Documentation

The Importance of Transposable Elements

TnCentral

References

Acknowledgements

The Importance of Transposable Elements

Transposable elements (TE) are key facilitators of bacterial adaptation and therefore are central players in the emergence of multiple antibacterial resistances such as resistance to antibiotics, heavy metals and to transmission of pathogenic traits. TE capture passenger genes using a number of mechanisms and transmit them to larger mobile genetic elements, plasmids, where they accumulate and are then transferred within and between bacterial populations. TE also contribute significantly to the on-going reorganization of bacterial genomes giving rise to new strains that are more and more adept at proliferating both in the environment and in hospitals. Understanding TE nature, distribution and action is therefore an indispensable part of the struggle to cope with the public health crisis of multiple antimicrobial resistance (AMR) ([1],[2]).

To understand the impact of TE on bacterial populations, to follow the flow of genes important in public health both in clinical and environmental settings and to provide some measure of understanding which might allow prediction of resistance transmission, it is essential to provide a detailed description and catalogue of TE structures and diversity.

This has already been undertaken for the simplest TE, the insertion sequences (IS), in the form of the online knowledge base ISfinder (https://www-is.biotoul.fr/index.php) ([3],[4]), an international resource for IS currently including over 5000 individual examples. The ISfinder platform also includes a set of software tools, ISsaga, allowing semi-automatic genome annotation for IS using the ISfinder database ([5]). Although movement of IS has a profound and continuous impact on genome organization and function due to their ability to rearrange DNA, regulate neighboring genes and generate mutations ([6–9]), they do not themselves generally carry integrated passenger genes.

There are a large number of significantly more complex TE, arguably even more important in the global emergence of AMR. These are generically called transposons and may carry multiple passenger genes, including some of the most clinically important antibiotic resistance genes. They are grouped into a number of distinct families with characteristic organizations (6). Like IS, their transposition activities facilitate the rapid spread of groups of antibiotic resistance genes and promote their horizontal transfer to other bacterial strains, species and genera via natural vectors such as conjugal plasmids and bacterial viruses. Yet another important aspect of their impact is their ability to assemble passenger genes into resistance clusters ([10],[11]). While there appears to be wide-spread appreciation that mobile plasmids are responsible for the spread of antibiotic resistance, fewer people are aware that IS and transposons are the conduit that transfers this information between chromosomes and plasmids.

It is crucial to stress the importance of educating the clinical world concerning transposition mechanisms in an easy to use way. Most scientists consider that IS/Tn all behave in the same way and believe that cataloguing them is simply "busy-work". However, a thorough understanding of how IS/Tn assemble antimicrobial resistance genes and effect rapid changes in plasmid vector structure is critical to understanding the increasingly efficient AMR spread observed today and combatting future AMR outbreaks (see Figure 1).

pAAC154 and family from 63 plasmids
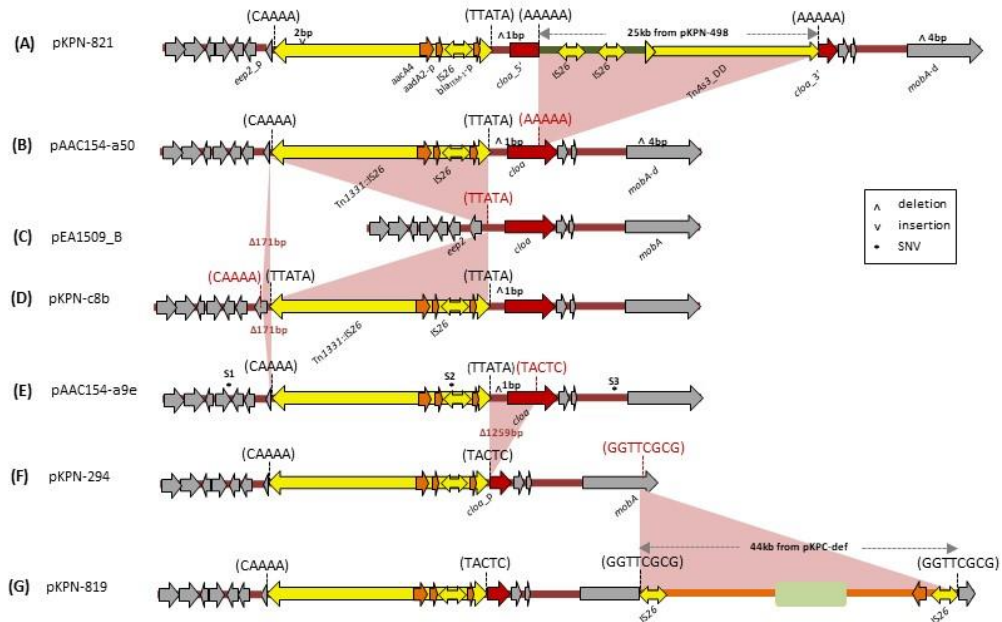
He et al 2016

