



HAL
open science

Low-redundancy sequencing of the entire *Lactococcus lactis* IL1403 genome.

Alexander Bolotin, Stéphane Mauger, Karine Malarme, S. Dusko Ehrlich,
Alexei Sorokin

► **To cite this version:**

Alexander Bolotin, Stéphane Mauger, Karine Malarme, S. Dusko Ehrlich, Alexei Sorokin. Low-redundancy sequencing of the entire *Lactococcus lactis* IL1403 genome.. *Antonie van Leeuwenhoek*, 1999, 76 (1-4), pp.27-76. 10.1023/a:1002048720611 . hal-04297789

HAL Id: hal-04297789

<https://hal.science/hal-04297789v1>

Submitted on 23 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

1 **Low-redundancy sequencing of the entire *Lactococcus lactis***

2 **IL1403 genome**

3 Alexander Bolotin, Stéphane Mauger, Karine Malarme, S. Dusko Ehrlich & Alexei Sorokin Génétique
4 Microbienne, INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France.

5 (Author for correspondence; E-mail: sorokine@biotec.jouy.inra.fr)

6

7 **Abstract**

8 *Lactococcus lactis* is an AT-rich gram positive bacterium phylogenetically close to the genus
9 *Streptococcus*. Various strains of *L. lactis* are used in dairy industry as starters for cheese making. *L.*
10 *lactis* is also one of the well characterized laboratory microorganisms, widely used for studies on
11 physiology of lactic acid bacteria. We describe here a low redundancy sequence of the genome of the
12 strain *L. lactis* IL1403. The strategy which we followed to determine the sequence consists of two main
13 steps. First, a limited number of plasmids and -phages that carry random segments of the genome were
14 sequenced. Second, sequences of the inserts were used for production of novel sequencing templates by
15 applying Multiplex Long Accurate PCR protocols. Using of these PCR product allowed to determine
16 the sequence of the entire 2.35 Mb genome with a very low redundancy, close to 2. The error rate of the
17 sequence is estimated to be below 1%. The correctness of the sequence assembly was confirmed by PCR
18 amplification of the entire *L. lactis* IL1403 genome, using a set of 266 oligonucleotides. Annotation of
19 the sequence was undertaken by using automatic gene prediction computer tools. This allowed to
20 identify 1495 protein-encoding genes, to locate them on the genome map and to classify their functions
21 on the basis of homology to known proteins. The function of about 700 genes expected to encode
22 proteins that lack homologs in data bases cannot be reliably predicted in this way. The approach which
23 we used eliminates high redundancy sequencing and mapping efforts, needed to obtain detailed and
24 comprehensive genetic and physical maps of a bacterium.

25 Availability of detailed genetic and physical maps of the *L. lactis* IL1403 genome provides many entries
26 to study metabolism and physiology of bacteria from this group. The presence of 42 copies of five
27 different IS elements in the IL1403 genome confirms the importance of these elements for genetic
28 exchange in *Lactococci*. These include two previously unknown elements, present at seven and fifteen

29 copies and designated IS1077 and IS983, respectively. Five potential or rudimentary prophages were
30 identified in the genome by detecting clusters of phage-related genes. The metabolic and regulatory
31 potential of *L. lactis* was evaluated by inspecting gene sets classified into different functional categories.
32 *L. lactis* has the genetic potential to synthesise 20 standard amino acids, purine and pyrimidine
33 nucleotides and at least four cofactors. Some of these metabolites, which are usually present in
34 chemically defined media, can probably be omitted. About twenty compounds can be used by *L. lactis*
35 as a sole carbon source. Some 83 regulators were revealed, indicating a regulatory potential close to that
36 of *Haemophilus influenzae*, a bacterium with a similar genome size. Unexpectedly, *L. lactis* has a
37 complete set of late competence genes, which may have concerted transcriptional regulation and
38 unleadered polycistronic mRNAs. These findings open new possibilities for developing genetic tools,
39 useful for studies of gene regulation in AT-rich gram positive bacteria and for engineering of new strains
40 for the dairy industry.

41

42 **Introduction**

43 *Lactococcus lactis* is probably the best characterised lactic acid bacterium, with highly developed
44 knowledge on physiology and genetics. It is therefore an obvious target for genome sequencing, since
45 the availability of the full sequence would greatly improve the efficiency of numerous studies that are
46 currently being conducted with this organism, or are planned in the near future. We decided to undertake
47 the *L. lactis* genome sequencing project in an academic setting, with resources of a single laboratory,
48 which obviates the need to raise a rather large amount of funding which would be required for sequence
49 determination by a private company, or by a consortium of academic laboratories. For this purpose, we
50 developed an approach that allowed us to determine the genome sequence with a redundancy much lower
51 than generally used, and at a commensurately lower cost. Due to the low redundancy, the sequence is of
52 a so-called diagnostic quality, allowing to detect essentially all genes that have homologs in other
53 organisms, but not the genes which may be present in the *Lactococcus* genome only.

54 The approach most often used for sequencing of an entire bacterial genome is based on a high-
55 throughput random clone sequencing (Fleischmann et al. 1995). This method is most efficient if the
56 facilities to produce raw reading data are not limiting. However, the speed of accumulation of new
57 sequencing data decreases exponentially with the linear increase of the number of randomly sequenced

58 templates (Lander & Waterman 1988). This becomes a limiting factor, in particular for relatively large
59 bacterial genomes. Furthermore, the assembly becomes exceedingly difficult if the genome contains
60 numerous repeated elements. The chromosome of *L. lactis* IL1403 was estimated to have a size of about
61 2.4 Mb (Le Bourgeois et al. 1992). Application of the whole-genome random sequencing and assembly
62 method (Fleischmann et al. 1995) would require producing of 30,000 random gel readings,
63 corresponding to six-fold coverage, and the sequence would be distributed in about 80 contigs. We
64 considered an alternative strategy that would consist in a much less massive random sequencing
65 followed by primer walking on templates generated by Long Range PCR (LR PCR, Cheng et al. 1994).
66 If sequence acquisition and assembly were carried out in parallel, template synthesis can be conducted
67 in such a way that the re-sequencing of the regions that were sequenced already is kept at a minimum.
68 We estimated that such approach may allow to sequence the whole *L. lactis* genome with a coverage of
69 only about two. Some 10,000 sequencing reactions instead of 30,000 would therefore be required. We
70 report here the *L. lactis* genome sequence determined by this approach.

71

72 **Low-redundancy genome sequencing**

73 The strategy of low-redundancy *L. lactis* genome sequencing is presented in Table 1. This strategy is a
74 compromise between completely random and directed approaches. Its objective was to reduce time and
75 effort or obtaining complete information about gene content and genome organisation. In the first stage,
76 a limited number of randomly chosen clones is sequenced, so that the speed of new sequence
77 accumulation keeps approximately constant. We estimated that this condition is fulfilled up to the
78 genome coverage of about one. A large number of small contigs, separated by sequencing gaps, is
79 present at the end of the first stage. In the second stage randomly chosen clones carrying large inserts
80 are sequenced by primer walking. This allows to close some sequencing gaps and extend the contigs, as
81 verified by sequence assembly, which is carried out in parallel with the acquisition of new sequence.
82 The redundancy is kept low by selecting the primers that correspond to the ends of the extended contigs
83 for the next round of primer walking. This stage is carried out until the rate of acquisition of the new
84 sequence is above about 1 new base for 3 sequenced bases. Finishing of the sequence is achieved by
85 applying another directed procedure, which was designated Multiplex Long Accurate PCR (MLA PCR)
86 and tested in the course of our participation in the *Bacillus subtilis* sequencing project (Sorokin et al.

87 1996; Bolotin et al. 1997; Lapidus et al. 1997). It involves mixing a large number of primers that
88 correspond to the ends of contigs. A product will be obtained each time when the distance between two
89 sites on the genome, corresponding to two contig ends, is less than the maximal size that can be
90 synthesised by LR PCR. The probability of producing such products, as a function of the genome size
91 and the number of primers present in LRPCR reaction is shown in Figure 1. For the genome of *L. lactis*
92 size this probability is between 0.5 and 1 if some 20 primers are mixed (that is at least half of PCR
93 reactions containing 20 randomly chosen primers will give an amplification product). It should be noted
94 that the efficiency of gap closure is expected to increase with the progression of the sequencing project,
95 since the number of contigs continuously decreases. This is in sharp contrast with the fully random
96 approach, where the efficiency of gap closure decreases exponentially with sequence acquisition. We
97 now believe that the long-insert library construction can be omitted.

98 Some statistical data of application of this strategy for the *L. lactis* sequencing are presented in Table 2.
99 A library consisting of 2854 clones with insert sizes in the range from 1 to 2 kb, was constructed by
100 using pBluescript II KS+ (Stratagene) or pSGMU2 (Errington1986) vectors. 2625 clones were
101 sequenced by forward (M13-21) and 2168 by reverse (M13RP1) primers with successful sequencing
102 rate of about 90%. After accumulating about 2100 kb of sequence 2357 oligonucleotides were
103 synthesised to fill the gaps between the forward and reverse sequences. A total of about 3.3 Mb of
104 sequence was thus acquired.

105 -FIXII vector (Stratagene) was used to construct along insert library. *L. lactis* chromosome was
106 partially digested with Sau3A, fractionated by centrifugation in sucrose gradient, treated with Klenow
107 polymerase in the presence of dGTP and dATP, and ligated with -FIXII arms cut by XhoI and treated
108 with Klenow in the presence of dCTP and dTTP. 262 phages were chosen at random and the ends of
109 inserts sequenced by the T7 primer (Stratagene). Of these, 122 phages which had a unique sequence
110 obtained by the T7 primer, were sequenced by the T3 primer (Stratagene). These 122 clones appeared
111 to constitute a nonredundant set and their inserts were sequenced by primer walking (see Table 2). Some
112 250 kb of sequence were obtained in this way.

113 MLA PCR was used to obtain new sequencing substrates. A critical step in the procedure is to identify
114 in the mixture the two primers that give a product useful for sequencing. The previously developed
115 protocol (Sorokin et al. 1996), where two steps were required for the identification, was modified here

116 so that only a single step is required. The principle of the modification is illustrated in Figure 2A. In this
117 theoretical example each reaction contains 8 different primers. A total of 16 primers is used (labelled 1–
118 16 on the left) and there are 16 reactions (labelled 1–16 on the top). To obtain a different primer
119 combination in each reaction one primer is replaced at a time with a new primer (for instance, reaction
120 8 contains primers 1–8, reaction 9 contains primers 2–9, etc). The primer couples which give a useful
121 product can be identified unambiguously, by simply identifying the reaction mixtures in which the
122 product is synthesized, as illustrated in the three lower rows. For instance, if a product were found only
123 in the reaction 8, primers 1 and 8 would be required for it, since only the reaction 8 contains these two
124 primers together. Alternatively, if a product were found in reactions 4 to 10, primers 3 and 4 would be
125 required. A general correlation between positive reactions and primer couples is given in the legend of
126 Figure 2. An example of application of this protocol to the IL1403 genome, using 32 oligonucleotides
127 in each reaction, is shown on Figure 2B (the analysis of 36 reactions out of a total of 96 that were
128 routinely run together is displayed). There are 4 different products in the reaction mixtures, ranging
129 between 1 and 10kb. The largest product appears in reactions 7 to 34, identifying that the primers 7 and
130 3 are required for its synthesis (see legend, Figure 2). The primers identified in this way were used to
131 sequence the ends of the PCR products and thus confirm the identification and to extend the sequence.
132 In total 1641 such sequencing reactions on the products with sizes from 1 to 20 kb were performed and
133 about 0.77 Mb of sequence was read (Table 2). This step allowed to finish the entire chromosome
134 assembling, giving a contig of 2.34 Mb. The total redundancy was close to 2 (Table 2).

135 To test the correctness of the assembling we carried out LR PCR amplification of the entire genome
136 using 266 oligonucleotides, predicted to be separated by distances 10–20 kb. The expected products were
137 obtained (Figure 3), which indicates that the assembling is correct. The sequences of these 266
138 oligonucleotides are presented at <http://spock.jouy.inra.fr>.

139 The average redundancy of the *L. lactis* genome sequence is two. Such low redundancy does not allow
140 to eliminate all errors by comparing multiple sequences corresponding to the same genome region. The
141 error rate of this sequence is therefore determined by individual gel reading quality. We counted the
142 number of ambiguities in the raw sequence data used for the assembling and found 46,954 ambiguities
143 for a total of 4,687,630 bases. From this estimation the accuracy of the *L. lactis* diagnostic sequence is
144 about 1% (one ambiguity/100 bases). However, assembling usually decreases the number of errors in

145 the final consensus sequence and the real error rate may be lower. We therefore carried out a different
146 quality test, aiming to detect the level of frame-shift errors in the open reading frames present in the
147 genome sequence. For this purpose we used the GENMARK programme (Borodovsky & Mc Ininch
148 1993), which predicts the genome regions statistically fit to encode proteins. Some 3544 protein coding
149 regions were thus detected. If the real density of genes in *L. lactis* were similar to that observed in other
150 bacterial genomes, we would expect only about 2300 such regions. There should therefore be over 1200
151 frame-shift errors in our sequence (the difference between 3544 regions detected and 2300 expected),
152 corresponding to about 1 error for two genes, or 2 kb. We suggest therefore that the overall error rate of
153 the diagnostic quality sequence after assembling, at a redundancy of 2, may be close to 1 in 1000.

154

155 **Genome analysis**

156 GC skew, replication origin and terminus. *L. lactis* IL1403 chromosome has the size of 2345 kb.
157 The average GC-content is 35.4%, its variation over the chromosome is shown on Figure 4. By analogy
158 with the analysis performed on the *B. subtilis* genome (Kunst et al. 1997), the deviation from the mean
159 GC content may indicate the presence of foreign insertion elements, such as transposons or prophages.
160 We did not detect, however, in the *L. lactis* genome, a correlation of GC-content with location of these
161 elements, determined by other approaches (see below). Several regions of highly elevated GC-content
162 were revealed (Figure 4). Six such peaks with GC-content higher than 46%, at positions 540, 1970,
163 2210, 2250, 2325 and 2340 (Figure 4, bottom) correspond to rRNA operons. Another GC rich region,
164 near position 2150 kb (Fig 4, bottom), corresponds to the cluster of ribosomal protein genes. We
165 conclude that the regions encoding ribosomes have a higher GC content than the other regions in *L.*
166 *lactis*. In two other AT-rich bacteria, *B. subtilis* and *Rickettsia prowazekii*, the regions encoding
167 ribosomal genes are not particularly GC-rich (Kunst et al. 1997; Andersson et al. 1998).

168 GC skew, replication origin and terminus. GC or AT skews are useful to predict positions of replication
169 origin and terminus (Lobry 1996a, b). The distributions of the values $(C-G)/(C+G)$ and $(A-T)/(A+T)$
170 over the chromosomal region in the vicinity of *dnaA* gene is shown in Figure 5, top. A sharp transition
171 between positive and negative values is observed, and indicates the presence of the replication origin
172 (Figure 5A). This region contains four *DnaA* boxes (Figure 6), which also indicates the presence of the
173 replication origin. We choose the start point for the presentation of the circular *L. lactis* genome map at

174 the middle of HindIII site near the replication origin and the map is oriented to be colinear with the
175 transcription direction of the *dnaA* and *dnaN* genes (see below, Figure 9).

176 GC and AT skews may also indicate location of the replication terminus. This is seen for *L. lactis*
177 genome from Figure 5, bottom. The transition between positive and negative values occurs near the
178 position 1260 kb. These data correlate with the replication terminus location based on orientation of
179 transcription of potential genes and distribution of chi-sites over the genome (see below).

180 IS-elements and prophages in *L. lactis* IL1403. Three different IS elements in the genome of *L. lactis*
181 IL1403 were known previously, designated IS981, IS982 and IS1076. Their number of copies (ten, one
182 and seven, respectively) and approximate locations were reported previously (Le Bourgeois et al. 1992,
183 1995; Polzin et al. 1993; Schouler et al. 1998). The element IS1076 was initially identified in the strain
184 *L. lactis* subsp. *lactis* CNRZ270 on the plasmid pUCL22 (Huang et al. 1991). Nucleotide sequence of
185 this element (Genbank accession X53013) shows complete identity to IS904, known from nisin-
186 producing lactococcus (Engelke et al. 1992). Our sequencing data confirmed that in all chromosome
187 locations where IS1076 was mapped, the nucleotide sequence identical to IS904 is present, and we
188 therefore keep the latter name on the map. We found, however, that another element, which we designate
189 IS1077, was also present in each of these seven sites. At two sites the IS1077 element is flanked by two
190 sequences identical to IS904. The orientation of IS904 relative to IS1077 is different in five cases out of
191 seven. A possible explanation for the fact that the two elements always occur together in this genome
192 would be that one of these elements carry a sequence essential for the transposition of the other. Since
193 IS904 can be found without IS1077 in other genomes, the essential sequence must be in IS904. By
194 searching Genbank we identified another copy of IS1077 in the plasmid pK214, (sequence accession
195 number is X92946). IS904 is also present on this plasmid.

196 Fifteen copies of an element, which was not described earlier for lactococcus species and is designated
197 here IS983, were detected in the IL1403 genome. The closest relative from other species is IS1239 from
198 *Streptococcus pyogenes* (Ashbauch et al. 1998), and, from other lactic acid bacteria, it is IS1070,
199 discovered in the plasmid pNZ63 in *Leuconostoc lactis* NZ6009 (Vaughan & de Vos 1995).

200 Chopin et al. (1989) detected in the chromosome of IL1403 two different prophages, bIL285 and
201 bIL286, which can be induced for lysogenic development by UV irradiation. Two more prophages,
202 bIL309 and bIL310, can also probably be induced if supplied in trans by some functions of the first two

203 phages (Chopin & Perreira, pers. comm.). We tried to identify potential prophages in the chromosome
204 by using homology search against a database containing known phage protein sequences. The database
205 consisted of 1219 protein sequences, including the complete set of 50 putative proteins derived from the
206 sequence of the rlt temperate phage of *L. lactis* (van Sinderen et al. 1996). A distribution of
207 nonredundant homology hits over *L. lactis* genome is shown on Figure 7. This distribution indicates the
208 presence of three regions, around 460, 1060 and 1410 kb, which may contain the pro-phages identified
209 by Chopin et al. (1989). Two areas, around 40 and 2005 kb, may indicate a fourth and a fifth prophage.
210 More detailed analysis of these areas will help to elucidate the correspondence between the prophages
211 detected by sequencing and by induction of the lysogenic development.

