Supplementary File

Table S1. Information about genomes analyzed in this study.

Species	Strain	Serogroup	ST	CC	Accession
N. meningitidis	Z2491	A	4	4	AL157959
N. meningitidis	WUE 2594	A	5	5	FR774048
N. meningitidis	G2136	B	8	8	CP002419
N. meningitidis	961-5945	B	153	8	AEQK00000000
N. meningitidis	K1207	C	11	11	ADWM00000000
N. meningitidis	S0108	C	11	11	ADWN00000000
N. meningitidis	FAM18	C	11	11	AM421808
N. meningitidis	ES14902	B	11	11	AEQI00000000
N. meningitidis	M6190	B	1988	11	AEQF00000000
N. meningitidis	NEM8013	C	177	18	FM999788
N. meningitidis	Nm6938	W135	22	22	AGBQ00000000
N. meningitidis	Nm2732	W135	22	22	AGBP00000000
N. meningitidis	Nm6756	Y	23	23	AGBR00000000
N. meningitidis	Nm8663	Y	23	23	AGBS00000000
N. meningitidis	NS44	Y	23	23	AEPJ00000000
N. meningitidis	H44/76	B	32	32	CP002420
N. meningitidis	CU385	B	33	32	AEQJ00000000
N. meningitidis	MC58	B	74	32	AE002098
N. meningitidis	M01-240149	B	41	41/44	CP002421
N. meningitidis	NZ-05/33	B	42	41/44	CP002424
N. meningitidis	M0579	B	43	41/44	AEQH00000000
N. meningitidis	α710	B	136	41/44	CP001561
N. meningitidis	OX99-30304	B	8094ξ	41/44	AEQE00000000
N. meningitidis	α14	NG-cnl	53	53	AM889136
N. meningitidis	Nm9418	C	60	60	ARYG00000000*
N. meningitidis	Nm10259	C	60	60	ARYF00000000*
N. meningitidis	Nm8187	Y	3923	167	AGBN00000000
N. meningitidis	Nm3127	Y	3980	167	AGBO00000000
N. meningitidis	N1568	NG†	751	181	AEQD00000000
N. meningitidis	M01-240355	B	213	213	CP002422
N. meningitidis	M04-240196	B	269	269	CP002423
N. meningitidis	M01-240013	B	1159φ	269	AEQL00000000
N. meningitidis	M13399	B	2976	269	AEQG00000000
N. meningitidis	Nm1140	NG-cnl	1136	1136	AGBT00000000
N. meningitidis	053442	C	4821	4821	CP000381
N. meningitidis	ATCC 13091	B	7355	-	AEEF00000000
N. gonorrhoeae	TCDC-NG08107	-	7363	-	CP002440
N. gonorrhoeae	NCCP11945	-	1901	-	CP001050
N. gonorrhoeae	FA1090	-	1899	-	AE004969
N. lactamica	020-06	-	640	640	HQ332785

*Sequenced in this study

†NG: Non-groupable

ξ Inferred directly from the genomic data, but reported as ST44 in Budroni et al. 2011

φ Inferred directly from the genomic data, but reported as ST275 in Budroni et al. 2011

Table S2. Phylogenetic incongruence between individual MLST locus and the concatenated sequence of the seven MLST loci.

	P-values	
Name	Individual sequence	Concatenated sequence
abcZ	4.5×10^{-08}	5.0×10^{-53}
adk	3.0×10^{-33}	1.1×10^{-08}
aroE	1.0×10^{-03}	1.5×10^{-06}
fumC	2.5×10^{-08}	1.0×10^{-54}
gdh	3.5×10^{-07}	5.2×10^{-55}
pdhC	5.0×10^{-27}	3.0×10^{-18}
pgm	5.0×10^{-09}	5.0×10^{-08}

The AU test was performed 10 times between each individual phylogeny and the concatenated-sequence phylogeny using both the individual sequence and the concatenated sequence of the seven MLST loci, and the median P-value was presented.

Table S3. Variable genes only present in Nm10259 or Nm9418.

Genome	Gene Name[†]	Panel in Fig.2	COG	Protein Function
Nm10259	*cas1*	A	COG1518L	hypothetical protein
	cas2	A	COG3512S	hypothetical protein
	csn1	A	COG3513S	hypothetical protein
	NMA1079[(2)]	B	-	hypothetical protein
	NMA1080	B	-	hypothetical protein
	NMA1081	B	-	hypothetical protein
	NMC0806	B	-	integral membrane protein
	NMB0432	C	COG0730R	hypothetical protein
	NMAA_1384	D	-	putative lipoprotein
	NMV_2010	E	-	hypothetical protein
	NMO0338	E	-	hypothetical protein
	NMC0889	E	-	integral membrane protein
	NMC0890	E	-	integral membrane protein
	nadA	F	COG0379H	adhesin/invasin autotransportor
	NMB1202	H	-	hypothetical protein
Nm9418	NMB1828	A	COG2819R	hypothetical protein
	NMB1829	A	COG4773P	TonB-dependent receptor
	NMB0856	B	-	hypothetical protein
	NMB0857	B	-	hypothetical protein
	NMB0858	B	-	hypothetical protein
	NMB0859	B	-	hypothetical protein
	NMB0860	B	-	hypothetical protein
	NMB0861	B	-	hypothetical protein
	NMB0863	B	-	hypothetical protein
	NMCC_0822	B	-	hypothetical protein
	NMC1730	C	COG5505S	integral membrane protein
	ddpX	C	COG2173M	D-alanyl-d-alanine dipeptidase
	NMB1665	D	COG3744S	hypothetical protein
	NMB1666	D	COG4118D	hypothetical protein
	hpuB	D	COG1629P	hemoglobin utilization protein
	NMO0341	E	-	hypothetical protein
	NMO0342	E	-	hypothetical protein
	N1568_0676	G	-	hypothetical protein
	NMB0644	I	-	hypothetical protein
	NMB0645	I	-	hypothetical protein

[†]Gene names were following other well annotated *Nm* genomes.