Figure 1. This figure represents the relationship between a group of plasmids, many isolated from patients during an outbreak of carbapenem resistant bacteria at the NIH Clinical Center (references [10] and [11]). The DNA sequences in brackets show small direct target repeats generated during transposon movement and act as an important guide to the way in which the plasmids have been rearranged during the course of the infections. This type of information will be crucial in developing the fourth stage of the project: semiautomatic annotation of plasmid groups of interest such as those isolated from clinical outbreaks.

TnCentral

There has been a need for a database for transposons providing similar comprehensiveness, transparency and usability that ISfinder provides for IS. TnCentral is a pilot transposon database conceived as a resource for transposons, their associated passenger genes, and their host organisms.

TnCentral: Content

Mobile Element Groups Covered

TnCentral initially focuses on two transposon groups—the Tn*3* transposon family (Figures 2, 4 and 5) and the composite (or compound) transposons composed of two IS flanking a variety of passenger genes --because they include some of the most clinically important AMR transposons (Figure 3 and 6).

The Tn3 family

Tn*3* family members form a tightly knit group. The basic Tn*3* family transposition module is composed of transposase and resolvase genes and two ends with related terminal inverted repeat DNA sequences, the IRs, of 38-40bp or sometimes even longer ([12](#)). There is a large (~1000 aa) DDE transposase, TnpA, significantly longer than the DDE transposases normally associated with Insertion Sequences (IS) (see ([7](#))). TnpA catalyzes the DNA cleavage and strand transfer reactions necessary for formation of a cointegrate transposition intermediate during replicative transposition. The cointegrate is composed of fused donor (with the transposon) and target (without the transposon) circular DNA molecules fused into a single circular molecule and separated by two directly repeated transposon copies, one at each junction ([13](#)).

A second feature of members of this transposon family is that they carry short (~100-150bp) DNA segments, *res* (for *res*olution) or *rst* (for *r*esolution site tnp*S* tnp*T* – see below; ([14](#))) at which site-specific recombination between each of the two Tn copies occurs to "resolve" the cointegrate into individual copies of the transposon donor and the target molecules each containing a single transposon copy (see ([15](#))). This highly efficient recombination system is assured by a transposon-specified sequence-specific recombinase enzyme: the resolvase.

There are at present three known major resolvase types: TnpR, TnpI, and TnpS+TnpT, distinguished, among other things, by the catalytic nucleophile involved in DNA phosphate bond cleavage and rejoining during recombination: TnpR, a classic serine (S)-site-specific recombinase (e.g. ([16](#))); TnpI, a tyrosine (Y) recombinase similar to phage integrases ([17](#))(see ([15](#)); and a heteromeric resolvase combining a tyrosine recombinase, TnpS, and a divergently expressed helper protein, TnpT, with no apparent homology to other proteins ([14,18](#)). The resolvase genes can be either co-linear, generally upstream of *tnpA* or divergent. In the former case the *res* site lies upstream of *tnpR* and in the latter case, between the divergent *tnpR* and *tnpA* genes. For relatives encoding TnpS and TnpT, the corresponding genes are divergent and the *res* (*rst*) site lies between *tnpS* and *tnpT*. Examples of these architectures are shown in Figure 1. Each *res* includes a number of short DNA sub-sequences which are recognised and bound by the cognate resolvases. These are different for different resolvase

systems (Figure 1) But where analysed, *res* sites also include promoters which drive both transposase and resolvase expression. Indeed, TnpR from Tn*3* was originally named for its ability to repress transposase expression by binding to these sites ([19](),[20]()).

