

Supplementary Material

Comparative genome analysis of *Lactococcus lactis* indicates niche adaptation and resolves genotype/phenotype disparity

Michiel Wels^{1,2}, Roland Siezen^{2,3,4}, Sacha van Hijum^{1,2,3}, William J. Kelly⁵, Herwig Bachmann^{1,2,6*}

¹ NIZO B.V., Ede, The Netherlands

² TI Food and Nutrition, Wageningen, The Netherlands

³ Centre for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences,

Radboud UMC, Nijmegen, The Netherlands.

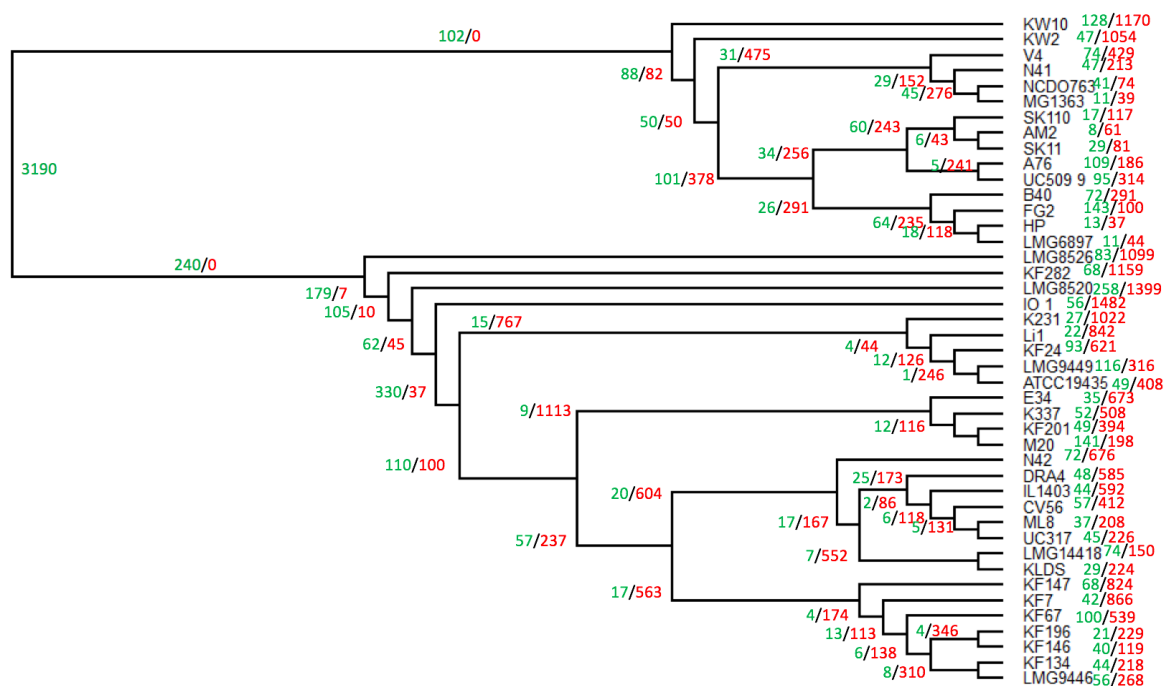
⁴ Microbial Bioinformatics, Ede, The Netherlands.

