

Research Article

Distribution of Genes Encoding Nucleoid-Associated Protein Homologs in Plasmids

Toshiharu Takeda,¹ Choong-Soo Yun,^{1,2} Masaki Shintani,³
Hisakazu Yamane,¹ and Hideaki Nojiri^{1,2}

¹Biotechnology Research Center, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

²Agricultural Bioinformatics Research Unit, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

³Japan Collection of Microorganisms, RIKEN Bioresource Center, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

Correspondence should be addressed to Hideaki Nojiri, anojiri@mail.ecc.u-tokyo.ac.jp

Received 14 October 2010; Accepted 27 November 2010

Academic Editor: Hiromi Nishida

Copyright © 2011 Toshiharu Takeda et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Bacterial nucleoid-associated proteins (NAPs) form nucleoprotein complexes and influence the expression of genes. Recent studies have shown that some plasmids carry genes encoding NAP homologs, which play important roles in transcriptional regulation networks between plasmids and host chromosomes. In this study, we determined the distributions of the well-known NAPs Fis, H-NS, HU, IHF, and Lrp and the newly found NAPs MvaT and NdpA among the whole-sequenced 1382 plasmids found in Gram-negative bacteria. Comparisons between NAP distributions and plasmid features (size, G+C content, and putative transferability) were also performed. We found that larger plasmids frequently have NAP gene homologs. Plasmids with H-NS gene homologs had less G+C content. It should be noted that plasmids with the NAP gene homolog also carried the relaxase gene involved in the conjugative transfer of plasmids more frequently than did those without the NAP gene homolog, implying that plasmid-encoded NAP homologs positively contribute to transmissible plasmids.

1. Introduction

Bacterial chromosomal DNA is folded to form a compacted structure, the nucleoid. The proteins involved in folding the chromosome are known as nucleoid-associated proteins (NAPs) [1, 2]. Because of their DNA-binding ability, NAPs can also play an important role in global gene regulation [1, 2]. Each well-known NAP in *Enterobacteriaceae* may be categorized as a “factor for inversion stimulation” (Fis), “histone-like nucleoid structuring protein” (H-NS), “histone-like protein from *Escherichia coli* strain U93” (HU), “integration host factor” (IHF), or “leucine-responsive regulatory protein” (Lrp) [1]. Fis is one of the most abundant NAPs in exponentially growing *E. coli* cells, and its role as a transcriptional regulator has been investigated [3]. H-NS binds DNA, especially A+T-rich regions including promoter regions or horizontally acquired DNA and acts as a global transcriptional repressor [4]. HU and IHF are similar in amino acid sequence level, and both are global regulators

[5, 6], although they have distinct DNA-binding activities: HU binds to DNA nonspecifically whereas IHF binds to a consensus sequence [7]. Lrp has a global influence on transcription regulation and is also involved in microbial virulence [8]. In addition to these well-known NAPs, many other NAPs are found not only in *Enterobacteriaceae* but also in other organisms. For instance, NdpA, a functionally unknown NAP, has been found in Gram-negative bacteria [9]. The MvaT family protein is the functional homolog of H-NS in *Pseudomonas* bacteria [10].

Horizontal gene transfer (HGT), which is mediated by transduction, transformation, and conjugation, plays an important role in the evolution of prokaryotic genomes [11, 12]. Genes acquired by HGT can provide beneficial functions such as resistance to antibiotics and advantages to their host under selective pressures [13]. However, the mechanisms underlying the integration of newly acquired genes into host regulatory networks are still unclear. Recent investigations have shown that some plasmids carry the genes

encoding NAP homologs, which play important roles in transcriptional regulation networks between plasmids and host chromosomes and in maintaining host cell fitness. For example, Doyle et al. [14] reported that plasmid-encoded H-NS-like protein has a “stealth” function that allows for plasmid transfer into host cells without disrupting host regulatory networks, maintaining host cell fitness. Yun and Suzuki et al. [15] reported that plasmid-encoded H-NS-like protein can also play a key role in optimizing gene transcription both on the plasmid and in the host chromosome.

In this study, we determined the distributions of NAP homologs among plasmids and discussed their roles in the maintenance of plasmid and host cell fitness.

2. Materials and Methods

2.1. Plasmid Database Collection and Local BLAST Analyses. The completely sequenced plasmid database was downloaded from the NCBI ftp site (<ftp://ftp.ncbi.nih.gov/genomes/Plasmids/>). Some duplicated sequence data of the same plasmids were removed from the database. Identification of plasmids that contain the genes encoding NAP homologs was performed using the local TBLASTN program (ver. 2.2.24, <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/>) under strict conditions (i.e., thresholds of 30% identity and 70% query coverage). The complete amino acid sequences of Fis (DDBJ/EMBL/GenBank accession no. AP_003801), H-NS (AP_001863), Hha (AP_001109), HU α (AP_003818), HU β (AP_001090), IHF α (AP_002332), IHF β (AP_001542), Lrp (AP_001519), and NdpA (P33920) from *E. coli* K-12 W3110 and MvaT (AAP33788) from *Pseudomonas aeruginosa* PAO1 were used as query sequences.

2.2. Bacterial Genome Analyses. The complete genome sequences of bacteria were downloaded from the NCBI ftp site (<ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>). The number of NAP genes on proteobacterial genomes was investigated using the TBLASTN program (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi) under strict conditions (i.e., thresholds of 30% identity and 70% query coverage).

2.3. Plasmid Classification. Plasmids in the database were classified into six groups according to their source organisms: Gram-negative, Gram-positive, archaeal, eukaryotic, viral, and unclassified. Putative transferability of each Gram-negative plasmid was determined by whether it carried the relaxase gene of each MOB family that Garcillán-Barcia et al. proposed [16]. Instead of using the local PSI-BLAST program (ver. 2.2.24, <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/>) as described by Garcillán-Barcia et al. [16], we used the local TBLASTN program.

3. Results and Discussion

3.1. Database Collection and Plasmid Classification by Origin. We downloaded the whole sequences of 2278 plasmids

from the NCBI ftp site (April 2010). Duplicated plasmids were removed manually, and the resultant 2260 plasmid sequences were used in this study. To understand what types of plasmids were included in the database, we classified them into six groups according to their source organisms. The database included 1382 Gram-negative, 725 Gram-positive, 81 archaeal, 43 eukaryotic, 1 viral, and 28 unclassified plasmids.

3.2. Identification of the Plasmids Containing NAP Gene Homologs. Using the amino acid sequences of well-known NAPs (Fis, H-NS, HU, IHF, and Lrp) and newly found NAPs (MvaT and NdpA), their distributions were surveyed for plasmids using the TBLASTN program. Some plasmids had ORFs showing sequence similarities to both HU and IHF. We adopted the one with the higher E value. Of 2260 plasmids, 155 (7%) contained the gene encoding NAP homolog. Of those, 116 (75%) contained only one NAP gene homolog and 39 (25%) contained more than one NAP gene homolog. No plasmids carried the Fis gene homolog. Twenty-two plasmids carried the H-NS gene homolog, and all of them had a Gram-negative origin (Table 1). Sixty-six plasmids had the HU gene homolog; of these, 51 had a Gram-negative origin and 15 had a Gram-positive origin (Table 2). Twenty-seven plasmids (25 with Gram-negative and 2 with Gram-positive origins) carried the IHF gene homolog (Table 3). Forty-eight plasmids (46 with Gram-negative, 1 with a Gram-positive, and 1 with an archaeal origin) carried the Lrp gene homolog (Table 4). Of these, 23 (48%) contained more than one Lrp gene homolog. On the other hand, MvaT and NdpA homologs were encoded on only 3 plasmids, and all of them were of Gram-negative origin (Table 5). Previously reported plasmids that are known to have NAP gene homologs were included in those 155 plasmids. These included R27 (NC_002305) and pHCM1 (NC_003384) [18, 19] with the H-NS gene homolog; pQBR103 (NC_009444) [20] with the HU and NdpA gene homologs; and pCAR1 (NC_004444) [21, 22] with the MvaT, HU, and NdpA gene homologs. These results indicated the adequacy of our search. Because we used NAPs from Gram-negative bacteria as query sequences, it may be reasonable that 136 (88%) of 155 plasmids with the NAP gene homolog belonged to the group isolated from Gram-negative bacteria. Therefore, in further studies we discussed the Gram-negative plasmid group.

3.3. Relationships between Plasmid Size and NAP Gene Homolog Distributions. We first compared the sizes of 136 plasmids with NAP gene homologs with those of all 1382 Gram-negative group plasmids. All 1382 plasmids could be divided into 4 groups according to size, small (<10 kb), intermediate (10 to 100 kb), large (100 kb to 1 Mb), and mega (>1 Mb) plasmids. The distribution of the 136 plasmids, each of which had one or more genes encoding NAP homologs, is shown in Figure 1(a): none of 415 small plasmids, 34 (5%) of 686 intermediate plasmids, 90 (33%) of 269 large plasmids, and 12 (100%) of 12 mega plasmids carried at least one NAP gene homolog. The average size of the 136 plasmids was larger (364 kb) than that of all 1382 plasmids

TABLE 1: Plasmids containing the gene encoding H-NS homolog^a.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	Classification ^d	MOB family ^e
1	NC_013972	<i>Erwinia amylovora</i> ATCC 49946	28243	50	66	99	3129	2728	–	–
pAsa5	NC_009350	<i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449	155098	54	46	99	941	534	–	MOB _F
pAsal5	NC_009352	<i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i>	18536	54	46	99	16890	16483	–	–
pEA29	NC_013957	<i>Erwinia amylovora</i> CFBP1430	28259	50	66	99	3129	2728	–	–
pEA29	NC_005706	<i>Erwinia amylovora</i>	28185	50	64	99	2991	2590	–	–
pEC-IMP	NC_012555	<i>Enterobacter cloacae</i>	318782	48	64	99	109370	108969	–	MOB _H
pEC-IMPQ	NC_012556	<i>Enterobacter cloacae</i>	324503	48	64	99	109370	108969	–	MOB _H
pEJ30	NC_004834	<i>Erwinia</i> sp. Ejp 556	29593	50	66	99	4651	4250	–	–
pEP36	NC_013263	<i>Erwinia pyrifoliae</i> Ep1/96	35909	50	66	99	25040	25441	–	–
pEP36	NC_004445	<i>Erwinia pyrifoliae</i> Ep1/96	35904	50	64	98	4675	4280	–	–
pET45	NC_010699	<i>Erwinia tasmaniensis</i> Et1/99	44694	51	52	93	37435	37809	–	MOB _F
pET49	NC_010697	<i>Erwinia tasmaniensis</i> Et1/99	48751	44	36	94	30821	31204	–	–
pHCM1	NC_003384	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi str. CT18	218160	48	61	99	131861	131460	–	MOB _H
pK2044	NC_006625	<i>Klebsiella pneumoniae</i> NTUH-K2044	224152	50	67	99	35717	36112	–	–
plasmid_153kb	NC_009705	<i>Yersinia pseudotuberculosis</i> IP 31758	153140	40	44	100	139846	140265	–	–
pLVPK	NC_005249	<i>Klebsiella pneumoniae</i> <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Choleraesuis	219385	50	67	99	114397	114792	–	–
pMAK1	NC_009981	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Choleraesuis	208409	47	61	99	60046	59645	–	MOB _H

TABLE 1: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	Classification ^d	MOB family ^e
pO1111_L1	NC_013365	<i>Escherichia coli</i> O111:H-str. 11128	204604	47	61	99	80175	79774	–	MOB _H
pSG1	NC_007713	<i>Sodalis glossinidius</i> str. “morstians”	83306	49	43	97	2533	2922	–	
R27	NC_002305	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi	180461	46	61	99	148225	148626	–	MOB _H
R478	NC_005211	<i>Serratia marcescens</i>	274762	46	64	99	111747	111346	–	MOB _H
Unnamed	NC_011148	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Agona str. SL483	37978	41	43	95	7671	7288	–	

^a This list is the result of a TBLASTN analysis using the amino acid sequence of H-NS as a query under strict conditions (i.e., thresholds of 30% identity and 70% query coverage). Besides these plasmids, pSf-R27 from *Shigella flexneri* 2a str. 24571 was completely sequenced by Wei et al. [17] and encodes the H-NS-like protein Sfh.

