Nucleic Acids Research, 2026, gkag1163
https://doi.org/10.1093/nar/gkag1163
Database issue

proMGE2—a planetary-scale resource of mobile genetic elements in microbial (meta)genomes

Anastasiia Grekova [1,2,†], Christian Schudoma [2,†], Ivica Letunic [3], Chan Yeong Kim [2], Daniel Podlesny [2], Anthony Fullam [2], Thomas S.B. Schmidt [2,4], Askarbek Orakov [2,5], Daniel Mende [6], Jonas Schiller [2,7], Katarina O.E. Erbstein [2,8], Bartosz Bartmanski [9], Maria Zimmermann-Kogadeeva [9], Michael Zimmermann [2], Michael Kuhn [2,*], Peer Bork [2,7,10], Supriya Khedkar [2,11,*]

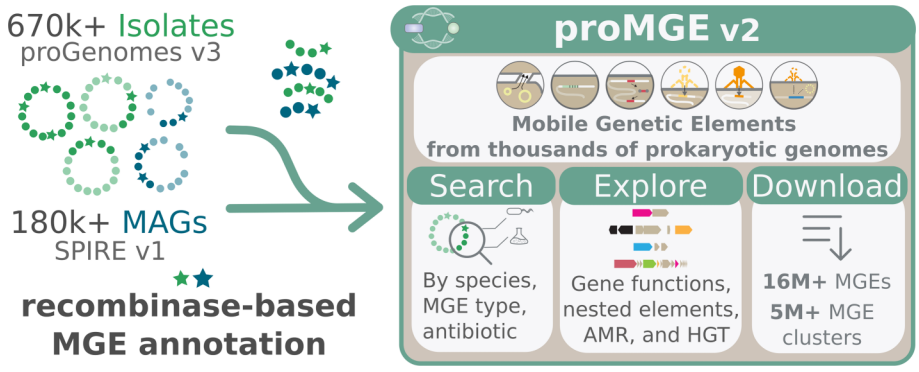
[1]Technical University of Munich, School of Life Sciences, 85354 Freising, Germany
[2]European Molecular Biology Laboratory, Molecular Systems Biology Unit, 69117 Heidelberg, Germany
[3]biobyte solutions GmbH, Bothestr 142, 69126 Heidelberg, Germany
[4]Present address: APC Microbiome Ireland, University College Cork, College Road, Cork T12 K8AF, Ireland
[5]Present address: Department of Biostatistics, Harvard T.H. Chan School of Public Health, MA, 02115 Boston, United States
[6]Human Biology–Microbiome–Quantum Research Center (WPI-Bio2Q), Keio University, Tokyo 160-8582, Japan
[7]Department of Bioinformatics, Biocenter, University of Würzburg, 97074 Würzburg, Germany
[8]Mathematisch-Naturwissenschaftliche Fakultät, Eberhard Karls Universität Tübingen, 72076 Tübingen, Germany
[9]European Molecular Biology Laboratory, Genome Biology Unit, 69117 Heidelberg, Germany
[10]Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany
[11]Center for Quantitative Analysis of Molecular and Cellular Biosystems (BioQuant), Heidelberg University, 69120 Heidelberg, Germany

*To whom correspondence should be addressed. Email: khedkar@bioquant.uni-heidelberg.de
Correspondence may also be addressed to Michael Kuhn. Email: mkuhn@embl.de
†The first two authors should be regarded as Joint First Authors.

Abstract

proMGE is a resource of prokaryotic mobile genetic elements (MGEs) that systematically catalogues MGEs from prokaryotes, their intercellular spread, antimicrobial resistance gene carriage, and MGE nesting interactions. We present proMGE2 featuring a 6X expansion, a total of >16.4 million MGEs from 670k high-quality isolate genomes, and 180k metagenome-assembled genomes (MAGs), spanning diverse MGE categories: transposons and insertion sequences, phages, phage-like elements, conjugative elements, mobility islands, and integrons. proMGE2 includes MGEs from MAGs derived from environmental microbiomes, encompassing ~50k samples across several major habitat types. proMGE2 provides a standardized MGE reference collection of >5 million non-redundant MGE sequences drawn from our curated database, and enables identification of MGEs in user-provided genomes. proMGE2 captures MGE diversity and their functional potential at unprecedented scale to support future research of mobile genetic elements and transfer of genetic material in microbes. The database is available at http://promge.embl.de/.

Graphical abstract



Received: September 15, 2026. **Revised:** October 19, 2026. **Accepted:** October 20, 2026
© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Mobile genetic elements (MGEs) are important drivers of survival and adaptation in microbes [1–5]. A wealth of experimental and computational data establishes their role in intercellular transfer of DNA and their effect on bacterial evolution [3, 4, 6, 7]. The genes transferred by MGEs (often referred to as cargo genes) can confer adaptive traits, e.g. affecting human health through antibiotic resistance genes that promote survival of infection-causing bacteria during antibiotic treatment [8–12]. Mobile genetic elements can carry genes encoding xenobiotic-degrading enzymes, thereby facilitating bioremediation in anthropogenically impacted environments, as well as genes relevant to biogeochemical cycles such as nitrogen fixation [13–17]. Moreover, MGEs drive bacterial community dynamics by carrying genes promoting competition and symbiosis [18]; these include toxin genes [5] and genes that facilitate mutualistic interactions among bacteria [19] respectively.

MGEs include diverse element types—phages, plasmids, and intergrative conjugative elements (ICEs) that can actively transfer inter-cellularly and also insertion sequences (IS), transposons, integrons, and other mobile elements that depend on phages, plasmids and ICEs for their transfer [20]. Altogether, mobile genetic elements exhibit tremendous diversity in terms of their mobility, selfishness, gene composition, length, transfer mechanisms, prevalence, and abundance [21, 22]. MGEs evolve fast, are prone to gene gains and losses, are mobile both vertically (to offsprings) and horizontally (to different organisms), and encode modules (sets of genes) associated with mobility and/or auxiliary functions [23]. MGE types that differ significantly in terms of their intercellular transfer machinery, genomic integration sites [24], or copy numbers [25, 26] can share modules [27]. This sharing and combination of different modules forms a continuum of mosaic MGEs [21, 27–29]. These features pose a challenge for both the identification of MGEs in genomic datasets and the disentangling of different MGE types [30].

Several computational methods and tools exist to annotate MGEs in genomic data. These capture experimentally well-studied MGEs [31] and typically focus on one mobile element type, for example IS elements [32], transposons [33, 34], plasmids [35], bacterial viruses [36, 37], ICEs [38–41], or integrons [42]. Therefore, it is often cumbersome to integrate outputs of these different tools [43] and they are typically not sufficient to capture MGE interactions mediated by hitchhiking such as nesting of transposons on plasmids, known to drive the spread of antibiotic resistance genes [44]. Detecting MGEs in metagenomes is a developing field focusing mainly on the identification of bacterial viruses [45] and plasmids [36]. Two major issues associated with metagenomic assemblies, namely genome incompleteness and contamination, make MGE identification in metagenome-assembled genomes (MAGs) challenging [46, 47]. As previously established [22], the need persists for a unified annotation and classification of all MGE types in order to enable and facilitate comprehensive comparative analyses of MGEs in the growing set of available microbial isolate genomes and MAGs.