[(2)]There are two copies of NMB1079 present in Nm10259.

Table S4. Nonvertical genes in ST60 that have been previously identified as putative virulence genes.

Gene Name	COG	Protein Function
ctrB	COG3524M	capsule polysaccharide export inner-membrane protein
tbp	COG1629P	transferrin-binding protein B
kdsB	COG1212M	3-deoxy-manno-octulosonate cytidylyltransferase
NMB1494	COG2879S	hypothetical protein NMB1494
lgtF	COG0463M	beta-1,4-glucosyltransferase
rfaK	COG0438M	alpha 1,2 N-acetylglucosamine transferase
gcp	COG0533O	DNA-binding/iron metalloprotein/AP endonuclease
pglC	COG0399M	pilin glycosylation protein
pglD	COG1086MG	pilin glycosylation protein
NMB1845	COG0526OC	thioredoxin
fetB	COG4607P	iron ABC transporter substrate-binding protein
porB	COG3203M	major outer membrane protein PIB

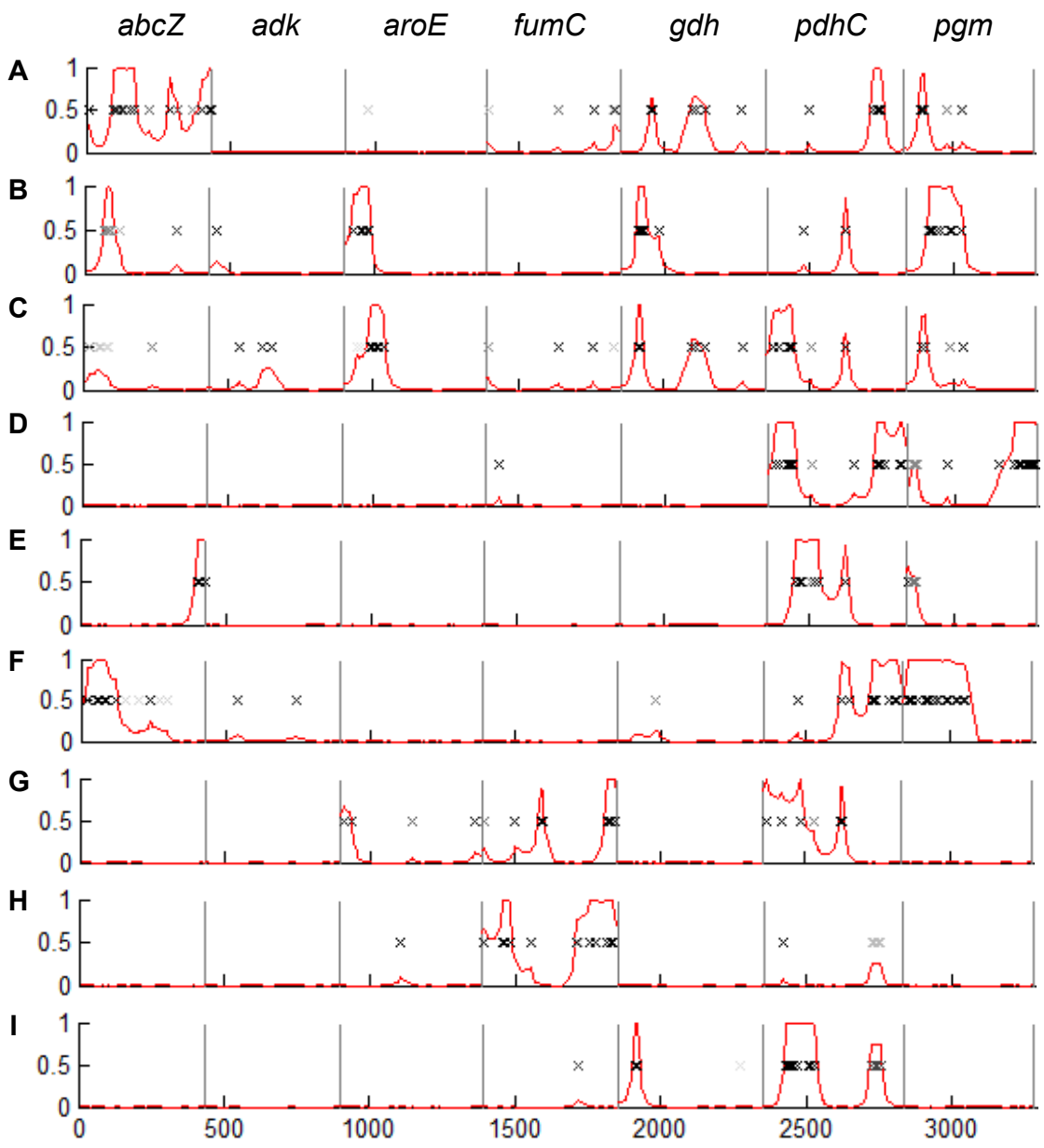

Figure S1. Recombination events identified by ClonalFrame, using the seven MLST loci, on the branches leading to individual CCs as shown in Figure 1C. The columns represented the seven MLST loci. "×" indicated substitutions inferred to have occurred in the respective branches. Red lines represented the probability for each nucleotide to have been imported by means of recombination. Values at the bottom represented nucleotide positions in the concatenated sequence of the seven MLST loci.