^b Average G+C content of the plasmid.

^c Reported TBLASTN identity to H-NS.

^d Plasmid classification according to its source organism (–, Gram-negative plasmid).

^e Plasmid classification according to its relaxase gene sequence as described by Garcillán-Barcia et al. [16].

TABLE 2: Plasmids containing the gene encoding HU homolog^a.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	Classification ^d	MOB family ^e
1	NC_006823	<i>Aromatoleum aromaticum</i> EbN1	207355	58	55	99	186175	185909	-	
1	NC_007949	<i>Polaromonas</i> sp. JS666	360405	57	52	99	61052	60786	-	MOB _H
1	NC_008010	<i>Deinococcus geothermalis</i> DSM 11300	574127	66	38	97	550805	550545	+	
1	NC_008503	<i>Lactococcus lactis</i> subsp. <i>cremoris</i> SK11	14041	34	37	94	9732	10007	+	MOB _P
1	NC_008242	<i>Chelativorans</i> sp. BNC1	343931	62	41	94	133932	133678	-	MOB _Q
2	NC_012529	<i>Deinococcus deserti</i> VCD115	314317	64	38	93	269648	269899	+	
3	NC_012528	<i>Deinococcus deserti</i> VCD115	396459	61	40	96	8700	8957	+	
Megaplasmid	NC_007974	<i>Cupriavidus metallidurans</i> CH34	2580084	64	51	99	1393415	1393149	-	MOB _V
Megaplasmid	NC_005863	<i>Desulfovibrio vulgaris</i> str. Hildenborough	202301	66	31	98	5502	5765	-	
Megaplasmid pDF308	NC_013940	<i>Deferribacter desulfuricans</i> SSM1	308544	24	41	100	253817	253548	-	
Megaplasmid pHG1	NC_005241	<i>Ralstonia eutropha</i> H16	452156	62	48	99	343060	342791	-	
p49879.1	NC_006907	<i>Leptospirillum ferrooxidans</i>	28878	58	47	99	3281	3015	-	MOB _Q
p49879.2	NC_006909	<i>Leptospirillum ferrooxidans</i>	28012	55	48	99	15858	15592	-	MOB _Q
pAH187_270	NC_011655	<i>Bacillus cereus</i> AH187	270082	34	59	100	113139	112870	+	
pAH820_272	NC_011777	<i>Bacillus cereus</i> AH820	272145	34	58	100	153060	152791	+	
pAM04528	NC_012693	<i>Salmonella enterica</i>	158213	52	57	99	14067	14333	-	MOB _H
pAOVO01	NC_008765	<i>Acidovorax</i> sp. JS42	72689	62	46	100	65140	64871	-	MOB _F
pAPA01-011	NC_013210	<i>Acetobacter pasteurianus</i> IFO 3283-01	191799	53	47	100	154736	154467	-	
pAR060302	NC_012692	<i>Escherichia coli</i>	166530	53	46	99	38442	38708	-	MOB _H
pAsa4	NC_009349	<i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449	166749	53	57	99	15755	16021	-	MOB _H
pAtS4c	NC_011984	<i>Agrobacterium vitis</i> S4	211620	59	45	94	141245	140991	-	MOB _Q
pAtS4e	NC_011981	<i>Agrobacterium vitis</i> S4	631775	57	41	94	40476	40222	-	MOB _Q
pBc239	NC_011973	<i>Bacillus cereus</i> Q1	239246	33	52	100	191895	192164	+	
pBF9343	NC_006873	<i>Bacteroides fragilis</i> NCTC 9343	36560	32	35	92	15803	15558	-	MOB _P
pBPHY01	NC_010625	<i>Burkholderia phymatum</i> STM815	1904893	62	43	99	826527	826252	-	

TABLE 2: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	Classification ^d	MOB family ^e
pBPHY02	NC_010627	<i>Burkholderia phymatum</i> STM815	595108	59	45	99	98625	98359	-	
pBtoxis	NC_010076	<i>Bacillus thuringiensis</i> serovar israelensis	127923	32	52	99	77382	77648	+	
pBWB401	NC_010180	<i>Bacillus weihenstephanensis</i> KBAB4	417054	34	59	100	338347	338078	+	
pCAR1	NC_004444	<i>Pseudomonas resinovorans</i>	199035	56	42	99	97809	98075	-	MOB _H
pCAUL01	NC_010335	<i>Caulobacter</i> sp. K31	233649	67	44	99	97598	97329	-	MOB _Q
pCER270	NC_010924	<i>Bacillus cereus</i>	270082	34	59	100	169548	169279	+	
pDBORO	NC_009137	<i>Lactococcus lactis</i> subsp. <i>lactis</i> bv. <i>diacetylactis</i>	16404	35	37	94	16387	16112	+	
pDVUL01	NC_008741	<i>Desulfovibrio vulgaris</i> DP4	198504	66	31	98	198317	198054	-	
pEH4H	NC_012690	<i>Escherichia coli</i>	148105	53	57	99	14067	14333	-	MOB _H
pG9842_209	NC_011775	<i>Bacillus cereus</i> G9842	209488	30	60	100	88828	88559	+	
pH308197_258	NC_011339	<i>Bacillus cereus</i> H3081.97	258484	34	59	100	83033	83302	+	
pHD5AT	NC_012752	<i>Candidatus Hamiltonella defensa</i> 5AT (<i>Acyrtosiphon pisum</i>)	59032	45	45	99	14981	15247	-	MOB _P
pIP1202	NC_009141	<i>Yersinia pestis</i> bv. <i>Orientalis</i> str. IP275	182913	53	57	99	14067	14333	-	MOB _H
plasmid 2	NC_007972	<i>Cupriavidus metallidurans</i> CH34	171459	61	46	99	125530	125261	-	
pMOL28	NC_006525	<i>Cupriavidus metallidurans</i> CH34	171461	61	46	99	51529	51798	-	
pMP118	NC_007930	<i>Lactobacillus salivarius</i> UCC118	242436	32	54	99	56763	56497	+	MOB _V
pNPUN02	NC_010632	<i>Nostoc punctiforme</i> PCC 73102	254918	41	44	99	74804	74538	-	MOB _V
pOANT02	NC_009670	<i>Ochrobactrum anthropi</i> ATCC 49188	101491	59	49	94	32700	32446	-	
pP91278	NC_008613	<i>Photobacterium damsela</i> subsp. <i>piscicida</i>	131520	52	57	99	125918	126184	-	MOB _H
pP99-018	NC_008612	<i>Photobacterium damsela</i> subsp. <i>piscicida</i>	150157	51	57	99	133314	133580	-	MOB _H
pPER272	NC_010921	<i>Bacillus cereus</i>	272145	34	58	100	153060	152791	+	
pPMA4326A	NC_005918	<i>Pseudomonas syringae</i> pv. <i>maculicola</i>	46697	55	42	99	1520	1786	-	
pPMA4326B	NC_005919	<i>Pseudomonas syringae</i> pv. <i>maculicola</i>	40110	55	45	99	1457	1723	-	

TABLE 2: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	Classification ^d	MOB family ^e
pQBR103	NC_009444	<i>Pseudomonas fluorescens</i> SBW25	425094	53	51	99	182862	183128	–	–
pRI32503	NC_012853	<i>Rhizobium leguminosarum</i> bv. trifolii WSM1325	516088	59	47	94	300662	300916	–	MOB _Q
pRA1	NC_012885	<i>Aeromonas hydrophila</i>	143963	51	58	99	15573	15839	–	MOB _H
pRALTA	NC_010529	<i>Cupriavidus taiwanensis</i>	557200	60	46	98	153542	153276	–	–
pREB1	NC_009926	<i>Acaryochloris marina</i> MBIC11017	374161	47	46	100	339743	340012	–	MOB _F
pREB2	NC_009927	<i>Acaryochloris marina</i> MBIC11017	356087	45	48	100	57583	57852	–	MOB _F
pREB3	NC_009928	<i>Acaryochloris marina</i> MBIC11017	273121	45	46	100	234682	234951	–	MOB _F
pRL7	NC_008382	<i>Rhizobium leguminosarum</i> bv. viciae 3841	151564	58	48	94	20484	20230	–	MOB _Q
pRLG203	NC_011370	<i>Rhizobium leguminosarum</i> bv. trifolii WSM2304	308747	58	49	94	141121	140867	–	–
pRp12D01	NC_012855	<i>Ralstonia pickettii</i> 12D	389779	58	37	99	321346	321080	–	MOB _H
pSG2	NC_007184	<i>Sodalis glossinidius</i>	27240	45	45	86	10072	9845	–	–
pSG3	NC_007186	<i>Sodalis glossinidius</i>	19201	51	51	100	13812	13543	–	–
pSN254	NC_009140	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport str. SL254	176473	53	57	99	14067	14333	–	MOB _H
pTiS4	NC_011982	<i>Agrobacterium vitis</i> S4	258824	57	41	94	27356	27102	–	MOB _Q
pTi-SAKURA	NC_002147	<i>Agrobacterium tumefaciens</i>	206479	56	44	94	83408	83154	–	MOB _Q
pVSAL840	NC_011311	<i>Altivibrio salmonicida</i> LF11238	83540	40	60	99	95763	95509	–	MOB _Q
pYRI	NC_009139	<i>Yersinia ruckeri</i>	158038	51	58	99	31361	31627	–	MOB _F
Ti	NC_003065	<i>Agrobacterium tumefaciens</i> str. C58	214233	57	44	94	77350	77084	–	MOB _H
							15070	15336	–	MOB _H
							139735	139481	–	MOB _Q

^aThis list is the result of a TBLASTN analysis using the amino acid sequence of HU α or HU β as a query under strict conditions (i.e., thresholds of 30% identity and 70% query coverage).
^bAverage G+C content of the plasmid.