A number of Tn3 members do not include a resolvase gene and therefore, although cointegrates are formed during transposition of these transposons, no efficient sequence-specific recombination occurs to resolve these structures. Instead, "resolution" depends on the homologous recombination system of the host which uses the directly repeated transposon copies as a substrate.

The complexity of these Tn resides in the diversity of other mobile elements incorporated into their structures (such as, Insertion Sequences (IS) and integrons as well as other Tn*3* family members – see ([15]())) (Figure 4) and other passenger genes. The most notorious of these genes are those for antibiotic and heavy metal resistance although other genes involved in organic catabolite degradation and virulence functions for both animals and plants (Figure 2) also form part of the Tn*3* family arsenal of passenger genes.
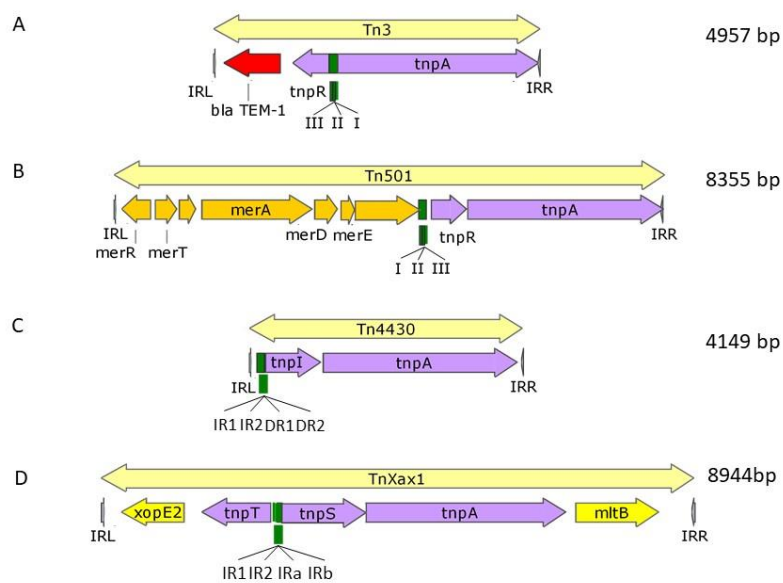

Figure 2. Selected Tn3 family members. Transposition-related genes are shown in purple, res sites in green, antibiotic resistance genes in red, heavy metal resistance genes in chrome yellow and plant pathogenicity genes in yellow. IRR and IRL are the terminal inverted repeats.

Compound transposons

There are several transposon types. Members of one major group, the compound transposons (Figure 3), are simply composed of a pair of IS flanking a DNA segment in either direct or inverted repeat. The flanking ISs, each composed of short terminal inverted repeat sequences encompassing gene(s) encoding transposition enzymes (transposase, Tpase, and regulatory gene products), can mobilise the intervening DNA segment. Classical transposons

of this type include Tn*5* (inverted flanking IS50), Tn10 (inverted flanking IS*10*), Tn*903* (inverted flanking IS*903*), Tn*602* (direct flanking IS*903*) as well as Tn*9* (direct flanking IS*1*). Since these
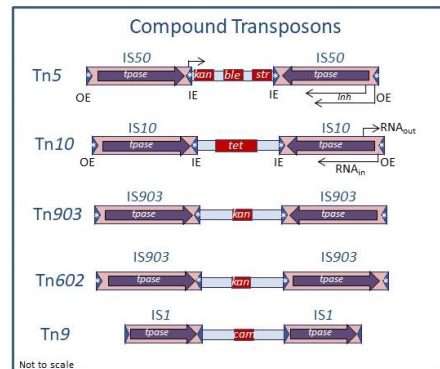
early examples, many other such composite transposons have been identified by experiment or from genome sequencing. These can carry a variety of phenotypic characters including antibiotic resistance but also genes involved in pathogenesis, symbiosis and xenobiotic catabolism. Some early observations concerning Tn*5* and Tn*10* suggest that the flanking ISs have mutated to render them less autonomous: for example, mutations within IS*10L* that inactivate its Tpase or mutation in one of the internal IR of IS*50L* that truncates the Tpase and generates a promoter for expression of the interstitial kanamycin resistance (*kan*) gene ([7]). Transposition of these compound transposons depends on the mechanism of transposition of the constituent IS and is subject to the same types of regulation. For example, insertion, like that of the constituent IS, will generally generate flanking target site duplications (TSD) whose length is particular to the IS.

