# *Plasmidfinder* and pMLST: *in silico* detection and typing of plasmids

P 2420

**Alessandra Carattoli[1], Ea Zankari[2], Aurora García-Fernández, Mette Volby Larsen[3], Ole Lund[3], Laura Villa[1], Frank Møller Aarestrup[2], and Henrik Hasman[2]**

[1]Department of Infectious, Parasitic and Immuno-Mediated Diseases, Instituto Superiore di Sanità, Rome, Italy.
[2]Danish Technical University, National Food Institute, Division for Epidemiology and Microbial Genomics, Lyngby, Denmark.
[3]Danish Technical University, Center for Biological Sequence Analysis, Dept. of Systems Biology, Lyngby, Denmark.

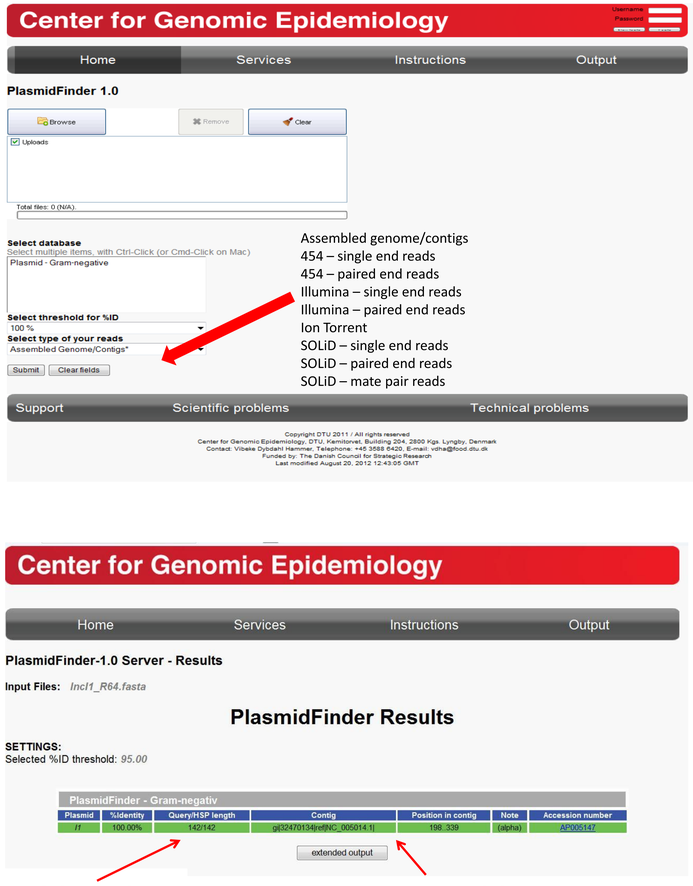
## Background

Bacteria carry plasmids, double stranded, circular or linear DNA molecule capable of autonomous replication. Many plasmids carry specific regions, called replicons, encoding functions that are able to activate and control replication .
Since 2005, a PCR-Based Replicon Typing (PBRT) scheme has been available, targeting in multiplex PCRs the replicons of the major plasmid families occurring in Enterobacteriaceae (Carattoli et al., JMM, 2005). This method was initially built on to detect the replicons of plasmids belonging to the 18 major incompatibility (Inc) groups. More recently, novel replicons and plasmid types were identified by whole-genome and plasmid high-throughput sequencing, extending the PBRT to the identification of 25 different replicons (Villa et al. 2010, Villa et al. 2012, Garcia-Fernandez et al. 2009, García-Fernández et al. 2010). Together with other specific characteristics of the bacterial strain (*i.e.* resistance gene content, Sequence Type by Multi-Locus Sequence-Typing, phylogroup, serotype etc.), the replicon content is currently used as a marker for comparative analysis of unrelated and related strains during epidemiological investigations.
Not all plasmid families occur at the same frequency in clinically relevant enterobacterial strains, but some families are prevalent. For highly frequent plasmids, sequence-based typing schemes were devised to identify related plasmid scaffolds. IncF, IncI1, IncN, IncHI2 and IncHI1 plasmids are currently subtyped by plasmid Multi Locus Sequence Typing (pMLST; http://pubmlst.org/plasmid/). With the recent rapid increase in whole-genome and plasmid sequence data generated by the high-throughput sequencing platforms there is the need to identify resistance genes and plasmids using raw sequence data or contigs generated by high-throughput sequencing of entire genomes. Thus, the current challenge is to extract the relevant information from the large amount of generated data. A web-based method, ResFinder that uses BLAST for identification of acquired antimicrobial resistance genes in whole-genome data has been recently developed (Zankari et al. 2012).