^cReported TBLASTN identity to HU.

^dPlasmid classification according to its source organism (–, Gram-negative plasmid; +, Gram-positive plasmid).

^ePlasmid classification according to its relaxase gene sequence as described by Garcillán-Barcia et al. [16].

TABLE 3: Plasmids containing the gene encoding IHF homolog^a.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	Classification ^d	MOB family ^e
At	NC_003064	<i>Agrobacterium tumefaciens</i> str. C58	542868	57	36	82	112654	112412	-	MOB _Q
Megaplasmid	NC_012811	<i>Methylobacterium extorquens</i> AM1	1261460	68	33	94	720582	720860	-	
p2META1	NC_012809	<i>Methylobacterium extorquens</i> AM1	37858	65	44	95	28369	28635	-	MOB _Q
pAAC101	NC_013206	<i>Alicyclobacillus acidocaldarius</i> subsp. <i>acidocaldarius</i> DSM 446	91726	54	43	80	62668	62432	+	
pACHL01	NC_011879	<i>Arthrobacter chlorophenolicus</i> A6	426858	64	32	92	408818	408546	+	
pALVIN02	NC_013862	<i>Allochromatium vinosum</i> DSM 180	39929	53	60	98	10902	10627	-	
pAph01	NC_013193	<i>Candidatus Accumulibacter phosphatis</i> clade IIA str. UW-1	167595	62	56	95	144197	144463	-	MOB _P
pAph03	NC_013191	<i>Candidatus Accumulibacter phosphatis</i> clade IIA str. UW-1	37695	59	58	97	5412	5140	-	
pAfk84b	NC_011990	<i>Agrobacterium radiobacter</i> K84	184668	59	38	86	54109	53855	-	MOB _Q
pAfk84c	NC_011987	<i>Agrobacterium radiobacter</i> K84	388169	57	43	93	340807	340532	-	
pAIS4b	NC_011991	<i>Agrobacterium vitis</i> S4	130435	56	46	93	10327	10052	-	
pBBta01	NC_009475	<i>Bradyrhizobium</i> sp. BTA1	228826	61	39	86	6642	6388	-	MOB _Q
pBFY46	NC_006297	<i>Bacteroides fragilis</i> YCH46	33716	34	35	89	25098	25343	-	MOB _P
pBIND01	NC_010580	<i>Beijerinckia indica</i> subsp. <i>indica</i> ATCC 9039	181736	56	36	77	179816	179601	-	MOB _F
pCHQ1	NC_014007	<i>Sphingobium japonicum</i> UT26S	190974	63	36	90	63111	63377	-	
pGLOV01	NC_010815	<i>Geobacter lovleyi</i> SZ	77113	53	38	92	41196	41468	-	
pM44601	NC_010373	<i>Methylobacterium</i> sp. 4-46	57951	65	35	97	7806	7534	-	
pMPOP01	NC_010727	<i>Methylobacterium populi</i> BJ001	25164	65	49	93	10635	10375	-	
pMIRAD03	NC_010514	<i>Methylobacterium radiotolerans</i> JCM 2831	42985	63	38	94	26778	26515	-	MOB _F

TABLE 3: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	Classification ^d	MOB family ^e
pMRAD04	NC_010517	<i>Methylobacterium radiotolerans</i> JCM 2831	37743	64	38	94	10763	10500	–	
pPRO1	NC_008607	<i>Pelobacter propionicus</i> DSM 2379	202397	48	41	94	129679	129957	–	
pRSPA01	NC_009429	<i>Rhodobacter sphaeroides</i> ATCC 17025	877879	68	49	97	783519	783791	–	
pSWIT01	NC_009507	<i>Sphingomonas wittichii</i> RW1	310228	64	40	95	106554	106820	–	MOB _F
pTcM1	NC_010600	<i>Acidithiobacillus caldus</i>	65158	57	56	89	35341	35069	–	MOB _P , MOB _Q
pXCV183	NC_007507	<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i> str. 85-10	182572	60	33	95	138753	138490	–	
Ti	NC_002377	<i>Agrobacterium tumefaciens</i>	194140	55	43	97	180164	180436	–	MOB _Q
Ti plasmid pTiBo542	NC_010929	<i>Agrobacterium tumefaciens</i>	244978	55	36	86	209743	209489	–	MOB _Q
					45	98	187204	187479		

^a This list is the result of a TBLASTN analysis using the amino acid sequence of IHF α or IHF β as a query under strict conditions (i.e., thresholds of 30% identity and 70% query coverage).

^b Average G+C content of the plasmid.

^c Reported TBLASTN identity to IHF.

^d Plasmid classification according to its source organism (–, Gram-negative plasmid; +, Gram-positive plasmid).

^e Plasmid classification according to its relaxase gene sequence as described by Garcillán-Barcia et al. [16].

TABLE 4: Plasmids containing the gene encoding Lrp homolog^a.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	Classification ^d	MOB family ^e
1	NC_008688	<i>Paracoccus denitrificans</i> PDI222	653815	67	41	92	252075	251623	–	
A	NC_009007	<i>Rhodobacter sphaeroides</i> 2.4.1	114045	69	39	93	30241	29789	–	MOB _F
B	NC_007488	<i>Rhodobacter sphaeroides</i> 2.4.1	114178	70	43	96	81861	81385	–	
bglu_1p	NC_012723	<i>Burkholderia glumae</i> BGR1	133591	61	36	90	124017	123577	–	
Megaplasmid	NC_008043	<i>Ruegeria</i> sp. TM1040	821788	59	41	84	143820	144233	–	
					41	91	687257	687706	–	
					36	91	734136	733690	–	
Megaplasmid	NC_007974	<i>Cupriavidus metallidurans</i> CH34	2580084	64	44	88	1171245	1170814	–	MOB _V
					40	91	1169702	1169256	–	
					38	97	1586726	1586250	–	
Megaplasmid	NC_006569	<i>Ruegeria pomeroyi</i> DSS-3	491611	63	36	88	356303	355869	–	MOB _C
Megaplasmid	NC_007336	<i>Ralstonia eutropha</i> JMP134	634917	61	35	93	377503	377045	–	
p42e	NC_007765	<i>Rhizobium etli</i> CFN 42	505334	62	34	71	255037	255384	–	
p42f	NC_007766	<i>Rhizobium etli</i> CFN 42	642517	61	45	88	436907	437341	–	
					43	91	406350	405901	–	
					41	85	491383	491799	–	
					39	95	210634	211098	–	
					39	96	199426	199899	–	
pAB510a	NC_013855	<i>Azospirillum</i> sp. B510	1455109	68	57	88	274908	275342	–	
					44	95	979549	980013	–	
					32	94	1180335	1179874	–	
pAB510b	NC_013856	<i>Azospirillum</i> sp. B510	723779	67	44	84	471830	472243	–	
					32	94	318139	318600	–	
pAB510c	NC_013857	<i>Azospirillum</i> sp. B510	681723	67	45	85	408064	407645	–	
					34	91	36385	36834	–	
pAB510d	NC_013858	<i>Azospirillum</i> sp. B510	628837	68	44	79	472768	472379	–	
					40	90	323184	322741	–	
					37	87	281438	281866	–	
					30	85	619027	618623	–	
pAtS4e	NC_011981	<i>Agrobacterium vitis</i> S4	631775	57	30	87	460443	460871	–	MOB _Q
					34	74	425247	424888	–	
pBPHY01	NC_010625	<i>Burkholderia phymatum</i> STM815	1904893	62	46	85	1153608	1154027	–	
pBPHY02	NC_010627	<i>Burkholderia phymatum</i> STM815	595108	59	41	91	271795	271346	–	

TABLE 4: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	Classification ^d	MOB family ^e
pC	NC_010997	<i>Rhizobium ethi</i> CIAT 652	1091523	61	46	88	617696	618130	–	MOB _Q
					42	90	609059	608619		
					39	95	417738	418202		
					42	79	714804	715193		
pCAUL01	NC_010335	<i>Caulobacter</i> sp. K31	233649	67	39	93	406570	407025		MOB _Q
pEST4011	NC_005793	<i>Achromobacter denitrificans</i>	76958	62	58	88	182479	182042	–	MOB _P
					58	88	41224	40793		
					43	98	34233	33802		
pGMI1000MP	NC_003296	<i>Ralstonia solanacearum</i> GM11000	2094509	67	46	93	1737958	1738437	–	
					33	71	822030	821572		
pHV4	NC_013966	<i>Haloferax volcanii</i> DS2	635786	62	58	88	401763	401410	Archaea	
pIJB1	NC_013666	<i>Burkholderia cepacia</i>	99448	63	33	88	74907	75338	–	MOB _P
pK2044	NC_006625	<i>Klebsiella pneumoniae</i> NTUH-K2044	224152	50	33	90	194643	195086	–	
pLVPK	NC_005249	<i>Klebsiella pneumoniae</i>	219385	50	33	90	46236	46679	–	
pMLa	NC_002679	<i>Mesorhizobium loti</i> MAFF303099	351911	59	32	93	185603	185148	–	
					30	89	207314	206877		
pMLb	NC_002682	<i>Mesorhizobium loti</i> MAFF303099	208315	60	37	93	24632	24177	–	
pNGR234a	NC_000914	<i>Rhizobium</i> sp. NGR234	536165	58	41	70	197189	196845	–	MOB _Q
					30	89	188867	188430		
					46	90	656547	656107	–	MOB _Q
pNGR234b	NC_012586	<i>Rhizobium</i> sp. NGR234	2430033	62	45	85	667494	667913	–	
					43	90	1038020	1038463		
					44	85	682796	683215		
					38	96	2400849	2401319		
					44	79	709104	708715		
					41	89	28336	28761		
					33	89	1108900	1109337		
					36	90	703213	702764		
					32	77	1112953	1112582		
pPNAP04	NC_008760	<i>Polaromonas naphthalenivorans</i> CJ2	143747	59	35	90	142511	142068	–	
pR132501	NC_012848	<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM1325	828924	60	47	88	234905	234471	–	MOB _Q
					44	86	386338	386760		

TABLE 4: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	Classification ^d	MOB family ^e
pRLTA	NC_010529	<i>Cupriavidus taiwanensis</i>	557200	60	39	93	645542	645087	–	
pRHL1	NC_008269	<i>Rhodococcus jostii</i> RHAI	1123075	65	36	91	854207	854656	+	
pRL12	NC_008378	<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841	870021	61	46	88	599116	598682	–	MOB _Q
pRL8	NC_008383	<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841	147463	59	33	87	70763	70344	–	MOB _Q
pRLG201	NC_011368	<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM2304	1266105	60	45	89	917573	917136	–	MOB _Q
pRSKD131A	NC_011962	<i>Rhodobacter sphaerooides</i> KDI131	157345	70	42	96	148295	147819	–	
pRSKD131B	NC_011960	<i>Rhodobacter sphaerooides</i> KDI131	103355	70	39	93	98400	97948	–	
pRSPA01	NC_009429	<i>Rhodobacter sphaerooides</i> ATCC 17025	877879	68	40	90	31309	30866	–	
pRSPH01	NC_009040	<i>Rhodobacter sphaerooides</i> ATCC 17029	122606	70	39	93	118088	118540	–	
pSMED01	NC_009620	<i>Sinorhizobium medicae</i> WSM419	1570951	61	40	77	143180	143557	–	MOB _Q
pSMED02	NC_009621	<i>Sinorhizobium medicae</i> WSM419	1245408	60	42	91	556486	556932	–	MOB _Q
pSMED03	NC_009622	<i>Sinorhizobium medicae</i> WSM419	219313	60	46	95	105044	105508	–	

TABLE 4: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	Classification ^d	MOB family ^e
pSmeSM11a	NC_013545	<i>Sinorhizobium meliloti</i>	144170	60	46	96	70449	70922	–	MOB _Q
pSymA	NC_003037	<i>Sinorhizobium meliloti</i> 1021	1354226	60	43	89	1060699	1060262	–	MOB _Q
pSymB	NC_003078	<i>Sinorhizobium meliloti</i> 1021	1683333	62	38	90	440778	440335	–	MOB _Q
pTIS4	NC_011982	<i>Agrobacterium vitis</i> S4	258824	57	36	89	29555	29992	–	MOB _Q
Unnamed	NC_011961	<i>Thermomicrobium roseum</i> DSM 5159	917738	66	42	79	96920	97309	–	MOB _P

^aThis list is the result of a TBLASTN analysis using the amino acid sequence of Lrp as a query under strict conditions (i.e., thresholds of 30% identity and 70% query coverage).