212

213 *Distribution of chi-sites over the chromosome.* Recent analyses suggested that so-called chi-sites, which
214 in *L. lactis* have sequence 5'-GCGCGTG-3', protect DNA from exonuclease degradation (Biswas et al.
215 1995). The same sequence was reported to modulate the activities of RexAB exonuclease corresponding
216 to RecBCD of *E. coli* and is thus involved in the regulation of homologous recombination (el Karoui et
217 al. 1998). In *E. coli* and *B. subtilis* such sequences are over-represented and most of them are found to
218 be co oriented with DNA replication (Burland et al. 1993; Chédin et al. 1998). The latter property was
219 correlated either with the involvement of RecBCD in recombination, and therefore preferring the same
220 DNA structures as the RecA protein (Tracy et al. 1998), or with a statistical bias of base distribution,
221 due to codon usage (Biaudet et al. 1998). The chi sequence is presented in *L. lactis* on average once per
222 13.2 kb (178/2350 kb) and tends to be co-oriented with replication (Figure 8). Interestingly, the density
223 of chi sequences increases up to one site per 9 kb (34/300 kb) in the region close to the replication origin.
224 Also, chi-site density decreases in the regions near 460, 1060, 1410 and 2000 kb, where the putative
225 prophages were detected (see above).

226

227 *RNA genes.* Six rRNA operons were mapped on the *L. lactis* chromosome (Tulloch et al. 1991; Le
228 Bourgeois et al. 1992) and one was sequenced (Beresford & Condon 1991; Chiaruttini & Milet 1993).
229 We did not resequence all other operons since the diagnostic sequencing would not reveal reliably any
230 difference. However, we confirmed map positions of these operons by PCR amplification of the
231 corresponding genome regions.

232 tRNA genes were identified with the computer programme tRNA scan-SE (Lowe & Eddy 1997). We
233 detected 62 tRNA genes, corresponding to the 20 amino acids. Most of the tRNA genes were found to
234 be organised in four large operons, three of them transcribed as parts of rrnB, rrnE and rrnF rRNA
235 operons. The trnA operon, encoding seven tRNAs, was described earlier (Nilsson & Johansen 1994).
236 Twenty genes encoding tRNA were found to be isolated from other tRNA genes. RNA component of
237 RNase P (rnpB gene) and 10S RNA (ssrA) were detected by homology at 1703 and 1458 kb,
238 respectively.

239

240 *Protein-coding genes.* Prediction of protein-coding regions by methods based on statistical properties
241 of DNA is limited for the sequences of diagnostic quality, since potential frameshifts will hamper the
242 prediction of true stop codons. It is, however, possible to predict genes encoding proteins homologous
243 to proteins from other organisms. We used for the comparison 30757 protein sequences, corresponding
244 to prokaryotic and phage genes, available from SwissProt (release of 17.02.98). Several home-made Perl
245 5.0 or UNIX shell scripts had been used for automatic extraction of *L. lactis* protein sequences and other
246 parameters from the standard BLASTx or FASTx output files and organising them into clickable tables
247 or genome map drawings. The extracted protein set of *L. lactis* genome was then again used for
248 homology searches, to find all paralogous proteins encoded in the genome. The largest protein from the
249 FASTx comparison, or the one, having BLASTx homology hits to the same Swissprot entry, was
250 considered as really encoded by a particular area of the *L. lactis* genome. The final protein set consists
251 of 1495 proteins. If the true number of genes in the *L. lactis* genome were close to 2300 (a density of
252 one gene per kb), our analysis would have identified about two thirds of them. A version of the
253 automatically generated ORF map of *L. lactis* IL1403 genome is shown on Figure 9. There is a strong
254 bias of gene orientations to be colinear with the direction of replication. The same phenomenon was
255 seen for another AT-rich Gram positive bacterium *B. subtilis* (Kunst et al. 1997), but not for the GC-
256 rich *M. tuberculosis* (Cole et al. 1998). It may be a characteristic of AT-rich gram positive
257 microorganisms.

258 Another automatic tool, based on the proposed functional classifications of genes in prokaryotic
259 genomes (Riley 1993; Fleischmann et al. 1995) and commentaries from SwissProt, has been used to
260 classify the genes. The programme uses list of functional categories and the relevant keywords with

261 assigned weights corresponding to each category for a given keyword. The advantages of such a
262 classification method are: it is completely automatic, it can be applied to genes with hypothetical
263 functions and needs the interaction with an expert only at the stage of construction of the keyword list.
264 After a keyword list was constructed and the programme had been educated y using several functionally
265 annotated genomes, it provided satisfactory results, which can be used by experts as a draft for more
266 precise functional annotation of the genome and classification of the genes. Any protein database can
267 be automatically classified by using this approach and then used in turn to classify a set of new proteins.
268 Presumably, such method of automatic functional classification of protein databases can be further
269 developed to give tools, which are completely independent from human expert and use different sources
270 for functional description of proteins. Parameters of homology with several proteins, rather than with
271 the best one, can be used to increase the performance of the approach for classification of hypothetical
272 proteins. The functional classification of predicted genes is colour-encoded on Figure 9 and summarised
273 in Figure 10 and in Table 3. A more detailed and systematically updated version of Table 3 can be found
274 at <http://spock.jouy.inra.fr>.

275

276 **Metabolic potential of *L. lactis* IL1403.**

277 The automatic procedure for protein finding, function assignment and classification, described above,
278 allows to estimate the potential of *L. lactis* as a free-living microorganism. Several of the detected
279 systems were characterised experimentally earlier, by biochemical, genetic or gene cloning means.

280

281 *Amino acid, vitamin and nucleotide biosynthesis*

282 Lactic acid bacteria are nutritionally fastidious. The presence of a gene for synthesis of a metabolite
283 does not mean that a strain of *L. lactis* where the gene was found will not require this metabolite in the
284 growth medium. The studies performed long ago (Anderson & Elliker 1953) and recently (Jensen &
285 Hammer 1993; Cocaign-Bousquet et al. 1995) showed that synthetic medium for *L. lactis* should contain
286 at least six amino acids (isoleucine, valine, leucine, histidine, methionine and glutamic acid) and seven
287 vitamins (biotin, pyridoxal, folic acid, riboflavin, niacinamide, thiamine and pantothenic acid). Our
288 analysis showed that *L. lactis* has a genetic potential to synthesise 20 standard amino acids and at least
289 four cofactors (folic acid, menaquinone, riboflavin and thioredoxin). The genes for biosynthesis of

290 aromatic amino acids, threonine, branched chain amino acids and histidine were cloned and
291 characterised earlier (Bardowski et al. 1992; Delorme et al. 1992; Godon et al. 1992; Griffin &
292 Gasson 1995; Madsen et al. 1996). The regulation of transcription of the tryptophane operon was studied
293 in detail (Raya et al. 1998; van de Guchte et al. 1998). The problem of fastidious nutritional requirements
294 of *L. lactis* strains has recently been addressed by application of the single-omission technique (Cocaign-
295 Bousquet et al. 1995; Novak et al. 1997) and genetic approaches (Lapujade et al. 1998). It was shown
296 also that auxotrophy of IL1403, used as a dairy strain, for histidine and branched-chain amino acids is
297 due to the recently acquired mutations (Delorme et al. 1993; Godon et al. 1993).

298 Biosynthesis and salvage of nucleotides and nucleosides in *L. lactis* were explored earlier and several
299 pyrimidine and purine biosynthetic genes were cloned and sequenced (Martinussen et al. 1994;
300 Andersen et al. 1996; Nilsson & Kilstrup 1998). It was demonstrated that milk does not contain sufficient
301 purine compounds to support growth and therefore de novo biosynthesis is necessary (Dickely et al.
302 1995). Nucleoside compounds are not needed in the minimal synthetic medium to maintain exponential
303 growth of *L. lactis* (Jensen & Hammer 1993). We detected 47 genes implicated in this metabolism. Most
304 of the genes for biosynthesis of purines are clustered near the purDEK operon, which was recently
305 characterised (Nilsson & Kilstrup 1998). A copy of IS983 is detected between the purDEK operon and
306 other genes of purine biosynthesis.

307

308 *Energy metabolism and transporters.*

309 The main feature of the energy metabolism in *L. lactis* is that fermented carbon sources are used mostly
310 to maintain the energetic potential of the cells (Poolman 1993; Cocaign-Bousquet et al. 1996). About
311 5% of the fermented sugar is used as the source of substrates for synthetic reactions (Novak et al. 1998).
312 *L. lactis* does not have a functional citric acid cycle and gluconeogenesis enzymes and is unable to carry
313 out many anaplerotic reactions. The exponential growth of this bacterium appears to be 'programmed'
314 to happen only in sugar rich complex media, like milk. As mentioned above, many metabolites have to
315 be added into the synthetic medium (Jensen & Hammer 1993), although the corresponding biosynthetic
316 genes seem to be present.

317 It is interesting to compare the genetic potential of *L. lactis* to grow on different carbon sources with the
318 available experimental data. The genetic potential can be estimated from the presence of the

319 biodegradation genes and the relevant transporters. IL1403 has genes which can be used to grow on
320 different carbon sources: glucose (glycolysis genes), fructose (positions 1519 and 2230 kb), N-acetyl
321 glucosamine (1032 kb), xylose (1550 kb), ribose (1685 kb), mannose (777 kb), gluconate (606 and 2253
322 kb), maltose (692, 700 and 1526 kb), lactose (2040 kb), galactose (2045 kb), mannitol (33 kb), different
323 -glucosides (186, 419, 830, 1490 and 1520 kb). Glucuronate or galacturonate catabolic operon (1670
324 kb) can be used for utilisation of pectin degradation products as a supplementary source of energy and
325 carbon. Components of sugar-specific enzyme II of phosphoenol-pyruvate dependent transport systems
326 were found for mannitol (30 kb), sucrose or trehalose (435 kb), fructose (984 and 1747 kb), mannose
327 (1748 kb) and -glucosides (175, 416, 830, 1144 and 1489 kb). These findings are summarised in Table
328 4, along with references to experimental data on the growth of *L. lactis* on different carbon sources. The
329 ability to grow more efficiently in particular conditions can be acquired by *L. lactis* by means of plasmid
330 transfer. This is well illustrated by experimental studies of the system needed for efficient growth of *L.*
331 *lactis* in milk. To overgrow other bacteria in this niche *L. lactis* needs efficient systems for utilisation of
332 lactose and peptides. Such genes, which include PEP-dependent transporter for lactose and proteinase
333 for peptides, were found in the plasmid pLP712 isolated from *Streptococcus lactis* NCDO 712 (Gasson
334 1983; Maeda & Gasson 1986). The system encoded by pLP712 contains a PEP-dependent transporter
335 for lactose. This allows efficient entry of lactose into glycolytic pathways through glucose, which is
336 phosphorylated by glucokinase, and galactose-6P, which is metabolised through the tagatose-6P
337 pathway, also encoded by pLP712. Sequence analysis of the IL1403 chromosome revealed that the genes
338 encoding the PTS-dependent pathway of lactose utilisation are absent in this strain. The chromosome
339 contains, however, another system for lactose utilisation, dependent on the transport by the product of
340 the *lacS* gene, encoding an HC symporter or galactose-lactose antiporter. This system was characterised
341 earlier for *Streptococcus thermophilus*. Special selection is needed for *S. thermophilus* strains to be able
342 to ferment galactose (Thomas & Craw, 1984; Hutkins et al. 1985). Strain IL1403 was isolated from
343 cheese starter culture and cured from plasmids for the purpose to be used as a general genetic engineering
344 host (Chopin et al. 1984; Simon et al. 1985). Probably one of the eliminated plasmids contained genes
345 for efficient utilisation of lactose, similar to the ones encoded by pLP712.

346 Chromosomal genes for galactose and lactose utilisation were cloned and characterised recently from
347 different *L. lactis* strains (Grossiord et al. 1998; Vaughan et al. 1998). It appeared that the strains can be

348 divided in two groups by their possession or lack of the inducible -galactosidase gene. Since we found
349 the gene in IL1403, this strain belongs to the former group. It is different in this respect from another
350 model strain cured of plasmids, MG1363 (Gasson 1983), which does not have -galactosidase.

351 Automatic analysis detected 14 genes involved in glycolysis, completing the description of this system
352 and revealed a second glyceraldehyde-phosphate dehydrogenase gene. It also confirmed the absence of
353 a complete citric acid cycle, revealing only the genes working around pyruvate, like citrate lyase (5
354 genes), malate oxidoreductase and pyruvate dehydrogenase (4 genes). Close analysis of the genes
355 automatically classified as relevant to glyconeogenesis confirmed that this process is not present in
356 *L.lactis*. No genes for fructose 1,6 bisphosphatase, phosphoenolpyruvate carboxy kinase or
357 phosphoenolpyruvate synthetase were found.

358 It is interesting to mention that all genes of the pentose-phosphate pathway were detected, but not a
359 putative trans-aldolase. It is not yet clear which gene is encoding this function in *B. subtilis*. It seems
360 that the structure of the corresponding enzyme in AT-rich gram positive bacteria is different from that
361 in gram negative and GC-rich gram positive bacteria. The corresponding gene in AT-rich gram positive
362 bacteria still has to be identified.

363

364 *Cell envelope*

365 The cell surface structure has a large impact on the properties of lactic acid bacteria relevant for their
366 application in milk industry, such as resistance to stress, cell lysis, sensitivity to phages antimicrobial
367 agents and influence on the product texture. However, very few studies, mostly related to the potential
368 application of lactic acid bacteria as a live vehicle for vaccine delivery, were reported (Pouwels et al.
369 1998). Our analysis revealed 73 genes implicated in cell envelope function, including 29 membrane
370 proteins, 23 murein succulus and peptidoglycan biosynthesis genes and 21 surface polysaccharide
371 biosynthesis genes. Operons for the biosynthesis of extracellular polysaccharides were identified
372 recently in the chromosome of *Streptococcus thermophilus* (Stingele et al. 1996) and on a plasmid in *L.*
373 *lactis* (van Kranenburg et al. 1997). We did not find similar genes in the chromosome of IL1403. This
374 may be another property of industrial interest which is encoded and propagated predominantly by
375 plasmids in *L. lactis*.

376

377 *Cellular processes*

378 *L. lactis* is not known as a good secretor of proteins compared to, for example, *Bacillus* species.
379 Optimised gene constructions allow *L. lactis* to produce up to 20 mg of foreign protein per litre (Wells
380 et al. 1993; Le Loir et al. 1998), compared to 800 mg produced by *B. subtilis* (Hemila et al. 1992). The
381 reasons maybe the lack of some components of the secretion machinery or the inefficiency of energy
382 metabolism. We did not detect components such as *secDF* genes, which are known to improve the
383 secretion efficiency in gram negative and gram positive bacteria (Pogliano & Beckwith 1994; Duong &
384 Wickner 1997; Bolhuis et al. 1998). Also, only one signal peptidase gene was found in IL1403, compared
385 to five detected in the *B. subtilis* genome (Kunst et al. 1997). Probably, the useful genes Figure 10.
386 Distribution of the detected genes by functional categories. Colours designating different categories are
387 defined in the legend of Figure 9. The number and the percentage of homology hits to SwissProt are
388 indicated for each category. Classification method is presented in the text. for protein secretion can be
389 acquired by lactococci from closely related species through plasmid transfer, if selective conditions for
390 this are created by industrial applications.

391 Natural competence for DNA-mediated transformation was not described for *L. lactis* species. However,
392 we found that the *L. lactis* genome contains four operons, *comE*, *comF*, *comC* and *comG*, which encode
393 nine genes that share strong similarity with the late competence genes from *B. subtilis* and *S.*
394 *pneumoniae*. Since *L. lactis* seems to possess a complete set of these genes and since no function other
395 than the involvement in DNA-mediated transformation is known for them, it is possible that *L. lactis*
396 can acquire natural competence. Sequence studies of late competence genes showed that they are quite
397 similar in different AT-rich gram positive bacteria, like *B. subtilis* or Streptococci (Dubnau 1997;
398 Campbell et al. 1998). It is apparent so that a big difference exists in the molecular mechanisms which
399 regulate the development of this process in Streptococci and Bacilli. In *B. subtilis* the regulator ComK
400 assembles the signals from earlier stages of competence development (van Sinderen et al. 1995; Hahn
401 et al. 1996). A functional counterpart of such a regulator was not yet discovered for Streptococci. A way
402 to discover the conditions for acquiring the competence may be the study of regulation of the late genes.
403 We analysed the regions preceding the first ORF in the four late competence operons. The corresponding
404 alignment is shown in Figure 11. It is apparent that a common feature exists for these regions, which
405 suggests that they may be recognised by a common transcription regulator. An interesting feature, which

406 seems to be unique for these genes, is the presence of a strictly conserved sequence AGTATG, in which
407 ATG always corresponds to the start codon of the first gene in each operon. This is different from the
408 usual ribosome binding site in AT-rich gram positive bacteria, which has a sequence (SD) homologous
409 to the 30-end of 16S rRNA (Vollenoweth & Rabinowitz1992). Several recent reports show that the
410 efficiency of translation may be not much lower for mRNAs that start directly from ATG codon than
411 for mRNAs with long leaders and strong SD-sequences (Resch et al.1996; Moll et al. 1998; Van Etten
412 & Janssen 1998), and that the ribosomal protein S1 is not needed for the translation of unleadered mRNA
413 (Tedin et al. 1997). It appears that the absence of the leader region is a common feature of mRNAs of
414 late competence genes in *L. lactis*.

415 Studies of the regulation of competence in Streptococci revealed the presence of a gene, encoding a
416 small peptide, that stimulates the induction of expression of the whole competence regulon (Havarstein
417 et al. 1995, 1997). In Streptococci the gene encoding competence-stimulating peptide (CSP) precursor
418 was always found to be the first gene of an operon encoding a two-component regulatory system, which
419 is involved in the induction of competence (Havarstein et al. 1996). Homology search over the *L. lactis*
420 genome sequence did not reveal any two-component system which resembles the one regulating
421 competence in Streptococci. However, we found an operon encoding proteins homologous to DprA or
422 Smf protein, supposed to be required for chromosomal transformation in *H. influenzae*, and DNA
423 topoisomerase I from *B. subtilis*. Interestingly, upstream of the *dprA* gene there is a sequence resembling
424 that found in the putative com-promoters, described above (Figure 11B). A small ORF between this
425 putative promoter and initiation codon of the *dprA* gene, may be a candidate for the gene of CSP
426 precursor, since the C-terminal part of the encoded peptide resembles that of CSP from Streptococci.
427 This potential CSP precursor does not have, however, a Gly-Gly pattern that is always found in the
428 Streptococcal CSP precursors (Havarstein et al.1997).

429

430 *Regulatory functions*

431 Regulation of gene expression in *L. lactis* was not yet thoroughly explored. A few examples of
432 experimental study of chromosomal gene regulation are known (Bardowski et al. 1994; Kilstrup &
433 Martinussen 1998;Luesink et al. 1998; Raya et al. 1998; van de Guchte et al. 1998; Vaughan et al. 1998;
434 Wouters et al. 1998).Automatic analysis revealed 83 genes potentially involved in regulation, which is

435 about 6% of the total number of the identified ORFs. This figure is about two times higher than that for
436 *M. genitalium* (Fraser et al. 1995), coincides with that found in *H. influenzae* (Fleischmann et al. 1995)
437 and is about two-fold lower than that found in *B. subtilis* using the same classification criteria (Lapidus
438 et al. 1997). These ratios support the hypothesis that the increase of genome size in bacteria results in
439 the proportional increase of the number of different regulatory systems.