⁵ Ashhurst, New Zealand

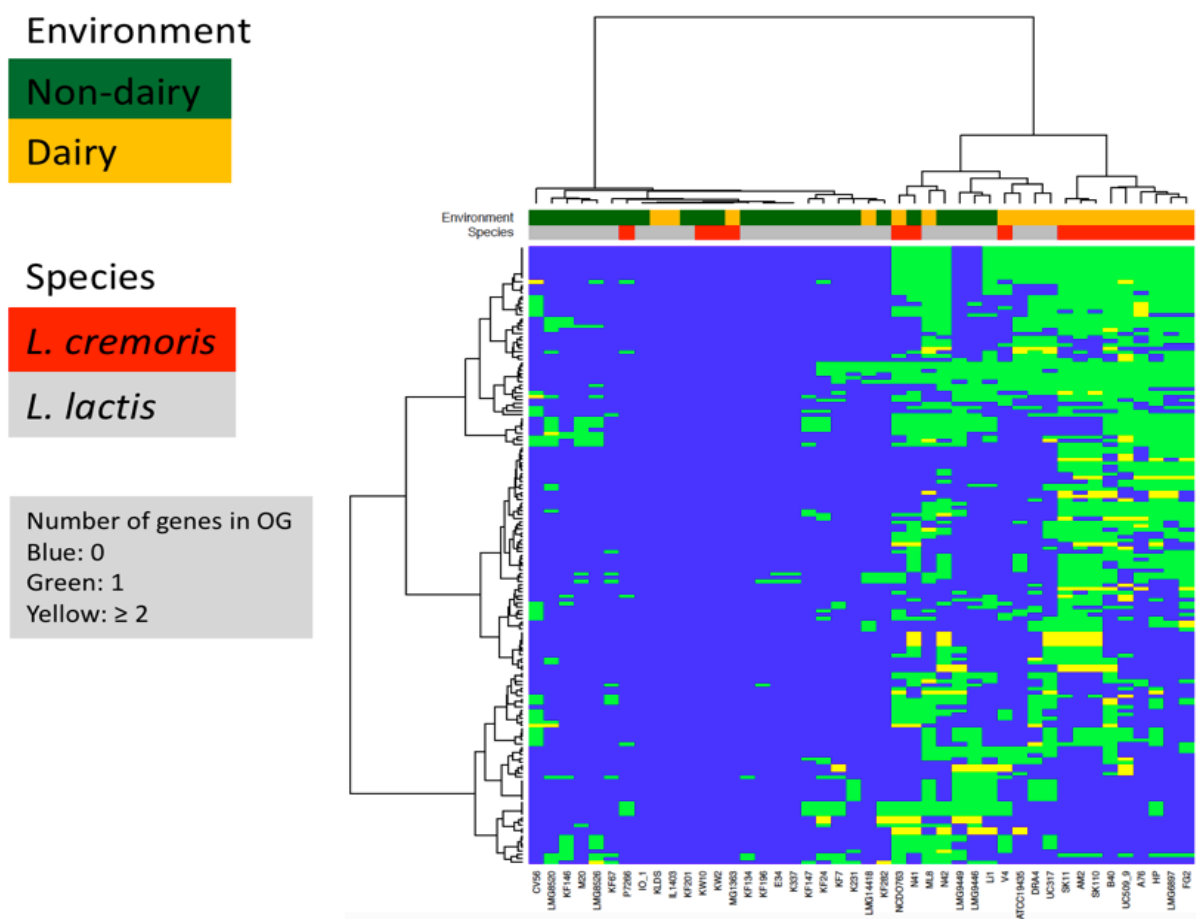
⁶ Systems Bioinformatics, VU University, Amsterdam, The Netherlands

*Correspondence: Herwig Bachmann, NIZO B.V., P.O. Box 20, 6710BA Ede, The Netherlands; M: +31 318 659652; E: Herwig.Bachmann@nizo.com