^bAverage G+C content of the plasmid.

^cReported TBLASTN identity to Lrp.

^dPlasmid classification according to its source organism (–, Gram-negative plasmid; +, Gram-positive plasmid).

^ePlasmid classification according to its relaxase gene sequence as described by Garcillán-Barcia et al. [16].

TABLE 5: Plasmids containing the gene encoding MvaT or NdpA homolog^a.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	MvaT		Query coverage (%)	Subject start	Subject end	Classification ^d	MOB family ^e
					Identity (%) ^c	MOB family ^e					
pCARI	NC-004444	<i>Pseudomonas resinovorans</i>	199035	56	61	98	77640	77993	–	–	MOB _H
pQBR103	NC-009444	<i>Pseudomonas fluorescens</i> SBW25	425094	53	61	96	98076	97717	–	–	–
pWW53	NC-008275	<i>Pseudomonas putida</i>	107929	57	61	98	8415	8768	–	–	–
NdpA											
p0908	NC-010113	<i>Vibrio</i> sp. 0908	81413	49	51	99	79731	78736	–	–	–
pCARI	NC-004444	<i>Pseudomonas resinovorans</i>	199035	56	36	98	95390	94395	–	–	MOB _H
pQBR103	NC-009444	<i>Pseudomonas fluorescens</i> SBW25	425094	53	31	99	161413	160400	–	–	–

^aThis list is the result of a TBLASTN analysis using the amino acid sequence of MvaT or NdpA as a query under strict conditions (i.e., thresholds of 30% identity and 70% query coverage).

^bAverage G+C content of the plasmid.

^cReported TBLASTN identity to MvaT or NdpA.

^dPlasmid classification according to its source organism (–, Gram-negative plasmid).

^ePlasmid classification according to its relaxase gene sequence as described by Garcillán-Barcia et al. [16].

TABLE 6: Gram-negative plasmids containing the gene encoding Hha-like protein^a.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	NAP gene homolog	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	MOB family ^d
55989p	NC_011752	<i>Escherichia coli</i> 55989	72482	46		53	92	10025	9828	
NR1	NC_009133	<i>Escherichia coli</i>	94289	52		53	92	87193	87390	MOB _F
p1658/97	NC_004998	<i>Escherichia coli</i>	125491	51		55	92	36419	36616	MOB _F
p1ESCUM	NC_011749	<i>Escherichia coli</i> UMN026	122301	50		53	92	53508	53311	MOB _F
p2ESCUM	NC_011739	<i>Escherichia coli</i> UMN026	33809	42		62	90	7682	7488	MOB _Q
p53638.226	NC_010719	<i>Escherichia coli</i> 53638	225683	48		55	92	67615	67418	MOB _F
pAPEC-O1-R	NC_009838	<i>Escherichia coli</i> APEC O1	241387	46		50	92	61389	61586	MOB _H
pAPEC-O2-CoIV	NC_007675	<i>Escherichia coli</i>	184501	49		55	92	3882	3685	MOB _F
pAPEC-O2-R	NC_006671	<i>Escherichia coli</i>	101375	53		53	92	4856	4659	MOB _F
pBS512.211	NC_010660	<i>Shigella boydii</i> CDC 3083-94	210919	46		55	89	190719	190910	MOB _F
pBS512.33	NC_010657	<i>Shigella boydii</i> CDC 3083-94	33103	41		62	90	2894	3088	
pCJ5-1a	NC_005327	<i>Escherichia coli</i>	92353	53		53	92	87490	87687	MOB _F
pCP301	NC_004851	<i>Shigella flexneri</i> 2a str. 301	221618	46		55	92	207828	208025	MOB _F
pCROD1	NC_013717	<i>Citrobacter rodentium</i> ICC168	54449	47		56	92	53220	53417	
pCROD2	NC_013718	<i>Citrobacter rodentium</i> ICC168	39265	42		62	90	15526	15332	
pCT02021853_74	NC_011204	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Dublin str. CT_02021853	74551	49		62	90	48482	48288	MOB _Q
pCTX-M3	NC_004464	<i>Citrobacter freundii</i>	89468	51		38	71	26136	26294	MOB _P
pCTXM360	NC_011641	<i>Klebsiella pneumoniae</i>	89468	51		31	96	40648	40439	
			68018	51		38	71	64551	64709	MOB _P
			68018	51		31	96	10927	10718	
pCVM29188_146	NC_011076	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Kentucky str. CVM29188	146811	49		53	92	18755	18558	MOB _F
pEC14_114	NC_013175	<i>Escherichia coli</i>	114222	51		53	92	113985	114182	MOB _F
pEC-IMP	NC_012555	<i>Enterobacter cloacae</i>	318782	48	H-NS	50	92	60491	60688	MOB _H
pEC-IMPQ	NC_012556	<i>Enterobacter cloacae</i>	324503	48	H-NS	50	92	60491	60688	MOB _H
pEG356	NC_013727	<i>Shigella sonnei</i>	70275	52		53	92	69444	69641	MOB _F
pEK499	NC_013122	<i>Escherichia coli</i>	117536	53		53	92	41985	42182	
pEK516	NC_013121	<i>Escherichia coli</i>	64471	53		53	92	31410	31213	

TABLE 6: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	NAP gene homolog	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	MOB family ^d
pEL60	NC_005246	<i>Erwinia amylovora</i>	60145	51		38	71	23187	23345	MOB _P
pEntH10407	NC_013507	<i>Escherichia coli</i> ETEC H10407	60145			31	96	37863	37654	
pHCM1	NC_003384	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi str. CT18	218160	48	H-NS	47	100	105911	106117	MOB _H
pK2044	NC_006625	<i>Klebsiella pneumoniae</i> NTUH-K2044	224152	50	H-NS, Lrp	45	85	143331	143528	
pK29	NC_010870	<i>Klebsiella pneumoniae</i>	269674	46		50	92	59322	59519	MOB _H
pKF3-70	NC_013542	<i>Klebsiella pneumoniae</i>	70057	52		53	92	15967	15770	MOB _F
pKF3-94	NC_013950	<i>Klebsiella pneumoniae</i>	94219	52		58	96	9596	9390	MOB _F
pKP187	NC_011282	<i>Klebsiella pneumoniae</i> 342	187922	47		64	96	110083	109877	
			187922			42	89	1550	1344	
pKPN3	NC_009649	<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> MGH 78578	175879	52		59	97	56930	56721	MOB _F
plasmid_153 kb	NC_009705	<i>Yersinia pseudotuberculosis</i> IP 31758	153140	40	H-NS	69	93	63342	63542	
pLVPK	NC_005249	<i>Klebsiella pneumoniae</i>	153140			56	92	49734	49931	
			219385	50	H-NS, Lrp	61	97	148056	147847	
			219385			45	85	214828	215025	
pMAK1	NC_009981	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Choleraesuis	208409	47	H-NS	47	100	49208	49414	MOB _H
pMAS2027	NC_013503	<i>Escherichia coli</i>	42644	43		62	90	19685	19491	MOB _Q
pO103	NC_013354	<i>Escherichia coli</i> O103:H2 str. 12009	75546	49		55	92	51727	51924	MOB _F
pO111_1	NC_013365	<i>Escherichia coli</i> O111:H- str. 11128	204604	47	H-NS	47	100	66925	67131	MOB _H
pO111_3	NC_013366	<i>Escherichia coli</i> O111:H- str. 11128	77690	50		55	92	11975	12172	MOB _F
pO157	NC_013010	<i>Escherichia coli</i> O157:H7 str. TW14359	94601	48		55	92	70792	70989	
pO157	NC_011350	<i>Escherichia coli</i> O157:H7 str. EC4115	94644	48		55	92	54735	54932	
pO157	NC_007414	<i>Escherichia coli</i> O157:H7 EDL933	92077	48		55	92	1667	1864	
pO157	NC_002128	<i>Escherichia coli</i> O157:H7 str. Sakai	92721	48		55	92	71183	71380	

TABLE 6: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	NAP gene homolog	Identity (%) ^c	Query coverage (%)	Subject start	Subject end	MOB family ^d
pO261	NC_011812	<i>Escherichia coli</i>	72946	51		53	92	66608	66805	MOB _F
pO86A1	NC_008460	<i>Escherichia coli</i>	120730	49		55	92	101598	101795	MOB _F
pOLA52	NC_010378	<i>Escherichia coli</i>	51602	46		62	90	12114	11920	MOB _Q
pOU1114	NC_010421	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Dublin	34595	41		62	90	5446	5252	MOB _Q
pOU1115	NC_010422	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Dublin	74589	49		62	90	37246	37052	MOB _Q
pSB4_227	NC_007608	<i>Shigella boydii</i> Sb227	126697	47		55	92	110688	110885	MOB _F
pSE11-1	NC_011419	<i>Escherichia coli</i> SE11	100021	50		56	92	58407	58210	MOB _P
pSE34	NC_010860	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Enteritidis	32950	41		62	90	21875	22069	MOB _Q
pSFO157	NC_009602	<i>Escherichia coli</i>	121239	50		52	75	1709	1870	MOB _F
pSG1	NC_007713	<i>Sodalis glossinidius</i> str. "morsitans"	83306	49	H-NS	48	92	2294	2491	
pSG1	NC_007182	<i>Sodalis glossinidius</i>	81553	49		48	92	56217	56414	
pSMS35_130	NC_010488	<i>Escherichia coli</i> SMS-3-5	130440	51		55	92	3364	3167	MOB _F
pSS_046	NC_007385	<i>Shigella sonnei</i> Ss046	214396	45		55	92	178363	178560	MOB _F
pUTI89	NC_007941	<i>Escherichia coli</i> UTT189	114230	51		53	92	113993	114190	MOB _F
pWR501	NC_002698	<i>Shigella flexneri</i>	221851	46		55	92	207534	207731	MOB _F
R100	NC_002134	<i>Shigella flexneri</i> 2b str. 222	94281	52		53	92	87185	87382	MOB _F
R27	NC_002305	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi	180461	46	H-NS	47	100	159402	159196	MOB _H
R478	NC_005211	<i>Serratia marcescens</i>	274762	46	H-NS	50	92	59426	59623	MOB _H
R721	NC_002525	<i>Escherichia coli</i>	75582	43		66	90	35285	35091	
Unnamed	NC_011148	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Agona str. SL483	37978	41	H-NS	42	93	1363	1163	

^aThis list is the result of a TBLASTN analysis using the amino acid sequence of Hha as a query under strict conditions (i.e., thresholds of 30% identity and 70% query coverage).