440 Seven two-component systems were detected. They are all organised in operons, consisting of a cognate
441 sensor kinase and a response regulator. The genes of five two-component systems were cloned recently
442 from the strain MG1363 by complementation of *E. coli* mutants (O'Connell-Motherway et al. 1997). All
443 of them were identified in our sequence.

444 An interesting class of regulatory genes are the RAS-like GTP-binding proteins. Several such regulators
445 were found in other bacteria, including the most intensively studied Era protein of *E. coli*, Obg of
446 *B. subtilis* and several others, usually revealed by genome sequencing. Exact biological functions of most
447 of these proteins are unknown, although they are sometimes essential and have pleiotropic functions.
448 Interestingly, although all these proteins are homologous and can be considered as a family of paralogs.
449 Each of the nine *L. lactis* proteins has its clear counterpart in *B. subtilis* or other bacteria. This suggests
450 that each of these proteins may have a unique biological function in the cell and that revealing it in one
451 microorganism will be directly applicable to the others. It would be interesting to test whether the
452 counterparts of the *B. subtilis* genes from *L. lactis* can complement the corresponding mutations in *B.*
453 *subtilis* and vice versa. A gene encoding the extra cytoplasmic function (ECF) σ -factor was detected in
454 *L. lactis*. This finding can be interesting for studies of regulation of stress response in *L. lactis*. It is not
455 clear if any of the 11 ECF σ -factors from *B. subtilis* has the same biological function as the one detected
456 in *L. lactis*. The most similar proteins are SigV, SigX and YlaC, but the biological function of none of
457 them has yet been established in *B. subtilis*.

458

459 *Replication, transcription and translation*

460 Seventy-seven, 30 and 134 genes were assigned to replication, transcription and translation functional
461 categories, respectively. It is apparent that the replication system of *L. lactis* is very similar to that of
462 *B. subtilis*. Counterparts of *dnaB* and *dnaD* genes, essential for DNA replication in *B. subtilis* and not
463 present in gram negative bacteria were detected. Two DNA polymerase III β -chain genes, one

464 corresponding to polC and another to dnaE of B. subtilis, were also detected in L. lactis. E. coli has only
465 the latter gene. The transcription and translation machinery does not seem to have any remarkable
466 difference compared to B. subtilis. Presumably B. subtilis with its well-developed genetic tools can be
467 a convenient host organism to study gene regulation of L. lactis systems.

468

469 **Conclusions and perspectives**

470 We present here the data on a diagnostic sequencing study of the L. lactis IL1403 genome. The
471 sequencing data with a low average redundancy (about 2) allowed to reveal 1495 proteins, encoded by
472 the L. lactis genome, and to classify them by functions predicted on the basis of homology to known
473 proteins. Some statistical properties of the L. lactis genome sequence were also calculated.

474 The data presented here allow to evaluate L. lactis as a model organism for studies of bacterial
475 metabolism, physiology and regulation. Compared to the well established models, like E. coli or B.
476 subtilis, L.lactis represents a simplified system and is, therefore, potentially more amenable for detailed
477 understanding. The absence of numerous genes, relative to E. coli or B. subtilis, should facilitate, in the
478 case of L. lactis, detecting of phenotypes associated with genes of unknown function. With regard to
479 even more simple microorganisms, such as Mycoplasma, the advantage of L. lactis is the ease of
480 manipulation and the absence of any pathogenicity for humans. These properties, with the complete
481 genome information available, place L.lactis among the best model bacteria, to be used for fundamental
482 biological research.

483

484 **Acknowledgements**

485 We are grateful to the L. lactis community of this lab, and especially to the team leaders Marie-Christine
486 and Alain Chopin, Emmanuelle Maguin, Pierre Renault and Alexandra Gruss for the interest to this
487 work and stimulating discussions. Pierre Renault, Patrick Duwat, Meriem El Karoui, Isabelle Poquet,
488 Catherine Schouler are gratefully acknowledged for the use of their unpublished sequence data in the
489 early stages of this project. We thank Philippe Bessieres, Franck Samson, Véronique Biaudet-Brunaud
490 and Nicole Vermeire (INRA, Jouy en Josas), and Gregory Kucherov⁷⁴ and Vladimir Grebinskii (INRIA,
491 Nancy) for the discussions and advice on programming and computing.

492

493 **References**

- 494 Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local alignment search tool. *J.*
495 *Mol. Biol.* 215(3): 403–410 Andersen PS, Martinussen J & Hammer K (1996) Sequence analysis and
496 identification of the pyrKDbF operon from *Lactococcus lactis* including a novel gene, pyrK, involved
497 in pyrimidine biosynthesis. *J Bacteriol* 178(16): 5005–5012
- 498 Anderson AW & Elliker PR (1953) The nutritional requirements of lactic streptococci isolated from
499 starter cultures. I. Growth in a synthetic medium. *J Dairy Science* 36: 161–167
- 500 Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UC, Podowski RM,
501 Näslund AK, Eriksson A-S, Winkler HH & Kurland CG (1998) The genome sequence of *Rickettsia*
502 *proWazekii* and the origin of mitochondria. *Nature* 396:133–140
- 503 Bardowski J, Ehrlich SD & Chopin A (1992) Tryptophan biosynthesis genes in *Lactococcus lactis*
504 subsp. *lactis*. *J. Bacteriol.* 174(20): 6563–7650
- 505 Bardowski J, Ehrlich SD & Chopin A (1994) BglR protein, which belongs to the BglG family of
506 transcriptional antiterminators, is involved in beta-glucoside utilization in *Lactococcus lactis*.
507 *J. Bacteriol.* 176(18): 5681–5685
- 508 Beresford T & Condon S (1991) Cloning and partial characterization of genes for ribosomal ribonucleic
509 acid in *Lactococcus lactis* subsp. *lactis*. *FEMS Microbiol. Lett.* 62(2–3): 319–323
- 510 Biaudet V, el Karoui M & Gruss A (1998) Codon usage can explain GT-rich islands surrounding Chi
511 sites on the *Escherichia coli* genome. *Mol. Microbiol.* 29(2): 666–669
- 512 Biswas I, Maguin E, Ehrlich SD & Gruss A (1995) A 7-base pair sequence protects DNA from
513 exonucleolytic degradation in *Lactococcus lactis*. *Proc. Natl. Acad. Sci. USA* 92(6): 2244–2248
- 514 Bolhuis A, Broekhuizen CP, Sorokin A, van Roosmalen ML, Venema G, Bron S, Quax WJ & van Dijk
515 JM (1998) SecDF of *Bacillus subtilis*, a molecular Siamese twin required for the efficient secretion of
516 proteins. *J. Biol. Chem.* 273(33): 21217–21224
- 517 Bolotin A, Sorokin A & Ehrlich SD (1996) Mapping of the 150 kb spoIIIC-pheA region of the *Bacillus*
518 *subtilis* chromosome using Long Accurate PCR and three yeast artificial chromosomes. *Microbiology*
519 142: 3017–3020

520 Borodovsky M & McIninch J (1993) GENMARK: a parallel gene recognition for both DNA strands.
521 *Comput. Chem.* 17: 123–133

522 Burland V, Plunkett G, Daniels DL & Blattner F R (1993) DNA sequence and analysis of 136 kilobases
523 of the *Escherichia coli* genome: organizational symmetry around the origin of replication. *Genomics*
524 16(3): 551–561

525 Campbell EA, Choi SY & Masure HR (1998) A competence regulon in *Streptococcus pneumoniae*
526 revealed by genomic analysis. *Mol. Microbiol.* 27(5): 929–939

527 Chedin F, Noirot P, Biaudet V & Ehrlich SD (1998) A five nucleotide sequence protects DNA from
528 exonucleolytic degradation by AddAB, the RecBCD analogue of *Bacillus subtilis*. *Mol. Microbiol.*
529 29(6): 1369–1377

530 Cheng S, Fockler C, Barnes WM & Higuchi R (1994) Effective amplification of long targets from cloned
531 inserts and human genomic DNA. *Proc. Natl. Acad. Sci. U.S.A.* 91(12): 5695–5699

532 Chiaruttini C & Milet M (1993) Gene organization, primary structure and RNA processing analysis of
533 a ribosomal RNA operon in *Lactococcus lactis*. *J. Mol. Biol.* 230(1): 57–76

534 Chopin A, Chopin MC, Moillo-Batt A & Langella P (1984) Two plasmid-determined restriction and
535 modification systems in *Streptococcus lactis*. *Plasmid* 11(3): 260–263

536 Chopin MC, Chopin A, Rouault A & Galleron N (1989) Insertion and amplification of foreign genes in
537 the *Lactococcus lactis* subsp. *lactis* chromosome. *Appl. Environ. Microbiol.* 55(7): 1769–1774

538 Cocaign-Bousquet M, Garrigues C, Novak L, Lindley ND & Loubiere P (1995) Rational development
539 of a simple synthetic medium for the sustained growth of *Lactococcus lactis*. *J. Appl. Bacteriol.* 79: 108–
540 116

541 Cocaign-Bousquet M, Garrigues C, Loubiere P & Lindley ND (1996) Physiology of pyruvate
542 metabolism in *Lactococcus lactis*. *Antonie Van Leeuwenhoek* 70(2–4): 253–267

543 Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S & Barry
544 C.E. et al (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
545 sequence. *Nature* 393(6685): 537–544

546 Delorme C, Ehrlich SD & Renault P (1992) Histidine biosynthesis genes in *Lactococcus lactis* subsp.
547 *lactis*. *J. Bacteriol.* 174(20): 6571–6579

548 Delorme C, Godon J-J, Ehrlich SD & Renault P (1993) Gene inactivation in *Lactococcus lactis*:
549 Histidine biosynthesis. *J. Bacteriol.* 175(14): 4391–4399

550 De Vos WM & Vaughan EE (1994) Genetics of lactose utilisation in lactic acid bacteria. *FEMS*
551 *Microbiol. Rev.* 15: 217–237

552 De Vos WM (1996) Metabolic engineering of sugar catabolism in lactic acid bacteria. *Antonie Van*
553 *Leeuwenhoek* 70(2–4): 223–242

554 Dickely F, Nilsson D, Hansen EB & Johansen E (1995) Isolation of *Lactococcus lactis* nonsense
555 suppressors and construction of a food-grade cloning vector. *Mol. Microbiol.* 15(5): 839–847

556 Dubnau D (1997) Binding and transport of transforming DNA by *Bacillus subtilis*: the role of type-IV
557 pilin-like proteins – *in vivo*. *Gene* 192(1): 191–198

558 Duong F & Wickner W (1997) The SecDFyajC domain of preprotein translocase controls preprotein
559 movement by regulating SecA membrane cycling. *EMBO J.* 16(16): 4871–4879

560 El Karoui M, Ehrlich D & Gruss A (1998) Identification of the lactococcal exonuclease/recombinase
561 and its modulation by the putative Chi sequence. *Proc. Natl. Acad. Sci. USA* 95(2): 626–631

562 Engelke G, Gutowski-Eckel Z, Hammelmann M & Entian KD (1992) Biosynthesis of the lantibiotic
563 nisin: genomic organization and membrane localization of the NisB protein. *Appl. Environ. Microbiol.*
564 58(11): 3730–3743

565 Errington J (1986) A general method for fusion of the *Escherichia coli* lacZ gene to chromosomal genes
566 in *Bacillus subtilis*. *J. Gen. Microbiol.* 132: 2953–2966

567 Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF,
568 Dougherty BA, Merrick JM *et al.* (1995) Whole-genome random sequencing and assembly of
569 *Haemophilus influenzae* Rd. *Science* 269(5223): 496–512

570 Gasson MJ (1983) Plasmid complements of *Streptococcus lactis* NCDO 712 and other lactic
571 streptococci after protoplast-induced curing. *J. Bacteriol.* 154(1): 1–9

572 Godon JJ, Chopin MC & Ehrlich SD (1992) Branched-chain amino acid biosynthesis genes in
573 *Lactococcus lactis* subsp. *lactis*. *J. Bacteriol.* 174(20): 6580–6589

574 Godon JJ, Delorme C, Bardowski J, Chopin MC, Ehrlich SD & Renault P (1993) Gene inactivation in
575 *Lactococcus lactis*: Branched-chain amino acid biosynthesis. *J. Bacteriol.* 175(14): 4383–4390

576 Griffin HG & Gasson MJ (1995) Genetic aspects of aromatic amino acid biosynthesis in *Lactococcus*
577 *lactis*. *Mol. Gen. Genet.* 246(1):119–127

578 Grossiord B, Vaughan EE, Luesink E & de Vos WM (1998) Genetics of galactose utilisation via the
579 Leloir pathway in lactic acid bacteria. *Le Lait* 78: 77–84

580 Hahn J, Luttinger A & Dubnau D (1996) Regulatory inputs for the synthesis of ComK, the competence
581 transcription factor of *Bacillus subtilis*. *Mol. Microbiol.* 21(4): 763–775

582 Havarstein LS, Coomaraswamy G & Morrison DA (1995) An unmodified heptadecapeptide pheromone
583 induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci.*
584 *USA* 92(24): 11140–11144

585 Havarstein LS, Gaustad P, Nes IF & Morrison DA (1996) Identification of the streptococcal
586 competence-pheromone receptor. *Mol. Microbiol.* 21(4): 863–869

587 Havarstein LS, Hakenbeck R & Ganstad P (1997) Natural competence in the genus *Streptococcus*:
588 Evidence that *Streptococci* can change phenotype by interspecies recombinational exchanges.
589 *J. Bacteriol.* 179: 6589–6594

590 Hemila H, Pakkanen R, Heikinheimo R, Palva ET & Palva I (1992) Expression of the *Erwinia carotovora*
591 a polygalacturonase encoding gene in *Bacillus subtilis*: role of signal peptide fusions on production of a
592 heterologous protein. *Gene* 116(1): 27–33

593 Huang DC, Novel M & Novel G (1991) A transposon-like element on the lactose plasmid of
594 *Lactococcus lactis* subsp. *lactis* Z270. *FEMS Microbiol. Lett.* 61(1): 101–106

595 Hutkins RW, Morris HA & McKay LL (1985) Galactokinase activity in *Streptococcus thermophilus*.
596 *Appl. Env. Microbiol.* 50:777–780

597 Jensen PR & Hammer K (1993) Minimal requirements for exponential growth of *Lactococcus lactis*.
598 *Appl. Env. Microbiol.* 59:4363–4366

599 Kilstrup M & Martinussen J (1998) A transcriptional activator, homologous to the *Bacillus subtilis* PurR
600 repressor, is required for expression of purine biosynthetic genes in *Lactococcus lactis*. *J. Bacteriol.*
601 180(15): 3907–3916

602 Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G et al. (1997) The complete genome sequence
603 of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 390: 249–256

604 Lander ES & Waterman MS (1988) Genomic mapping by fingerprinting random clones: a mathematical
605 analysis. *Genomics* 2(3):231–239

606 Lapidus A, Galleron N, Sorokin A & Ehrlich SD (1997) Sequencing and functional annotation of the
607 *Bacillus subtilis* genes in the 200kb *rrnB-dnaB* region. *Microbiology* 143: 3431–3441

608 Lapujade P, Coccagn-Bousquet M & Loubiere P (1998) Glutamate biosynthesis in *Lactococcus lactis*
609 subsp *lactis* NCDO 2118. *Appl. Environ. Microbiol* 64: 2485–2489

610 Le Bourgeois P, Lautier M, Mata M & Ritzenthaler P (1992) Physical and genetic map of the
611 chromosome of *Lactococcus lactis* subsp. *lactis* IL1403. *J. Bacteriol.* 174(21): 6752–6762

612 Le Bourgeois P, Lautier M, van den Berghe L, Gasson MJ & Ritzenthaler P (1995) Physical and genetic
613 map of the *Lactococcus lactis* subsp. *cremoris* MG1363 chromosome: comparison with that of
614 *Lactococcus lactis* subsp. *lactis* IL 1403 reveals a large genome inversion. *J. Bacteriol.* 177(10): 2840–
615 2850

616 Le Loir Y, Gruss A, Ehrlich SD & Langella P (1998) A nine residue synthetic propeptide enhances
617 secretion efficiency of heterologous proteins in *Lactococcus lactis*. *J. Bacteriol.* 180(7):1895–1903

618 Lobry JR (1996a) Origin of replication of *Mycoplasma genitalium*. *Nature* 272: 745–746

619 Lobry JR (1996b) Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.*
620 13: 660–665

621 Lowe TM & Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes
622 in genomic sequence. *Nucleic Acids Res.* 25(5): 955–964

623 Luesink EJ, van Herpen REMA, Grossiord BP, Kuipers OP & deVos WM (1998) Transcriptional
624 activation of the glycolytic *Las* operon and catabolite repression of the *gal* operon in *Lactococcus lactis*
625 are mediated by the catabolite control protein CcpA. *Mol. Microbiol.* 30(4): 789–798

626 Madsen SM, Albrechtsen B, Hansen EB & Israelsen H (1996) Cloning and transcriptional analysis of
627 two threonine biosynthetic genes from *Lactococcus lactis* MG1614. *J. Bacteriol.* 178(13):3689–3694

628 Maeda S & Gasson MJ (1986) Cloning, expression and location of the *Streptococcus lactis* gene for
629 phospho-beta-D-galactosidase. *J. Gen. Microbiol.* 132 (2): 331–340

630 Martinussen J, Andersen PS & Hammer K (1994) Nucleotide metabolism in *Lactococcus lactis*: salvage
631 pathways of exogenous pyrimidines. *J. Bacteriol.* 176(5): 1514–1516

632 Moll I, Resch A & Blasi U (1998) Discrimination of 50-terminalstart codons by translation initiation
633 factor 3 is mediated by ribosomal protein S1. *FEBS Lett.* 436(2): 213–217

634 Nilsson D & Johansen E (1994) A conserved sequence in tRNA and rRNA promoters of *Lactococcus*
635 *lactis*. *Biochim Biophys Acta* 1219(1): 141–144

636 Nilsson D & Kilstrup M (1998) Cloning and expression of the *Lactococcus lactis* purDEK genes,
637 required for growth in milk. *Appl. Environ. Microbiol.* 64: 4321–4327

638 Novak L, Cocaign-Bousquet M, Lindley ND & Loubiere P (1997) Metabolism and energetics of
639 *Lactococcus lactis* during growth in complex or synthetic media. *Appl. Environ. Microbiol.* 63:2665–
640 2670

641 Novak L, Cocaign-Bousquet M, Lindley ND & Loubiere P (1998) Cométabolisme sucre-acides aminés
642 chez *Lactococcus lactis* (inFrench). *Le Lait* 78: 17–22

643 O’Connell-Motherway M, Fitzgerald GF & van Sinderen D (1997) Cloning and sequence analysis of
644 putative histidine protein kinases isolated from *Lactococcus lactis* MG1363. *Appl. Environ. Microbiol.*
645 63(6): 2454–2459

646 Pearson WR, Wood T, Zhang Z & Miller W (1997) Comparison of DNA sequences with protein
647 sequences. *Genomics* 46(1): 24–36 Pogliano JA & Beckwith J (1994) SecD and SecF facilitate protein
648 export in *Escherichia coli*. *EMBO J.* 13(3): 554–561