Here we describe the design of easy-to-use web-tool useful for the rapid identification of plasmids in bacteria of interest for clinical microbiology investigations. *Plasmidfinder* is thought for the identification of plasmids using directly high-throughput raw reads, assembled contigs or assembled Sanger sequences; the same raw data can be also used for the *in-silico* application of the pMLST.

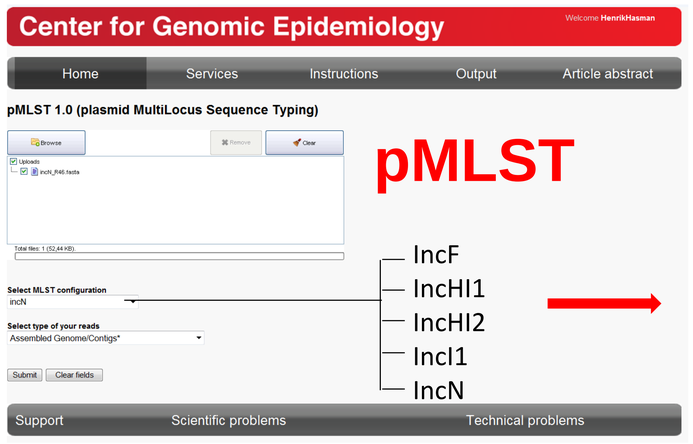
## Methods

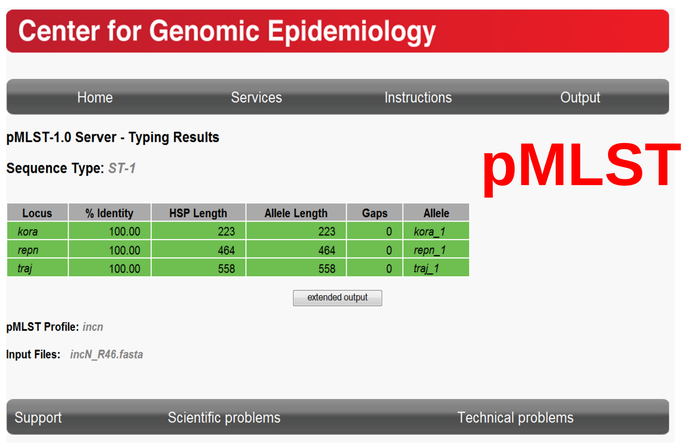
A total of 758 sequences were collected at the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov/nuccore/), corresponding to no-redundant, complete sequences of plasmids identified in bacterial species belonging to the genera Enterobacteriaceae. Among them, 175 plasmids were identified in bacterial species that were endosymbiont of insects that were excluded by the current version of the Plasmidfinder database .
Among the 584 plasmid sequences of interest 243 were small plasmids showing <10 kb in size and 341 were large plasmids. The 584 complete plasmid sequences were firstly aligned against 38 DNA sequences of previously characterized replicons: 25 replicons currently identified by PBRT, and 13 replicons previously described in IncHI1, IncHI2, IncX, IncN2, and IncF plasmids of *Escherichia coli*, *Salmonella* spp., *Yersinia* spp., *Citrobacter freundii* and *Klebsiella pneumoniae* (Table 1). Using these probes a total of 282 plasmids from GenBank (281 large and 1 small) were successfully recognized, showing >95% nucleotide identity, >96% coverage with the reference replicon sequences.
Of the 302 plasmids that were not recognized by the existing 38 sequence-probes, 60 were large plasmids (>10 Kb in size) and 242 were small plasmids, with size ranging from 1,099 to 9,957 bp.
In all the 60 large plasmids a *repA* gene was identified and 32 probes were devised to recognize these replicons at >95% nucleotide identity (Table 2).
The 242 fully sequenced small plasmids were classified in 30 families of homology using the >85% nucleotide identity, This classification was based on multiple phylogenetic analysis of the *repA*, RNAI, oriT and *rop* loci (Table 3).
In conclusion, a total of 75 novel probes, 45 and 30 recognizing large and small plasmids, respectively (Tables 2 and 3) were devised in this study and included in the Plasmidfinder database with the 38 sequences of previously studied replicons, obtaining a database of 113 specific plasmid probes.