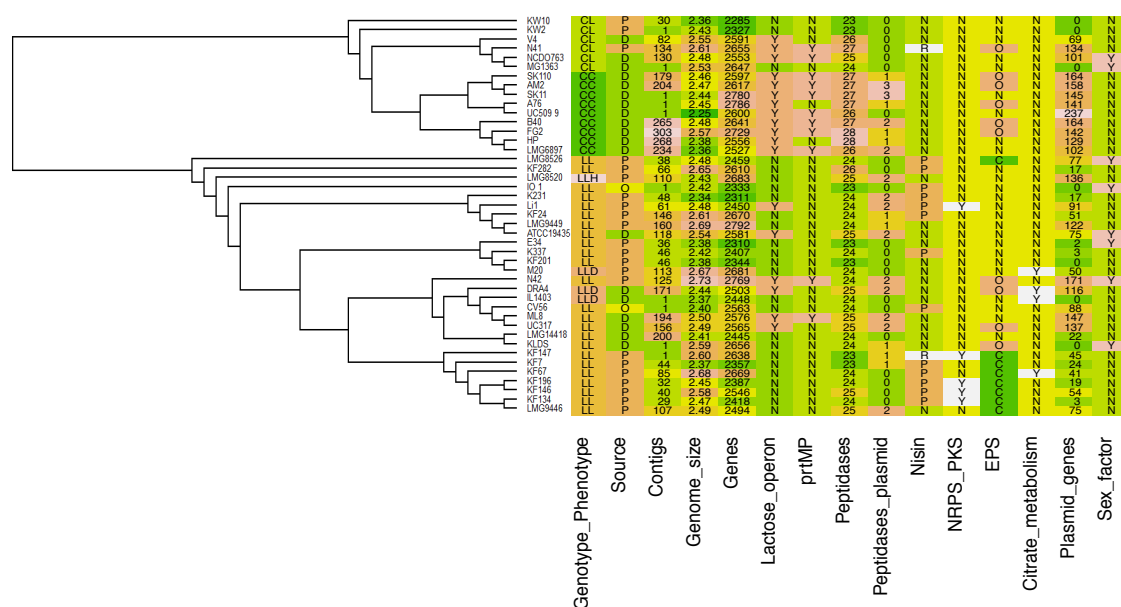
Supplemental Figures



Supplementary Figure 1: Phylogenetic tree based on the single nucleotide polymorphisms (SNPs) in the *L. lactis* core genes. Gene gain and gene loss at the individual nodes are given in green and red respectively.



Supplementary Figure 2: Heatmap of orthologous groups (left dendrogram) identified as plasmid encoded genes. Isolation environment and species information are indicated at the top of the figure (LMG8520 is the type strain *L. lactis* ssp. *hordniae*). Each row represents one OG. Each column represents one strain. For details see Supplementary Table 1/Sheet 5.



Supplementary Figure 3: Phylogenetic tree and its relation to a selected set of phenotypic properties. Abbreviations: Genotype-Phenotype CL=*cremoris-lactis*, CC=*cremoris-cremoris*, LL= *lactis-lactis*, LLH=*lactis-lactis-hordniae*, LLD=*lactis-lactis-diacetylactis*; Source D=dairy, P=plant material, O=other; Nisin P=producer, R=resistant, N=contains no genes from the nisin cluster; EPS C=conserved on chromosome, O=other position on chromosome or plasmid-located, N= contains no EPS genes; in all other columns y=yes/present, n=none/absent. The branch of the dendrogram on the length does not reflect the phylogenetic distance (see Figure 2 for phylogenetic distance). Colors are only for easier visualization of differences within columns.

```

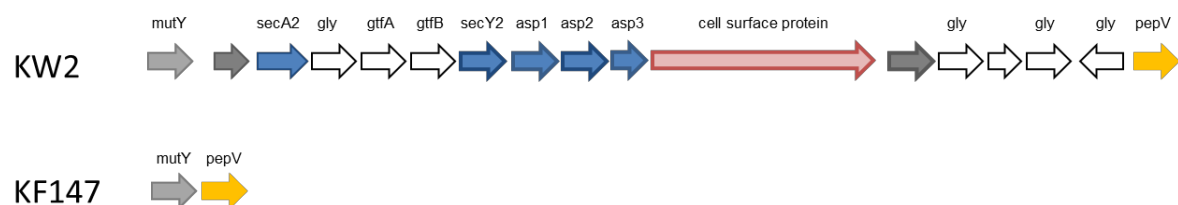
-SqbB: MNEFKKIGLENEFNLI-EDTEQLE-LMSGAGWTGVAKSITSCSLLSIGLGNDGWVCTWTAECQATCR
-LlbB: MKNLKNSGLQEDYLSLI-EETKDLE-LISGAGWGGVAKSITSCSILSITLGNNGWICTWTAECQKCCK
-SmbB: MKEIQKAGLQEE-LSILMDDANNLEQLTAGIG-TTVVNS-----TFSIVLGNKGYICTVTVECMRNCSK

-SqbA: MEHNNLLINISEEEMQKNMVLLIEDEDQELSGASTPGCAWAAVSAISTVSALFQITTACCTTRCYHP
-LlbA: MKNNNLLDVTDEVMKKNIVTLIEDEDLNI SGAGTPA-GITVV---ITIGLLQVTTACTSRCWHP
-SmbA: MKSNLLKINNVTEMEKNMVTLLIKDEDEMELAGGSTPACCAIGVV---GITVAVTGISETACTSRCINK

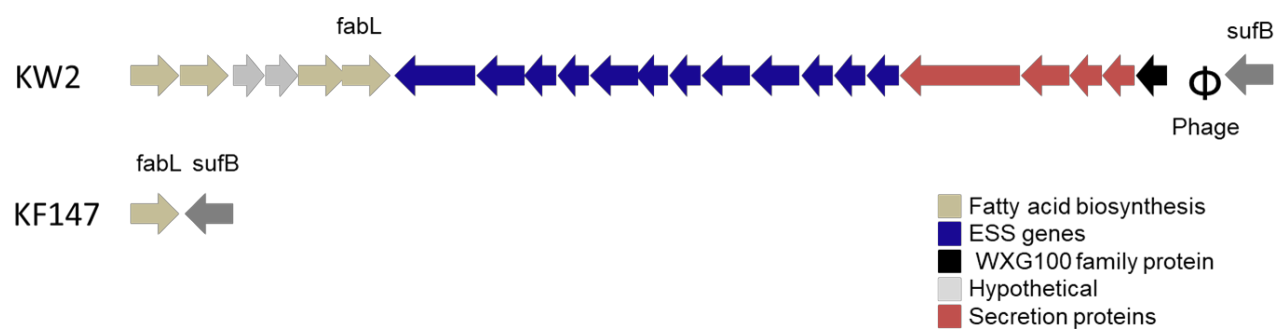
```