649 Polzin KM, Romero D, Shimizu-Kadota M, Klaenhammer TR & McKay LL (1993) Copy number and
650 location of insertion sequences ISS1 and IS981 in lactococci and several other lactic acid bacteria. *J.*
651 *Dairy Sci.* 76(5): 1243–1252

652 Poolman B (1993) Energy transduction in lactic acid bacteria. *FEMS Microbiol. Rev.* 12(1-3): 125–147

653 Pouwels PH, Leer RJ, Shaw M, Heijne den Bak-Glashouwer MJ, Tielen FD, Smit E, Martinez B, Jore J
654 & Conway PL (1998) Lactic acid bacteria as antigen delivery vehicles for oral immunization purposes.
655 *Int. J. Food. Microbiol.* 41(2): 155–167

656 Raya R, Bardowski J, Andersen PS, Ehrlich SD & Chopin A (1998) Multiple transcriptional control of
657 the *Lactococcus lactis* trp operon. *J. Bacteriol.* 180(12): 3174–3180

658 Resch A, Tedin K, Grundling A, Mundlein A & Blasi U (1996) Downstream box-anti-downstream box
659 interactions are dispensable for translation initiation of leaderless mRNAs. *EMBO J.* 15(17): 4740–4748

660 Riley M (1993) Functions of the gene products of *Escherichia coli*. *Microbiol. Rev.* 57(4): 862–952

661 Schouler C, Gautier M, Ehrlich SD & Chopin MC (1998) Combinational variation of restriction
662 modification specificities in *Lactococcus lactis*. *Mol. Microbiol.* 28(1): 169–178

663 Simon D, Rouault A & Chopin MC (1985) Protoplast transformation of group N streptococci with
664 cryptic plasmids. *FEMSMicrobiol. Lett.* 26: 239–241

665 Sorokin A, Lapidus A, Capuano V, Galleron N, Pujic P & Ehrlich SD (1996) A new approach using
666 multiplex long accurate PCR and yeast artificial chromosomes for bacterial chromosome mapping and
667 sequencing. *Genome Res.* 6(5): 448–453

668 Stinglele F, Neeser JR & Mollet B (1996) Identification and characterization of the *eps*
669 (Exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6. *J. Bacteriol.* 178(6): 1680–
670 1690

671 Tedin K, Rosch A & Blösi U (1997) Requirements for ribosomal protein S1 for translation initiation of
672 mRNAs with and without a 5' leader sequence. *Mol. Microbiol.* 25(1): 189–199

673 Thomas TD & Craw VF (1984) Selection of galactose-fermenting *Streptococcus thermophilus* in
674 lactose-limited chemostat cultures. *Appl. Env. Microbiol.* 48: 186–191

675 Tracy RB, Chedin F & Kowalczykowski SC (1997) The recombination hot spot *chi* is embedded within
676 islands of preferred DNA pairing sequences in the *E. coli* genome. *Cell* 90(2): 205–206

677 Tulloch DL, Finch LR, Hillier AJ & Davidson BE (1991) Physical map of the chromosome of
678 *Lactococcus lactis* subsp. *Lactis* DL11 and localization of six putative rRNA operons. *J.*
679 *Bacteriol.* 173(9): 2768–7275

680 Van de Guchte M, Ehrlich DS & Chopin A (1998) tRNA^{Trp} as a key element of antitermination in the
681 *Lactococcus lactis* *trp* operon. *Mol. Microbiol.* 29(1): 61–74

682 Van Etten WJ & Janssen GR (1998) An AUG initiation codon, not codon-anticodon complementarity,
683 is required for the translation of unleadered mRNA in *Escherichia coli*. *Mol. Microbiol.* 27(5):987–1001

684 Van Kranenburg R, Marugg JD, van Swam II, Willem NJ & de Vos WM (1997) Molecular
685 characterization of the plasmid-encoded *eps* gene cluster essential for exopolysaccharide biosynthesis
686 in *Lactococcus lactis*. *Mol. Microbiol.* 24(2): 387–397

687 Van Sinderen D, Karsens H, Kok J, Terpstra P, Ruiters MH, VenemaG & Nauta A (1996) Sequence
688 analysis and molecular characterization of the temperate lactococcal bacteriophage *r1t*. *Mol. Microbiol.*
689 19(6): 1343–1355

690 Van Sinderen D, Luttinger A, Kong L, Dubnau D, Venema G & Hamoen L (1995) comK encodes the
 691 competence transcription factor, the key regulatory protein for competence development in *Bacillus*
 692 *subtilis*. *Mol. Microbiol.* 15(3): 455–462

693 Vaughan EE & de Vos WM (1995) Identification and characterization of the insertion element IS1070
 694 from *Leuconostoc lactis* NZ6009. *Gene* 155 (1): 95–100

695 Vaughan EE, Pridmore RD & Mollet B (1998) Transcriptional regulation and evolution of lactose genes
 696 in the galactose-lactose operon of *Lactococcus lactis* NCDO 2054. *J. Bacteriol.* 180(18):4893–4902

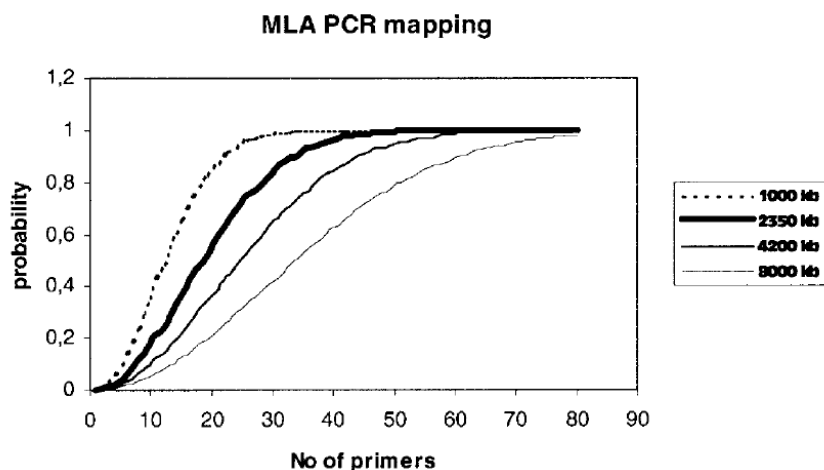
697 Vellanoweth RL & Rabinowitz JC (1992) The influence of ribosome-binding-site elements on
 698 translational efficiency in *Bacillus subtilis* and *Escherichia coli* in vivo. *Mol. Microbiol.* 6(9):1105–
 699 1114

700 Venter JC, Smith HO & Hood L (1996) A new strategy for genome sequencing. *Nature* 381: 364–
 701 366

701 Wells JM, Wilson PW, Norton PM & Le Page RW (1993) A model system for the investigation of
 702 heterologous protein secretion pathways in *Lactococcus lactis*. *Appl. Environ. Microbiol.* 59(11): 3954–
 703 3959

704 Wouters JA, Sanders JW, Kok J, de Vos WM, Kuipers OP & Abee T (1998) Clustered organization and
 705 transcriptional analysis of a family of five *csp* genes of *Lactococcus lactis* MG1363. *Microbiology* 144
 706 (10): 2885–2893

707



708 *Figure 1.* Probability of successful amplification in MLA PCR as a function of the number of oligonucleotides used in the same mixture. The calculation was done by generating the results of 10,000 independent random experiments and averaging these results. The longest distance between primers still allowing the amplification was supposed to be 20 kb. The genome sizes used for calculation are: 1 Mb (dashed line); 2.35 Mb (bold line); 4.2 Mb (thin line); 8 Mb (dotted line).

A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	+	+	+	+	+	+	+	+								
2		+	+	+	+	+	+	+	+							
3			+	+	+	+	+	+	+	+						
4				+	+	+	+	+	+	+	+					
5					+	+	+	+	+	+	+	+				
6						+	+	+	+	+	+	+	+			
7							+	+	+	+	+	+	+	+		
8								+	+	+	+	+	+	+	+	
9									+	+	+	+	+	+	+	+
10	+									+	+	+	+	+	+	+
11	+	+									+	+	+	+	+	+
12	+	+	+									+	+	+	+	+
13	+	+	+	+									+	+	+	+
14	+	+	+	+	+									+	+	+
15	+	+	+	+	+	+									+	+
16	+	+	+	+	+	+	+									+
Product 1x8	n	n	n	n	n	n	n	y	n	n	n	n	n	n	n	n
Product 3x4	n	n	n	y	y	y	y	y	y	y	n	n	n	n	n	n
Product 12x15	y	y	y	n	n	n	n	n	n	n	n	n	n	n	y	y

B

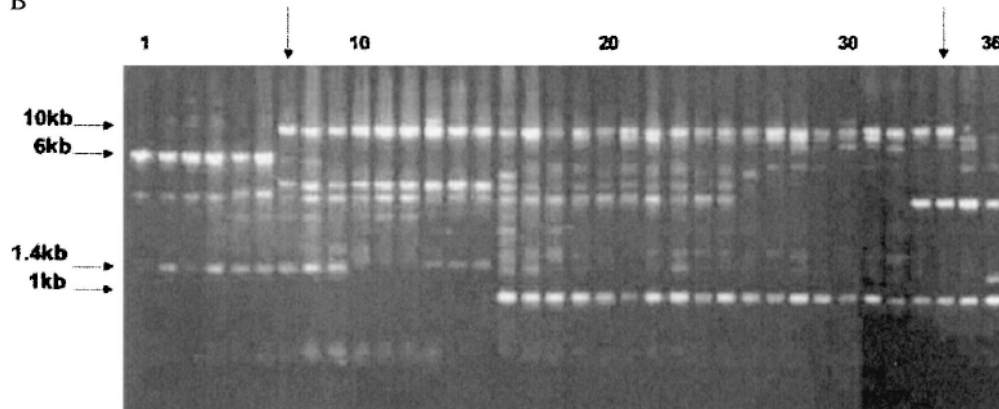


Figure 2. MLA PCR procedure. (A) The table illustrates a theoretical experiment with a total of 16 (M) primers, where each reaction mixture contains 8 (N) primers. The numbers at the top designate reaction mixtures, the numbers on the left designate individual primers, the '+' sign indicates presence of a given primer in a reaction mixture. The three last rows indicate the reaction mixtures in which amplification does (y) or does not (n) take place, if the primer couples 1 and 8, 3 and 4 or 12 and 15, respectively, give a product. To identify the primers which give a product from the analysis of the reactions the following rule is used: If N_1 and N_2 are the first and the last positive reaction, respectively, the primers giving the product will be: $P_1 = N_1$; $P_2 = N_2 - N + 1$, if $N_2 > N_1$ (the cases 1x8 and 3x4), or $P_2 = M + N_2 - N + 1$, if $N_2 < N_1$ (the case 12x15). (B) Part of a real experiment with 96 primers, where each reaction contained 32 primers and *L. lactis* IL1403 chromosome as template. Sizes and positions of identified fragments are shown on the left. Numbers on the top designate the reaction mix. Arrows on the top indicate the reactions that are used to identify the primers which gave the 10 kb fragment. These primers are $P_1 = N_1 = 7$, and $P_2 = N_2 - N + 1 = 34 - 32 + 1 = 3$.

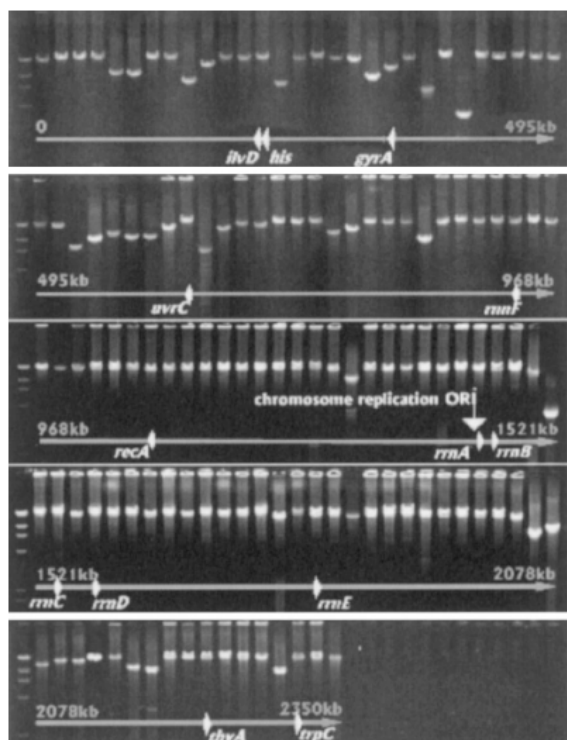


Figure 3. Whole-genome PCR amplification of the *L. lactis* IL1403 chromosome. 266 oligonucleotides were used in 133 LR PCR reactions to amplify the entire genome. Correspondence of the sizes of amplified fragments and the physical map predicted from the sequence (Figure 9) indicates the correctness of sequence assembling. Left slot on each panel shows size marker (DNA of phage λ cut by *Hind*III). The PCR product designations correspond to the ones shown on Figure 9. The first panel from the top, left to the right: BA7, BA8, BB1, BB2, CC1, CC3, BB3, BB4, CC5, CC6, BB5, BB6, BB7, BA1, CB8, BB8, BC1, BC2, CL1, CL2, BC4, BA2, BC5, BC6, BC7, CB3, BC8, BD1, BD2; second panel: BD3, CB4, CC7, CC8, CD1, CD2, CD3, CG4, BD4, CD4, cg7, BD5, BD6, BD7, BD8, BE1, CH4, CE1, BE2, BE3, BE4, BE5, BE6, BE7, BE8, BF1, BF2, BF3, BF4; third panel: BF5, BA3, BF6, BF7, BF8, BG1, BG2, BG3, BG4, BG5, BG6, BG7, BG8, BH1, BH2, BH3, BH4, BA4, BH5, BH6, BH7, BH8, CB5, BI1, BI2, BI3, BI4, DA1, DA2; fourth panel: BI5, BI6, BI7, BI8, BJ1, BJ2, BJ3, BJ4, BJ5, BJ6, BJ7, BJ8, BK1, CB7, BK2, BK3, BK4, BA5, BK5, BK6, BK7, BK8, BL1, BL2, BL3, BL4, BL5, BL6, CE3; fifth panel: CE4, CK5, CK7, BA6, BL8, CE7, CE8, CA1, CA2, CA3, CA4, CA5, CA6, CA7, CA8, CB1, CF3. The sequences of the used oligonucleotides can be found at <http://spock.jouy.inra.fr>. Chromosome drawing, which is complementary to the one shown on Figure 9, approximately corresponds to the PCR fragments above it.

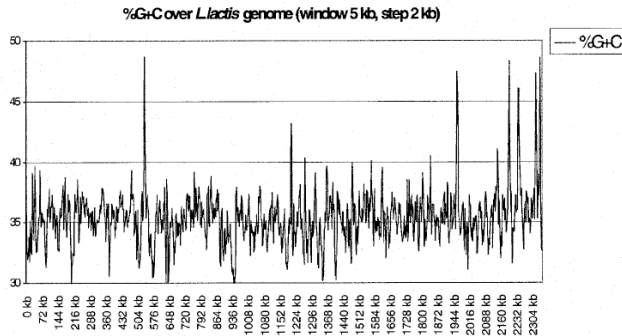
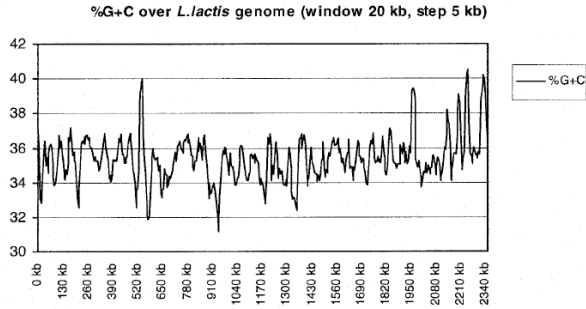


Figure 4. Distribution of the GC-content over the *L. lactis* chromosome. Zero point corresponds to the position of the origin of replication. Top panel: a sliding window of 20 kb and a step of 5 kb were used. Bottom panel: a sliding window of 5 kb and the step of 2 kb were used. The peaks near coordinates 540, 1970, 2210, 2250, 2325 and 2340 correspond to rRNA operons. A peak near 2150 kb corresponds to the cluster of ribosomal protein genes.

711

712

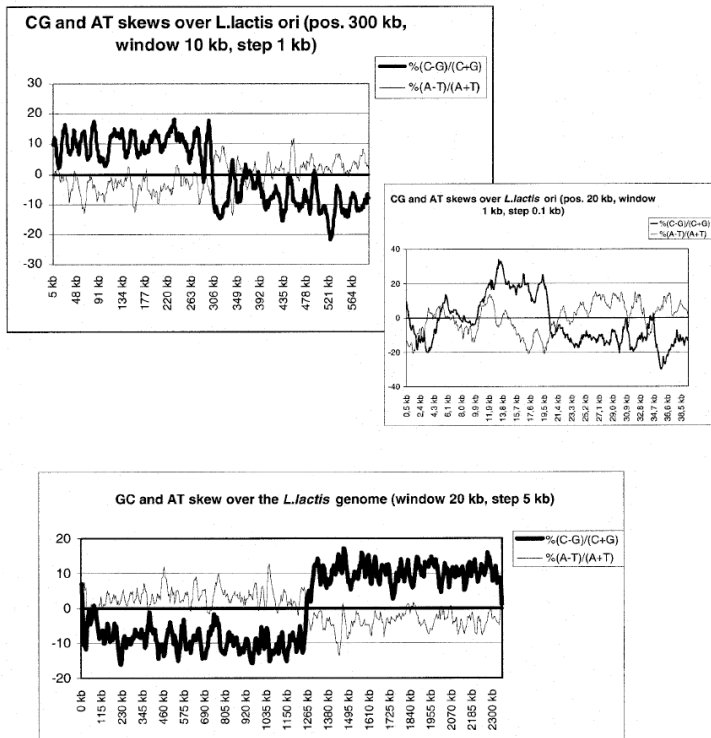


Figure 5. GC and AT skew distribution over the *L. lactis* genome. Top panel: a sliding window of 10 kb and step of 1 kb were used. Inset: a sliding window of 1 kb and step of 0.1 kb, were used. In both cases a region near the *dnaA* gene was analysed. Change of the skew distribution sign indicates the position of replication origin. Bottom panel: a sliding window 20 kb and a step of 5 kb were used. Zero coordinate is placed at the replication origin. The whole genome was analysed. Change of the skew distribution sign indicates the position of replication terminus.