Figure 3. Examples of the well-known compound transposons Tn5, Tn10, Tn903, Tn602 and Tn9 showing the flanking IS which bestow transposition mobility.

Scripts for identifying transposons

1) Tn3 family. We have developed software—Tn3finder—to automatically identify Tn*3* family transposons in nucleotide sequence databases. Running Tn3finder against complete assembled bacterial genomes in RefSeq (a comprehensive, non-redundant set of sequences at NCBI designed to be used as reference standards), resulted in identification of several hundred, largely unannotated, potential Tn*3* family sequences, of which we have so far manually validated nearly the entire collection. These will be included as their annotation proceeds.

2) Compound transposons. A second script, TnCompfinder, has been written to identify potential composite or compound transposons. Several hundred potential compound transposons in which a central segment of DNA is flanked by two related IS have been identified in addition to those already described in the literature. These will also be included as they are completely annotated.

Both can be downloaded as Zip files with instructions for their use.

These scripts can be easily adapted to identify other transposon families (e.g. Tn*7*) and passenger genes.

General information on transposable elements

In addition, it is intended that TnCentral will incorporate a large general section on transposition mechanism and regulation imported and updated from ISfinder ([https://www-is.biotoul.fr/general_information.php; written by MC](https://www-is.biotoul.fr/general_information.php)). This will be enhanced by detailed descriptions of individual groups of and transposons.

TnCentral: Organization and Features

TnCentral is a Web-based database designed to provide a detailed annotation of bacterial transposons with the same rigour as ISfinder (https://www-is.biotoul.fr/index.php) for bacterial insertion sequences.

Search functions

The interface provides a variety of search functions.

The transposon collection can be searched using the transposon name, synonyms which may have been used in the literature, the type of mobile genetic element (insertion sequence, transposon or integron), the family and subgroup to which it belongs, the host organism, country of identification and date of identification. The latter three search terms will eventually useful in for epidemiological tracking. It is also possible to search for passenger genes such as antibiotic resistance and heavy metal resistance genes and retrieve information on the transposons in which they are found.

BLAST functions

The platform also provides a BLAST facility for entries in the TnCentral database and provides links out to BLAST of the

ISfinder (https://isfinder.biotoul.fr/blast.php),

NCBI ([https://blast.ncbi.nlm.nih.gov/Blast.cgi)](https://blast.ncbi.nlm.nih.gov/Blast.cgi),

Comprehensive Antibiotic Resistance Database (CARD; [https://card.mcmaster.ca/analyze/blast)](https://card.mcmaster.ca/analyze/blast)

and the Toxin-Anitoxin (TADB; http://202.120.12.135/TADB2/index.php) databases.

Transposon entry page

Each transposon has an entry page which includes a graphic representation of the annotated sequence with color-coded features. Annotations include:

1. Host information: this includes the species, strain, and plasmid/chromosome in which the transposon was found as well as the date and geographic location of the isolate.
2. Protein coding genes: these include transposases (Figure 4, #1-#3), accessory genes that support transpositions (e.g., resolvases, Figure 4, #4-#5), and passenger genes, including antibiotic resistance (Figure 4, #6) and heavy metal resistance genes. Transposable Elements (TE): these include the Tn*3*-family or compound transposons themselves as well as any additional TE that are "nested" within. For example, the Tn*3*-family transposon Tn*4401a* (Figure 4, #7) contains two IS nested within it: IS*Kpn7* (Figure 4, #8), and IS*Kpn6* (Figure 4, #9).
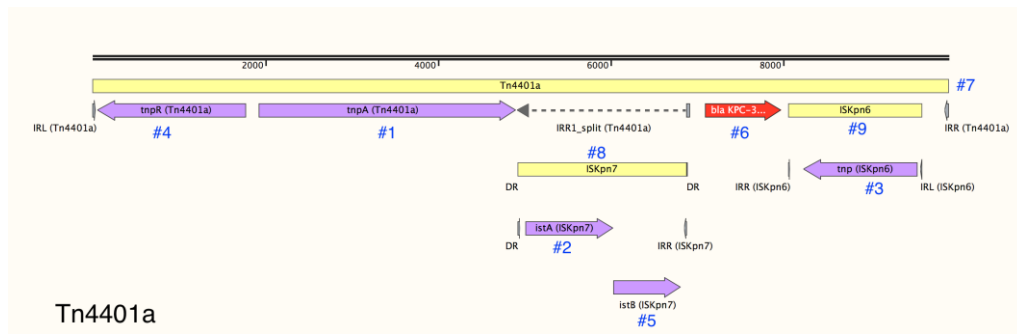
Figure 4. Annotated features of transposon Tn4401a, a member of the Tn3 family including transposases (#1-#3), accessory genes (#4-#5), passenger genes (#6), and mobile elements (#7-#9).