Here, we present an updated version of proMGE, a comprehensive resource of MGEs in prokaryotic genomes. proMGE2 facilitates studying the diversity, phylogenetic distribution, nesting interactions, and cargo gene functions of MGEs from isolate genomes and MAGs from environmental microbiomes. The resource provides high-resolution MGE annotations for six mobile element categories: transposable elements and insertion sequences (IS_Tn), phages, phage-like elements, conjugative elements (CE), mobility islands (MI), and integrons. proMGE2 is the first MGE resource to systematically catalogue different MGE types from MAGs. Compared to the first version, we expanded the number of annotations from 2 million to over 16.4 million MGEs derived from >670k isolate genomes from proGenomes3 (PG3) [48] and >180k high-quality MAGs derived from host-associated and environmental microbiomes encompassing almost 50k samples from over 500 studies across several major habitat types in SPIRE v1 (SPIRE) [49]. The improved search functionality stratifies species by isolate genome and MAG, supporting comparative investigation of MGEs based on genome source. For all catalogued MGEs, proMGE2 continues to provide functionality to investigate horizontal transfer potential, antimicrobial resistance (AMR) gene carriage, and nesting interactions among reported elements. As a standardized reference collection, proMGE2 provides a non-redundant set of genomic sequences for >5 million MGEs. In addition to the existing interoperability with PG3 [48], SMART [50], and the Comprehensive Antibiotic Resistance Database (CARD) [51], this version integrates MAGs from SPIRE [49], leveraging environmental data, and habitat descriptions for individual MAGs, as well as MGE visualisation through SPIRE's interactive MGE viewer. With its improved annotation feature, proMGE2 now supports annotation of MGEs in user-provided genomes (isolates and MAGs) and genomic regions.

Database construction

Isolate genomes and MAGs

We utilized >670K isolate genomes from PG3 and >180K MAGs from SPIRE for MGE annotation (Supplementary Table S1). Only high-quality isolate genomes and MAGs were included (completeness ≥90% and contamination ≤5%) [52]. We obtained species information from the native PG3 and SPIRE species assignment (Supplementary Table S2). MAGs were more fragmented and in general shorter than isolates (Supplementary Fig. S1A and B). We only included genomes with resolved GTDB version r207 [53] species level taxonomy, as unresolved genomes were not suitable for subsequent pangenome construction (Supplementary Fig. S2C). All included genomes passed the GUNC [54] check to ensure taxonomically consistent lineages and avoid chimeric genomes (Supplementary Table S1).

The collection of MAGs passing the above quality controls spans prokaryotic diversity from >6K species, originating from 47 242 samples, 545 studies covering a wide array of microbial environments across Earth. To ensure consistent habitat annotation based on the sample's taxonomy we made use of the habitat clustering framework by Kim, Podlesny and Schiller *et al.* [55] and assigned the used SPIRE samples to one of the 40 habitat clusters ranging from human and host-associated to extreme environments.

MGE detection and annotation

To annotate mobile genetic elements in isolate genomes we applied our previously developed framework [22]. Briefly, for every isolate genome, we identified mobile element recombinase marker genes using 68 high accuracy profile Hidden Markov Models (HMMs) for MGE-specific recombinases [56] and

computed species pangenomes for species with at least two genomes in PG3 (pg3-pangenome) (Supplementary Fig. S2B). Using this pangenome information we defined recombinase islands (MGE boundaries) following the principle that an MGE is a region of consecutive accessory genes containing a mobile element recombinase (see supporting materials, method section for details) [57, 58]. Based on the association of recombinases with known MGE types we assigned the identified recombinase islands to six MGE categories: phage, phage-like, insertion sequences or transposons (IS_Tn), integrons, CE, or MI.

For annotating MGEs in MAGs we adapted the above framework as described below.

1) As for isolate genomes, we annotated recombinase marker genes using 68 previously described mobile element recombinase profile HMMs [22, 56] using HMMER [59] and identified 3 661 831 recombinase genes in ∼180K MAGs.
2) To infer MGE boundaries using our pangenome approach, we focused on MAGs whose species were represented in pg3-pangenome, in order to ensure inclusion of species with at least two member genomes (isolates or MAGs). To account for accessory genome inflation commonly encountered in MAGs [60] (Supporting material, Methods, and Supplementary Fig. S2B) due to their high incompleteness and fragmentation compared to isolate genomes (Supplementary Fig. S1) we did the following: for each MAG in a species cluster, we classified genes as core if they were present in >80% of the genomes in the cluster (compared to presence in 95% of genomes for isolate genomes), and the remaining genes as accessory (see Supporting material and Methods). Based on this pangenome information we determined MGE boundaries as described above for isolate genomes. Higher accessory-gene fractions were not correlated with genome size in both isolates and MAGs (Supplementary Fig. S2E). The median accessory-gene fraction was higher in MAGs than in isolate genomes belonging to the same species cluster (Supplementary Fig. S2D). Approximately 30% of MAGs (n = 63 700) compared to ∼20% of isolate genomes (n = 177 470) showed an extreme accessory gene content (≤0.2 or ≥0.8) (Supplementary Table S1), (Supplementary Fig. S2A). These mainly belonged to species with few sequenced genomes (Supplementary Fig. S2B and C) and were filtered out from our dataset (Supplementary Table S1). For the remaining 185 314 MAGs, we defined recombinase islands as described above for isolate genomes.
3) We assigned MGE categories to each recombinase island based on the subfamily of recombinase marker genes as for isolate genomes. For recombinase islands containing phage- or CE-specific recombinases, we used additional evidence, in particular annotations of phage structural genes or, respectively, conjugation machinery genes, as described previously [22]. Phage structural genes were identified by parsing eggNOG-mapper v2.1.2 [61] output for specific terms associated with phage structural genes and excluding potential false positive hits. Conjugation machinery genes were detected by parsing the output of MacSyFinder v2.1.4 with CONJScan models v2.1 (https://github.com/macsy-models/CONJScan, commit d5fc1e3) [62–64] for the presence of CONJ type systems. We also used this information to accurately disentangle the fraction of MGE categories that share recombinases. Recombinase islands harboring a cellular recombinase were categorized as "cellular" like in proMGE version 1 [22].

The final MGE category assignment integrated the recombinase signal together with phage and conjugation machinery signals (where applicable) and performed rule-based assignment as described previously [22]. The MGE categories are based on known MGE types: IS_Tn (transposable elements) include insertion sequences and transposons; phage and phage-like elements correspond to integrated phages and their derivatives; CE (conjugative elements) include conjugative plasmids and integrative conjugative elements (ICEs); MI (mobility islands) include MGEs that encode CE associated recombinases or recombinases common to both phages and CEs that cannot be resolved based on machinery genes; the integron category corresponds to integrons. We defined an additional category called "nested" elements to describe recombinase islands that can harbor several recombinases; these often indicate nesting of different MGE categories [21]. For example, "ce:1; is_tn:1" is an IS element/transposon embedded in a conjugative element that can enable intercellular mobility of IS_Tn [21, 44, 65]. MGEs found in MAGs tended to be longer than those recovered from isolate genomes, potentially due to their boundaries coinciding more frequently with the contig boundaries, which is a possible consequence of high fragmentation and/or high accessory gene content of MAGs (Supplementary Fig. S3A and B). Hence, we introduced confidence levels that capture whether the MGE boundary is derived from pangenome or from a contig boundary. The confidence level is calculated for every MGE and ranges from 1 to 3 (1 being the lowest and 3 highest). It scores, presence of recombinase (+1); pangenome size >5 (+1); MGE boundary flanked by a gene and not a contig boundary on the 5′ end (+0.5), and/or on the 3′ end (+0.5) (Supplementary Fig. S3C). Our framework detects DNA-based, recombinase-encoding MGEs associated with prokaryotic genomes and metagenome-assembled genomes. We implemented the MGE annotation framework as a Nextflow pipeline ("proMGE-flow") to ensure reproducibility and robustness. The workflow is hosted on the CloWM platform [66].

Antibiotic resistance gene annotation

To annotate AMR genes, we utilised predictions from abricate v1.0.1 run using default parameters (https://github.com/tseemann/abricate) [67] against the MEGARes database 2.0.0 [68]. We integrated AMR annotations with Antibiotic Resistance Ontology (ARO) terms [51] and assigned antibiotic resistance classes using argNorm version 1.1.0 (https://github.com/BigDataBiology/argNorm) [69].