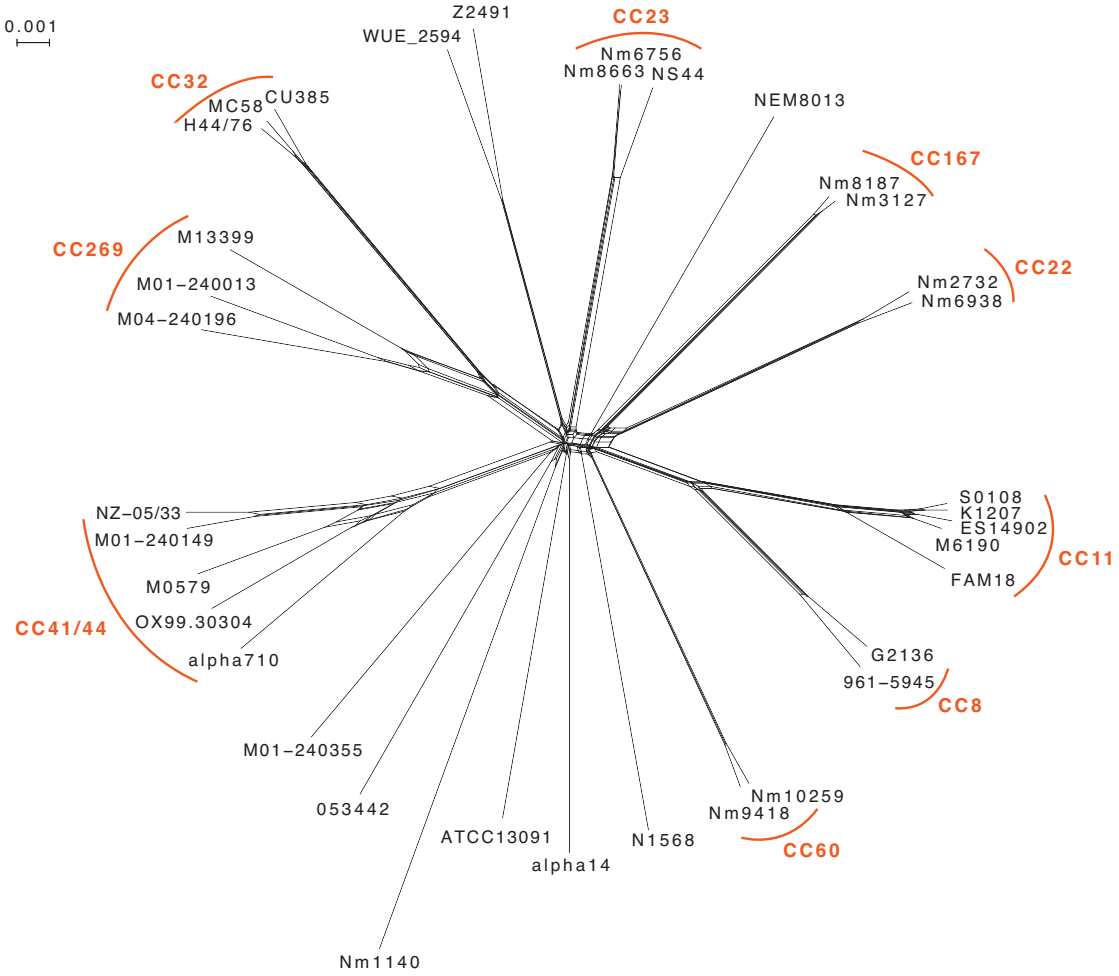

Figure S2. Neighbour-net networks of the concatenated sequences of the 1090 single-copy genes present in all studied *Nm* genomes. Nucleotide distance was corrected using the F84 substitution matrix.

