

## SI Appendix

# Structural basis of a novel histidine-DNA nicking/joining mechanism for gene transfer and promiscuous spread of antibiotic resistance

Radoslaw Pluta<sup>a,b,1,2</sup>, D. Roeland Boer<sup>a,b,1,3</sup>, Fabián Lorenzo-Díaz<sup>c,4</sup>, Silvia Russi<sup>a,b,5</sup>, Hansel Gómez<sup>a,d</sup>, Cris Fernández-López<sup>c</sup>, Rosa Pérez-Luque<sup>a,b</sup>, Modesto Orozco<sup>a,d,e</sup>, Manuel Espinosa<sup>c</sup> and Miquel Coll<sup>a,b,6</sup>

## SI Appendix, Material and Methods

### Bacterial strains, plasmids and mutagenesis

*Escherichia coli* BL21 (DE3) ( $\lambda$ DE3 (*lacI lacUV5-T7 gene 1 ind1 sam7 nin5*)  $F^- dcm ompT hsdS(r_B^- m_B^+)$  *gal*) was used for purification of MobM and MobMN199 proteins. *E. coli* B834 (DE3) ( $\lambda$ DE3 (*lacI lacUV5-T7 gene 1 ind1 sam7 nin5*)  $F^- dcm ompT hsdS(r_B^- m_B^+)$  *gal met*) was employed for purification of MobMN199 with SeMet (1). *E. coli* M15 / pREP4 (*nalS, strS, rifS, thi^-, lac^-, ara^+, gal^+, mtI^-, F^-, recA^+, uvr^+, lon^+*) was used to express the C-terminally His-tagged MobMN199 mutants H22A, H22Y, R25, Y44F, E129A and E129Q. Purified pMV158 supercoiled DNA and other plasmids used for expression or as substrates of MobM and MobMN199 proteins were prepared by two consecutive CsCl gradients (2). To obtain the MobMN199 variants, point mutations were introduced using the GeneTailor Site-Directed Mutagenesis System (Invitrogen).

### Protein expression and purification

Expression and purification of full-length MobM and of MobMN199 was performed as described previously (3). Briefly, the plasmid used for controlled expression of full length MobM protein was pLGM2 (based on pET5 vector; 4). The plasmid used to purify MobMN199 was pET24(b) (Novagen). The proteins were purified from 4 l (TY-Km at 37°C) of *E. coli* cultures after induction with 1 mM IPTG (30 min) and addition of rifampicin (200 mg/ml, 90 min). Cells were concentrated 100-fold, resuspended in buffer A (20 mM Tris-HCl pH 7.6, 1mM EDTA, 1mM dithiothreitol, 5% (v/v) glycerol) containing 1M NaCl and a tablet of protease inhibitor cocktail, and lysed by passage through a French pressure cell. The cell extract was precipitated with 0.2% (v/v) polyethyleneimine, and proteins in the supernatant were precipitated at 70% (w/v) ammonium sulphate saturation. Proteins in the precipitate were dissolved in buffer A with 500mM NaCl and purified by two chromatographic steps (heparin-agarose, and HiLoad Superdex 200 gel-filtration). Fractions containing pure protein (>98%) were pooled, concentrated until the final concentration was 5 mg/ml protein and stored at -80°C. In these conditions, the protein retained full activity for at least 1 year. Edman's sequential degradation was used to determine the first 10 residues of the N-

terminal sequence of the proteins, and their concentration was calculated by spectrophotometry and by determination of their amino acid composition. The mutant proteins had a His-tag placed at the C-terminal of the protein and were purified with Ni-affinity chromatography.

### **Relaxation activity and relaxosome formation (covalent pull-down) assays**

Assays of relaxation of scDNA by purified proteins were performed essentially as reported (3, 4). Purified DNA from pMV158 (500 ng) was treated with 100, 200, or 400 nM concentrations of purified MobMN199 proteins (wt or mutants) in reaction mixtures of 20  $\mu$ l volume (25 mM Tris-HCl pH 7.6, 0.1 mM EDTA, 10% glycerol (v/v), 1 mM DTT, NaCl 50 mM, 8 mM MnCl<sub>2</sub>). Samples were incubated and treated as reported (3). Generation of open circular forms (FII) by the nicking activity of the various relaxase variants was tested by electrophoresis on 1% (w/v) agarose gels and staining with 1  $\mu$ g ml<sup>-1</sup> ethidium bromide.

For relaxosome formation assays, supercoiled pMV158 DNA (8 nM) was relaxed with 250 nM of protein MobMN199 (wt or mutants) in 20  $\mu$ l reaction volumes. Reactions were stopped by the addition of 1% SDS, and stable DNA-protein complexes were precipitated by treatment with 180 mM KCl, at 0°C, for 10 min (34). The SDS binds protein but not free DNA, and the addition of KCl would generate a potassium-SDS precipitate. Thus, this method would selectively precipitate detergent-resistant protein-DNA complexes, leaving free DNA in the supernatant. After centrifugation, the DNA of the pellets was precipitated with NaCl (0.3 M) and ethanol, obtaining the DNA-protein complexes. Samples were resuspended in Tris-HCl (10 mM, pH 8.0)-EDTA (1 mM) buffer and incubated for 30 min at 37°C with 0.1 mg ml<sup>-1</sup> proteinase K, and then loaded in 1% agarose gels (5). Separation of the plasmid forms was achieved by electrophoresis on 1% agarose in the presence of 1  $\mu$ g ml<sup>-1</sup> ethidium bromide, at 100 V for 1 h.

### **Mobilization experiments**

*In vivo* mobilization of pMV158 and derivatives between pneumococcal strains was done essentially as described (6); all these assays were performed five times under independent conditions.

### **MobM-DNA complexes preparation**

For the preparation of the native and SeMet-labeled MobMN199-DNA26 complexes, protein in buffer B (500 mM NaCl, 20 mM Tris-HCl pH 7.6, 1 mM EDTA, 1% glycerol, 1 mM DTT) was mixed at 1:1.2 protein:DNA stoichiometry with a previously annealed oligonucleotide dna26 encompassing the 26 nucleotides upstream from the pMV158<sub>oriT</sub> cleavage site (Fig. 1C; all oligonucleotides are from Biomers, Ulm, Germany) and incubated on ice for 1h. The other complexes were prepared using a similar procedure but mixing MobMN199 in buffer M (100 mM

NaCl, 20 mM Tris-HCl, pH 7.5, 15 mM MnCl<sub>2</sub>·4H<sub>2</sub>O, 1% glycerol, 1 mM DTT) with the respective dsDNAs (Fig. 1C). All oligonucleotides were previously annealed, except for nic0-SPO, because the thiophosphate group was not stable at high temperatures. The resulting complexes were purified by size-exclusion chromatography, concentrated to 4 mg ml<sup>-1</sup> and used for crystallization trials.

### **Crystallization**

Crystals of the SeMet MobMN199-DNA26 and the MobMN199-DNA26 complexes were grown by sitting-drop vapor diffusion of the complex in buffer B against a crystallization buffer containing 100 mM NaAc, pH 4.6 and 10% PEG 6000, at 4°C. Crystals of the MobMN199 Nic0, Nic0+P and Nic0+SP complexes were grown by sitting-drop vapor diffusion of the complex in sample buffer M at 20°C against a crystallization buffer containing 100 mM NaAc, pH 4.6 and 18-23% PEG 6000. Crystals of the Nic0+1 complex were obtained in a similar way against a crystallization buffer containing 100mM MES, pH 6.5, 200mM MgAcetate·4H<sub>2</sub>O and 20% PEG 8000. Microseeding was used for Nic0+P and Nic0+SP complexes crystal growth. All crystals were harvested in cryo-solutions containing crystallization cocktail supplemented with 15% glycerol. For the high pH data, the crystals were soaked in cryo-solutions in which the original pH 4.6 buffer was replaced by buffers of a different pH. For the Nic0\_A structure 100 mM NaAc at pH 5.5 was used and for the Nic0+SP structure 100 mM NaAc at pH 6.8.