^bAverage G+C content of the plasmid.

^cReported TBLASTN identity to Hha.

^dPlasmid classification according to its relaxase gene sequence as described by Garcillán-Barcia et al. [16].

TABLE 7: MOB_H-family plasmids of Gram-negative origin^a.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	NAP gene homolog	Identity (%) ^c	Query coverage (%)	Subject start	Subject end
1	NC_007949	<i>Polaromonas</i> sp. JS666	360405	57	HU	52	99	61052	60786
1	NC_008573	<i>Shewanella</i> sp. ANA-3	278942	46					
2	NC_007950	<i>Polaromonas</i> sp. JS666	338007	60					
ICEhin1056	NC_011409	<i>Haemophilus influenzae</i>	59393	39					
pAM04528	NC_012693	<i>Salmonella enterica</i>	158213	52	HU	57	99	14067	14333
pAPEC-O1-R	NC_009838	<i>Escherichia coli</i> APEC O1	241387	46	HU	57	99	15755	16021
pAR060302	NC_012692	<i>Escherichia coli</i>	166530	53	HU	60	99	26844	26578
pAsa4	NC_009349	<i>Aeromonas salmonicida</i>	166749	53	HU	61	98	77640	77993
pCAR1	NC_004444	subsp. <i>salmonicida</i> A449	199035	56	MvaT	36	98	95390	94395
		<i>Pseudomonas resinovorans</i>			NdpA	42	99	97809	98075
pEC-IMP	NC_012555	<i>Enterobacter cloacae</i>	318782	48	H-NS	64	99	109370	108969
pEC-IMPQ	NC_012556	<i>Enterobacter cloacae</i>	324503	48	H-NS	64	99	109370	108969
peH4H	NC_012690	<i>Escherichia coli</i>	148105	53	HU	57	99	14067	14333
pHCM1	NC_003384	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi str. CT18	218160	48	H-NS	61	99	131861	131460
pIP1202	NC_009141	<i>Yersinia pestis</i> bv. <i>Orientalis</i> str. IP275	182913	53	HU	57	99	14067	14333
pK29	NC_010870	<i>Klebsiella pneumoniae</i>	269674	46					
plasmid1	NC_007901	<i>Rhodoferrax ferritducens</i> T118	257447	54					
pMAK1	NC_009981	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Choleraesuis	208409	47	H-NS	61	99	60046	59645
pMAQU02	NC_008739	<i>Marinobacter aquaeolei</i> VT8	213290	53					
pO111-1	NC_013365	<i>Escherichia coli</i> O111:H-str. 11128	204604	47	H-NS	61	99	80175	79774
pP91278	NC_008613	<i>Photobacterium damsela</i> subsp. <i>Piscitida</i>	131520	52	HU	57	99	125918	126184

TABLE 7: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	NAP gene homolog	Identity (%) ^c	Query coverage (%)	Subject start	Subject end
pP99-018	NC_008612	<i>Photobacterium damisela</i> subsp. <i>piscicida</i>	150157	51	HU	57	99	133314	133580
pRA1	NC_012885	<i>Aeromonas hydrophila</i>	143963	51	HU	58	99	15573	15839
pRp12D01	NC_012855	<i>Ralstonia pickettii</i> 12D	389779	58	HU	37	99	321346	321080
pSN254	NC_009140	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport str. SL254	176473	53	HU	57	99	14067	14333
pTK9001	NC_013930	<i>Thioalkalivibrio</i> sp. K90mix	240256	62					
pYR1	NC_009139	<i>Yersinia ruckeri</i>	158038	51	HU	57	99	15070	15336
R27	NC_002305	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi	180461	46	H-NS	61	99	148225	148626
R478	NC_005211	<i>Serratia marcescens</i>	274762	46	H-NS	64	99	111747	111346
Rts1	NC_003905	<i>Proteus vulgaris</i>	217182	46					

^aThis list is the result of a TBLASTN analysis using the 300 N-terminal amino acid sequence of protein TraI_R27 as a query under strict conditions (i.e., thresholds of 30% identity and 70% query coverage).

^bAverage G+C content of the plasmid.

^cReported TBLASTN identity to each NAP.

TABLE 8: MOB_Q-family plasmids of Gram-negative origin^a.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	NAP gene homolog	Identity (%) ^c	Query coverage (%)	Subject start	Subject end
1	NC_008242	<i>Chelativorans</i> sp. BNC1	343931	62	HU	41	94	133932	133678
3	NC_007617	<i>Nitrosospira multiformis</i> ATCC 25196	14159	50					
3	NC_007961	<i>Nitrobacter hamburgensis</i> X14	121408	62					
At	NC_003064	<i>Agrobacterium tumefaciens</i> str. C58	542868	57	IHF	36	82	112654	112412
C	NC_010542	<i>Cyanotheca</i> sp. ATCC 51142	14685	38					
ColE9-J	NC_011977	<i>Escherichia coli</i>	7577	50					
DN1	NC_002636	<i>Dichelobacter nodosus</i>	5112	62					
F plasmid	NC_008036	<i>Sphingopyxis alaskensis</i> RB2256	28543	60					
p11745	NC_013546	<i>Actinobacillus pleuropneumoniae</i>	5486	38					
p12494	NC_010889	<i>Actinobacillus pleuropneumoniae</i>	14393	33					
p1ABAYE	NC_010401	<i>Acinetobacter baumannii</i> AYE	5644	35					
p1META1	NC_012807	<i>Methylobacterium extorquens</i> AM1	44195	68					
p1METDI	NC_012987	<i>Methylobacterium extorquens</i> DM4	141504	65					
p2007057	NC_011897	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Bovismorbificans	4270	47					
p2ABSDF	NC_010396	<i>Acinetobacter baumannii</i> SDF	25014	35					
p2ESCUM	NC_011739	<i>Escherichia coli</i> UMN026	33809	42					
p2META1	NC_012809	<i>Methylobacterium extorquens</i> AM1	37858	65	IHF	44	95	28369	28635
p3ABSDF	NC_010398	<i>Acinetobacter baumannii</i> SDF	24922	34					
p42a	NC_007762	<i>Rhizobium etli</i> CFN 42	194229	58					
p49879.1	NC_006907	<i>Leptospirillum ferrooxidans</i>	28878	58	HU	47	99	3281	3015

TABLE 8: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	NAP gene homolog	Identity (%) ^c	Query coverage (%)	Subject start	Subject end
p49879.2	NC_006909	<i>Leptospirillum ferrooxidans</i>	28012	55	HU	48	99	15858	15592
pAb5S9	NC_009476	<i>Aeromonas bestiarum</i>	24716	54					
pACRY07	NC_009473	<i>Acidiphilium cryptum</i> JF-5	5629	58					
pAgK84	NC_011994	<i>Agrobacterium radiobacter</i> K84	44420	54					
pAM5	NC_008691	<i>Acidiphilium multivorum</i>	5161	58					
pAMI2	NC_010847	<i>Paracoccus aminophilus</i>	18563	62					
pAMI3	NC_013513	<i>Paracoccus aminophilus</i>	5575	61					
pAPA01-030	NC_013212	<i>Acetobacter pasteurianus</i> IFO 3283-01	49961	54					
pAPA01-040	NC_013213	<i>Acetobacter pasteurianus</i> IFO 3283-01	3204	54					
pAUK84b	NC_011990	<i>Agrobacterium radiobacter</i> K84	184668	59	IHF	38	86	54109	53855
pAtS4b	NC_011991	<i>Agrobacterium vitis</i> S4	130435	56	IHF	47	97	44880	45152
pAtS4c	NC_011984	<i>Agrobacterium vitis</i> S4	211620	59	HU	45	94	141245	140991
pAtS4e	NC_011981	<i>Agrobacterium vitis</i> S4	631775	57	HU	41	94	40476	40222
pAV2	NC_010310	<i>Acinetobacter venetianus</i>	15135	36	Lrp	30	87	460443	460871
pB	NC_010996	<i>Rhizobium ethi</i> CIAT 652	429111	58	Lrp	34	74	425247	424888
pBGR3	NC_012847	<i>Bartonella grahamii</i> as4aup	28192	36					
pBS512.5	NC_010659	<i>Shigella boydii</i> CDC 3083-94	5114	46					
pC	NC_010997	<i>Rhizobium ethi</i> CIAT 652	1091523	61	Lrp	46	88	617696	618130
pCAUL01	NC_010335	<i>Caulobacter</i> sp. K31	233649	67	Lrp	42	90	609059	608619
					Lrp	39	95	417738	418202
					Lrp	42	79	714804	715193
					Lrp	39	93	406570	407025
					HU	44	99	97598	97329
					Lrp	34	89	182479	182042

TABLE 8: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	NAP gene homolog	Identity (%) ^c	Query coverage (%)	Subject start	Subject end
pCAUL02	NC_010333	<i>Caulobacter</i> sp. K31	177878	64					
pCCK1900	NC_011378	<i>Pasteurella multocida</i>	10226	61					
pCCK381	NC_006994	<i>Pasteurella multocida</i>	10874	61					
pCFPG4	NC_011563	<i>Candidatus Azobacteroides pseudotrichonympha</i> genomovar. CFP2	4149	44					
pCHE-A	NC_012006	<i>Enterobacter cloacae</i>	7560	60					
pColE8	NC_012882	<i>Escherichia coli</i>	6751	51					
pCROD3	NC_013719	<i>Citrobacter rodentium</i> ICC168	3910	51					
pCT02021853_74	NC_011204	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Dublin str. CT_02021853	74551	49					
pCVM19633_4	NC_011093	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Schwarzengrund str. CVM19633	4585	48					
pDSHI01	NC_009955	<i>Dinoroseobacter sibiriae</i> DFL 12	190506	60					
pET09	NC_010695	<i>Erwinia tasmaniensis</i> Et1/99	9299	47					
pGDIA01	NC_011367	<i>Gluconacetobacter diazotrophicus</i> PAI 5	27455	59					
pGOX3	NC_006674	<i>Gluconobacter oxydans</i> 621H	14547	56					
pHCG3	NC_005873	<i>Oligotropha carboxidovorans</i> OM5	133058	61					
pHRM2a	NC_012109	<i>Desulfobacterium autotrophicum</i> HRM2	68709	42					
pIGJC156	NC_009781	<i>Escherichia coli</i>	5146	47					
pIGMS5	NC_010883	<i>Escherichia coli</i>	6750	51					
pIGWZ12	NC_010885	<i>Escherichia coli</i>	4072	50					
pISP3	NC_013970	<i>Sphingomonas</i> sp. MM-1	43398	63					
pJD4	NC_002098	<i>Neisseria gonorrhoeae</i>	7426	38					
plasmid1	NC_007801	<i>Janaschia</i> sp. CCS1	86072	58					