3. Repeat elements: these include the terminal inverted repeats (IRL and IRR) of the transposon and sometimes internal repeat elements. If the transposons contain "nested" TE, the repeats of these elements are annotated as well.
4. Recombination sites: these sites (*res* sites), which are necessary to resolve Tn*3*-family transposition intermediates, are not well-characterized in most transposons. However, *res* sites and their component sub-sites are annotated when known (Figure 5, green bars).
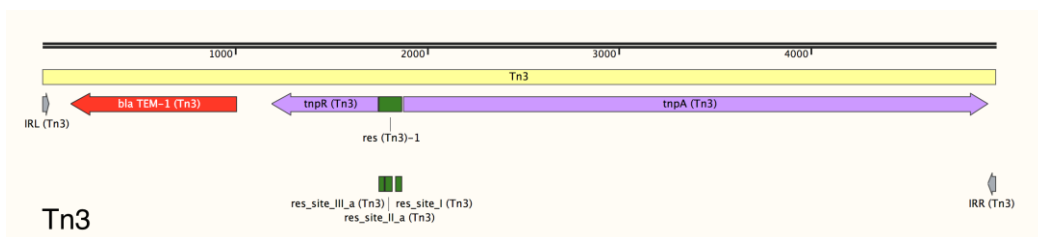
Figure 5. Annotated features of transposon Tn3, including recombination (res) sites (green bars).

TnCentral also includes a number of composite transposons (Fig 4 and 5) which are annotated using the same rules as for Tn*3* family transposons and can be interrogated using the same search terms.
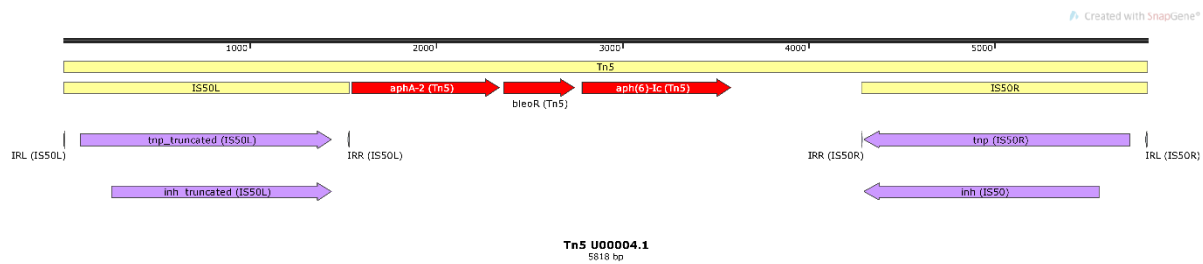

Figure 6. The TnCentral annotated copy of Tn5.

Annotation file

Finally, for each transposon, the user can download a file in GenBank format, which includes the| annotations added by TnCentral.