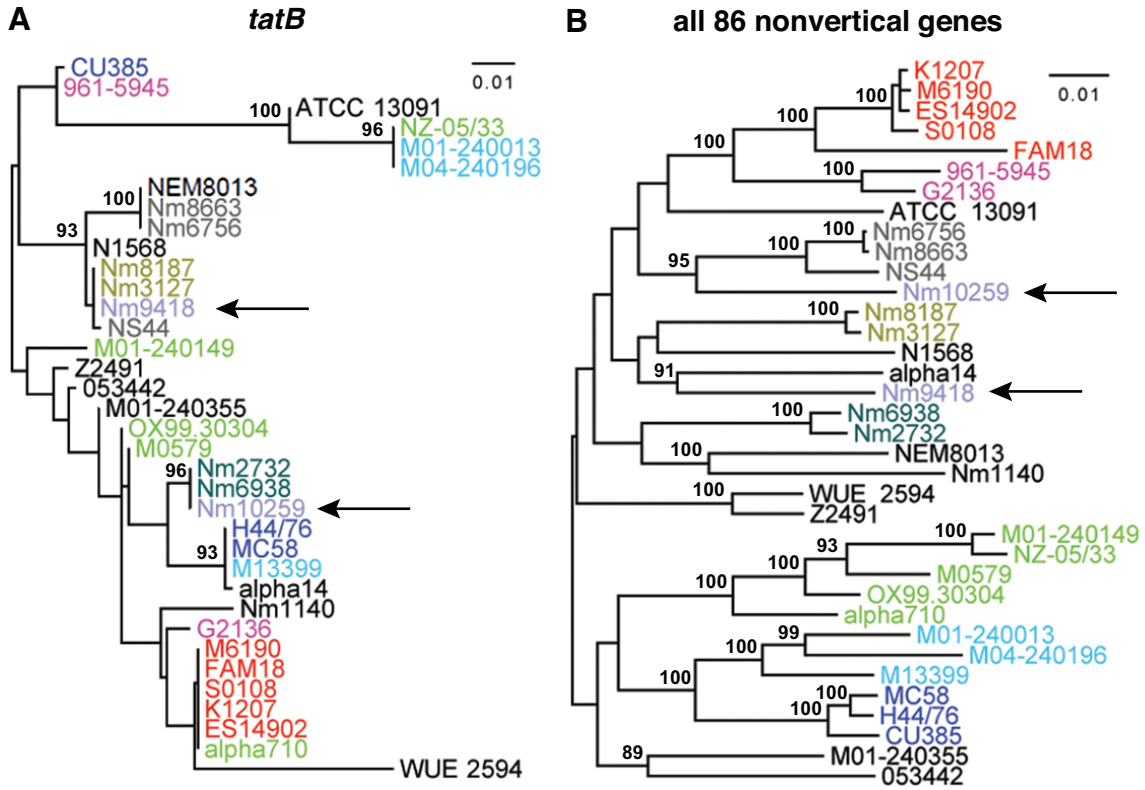

Figure S3. Maximum likelihood trees of one representative nonvertical gene in ST60, *tatB* (A) and the concatenated sequences of all the 86 nonvertical genes presented in all *Nm* strains (B). The two ST60 strains (by arrows) were well separated on both phylogenies. Strains are color-coded as in Figure 1. 100 bootstrap iterations were performed for each phylogeny and bootstrap values >85 are shown.

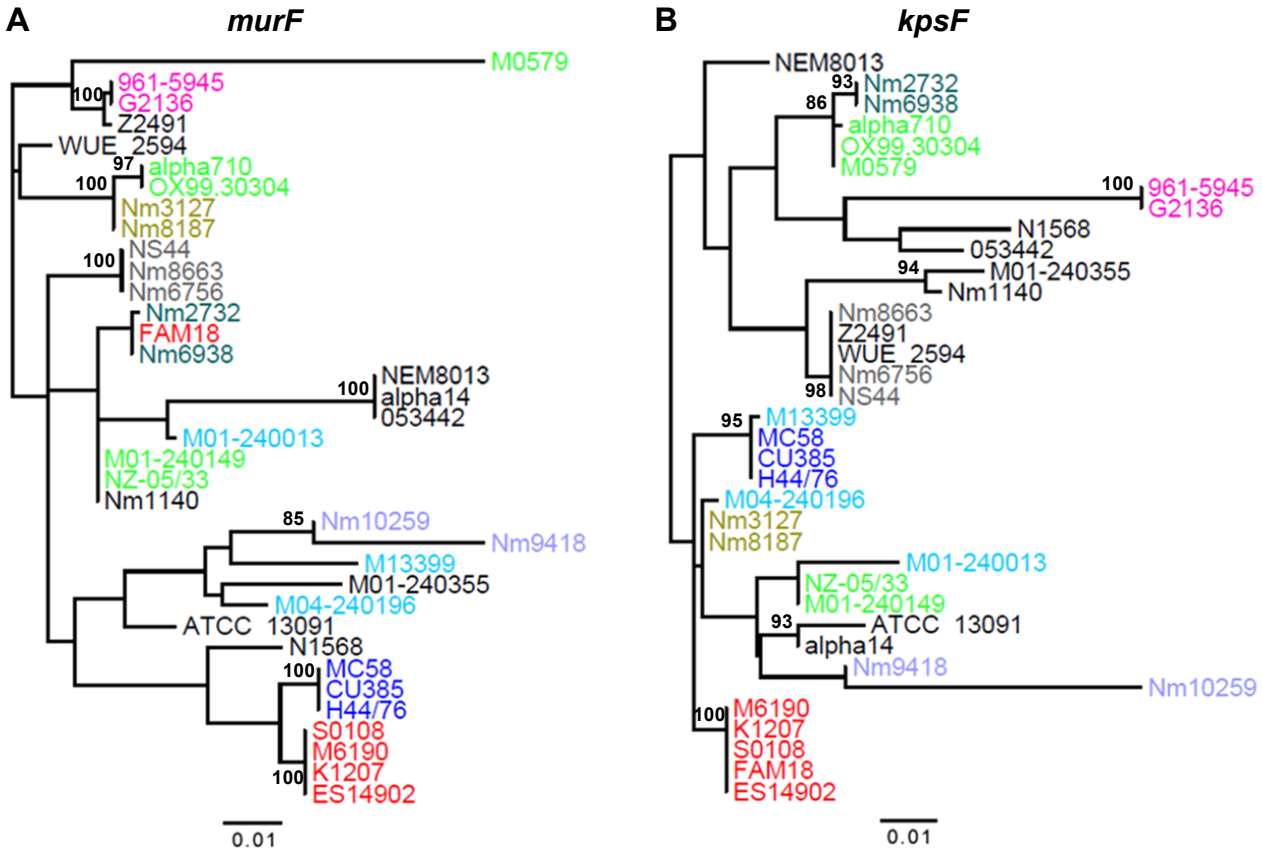

Figure S4. Maximum likelihood trees of two nonvertical ST60 genes, *murF* (A) and *kpsF* (B). On both phylogenies, Nm10259 and Nm9418 were clustered together, but there was a significantly elongated branch leading to Nm10259 or Nm9418. Strains are color-coded as in Figure 1. 100 bootstrap iterations were performed for each phylogeny and bootstrap values >85 are shown.