## Plasmid database

Table 1

| Probe name_Acc.no | Position | Locus targeted | Reference |
|---|---|---|---|
| A/C2_JN157804 | 59363-58947 | *repA* | Carattoli et al. 2005 |
| B/O/K/Z_GU256641 | 6440-6281 | RNAI | Carattoli et al. 2005 |
| FIA_AP001918 | 48305-48692 | *repFIA* | Carattoli et al. 2005 |
| FIAHI1_AF250878 | 157744-157357 | *repFIA* | Sherburne et al. 2000 |
| FIB_AP001918 | 37699-37018 | *repFIB* | Carattoli et al. 2005 |
| FIBKPN3_JN233704 | 76861-77420 | *repFIB* | García-Fernández et al. 2012 |
| FIBKPQIL_JN233705 | 85120-85959 | *repFIB* | García-Fernández et al. 2012 |
| FIBM_JN420336 | 29199-29637 | *repFIB* | Villa et al. 2012 |
| FIBpHCM2_AL513384 | 104938-105812 | *repFIB* | Parkhill et al. 2001 |
| FIBs_FN432031 | 14696-14054 | *repFIB* | Villa et al 2010 |
| FIC_AP001918 | 3421-3660 | *repFIC* | Carattoli et al. 2005 |
| FII_AY458016 | 89119-88859 | RNAI-FII | Carattoli et al. 2005 |
| FIIK_CP000648 | 5614-5761 | RNAI-FII | Villa et al 2010 |
| FIIS_CP000858 | 26564-26303 | *repFII* | Villa et al. 2010 |
| FIIYp_CP000670 | 114633-114862 | *repFII* | Villa et al. 2010 |
| FIIYps_CP001049 | 20297-20071 | RNAI-FII | Villa et al. 2010 |
| HI1A_AF250878 | 40122-39702 | *repHIA* | Sherburne et al. 2000 |
| HI1ACIT_JX182975 | 220487-220906 | *repHIA* | Dolejska et al. 2012 |
| HI1B_AF250878 | 54421-54960 | *repHIB* | Dolejska et al. 2012 |
| HI2_EX664015 | 209836-209510 | *repHI2* | Carattoli et al. 2005 |
| HI2A_EX664015 | 211-840 | *repHIA* | Gilmour et al. 2004 |
| HIB_3_JN420336 | 126037-126606 | *repHIB* | Villa et al. 2012 |
| HIBCIT_JX182975 | 203748-203211 | *repHIB* | Dolejska et al. 2012 |
| I1(alpha)_AP005147 | 198-339 | RNAI | Carattoli et al. 2005 |
| L/M_AF550415 | 54207-54947 | *repA* | Carattoli et al. 2005 |
| N_AY046276 | 31781-32294 | *repA* | Carattoli et al. 2005 |
| N2_JF785549 | 14890-15366 | *repA* | Poirel et al. 2011 |
| P_1_L27758 | 12765-12232 | Control of *repA* | Carattoli et al. 2005 |
| R_DQ449578 | 19367-19617 | *repA* | García-Fernández et al. 2009 |
| T_AP004237 | 423-1172 | *repA* | Carattoli et al. 2005 |
| U_DQ401103 | 51-615 | *repA* | Carattoli et al. 2005 |
| W_EF633507 | 25644-25886 | *repA* | Carattoli et al. 2005 |
| X1_EU370913 | 35767-36140 | Fi *repA* | Hansen et al. 2004 |
| X2_JQ269335 | 9035-8662 | Fi *repA* | Carattoli et al. 2005 |
| X3_JN247852 | 6440-6067 | Fi *repA* | Johnson et al. 2012 |
| X3pEC14_JN935899 | 7-380 | Fi *repA* | Johnson et al. 2012 |
| X4_JN194214 | 967-1340 | Fi *repA* | Johnson et al. 2012 |
| Y_K02380 | 1075-1839 | *repA* | Carattoli et al. 2005 |