713

714

```

tattaggacgcccgcagccagtcgggtacaagtaccacaccttcagcgcctcaacttgcattg 2344620
ataatccgtgcccggcgtgggtcagccatgttctatgggtggaaagtcgaggttgaactgtac
cgttcgtccctggccagcagatcaaacctcaaaataaagtatttaagtcaccttagtgactg 2344680
gcaagcaggactcgtctcctagtttgaagatttatttcataaactcagtggaatcaactgac
<-----| 16S rRNA rrrA
gcttggcttccattatgattgacagatttttgaatgcttctactacatcacatgagctca 2344740
cgaacagaaagttaaacactaactgctcaaaaactacatagaatgattgtagtactcagt
ttctcttttattcagtttcaaatggctagcttatcgtctcaagccctcgtctctcttggc 2344800
aagaagaaatagtcacaaagtaccagatcgaatagcgaagtctcgggagcagagaacgc
acaactattatctctatcaaacctaacctaacctgtcaagaattactgtgactlltat 2344860
tgttgataataatagatagttcggattgattgacagttcttaattgacactgaaaata
accttttttattctagcttctctctctctctctctctctctctctctctctctctctct 2344920
tggaaaaaatagaaatgcaagcgaaaaaaggagggacataaagcgaataaagtcggtg
cactagctttggtaattaaaaaaatagctcttttttaactgtctttcttttttttttaa 2344980
gtgatcaaaccttaatttttttaacgaaaaaattgaaagaaagaaaaaatt
taagataaggtgcttatttagtctctctctctctctctctctctctctctctctctctct 2345040
attctattccagaaataaacgagaagaaatggaatactcttatttttaagaactaaaaa
acgtagaatagctgttttaaacctaacctaaattttttataaastgtagaataatgaa 2345100
tgcattctactacaaaaattgaattgattaaaaataatattatcatcttttaactt
ttctcttttatttattgcaaaatagttgtctctatattccgtcttataatagtagaAAG 0
aagagaaaaataaatcagtttataaacagsgatataaggcagaataatatactctTTC
HindIII
CTTaaatttaacaatcattttcaaatcttagattctattaccattagagataaaat 60
GAAttttaaattgtagataaaagtttaagaatcaagataatgtatctctattttata
ccccaaagatagaaaagactgtctaaaagctgtgtaaattttcttaattttctttt 120
gggtttctatcttllctgacagatttctgacacctttaaagaattaaaaagaaaa
tccacagtaaatataatataaattgaacttttttgattgtggataagctctgtagcat
agggtcatttattagatatttaaacgaaaaaaactaacacattctcgacatcgtga 180
DnaA box
ttgatattactcattttttctgtggataaagaaactttacacattatccacagttttg 240
aactataatagtaaaaaagacacattcttttgaantgttaantaggtgtaaac
DnaA box
ttaataacttttcttatggtgaaaaactttctttgttaattcaacagactgtgaaa 300
aattattgaaaagaatcaccctttgaaagaaacaaataaagtgtctgacacctt
DnaA box M dnaA
aatataatgtaaatgatataaaggttatcaacaactaaaaagtaaggagataatag 360
ttatattcaacttactataattactcaatagttgtttgattttctctctctctatac
A S L N E N Q K F W A R V T E L A Q Q S dnaA 420
gcattcccttaatgaaaaataaaaaatttggcccggttaaccgaactagctcgcagagt
cgtagggaattacttttagtttttaaacccggccacattggcttgatcagctcgtca
I G K Q A Y D F F I E P A Q L M S V E Q dnaA 480
atcgttaaacagcttatgatttttttattgacctgctcaatbaatgtctcgcagcaa
tagccattgttcgaactcaaaaaataactcggcaggttaattacagacagctcgtt

```

Figure 6. Sequence of the *L. lactis* chromosome near the origin of replication. The *HindIII* site shown in bold was chosen as a zero point to represent the genetic map of *L. lactis* IL1403. Putative DnaA boxes are shown in bold. The beginning of the *dnaA* gene and of the *rrnA* genes are shown. Numbering corresponds to that used in the map shown on Figure 9.

715

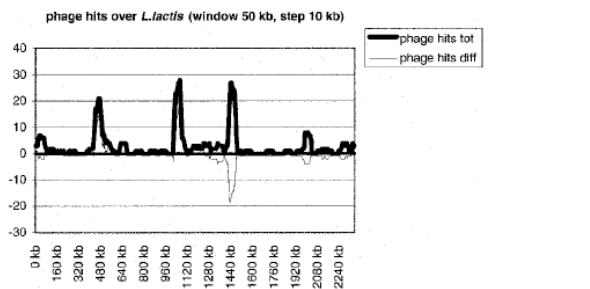


Figure 7. Identification of putative prophages in the *L. lactis* IL1403 genome. The distribution shows the number of phage protein hits detected with a sliding window of 50 kb with a step of 10 kb, over the whole genome. Total number of hits on both strands is shown by bold line. Thin line shows difference of the numbers of hits detected separately on each strand and indicates the orientation of most of the phage genes.

716

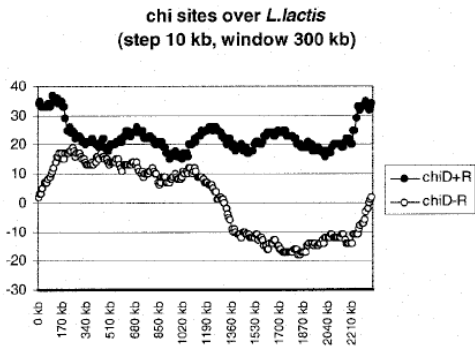

















Figure 8. Distribution of chi-sites over the *L. lactis* genome. The distribution represents the number of chi sequences detected with a sliding window of 300 kb with a step of 10 kb, over the whole genome. Filled circles show total numbers of chi-sites in the window, open circles show the difference of the numbers of sites detected separately on each strand, referred to as D and R.

717

Figure 9. Linear representation of the *L.lactis* IL1403 genome.

The map shows predicted protein-encoding genes, IS-elements, RNA genes and the ends of regions amplified by PCR (Fig.3). The genes are shown as arrows, with the best BLASTx hit SwissProt name indicated above. Colours correspond to the functional categories assigned by the automatic gene classification method described in the text: amino acid biosynthesis (), biosynthesis of cofactors (), cell envelope (), cellular processes (), central intermediary metabolism (), energy metabolism (), fatty acid and phospholipid metabolism (), purines, pyrimidines, nucleosides and nucleotides (), regulatory functions (), replication (), transcription (), translation (), transport and binding proteins (), other categories (), hypothetical and unknown (). The arrows outlined by bold line correspond to the proteins whose size differ less than 20% of their Swissprot homologue. IS-elements and rRNA genes are shown as black arrows with corresponding designation inside. Arrow directions for IS-elements are arbitrary, indicating orientations. tRNA genes are shown as small black triangles. The ends of regions amplified by PCR are indicated by open arrows below the scale line. Identifier and size of the region are indicated near the left end.

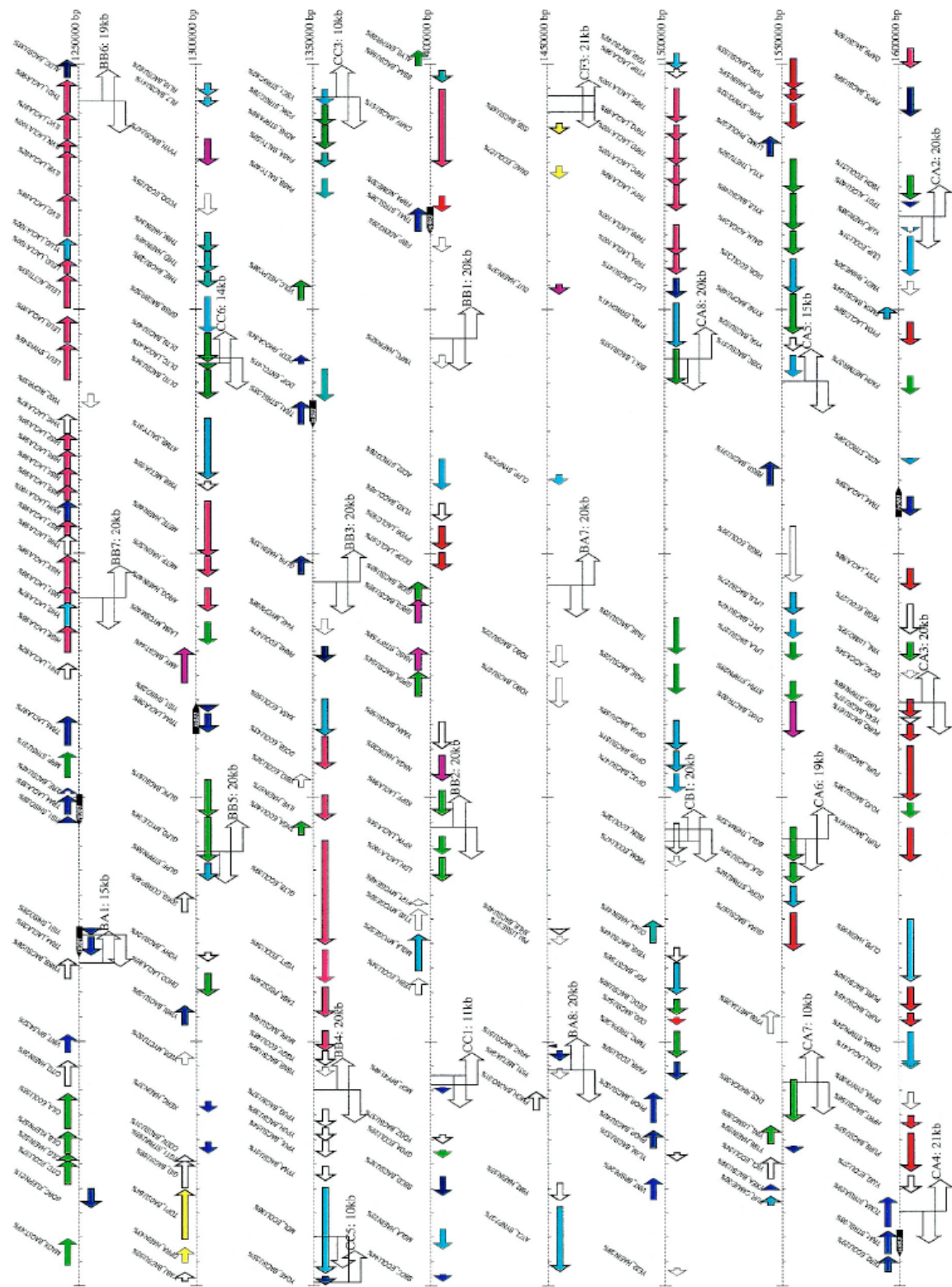


Figure 9.

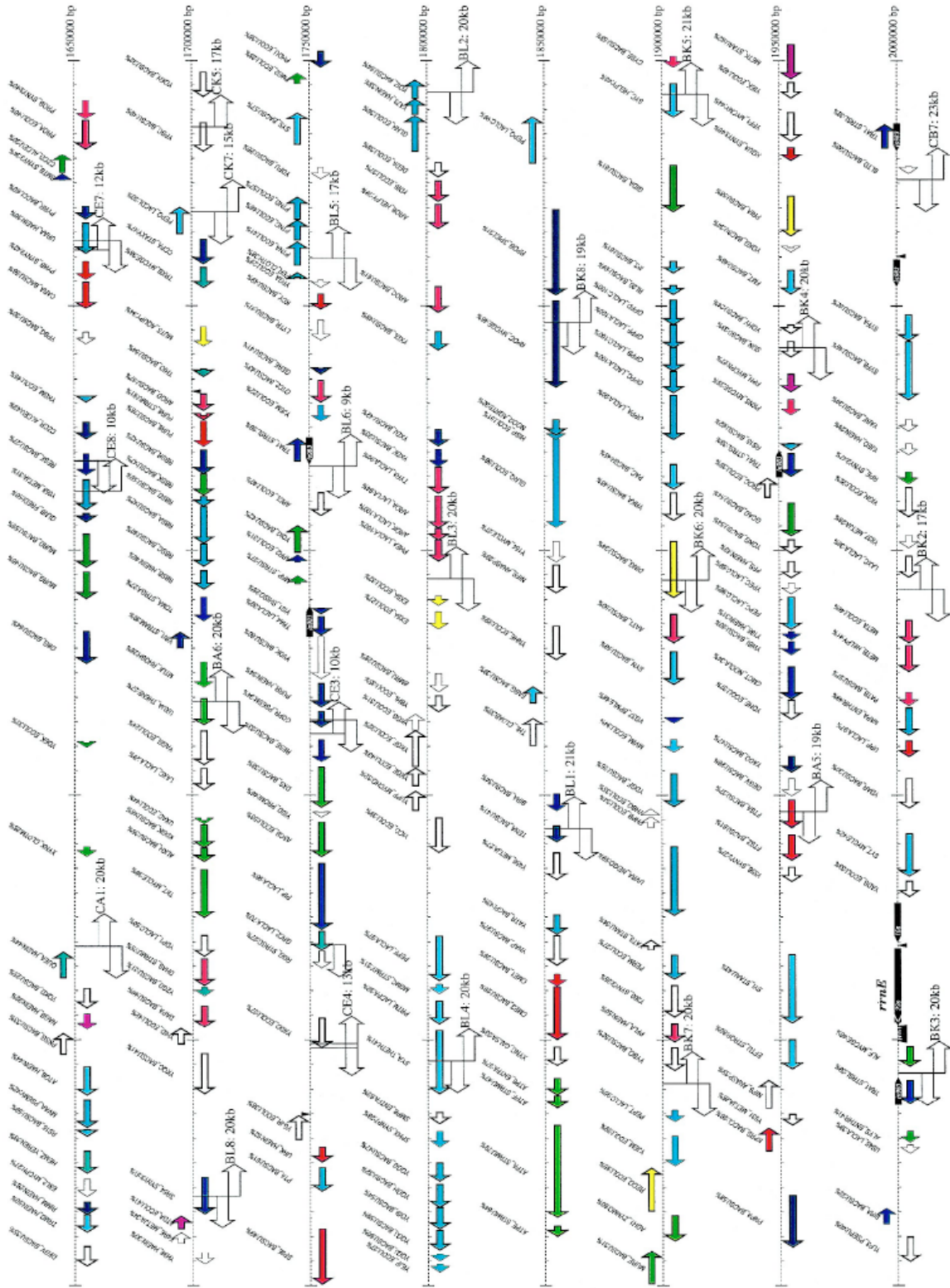


Figure 9.

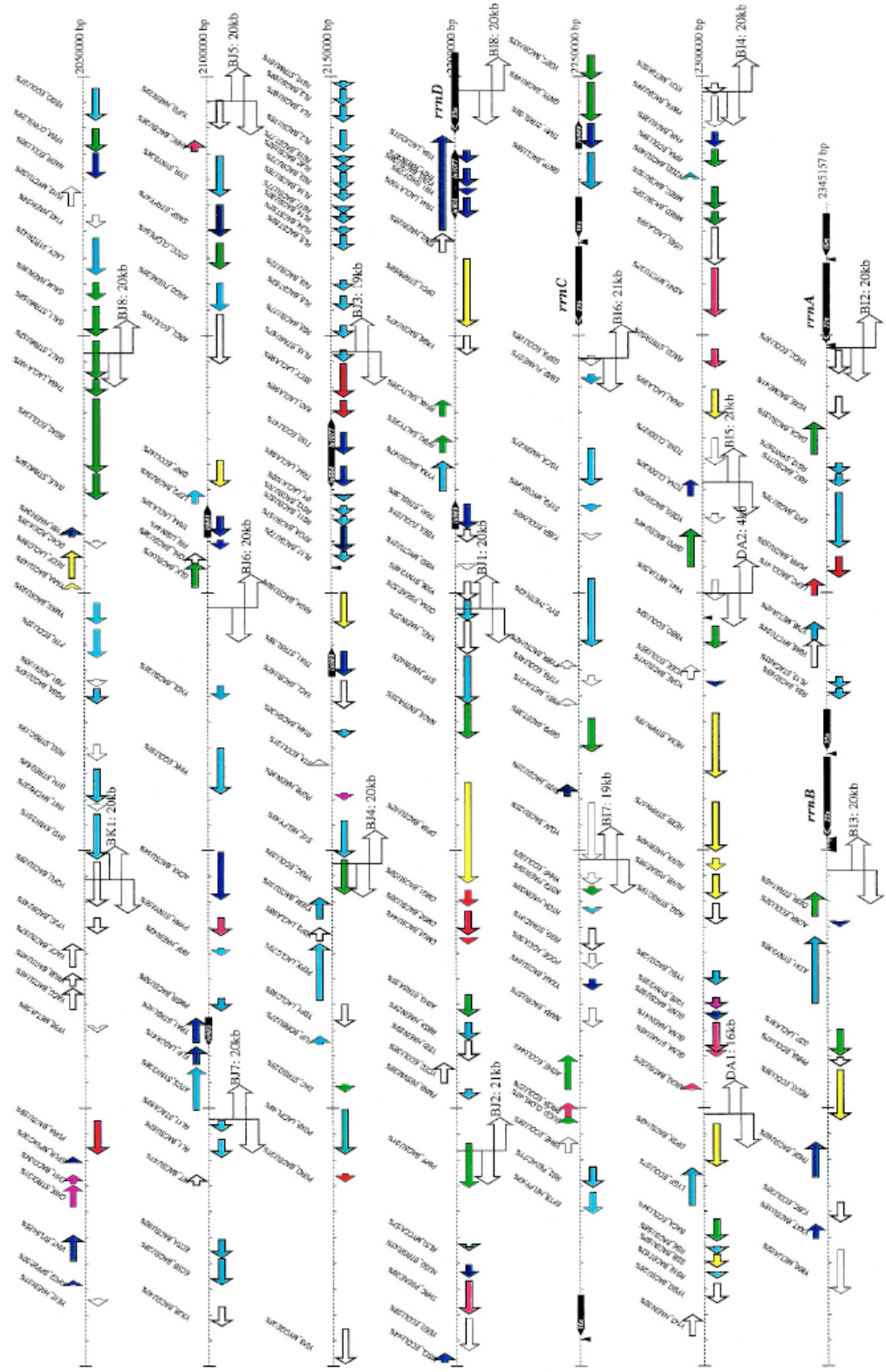


Figure 9.



Figure 9.

724

725

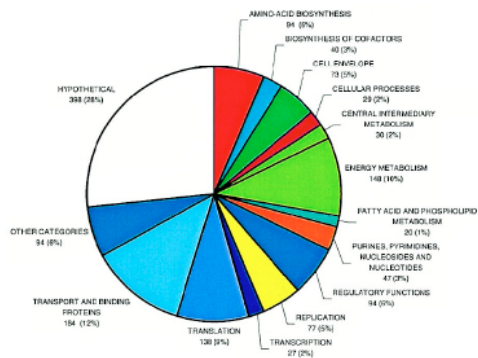


Figure 10. Distribution of the detected genes by functional categories. Colours designating different categories are defined in the legend of Figure 9. The number and the percentage of homology hits to SwissProt are indicated for each category. Classification method is presented in the text.