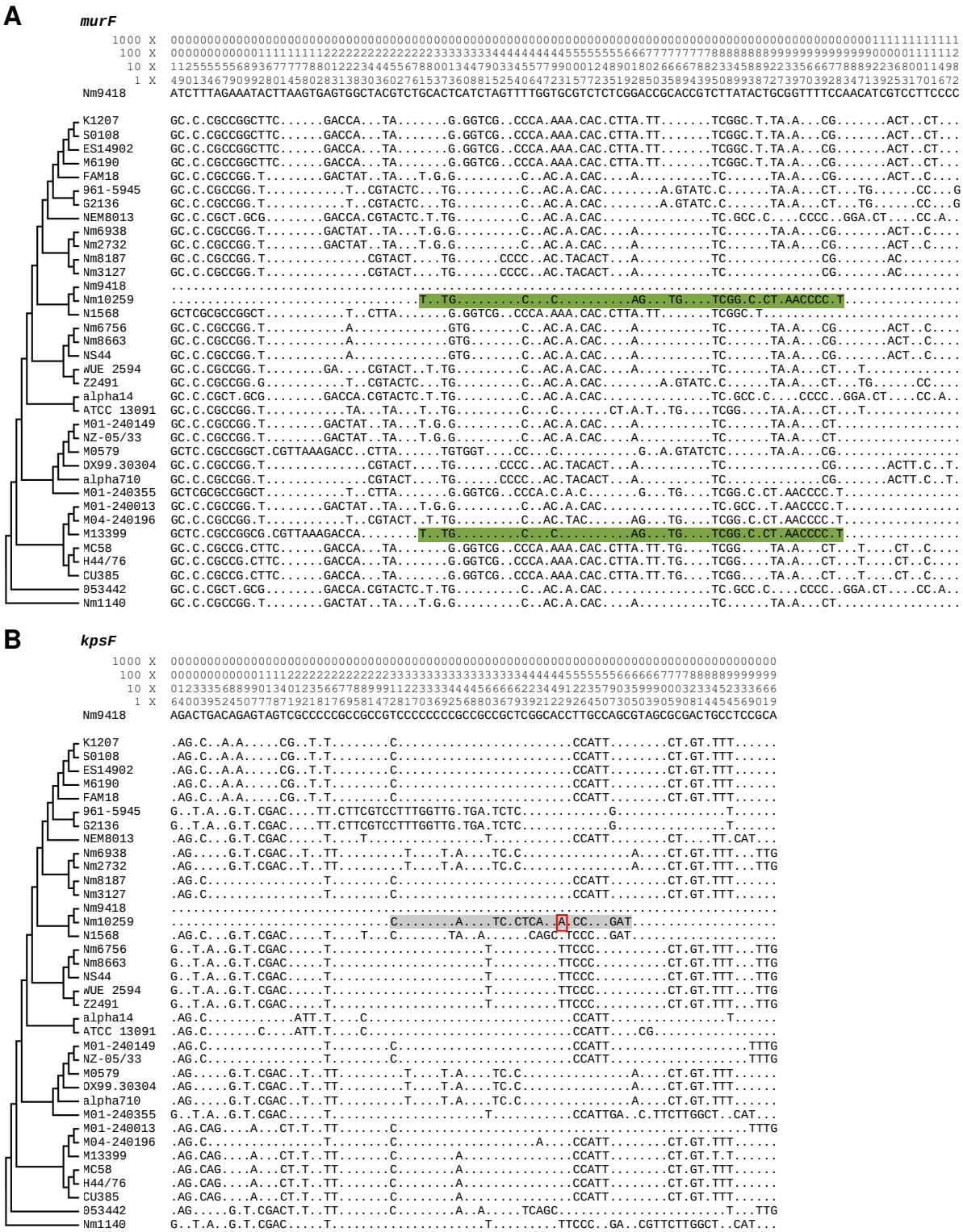

Figure S5. Sequence alignments of the *murF* (A) and *kpsF* (B) genes. To highlight the difference between Nm10259 and Nm9418, sequences in Nm9418 were used as the reference. Only informative sites were shown with coordinates at the top. The topology on the left was the same as Figure 1A. Regions in Nm10259 different from Nm9418 are highlighted, and the nucleotide position 492 of the *kpsF* gene in Nm10259 is boxed, as it was not found in any other *Nm* genomes.

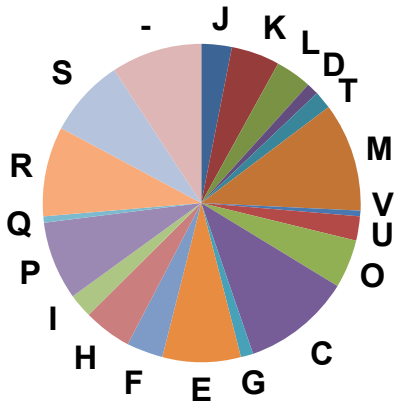

Figure S6. COG functional categories of the 152 nonvertical genes in ST-60. The major functional categories are information storage and processing (COG categories J, K and L), cellular processes and signaling (COG categories D, V, T, M, U and O), metabolism (COG categories C, G, E, F, H, I, P and Q), and poorly characterized (COG categories R and S). '-' refers to no information in COG.