TABLE 8: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	NAP gene homolog	Identity (%) ^c	Query coverage (%)	Subject start	Subject end
pLD-TEX-KL	NC_009966	<i>Fluoribacter dumoffii</i>	66512	39					
pMAC	NC_006877	<i>Acinetobacter baumannii</i>	9540	35					
pMAS2027	NC_013503	<i>Escherichia coli</i>	42644	43					
pMbo4.6	NC_013500	<i>Moraxella bovis</i>	4658	39					
pMCHL01	NC_011758	<i>Methylobacterium chloromethanicum</i> CM4	380207	66					
pMG160	NC_004527	<i>Rhodobacter blasticus</i>	3431	67					
pMG828-2	NC_008487	<i>Escherichia coli</i>	4091	50					
pMG828-4	NC_008489	<i>Escherichia coli</i>	7462	48					
pMIMCU1	NC_013056	<i>Acinetobacter calcoaceticus</i>	8771	35					
pMIMCU2	NC_013506	<i>Acinetobacter baumannii</i>	10270	36					
pMRAD01	NC_010510	<i>Methylobacterium radiotolerans</i> JCM 2831	586164	70					
pMS260	NC_005312	<i>Actinobacillus pleuropneumoniae</i>	8124	61					
pNGR234a	NC_000914	<i>Rhizobium</i> sp. NGR234	536165	59	Lrp	41	70	197189	196845
pNGR234b	NC_012586	<i>Rhizobium</i> sp. NGR234	2430033	62	Lrp	30	89	188867	188430
pNL2	NC_009427	<i>Novosphingobium aromaticivorans</i> DSM 12444	487268	66	Lrp	46	90	656547	656107
pO111_4	NC_013367	<i>Escherichia coli</i> O111:H-str. 11128	8140	50	Lrp	45	85	667494	667913
pO26-S4	NC_011228	<i>Escherichia coli</i>	6758	51	Lrp	43	90	1038020	1038463
pOLA52	NC_010378	<i>Escherichia coli</i>	51602	46	Lrp	44	85	682796	683215
pOU114	NC_010421	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Dublin	34595	42	Lrp	38	96	2400849	2401319
					Lrp	44	79	709104	708715
					Lrp	41	89	28336	28761
					Lrp	33	89	1108900	1109337
					Lrp	36	90	703213	702764
					Lrp	32	77	1112953	1112582

TABLE 8: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	NAP gene homolog	Identity (%) ^c	Query coverage (%)	Subject start	Subject end
pOU1115	NC_010422	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Dublin	74589	49					
pP	NC_003455	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Enteritidis</i>	4301	50					
pP742405	NC_011733	<i>Cyanotheca</i> sp. PCC 7424	18083	38					
pP742406	NC_011734	<i>Cyanotheca</i> sp. PCC 7424	15219	40					
pPMA4326C	NC_005921	<i>Pseudomonas syringae</i> pv. <i>maculicola</i>	8244	53					
pPNAP07	NC_008763	<i>Polaromonas naphthalenivorans</i> CJ2	9898	57					
pPRO2	NC_008608	<i>Pelobacter propionicus</i> DSM 2379	30722	56					
pPT1	NC_002143	<i>Comamonas testosteroni</i>	15398	56					
pR132501	NC_012848	<i>Rhizobium leguminosarum</i> bv. trifolii WSM1325	828924	60	Lrp	47	88	234905	234471
pR132502	NC_012858	<i>Rhizobium leguminosarum</i> bv. trifolii WSM1325	660973	61					
pR132503	NC_012853	<i>Rhizobium leguminosarum</i> bv. trifolii WSM1325	516088	59	Lrp	44	86	386338	386760
pR132504	NC_012852	<i>Rhizobium leguminosarum</i> bv. trifolii WSM1325	350312	61	Lrp	39	93	645542	645087
pR132505	NC_012854	<i>Rhizobium leguminosarum</i> bv. trifolii WSM1325	294782	60	Lrp	42	79	147165	146776
pRF	NC_007110	<i>Rickettsia felis</i> URRWXCal2	62829	34					
pRFdelta	NC_007111	<i>Rickettsia felis</i> URRWXCal2	39263	33					
pRi1724	NC_002575	<i>Agrobacterium rhizogenes</i>	217594	57					
pRi2659	NC_010841	<i>Agrobacterium rhizogenes</i>	185462	58					
pRL10	NC_008381	<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841	488135	60					
pRL11	NC_008384	<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841	684202	61					
pRL12	NC_008378	<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841	870021	61	Lrp	46	88	599116	598682

TABLE 8: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	NAP gene homolog	Identity (%) ^c	Query coverage (%)	Subject start	Subject end
pRL7	NC_008382	<i>Rhizobium leguminosarum</i> bv. viciae 3841	151564	58	Lrp	43	88	658287	658718
pRL8	NC_008383	<i>Rhizobium leguminosarum</i> bv. viciae 3841	147463	59	Lrp	42	79	450080	449691
pRLG201	NC_011368	<i>Rhizobium leguminosarum</i> bv. trifolii WSM2304	1266105	60	Lrp	45	87	70763	70344
pRM	NC_010927	<i>Rickettsia monacensis</i>	23486	32	Lrp	44	89	917573	917136
pSC101	NC_002056	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhimurium	9263	51	Lrp	44	85	41998	42417
pSE11-6	NC_011411	<i>Escherichia coli</i> SE11	4082	49	Lrp	44	79	473039	472650
pSE34	NC_010860	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Enteritidis	32950	41	Lrp	40	93	1162146	1161691
pSMED01	NC_009620	<i>Sinorhizobium medicae</i> WSM419	1570951	62	Lrp	40	93	1150939	1150484
pSMED02	NC_009621	<i>Sinorhizobium medicae</i> WSM419	1245408	60	Lrp	34	89	574284	573847
pSmeSM11a	NC_013545	<i>Sinorhizobium meliloti</i>	144170	60	Lrp	42	91	556486	556932
pSmeSM11b	NC_010865	<i>Sinorhizobium meliloti</i> SM11	181251	59	Lrp	40	91	842324	842758
pSMS35-4	NC_010486	<i>Escherichia coli</i> SMS-3-5	4074	50	Lrp	31	87	22345	21917
pSx-Qyy	NC_006826	<i>Sphingobium xenophagum</i>	5683	56	Lrp	46	96	70449	70922
pSymA	NC_003037	<i>Sinorhizobium meliloti</i> 1021	1354226	60	Lrp				
pSymB	NC_003078	<i>Sinorhizobium meliloti</i> 1021	1683333	62	Lrp	38	90	440778	440335

TABLE 8: Continued.

Plasmid name	Accession no.	Source organism	Length (nt)	G+C content (%) ^b	NAP gene homolog	Identity (%) ^c	Query coverage (%)	Subject start	Subject end
pTB3	NC_008388	<i>Roseobacter denitrificans</i> OCh 114	16575	55	Lrp	36	89	29555	29992
pTcM1	NC_010600	<i>Acidithiobacillus caldus</i>	65158	57	IHF	56	89	25186	25449
pTIS4	NC_011982	<i>Agrobacterium vitis</i> S4	258824	57	HU	41	94	27356	27102
					HU	40	94	83408	83154
					Lrp	42	79	96920	97309
pTi-SAKURA	NC_002147	<i>Agrobacterium tumefaciens</i>	206479	56	HU	44	94	95763	95509
pUT1	NC_014005	<i>Sphingobium japonicum</i> UT26S	31776	64					
pUT2	NC_014009	<i>Sphingobium japonicum</i> UT26S	5398	61					
pXAUT01	NC_009717	<i>Xanthobacter autotrophicus</i> Py2	316164	65					
pXCV19	NC_007505	<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i> str. 85-10	19146	60					
pXF51	NC_002490	<i>Xylella fastidiosa</i> 9a5c	51158	50					
pYAN-1	NC_008246	<i>Sphingobium yanoikuyae</i>	5182	62					
pYAN-2	NC_008247	<i>Sphingobium yanoikuyae</i>	4924	64					
RSF1010	NC_001740	<i>Escherichia coli</i>	8684	61					
Symbiotic plasmid p42d	NC_004041	<i>Rhizobium etli</i> CFN 42	371254	58					
Ti	NC_002377	<i>Agrobacterium tumefaciens</i>	194140	55	IHF	43	97	180164	180436
Ti	NC_003065	<i>Agrobacterium tumefaciens</i> str. C58	214233	57	HU	44	94	139735	139481
Ti plasmid pTiBo542	NC_010929	<i>Agrobacterium tumefaciens</i>	244978	55	IHF	36	86	209743	209489
Unnamed	NC_011143	<i>Phenylobacterium zucineum</i> HLK1	382976	69	IHF	45	98	187204	187479

^aThis list is the result of a TBLASTN analysis using the 300 N-terminal amino acid sequence of protein MobA_RSF1010 as a query under strict conditions (i.e., thresholds of 30% identity and 70% query coverage).

^bAverage G+C content of the plasmid.

^cReported TBLASTN identity to each NAP.