### **X-ray data collection, structure determination and refinement**

The details of the data collection at beamlines ID14-4, ID23-1 and ID23-2 of the European Synchrotron Radiation Facility (ESRF) and at beamline PROXIMA1 of the French Synchrotron Facility Soleil are summarized in Table S4. All data were processed with iMOSFLM (7) and scaled with SCALA (8). The experimental phases were obtained by the SAD method, using the SeMet-DNA26 data set, collected at the Se absorption edge (0.9795Å). Seven Se sites were located in the asymmetric unit using the program SHELXD (9) and protein phases were calculated with PHASER (10), improving the initial electron density map by density modification with the program PIRATE (11). 30% of the polypeptidic chain was automatically traced with the program RESOLVE (12) and completed manually with the program COOT (13). This initial SeMet structure was trimmed, removing the side-chain of H22, the active site metal ion and bases G8-A9-A10-T11 and G24-T25, before using it for molecular replacement with PHASER against the data of the remaining complexes.

Refinement of all structures was performed with the program REFMAC 5.5.0102 (14), interspersed with the manual adjustment of the model to the electron density map using the program COOT (13). Further model improvement and validation was done using the MolProbity

web server (15). Final refinement statistics are shown in Table S4. All structures and corresponding structure factors were deposited in the PDB with accession codes 4LVI, 4LVJ, 4LVK, 4LVL, 4LVM and 5N2Q. Images were prepared with PYMOL (16).

### **Unambiguous assignment of the identity of the Mn<sup>2+</sup> ion**

Crystals of the Nic0 complex were diffracted at BM14 (ESRF, Grenoble, France). An x-ray absorption scan revealed an absorption peak near the Mn K-edge ( $E=6.550$  keV). Two complete anomalous datasets were measured at  $E=6.519$  keV and at  $E=6.550$  keV. The anomalous map calculated for the higher energy dataset (Fig. S2E) showed a strong peak ( $\rho > 8\sigma$ ) at the position of the active site metal ion, whereas no discernable contribution was observed for the lower energy near-edge data. Since no other element has such absorption behavior at these wavelengths, the element of the atom bound at the active site is Mn.

### **Theoretical calculations**

The crystallographic structures were used to model a pre-catalytic complex and conduct theoretical calculations -including classical Molecular Dynamics (MD) and Quantum Mechanics/Molecular Mechanics (QM/MM) methods- in the study of the catalytic mechanism of MobM. We followed a standard equilibration protocol including energy minimizations and classical MM Molecular Dynamics (MD) simulations steps with AMBER14 (17) to relax and equilibrate the systems using the amber force field ff14SB (18) and Parmbsc1 (19) for protein and DNA, respectively. The final equilibrated structures were used as starting point for the Quantum Mechanics/Molecular Mechanics (QM/MM) calculations where a total of 73 or 91 atoms were considered in the QM subset for the dynamic (i.e. Umbrella Sampling) and geometry optimization calculations, respectively (Fig. S5A).

The catalytic process was studied by using a combination of distances between the nucleophile H22 NE2 nitrogen and the Gua26 O3' oxygen atom with respect to the reaction center (i.e. phosphorus atom of the scissile phosphate, Fig. 6A-B), thus modeling the bond forming and breaking processes. This reaction coordinate was scanned in steps of  $0.25 \text{ \AA}$  in Umbrella Sampling (US) and geometry optimizations to obtain free (FES) and potential energy surfaces (PES), respectively. The PES scan provided the starting structure for subsequent full optimization of the transition state (TS) and product of the catalytic reaction. Frequency calculations were performed to ensure that one imaginary frequency and a suitable transition vector characterize the optimized TS structure. The VMD program (20) was used to obtain the H22Y mutant of the enzyme, the relevant geometrical parameters and the figures showing molecular structures.

## Detailed description of the modeling process, MD protocols and QM/MM calculations

We added the TTCT polynucleotide to the 5' terminus of the single-stranded DNA molecule present in available structures to model the enzyme-substrate (ES) complexes addressed in the mechanistic studies. More specifically, we considered crystal structures Nic0+1/molA and Nic0+SP to model a Michaelis complex, where H22 was initially coordinating or not the  $\text{Mn}^{2+}$  ion, respectively.

The protonation states of the titratable residues of the protein (Asp, Arg, Glu, His, Lys) were chosen on the basis of the  $\text{p}K_{\text{a}}$  values predicted by the empirical H++ v3.1 procedure (21). The LEaP module of the AMBER 14 software suite (17) was used to solvate the systems with a truncated octahedral box of TIP3P water molecules (margin distance of 12 Å; final size of ~34000 atoms), and 10  $\text{Na}^+$  ions were added to neutralize the systems.

A series of minimizations (1000 optimization steps of each steepest descent and conjugate gradient method) of hydrogen atoms, water molecules, DNA substrate, and finally the full system were performed. Later on, the system was gradually heated to 300 K over 100 ps, during which a weak constraint of  $5 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$  was applied to the alpha carbon and phosphorus atoms of the protein and DNA substrate. Then, equilibration and production of MD simulations of 100 ps and 5 ns, respectively, were performed using the NPT ensemble.

All minimizations and equilibration step simulations used a non-bonded cut-off of 15Å in the short-range electrostatic and Lennard-Jones interaction calculations. The Berendsen barostat (22) was used to maintain a constant pressure of 1 atm, and the temperature (300 K) was controlled using Langevin dynamics with a collision frequency of  $5 \text{ ps}^{-1}$ . The Particle Mesh Ewald (PME) method (23) was applied to calculate long-range electrostatic interactions. The time step was set to 2 fs, and all bonds involving hydrogen atoms were constrained using the SHAKE algorithm (24) during the MD simulations.

Having obtained equilibrated structures, we removed all the solvent molecules that were more than 35 Å away from the manganese ion in order to reduce the system size to ~20000 atoms before performing QM/MM calculations. Several QM-MM partitions were considered in test calculations: including just the deoxyribose rings of nucleotides G26 and phosphate group from T27, side chain of E129 and the divalent ion (37 atoms); incorporating the full coordination sphere of the ion (i.e. side chain of residues H126, H133, H135; 73 atoms); adding the side chain of R24 to the previous selection (91 atoms); and, finally, considering full G26 and T27 and terminal groups from neighboring nucleotides (148 atoms). The resulting energy profiles were very similar in the last three QM/MM partition schemes, and for the sake of efficiency, we ended up considering the one with 73 and 91 atoms in the QM subset for the QM/MM dynamic and geometry optimization calculations, respectively (Fig. S5A).

An electronic embedding scheme (25) was adopted in the QM/MM calculations, and no cut-

offs were introduced for the non-bonding MM and QM/MM interactions. Hydrogen link atoms were used to treat the QM/MM boundary with the charge shift model (26, 27). Moreover, a sextuplet state was considered for the Mn<sup>2+</sup> ion.

Orca v3.03 (28) was used in interface with AMBER (29) for the QM/MM MD simulations corresponding to the Umbrella Sampling (US) calculations. A force constant of 100 kcal/(mol Å<sup>2</sup>) was applied to the reaction coordinate in each of the resulting 16 US windows. A total of 20 ps of QM(BLYP (30, 31) / 6-31g(d) (32-34) /MM were run for each sampling window, where the last 17 were considered when computing the Potential of Mean Force (PMF) with the vFEP method (35).

The QM/MM geometry optimizations were performed with the modular program package ChemShell (36). TURBOMOLE (37) was used for the energy/gradient calculations in the QM partition at the BP86 (30, 38-40) /def2-SVP (41, 42) level of theory and taking advantage of the Resolution-of-the-Identity (RI) approximation (41, 43). MM energies and gradients were retrieved from the DL\_POLY (44) program. Algorithms L-BFGS (27, 45) and P-RFO (46, 47) were used in minima and transition state (TS) searches, respectively. Both methodologies are implemented in the HDLCopt (48) module of ChemShell.