Supplementary Figure 4: Sequence alignment of the lantibiotic precursors A and B of *L. lactis* (Ll), *S. mutans* (Sm), and *S. gallolyticus*. Identical residues are colored green, while Cys residues probably involved in cyclization are colored yellow.

A. SecA2 system



B. ESX system



Supplementary Figure 5: Accessory protein secretion systems found in *Lactococcus lactis* subsp. *cremoris* KW2 (Panel A) and ESX system (Panel B).

Supplementary Tables

(if only legends are given the Tables are supplied as separate Excel files)

Supplementary Table 1: OG tables including all analyzed OGs (Sheet 1), subspecies specific OGs that only occur in either *ssp. lactis* or *ssp. cremoris* (Sheet 2) and detailed lists of OGs given in the heatmaps of Figure 3 (Sheet 3), Figure 4 (Sheet 4) and Supplementary Figure 2 (Sheet 5). See Excel file.

The alignments of all OGs are deposited at zenodo.com under DOI: 10.5281/zenodo.1471674

<https://doi.org/10.5281/zenodo.1471674>

Supplementary Table 2: Pseudogenes occurring in more than 8 out of 43 strains (Sheet1) and an overrepresentation analysis (Sheet 2). See Excel file.

Supplementary Table 3: Glycoside hydrolase gene cluster found in *L. lactis* strains.

+ gene present; - pseudogene; no-symbol gene absent;