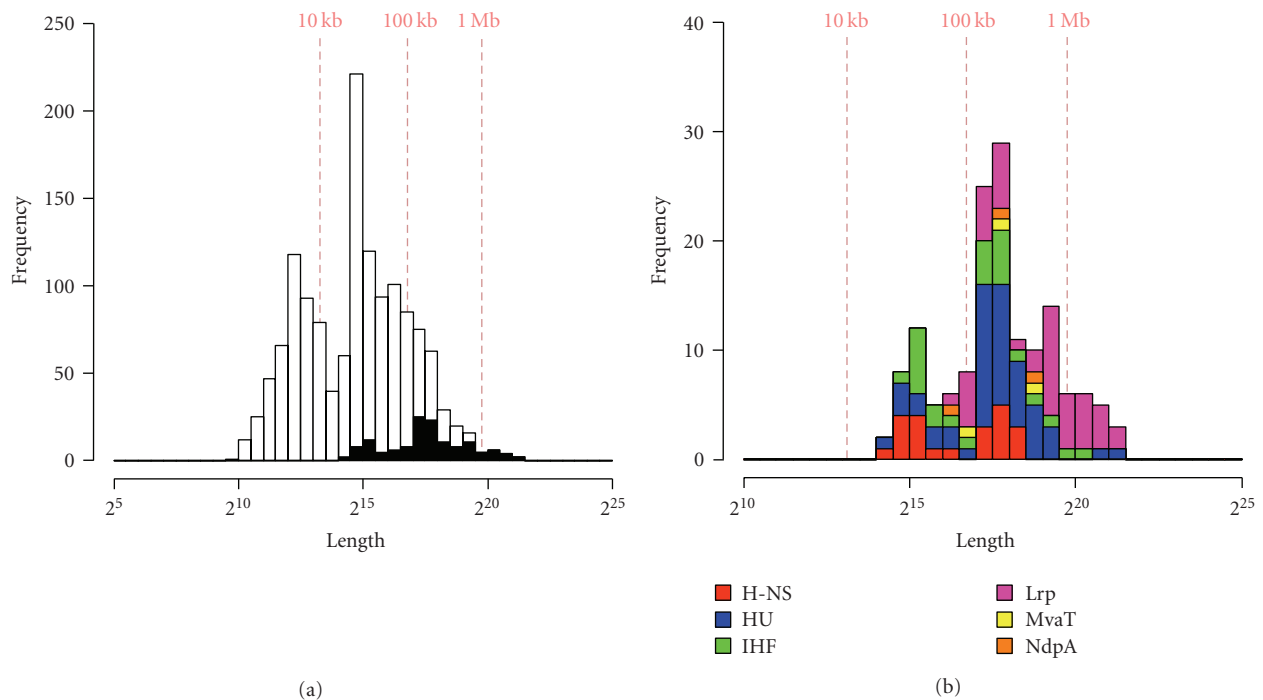


FIGURE 1: Size comparison of the Gram-negative plasmids with and without NAP gene homologs. (a) A total of 136 Gram-negative plasmids with one or more NAP gene homologs and 1246 Gram-negative plasmids without NAP gene homologs are shown by black and white bars, respectively. (b) Gram-negative plasmids with each NAP gene homolog are as follows: H-NS, red; HU, blue; IHF, green; Lrp, purple; MvaT, yellow; and NdpA, orange.

(83 kb). These results suggest that larger plasmids, especially >100 kb, frequently have NAP gene homologs. Carrying large plasmids may reduce host fitness more than carrying small plasmids because the former have more genes that can disrupt transcriptional networks in the host cell. In addition, large plasmids may have more binding sites for NAPs than small plasmids. Because chromosome-encoded NAPs bind to both chromosomes and plasmids, carrying large plasmids may also result in a reduction in the binding of NAPs to the host chromosome, causing undesirable effects on the host cell. Plasmid-encoded NAP homologs may interact with chromosome-encoded NAPs, coordinately sustain the structure of both chromosome and plasmid, and regulate the transcriptional regulation network [23]. In fact, recent studies have shown that some plasmid-encoded NAP homologs can complement the depletion of chromosomal NAPs and optimize gene transcription both on plasmids and in the host chromosome [14, 15, 24]. Thus, larger plasmids may have NAP gene homologs to maintain host cell fitness. In addition, the average size of the 38 plasmids containing more than one NAP gene homolog was larger (790 kb) than that of the 98 plasmids containing only one NAP gene homolog (199 kb). This suggests that particularly large plasmids have many NAP gene homologs to maintain themselves in the host cell.

Distributions of the NAP genes on proteobacterial genomes were also surveyed using the TBLASTN program. The average size of the completely sequenced bacterial genomes was 3.25 Mb and 1054 NAP genes (100, Fis; 125, H-NS; 236, HU; 247, IHF; 127, Lrp; 119, MvaT; and 100, NdpA)

were found in 588 proteobacterial genomes. Frequency of NAP genes in plasmids was higher (1 per 236 kb) than that in proteobacterial genomes (1 per 1.8 Mb), also suggesting that larger plasmids frequently have NAP gene homologs to minimize their negative effects on the host cell.

Of the plasmids with the NAP gene homolog, the average size of those with the H-NS gene homolog was relatively small (132 kb) while that of those with the Lrp gene homolog was relatively large (725 kb). The average sizes of those with the other NAP gene homologs were as follows: HU (301 kb), IHF (230 kb), MvaT (244 kb), and NdpA (235 kb) (Figure 1(b)). H-NS exists in an oligomeric form and binds to DNA, especially A+T-rich regions, by bridging it [25]. This function may be important for regulating gene expression on relatively small plasmids among those with the NAP gene homolog. The activity of H-NS can also be modulated by Hha-like proteins [26]. Intriguingly, TBLASTN analysis showed that 12 (55%) of 22 plasmids with the H-NS gene homolog also carried gene encoding Hha-like protein although only 65 (5%) of all 1382 plasmids carried Hha-like protein gene (Table 6). This suggests the close relationship of H-NS and Hha-like protein. On the other hand, Lrp exists in dimeric, octameric, and hexadecameric forms and compacts DNA by wrapping it [27]. This distinctive DNA-binding ability may be essential for maintaining the structure of particularly larger plasmids.

3.4. Relationships between Plasmid G+C Content and NAP Gene Homolog Distributions. Next, we surveyed the G+C content of the Gram-negative group plasmids with and

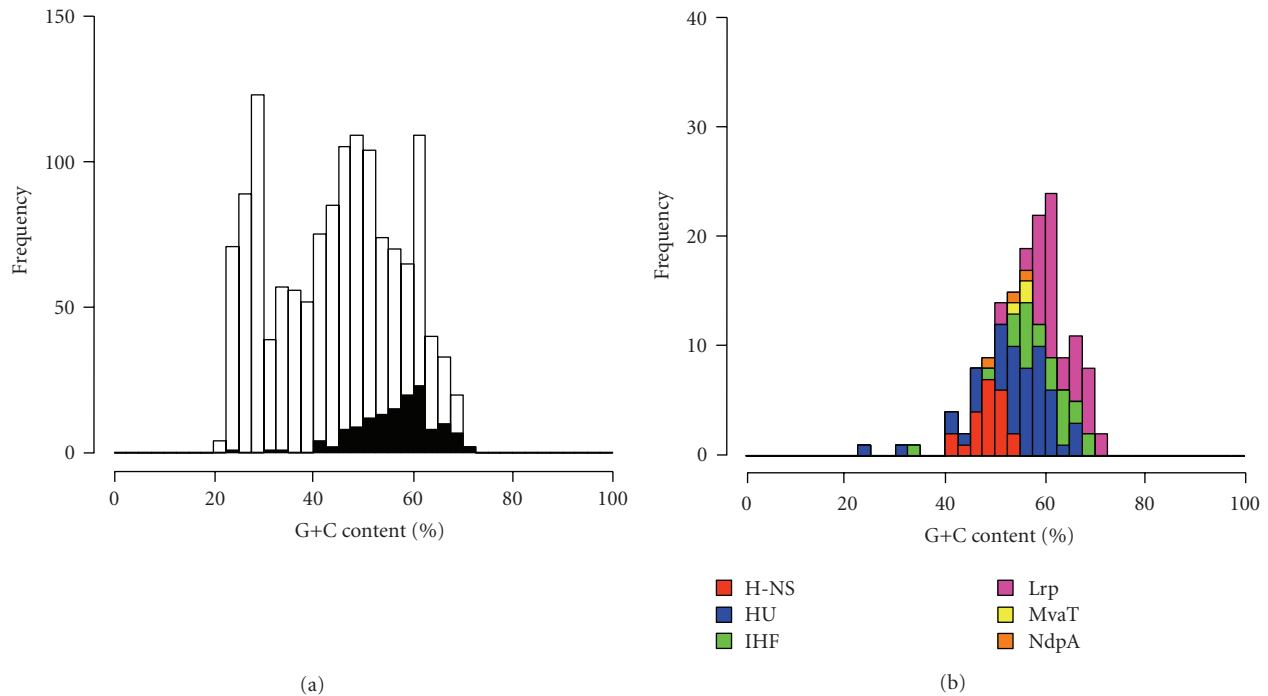


FIGURE 2: G+C content comparison of the Gram-negative plasmids with and without NAP gene homologs. (a) A total of 136 Gram-negative plasmids with one or more NAP gene homologs and 1246 Gram-negative plasmids without NAP gene homologs are shown by black and white bars, respectively. (b) Gram-negative plasmids with each NAP gene homolog are as follows: H-NS, red; HU, blue; IHF, green; Lrp, purple; MvaT, yellow; and NdpA, orange.

without NAP gene homologs. The average G+C content of the 136 plasmids with NAP gene homologs was higher (56.4%) than that of all 1382 plasmids (44.8%) (Figure 2(a)). Note that the average G+C content of large and mega plasmids (55.0% and 62.9%, resp.) was higher than that of small and intermediate plasmids (44.8% and 40.4%). Considering that larger plasmids frequently had NAP gene homologs, this seems reasonable. Nevertheless, plasmids with H-NS gene homologs had a lower G+C content (48.3%) than did those with other NAP gene homologs, including HU (54.2%), IHF (58.7%), Lrp (62.3%), MvaT (55.6%), and NdpA (52.9%) (Figure 2(b)). H-NS family protein binds A+T-rich regions not only on chromosomes but also on plasmids [15]. Acquisition of a large A+T-rich plasmid with many H-NS binding sites may result in a reduction in the binding of H-NS to the host chromosome and host cell fitness [14]. It is therefore possible that large A+T-rich plasmids may have to supply another H-NS encoded on themselves to minimize the effect on the host cell. On the other hand, although MvaT-family proteins are the functional homolog of H-NS [10, 15], plasmids containing the MvaT gene homolog were not particularly low in G+C content. Although only three plasmids contained the MvaT gene homolog and thus we cannot discuss this interesting phenomenon in detail, the difference between H-NS and MvaT may be derived from their different origin or host bacteria.

3.5. Relationships between Plasmid Transferability and NAP Gene Homolog Distributions. Conjugative transfer is an

essential function of plasmids, through which they play an important role in bacterial evolution and host cell behavior [11, 12]. Relaxase is an essential protein for plasmid transmission involved in the cleavage of the transferring DNA at the origin of transfer (*oriT*) site, and plasmids with relaxase genes are thought to be transmissible. Garcillán-Barcia et al. [16] proposed that transmissible plasmids can be classified into 6 MOB families (MOB_C, MOB_F, MOB_H, MOB_P, MOB_Q, and MOB_V) according to the amino acid sequences of 6 prototype relaxase proteins. MOB_F and MOB_H families are predominantly composed of conjugative plasmids, also called self-transmissible plasmids, and the other 4 families are composed of both mobilizable and conjugative plasmids. Recent studies have reported that plasmid-encoded H-NS family proteins have a “stealth” function and aide horizontal transfer of plasmids [14, 15]. Other NAPs also act as global transcriptional regulators and may regulate expression of genes involved in plasmid transmission. To discuss the relationship between NAP gene homolog distribution and plasmid transferability, we determined the distribution of genes encoding relaxase proteins in Gram-negative plasmids according to the classification by Garcillán-Barcia et al. [16]. Four hundred and nine (30%) of 1382 Gram-negative plasmids carried relaxase genes, and 71 (17%) of those 409 plasmids carried NAP gene homologs. Note that 71 (52%) of 136 plasmids with NAP gene homologs carried relaxase genes. This indicates that plasmids with NAP gene homologs frequently carried the relaxase genes than did those without NAP gene homologs. This phenomenon may be related to the average size of the plasmids. That of the 409 plasmids with

relaxase genes was relatively larger (145 kb) than that of all 1382 plasmids (83 kb), corresponding to the fact that larger plasmids frequently had NAP gene homologs.