## References

1. Studier FW, Rosenberg AH, Dunn JJ, & Dubendorff JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Meth. Enzymol.* 185:60-89.
2. del Solar G, Díaz R, & Espinosa M (1987) Replication of the streptococcal plasmid pMV158 and derivatives in cell-free extracts of *Escherichia coli*. *Mol. Gen. Genet.* 206:428-435.
3. Lorenzo-Díaz F, *et al.* (2011) The MobM-relaxase domain of plasmid pMV158: thermal stability and activity upon Mn<sup>2+</sup>-and DNA specific-binding. *Nucleic Acids Res.* 39:4315-4329.
4. Guzmán LM & Espinosa M (1997) The mobilization protein, MobM, of the streptococcal plasmid pMV158 specifically cleaves supercoiled DNA at the plasmid *oriT*. *J Mol Biol* 266(4):688-702.
5. Trask DK, DiDonato JA, & Muller MT (1984) Rapid detection and isolation of covalent DNA/protein complexes: application to topoisomerase I and II. *EMBO J* 3:671-676.
6. Lorenzo-Díaz F & Espinosa M (2009) Large-scale filter mating assay for intra- and inter-specific conjugal transfer of the promiscuous plasmid pMV158 in Gram-positive bacteria. *Plasmid* 61(1):65-70.
7. Leslie AGW & Powell HR (2007) Processing Diffraction Data with Mosflm. *Evolving Methods for Macromolecular Crystallography*, eds Read RJ & Susman JL, Vol 245, pp 41-51.
8. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D Biological Crystallography* 62(Pt 1):72-82.
9. Schneider TR & Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallographica Section D Biological Crystallography* 58(Pt 10 Pt 2):1772-1779.
10. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *Journal of applied crystallography* 40(Pt 4):658-674.
11. Cowtan K (2001) Fast Fourier feature recognition. *Acta Crystallographica Section D Biological Crystallography* 57(Pt 10):1435-1444.
12. Terwilliger TC (2003) SOLVE and RESOLVE: automated structure solution and density modification. *Methods in Enzymology* 374:22-37.
13. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography* 60(Pt 12 Pt 1):2126-2132.
14. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D Biological Crystallography* 53(Pt 3):240-255.
15. Davis IW, *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35(Web Server issue):W375-383.
16. DeLano WL (2002) The Pymol Molecular Graphics System. *on World Wide Web* <http://www.pymol.org>.
17. Case DA, *et al.* (2014) AMBER 14 (University of San Francisco).
18. Maier JA, *et al.* (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from

- ff99SB. *Journal of Chemical Theory and Computation* 11(8):3696-3713.
19. Ivani I, et al. (2016) Parmbsc1: a refined force field for DNA simulations. *Nat Meth* 13(1):55-58.
  20. Foster JP & Weinhold F (1980) Natural hybrid orbitals. *Journal of the American Chemical Society* 102(24):7211-7218.
  21. Anandkrishnan R, Aguilar B, Onufriev AV (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* 40:W537-W541.
  22. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81: 3684-3690.
  23. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An Nlog (N) method for Ewald sums in large systems. *J Chem Phys* 98:10089-10092.
  24. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 23: 327-341.
  25. Dirk B, Walter T (1996) Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J Phys Chem* 100:10580-10594.
  26. de Vries AH, et al. (1999) Zeolite structure and reactivity by combined quantum-chemical-classical calculations. *J Phys Chem B* 103:6133-6141
  27. Sherwood P, et al. (1997) Computer simulation of zeolite structure and reactivity using embedded cluster methods. *Farad Discuss* 106:79-92.
  28. Neese F (2012) The ORCA program system. *Wiley Interdiscip Rev Comput Mol Sci* 2:73-78.
  29. Götz AW, Clark MA, Walker RC (2014) An extensible interface for QM/MM molecular dynamics simulations with AMBER. *J Comput Chem* 35:95-108.
  30. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A* 38:3098-3100.
  31. Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 37:785-789.
  32. Francl MM, et al. (1982) Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J Chem Phys* 77:3654-3665.
  33. Hariharan PCP, Pople JA (1973) The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor Chim Acta* 28:213-222.
  34. Rassolov VA, Pople JA, Ratner MA, Windus TL (1998) 6-31G\* basis set for atoms K through Zn. *J Chem Phys* 109:1223-1229.
  35. Lee T-S, Radak BK, Pabis A, York DM (2013) A New Maximum Likelihood Approach for Free Energy Profile Construction from Molecular Simulations. *J Chem Theory Comput* 9:153-164.
  36. Sherwood P, et al. (2003) QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J Mol Struct-THEOCHEM* 632:1-28.
  37. Ahlrichs R, Bär M, Häser M, Horn H, Kölmel C (1989) Electronic structure calculations on workstation computers: The program system turbomole. *Chem Phys Lett* 162:165-169.
  38. Perdew JP (1986) Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys Rev B* 33:8822-8824.
  39. Slater JC (1951) A simplification of the Hartree-Fock method. *Phys Rev* 81:385-390.
  40. Vosko SH, Wilk L, Nusair M (1980) Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can J Phys* 58:1200-1211.
  41. Karin E, Florian W, Oliver T, Reinhart A (1997) Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. *Theor Chem Acc* 97:119-124.
  42. Weigend F, Ahlrichs R (2005) Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. *Phys Chem Chem Phys* 7:3297-3305.
  43. Eichkorn K, Treutler O, Öhm H, Häser M, Ahlrichs R (1995) Auxiliary basis sets to approximate Coulomb potentials. *Chem Phys Lett* 240:283-289.
  44. Smith W, Forester TR (1996) DL\_POLY\_2.0: a general-purpose parallel molecular dynamics simulation package. *J Mol Graph* 14:136-141.
  45. Nocedal J (1980) Updating Quasi-Newton Matrices with Limited Storage. *Math Comp* 35:773-782.
  46. Baker J (1986) An algorithm for the location of transition states. *J Comput Chem* 7:385-395.
  47. Banerjee A, Adams N, Simons J, Shepard R (1985) Search for stationary points on surfaces. *J Phys Chem* 89:52-57.
  48. Billeter SR, Turner AJ, Thiel W (2000) Linear scaling geometry optimisation and transition state search in hybrid delocalised internal coordinates. *Phys Chem Chem Phys* 2:2177-2186.

## SI Appendix, Supplementary Tables

**Table S1. Distribution of relaxase families in Gram-positive human pathogens<sup>a</sup>.**

MOB (CDD) Family <sup>b</sup>	E-value threshold for specific and non-specific hits <sup>c</sup>	<i>Staph. aureus</i> (% of all <i>S.a.</i> hits) <sup>d</sup>	<i>Streptococcus</i>			<i>Enterococcus</i>		Firmicutes	Bacteria	Sum of CDC G+ pathogens (columns III - VIII)	Percentage of <i>S. aureus</i> among all bacterial hits	Percentage of CDC G+ pathogens among all bacterial hits
			<i>agalactiae</i>	<i>pneumoniae</i>	<i>pyogenes</i>	<i>faecalis</i>	<i>faecium</i>					
MOB <sub>V</sub> (Mob_Pre)	1.5e-44	3,432 (83%)	453	86	11	200	109	5,765	5,838	4,291	59%	74%
	1.0e-2	5,716 (85%)	505	256	11	229	419	11,668	14,974	7,136	38%	48%
MOB <sub>P</sub> (Relaxase)	7.7e-50	394 (9%)	7	706	0	24	224	2,211	5,893	1,355	7%	23%
	1.0e-2	462 (7%)	735	7,021	147	461	523	17,810	45,468	9,349	1%	21%
MOB <sub>Q</sub> (MobA_MobL)	1.7e-73	323 (8%)	58	2,499	19	50	24	5,920	10,181	2,973	3%	29%
	1.0e-2	482 (7%)	58	2,906	22	56	27	7,799	23,434	3,551	2%	15%
MOB <sub>F</sub> (TrwC)	5.2e-15	0	0	0	0	0	0	30	6,546	0	0%	0%
	1.0e-2	0	0	0	0	0	0	66	16,029	0	0%	0%
MOB <sub>H</sub> (Tral_2_C)	8.4e-29	0	0	2	0	0	0	2	5,632	2	0%	0%
	1.0e-2	0	0	2	0	0	0	64	10,397	2	0%	0%
MOB <sub>C</sub> (Replic_Relax)	4.2e-27	0	0	20	0	0	0	1,461	2,650	20	0%	1%
	1.0e-2	45 (1%)	0	1,541	0	298	11	5,985	10,146	1,895	1%	19%

<sup>a</sup>Pathogen list was taken from the CDC Report “Antibiotic resistance threats in the United States, 2013”.