Table 2

| Probe name_Acc.no | Position | Locus targeted | Reference |
|---|---|---|---|
| A/C1_FJ705807 | 684-1100 | *repA* | This study |
| FIBpB171_AB034946 | 250-892 | *repFIB* | This study |
| FIBpCTU1_FN543094 | 2655-3347 | *repFIB* | This study |
| FIBpCTU3_FN543096 | 142054-142620 | *repFIB* | This study |
| FIBpECLA_CP001919 | 211-771 | *repFIB* | This study |
| FIBpENTAS01_CP003027 | 141679-142239 | *repFIB* | This study |
| FIBpENTE01_CP000654 | 142054-142620 | *repFIB* | This study |
| FIBpKPHS1_CP003223 | 201-761 | *repFIB* | This study |
| FIBpLF82_CU638872 | 211-771 | *repFIB* | This study |
| FIBpO111_AP010962 | 54482-53598 | *repFIB* | This study |
| FIBp14_JQ418538 | 34250-33989 | *repFII* | This study |
| FIBp2B1J_HG706666 | 545-286 | RNAI-FII | This study |
| FIBp96A_JQ418521 | 42799-43332 | *repFII* | This study |
| FIBpCoo_CR942285 | 27847-27586 | RNAI-FII | This study |
| FIBpCRY_NC_005814 | 739-1331 | *repFII* | This study |
| FIBpCTU2_FN543095 | 8327-8903 | *repFII* | This study |
| FIBpECLA_CP001919 | 3-749 | *repFII* | This study |
| FIBpENTA_CP003027 | 84306-84866 | RNAI-FII | This study |
| FIBpMET1_EU383016 | 690-1266 | *repFII* | This study |
| FIBpRSB107_AJ851089 | 21949-21689 | RNAI-FII | This study |
| FIBpSE11_AP009242 | 75348-75085 | RNAI-FII | This study |
| FIIpseudo_NC_011759 | 76101-76490 | *repFII* | This study |
| FIIpYYal2790_AY150843 | 121-794 | *repFII* | This study |
| FIISARC14_JQ418540 | 36938-37382 | *repFII* | This study |
| FIISerratia_NC_009829 | 40691-40968 | *repFII* | This study |
| I2(delta)_AP005277 | 207-522 | *repR* | This study |
| L/MpMU407_U27345 | 22-760 | *repC* | This study |
| P_6_JF785550 | 1477-2282 | *repA* | This study |
| P_beta_U67194 | 15532-16163 | *repA* | This study |
| X4_2_FN543504 | 6671-7382 | Pi *repA* | This study |
| X5_CP002179 | 34413-34786 | Pi *repA* | This study |
| X6_AM942760 | 1-374 | Pi *repA* | This study |
| pADAP_AF135182 | 113796-114335 | *repA* | This study |
| pENTAS02_CP003028 | 25220-24242 | *repA* | This study |
| pESA2_CP000784 | 2497-3246 | *repA* | This study |
| pIP31758_CP000718 | 639-1556 | *repA* | This study |
| pIP31758_CP000719 | 150807-151715 | *repA* | This study |
| pIP32953_BX936400 | 262-1188 | *repA* | This study |
| pSL483_CP001137 | 23947-22953 | *repA* | This study |
| pXuzhou21_CP001927 | 71-791 | *repA* | This study |
| pYE854_AM905950 | 77347-78325 | *repA* | This study |
| repH7_NC_011351 | 93-798 | *repA* | This study |
| repJava_NC_015068 | 179-712 | *repA* | This study |
| repSerratia_NC_015972 | 17401-18023 | *repA* | This study |

Table 3

| Probe name_Acc. no. | Positions | Locus targeted | References |
|---|---|---|---|
| ColM_X54106S | 2368-2617 | ORF4 | This study |
| ColA_X01654 | 1842-2141 | Intergenic region | This study |
| ColE10_X01654 | 5889-6223 | OriT | This study |
| ColpJ_AF268389 | 488-1077 | *repA* | This study |
| ColGM8S_DQ916145 | 5703-6033 | *repA* | This study |
| ColM6282_DQ995352 | 3222-3568 | *repA* | This study |
| ColBS512_NC_010656 | 1081-1560 | *repA* | This study |
| ColGJC156_NC_009781 | 1249-1628 | *repA* | This study |
| ColFST258_JN247853 | 1638-1898 | Intergenic region | This study |
| ColMG831_NC_011406 | 700-1233 | ORF2 | This study |
| ColIP843_NC_005015 | 6717-6908 | RNAI | This study |
| ColHRGK_AY543071 | 2041-2348 | Intergenic region | This study |
| ColKP3_JN205800 | 5873-6317 | *repA* | This study |
| ColKPH56_NC_016841 | 526-862 | *repA* | This study |
| ColMG828_NC_008486 | 882-1297 | *repA* | This study |
| ColMGD2_NC_003789 | 726-1335 | *repA* | This study |
| ColMP18_NC_013652 | 175-708 | *repA* | This study |
| ColPCP1_AE017046 | 3470-3203 | *rop* | This study |
| ColPnuII_AF305615 | 2886-3092 | Intergenic region | This study |
| ColpWES_DQ268764 | 10151-10545 | *repA* | This study |
| ColRNAI_CP002187 | 1273-1489 | RNAI | This study |
| ColSD853_NC_015392 | 1363-1896 | *repA* | This study |
| ColSh_NC_009344 | 1008-1697 | *repA* | This study |
| ColSL476_NC_011082 | 182-343 | *rop* | This study |
| ColVCM04_HM231165 | 776-1290 | *repA* | This study |
| ColYC_NC_002144 | 5043-5481 | *repA* | This study |
| ColYe4449_FJ696405 | 3664-4221 | *repA* | This study |
| ColYF27601_JF937655 | 1-220 | RNAI | This study |
| Q1_FJ696404 | 4547-4996 | *repA* | This study |
| Q2_HE654726 | 2181-2630 | *repA* | This study |