Glycosylhydrolase OG	GH115 OG_2205	GH1 OG_34	GH85, CBM32 OG_1824	GH13 OG_24	GH20 OG_1863	GH38 OG_1884	GH125 OG_1883	GH92 OG_1881	GH43 OG_3849	GH43 OG_871	GH11 OG_2465
ssp. cremoris											
KW2	-	+	-	+	+	+	+	+		+	+
KW10	+	+	+	+	+	+	+	+		+	+
V4	+	+	+	+	+	+	+	+		+	+
N41	-	+	+	+	-	-	+	+	-	+	+
MG1363	-	-									
NCDO763	-	-									
A76	+	+									+
UC509.9	+										-
SK11	-	+									+
SK110	-	+									+
AM2	-	+									+
B40	+										-
HP	+										+
FG2	+										-
LMG6897	+										+
ssp. lactis											
KF147	+	+	+	+	+	+	+	+	+	+	
M20		+	+	+	+	+	+	+		+	
KF201		+	+	+	-	+	+	+		+	
K337	+	+		+	-	+	+	+		+	
KF67		+	+	+	+	+	+	+		+	
KF134		+	+	+	+	+	+	+		+	
KF146		+	+	+	+	+	+	+		+	
KF196		+	+	+	+	+	+	+		+	
E34		+	+	+	+	+	+	+		+	
KF7		+	-	+	+	+	+	+		+	
LMG9446		+	+	+	-	+	+	+		+	
LMG8526		+	+	+	+	+	+	+	+	+	
KF282	+	+	+	+	+	+	+	+	+	+	
LMG8520		+	-	-	-	+	-	+		-	
N42		+	+	+	+	+	+	+		+	
ATCC19435		+	+	+	+	+	+	+		+	
K231		+	+	+	+	+	+	+		+	
Li-1		+	+	+	+	+	+	+		+	
IO-1		+	+	+	+	+	+	+		+	
KF24		+	+	+	+	+	+	+		+	
LMG9447		+	+	+	-	+	+	+		+	
UC317		+	+	+	+	+	+	+		+	
ML8		+	+	+	+	+	+	+		+	
CV56		+	+	+	+	+	+	+		+	
IL1403		+		+	+	+	+	+		+	
DRA4		+		+	+	+	+	+		+	
LMG14418		+		+	+	+	+	+		+	

Supplementary Table 4: Genes involved in acid stress. Symbols: +, intact gene; -, gene missing; P, pseudogene; T/+, transposase inserted adjacent to intact gene; P/T, transposase inserted adjacent to a pseudogene.

OG		KW2	KW10	V4	N41	NCD0763	MG1363	AM2	SK11	SK110	A76	UC509.9	B40	FG2	HP	LMG6897
Genotype		cremoris							cremoris							
Phenotype		lactis							cremoris							
Arginine deiminase system																
argR	OG_1521	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
argS	OG_1099	+	+	+	+	+	+	+	+	+	+	T/+	+	+	+	+
arcA	OG_1100	+	+	+	+	+	+	P	P	P	+	+	P	P	P	P
arcB	OG_1101	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
arcD1	OG_11	+	+	+	+	+	+	+	+	+	+	+	P	P	P	P
arcC1	OG_1761	+	+	+	+	+	+	+	+	+	+	+	P	P	P	P
arcC2	OG_1522	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
arcT	OG_162	+	+	+	+	+	+	+	+	+	P/T	+	+	+	+	+
arcD2	OG_11	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Agmatine deiminase system																
aguR	OG_1903	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
aguB	OG_1902	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
aguD	OG_1901	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
aguA	OG_1922	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
aguC	OG_1921	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
Glutamate decarboxylase system																
gadB	OG_1672	+	-	P	+	+	+	P	P	P	P	P	P	P	P	P
gadC	OG_1701	+	-	+	+	+	+	+	+	+	+	+	+	P	P	P
gadR	OG_1614	+	-	+	+	+	+	+	+	+	+	P	+	+	+	+
Malolactic system																
mleR	OG_1447	+	+	+	P	+	+	+	+	+	P	P	+	P	P	P
mleS	OG_94	+	+	+	+	+	+	P	P	P	P	+	P	P	P	P
mleP	OG_944	+	+	+	+	+	+	+	+	+	P	P	+	+	+	+

Supplementary Table 5: Occurrence of genes predicted to be involved in maltose and trehalose utilization in the genomes of *L. lactis* ssp *cremoris* strains. Symbols: +, intact gene; -, gene missing; P, pseudogene; (+) intact gene present but predicted protein has only ~60% amino acid identity to that from KW2.