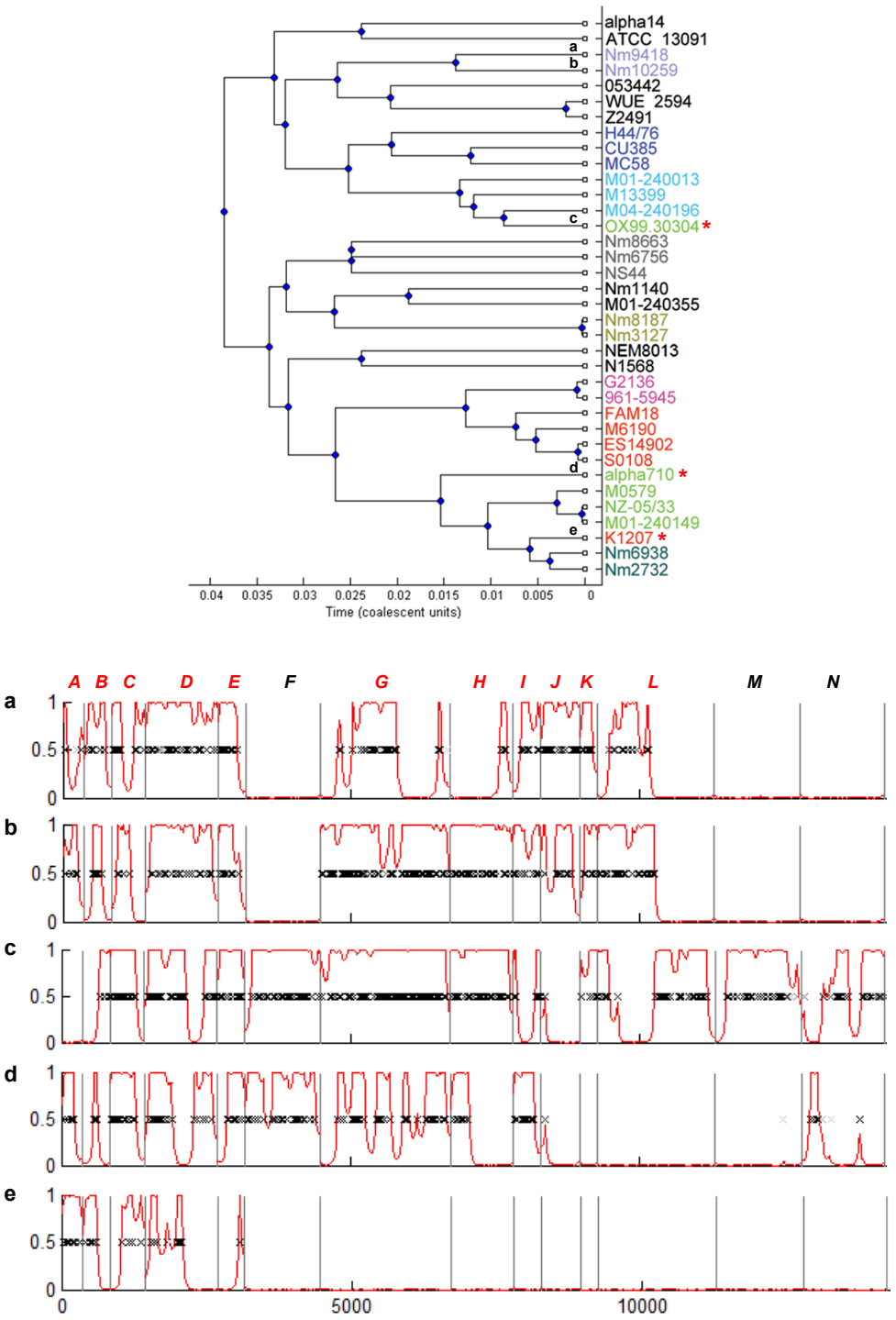

Figure S7. Recombination events in the *nuo* gene cluster identified by ClonalFrame. A majority rule consensus tree was constructed by ClonalFrame using the concatenated sequences of the 14 *nuo* genes (upper panel). Strains not grouped with other strains of the belonged CC were labeled with a red asterisk (*). The columns represented the 14 genes in the *nuo* gene locus. In each row, "×" indicated substitutions inferred to have occurred in the respective branches as labeled on the tree. Red lines represented the probability for each nucleotide to have been imported by means of recombination. Values at the bottom represented nucleotide positions in the concatenated sequence of the 14 *nuo* genes.

