462 (2021).

27. Lynch, T. et al. Molecular evolution of a Klebsiella pneumoniae ST278 isolate harboring blaNDM-7 and involved in nosocomial transmission. J. Infect. Dis. 214, 798–806 (2016).

28. Huang, T. W. et al. Copy number change of the NDM-1 sequence in a blaNDM-bearing contig is available on Figshare (https://doi.org/10.5522/04/16594784). Source Data are provided with this paper.

Received: 27 September 2021; Accepted: 14 February 2022; Published online: 03 March 2022
51. Sohn, J. I. & Nam, J. W. The present and future of de novo whole-genome assembly. Bioinformatics 19, 23–40 (2018).
52. Harmey, C. J. & Hall, R. M. An analysis of the IS6110 family of insertion sequences: is it a single family? Microb. Genomics 5, e000291 (2019).
53. He, S. et al. Insertion sequence IS6110 reorganizes plasmids in clinically isolated multidrug-resistant bacteria by replicative transposition. MBio 6, 1–14 (2015).
54. He, D. D. et al. Antimicrobial resistance-encoding plasmid clusters with heterogeneous MDR regions driven by IS6110 in a single Escherichia coli isolate. Antimicrob. Chemother. 74, 1511–1516 (2019).
55. Wang, R. et al. The global distribution and spread of the mobilized colistin resistance gene mcr-1. Nat. Commun. 9, 1–9 (2018).
56. Acman, M., van Dorp, L., Santini, J. M. & Balloux, F. Large-scale network analysis captures biological features of bacterial plasmids. Nat. Commun. 11, 1–11 (2020).
57. Shaw, L. et al. Niche and local geography shape the pan-genome of wastewater-and livestock-associated Enterobacteriaceae. Sci. Adv. 7, 15 https://doi.org/10.1126/sciadv.abj3668 (2021).
58. Redondo-Salvo, S. et al. Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. Nat. Commun. 11, 1–13 (2020).
59. Sheppard, A. E. et al. Nested Russian doll-like genetic mobility drives rapid dissemination of the carbapenem resistance gene blaKPC. Antimicrob. Agents Chemother. 60, 3767–3778 (2016).
60. Pruiti, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35, D61–D65 (2007).
61. Luhmann, N., Holley, G. & Achtman, M. BlastFrost: fast querying of biomolecular interaction networks. Bioinformatics 31, 22 (2017).
62. Bradley, P., den Bakker, H. C., Rocha, E. P. C., McVean, G. & Iqbal, Z. Ultrafast search of all deposited bacterial and viral genomic data. Nat. Biotechnol. 37, 152–159 (2019).
63. Wang, R. et al. The prevalence of colistin resistance in bacteria revealed by a global map of their plasmids. Mol. Biol. Evol. 37, 2068–2069 (2021).
64. Wang, Q. et al. Phenotypic and genotypic characterization of Carbapenem-resistant Enterobacteriaceae: data from a longitudinal large-scale CRE study in China (2012–2016). Clin. Infect. Dis. 67, S196–S205 (2018).
65. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13, 1–22 (2017).
66. Buskirk, J. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477 (2012).
67. Bushnell, B. BBMap: a fast, accurate, splice-aware aligner. https://www.osti.gov/biblio/1241166 (2014).
68. Chamberlain, S. A. & Szöcs, E. taxize: taxonomic search and retrieval in R. F1000Research 3, 1–11 (2014).
69. Seemann, T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069 (2014).
70. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31, 3691–3693 (2015).
71. Orlek, A. et al. Plasmid classification in an era of whole-genome sequencing: application in studies of antibiotic resistance epidemiology. Front. Microbiol. 8, 1–10 (2017).
72. Acman, M., van Dorp, L., Santini, J. M. & Balloux, F. Large-scale network analysis captures biological features of bacterial plasmids. Nat. Commun. 11, 1–11 (2020).
73. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 17, 1–14 (2016).
74. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. Proc. Natl Acad. Sci. USA. 105, 1118–1123 (2008).
75. Csardi, G. & Nepusz, T. The Igraph software package for complex network. Res. Interjournl, complex Syst. 1095.5, 1–9 (2006).
76. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).
77. Ben-Dor, A., Shamir, R. & Yakhini, Z. Clustering gene expression patterns. 6, 281–297, https://home.ierubertpub.cmn (2004).
78. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539 (2011).