			KW2	KW10	V4	N41	NCDO763	MG1363	AM2	SK11	SK110	A76	UC509.9	B40	FG2	HP	LMG6897
Function	Protein	OG															
Genotype			cremoris					cremoris									
Phenotype			lactis					cremoris									
Trehalose utilization																	
Trehalose operon transcriptional repressor	GntR	OG_989	+	+	+	+	+	+	+	+	+	+	P	+	+	+	+
PTS system, trehalose-specific IIA component(EC 2.7.1.69)	PTS IIA	OG_138	+	+	+	+	+	+	+	+	+	+	P	+	P	P	P
PTS system, trehalose-specific IIB component(EC 2.7.1.69) / PTS system, trehalose-specific	PTS IIBC	OG_990	+	+	+	+	+	+	+	+	+	P	+	P	+	+	+
Trehalose 6-phosphate phosphorylase (EC2.4.1.216)	GH65	OG_1490	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Beta-phosphoglucumutase (EC 5.4.2.6)	PGM	OG_991	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Maltose utilization (1)																	
Maltose operon transcriptional repressor MalR, LacI family	LacI	OG_1364	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Maltose phosphorylase (EC 2.4.1.8) / Trehalosephosphorylase (EC 2.4.1.64)	GH65	OG_1365	+	+	+	+	+	+	+	+	+	+	P	P	P	P	P
Alpha-glucosidase (EC 3.2.1.20)	GH13	OG_1753	+	+	+	+	+	+	+	+	+	P	-	-	-	-	-
Alpha-amylase (EC 3.2.1.1)	GH13*	OG_1768	+	+	+	+	+	+	+	+	+	P	-	-	-	-	-
Maltose O-acetyltransferase (EC 2.3.1.79)	MOAT	OG_1754	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
Glucan 1,6-alpha-glucosidase (EC 3.2.1.70)	GH13	OG_24	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
Neopullulanase (EC 3.2.1.135)	GH13	OG_1755	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
Maltose/maltodextrin ABC transporter, substratebinding periplasmic protein MalE	Maltose SBP	OG_1756	+	(+)	+	+	+	(+)	(+)	(+)	(+)	(+)	-	-	-	-	-
Maltose/maltodextrin ABC transporter, permeaseprotein MalF	Maltose P	OG_1757	+	(+)	+	+	+	(+)	(+)	(+)	(+)	(+)	-	-	-	-	-
Maltose/maltodextrin ABC transporter, permeaseprotein MalG	Maltose P	OG_1758	+	(+)	+	+	+	(+)	(+)	(+)	(+)	P	-	-	-	-	-
Maltose utilization (2)																	
Transcriptional regulator, LacI family	LacI	OG_3122	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-
membrane protein	HP	OG_3121	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-
hypothetical protein	HP	OG_3120	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-
Maltose phosphorylase (EC 2.4.1.8) / Trehalosephosphorylase (EC 2.4.1.64)	GH65	OG_3119	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-
Maltose/maltodextrin ABC transporter, substratebinding periplasmic protein MalE	SBP	OG_3118	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-
Maltose/maltodextrin ABC transporter, permeaseprotein MalF	P	OG_3093	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-
Maltose/maltodextrin ABC transporter, permeaseprotein MalG	P	OG_2956	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-

* Gene has a signal peptide sequence and is predicted to encode a secreted enzyme.

Supplementary Table 6: Chromosomal genes linked to citrate metabolism found in sequenced strains. Genes from IL1403, DRA4 and M20 are homologous and show >99% amino acid identity. The genes from KF67 show only 54-79% amino acid identity relative to IL1403 and presumably have a different origin.

Gene	OG	IL1403	DRA4	M20	KF67
Acetolactate synthase*	OG_753	+	+	+	+
CitM	OG_3414	+	+	+	+
Malate permease (CitP)	OG_2951	P	P	+	+
CitR	OG_3413	+	+	+	+
CitC	OG_3412	+	+	+	+
CitD	OG_3411	+	+	+	+
CitE	OG_3410	+	+	+	+
CitF	OG_3409	+	+	+	+
CitX	OG_3572	P	+	+	Fused CitXG
CitG	OG_3408	-	P	+	

*The citrate utilization genes are immediately downstream from acetolactate synthase (*als*) in IL1403 and DRA4. In M20 the *als* and citrate metabolism genes are on separate small contigs. In KF67 the citrate metabolism genes are not adjacent to *als* and the *citX* and *citG* genes are fused.