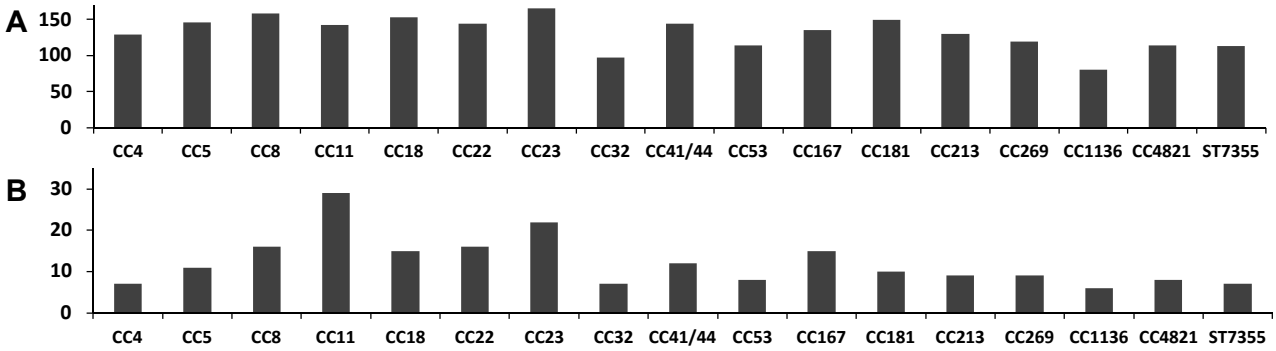

Figure S8. Distribution of recombination partners in CCs of identical genes (A) and nonvertical genes (B) between Nm10259 and Nm9418. Unlike in Figure 4, the recombination partners were defined by the nearest neighbors of the Nm10259 genes and Nm9418 genes. The total number of recombination partners in each CC is presented.

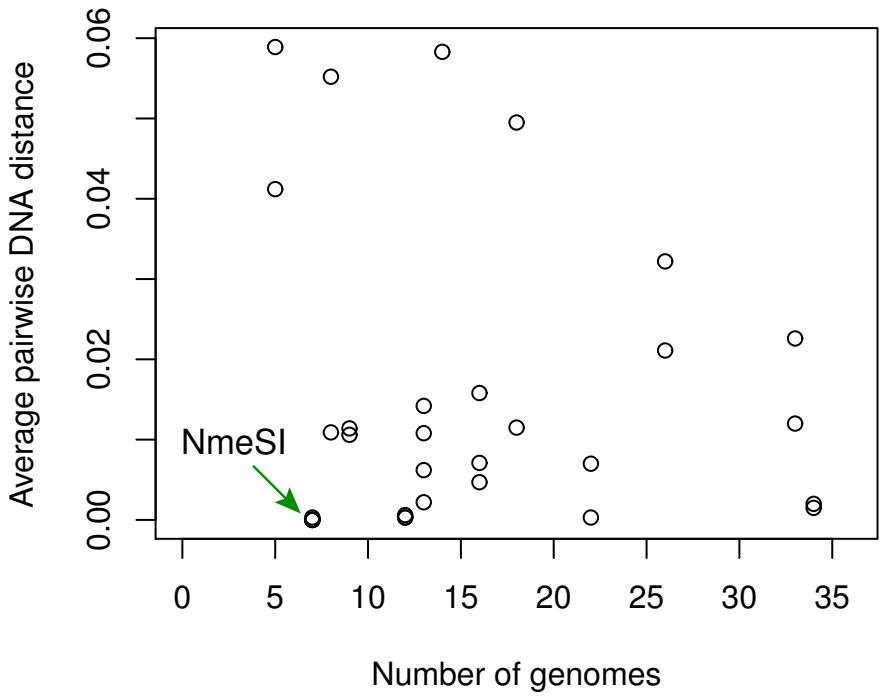

Figure S9. No significant correlation between sequence divergence and distribution of the RMS genes. Spearman's rank correlation test was performed between the average pairwise DNA distance of each RMS gene and the number of present genomes (P=0.14). All RMS genes present in at least five genomes were analyzed (the same dataset as in the AU test in Table 2). As labeled, all the NmeSI genes were identical at the sequence level.