<sup>b</sup>Data taken from the NCBI Conserved Domain Database (20.02.2017).

<sup>c</sup>**E-value** (Expected Value) is a parameter that describes the number of hits that can be expected by chance when searching a database of a particular size. False positive results should be very rare with the default setting of 0.01. **Non-specific hits** (“**Related Proteins**”) are hits that meet or exceed the [RPS-BLAST](#) threshold for statistical significance (default E-value cutoff of 0.01). **Specific hits** (“**Specific Proteins**”) are the top-ranking [RPS-BLAST](#) hits that meet or exceed a domain-specific E-value threshold. The domain-specific E-value threshold represents a very high confidence that the query sequence belongs to the same protein family as the sequences (“**Representatives**”) used to create the domain model. ([https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd\\_help.shtml](https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd_help.shtml)).

<sup>d</sup>The pools of the specific hits and of the non-specific hits are compared separately.



**Table S2. Percentage<sup>a</sup> of genomes (plasmids included) of Gram-positive human pathogens that contain relaxases<sup>b</sup>.**

MOB (CDD) Family <sup>b</sup>	E-value threshold: specific non-specific	<i>Staphylococcus</i>	<i>Streptococcus</i>			<i>Enterococcus</i>	
		<i>aureus</i>	<i>agalactiae</i>	<i>pneumoniae</i>	<i>pyogenes</i>	<i>faecalis</i>	<i>faecium</i>
MOB <sub>V</sub> (Mob_Pre)	4.9e-43	44%	50%	1%	3%	39%	18%
	1.0e-2	73%	56%	3%	3%	45%	68%
MOB <sub>P</sub> (Relaxase)	4.9e-70	5%	1%	10%	0%	5%	36%
	1.0e-2	6%	81%	95%	43%	90%	85%
MOB <sub>Q</sub> (MobA_MobL)	5.8e-51	4%	6%	34%	6%	10%	4%
	1.0e-2	6%	6%	39%	6%	11%	4%
MOB <sub>F</sub> (TrwC)	3.8e-52	0%	0%	0%	0%	0%	0%
	1.0e-2	0%	0%	0%	0%	0%	0%
MOB <sub>H</sub> (TraI_2_C)	8.4e-29	0%	0%	0%	0%	0%	0%
	1.0e-2	0%	0%	0%	0%	0%	0%
MOB <sub>C</sub> (Replic_Relax)	4.2e-27	0%	0%	0%	0%	0%	0%
	1.0e-2	1%	0%	21%	0%	58%	2%

<sup>a</sup>The number of genomes was taken from the NCBI Genome recourse data ([www.ncbi.nlm.nih.gov/genome/](http://www.ncbi.nlm.nih.gov/genome/)):  
*Staphylococcus aureus*: Genome Assembly and Annotation report [7577]; Plasmid Annotation Report [286] = 7863  
*Streptococcus agalactiae*: Genome Assembly and Annotation report [895]; Plasmid Annotation Report [8] = 903  
*Streptococcus pneumoniae*: Genome Assembly and Annotation report [7396]; Plasmid Annotation Report [4] = 7400  
*Streptococcus pyogenes*: Genome Assembly and Annotation report [334]; Plasmid Annotation Report [8] = 342  
*Enterococcus faecium*: Genome Assembly and Annotation report [518]; Plasmid Annotation Report [96] = 614  
*Enterococcus faecalis*: Genome Assembly and Annotation report [484]; Plasmid Annotation Report [29] = 513

<sup>b</sup>See Table S1 legends for definitions

**Table S3. Taxonomy Report on the pMV158 *oriT* sequence<sup>a</sup>.**

Taxonomy	Number of hits
root	118
. Bacteria	104
.. Bacilli	89
... Bacillales	71
.... Staphylococcus	65
..... Staphylococcus epidermidis	3
..... Staphylococcus aureus	29
..... Staphylococcus aureus subsp. aureus	11
..... Staphylococcus aureus subsp. aureus RN4220	1
..... Staphylococcus aureus subsp. aureus CN1	1
..... Staphylococcus aureus subsp. aureus ST398	4
..... Staphylococcus aureus subsp. aureus ECT-R 2	1
..... Staphylococcus aureus subsp. aureus str. JKD6008	1
..... Staphylococcus aureus subsp. aureus Mu3	1
..... Staphylococcus aureus subsp. aureus Mu50	1
..... Staphylococcus aureus subsp. aureus N315	1
..... Staphylococcus aureus subsp. aureus MRSA252	1
..... Staphylococcus aureus 04-02981	1
..... Staphylococcus aureus CA-347	1
..... Staphylococcus hyicus	2
..... Staphylococcus cohnii	2
..... Staphylococcus schleiferi	1
..... Staphylococcus saprophyticus	1
..... Staphylococcus sciuri	1
..... Staphylococcus arlettae	1
.... Bacillaceae	5
.... Bacillus	4
..... Bacillus subtilis	1
..... Bacillus cereus	1
..... Bacillus sp.	2
..... Geobacillus stearothermophilus	1
.... Listeria monocytogenes	1
.. Lactobacillales	18
... Enterococcus	12
.... Enterococcus hirae	1
.... Enterococcus faecium	4
..... Enterococcus faecium DO	1
.... Enterococcus cecorum	5
.... Enterococcus faecalis	1
.... Lactobacillus reuteri	2
.... Streptococcus	4
..... Streptococcus pasteurianus ATCC 43144	1
..... Streptococcus gallolyticus UCN34	1
..... Streptococcus suis BM407	1
..... Streptococcus agalactiae	1
.. environmental samples	11
... uncultured bacterium	10
... uncultured bacterium MID12	1
.. Proteobacteria	4
... Enterobacteriaceae	3
.... Proteus vulgaris	1
.... Escherichia coli	2
... Burkholderia sp. FDS-1	1
. other sequences	14
.. artificial sequences	11
... vectors	10
.... Cloning vector pME10	1
.... S. aureus-E. coli shuttle vector pBUS1-Pcap-HC	1
.... Expression vector pME9	1
.... Shuttle vector pME8	1
.... Expression vector pME10	1
.... GFP expression vector pHapII	1
.... Shuttle vector pMK3	1
.... Shuttle expression-secretion vector pP43NMK	1
.... Cloning vector pND50	1
.... Shuttle vector pHY320PLK	1
... synthetic construct	1
.. plasmids	3
... Plasmid pUB110	2
... Plasmid pMV158	1

<sup>a</sup>Data from NCBI Nucleotide BLAST. The table shows only hits that are at least 95% identical to *oriT*<sub>pMV158</sub>, i.e. at least 39 bases out of 41 must be identical. (Sequenced used: ACACACTTTATGAATATAAAGTATAGTGTGTTACTTTA; 41 nucleotides).