MGE clustering and putative horizontal gene transfer detection

To build a non-redundant MGE dataset, we combined all MGEs from MAGs and isolate genomes into a single dataset. We then calculated MD5 hashes for each sequence and its reverse complement and grouped sequences with matching hashes into 100% sequence identity clusters. For each such

Table 1. Overview of the mobile genetic elements and taxonomy of their hosts covered by proMGE version 2

Dataset	IS_Tn	Phage	Phage like	CE	Integron	MI	Cellular	Nested	Total MGEs [a]	Genomes	Species clusters
Isolates	7 370 860	741 511	1 743 595	646 582	29 399	1 107 932	1 141 054	2 386 508	14 026 387	673 876	5143
Predicted AMR (Isolates)	271 600	4932	13 601	57 523	15 025	16 097	8	246 006	624 784	185 617	1103
Predicted HGT (Isolates)	232 108	267	7745	4742	2463	20 190	0	NA	267 515	168 570	799
MAGs	992 607	55 159	333 696	227 692	5815	561 066	235 128	250 330	2 426 365	181 447	7023
Predicted AMR (MAGs)	2253	210	339	3182	24	1517	30	2448	9973	6809	752
Predicted HGT (MAGs)	1650	2	228	420	1	654	0	NA	2955	2803	463
All	8 363 467	796 670	2 077 291	874 274	35 214	1 668 998	1 376 182	2 636 838	16 452 752	855 323	10 615
MGE100 clusters	2 168 187	318 067	669 275	450 952	13 505	763 217	246 844	1 033 915	5 419 060 [b]	NA	NA

[a] without cellular
[b] 1942 unassigned

cluster, we only retained one representative, establishing a final, non-redundant dataset (MGE100). For hybrid MGE100 clusters, i.e. those comprising MGEs recovered from both isolates and MAGs, we chose isolate-derived MGEs as representatives (Supplementary Fig. S4A and C). For a minority of the MGEs that were part of the same MGE100 cluster, we observed differences in their gene calls (varying gene lengths and/or numbers). These gene predictions were performed (at genome level) on the respective genomes from which the MGEs in the cluster originated. The resulting gene call inconsistencies resulted in differences in functional annotations of machinery genes leading to inconsistent MGE assignments among members of the same cluster. For example, the annotation of a phage structural gene led to an element being classified as a phage, while cluster members lacking this annotation were classified as phage-like. MGE100 clusters with such conflicts constitute 0.02% of MGE100 clusters (1942 clusters), corresponding to ∼0.47% of total MGEs (∼80k MGEs). These clusters were labelled as "unassigned."

Horizontal gene transfer (HGT) of mobile genetic elements was defined as the occurrence of members of the same MGE100 cluster in genomes belonging to different taxonomic families, indicating lower limits of recent horizontal transfer across higher taxonomic levels.

MGE annotation of user-submitted genomes

To enable MGE annotation in a user-provided genome (isolate or MAG) and contextualize the mobilome in our comprehensive MGE catalogue, we implemented a new "Annotate" functionality. Now, users can provide a FASTA file of a high-quality isolate or MAG to annotate MGEs belonging to different MGE categories and obtain the complete MGE profile in GFF format. The typical annotation time for an average-sized input genome is ∼1 h and the results can be downloaded directly from the "Annotate" page.

Database content

Global distribution of mobile genetic elements

proMGE2 is a comprehensive resource of the global distribution of MGEs from ∼850k genomes, >10k species, and >500 studies (Table 1). It represents MGEs across a broad phylogenetic and environmental diversity of prokaryotic isolate genomes and MAGs (Fig. 1A and B, Supplementary Figs. S4B and S5). Most of the MGE diversity in proMGE2 is represented by multiple copies of insertion sequences and transposons (Supplementary Fig. S4A), (Table 1). The identified ∼8.3 million IS elements and transposons constitute the most dominant MGE category across isolate genomes (∼52%) and MAGs (∼41%). The other MGE categories (∼800k phages, ∼2 million phage-like elements, ∼870k CEs, ∼1.7 million MIs, and ∼35k integrons) show distinct patterns depending on genome source. For example, compared to isolates, MAGs are annotated with higher numbers of MIs and phage-like elements. In addition, proMGE2 contains ∼2.6 million nested elements (Table 1), exhibiting co-localisation of MGE categories. In particular, this includes hitchhiking of integrons, IS elements, and transposons with each other and other MGE categories. Nested elements are the most prevalent type after IS elements and transposons, representing ∼16% of MGEs within proMGE2 and showcasing the prevalence of MGE-MGE co-localisation and hitchhiking interactions across prokaryotic (meta)genomes. Furthermore, the database enables exploring MGEs from diverse environmental MAGs from disparate habitats, unlocking comparisons of the mobilome of thousands of samples around the globe (Fig. 1A). Redundancy reduction of MGEs revealed high MGE sequence redundancy in the isolate mobilome when compared to MAGs (Supplementary Fig. S4A), resulting in ∼5 million non-redundant MGE sequences or MGE100 clusters. This is in-line with the previously reported lower phylogenetic diversity of isolate genomes compared with MAGs [70]. MGE100 clusters contextualize MGE abundance across genomes, taxa, and samples and enable comparative analysis of potential MGE-mediated horizontal gene transfer events across higher taxonomic levels (Table 1).

Functional repertoire of MGEs

Alongside MGE annotations, proMGE2 provides functional annotations of the MGE-encoded genes (Fig. 2A and B) and classifies them into predicted machinery (recombinase genes—MGE detection markers, phage structural genes, and conjugation machinery genes) (Fig. 2C) and the rest as cargo genes. This enables the comparison of not only the distribution of MGE categories across habitats, but also the functional composition of genes encoded within different mobile elements (Fig. 3).

MGE categories were defined primarily by recombinase evidence and refined using genomic-context information. Hence,