TnCentral Curation Workflow

The TnCentral curation workflow is depicted in Figure 6. TnFinder scripts are run against RefSeq and other sequences databases and GenBank files potentially containing TE are retrieved. TE sequences are isolated and annotated using SnapGene, a software tool for visualizing and documenting nucleotide sequences and their features. Features of interest (i.e., protein coding genes, TE, repeat elements, and recombination sites) are annotated according to detailed curation guidelines (see section "For Curators"). Fully annotated features are saved in a SnapGene Custom Library. New transposon sequences can be searched against this library, enabling detection of features previously identified in other TEs. All annotated TE files are checked by a second curator to ensure that they are complete and consistent with the curation guidelines. An image file showing a color-coded map of TE features and an enhanced GenBank file containing all annotations are exported from SnapGene. Information from the GenBank file is used to populate the TnCentral database, which, in turn, serves as the backend for the TnCentral web portal.
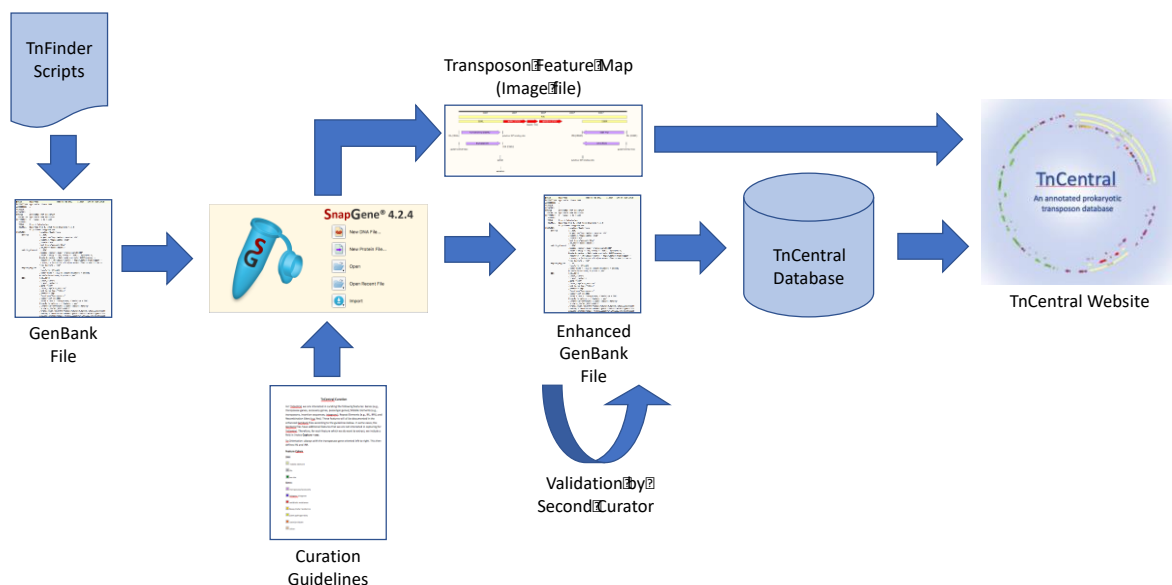
Figure 7. The TnCentral curation workflow.

Future developments: TnCentral 2.0

Although there are resources devoted to curating and tracking the spread of AMR (CARD ([21](#)), ResFinder ([22](#)), and AMRFinder (https://www.ncbi.nlm.nih.gov/ pathogens/antimicrobial-resistance/AMRFinder/)), these resources do not include either the transposon or the IS context. This is a key limitation as transposon movement and inter-transposon/plasmid recombination are the major driving forces of on-going plasmid evolution and accumulation of multiple antibiotic resistance genes. Moreover, recent advances in long-read (Single Molecule, Real-Time or SMRT) sequencing, such as PacBio and Oxford Nanopore technology, are producing reliable sequence data on AMR plasmids, giving us an unprecedented opportunity to track their on-going evolution. Thus, there is a need for a platform for analysis of clinical plasmid sequences with respect to the combinations of AMR-carrying transposons they contain. Such a platform will significantly facilitate understanding of how particular AMR combinations have arisen, what combinations are a risk of arising in the future, as well as which new hosts may gain AMR.