**Table S4. Crystallographic data table.**

Structure	SeMet-DNA26	DNA26	Nic0_B	Nic0_A	Nic0+P	Nic0+SP	Nic0+1
PDB code	-	5N2Q	4LVI	4LVJ	4LVK	4LVL	4LVM
				<b>Data collection</b>			
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub>	<i>P</i> 6 <sub>1</sub> 22	<i>P</i> 6 <sub>1</sub> 22	<i>P</i> 6 <sub>1</sub> 22	<i>P</i> 6 <sub>1</sub> 22	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
				<b>Cell dimensions</b>			
<i>a</i> , <i>b</i> , <i>c</i> (Å); angles (°)	54.57, 65.26, 77.03	43.56, 52.85, 56.04; <i>β</i> = 95.88	111.06, 111.06, 90.74	112.02, 112.02, 90.61	112.74, 112.74, 91.66	112.42, 112.42, 90.65	67.34, 130.12, 87.38
Resolution (Å)	30.00-2.60 (2.74-2.60)	55.73-2.00 (2.11-2.00)	48.10-1.90 (2.00-1.90)	41.00-2.17 (2.29-2.17)	35.60-2.38 (2.51-2.38)	41.10-2.20 (2.32-2.20)	28.40-3.10 (3.18-3.10)
<i>R</i> <sub>merge</sub> (%)	7.7 (27.7)	9.3 (43.6)	10.9 (65.3)	15.3 (70.2)	11.3 (67.1)	15.9 (62.8)	9.2 (42.4)
< <i>I</i> / $\sigma$ ( <i>I</i> )>	15.8 (5.9)	10.8 (2.4)	8.2 (1.4)	7.3 (1.8)	12.2 (2.8)	7.0 (2.4)	9.4 (2.9)
Completeness (%)	99.8 (100.0)	98.6 (90.7)	99.8 (99.6)	98.2 (100.0)	99.3 (97.8)	100 (100)	98.7 (98.1)
Multiplicity	5.8 (6.0)	4.1 (3.5)	5.0 (3.3)	4.2 (4.2)	8.4 (8.1)	4.1 (4.2)	4.1 (4.2)
Anomalous completeness	99.8 (100.0)	-	-	-	-	-	-
Anomalous multiplicity	3.2 (3.2)	-	-	-	-	-	-
				<b>Refinement</b>			
Resolution (Å)	-	55.73 - 2.00	48.10-1.90	35.21-2.17	35.60-2.38	36.79-2.20	28.41-3.10
No. reflections	-	16179	25154	16928	13543	16824	13500
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	-	16.6/23.8	20.6/24.8	18.6/23.2	20.8/25.6	17.9/22.5	25.5/29.4
<b>Number of atoms</b>	-	2277	2224	2219	2109	2279	4031
Protein	-	1576	1634	1661	1572	1672	3144
DNA	-	357	449	449	453	453	864
Ligands/water	-	344	141	109	84	154	23
				<b>B-factors</b>			
Wilson	-	26.8	28.3	28.3	52.9	26.9	76.9
Overall	-	31.7	28.5	26.9	23.3	22.9	58.7
Protein	-	31.4	27.8	27.1	24.4	22.4	59.9
DNA	-	32.3	24.9	26.3	20.2	23.6	57.5
Ligands/water	-	32.5	47.9	27.1	23.2	26.1	37.2
				<b>R.M.S. deviations</b>			
Bond lengths (Å)	-	0.022	0.014	0.021	0.013	0.017	0.008
Bond angles (°)	-	2.135	1.74	2.13	1.63	1.78	1.41
				<b>MolProbity scores</b>			
Overall score	-	1.7 (92 <sup>nd</sup> percentile)	1.6 (91 <sup>st</sup> percentile)	1.8 (93 <sup>rd</sup> percentile)	1.2 (100 <sup>th</sup> percentile)	1.5 (99 <sup>th</sup> percentile)	9.6 (96 <sup>th</sup> percentile)
All-atom clashscore	-	5.1 (97 <sup>th</sup> percentile)	3.8 (98 <sup>th</sup> percentile)	9.7 (90 <sup>th</sup> percentile)	2.1 (100 <sup>th</sup> percentile)	4.7 (98 <sup>th</sup> percentile)	9.6 (96 <sup>th</sup> percentile)
Ramachandran outlier (%)	-	0	0	0	0	0	0
Ramachandran favoured (%)	-	94.1	96.2	96.3	96.3	96.4	93.4

**Table S5. Transfer efficiencies of MobM and Mob variants.**

Protein variant	Transfer efficiency (transconjugants/donor cell)
wt MobM	2,5x10 <sup>-4</sup>
H22A MobM	Below detection limit, <1,5x10 <sup>-10</sup>
D120L Mob_pBHR1*	Below detection limit, <1,0x10 <sup>-8</sup>
E121G Mob_pBHR1*	Below detection limit, <1,0x10 <sup>-8</sup>

\* data from Szpirer *et al.*, 2001 for Mob\_pBHR1 relaxase (MOBV family); Mob\_pBHR1 D120 and E121 are equivalents to MobM D128 and E129

## SI Appendix, Supplementary Figures Legends

**Fig. S1. Distribution of relaxases in NCBI Conserved Domain Database for *S. aureus* and other bacteria.** See Table S1 legend for definition.

**Fig. S2. MobM interactions with DNA and manganese ion.** Cartoon presentation of the two types of MobM199-DNA complexes determined in this study, the Nic0 (A) and DNA26 (B) types. (C) DNA26 structure and a symmetry related interacting complex. (D) Symmetry related mates interactions in the active site of the DNA26 structure. (E) Unambiguous assignment of Mn<sup>2+</sup> ion identity in the Nic0\_A structure. (F) MobM electrostatic surface potential (Nic0+SPO structure).

**Fig. S3. MobM-DNA structural and conservation details.** (A) DNA hairpin minor groove penetration by and conserved R71 (MOB<sub>V1</sub> subfamily) and R74 (MOB<sub>V</sub> family). (B) Stereo cartoon-ribbon-sticks view on MobM-ssDNA interactions (Nic0\_A structure). (C) Inverted repeats (IR) and possible DNA hairpins within the plasmid pMV158 *oriT*. (D) Mob\_Pre/MOB<sub>V</sub> family conserved residues and motifs. In the upper part, residue frequency bars are presented; the darkness of green correlates with the highest scores in the position-specific scoring matrix (PSSM). The panel was adapted from the NCBI CDD PSSM Viewer.