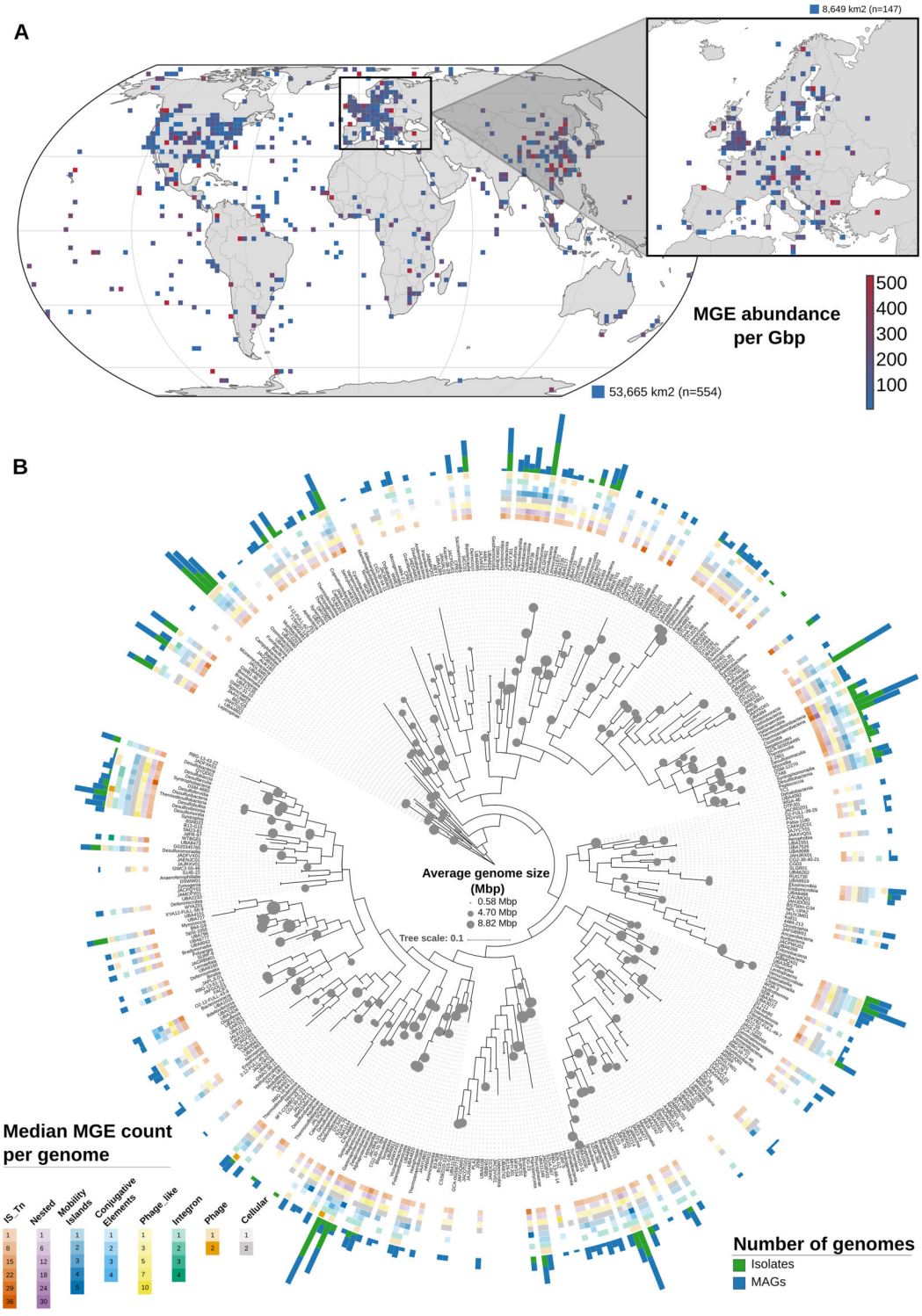


Figure 1. Global distribution of MGEs. (**A**) Global MGE abundance across high-quality MAGs from SPIRE samples. Global MGE abundance was estimated for 2 091 910 MGEs (excluding "cellular") detected across 35 011 SPIRE samples with valid geographic coordinates extracted from Metalog [72]. Samples were projected using the Equal Earth projection and aggregated into 554 equal-area grid cells of ~231.7 km per side (~92.7 km per side for the Europe zoom in featuring 8933 samples). For each grid cell, the MGE detection rate was calculated as the total number of detected MGEs (subtypes of "nested" elements counted as one MGE, e.g. "ce:1; is_tn:1" results in an MGE count equal to one) divided by the total binned genome size across all samples in that cell, and is reported as MGEs per gigabasepair (MGEs/Gbp). This normalization accounts for differences in sequencing depth, assembled genome content, and genome length across samples. Colors show the MGE detection rate clipped to the 5th–95th percentile range, corresponding to 25–548.6 MGEs/Gbp; the full observed range of mobilome positive grid cells was 6.02–7 293.01 MGEs/Gbp to account for uneven geographic sampling distribution (Supplementary Fig. S4B). (**B**) MGE abundances across bacterial taxonomic classes. The color-coded rings show median MGE count per genome. The MGE count of the "nested" category is calculated as a sum of individual elements e.g. "ce:1, is_tn:1" results in an MGE count equal to two, showing the potential nesting of several MGEs including multi-copy transposons present across different clades. The outer ring shows the number of isolate genomes (green) and MAGs (blue). The relative average genome sizes across bacterial classes are shown as gray circles. The source tree can be viewed through iTOL page https://itol.embl.de/tree/10116327340051791200843.