726

A

```

aaattttcggctttttGTTACATTaactgAttaTTTCGTATAgtct AGTATG. . . comGA
tgcatttaacaattttGTTACATTttaacAgcaTTTCGTATAgast AGTATG. . . comFA
ttataattcaactcttgGaaACITTCcaagAgttTTTCGTATAanaa AGTATG. . . comEA
tttcgctaratattaaGTTACAtTcttctATTTTCGTATAgattAGTATG. . . comC

consensus      GTTACATTnnnnAnnTTTCGTATAnnnn AGT

```

B

```

tgcagtaatttttttGTTACAagtcagtcctttTTTCGTATAgaaa Attataagccaatttttaaattg
                                     M I T N P D L Y R W K K A
ctgacaanagctgtcagtaaaaaatGGAGGaatATGatacaaatttgatttatctgttgaaaaaagct
G M T N L G V N K L L K F F H R Y D K K stop
ggcctgaccaatttaggagtaataaaccttttaaaatttttcatagatatgataaaaaTAAGtttaagac

```

Figure 11. Regulatory regions of the *L. lactis* competence genes. (A) Sequences of the regions upstream of the first gene in the four putative competence operons. Sequences corresponding to the recognition site of a hypothetical regulator are shown in bold capital letters. Start codons of the competence genes are italicised. Consensus sequence of the recognition site is shown at the bottom. (B) Sequence of the upstream region of the *dprA* gene. Putative recognition site of the hypothetical competence regulator is shown in bold capital letters. Conceptual translation of a small ORF upstream of the *dprA* gene is shown.

727

Table 1. Low-redundancy whole-genome sequencing strategy

Stage	Description
Random small insert and large insert library construction	Shearing DNA randomly to 1–2 kb and 15–20 kb, respectively. Cloning in multicopy plasmids and λ -based FIXII vector.
Random sequencing of a limited number of plasmid inserts	Sequencing of plasmid inserts up to coverage of about one. Using forward (M13-21), reverse (M13RP1) and middle primers. The latter are chosen to sequence the middle of plasmid inserts, when necessary. Assembling is done in parallel with sequencing.
Random sequencing of a number of large inserts by primer walking.	Sequencing of large inserts by using synthetic oligonucleotides. Assembling is done in parallel with sequencing to avoid too high redundancy.
Synthesis of new sequencing substrates by MLA PCR approach.	Application of Multiplex Long Accurate PCR protocol for random generation of new sequencing substrates.
Sequencing of MLA PCR products.	Sequencing of MLA PCR products by primer walking. Assembling is done in parallel with sequencing.
Iteration	Iteration of the last two stages until the whole contiguous genome sequence is obtained.
Annotation	On-line annotation is done in parallel with the sequencing progress by using automatic computer tools based on BLASTx and FASTx algorithms.

728

Table 2. Summary of low-redundancy sequencing of the *L. lactis* IL1403 genome*

Small insert plasmids	2854
Average insert length (bp)	1520
Forward sequence reactions (M13-21 primer)	2625
Average useful sequence length (bp)	462
Reverse sequence reactions (M13RP1 primer)	2168
Average useful sequence length (bp)	422
Sequencing reactions in the middle of insert	2357
Average read length (bp)	494
Total characters from plasmid insert sequencing (bp)	3292004
Number of selected λ -clones	122
Average insert length (kb)	15.6
Number of gel readings from λ -inserts	622
Average read length (bp)	409
Total characters from λ -insert reads (bp)	254398
Number of gel readings from LR PCR products	1641
Average read length (bp)	470
Total characters read from LR PCR products (bp)	771270
Total oligonucleotides used	4620
Total gel readings	10235
Average read length (bp)	458
Total characters read (bp)	4687630
Size of final contig (bp)	2345160
Undefined letters in the final contig (bp)	3563

*The figures were estimated by analysis of the assembling database and are therefore approximate ($\pm 5\%$).

Table 3. Functional classification of the *L. lactis* genes. Genes are ordered by functional categories and their positions on the chromosome. Position on the chromosome (in kb) and the percentage of identity with the most homologous Swissprot entry is also shown

BIO01. AMINO-ACID BIOSYNTHESIS			95 genes
BIO01.01 General			2 genes
DHSS_SYNP1	29%	485	SOLUBLE HYDROGENASE, SMALL SUBUNIT (EC 1.12.-.-)
PATB_BACSU	57%	1974	PUTATIVE AMINOTRANSFERASE B (EC 2.6.1.-)
BIO01.02 Aromatic amino-acid family			19 genes
AROF_ECOLI	41%	117	PHOSPHO-2-DEHYDRO-3-DEOXYHEPTONATE ALDOLASE, TYR-SENSITIVE (EC 4.1.2.15)
TRPE_THEMA	26%	736	ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27)
AROG_HAEIN	40%	1278	PHOSPHO-2-DEHYDRO-3-DEOXYHEPTONATE ALDOLASE (EC 4.1.2.15)
TRPA_LACLA	100%	1492	TRYPTOPHAN SYNTHASE ALPHA CHAIN (EC 4.2.1.20)
TRPB_LACLA	100%	1493	TRYPTOPHAN SYNTHASE BETA CHAIN (EC 4.2.1.20)
TRPF_LACLA	99%	1495	N-(5*-PHOSPHORIBOSYL)ANTHRANILATE ISOMERASE (EC 5.3.1.24)
TRPC_LACLA	100%	1495	INDOLE-3-GLYCEROL PHOSPHATE SYNTHASE (EC 4.1.1.48)
TRPD_LACLA	100%	1496	ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.18)
TRPG_LACLA	99%	1497	ANTHRANILATE SYNTHASE COMPONENT II (EC 4.1.3.27)
TRPE_LACLA	100%	1498	ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27)
AROD_BACSU	37%	1686	3-DEHYDROQUINATE DEHYDRATASE (EC 4.2.1.10)
PHEA_LACLA	100%	1780	PREPHENATE DEHYDRATASE (EC 4.2.1.51)
AROK_LACLA	100%	1781	SHIKIMATE KINASE (EC 2.7.1.71)
AROA_LACLA	94%	1782	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19)
TYRA_LACLA	90%	1783	PREPHENATE DEHYDROGENASE (EC 1.3.1.12)
AROC_BACSU	61%	1790	CHORISMATE SYNTHASE (EC 4.6.1.4)
AROB_HELPY	39%	1794	3-DEHYDROQUINATE SYNTHASE (EC 4.6.1.3)
YDIB_ECOLI	37%	1795	HYPOTHETICAL 31.2 KD PROTEIN IN LPP-AROD INTERGENIC REGION
AROC_BACSU	32%	2261	PHOSPHO-2-DEHYDRO-3-DEOXYHEPTONATE ALDOLASE (EC 4.1.2.15)
BIO01.03 Aspartate family			19 genes
AAT2_METJA	36%	57	PUTATIVE ASPARTATE AMINOTRANSFERASE 2 (EC 2.6.1.1)
ASSY_SYNY3	57%	125	ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5)
AAT_BACSP	32%	161	ASPARTATE AMINOTRANSFERASE (EC 2.6.1.1)
ASNB_ECOLI	40%	354	ASPARAGINE SYNTHETASE B (GLUTAMINE-HYDROLYZING) (EC 6.3.5.4)
AK_METJA	30%	746	PROBABLE ASPARTOKINASE (EC 2.7.2.4)
METB_HELPY	65%	789	CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9)
DHOM_LACLA	93%	1169	HOMOSERINE DEHYDROGENASE (EC 1.1.1.3)
KHSE_LACLA	88%	1170	HOMOSERINE KINASE (EC 2.7.1.39)
METF_HAEIN	32%	1279	5,10-METHYLENETETRAHYDROFOLATE REDUCTASE (EC 1.7.99.5)
METE_HAEIN	46%	1281	5-METHYLTETRAHYDROPTEROYLTRIGLUTAMATE--HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.14)
DAPB_BACSU	50%	1600	DIHYDRODIPICOLINATE REDUCTASE (EC 1.3.1.26)
DAPA_BACSU	46%	1661	DIHYDRODIPICOLINATE SYNTHASE (EC 4.2.1.52)
DHAS_STRMU	75%	1663	ASPARTATE-SEMIALDEHYDE DEHYDROGENASE (EC 1.2.1.11)
AAT1_BACSU	50%	1877	ASPARTATE AMINOTRANSFERASE (EC 2.6.1.1)
METB_HELPY	41%	1976	CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9)
META_ECOLI	46%	1977	HOMOSERINE O-SUCCINYLTRANSFERASE (EC 2.3.1.46)
PYRH_SYNY3	56%	2067	URIDYLATE KINASE (EC 2.7.4.-)
THRC_PSEAE	38%	2153	THREONINE SYNTHASE (EC 4.2.99.2)
ASNH_MYCTU	37%	2292	PUTATIVE ASPARAGINE SYNTHETASE (GLUTAMINE-HYDROLYZING) (EC 6.3.5.4)
BIO01.04 Branched chain family			11 genes
ILVX_BACSU	52%	1199	ACETOLACTATE SYNTHASE (EC 4.1.3.18)
LEU1_SYNY3	45%	1238	2-ISOPROPYLMALATE SYNTHASE (EC 4.1.3.12)
LEU3_LACLA	99%	1239	3-ISOPROPYLMALATE DEHYDROGENASE (EC 1.1.1.85)
LEU2_ACTTI	53%	1241	3-ISOPROPYLMALATE DEHYDRATASE (EC 4.2.1.33)
LEUD_LACLA	100%	1242	3-ISOPROPYLMALATE DEHYDRATASE SMALL SUBUNIT (EC 4.2.1.33)
ILVD_LACLA	98%	1244	DIHYDROXY-ACID DEHYDRATASE (EC 4.2.1.9)
ILVB_LACLA	82%	1246	ACETOLACTATE SYNTHASE LARGE SUBUNIT (EC 4.1.3.18)
ILVN_LACLA	100%	1247	ACETOLACTATE SYNTHASE SMALL SUBUNIT (EC 4.1.3.18)
ILVC_LACLA	97%	1247	KETOL-ACID REDUCTOISOMERASE (EC 1.1.1.86)
THD1_LACLA	99%	1249	THREONINE DEHYDRATASE BIOSYNTHETIC (EC 4.2.1.16)

Table 3.

ILVE_HAEIN	67%	1320	BRANCHED-CHAIN AMINO ACID AMINOTRANSFERASE (EC 2.6.1.42)
BIO01.05 Glutamate family			21 genes
ARLY_CAMJE	45%	126	ARGININOSUCCINATE LYASE (EC 4.3.2.1)
ARGE_ECOLI	26%	557	ACETYLORNITHINE DEACETYLASE (EC 3.5.1.16)
ARGC_LACPL	48%	802	N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE (EC 1.2.1.38)
ARGJ_BACSU	49%	803	GLUTAMATE N-ACETYLTTRANSFERASE (EC 2.3.1.35)
ARGD_SYNY3	42%	804	ACETYLORNITHINE AMINOTRANSFERASE (EC 2.6.1.11)
ARGB_METJA	42%	805	PROBABLE ACETYLGLUTAMATE KINASE (EC 2.7.2.8)
AHRC_BACSU	37%	880	ARGININE HYDROXIMATE RESISTANCE PROTEIN
MURI_BACSU	46%	1310	GLUTAMATE RACEMASE (EC 5.1.1.3)
TABA_PSESZ	40%	1312	TABA PROTEIN
YGFT_ECOLI	34%	1313	HYPOTHETICAL 69.7 KD PROTEIN IN KDOI-LYSS INTERGENIC REGION
GLTB_ECOLI	39%	1316	GLUTAMATE SYNTHASE (NADPH) LARGE CHAIN PRECURSOR (EC 1.4.1.13)
DCEB_ECOLI	43%	1322	GLUTAMATE DECARBOXYLASE BETA (EC 4.1.1.15)
CARY_BACSU	51%	1397	CARBAMOYL-PHOSPHATE SYNTHASE, ARGININE-SPECIFIC, LARGE CHAIN
(EC 6.3.5.5)			
PROA_ECOLI	46%	1647	GAMMA-GLUTAMYL PHOSPHATE REDUCTASE (GPR) (EC 1.2.1.41)
PROB_SYNY3	40%	1648	GLUTAMATE 5-KINASE (EC 2.7.2.11)
OTC2_BACSU	43%	1736	ORNITHINE CARBAMOYLTRANSFERASE CHAIN P (EC 2.1.3.3)
AHRC_BACSU	36%	2097	ARGININE HYDROXIMATE RESISTANCE PROTEIN
PROA_ECOLI	22%	2210	GAMMA-GLUTAMYL PHOSPHATE REDUCTASE (GPR) (EC 1.2.1.41)
GLNA_STAAU	45%	2262	GLUTAMINE SYNTHETASE (EC 6.3.1.2)
GLNA_HAEIN	41%	2263	GLUTAMINE SYNTHETASE (EC 6.3.1.2)
RACD_STRTR	53%	2289	ASPARTATE RACEMASE (EC 5.1.1.13)
BIO01.06 Histidine family			9 genes
PTIP_ECOLI	27%	118	PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE PTSP (EC 2.7.3.9)
HIS8_LACLA	96%	1226	HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9)
HIS1_LACLA	99%	1228	ATP PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.17)
HISX_LACLA	98%	1229	HISTIDINOL DEHYDROGENASE (EC 1.1.1.23)
HIS7_LACLA	98%	1231	IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE (EC 4.2.1.19)
HIS5_LACLA	99%	1232	AMIDOTRANSFERASE HIS (EC 2.4.2.-)
HIS4_LACLA	96%	1233	PHOSPHORIBOSYLFORMIMINO-5-AMINOIMIDAZOLE CARBOXAMIDE RIBOTIDE ISOMERASE (EC 5.3.1.16)
HIS6_LACLA	98%	1234	HISF PROTEIN (CYCLASE)
HIS2_LACLA	99%	1235	PHOSPHORIBOSYL-AMP CYCLOHYDROLASE (EC 3.5.4.19) / PHOSPHORIBOSYL-ATP PYROPHOSPHOHYDROLASE (EC 3.6.1.31)
BIO01.07 Pyruvate family			2 genes
DDL_ENTHR	54%	339	D-ALANINE--D-ALANINE LIGASE (EC 6.3.2.4)
ALR_BACSU	43%	860	ALANINE RACEMASE (EC 5.1.1.1)
BIO01.08 Serine family			12 genes
SERC_ECOLI	32%	463	PHOSPHOSERINE AMINOTRANSFERASE (EC 2.6.1.52)
GLYA_HELPY	30%	471	SERINE HYDROXYMETHYLTRANSFERASE (EC 2.1.2.1)
CYSK_BACSU	56%	524	CYSTEINE SYNTHASE (EC 4.2.99.8)
GLYA_BACSU	61%	589	SERINE HYDROXYMETHYLTRANSFERASE (EC 2.1.2.1)
SERC_BACCI	47%	591	PHOSPHOSERINE AMINOTRANSFERASE (EC 2.6.1.52)
SERA_HAEIN	32%	592	D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95)
SERB_HAEIN	42%	593	PHOSPHOSERINE PHOSPHATASE (EC 3.1.3.3)
CYSK_BACSU	53%	790	CYSTEINE SYNTHASE (EC 4.2.99.8)
CBH_CLOPE	34%	1147	CHOLYLGLYCINE HYDROLASE (EC 3.5.1.24)
PFLA_HAEIN	50%	1860	PYRUVATE FORMATE-LYASE 1 ACTIVATING ENZYME (EC 1.97.1.4)
CYSE_BACSU	58%	1900	SERINE ACETYLTTRANSFERASE (EC 2.3.1.30)
PKNS_MYCGE	35%	1936	PUTATIVE SERINE/THREONINE-PROTEIN KINASE (EC 2.7.1.-)
BIO02. BIOSYNTHESIS OF COFACTORS, PROSTHETIC GROUPS, AND CARRIERS			40 genes
BIO02.01 Biotin			3 genes
BIOY_BACSH	34%	313	BIOY PROTEIN PRECURSOR
BIOY_BACSH	28%	321	BIOY PROTEIN PRECURSOR
ACCD_SYNP7	53%	787	ACETYL-COENZYME A CARBOXYLASE CARBOXYL TRANSFERASE SUBUNIT BETA (EC 6.4.1.2)
BIO02.02 Folic acid			6 genes
HPPK_BACSU	42%	1164	2-AMINO-4-HYDROXY-6-HYDROXYMETHYLDIHYDROPTERIDINE
PYROPHOSPHOKINASE (EC 2.7.6.3)			
GCH1_BACSU	37%	1164	GTP CYCLOHYDROLASE I (EC 3.5.4.16)
DHPS_NEIME	34%	1165	DIHYDROPTEROATE SYNTHASE (EC 2.5.1.15)
FOLC_ECOLI	32%	1167	POLYLPOLYGLUTAMATE SYNTHASE (EC 6.3.2.17)
PABB_SALTY	40%	1345	PARA-AMINOBENZOATE SYNTHASE COMPONENT I (EC 4.1.3.-)
PABA_SALTY	50%	1346	PARA-AMINOBENZOATE SYNTHASE GLUTAMINE AMIDOTRANSFERASE COMPONENT II (EC 4.1.3.-)
BIO02.03 Heme and porphyrin			5 genes

Table 3.