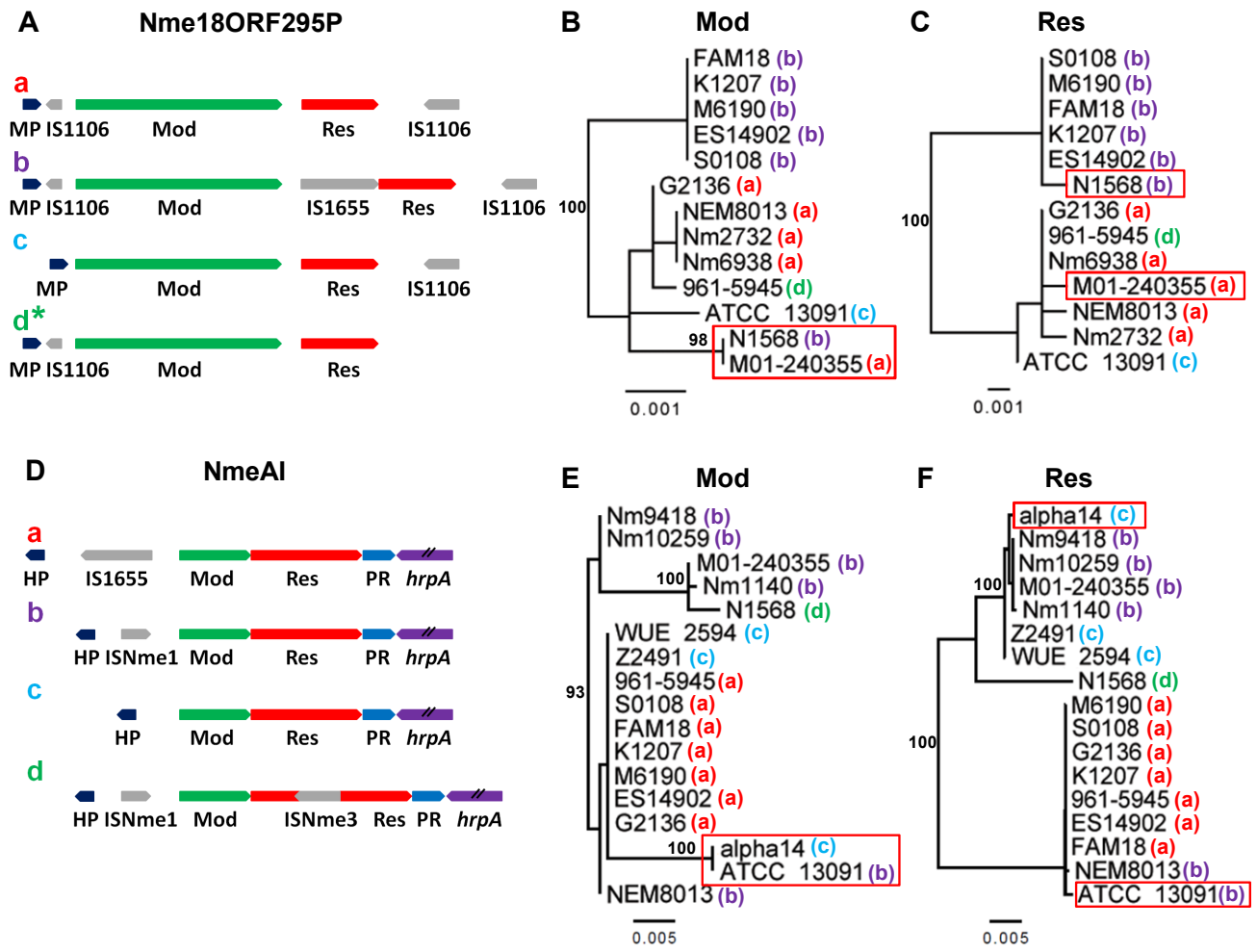

Figure S10. Plylogentic incongruence of the RMS genes in Nme18ORF295P (A-C) and NmeAI (D-F). All possible gene organizations of each RMS were shown on the left, as A for Nme18ORF295P and D for NmeAI. Genes were drawn as arrows, and empty spaces between the arrows represented intergenic regions. Some abbreviations were used: Mod, modification gene; Res, restriction gene; PR, patch repair gene; MP, probable inner membrane protein; HP, conserved hypothetical protein. In the 961-5945 genome (shown as d*), the gene after the restriction gene could not be identified due to the end of the contig. Maximum likelihood trees constructed for both the modification and restriction genes of Nme18ORF295P (B and C) and NmeAI (E and F). 100 bootstrap iterations were performed in each phylogenetic construction, and bootstrap values >85 are shown. Strains that showed strikingly different positions between the modification and restriction genes are boxed. The letters in parenthesis beside the strain names are the gene organization in each strain.

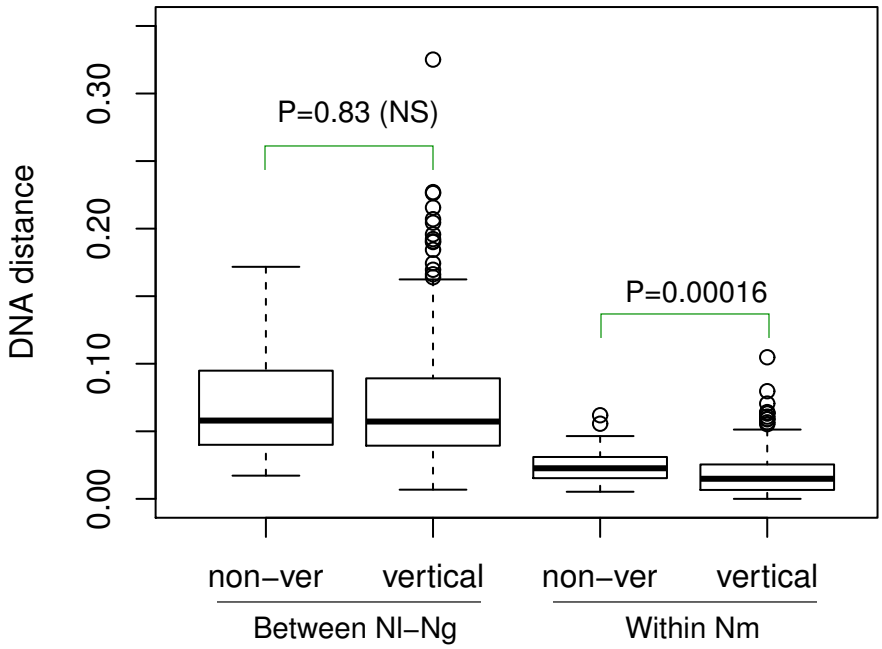

Figure S11. No significant acceleration of DNA distance between the *Ng* and *Nl* (*N. lactamica*) homologs of the nonvertical ST60 genes. All genes analyzed here were present in all 36 *Nm* genomes, all three *Ng* genomes and the *Nl* genome as well. The distance between *Nl-Ng* was the average between *Nl* and each of the three *Ng* genomes, and the distance within *Nm* was the average of all possible pair-wise distance among the 36 *Nm* genomes. P-values were calculated using the Wilcoxon signed-rank test.