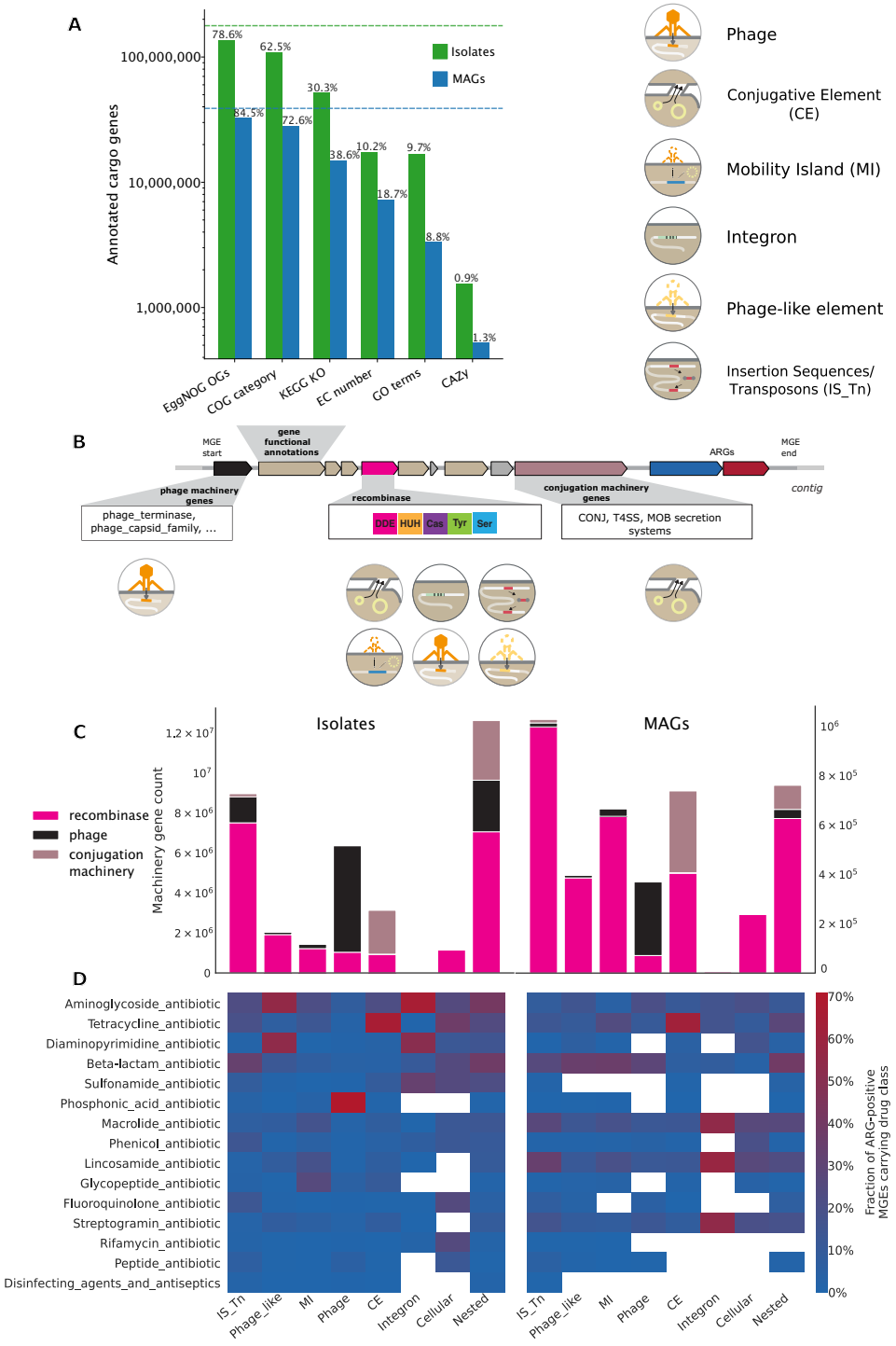


Figure 2. Functional repertoire of MGEs in proMGE2. (**A**) Functional annotation coverage of MGE cargo genes. Bar chart showing the number of cargo gene instances with at least one functional annotation, grouped by annotation type (EggNOG orthologous groups [73] , COG category [74] , KEGG KO [75] , EC number [76] , GO terms [77, 78] , CAZy [79]), for isolate genomes (green) and MAGs (blue). Dashed horizontal lines indicate the total number of cargo gene instances per source. Percentages above each bar denote annotation coverage as a fraction of all cargo gene instances in that source. Cargo genes were defined as all MGE genes unrelated to MGE classification. Functional gene annotation was performed with eggNOG-mapper (using default parameters, extended by -m diamond -dmnd_algo 0). (**B**) MGE profile. Representation of MGE features available as part of proMGE2—recombinases, conjugation machinery genes, phage structural genes, AMR genes, and functional annotations. The gene content of each MGE together with its genomic context can be visualized using the interactive "MGE Viewer" integrated to the SPIRE resource or downloaded for representative sequences in GFF format. (**C**) Machinery gene composition per MGE type. Stacked bar charts showing the total number of machinery genes attributed to each MGE machinery type (recombinase, phage structural genes, and conjugation machinery genes) across MGE categories, shown separately for isolate genomes and MAGs. (**D**) Drug-class composition of AMR cargo across MGE categories. Fraction of AMR-positive MGEs carrying genes conferring resistance to each of the top 15 drug classes, shown separately for isolates (left) and MAGs (right). Each cell represents the proportion of AMR-positive MGEs in a given category that carry at least one gene conferring resistance to the indicated drug class. White cells indicate no MGEs in that category carry genes for the respective drug class. The color scale ranges from blue (low fraction) to red (high fraction). Drug classes were assigned via CARD [51] ARO terms and are ordered by decreasing total prevalence across isolates.

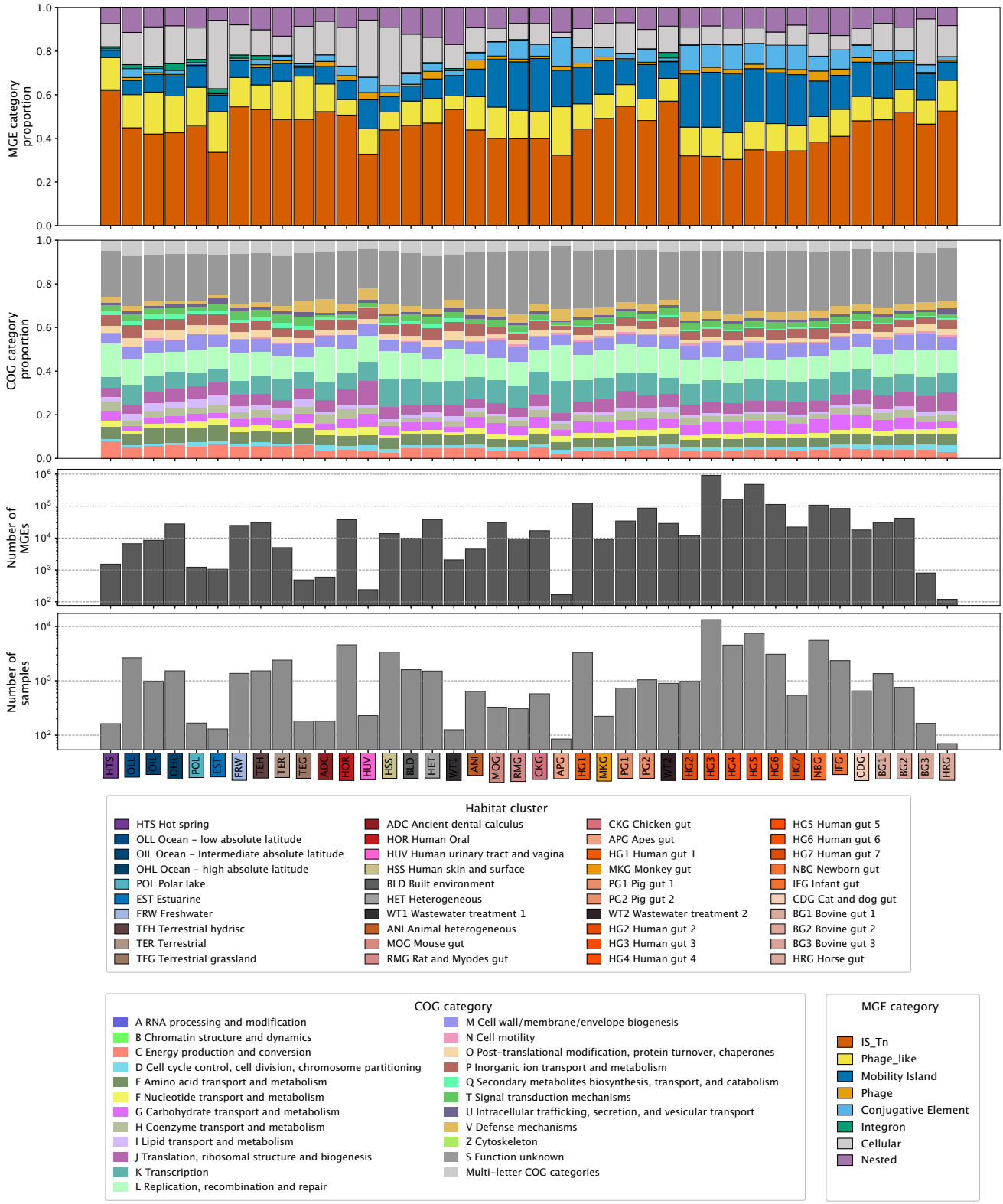


Figure 3. Mobilome composition across environmental habitat clusters. Stacked bar charts show the proportional distribution of 2 537 619 MGEs in SPIRE MAGs from 43 958 samples across disparate habitat clusters (top panel) and functional profiles of cargo genes from MGEs of seven main MGE categories, excluding cellular recombinases (second panel). Excluding "cellular," total number of MGEs (third panel) and respective samples (bottom panel) per habitat cluster are shown on a log scale. The MGE count of the "nested" category in the top panel is calculated as a sum of individual elements e.g. "ce:1;is_tn:1" results in an MGE count equal to two. Habitat clusters are ordered and colored according to the cluster identity in the habitat clustering framework by Kim, Podlesny, and Schiller *et al.* [55].

machinery categories represent detected contextual modules rather than mutually exclusive class-defining features (Fig. 2C). Irrespective of the differences in the number of predicted MGEs in isolates and MAGs they show similar trends in machinery gene distribution across different MGE categories. For instance, nested and IS_Tn elements contain the largest number of machinery genes, with substantial recombinase contributions, and additional phage structural or conjugation machinery components, while MGEs classified as phages are enriched in phage structural genes. The observed cross-category signals are consistent with the modular, mosaic nature of prokaryotic MGEs [27, 28].

proMGE2 also provides AMR annotations contextualized with the CARD ontology (Fig. 2D). In total, we predicted ∼630K MGEs (∼4% of all MGEs excluding "cellular") to contain at least one AMR annotation. Those MGEs belong to 1855 species (∼17% of all species) from isolates and MAGs conferring resistance against 23 drug classes.

Use cases

proMGE2 provides access to ∼16.4 million prokaryotic MGEs with consistent information on their genome source, recombinase markers, potential for horizontal gene transfer, nesting interactions, functional annotations, and burden of antibiotic resistance genes.

Targeted search

A search "By Species" provides a complete overview of the MGE profile of a species of interest and a summary of its mobilome features. On the other hand, searching by "MGE category" (is_tn, phage, phage_like, ce, mi, or integron; available on the landing page) lists all genomes containing MGEs of the queried MGE category. Finally, a user can query "By antibiotic" to list MGEs in the database that carry antibiotic resistance genes. Access to such MGEs facilitates downstream investigations into the risk assessment of antibiotic resistance spread by pathogens.

Annotate MGEs

To facilitate MGE annotation in sequenced (meta)genomes and contigs uncovered by proMGE2 we provide (i) a non-redundant MGE reference catalogue of >5.4 million MGEs and >245K cellular recombinase islands for mapping (Table 1), alongside functional annotations of their MGE genes and (ii) an "Annotate" feature that can predict MGEs for a given genomic sequence.