We propose to build a comprehensive mobile genetic element resource, TnCentral 2.0, taking advantage of the gold-standard transposon knowledge in TnCentral, so that the evolution of bacterial AMR can be understood in terms of IS and transposon action (aka the mobilome). It is our hope that with the patterns of changes learned, it will be possible to predict the emergence of future clinical threats giving us a headroom to mount defences in time. Specifically, we plan to use a "three layered approach" (see Figure 8).
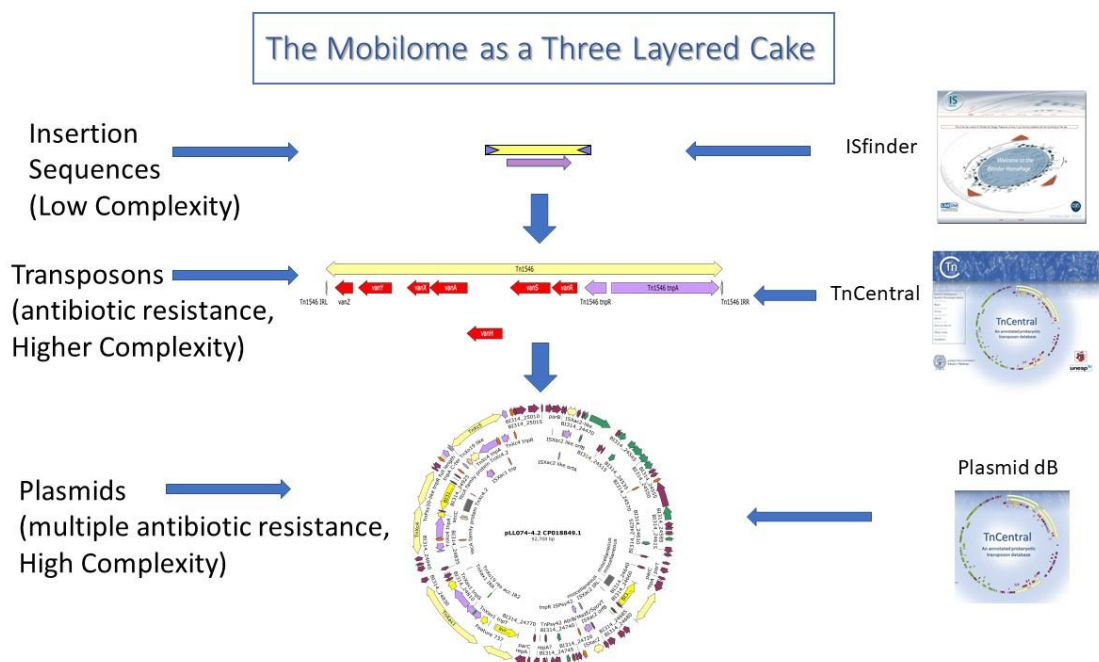
Figure 8. This figure illustrates the general flow of the project (from top to bottom). ISfinder (top layer) and TnCentral (middle layer) are already operational. Together, the two databases will provide the information to semiautomatically annotate the mobilome of plasmids (bottom layer). The final stage (not shown) would be to use these databases to semiautomatically annotate plasmid groups of interest such as those isolated from various clinical outbreaks using software built upon ISsaga, our tool for semi-automated annotation of IS.

Specific Aims.

Aim 1. Expand Tn3finder to identify a variety of clinically relevant AMR transposon sequences. Curate these transposons using the already developed TnCentral framework. Integrate ISfinder data on IS elements, which will be essential for characterizing derivative transposon sequences.

Aim 2. Using ISsaga as a starting point, develop tools for transposon analysis in long-read sequence data of plasmids from clinical samples. This includes: (i) annotating clinically important transposons and their flanking characteristic sequence signatures in AMR plasmid populations to facilitate reconstruction of the order of transposition and recombination events that led to multiple AMR acquisition (see Figures 1 and 2 and (10,11,23–25); (ii) building a precisely annotated AMR plasmid database; (iii) comparing plasmid populations from different AMR outbreaks to identify patterns in the emergence of multiple AMR; and (iv) making predictions about how plasmid populations might evolve additional resistances in the future. We will collaborate with researchers who have sequencing data on clinical samples and hands-on expertise in AMR (Drs John Dekker, NIHCC and NIAID and Patrick McGann, Erik Snesrud, WRIAR), as well as experts with specific knowledge of transposition mechanisms (Dr. Fred Dyda, NIDDK) and of specific transposon families or non-AMR passenger genes (Prof. Bernard Hallet, Université de Louvain la Neuve, Belgium; Prof Laurence van Melderen, Université Libre de Bruxelles, Belgium; Prof. Joseph Peters, Cornell University; Prof. Phoebe Rice, University of Chicago).

Aim 3. Develop TnCentral 2.0 as an interactive data sharing repository where the clinical and basic research community can search, visualize and interpret their own data on AMR transposons, plasmids, and host organisms, as well as analyze their own sequence data. Establish a framework for community users to incorporate their data into TnCentral 2.0, thereby strengthening the resource for future analyses.

References

1. Spellberg B, Guidos R, Gilbert D, Bradley J, Boucher HW, Scheld WM, et al. The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. Clin Infect Dis. 2008 Jan 15;46(2):155–164.

2. O'Neill J. Tackling Drug-Resistant Infections Globally:' ' Final Report and Recommendations. London?: Welcome Foundation HM Government UK; 2016 May.