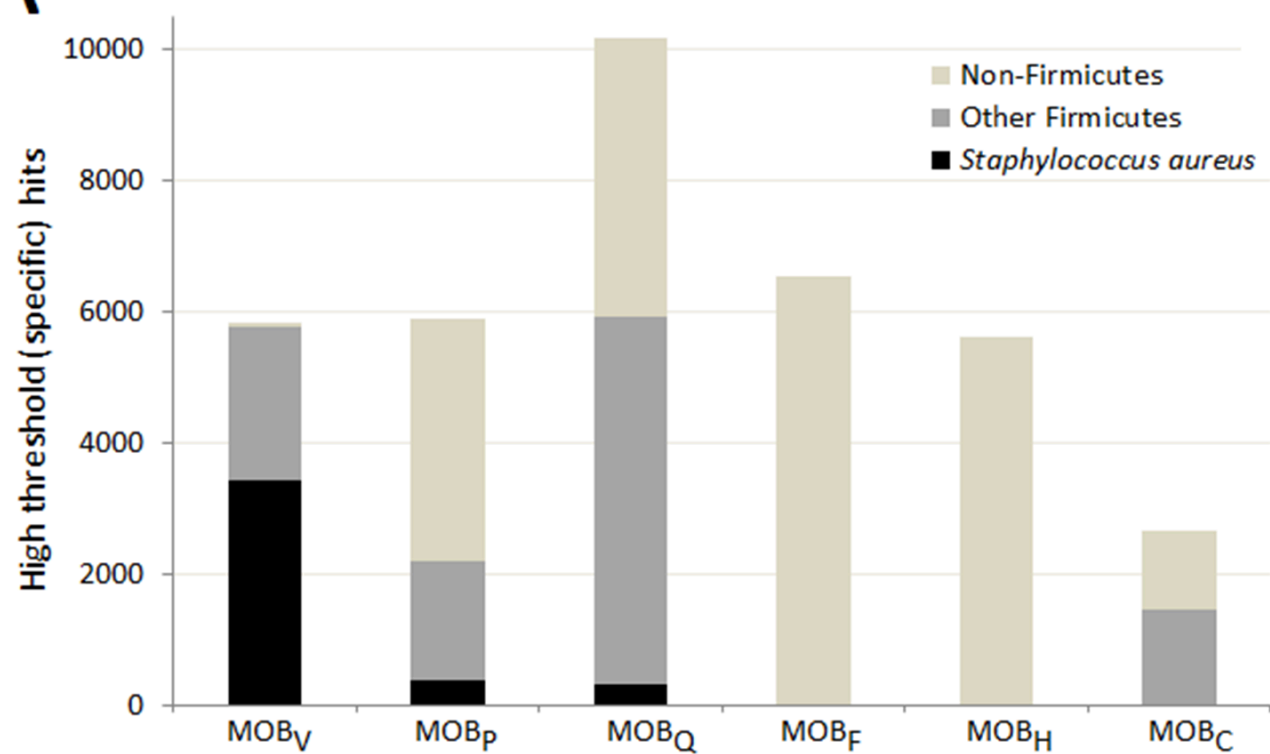
**Fig.S4. MobM-DNA superposition with other relaxases.** (A) MobM-DNA superposition with (A) MOB<sub>Q</sub> relaxases and (B) MOB<sub>F</sub> relaxases. Superposition of the β-turn RxD/N motif and other DNA hairpin-interacting elements of MobM DNA26 structure and (C) MOB<sub>Q</sub> relaxases and MobM Nic0+SPO structure (D) MOB<sub>F</sub> relaxases.

**Fig. S5. Details of theoretical calculations of MobM proteins nucleolytic reaction.** (A) Model system used in the QM/MM geometry optimizations (91 atoms considered in the QM partition plus the hydrogen link atoms). (B) Optimized structure of the transition state (TS) and products (P) of the nucleolytic reaction. (C) Optimized structure of the reactants considering H22 as initially coordinating the divalent ion. In (C) and (D) key interatomic distances (in Å) are labeled in red and numbers below the residue names indicate the coordination distance respect to the Mn<sup>2+</sup> ion. (D) MobM\_wt and MobM\_R25A nucleolytic reaction QM/MM potential energy profiles (PES). The referred Arg and Ala residues were included in the MM partition.