Interoperability

To enable comparative analysis, proMGE2 provides interoperability with input data sources, i.e. each MGE can be traced back to its host genome source—PG3 or SPIRE. Recombinase subfamilies are interlinked with the SMART database and AMR is contextualized via the CARD ontology. SPIRE MGEs are linked to the genomic viewer on the SPIRE webpage where all functional annotations of an MGE as well as its genomic context can be inspected. This integration allows in-depth investigations of MGEs across habitats and provides access to the metadata of the corresponding samples.

Exploration of the mobilome across habitats

To demonstrate the breadth of MGEs represented in proMGE2 across diverse habitats, we obtained corresponding habitat information for MAGs from SPIRE (see above). Different MGE categories were broadly represented across habitats, with most categories detected in the different habitat groups (Fig. 3). Similarly, the functional repertoire of MGE cargo genes from environmental MAGs can be explored across habitats, revealing a substantial fraction of genes with unknown or unassigned functions (Fig. 3). The distribution of annotated functions across diverse habitats illustrates the diverse functional potential carried by MGEs, while the prevalence of unassigned functions highlights a substantial unexplored component of MGE gene content. By integrating MGE classification, cargo gene annotations, and habitat metadata, proMGE2 facilitates comparative analyses of mobilome composition and functional potential across host-associated, aquatic, terrestrial, engineered, and extreme environments.

Usage considerations

The following considerations should be taken into account while exploring the catalogued MGEs:

First, our MGE annotation framework is based on accurate and sensitive detection of recombinases. The set of HMMs used for recombinase detection were trained on isolate proteomes and may therefore have reduced sensitivity to divergent recombinases such as those found in environmental MAGs. As a result, MGE counts from MAGs should be considered as conservative estimates with a potential for expansion.

Second, the accuracy of MGE boundary estimation depends on the number of genomes represented within the pangenome [22], hence continued generation of high-quality MAGs will improve pangenome coverage and consequently MGE boundary estimation. Third, the habitat clusters differ substantially in the number of contributing samples (Figs 1A and 3 and Supplementary Fig. S4B) and as a consequence in the number of detected MGEs. This variation should be considered when interpreting habitat-specific patterns, because well-sampled habitats provide more robust estimates of MGE and cargo composition, whereas sparsely sampled habitats may capture more specialized mobilome profiles (Fig. 3). Integration of annotated MGEs from MAGs with SPIRE enables interpretation of habitat-associated patterns in the context of sampling depth. Finally, it is important to note that proMGE2 reports MGEs associated with MAGs, hence free plasmids and viruses not represented in metagenomic assemblies are not captured.

Conclusion and outlook

proMGE2 is a resource of ∼16.4 million prokaryotic Mobile Genetic Elements that allows explorative analysis of MGEs integrated with rich contextual information of their functional features. Together with annotation of MGEs in user provided genomes, proMGE2 facilitates community-driven exploration of mobilome content to advance our understanding of the diverse role of MGEs and their interactions in microbes and microbial communities. Future versions of proMGE will include new releases of proGenomes and SPIRE expanding coverage of MGEs and their cargo function repertoire.

Acknowledgements

The authors thank Yan-Ping Yuan (EMBL) and Jurij Pečar (EMBL ITS) for technical support. We are indebted to the EMBL Heidelberg HPC Cluster [71] and its user community, both of which had to endure multiple periods of heavy computational load while we were hunting for MGEs in proGenomes3 and SPIRE. Computations for the proMGE2 "Annotate" feature are powered by the cloud infrastructure of the German Network of Bioinformatics (https://www.denbi.de/cloud) via the CloWM platform (https://clowm.bi.denbi.de). We sincerely thank Dr John Lees (European Bioinformatics Institute, EMBL-EBI), Dr Laura Carroll (Umeå University), Dr James Hall (University of Liverpool), and Prof. Lara Urban (University of Zurich) for their valuable suggestions and feedback. The authors express their deep gratitude to Peer Bork for his guidance and support over the years. Peer passed away in January 2026 before the paper was submitted and therefore was not able to approve the submitted and final accepted versions and cannot be held accountable for the work post-publication. During the preparation of this work, the authors used ChatGPT (OpenAI) to improve language and clarity. After the usage of this tool, the authors reviewed and edited the contents as needed and take full responsibility for the content of the article. All authors have read and approved the manuscript.

Author contributions: Conceptualization: A.G., P.B. (lead), S.K. (lead). Data curation: A.G., C.S. (lead), C.Y.K., D.P., A.O., A.F., T.S.S., D.M., and J.S. Formal analysis: A.G. (lead), C.S. Investigation: A.G., C.S., and S.K. Methodology: A.G., C.S., and S.K. Resources: C.S., I.L., A.F., T.S.B.S., D.M., M.Z., and M.K. Software: A.G., I.L., C.S. (lead). Validation: A.G. (lead), C.S. (lead). Visualization: A.G. (lead), K.O.E.E., and B.B. Project administration: S.K. (lead). Supervision: M.Z.-K., M.Z., M.K., P.B., and S.K. (lead). Funding acquisition: P.B. (lead), S.K. Writing—original draft: A.G. (lead), S.K. Writing—review & editing: A.G., C.S., and S.K.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was supported by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de.NBI and by the Federal Ministry of Research, Technology and Space in the frame of de.NBI & ELIXIR-DE (W-de.NBI-014), and the European Molecular Biology Laboratory (EMBL). This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project numbers 555128266 and 460129525 (NFDI4Microbiota), and partially funded by the EMBL Planetary Biology Transversal Theme through a seed grant. Funding to pay the Open Access publication charges for this article was provided by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 460129525 (NFDI4Microbiota).

Data availability

All data are freely available at http://promge.embl.de/ under the Creative Commons Attribution-ShareAlike 4.0 International License. All genomes and MAGs underlying proMGE2 are publicly available via proGenomes3 (http://progenomes3.embl.de/) and SPIRE (https://spire.embl.de/). No new sequencing data was generated in this study. The workflow is available on CloWM (https://clowm.bi.denbi.de) [66]. For running the workflow locally, proMGEflow can be obtained from Github (https://github.com/grp-bork/promgeflow) and Zenodo (https://doi.org/10.5281/zenodo.23923727).