Four hundred and nine plasmids were classified into each MOB family (13, MOB_C; 128, MOB_F; 29, MOB_H; 86, MOB_P; 131, MOB_Q; and 26, MOB_V). Plasmid 1 (NC_008545) was classified into both the MOB_C and MOB_F families. In addition, the MOB_P, MOB_Q, and MOB_V families were partially overlapped as described by Garcillán-Barcia et al. [16]. Seventy-one plasmids with NAP gene homologs were contained in each MOB family (1, MOB_C; 11, MOB_F; 20, MOB_H; 8, MOB_P; 30, MOB_Q; and 2, MOB_V). Intriguingly, 20 (69%) of 29 MOB_H-family plasmids encoded some NAP homologs, and most of them were H-NS or HU (Table 7). The MOB_H family was composed of predominantly large conjugative plasmids, such as the IncHI1 group of plasmids, suggesting that HU may also contribute to plasmid transmission as does H-NS. Furthermore, 30 (23%) of 131 MOB_Q-family plasmids also contained some NAP gene homologs, and 15 (50%) of those carried Lrp gene homologs (Table 8). The MOB_Q family was composed of both mobilizable and conjugative plasmids, such as those of *Rhizobium* and *Agrobacterium*, implying that Lrp may also affect plasmid conjugation. In the other MOB families, plasmids containing NAP gene homologs were less than 10% (8%, MOB_C; 9%, MOB_F; 9%, MOB_P; and 8%, MOB_V). This phenomenon may also be related to the average size of the plasmids contained in each MOB family. MOB_H (220 kb) and MOB_Q (198 kb) were larger than MOB_C (78 kb), MOB_F (117 kb), MOB_P (87 kb), and MOB_V (149 kb). On the other hand, the average G+C content of all plasmids belonging to each MOB family was as follows: MOB_C (52%), MOB_F (52%), MOB_H (51%), MOB_P (53%), MOB_Q (54%), and MOB_V (46%). No relationship between the distribution of NAP gene homologs of each MOB family and the G+C content of plasmids was found.

3.6. Conclusions. We compared the distribution of NAP gene homologs among plasmids and plasmid features. Larger plasmids frequently had NAP gene homologs, possibly to maintain themselves and host cell fitness. Plasmids with NAP gene homologs also frequently carried relaxase genes. Although this may be related to their relatively larger sizes, together with the fact that NAPs affect global gene regulation, it is likely that NAPs contribute to plasmid transmission. Considering the fact that NAPs encoded on plasmids actually help the host cell to integrate newly acquired genes into host regulatory networks [14, 15], large plasmids with NAP gene homologs may be generally more beneficial not only for the host cell, but also for their own existence.

NAP homologs encoded on plasmids can interact with different types of NAPs encoded on the host chromosome and cooperatively regulate host transcriptional networks. Understanding these mechanisms in more detail will shed light on the meanings of the distributions of NAPs on plasmids and chromosomes. Comprehensive analysis of their binding sites in the host and plasmid genomes will help us to understand the relationships between G+C content and the presence of NAPs. Such information will explain

how bacteria adapt and evolve by acquiring foreign genes by HGT.

References

- [1] C. J. Dorman, "Chapter 2 nucleoid-associated proteins and bacterial physiology," *Advances in Applied Microbiology*, vol. 67, pp. 47–64, 2009.
- [2] S. C. Dillon and C. J. Dorman, "Bacterial nucleoid-associated proteins, nucleoid structure and gene expression," *Nature Reviews Microbiology*, vol. 8, no. 3, pp. 185–195, 2010.
- [3] M. D. Bradley, M. B. Beach, A. P. J. de Koning, T. S. Pratt, and R. Osuna, "Effects of Fis on *Escherichia coli* gene expression during different growth stages," *Microbiology*, vol. 153, no. 9, pp. 2922–2940, 2007.
- [4] W. W. Navarre, S. Porwollik, Y. Wang et al., "Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*," *Science*, vol. 313, no. 5784, pp. 236–238, 2006.
- [5] J. Oberto, S. Nabti, V. Jooste, H. Mignot, and J. Rouviere-Yaniv, "The HU regulon is composed of genes responding to anaerobiosis, acid stress, high osmolarity and SOS induction," *PLoS One*, vol. 4, no. 2, article e4367, 2009.
- [6] M. W. Mangan, S. Lucchini, V. Danino, T. Ó. Cróinín, J. C. D. Hinton, and C. J. Dorman, "The integration host factor (IHF) integrates stationary-phase and virulence gene expression in *Salmonella enterica* serovar Typhimurium," *Molecular Microbiology*, vol. 59, no. 6, pp. 1831–1847, 2006.
- [7] K. K. Swinger and P. A. Rice, "IHF and HU: flexible architects of bent DNA," *Current Opinion in Structural Biology*, vol. 14, no. 1, pp. 28–35, 2004.
- [8] B. K. Cho, C. L. Barrett, E. M. Knight, Y. S. Park, and B. Ø. Palsson, "Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 49, pp. 19462–19467, 2008.
- [9] L. D. Murphy, J. L. Rosner, S. B. Zimmerman, and D. Esposito, "Identification of two new proteins in spermidine nucleoids isolated from *Escherichia coli*," *Journal of Bacteriology*, vol. 181, no. 12, pp. 3842–3844, 1999.
- [10] C. Tendeng, O. A. Soutourina, A. Danchin, and P. N. Bertin, "MvaT proteins in *Pseudomonas* spp.: a novel class of H-NS-like proteins," *Microbiology*, vol. 149, no. 11, pp. 3047–3050, 2003.
- [11] L. S. Frost, R. Leplae, A. O. Summers, and A. Toussaint, "Mobile genetic elements: the agents of open source evolution," *Nature Reviews Microbiology*, vol. 3, no. 9, pp. 722–732, 2005.
- [12] C. M. Thomas and K. M. Nielsen, "Mechanisms of, and barriers to, horizontal gene transfer between bacteria," *Nature Reviews Microbiology*, vol. 3, no. 9, pp. 711–721, 2005.
- [13] A. Carattoli, "Plasmid-mediated antimicrobial resistance in *Salmonella enterica*," *Current Issues in Molecular Biology*, vol. 5, no. 4, pp. 113–122, 2003.
- [14] M. Doyle, M. Fookes, A. L. Ivens, M. W. Mangan, J. Wain, and C. J. Dorman, "An H-NS-like stealth protein aids horizontal DNA transmission in bacteria," *Science*, vol. 315, no. 5809, pp. 251–252, 2007.
- [15] C.-S. Yun, C. Suzuki, K. Naito et al., "Pmr, a histone-like protein H1 (H-NS) family protein encoded by the IncP-7 plasmid pCAR1, is a key global regulator that alters host function," *Journal of Bacteriology*, vol. 192, no. 18, pp. 4720–4731, 2010.

- [16] M. P. Garcillán-Barcia, M. V. Francia, and F. de la Cruz, "The diversity of conjugative relaxases and its application in plasmid classification," *FEMS Microbiology Reviews*, vol. 33, no. 3, pp. 657–687, 2009.
- [17] J. Wei, M. B. Goldberg, V. Burland et al., "Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T," *Infection and Immunity*, vol. 71, no. 5, pp. 2775–2786, 2003.
- [18] C. K. Sherburne, T. D. Lawley, M. W. Gilmour et al., "The complete DNA sequence and analysis of R27, a large IncHI plasmid from *Salmonella typhi* that is temperature sensitive for transfer," *Nucleic Acids Research*, vol. 28, no. 10, pp. 2177–2186, 2000.
- [19] J. Wain, L. T. D. Nga, C. Kidgell et al., "Molecular analysis of incHI1 antimicrobial resistance plasmids from *Salmonella* serovar Typhi strains associated with typhoid fever," *Antimicrobial Agents and Chemotherapy*, vol. 47, no. 9, pp. 2732–2739, 2003.
- [20] A. Tett, A. J. Spiers, L. C. Crossman et al., "Sequence-based analysis of pQBR103; a representative of a unique, transfer-proficient mega plasmid resident in the microbial community of sugar beet," *ISME Journal*, vol. 1, no. 4, pp. 331–340, 2007.
- [21] K. Maeda, H. Nojiri, M. Shintani, T. Yoshida, H. Habe, and T. Omori, "Complete nucleotide sequence of carbazole/dioxin-degrading plasmid pCAR1 in *Pseudomonas resinovorans* strain CA10 indicates its mosaicity and the presence of large catabolic transposon Tn4676," *Journal of Molecular Biology*, vol. 326, no. 1, pp. 21–33, 2003.
- [22] Y. Takahashi, M. Shintani, H. Yamane, and H. Nojiri, "The complete nucleotide sequence of pCAR2: pCAR2 and pCAR1 were structurally identical incP-7 carbazole degradative plasmids," *Bioscience, Biotechnology and Biochemistry*, vol. 73, no. 3, pp. 744–746, 2009.
- [23] P. Deighan, C. Beloin, and C. J. Dorman, "Three-way interactions among the Sfh, StpA and H-NS nucleoid-structuring proteins of *Shigella flexneri* 2a strain 2457T," *Molecular Microbiology*, vol. 48, no. 5, pp. 1401–1416, 2003.
- [24] S. C. Dillon, A. D. S. Cameron, K. Hokamp, S. Lucchini, J. C. D. Hinton, and C. J. Dorman, "Genome-wide analysis of the H-NS and Sfh regulatory networks in *Salmonella Typhimurium* identifies a plasmid-encoded transcription silencing mechanism," *Molecular Microbiology*, vol. 76, no. 5, pp. 1250–1265, 2010.
- [25] R. T. Dame, M. C. Noom, and G. J. L. Wuite, "Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation," *Nature*, vol. 444, no. 7117, pp. 387–390, 2006.
- [26] C. Madrid, C. Balsalobre, J. García, and A. Juárez, "The novel Hha/YmoA family of nucleoid-associated proteins: use of structural mimicry to modulate the activity of the H-NS family of proteins," *Molecular Microbiology*, vol. 63, no. 1, pp. 7–14, 2007.
- [27] S. de los Rios and J. J. Perona, "Structure of the *Escherichia coli* leucine-responsive regulatory protein Lrp reveals a novel octameric assembly," *Journal of Molecular Biology*, vol. 366, no. 5, pp. 1589–1602, 2007.