HEMK_HAEIN	46%	586	HEMK PROTEIN HOMOLOG
BPAL_STRAU	29%	832	NON-HAEM BROMOPEROXIDASE BPO-A1 (EC 1.11.1.-)
HEMN_BACSU	46%	1151	PROBABLE OXYGEN-INDEPENDENT COPROPORPHYRINOGEN III OXIDASE (EC 1.-.-.-)
BCHE_RHOCA	26%	1179	MAGNESIUM-PROTOPORPHYRIN IX MONOMETHYL ESTER OXIDATIVE CYCLASE 66 KD SUBUNIT
HEMZ_YEREN	39%	1605	FERROCHELATASE (EC 4.99.1.1)
BIO02.04 Lipoate		0 genes	
BIO02.05 Menaquinone and ubiquinone		4 genes	
MENB_BACSU	73%	732	NAPHTHOATE SYNTHASE (EC 4.1.3.36)
MEND_BACSU	42%	735	2-SUCCINYL-6-HYDROXY-2,4-CYCLOHEXADIENE-1-CARBOXYLATE SYNTHASE (SHHC SYNTHASE) / 2-OXOGLUTARATE DECARBOXYLASE (EC 4.1.1.71)
QUEA_HAEIN	44%	1613	S-ADENOSYLMETHIONINE:TRNA RIBOSYLTRANSFERASE-ISOMERASE (EC 5.-.-.-)
GRC2_LACLA	70%	1714	PROBABLE MENAQUINONE BIOSYNTHESIS METHYLTRANSFERASE (EC 2.1.1.-)
BIO02.06 Molybdopterin		0 genes	
BIO02.07 Pantothenate		1 genes	
COAA_HAEIN	43%	1464	PANTOTHENATE KINASE (EC 2.7.1.33)
BIO02.08 Pyridoxine		0 genes	
BIO02.09 Riboflavin		5 genes	
RIBD_ACTPL	51%	1020	RIBOFLAVIN-SPECIFIC DEAMINASE (EC 3.5.4.-)
RISA_ACTPL	54%	1021	RIBOFLAVIN SYNTHASE ALPHA CHAIN (EC 2.5.1.9)
GCH2_ACTPL	60%	1022	GTP CYCLOHYDROLASE II (EC 3.5.4.25) / 3,4-DIHYDROXY-2-BUTANONE 4- PHOSPHATE SYNTHASE
RISB_ACTPL	68%	1022	6,7-DIMETHYL-8-RIBITYLLUMAZINE SYNTHASE (EC 2.5.1.9)
RIBF_CORAM	38%	1139	RIBOFLAVIN KINASE (EC 2.7.1.26)
BIO02.10 Thioredoxin, glutaredoxin, and glutathione		5 genes	
GSHR_HAEIN	30%	861	GLUTATHIONE REDUCTASE (EC 1.6.4.2)
TRXB_CLOLI	47%	963	THIOREDOXIN REDUCTASE (EC 1.6.4.5)
BSAA_BACSU	56%	1400	GLUTATHIONE PEROXIDASE HOMOLOG BSAA
THIO_BACSU	54%	1667	THIOREDOXIN (TRX)
TRXB_MYCGE	36%	1691	THIOREDOXIN REDUCTASE (EC 1.6.4.5)
BIO02.11 Thiamine		10 genes	
YZGD_BACSU	26%	165	HYPOTHETICAL 45.4 KD PROTEIN IN THIAMINASE I 5'REGION
THID_HAEIN	30%	493	PHOSPHOMETHYLPYRIMIDINE KINASE (EC 2.7.4.7)
YZGD_BACSU	42%	1092	HYPOTHETICAL 45.4 KD PROTEIN IN THIAMINASE I 5'REGION
THIE_BACSU	39%	1291	THIAMIN-PHOSPHATE PYROPHOSPHORYLASE (EC 2.5.1.3)
THID_HAEIN	48%	1292	PHOSPHOMETHYLPYRIMIDINE KINASE (EC 2.7.4.7)
THIM_HAEIN	34%	1293	HYDROXYETHYLTHIAZOLE KINASE (EC 2.7.1.50)
DCIP_ENTCL	41%	1337	INDOLE-3-PYRUVATE DECARBOXYLASE (EC 4.1.1.74)
YZGD_BACSU	31%	1662	HYPOTHETICAL 45.4 KD PROTEIN IN THIAMINASE I 5'REGION
POXB_LACEL	46%	2109	PYRUVATE OXIDASE (EC 1.2.3.3)
YZGD_BACSU	40%	2296	HYPOTHETICAL 45.4 KD PROTEIN IN THIAMINASE I 5'REGION
BIO02.12 Pyridine nucleotides		1 genes	
NADE_ECOLI	65%	1107	NH(3)-DEPENDENT NAD
BIO03. CELL ENVELOPE		73 genes	
BIO03.01 Membranes, lipoproteins, and porins		29 genes	
YQJG_BACSU	32%	129	HYPOTHETICAL 30.7 KD LIPOPROTEIN IN GINQ-ANSR INTERGENIC REGION PRECURSOR
MRP_STRSU	29%	147	MURAMIDASE-RELEASED PROTEIN PRECURSOR (136 KD SURFACE PROTEIN)
WAPA_STRMU	26%	149	WALL-ASSOCIATED PROTEIN PRECURSOR
TEE6_STRPY	39%	152	TRYPSIN-RESISTANT SURFACE T6 PROTEIN PRECURSOR
RFAL_KLEPN	29%	201	O-ANTIGEN EXPORT SYSTEM PERMEASE PROTEIN RFBA
PLPC_PASHA	42%	315	OUTER MEMBRANE LIPOPROTEIN 3 PRECURSOR (PLP3)
PLPA_PASHA	36%	316	OUTER MEMBRANE LIPOPROTEIN 1 PRECURSOR (PLP1)
HLPA_HAEIN	38%	317	28 KD OUTER MEMBRANE PROTEIN PRECURSOR
YAEK_ECOLI	37%	318	HYPOTHETICAL 29.4 KD LIPOPROTEIN IN RCSP-ABC INTERGENIC REGION PRECURSOR
YOJL_ECOLI	27%	370	HYPOTHETICAL 38.5 KD LIPOPROTEIN IN ADA-OMPC INTERGENIC REGION PRECURSOR
CSTA_MYCTU	51%	412	CARBON STARVATION PROTEIN A HOMOLOG
SPR_ECOLI	35%	480	LIPOPROTEIN SPR PRECURSOR
60IM_PSEPU	29%	562	60 KD INNER-MEMBRANE PROTEIN
CRTX_RHOCA	33%	571	CRTX PROTEIN
LGT_ECOLI	25%	604	PROLIPOPROTEIN DIACYLGLYCERYL TRANSFERASE (EC 2.4.99.-)

Table 3.

WAPA_STRMU	26%	682	WALL-ASSOCIATED PROTEIN PRECURSOR
CAIC_ECOLI	26%	731	PROBABLE CROTONOBETAINE/CARNITINE-COA LIGASE (EC 6.3.2.-)
SP5E_BACSU	38%	906	STAGE V SPORULATION PROTEIN F
YD57_METJA	32%	935	PUTATIVE POTASSIUM CHANNEL PROTEIN MJ1357
MRP_STRSU	31%	1221	MURAMIDASE-RELEASED PROTEIN PRECURSOR (136 KD SURFACE PROTEIN)
CD59_HSVSA	32%	1299	SURFACE GLYCOPROTEIN CD59 HOMOLOG PRECURSOR
ADHS_STRPA	68%	1347	ADHESIN B PRECURSOR (SALIVA-BINDING PROTEIN)
P29K_STRGC	76%	1348	29 KD MEMBRANE PROTEIN IN PSAA 5'REGION (ORF1)
TMPC_TREPA	36%	1460	MEMBRANE LIPOPROTEIN TMPC PRECURSOR (MEMBRANE PROTEIN C)
LPLA_BACSU	27%	1526	LIPOPROTEIN LPLA PRECURSOR
YDEK_ECOLI	33%	1622	HYPOTHETICAL 136.5 KD LIPOPROTEIN IN HIPB-UXAB INTERGENIC REGION PRECURSOR (ORFT)
CZCD_ALCEU	30%	1646	CATION EFFLUX SYSTEM PROTEIN CZCD
MRP_STRSU	37%	1729	MURAMIDASE-RELEASED PROTEIN PRECURSOR (136 KD SURFACE PROTEIN)
ADHS_STRSA	35%	2164	ADHESIN B PRECURSOR (SALIVA-BINDING PROTEIN)
BIO03.02 Murein sacculus and peptidoglycan			23 genes
MURA_BACSU	55%	311	PROBABLE UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.7)
PBP2_STRPN	46%	336	PENICILLIN-BINDING PROTEIN 2B
MURF_BACSU	39%	340	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2,6-DIAMINOPIMELATE--D-ALANYL-D-ALANYL LIGASE (EC 6.3.2.15)
PBPD_BACSU	28%	390	PENICILLIN-BINDING PROTEIN 4 PRECURSOR (PBP 4)
PBPA_STROR	54%	528	PENICILLIN-BINDING PROTEIN 1A (PBP-1A)
MURA_ACICA	49%	532	UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.7)
RODA_HELPY	32%	661	ROD SHAPE-DETERMINING PROTEIN RODA
PBPF_BACSU	27%	888	PENICILLIN-BINDING PROTEIN 2B (PBP-2B)
MRAY_BACSU	44%	889	PHOSPHO-N-ACETYLMURAMOYL-PENTAPEPTIDE-TRANSFERASE (EC 2.7.8.13)
RODA_ECOLI	28%	914	ROD SHAPE-DETERMINING PROTEIN RODA
MURB_BACSU	39%	1172	UDP-N-ACETYLENOLPYRUVOYLGLUCOSAMINE REDUCTASE (EC 1.1.1.158)
ALYS_ENTHR	39%	1400	AUTOLYSIN PRECURSOR (EC 3.5.1.28)
MURG_BACSU	45%	1629	UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE) PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE (EC 2.4.1.-)
MURD_BACSU	55%	1630	UDP-N-ACETYLMURAMOYLALANINE--D-GLUTAMATE LIGASE (EC 6.3.2.9)
MURE_BACSU	31%	1851	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMATE--2,6-DIAMINOPIMELATE LIGASE (EC 6.3.2.13)
GCAD_BACSU	56%	1931	UDP-N-ACETYLGLUCOSAMINE PYROPHOSPHORYLASE (EC 2.7.7.23)
ALYS_ENTHR	41%	1956	AUTOLYSIN PRECURSOR (EC 3.5.1.28)
DAC_STRSQ	25%	2111	D-ALANYL-D-ALANINE CARBOXYPEPTIDASE PRECURSOR (EC 3.4.16.4)
PBPF_BACSU	31%	2158	PENICILLIN-BINDING PROTEIN 1A (PBP-1A)
BACA_ECOLI	34%	2255	BACITRACIN RESISTANCE PROTEIN (PUTATIVE UNDECAPRENOL KINASE) (EC 2.7.1.66)
MRED_BACSU	22%	2294	ROD SHAPE-DETERMINING PROTEIN MRED
MREC_BACSU	30%	2295	ROD SHAPE-DETERMINING PROTEIN MREC
DACA_BACSU	35%	2336	D-ALANYL-D-ALANINE CARBOXYPEPTIDASE PRECURSOR (EC 3.4.16.4)
BIO03.03 Surface polysaccharides, lipopolysaccharides and antigens			21 genes
DLTE_BACSU	33%	142	DLTE PROTEIN
RFBC_SHIFL	35%	196	DTDP-4-DEHYDRORHAMNOSE 3,5-EPIMERASE (EC 5.1.3.13)
RBB2_ECOLI	46%	197	DTDP-GLUCOSE 4,6-DEHYDRATASE (EC 4.2.1.46)
RFBG_SHIFL	36%	200	DTDP-RHAMNOSYL TRANSFERASE RFBG (EC 2.-.-.-)
LICD_HAEIN	31%	213	LICD PROTEIN
TAGD_BACSU	55%	217	GLYCEROL-3-PHOSPHATE CYTIDYLYLTRANSFERASE (EC 2.7.7.39)
TAGH_BACSU	48%	935	TEICHOIC ACID TRANSLOCATION ATP-BINDING PROTEIN TAGH
TAGG_BACSU	29%	937	TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN TAGG
TAGE_BACSU	29%	941	TEICHOIC ACID BIOSYNTHESIS PROTEIN B PRECURSOR
TAGE_BACSU	27%	944	TEICHOIC ACID BIOSYNTHESIS PROTEIN B PRECURSOR
TAGD_BACSU	49%	948	GLYCEROL-3-PHOSPHATE CYTIDYLYLTRANSFERASE (EC 2.7.7.39)
TAGE_BACSU	38%	950	TEICHOIC ACID BIOSYNTHESIS PROTEIN F
TAGE_BACSU	30%	951	TEICHOIC ACID BIOSYNTHESIS PROTEIN B PRECURSOR
DLTD_BACSU	34%	1287	PROTEIN DLTB PRECURSOR
DLTC_LACCA	43%	1288	D-ALANYL CARRIER PROTEIN (DCP)
DLTB_BACSU	49%	1288	DLTB PROTEIN
TAGE_BACSU	25%	1475	PROBABLE POLY(GLYCEROL-PHOSPHATE) ALPHA-GLUCOSYLTRANSFERASE (EC 2.4.1.52)
TAGE_BACSU	23%	1477	PROBABLE POLY(GLYCEROL-PHOSPHATE) ALPHA-GLUCOSYLTRANSFERASE (EC 2.4.1.52)
RFBU_SALTY	25%	2186	RFBU PROTEIN
RFAK_SALTY	28%	2187	LIPOPOLYSACCHARIDE 1,2-N-ACETYLGUCOSAMINETRANSFERASE (EC 2.4.1.56)
KDTB_HAEIN	39%	2218	LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS PROTEIN KDTB
BIO03.04 Surface structures			0 genes
BIO04. CELLULAR PROCESSES			29 genes
BIO04.01 Cell division			8 genes

Table 3.

DIVC_BACSU	31%	22	CELL DIVISION PROTEIN DIVIC
FTSH_LACLA	98%	27	CELL DIVISION PROTEIN FTSH HOMOLOG (EC 3.4.24.-)
SP0J_BACSU	44%	96	STAGE 0 SPORULATION PROTEIN J
FTSY_BACSU	58%	823	CELL DIVISION PROTEIN FTSY HOMOLOG
FTSE_ECOLI	40%	997	CELL DIVISION ATP-BINDING PROTEIN FTSE
SP3E_BACSU	49%	1701	STAGE III SPORULATION PROTEIN E
FTS2_BACSU	61%	1918	CELL DIVISION PROTEIN FTSZ
FTSA_BACSU	37%	1919	CELL DIVISION PROTEIN FTSA
BIO04.02 Cell killing			3 genes
HLY3_BACCE	44%	495	HEMOLYSIN III (HLY-III)
YQXC_BACSU	60%	879	HYPOTHETICAL 29.7 KD PROTEIN IN FOLD-AHRC INTERGENIC REGION
FRPA_NEIME	30%	1394	IRON-REGULATED PROTEIN FRPA
BIO04.03 Chaperones			4 genes
SUGE_ECOLI	46%	122	SUGE PROTEIN
CHI0_LACLA	100%	396	10 KD CHAPERONIN (PROTEIN CPN10)
CH60_LACLA	97%	397	60 KD CHAPERONIN (PROTEIN CPN60)
DNAK_LACLA	99%	977	DNAK PROTEIN (HEAT SHOCK PROTEIN 70)
BIO04.04 Detoxification			1 genes
SODM_LACLA	100%	410	SUPEROXIDE DISMUTASE (MN) (EC 1.15.1.1)
BIO04.05 Protein and peptide secretion			5 genes
SECA_STAAU	51%	115	PREPROTEIN TRANSLOCASE SECA SUBUNIT
TIG_BACSU	52%	534	TRIGGER FACTOR (TF)
LSPA_STAAU	43%	1023	LIPOPROTEIN SIGNAL PEPTIDASE (EC 3.4.23.36)
SECY_LACLA	98%	2138	PREPROTEIN TRANSLOCASE SECY SUBUNIT
LEPC_BACCL	41%	2330	SIGNAL PEPTIDASE I (EC 3.4.21.89)
BIO04.06 Transformation			8 genes
CMA2_BACSU	46%	729	COMA OPERON PROTEIN 2
CMF3_BACSU	37%	1094	COMF OPERON PROTEIN 3
CMF1_BACSU	37%	1095	COMF OPERON PROTEIN 1
CME3_BACSU	28%	1811	COME OPERON PROTEIN 3
CME1_BACSU	39%	1812	COME OPERON PROTEIN 1
CMG3_BACSU	44%	2166	COMG OPERON PROTEIN 3 PRECURSOR
CMG2_BACSU	30%	2167	COMG OPERON PROTEIN 2
CMG1_BACSU	39%	2168	COMG OPERON PROTEIN 1
BIO04.07 Other			0 genes
BIO05. CENTRAL INTERMEDIARY METABOLISM			30 genes
BIO05.01 General			7 genes
GRC3_BACSU	31%	184	PROBABLE HEPTAPRENYL DIPHOSPHATE SYNTHASE COMPONENT II (EC 2.5.1.30)
STR1_STRGR	42%	198	DTDP-4-DEHYDRORHAMNOSE REDUCTASE (EC 1.1.1.133)
ISPA_BACSU	47%	878	GERANYLTRANSTRANSFERASE (EC 2.5.1.10)
YWNE_BACSU	48%	985	HYPOTHETICAL 55.8 KD PROTEIN IN SPOIIQ-MTA INTERGENIC REGION
YWNE_BACSU	44%	1185	HYPOTHETICAL 55.8 KD PROTEIN IN SPOIIQ-MTA INTERGENIC REGION
GRC3_BACSU	35%	1378	PROBABLE HEPTAPRENYL DIPHOSPHATE SYNTHASE COMPONENT II (EC 2.5.1.30)
METK_STAAU	62%	1950	S-ADENOSYLMETHIONINE SYNTHETASE (EC 2.5.1.6)
BIO05.02 Amino sugars			8 genes
NYLA_PSES8	32%	167	6-AMINOHEXANOATE-CYCLIC-DIMER HYDROLASE (EC 3.5.2.12)
NODB_RHILV	37%	279	CHITOLIGOSACCHARIDE DEACETYLASE (EC 3.5.1.-)
GLMS_MYCLE	36%	1032	PROBABLE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE (ISOMERIZING) (EC 2.6.1.16)
NAGD_ECOLI	28%	1154	NAGD PROTEIN
YVYH_BACSU	47%	1296	POTATIVE UDP-N-ACETYLGLUCOSAMINE 2-EPIMERASE (EC 5.1.3.14)
NAGA_HAEIN	30%	1371	N-ACETYLGLUCOSAMINE-6-PHOSPHATE DEACETYLASE (EC 3.5.1.25)
HASC_STRPY	58%	1376	UTP--GLUCOSE-1-PHOSPHATE URIDYLTRANSFERASE (EC 2.7.7.9)
NAGB_HAEIN	32%	1611	GLUCOSAMINE-6-PHOSPHATE ISOMERASE (EC 5.3.1.10)
BIO05.03 Degradation of polysaccharides			5 genes
NEPU_THEVU	30%	701	NEOPULLULANASE (EC 3.2.1.135)
AMY_BACST	44%	1275	ALPHA-AMYLASE PRECURSOR (EC 3.2.1.1)
O16G_BACTR	60%	1523	OLIGO-1,6-GLUCOSIDASE (EC 3.2.1.10)
CHIX_STROI	31%	2007	EXOCHITINASE 1 PRECURSOR (EC 3.2.1.14)
CHI1_BACCI	34%	2007	CHITINASE A1 PRECURSOR (EC 3.2.1.14)
BIO05.04 Phosphorus compounds			9 genes
DOT_HAEIN	40%	179	DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
DOT_HAEIN	38%	456	DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
SUHB_ECOLI	29%	531	EXTRAGENIC SUPPRESSOR PROTEIN SUHB

Table 3.