References

1. Frost LS, Leplae R, Summers AO *et al.* Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 2005;3:722–32. https://doi.org/10.1038/nrmicro1235
2. Escudero JA, Loot C, Nivina A *et al.* The integron: adaptation on demand. *Microbiol Spectr* 2015;3:10.1128/microbiolspec.mdna3–0019-2014. https://doi.org/10.1128/microbiolspec.mdna3-0019–2014
3. Tokuda M, Shintani M. Microbial evolution through horizontal gene transfer by mobile genetic elements. *Microbiol Biotechnol* 2024;17:e14408. https://doi.org/10.1111/1751-7915.14408
4. Weisberg AJ, Chang JH. Mobile genetic element flexibility as an underlying principle to bacterial evolution. *Annu Rev Microbiol* 2023;77:603–24. https://doi.org/10.1146/annurev-micro-032521-022006
5. Rajabal V, Ghaly TM, Egidi E *et al.* Exploring the role of mobile genetic elements in shaping plant–bacterial interactions for sustainable agriculture and ecosystem health. *Plants People Planet* 2024;6:408–20. https://doi.org/10.1002/ppp3.10448
6. Hall JPJ, Harrison E, Baltrus DA. Introduction: the secret lives of microbial mobile genetic elements. *Phil Trans R Soc B* 2022;377:20200460. https://doi.org/10.1098/rstb.2020.0460
7. Moura De Sousa J, Lourenço M, Gordo I. Horizontal gene transfer among host-associated microbes. *Cell Host Microbe* 2023;31:513–27. https://doi.org/10.1016/j.chom.2023.03.017
8. Olsen NS, Riber L. Metagenomics as a transformative tool for antibiotic resistance surveillance: highlighting the impact of mobile genetic elements with a focus on the complex role of phages. *Antibiotics* 2025;14:296. https://doi.org/10.3390/antibiotics14030296
9. Partridge SR, Kwong SM, Firth N *et al.* Mobile genetic elements associated with antimicrobial resistance. *Clin Microbiol Rev* 2018;31:10.1128/cmr.00088–17. https://doi.org/10.1128/CMR.00088-17
10. Ghaly TM, Gillings MR. New perspectives on mobile genetic elements: a paradigm shift for managing the antibiotic resistance crisis. *Phil Trans R Soc B* 2021;377:20200462. https://doi.org/10.1098/rstb.2020.0462
11. Forster SC, Liu J, Kumar N *et al.* Strain-level characterization of broad host range mobile genetic elements transferring antibiotic resistance from the human microbiome. *Nat Commun* 2022;13:1445. https://doi.org/10.1038/s41467-022-29096-9
12. Pan T, Li Q. Mobile genetic elements in *Klebsiella pneumoniae*. *J Bacteriol* 2025;207:e00012–25. https://doi.org/10.1128/jb.00012-25
13. Rankin DJ, Rocha EPC, Brown SP. What traits are carried on mobile genetic elements, and why? *Heredity* 2011;106:1–10. https://doi.org/10.1038/hdy.2010.24
14. Norman A, Hansen LH, Sørensen SJ. Conjugative plasmids: vessels of the communal gene pool. *Phil Trans R Soc B* 2009;364:2275–89. https://doi.org/10.1098/rstb.2009.0037
15. Dennis JJ. The evolution of IncP catabolic plasmids. *Curr Opin Biotechnol* 2005;16:291–8. https://doi.org/10.1016/j.copbio.2005.04.002

16. Nojiri H, Shintani M, Omori T. Divergence of mobile genetic elements involved in the distribution of xenobiotic-catabolic capacity. *Appl Microbiol Biotechnol* 2004;64:154–74. https://doi.org/10.1007/s00253-003-1509-y
17. Top EM, Springael D, Boon N. Catabolic mobile genetic elements and their potential use in bioaugmentation of polluted soils and waters. *FEMS Microbiol Ecol* 2002;42:199–208. https://doi.org/10.1111/j.1574-6941.2002.tb01009.x
18. Schubert K, Shosanya T, García-Bayona L. The role of mobile genetic elements in adaptation of the microbiota to the dynamic human gut ecosystem. *Curr Opin Microbiol* 2025;88:102675. https://doi.org/10.1016/j.mib.2025.102675
19. Jiang X, Hall AB, Xavier RJ *et al.* Comprehensive analysis of chromosomal mobile genetic elements in the gut microbiome reveals phylum-level niche-adaptive gene pools. *PLoS One* 2019;14:e0223680. https://doi.org/10.1371/journal.pone.0223680
20. Frost LS, Leplae R, Summers AO *et al.* Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 2005;3:722–32. https://doi.org/10.1038/nrmicro1235
21. Lang AS, Buchan A, Burrus V. Interactions and evolutionary relationships among bacterial mobile genetic elements. *Nat Rev Micro* 2025;23:423–38. https://doi.org/10.1038/s41579-025-01157-y
22. Khedkar S, Smyshlyaev G, Letunic I *et al.* Landscape of mobile genetic elements and their antibiotic resistance cargo in prokaryotic genomes. *Nucleic Acids Res* 2022;50:3155–68. https://doi.org/10.1093/nar/gkac163
23. Bean EL, Herman C, Anderson ME *et al.* Biology and engineering of integrative and conjugative elements: construction and analyses of hybrid ICEs reveal element functions that affect species-specific efficiencies. *PLoS Genet* 2022;18:e1009998. https://doi.org/10.1371/journal.pgen.1009998
24. Arbel-Goren R, McKeithen-Mead SA, Voglmaier D *et al.* Target search by an imported conjugative DNA element for a unique integration site along a bacterial chromosome during horizontal gene transfer. *Nucleic Acids Res* 2023;51:3116–29. https://doi.org/10.1093/nar/gkad068
25. Wang H, Joffré E. Plasmid copy number as a modulator in bacterial pathogenesis and antibiotic resistance. *NPJ Antimicrob Resist* 2025;3:72. https://doi.org/10.1038/s44259-025-00145-9
26. Siguier P, Gourbeyre E, Varani A *et al.* Everyman's Guide to Bacterial Insertion Sequences. *Microbiol Spectr* 2015;3:10.1128/microbiolspec.mdna3–0030-2014. https://doi.org/10.1128/microbiolspec.mdna3-0030–2014
27. Toussaint A, Merlin C. Mobile Elements as a combination of functional modules. *Plasmid* 2002;47:26–35. https://doi.org/10.1006/plas.2001.1552
28. Mark Osborn A, Böltner D. When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid* 2002;48:202–12. https://doi.org/10.1016/S0147-619X(02)00117-8
29. Kieffer N, Hipólito A, Ortiz-Miravalles L *et al.* Mobile integrons encode phage defense systems. *Science* 2025;388:eads0915. https://doi.org/10.1126/science.ads0915
30. Pfeifer E, Rocha EPC. Phage-plasmids promote recombination and emergence of phages and plasmids. *Nat Commun* 2024;15:1545. https://doi.org/10.1038/s41467-024-45757-3
31. Grodner B, Shi H, Farchione O *et al.* Spatial mapping of mobile genetic elements and their bacterial hosts in complex microbiomes. *Nat Microbiol* 2024;9:2262–77. https://doi.org/10.1038/s41564-024-01735-5
32. Siguier P, Perochon J, Lestrade L *et al.* ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 2006;34:D32–6. https://doi.org/10.1093/nar/gkj014
33. Tansirichaiya S, Rahman MA, Roberts AP. The Transposon Registry. *Mobile DNA* 2019;10:40. https://doi.org/10.1186/s13100-019-0182-3
34. Dotto BR, Carvalho EL, Silva AF *et al.* HTT-DB: horizontally transferred transposable elements database. *Bioinformatics* 2015;31:2915–7. https://doi.org/10.1093/bioinformatics/btv281
35. Redondo-Salvo S, Bartomeus-Peñalver R, Vielva L *et al.* COPLA, a taxonomic classifier of plasmids. *BMC Bioinfomatics* 2021;22:390. https://doi.org/10.1186/s12859-021-04299-x
36. Camargo AP, Roux S, Schulz F *et al.* Identification of mobile genetic elements with geNomad. *Nat Biotechnol* 2024;42:1303–12. https://doi.org/10.1038/s41587-023-01953-y
37. Arndt D, Grant JR, Marcu A *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44:W16–21. https://doi.org/10.1093/nar/gkw387
38. Botelho J. Defense systems are pervasive across chromosomally integrated mobile genetic elements and are inversely correlated to virulence and antimicrobial resistance. *Nucleic Acids Res* 2023;51:4385–97. https://doi.org/10.1093/nar/gkad282
39. Durrant MG, Li MM, Siranosian BA *et al.* A bioinformatic analysis of integrative mobile genetic elements highlights their role in bacterial adaptation. *Cell Host Microbe* 2020;27:140–53.e9. https://doi.org/10.1016/j.chom.2019.10.022
40. Bi D, Xu Z, Harrison EM *et al.* ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. *Nucleic Acids Res* 2012;40:D621–6. https://doi.org/10.1093/nar/gkr846
41. Liu M, Li X, Xie Y *et al.* ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res* 2019;47:D660–5. https://doi.org/10.1093/nar/gky1123
42. Néron B, Littner E, Haudiquet M *et al.* IntegronFinder 2.0: identification and Analysis of integrons across bacteria, with a focus on antibiotic resistance in *Klebsiella*. *Microorganisms* 2022;10:700. https://doi.org/10.3390/microorganisms10040700
43. Rojas-Villalobos C, Ossandon FJ, Castillo-Vilcahuaman C *et al.* MOBHunter: a data integration platform for identification and classification of mobile genetic elements in microbial genomes. *Nucleic Acids Res* 2025;53:W398–407. https://doi.org/10.1093/nar/gkaf396
44. Che Y, Yang Y, Xu X *et al.* Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. *Proc Natl Acad Sci USA* 2021;118:e2008731118. https://doi.org/10.1073/pnas.2008731118
45. Nishijima S, Fullam A, Schmidt TSB *et al.* VIRE: a metagenome-derived, planetary-scale virome resource with environmental context. *Nucleic Acids Res* 2026;54:D902–11. https://doi.org/10.1093/nar/gkaf1225
46. Kerkvliet JJ, Bossers A, Kers JG *et al.* Metagenomic assembly is the main bottleneck in the identification of mobile genetic elements. *PeerJ* 2024;12:e16695. https://doi.org/10.7717/peerj.16695
47. Qi Q, Rajabal V, Ghaly TM *et al.* Identification of integrons and gene cassette-associated recombination sites in bacteriophage genomes. *Front Microbiol* 2023;14:1091391.
48. Fullam A, Letunic I, Schmidt TSB *et al.* proGenomes3: approaching one million accurately and consistently annotated high-quality prokaryotic genomes. *Nucleic Acids Res* 2023;51:D760–6. https://doi.org/10.1093/nar/gkac1078
49. Schmidt TSB, Fullam A, Ferretti P *et al.* SPIRE: a Searchable, Planetary-scale mIcrobiome REsource. *Nucleic Acids Res* 2024;52:D777–83. https://doi.org/10.1093/nar/gkad943
50. Letunic I, Bork P. SMART v10: three decades of the protein domain annotation resource. *Nucleic Acids Res* 2026;54:D499–503. https://doi.org/10.1093/nar/gkaf1023
51. Alcock BP, Huynh W, Chalil R *et al.* CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res* 2023;51:D690–9. https://doi.org/10.1093/nar/gkac920
52. Parks DH, Imelfort M, Skennerton CT *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;25:1043–55. https://doi.org/10.1101/gr.186072.114