Supplementary Table 8: Details of surface polysaccharide gene clusters. Each worksheet presents different sets of strains compared to reference strain KF147. Color codes: Green, putative plasmid contigs; grey, genes bordering polysaccharide gene clusters; light blue, pseudogenes. See Excel file.

Supplementary Table 9: Grouping of strains with a similar RGP cluster (rhamnose-glucose polysaccharide)

Group 1	DRA4	LMG852	KLDS	N42	LMG144		
Group 2	ATCC19	K231	KF24	Li1	LMG944		
Group 3	KF7	IL1403	KF147				
Group 4	KF196	KF134	KF146	LMG944			
Group 5	KF67	ML8	UC317	CV56	UC509_9		
Group 6	LMG852	E34					
Group 7	FG2	HP	LMG689	NCDO76		MG1363	
Group 8	B40	N41				KW2	
Group 9	SK110	SK11	AM2	KF282			
Ungroupable	A76	V4	IO_1	K337	KW10	KF201	M2

Supplementary Table 10: Prophage insertion location (see details on numbering below the table). Integration loci 1-7 where described earlier (Kelly et al 2013) while loci 8-12 were identified here.

Strain	Prophage insertion location											
	1	2	3	4	5	6	7	8	9	10	11	12
KW10												
KW2					X							
V4		X	X	X	X				X			
N41						X						
NCD0763		X				X						
MG1363		X				X						
SK110						X						
AM2				X		X						
SK11		X		X		X						
A76	X				X	X						
UC509.9	X					X						
B40		X	X		X	X						
FG2				X		X						
HP				X		X						
LMG6897				X		X						
LMG8526										X	X	
KF282				X	X		X				X	
LMG8520	X											
IO-1		X		X								
K231			X		X				X			
Li-1		X										
KF24					X		X		X			X
LMG9449				X		X						
ATCC19435	X		X	X		X						
E34					X							
K337									X			
KF201		X	X									
M20		X							X			
N42	X		X				X					
DRA4	X							X				
IL1403	X		X	X								
CV56				X		X	X					
ML8			X	X		X	X					
UC317						X	X			X		
LMG14418	X					X		X				
KLDS						X	X	X			X	
KF147			X	X								
KF7					X							
KF67						X	X		X			
KF196						X						
KF146						X						X
KF134						X			X			
LMG9446		X			X							

Insertion location numbering:

1 between *lytR* (COG1316) and *truA* (COG0101);
 2 between *rbsB* (COG1879) and *uxuB* (COG0246);
 3 between *carB* (COG0458) and *coaA* (COG1972);
 4 between *rex* (COG2344) and *radC* (COG2003);
 5 between *fabK* (COG2070) and *sufB* (COG0719);
 6 between *sunL* (COG0144) and *fmt* (COG0223);
 7 between *comGB* (COG1459) and *comGC* (COG4537);
 8 between *deoD* (COG0813) and *fhs* (COG2759);
 9 between *rplS* (COG0335) and *yijG* (COG1215);
 10 between *adk* (COG0563) and *rpsM* (COG0099);
 11 between *carA* (COG0505) and *rpmE* (COG0254);
 12 between *ligA* (COG0272) and *ptcC* (COG1455);

Supplementary Table 11: Sex factor gene clusters. The four worksheets represent different sets of strains with similar sex factor gene clusters. The putative sex factor gene cluster is indicated as SexF in the column “cluster”. See Excel file.

Supplementary Data in online repository

The following data was deposited at zenodo.com under the DOI: 10.5281/zenodo.1471674

<https://doi.org/10.5281/zenodo.1471674>

FILE NAME	FILE DESCRIPTION
alignments.zip	This .zip file contains the individual fasta files and the alignmens from all OGs used in this study.
GTM_output.zip	This .zip file contains all output files from the genotype-phenotype matching done with the phenotypes described by Bayjanov <i>et al.</i>