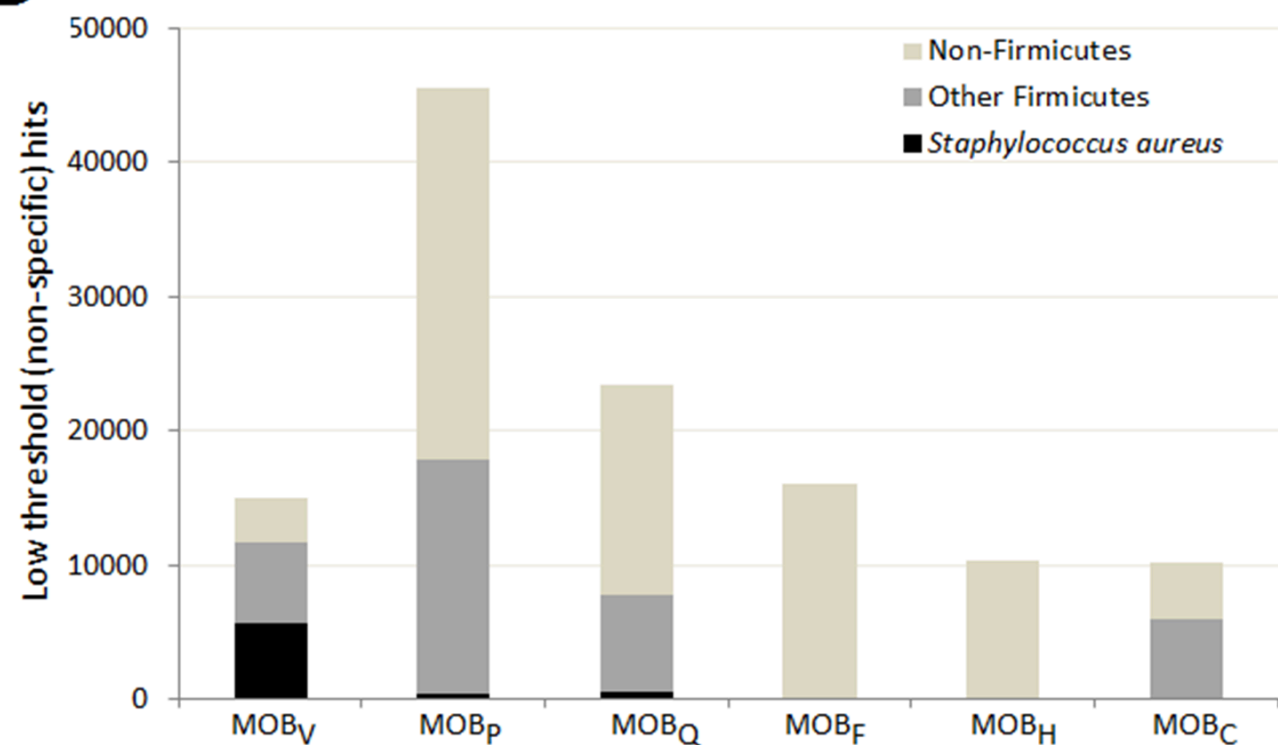
## SI Appendix, Supplementary Figures

# Fig.S1

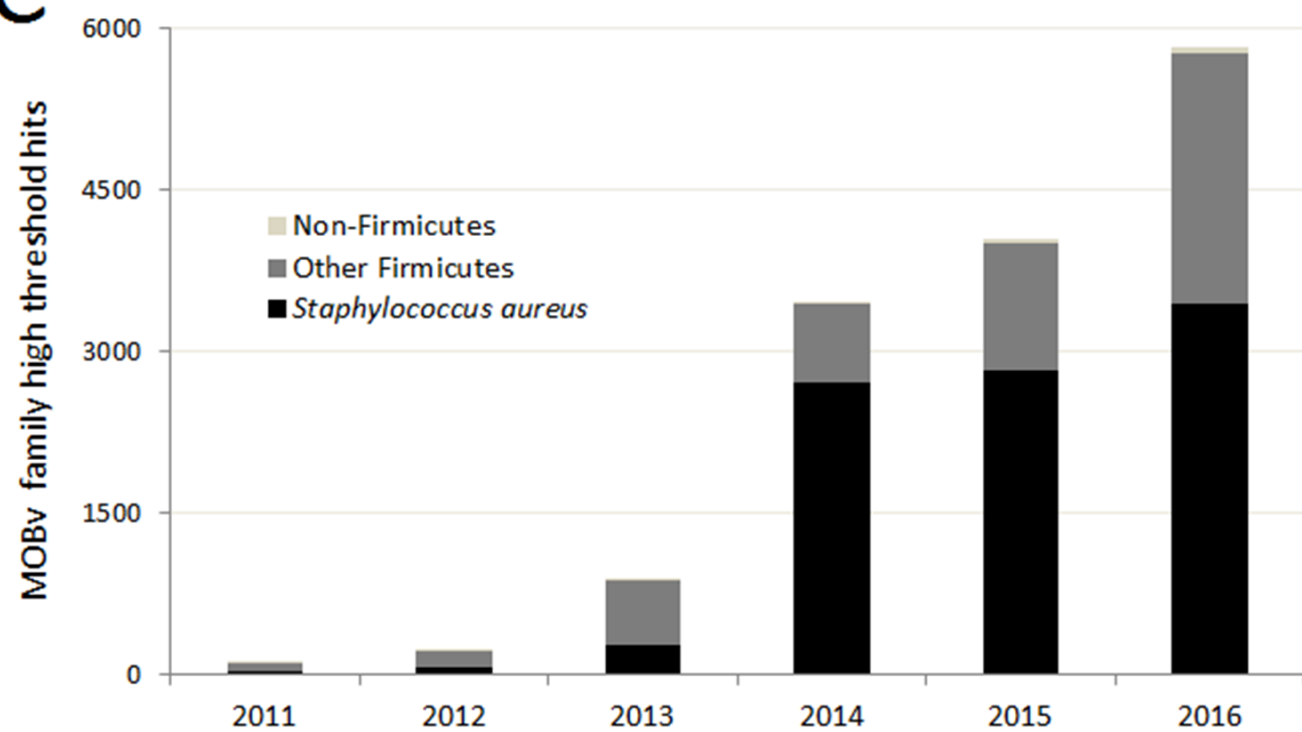
## A



## B



## C



## D

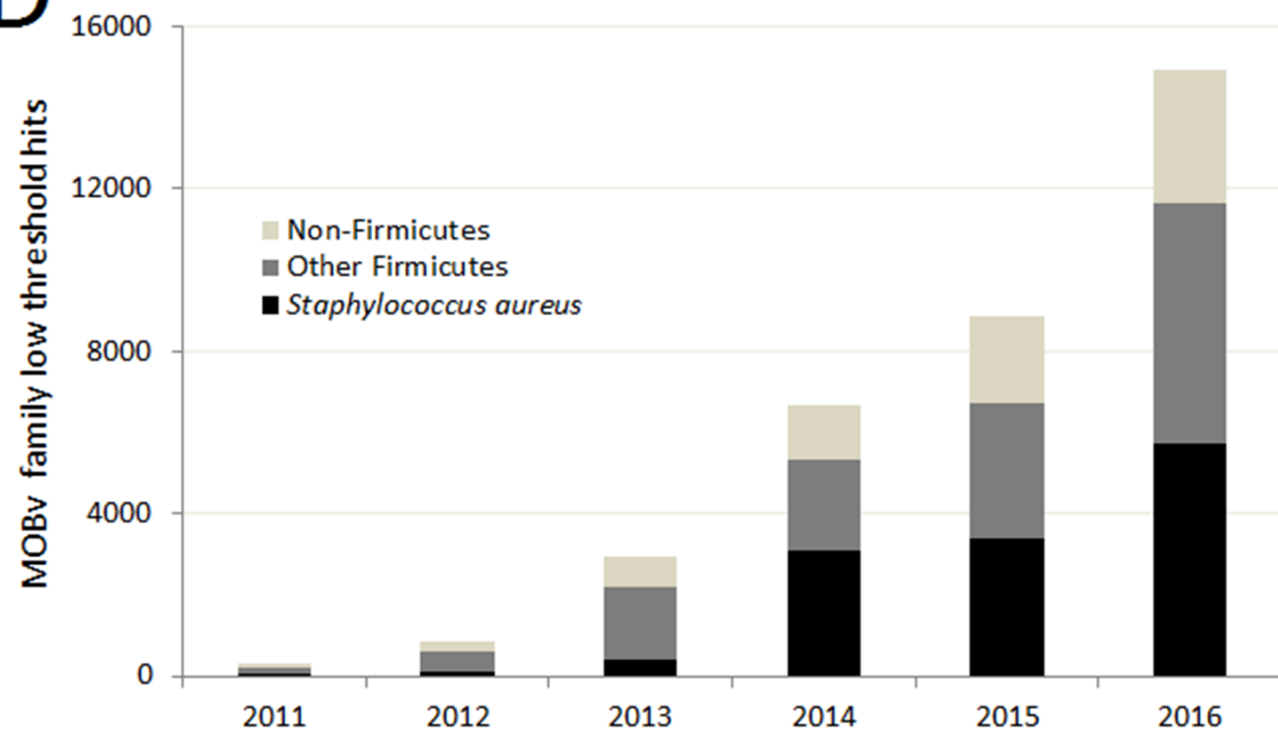


Fig.S2

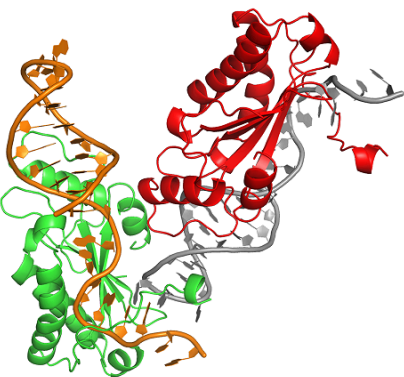
A



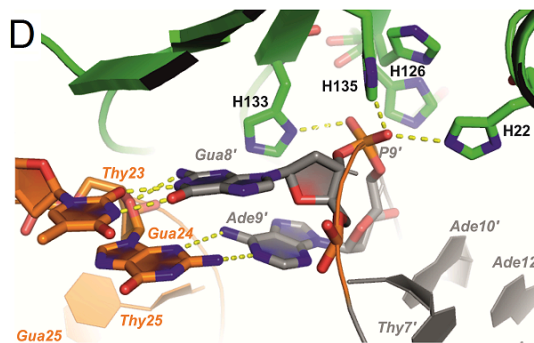
B



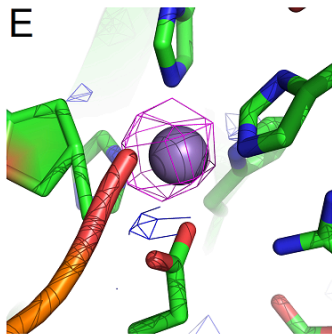
C



D



E



F

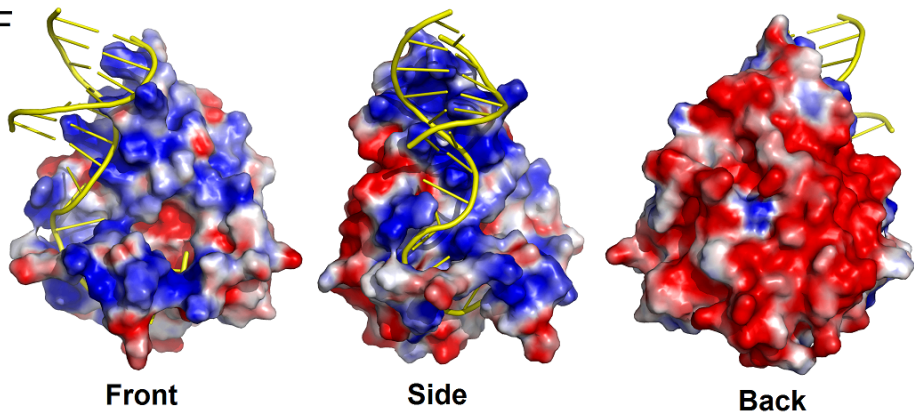
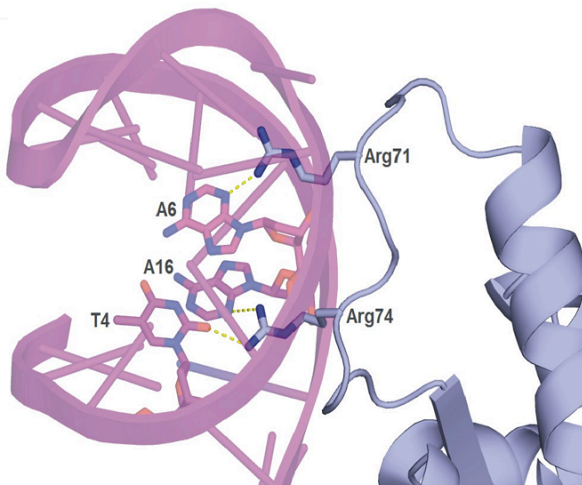
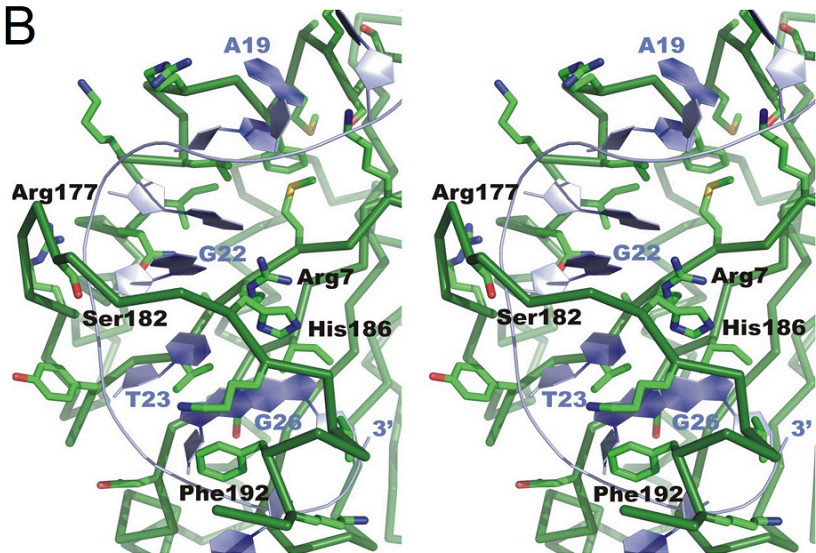