53. Pallen MJ, Rodriguez-R LM, Alikhan N-F. Naming the unnamed: over 65 000 Candidatus names for unnamed *Archaea* and bacteria in the Genome Taxonomy Database. *Int J Syst Evol Microbiol* 2022;72:005482. https://doi.org/10.1099/ijsem.0.005482
54. Orakov A, Fullam A, Coelho LP *et al.* GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biol* 2021;22:178. https://doi.org/10.1186/s13059-021-02393-0
55. Kim CY, Podlesny D, Schiller J *et al.* Planetary microbiome structure and generalist-driven gene flow across disparate habitats. *Cell* 2026;189:2073–91.e21. https://doi.org/10.1016/j.cell.2025.12.051
56. Smyshlyaev G, Bateman A, Barabas O. Sequence analysis of tyrosine recombinases allows annotation of mobile genetic elements in prokaryotic genomes. *Mol Syst Biol* 2021;17:e9880. https://doi.org/10.15252/msb.20209880
57. Bobay L-M, Rocha EPC, Touchon M. The adaptation of temperate bacteriophages to their host genomes. *Mol Biol Evol* 2013;30:737–51. https://doi.org/10.1093/molbev/mss279
58. Everitt RG, Didelot X, Batty EM *et al.* Mobile elements drive recombination hotspots in the core genome of Staphylococcus aureus. *Nat Commun* 2014;5:3956. https://doi.org/10.1038/ncomms4956
59. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011;7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195
60. Hellewell J, Horsfield ST, von Wachsmann J *et al.* CELEBRIMBOR: core and accessory genes from metagenomes. *Bioinformatics* 2024;40:btae542. https://doi.org/10.1093/bioinformatics/btae542
61. Cantalapiedra CP, Hernández-Plaza A, Letunic I *et al.* eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 2021;38:5825–9. https://doi.org/10.1093/molbev/msab293
62. Cury J, Abby SS, Doppelt-Azeroual O *et al.* Identifying conjugative plasmids and integrative conjugative elements with CONJscan. In:de la Cruz F. (ed), *Horizontal Gene Transfer: Methods and Protocols*. New York, NY, Springer US, 2020, pp. 265–83. https://doi.org/10.1007/978-1-4939-9877-7 10.1007/978-1-4939-9877-7_19
63. Abby SS, Néron B, Ménager H *et al.* MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-cas systems. *PLoS One* 2014;9:e110726. https://doi.org/10.1371/journal.pone.0110726
64. Néron B, Denise R, Coluzzi C *et al.* MacSyFinder v2: improved modelling and search engine to identify molecular systems in genomes. *Peer Community Journal* 2023;3:e28.
65. Ha Y, Maddamsetti R, Chen X *et al.* Transposon-plasmid nesting enables fast response to fluctuating environments. bioRxiv, https://doi.org/10.1101/2025.06.04.657954, 8 June 2025, preprint: not peer reviewed.
66. Göbel D, Stoye J, Sczyrba A *et al.* The cloud-based workflow manager (CloWM)—an integrated platform for highly scalable workflow execution. 2024.
67. Seemann T. [Computer software]. *GitHub* https://github.com/tseemann/abricate (tag: v1.0.1, 19 April 2020).
68. Bonin N, Doster E, Worley H *et al.* MEGARes and AMR++, v3.0: an updated comprehensive database of antimicrobial resistance determinants and an improved software pipeline for classification using high-throughput sequencing. *Nucleic Acids Res* 2022;51:D744–52. https://doi.org/10.1093/nar/gkac1047
69. Ugarcina Perovic S, Ramji V, Chong H *et al.* argNorm: normalization of antibiotic resistance gene annotations to the Antibiotic Resistance Ontology (ARO). *Bioinformatics* 2025;41:btaf173. https://doi.org/10.1093/bioinformatics/btaf173
70. Wu D, Seshadri R, Kyrpides NC *et al.* A metagenomic perspective on the microbial prokaryotic genome census. *Sci Adv* 2025;11:eadq2166. https://doi.org/10.1126/sciadv.adq2166
71. Pečar J, Lueck R, Wahlers M. EMBL Heidelberg HPC Cluster. [Computer software]. *Zenodo* https://doi.org/10.5281/zenodo.12785830 2020.
72. Kuhn M, Schmidt TSB, Ferretti P *et al.* Metalog: curated and harmonised contextual data for global metagenomics samples. *Nucleic Acids Res* 2026;54:D826–34. https://doi.org/10.1093/nar/gkaf1118
73. Huerta-Cepas J, Szklarczyk D, Heller D *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 2019;47:D309–14. https://doi.org/10.1093/nar/gky1085
74. Galperin MY, Vera Alvarez R, Karamycheva S *et al.* COG database update 2024. *Nucleic Acids Res* 2025;53:D356–63. https://doi.org/10.1093/nar/gkae983
75. Kanehisa M, Furumichi M, Tanabe M *et al.* KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;45:D353–61. https://doi.org/10.1093/nar/gkw1092
76. McDonald AG, Boyce S, Tipton KF. ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res* 2009;37:D593–7. https://doi.org/10.1093/nar/gkn582
77. The Gene Ontology Consortium The Gene Ontology knowledgebase in 2026. *Nucleic Acids Res* 2026;54:D1779–92. https://doi.org/10.1093/nar/gkaf1292
78. Ashburner M, Ball CA, Blake JA *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* 2000;25:25–9. https://doi.org/10.1038/75556
79. Levasseur A, Drula E, Lombard V *et al.* Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels* 2013;6:41. https://doi.org/10.1186/1754-6834-6-41

Received: September 15, 2026. **Revised:** October 19, 2026. **Accepted:** October 20, 2026
© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.