3. Siguier P, Varani A, Perochon J, Chandler M. Exploring bacterial insertion sequences with ISfinder: objectives, uses, and future developments. Methods Mol Biol. 2012;859:91–103.

4. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D32–6.

5. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol. 2011 Mar 28;12(3):R30.

6. Craig NL, Chandler M, Gellert M, Lambowitz A, Rice PA, Sandmeyer S. Mobile DNA III. 3rd ed. Washington, DC: ASM; 2015.

7. Mahillon J, Chandler M. Insertion sequences. Microbiol Mol Biol Rev. 1998 Sep;62(3):725–774.

8. Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev. 2014 Sep;38(5):865–891.

9. Siguier P, Gourbeyre E, Varani A, Ton-Hoang B, Chandler M. Everyman's guide to bacterial insertion sequences. Microbiol Spectr. 2015 Apr;3(2):MDNA3–0030.

10. He S, Chandler M, Varani AM, Hickman AB, Dekker JP, Dyda F. Mechanisms of Evolution in High-Consequence Drug Resistance Plasmids. MBio. 2016 Dec 6;7(6).

11. He S, Hickman AB, Varani AM, Siguier P, Chandler M, Dekker JP, et al. Insertion Sequence IS26 Reorganizes Plasmids in Clinically Isolated Multidrug-Resistant Bacteria by Replicative Transposition. MBio. 2015 Jun 9;6(3):e00762.

12. Ferreira RM, de Oliveira ACP, Moreira LM, Belasque J, Gourbeyre E, Siguier P, et al. A TALE of transposition: Tn3-like transposons play a major role in the spread of pathogenicity determinants of Xanthomonas citri and other xanthomonads. MBio. 2015 Feb 17;6(1):e02505–14.

13. Shapiro JA. Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proc Natl Acad Sci USA. 1979 Apr;76(4):1933–1937.

14. Yano H, Genka H, Ohtsubo Y, Nagata Y, Top EM, Tsuda M. Cointegrate-resolution of toluene-catabolic transposon Tn4651: determination of crossover site and the segment required for full resolution activity. Plasmid. 2013 Jan;69(1):24–35.

15. Nicolas E, Lambin M, Dandoy D, Galloy C, Nguyen N, Oger CA, et al. The Tn3-family of Replicative Transposons. Microbiol Spectr. 2015 Aug;3(4).

16. Stark WM, Sherratt DJ, Boocock MR. Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions. Cell. 1989 Aug 25;58(4):779–790.

17. Mahillon J, Lereclus D. Structural and functional analysis of Tn4430: identification of an integrase-like protein involved in the co-integrate-resolution process. EMBO J. 1988 May;7(5):1515–1526.

18. Genka H, Nagata Y, Tsuda M. Site-specific recombination system encoded by toluene catabolic transposon Tn4651. J Bacteriol. 2002 Sep;184(17):4757–4766.

19. Heffron F, McCarthy BJ, Ohtsubo H, Ohtsubo E. DNA sequence analysis of the transposon Tn3: three genes and three sites involved in transposition of Tn3. Cell. 1979 Dec;18(4):1153–1163.

20. Gill R, Heffron F, Dougan G, Falkow S. Analysis of sequences transposed by complementation of two classes of transposition-deficient mutants of Tn3. J Bacteriol. 1978 Nov;136(2):742–756.

21. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 2017 Jan 4;45(D1):D566–D573.

22. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012 Nov;67(11):2640–2644.

23. Snesrud E, He S, Chandler M, Dekker JP, Hickman AB, McGann P, et al. A Model for Transposition of the Colistin Resistance Gene mcr-1 by ISApl1. Antimicrob Agents Chemother. 2016 Nov;60(11):6973–6976.

24. Snesrud E, McGann P, Chandler M. The Birth and Demise of the ISApl1-mcr-1-ISApl1 Composite Transposon: the Vehicle for Transferable Colistin Resistance. MBio. 2018 Feb 13;9(1).

25. Weingarten RA, Johnson RC, Conlan S, Ramsburg AM, Dekker JP, Lau AF, et al. Genomic analysis of hospital plumbing reveals diverse reservoir of bacterial plasmids conferring carbapenem resistance. MBio. 2018 Feb 6;9(1).

Acknowledgements