Fig. S3

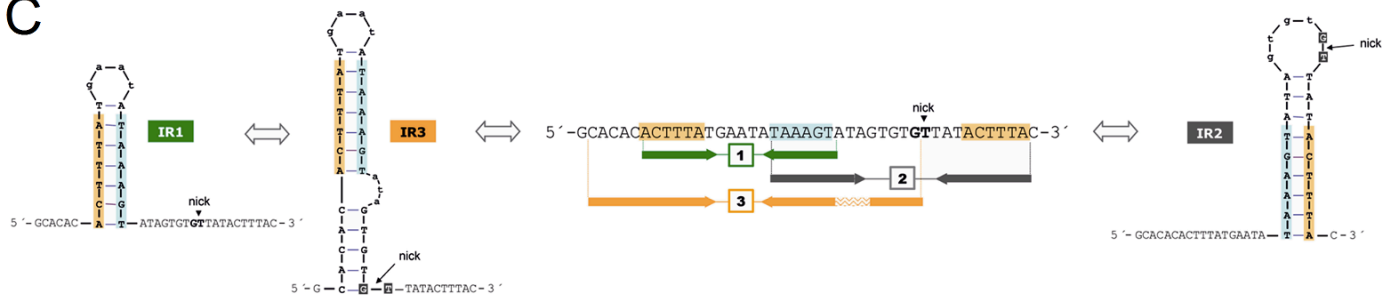
A



B



C



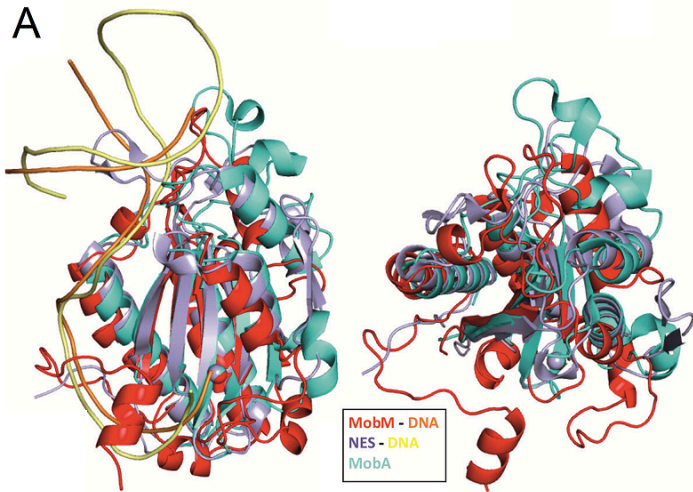
D

Motif I			II		RxD/N	Motif III																
H	R	K	H	R	N	R	Y	G	R	H	D	E	P	H	H	H	P	Q	R	G	H	F
Consensus sequence - most frequently occurring residue at each position																						
A	R	K	H	R	N	R	F	G	A	H	D	E	P	H	H	H	P	Q	R	G	H	F
7	11	23	26	34	74	91	116	123	130	120	129	132	133	134	135	140	163	179	180	180	194	
NA	7	10	22	25	43	74	91	116	123	126	128	129	132	133	134	135	140	161	177	178	186	192
Query Sequence: g 130889 s IP139 25.1 PRE_STRAG RecName: Full=Plasmid recombination enzyme; AltName: Full=Mobilizati...																						

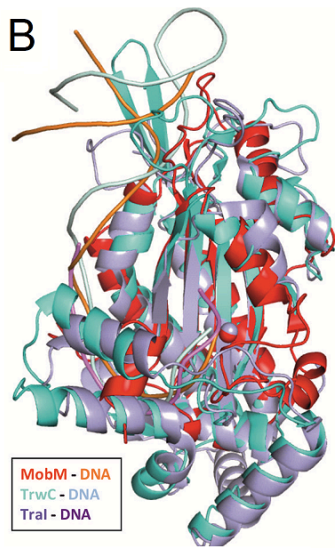


Fig. S4

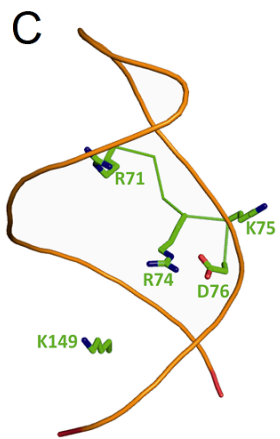
A



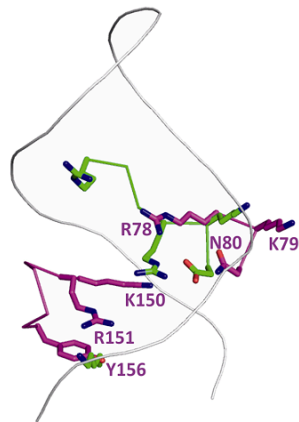
B



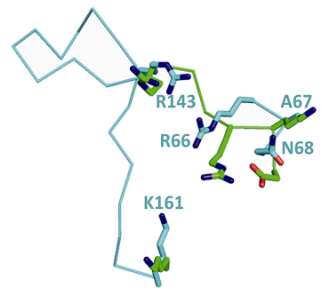
C



MobM-dsDNA

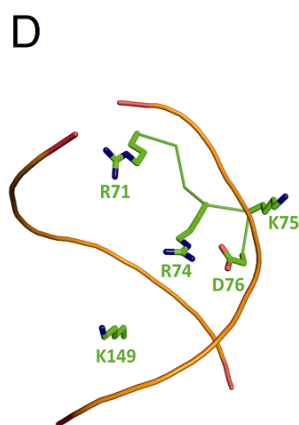


NES-dsDNA

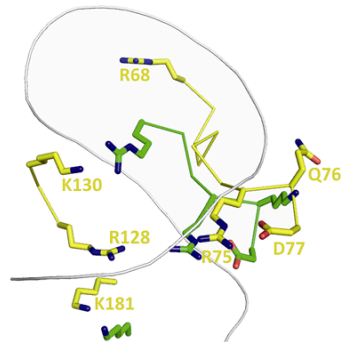


MobA

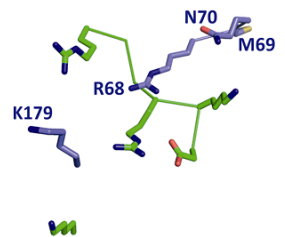
D



MobM-dsDNA



TrwC-dsDNA



TraI\_pF