APL_LACLA	96%	717	ALKALINE PHOSPHATASE LIKE PROTEIN
DUT_HAEIN	38%	1046	DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
DUT_HAEIN	37%	1441	DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
PPH_MYCPN	27%	1937	PUTATIVE PROTEIN PHOSPHATASE (EC 3.1.3.16)
PGPB_HAEIN	36%	2122	PHOSPHATIDYLGLYCEROPHOSPHATASE B (EC 3.1.3.27)
Y328_SYNY3	35%	2264	PUTATIVE LOW MOLECULAR WEIGHT PROTEIN-TYROSINE-PHOSPHATASE (EC 3.1.3.48)
BIO05.05 Polyamine biosynthesis 1 genes			
ATDA_ECOLI	41%	1653	SPERMIDINE N1-ACETYLTRANSFERASE (EC 2.3.1.57)
BIO05.06 Other 0 genes			
BIO06. ENERGY METABOLISM 148 genes			
BIO06.01 Aerobic 32 genes			
MPLD_STRMU	61%	34	MANNITOL-1-PHOSPHATE 5-DEHYDROGENASE (EC 1.1.1.17)
ODO2_MYCTU	34%	61	DIHYDROLIPOAMIDE SUCCINYLTRANSFERASE COMPONENT (E2) OF 2-OXOGLUTARATE DEHYDROGENASE COMPLEX (EC 2.3.1.61)
VDLC_HELPU	33%	78	PROBABLE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE VDLC (EC 1.-.-)
YOXD_BACSU	36%	141	HYPOTHETICAL OXIDOREDUCTASE IN RTP 5'REGION (EC 1.-.-)
NQOE_THETH	26%	159	NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 14 (EC 1.6.5.3)
DHNA_BACSU	50%	334	NADH DEHYDROGENASE (EC 1.6.99.3)
LDH_BACSU	54%	377	L-LACTATE DEHYDROGENASE (EC 1.1.1.27)
NAOX_ENTPA	39%	395	NADH OXIDASE (EC 1.6.-.-)
LDH_LISMO	24%	488	L-LACTATE DEHYDROGENASE (EC 1.1.1.27)
YBBO_ECOLI	29%	617	HYPOTHETICAL OXIDOREDUCTASE IN USHA-TESA INTERGENIC REGION (EC 1.-.-)
NAOX_METJA	34%	792	PUTATIVE NADH OXIDASE (EC 1.6.-.-)
YWPD_BACSU	38%	799	HYPOTHETICAL OXIDOREDUCTASE IN ROCC-PTA INTERGENIC REGION (EC 1.-.-)
BUDC_KLETE	53%	916	ACETOIN(DIACETYL) REDUCTASE (EC 1.1.1.5)
PHBB_CHRVI	37%	916	ACETOACETYL-COA REDUCTASE (EC 1.1.1.36)
FRDA_SHEPU	45%	1136	FUMARATE REDUCTASE FLAVOPROTEIN SUBUNIT PRECURSOR (EC 1.3.99.1)
LDH_LACCA	37%	1140	L-LACTATE DEHYDROGENASE (EC 1.1.1.27)
GLPD_MYCLE	34%	1268	GLYCEROL-3-PHOSPHATE DEHYDROGENASE (EC 1.1.99.5)
LACM_MYCSM	40%	1277	LACTATE 2-MONOXYGENASE (EC 1.13.12.4)
VDLC_HELPU	36%	1341	PROBABLE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE VDLC (EC 1.-.-)
GPDA_ECOLI	25%	1355	GLYCEROL-3-PHOSPHATE DEHYDROGENASE (NAD+) (EC 1.1.1.8)
LDH_LACLA	100%	1367	L-LACTATE DEHYDROGENASE (EC 1.1.1.27)
GPDA_BACSU	34%	1375	GLYCEROL-3-PHOSPHATE DEHYDROGENASE (NAD)
YINL_LISMO	36%	1506	HYPOTHETICAL OXIDOREDUCTASE IN INLA 5'REGION (EC 1.-.-)
YJQJ_BACSU	38%	1569	HYPOTHETICAL OXIDOREDUCTASE IN GLNQ-ANSR INTERGENIC REGION (EC 1.-.-)
YINL_LISMO	72%	1576	HYPOTHETICAL OXIDOREDUCTASE IN INLA 5'REGION (EC 1.-.-)
MTLK_RHOSH	28%	1675	MANNITOL 2-DEHYDROGENASE (EC 1.1.1.67)
YQIG_BACSU	42%	1730	PROBABLE NADH-DEPENDENT FLAVIN OXIDOREDUCTASE YQIG (EC 1.-.-)
YPPA_CORGL	29%	2048	HYPOTHETICAL 33.0 KD PROTEIN IN PROB-PROA INTERGENIC REGION
YKGC_ECOLI	39%	2119	PROBABLE PYRIDINE NUCLEOTIDE-DISULFIDE OXIDOREDUCTASE IN EAER-BETA INTERGENIC REGION
NAOX_ENTPA	35%	2175	NADH OXIDASE (EC 1.6.-.-)
YBBQ_ECOLI	33%	2278	HYPOTHETICAL 30.8 KD PROTEIN IN GIP-FDRA INTERGENIC REGION
OXIR_STRAT	40%	2318	PROBABLE OXIDOREDUCTASE (EC 1.-.-)
BIO06.02 Amino acids and amines 5 genes			
DHL2_LACCO	57%	488	L-2-HYDROXYISOCAPROATE DEHYDROGENASE (EC 1.1.1.-)
ASPG_BACLI	34%	741	L-ASPARAGINASE (EC 3.5.1.1)
OTCC_CLOPE	52%	806	ORNITHINE CARBAMOYLTRANSFERASE, CATABOLIC (EC 2.1.3.3)
DHCO_LACLA	86%	1262	N5-(CARBOXYETHYL)ORNITHINE SYNTHASE (EC 1.5.1.24)
OTCC_CLOPE	54%	2093	ORNITHINE CARBAMOYLTRANSFERASE, CATABOLIC (EC 2.1.3.3)
BIO06.03 Anaerobic 2 genes			
NRDD_ECOLI	52%	269	ANAEROBIC RIBONUCLEOSIDE-TRIPHOSPHATE REDUCTASE (EC 1.17.4.2)
NRDG_ECOLI	46%	270	ANAEROBIC RIBONUCLEOSIDE-TRIPHOSPHATE REDUCTASE ACTIVATING PROTEIN (EC 1.97.1.-)
BIO06.04 ATP-proton motive force interconversion 6 genes			
ATP6_MYCGE	28%	1087	ATP SYNTHASE A CHAIN (EC 3.6.1.34)
ATPE_STRMU	64%	1802	ATP SYNTHASE EPSILON CHAIN (EC 3.6.1.34)
ATPA_STRMU	79%	1804	ATP SYNTHASE ALPHA CHAIN (EC 3.6.1.34)
ATPA_SYNP1	65%	1806	ATP SYNTHASE ALPHA CHAIN (EC 3.6.1.34)
ATPF_STRMU	47%	1807	ATP SYNTHASE B CHAIN (EC 3.6.1.34)
ATP6_ENTPA	37%	1808	ATP SYNTHASE A CHAIN (EC 3.6.1.34)
BIO06.05 Electron transport 6 genes			

Table 3.

NIFJ_ANASP	44%	428	PYRUVATE-FLAVODOXIN OXIDOREDUCTASE (EC 1.-.-.-)
CYDA_BACSU	49%	706	CYTOCHROME D UBIQUINOL OXIDASE SUBUNIT I (EC 1.10.3.-)
CYDB_BACSU	43%	707	CYTOCHROME D UBIQUINOL OXIDASE SUBUNIT II (EC 1.10.3.-)
DHNA_HAEIN	28%	838	NADH DEHYDROGENASE (EC 1.6.99.3)
DHNA_HAEIN	32%	839	NADH DEHYDROGENASE (EC 1.6.99.3)
FIXC_ECOLI	37%	1193	FIXC PROTEIN
BIO06.06 Entner-Doudoroff			1 genes
KDGR_BACSU	46%	1668	2-DEHYDRO-3-DEOXYGLUCONOKINASE (EC 2.7.1.45)
BIO06.07 Fermentation			8 genes
QOR_PSEAE	32%	231	QUINONE OXIDOREDUCTASE (EC 1.6.5.5)
QOR_PSEAE	32%	721	QUINONE OXIDOREDUCTASE (EC 1.6.5.5)
DHSO_BACSU	32%	915	SORBITOL DEHYDROGENASE (EC 1.1.1.14)
FADH_METMR	37%	1587	GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE (EC 1.2.1.1)
YBDH_ECOLI	31%	1595	HYPOTHETICAL OXIDOREDUCTASE IN CSTA-AHPC INTERGENIC REGION
ADH1_ZYMMO	60%	1852	ALCOHOL DEHYDROGENASE I (EC 1.1.1.1)
SUCD_CLOKL	40%	2210	SUCCINATE-SEMIALDEHYDE DEHYDROGENASE (NAD)
ADHE_ECOLI	44%	2211	ALCOHOL DEHYDROGENASE (EC 1.1.1.1)
BIO06.08 Gluconogenesis			3 genes
DCOA_KLEPN	40%	664	OXALOACETATE DECARBOXYLASE ALPHA CHAIN (EC 4.1.1.3)
SDHM_ECOLI	36%	842	L-SERINE DEHYDRATASE 2 (EC 4.2.1.13)
G6PB_BACST	39%	2224	GLUCOSE-6-PHOSPHATE ISOMERASE B (GPI B) (EC 5.3.1.9)
BIO06.09 Glycolysis			14 genes
PGK_THEMA	57%	241	PHOSPHOGLYCERATE KINASE (EC 2.7.2.3)
ENO_MYCPN	52%	272	ENOLASE (EC 4.2.1.11)
PMGY_ZYMMO	62%	332	PHOSPHOGLYCERATE MUTASE (EC 5.4.2.1)
PGMB_LACLA	100%	440	BETA-PHOSPHOGLUCOMUTASE (EC 5.4.2.6)
G3P_LACLA	99%	552	GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12)
ENO_BACSU	70%	631	ENOLASE (EC 4.2.1.11)
ENO_HAEIN	55%	649	ENOLASE (EC 4.2.1.11)
PGMB_LACLA	32%	748	BETA-PHOSPHOGLUCOMUTASE (EC 5.4.2.6)
TPIS_LACLA	99%	1146	TRIOSEPHOSPHATE ISOMERASE (EC 5.3.1.1)
KPYK_LACLA	54%	1368	PYRUVATE KINASE (EC 2.7.1.40)
K6PF_LACLA	99%	1370	6-PHOSPHOFRUCTOKINASE (EC 2.7.1.11)
PMG2_ECOLI	28%	1749	PROBABLE PHOSPHOGLYCERATE MUTASE 2 (EC 5.4.2.1)
ALF_MYCGE	46%	1959	FRUCTOSE-BISPHOSPHATE ALDOLASE (EC 4.1.2.13)
G3P_LACLA	81%	2313	GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12)
BIO06.10 Pentose phosphate pathway			5 genes
DXS_RHOCA	38%	1508	PROBABLE 1-DEOXYXYLULOSE-5-PHOSPHATE SYNTHASE (DXP SYNTHASE)
TKT_MYCLE	38%	1666	TRANSKETOLASE (EC 2.2.1.1)
DXS_BACSU	38%	1720	PROBABLE 1-DEOXYXYLULOSE-5-PHOSPHATE SYNTHASE (DXP SYNTHASE)
G6PD_BACSU	48%	2282	GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49)
RPIA_ECOLI	39%	2297	RIBOSE 5-PHOSPHATE ISOMERASE A (EC 5.3.1.6)
BIO06.11 Pyruvate dehydrogenase			4 genes
DLDI_BACST	55%	60	LIPOAMIDE DEHYDROGENASE COMPONENT (E3) OF PYRUVATE
DEHYDROGENASE COMPLEX (EC 1.8.1.4)			
ODPB_BACST	62%	63	PYRUVATE DEHYDROGENASE E1 COMPONENT, BETA SUBUNIT (EC 1.2.4.1)
ODPA_BACSU	51%	64	PYRUVATE DEHYDROGENASE E1 COMPONENT, ALPHA SUBUNIT (EC 1.2.4.1)
DLDH_ACHLA	26%	840	LIPOAMIDE DEHYDROGENASE COMPONENT (E3) OF PYRUVATE
DEHYDROGENASE COMPLEX (EC 1.8.1.4)			
BIO06.12 Sugars			53 genes
GLGB_SYNY3	34%	144	1,4-ALPHA-GLUCAN BRANCHING ENZYME (EC 2.4.1.18)
BGL1_BACSU	24%	174	BETA-GLUCOSIDASE (EC 3.2.1.21)
BGLA_BACSU	41%	178	6-PHOSPHO-BETA-GLUCOSIDASE (EC 3.2.1.86)
RMLA_STRMU	88%	195	GLUCOSE-1-PHOSPHATE THYMIDYLTRANSFERASE (EC 2.7.7.24)
CAP1_SPAAU	27%	210	CAP1 PROTEIN
AHPC_BACSU	76%	333	ALKYL HYDROPEROXIDE REDUCTASE C22 PROTEIN (EC 1.6.4.-)
GUNM_CLOTM	31%	392	ENDOGLUCANASE M (EC 3.2.1.4)
CASB_KLEOX	51%	420	PHOSPHO-CELLOBIASE (EC 3.2.1.-)
XYNB_BUTFI	29%	440	ENDO-1,4-BETA-XYLANASE B (EC 3.2.1.8)
6PGD_SALTY	68%	606	6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING (EC 1.1.1.44)
GAL7_LACHE	36%	614	GALACTOSE-1-PHOSPHATE URIDYLTRANSFERASE (EC 2.7.7.10)
PFL_STRMU	83%	656	FORMATE ACETYLTRANSFERASE (EC 2.3.1.54)
NALQ_ECOLI	29%	692	4-ALPHA-GLUCANOTRANSFERASE (EC 2.4.1.25)
GLGC_BACST	57%	694	GLUCOSE-1-PHOSPHATE ADENYLYLTRANSFERASE (EC 2.7.7.27)
GLGD_BACST	31%	695	GLYCOGEN BIOSYNTHESIS PROTEIN GLGD
GLGA_BACSU	46%	697	GLYCOGEN SYNTHASE (EC 2.4.1.21)
PHSG_BACSU	44%	699	GLYCOGEN PHOSPHORYLASE (EC 2.4.1.1)
BGLA_BACSU	54%	831	6-PHOSPHO-BETA-GLUCOSIDASE (EC 3.2.1.86)
GAL7_STRLI	24%	849	GALACTOSE-1-PHOSPHATE URIDYLTRANSFERASE (EC 2.7.7.10)
UDP_HAEIN	39%	853	URIDINE PHOSPHORYLASE (EC 2.4.2.3)

Table 3.

WCAL_SALTY	35%	947	PUTATIVE COLANIC ACID BIOSYNTHESIS GLYCOSYL TRANSFERASE WCAL
DBOD_STRTR	72%	954	PURINE NUCLEOSIDE PHOSPHORYLASE (EC 2.4.2.1)
LACC_STRMU	33%	982	TAGATOSE-6-PHOSPHATE KINASE (EC 2.7.1.-)
3MGA_HAEIN	37%	1134	DNA-3-METHYLADENINE GLYCOSIDASE (EC 3.2.2.20)
GLPK_BACSU	61%	1270	GLYCEROL KINASE (EC 2.7.1.30)
THGA_ECOLI	40%	1319	GALACTOSIDE ACETYLTRANSFERASE (EC 2.3.1.18)
GDDB_BACSU	60%	1378	GLUCOSE INHIBITED DIVISION PROTEIN B
DEOC_BACSU	60%	1461	DEOXYRIBOSE-PHOSPHATE ALDOLASE (EC 4.1.2.4)
BGL1_BACSU	56%	1488	BETA-GLUCOSIDASE (EC 3.2.1.21)
GLK_BACSU	34%	1517	GLUCOKINASE (EC 2.7.1.2)
BGLA_THEMA	33%	1518	BETA-GLUCOSIDASE A (EC 3.2.1.21)
STRH_STRPN	29%	1524	BETA-N-ACETYLHEXOSAMINIDASE PRECURSOR (EC 3.2.1.52)
XYNB_BACPU	48%	1540	BETA-XYLOSIDASE (EC 3.2.1.37)
GALM_ACICA	29%	1543	ALDOSE 1-EPIMERASE PRECURSOR (EC 5.1.3.3)
XYLB_BACSU	49%	1544	XYLULOSE KINASE (EC 2.7.1.17)
XYLA_THETU	50%	1545	XYLOSE ISOMERASE (EC 5.3.1.5)
XYNX_CLOTM	28%	1618	EXOGLUCANASE XYNX PRECURSOR (EC 3.2.1.91)
ALKH_BACSU	35%	1668	4-HYDROXY-2-OXOGLUTARATE ALDOLASE (EC 4.1.3.16)
UXAC_ECOLI	44%	1669	URONATE ISOMERASE (EC 5.3.1.12)
UXUA_THENE	57%	1673	MANNONATE DEHYDRATASE (EC 4.2.1.8)
RBSK_BACSU	47%	1683	RIBOKINASE (EC 2.7.1.15)
ASCB_ECOLI	53%	1718	6-PHOSPHO-BETA-GLUCOSIDASE ASCB (EC 3.2.1.86)
GIDA_BACSU	61%	1895	GLUCOSE INHIBITED DIVISION PROTEIN A
RPE_SYNY3	47%	1983	RIBULOSE-PHOSPHATE 3-EPIMERASE (EC 5.1.3.1)
GALE_STRMU	60%	2034	UDP-GLUCOSE 4-EPIMERASE (EC 5.1.3.2)
BGA2_ECOLI	34%	2036	EVOLVED BETA-GALACTOSIDASE ALPHA-SUBUNIT (EC 3.2.1.23)
THGA_LACLA	100%	2038	GALACTOSIDE O-ACETYLTRANSFERASE (EC 2.3.1.18)
GAL7_STRMU	52%	2039	GALACTOSE-1-PHOSPHATE URIDYLTRANSFERASE (EC 2.7.7.10)
GAL1_STRMU	58%	2041	GALACTOKINASE (EC 2.7.1.6)
GALM_HARIN	36%	2042	ALDOSE 1-EPIMERASE (EC 5.1.3.3)
GLK_BACSU	47%	2081	GLUCOKINASE (EC 2.7.1.2)
GNTK_BACSU	49%	2249	GLUCONOXINASE (EC 2.7.1.12)
YQEC_BACSU	53%	2250	HYPOTHETICAL 32.8 KD PROTEIN IN NUCB-AROD INTERGENIC REGION
BIO06.13 TCA cycle			
CISY_THIFE	29%	665	9 genes CITRATE SYNTHASE (EC 4.1.3.7)
ACON_BRAJA	51%	667	ACONITATE HYDRATASE (EC 4.2.1.3)
LDH_STRMU	60%	669	ISOCITRATE DEHYDROGENASE (NADP) (EC 1.1.1.42)
MLES_LACLA	98%	920	MALOLACTIC ENZYME (EC 1.-.-.-)
MAOX_BACST	49%	1201	MALATE OXIDOREDUCTASE (NAD) (EC 1.1.1.38)
CITC_ECOLI	37%	1205	[CITRATE (PRO-3S)-LYASE] LIGASE (EC 6.2.1.22)
CILG_HAEIN	52%	1205	CITRATE LYASE ACYL CARRIER PROTEIN (CITRATE LYASE GAMMA CHAIN)
CILB_KLEFN	52%	1206	CITRATE LYASE BETA CHAIN (EC 4.1.3.6)
CILA_ECOLI	56%	1207	CITRATE LYASE ALPHA CHAIN (EC 4.1.3.6)
BIO07. FATTY ACID AND PHOSPHOLIPID METABOLISM			
		20 genes	
BIO07.01 General			
LPLA_MYCPN	36%	65	20 genes PROBABLE LIPOATE-PROTEIN LIGASE A (EC 6.-.-.-)
PLSX_BACSU	47%	71	FATTY ACID/PHOSPHOLIPID SYNTHESIS PROTEIN PLSX HOMOLOG
GLPQ_BACSU	33%	217	GLYCEROPHOSPHORYL DIESTER PHOSPHODIESTERASE (EC 3.1.4.46)
FABZ_ECOLI	42%	558	(3R)-HYDROXYMYRISTOYL-[ACYL CARRIER PROTEIN] DEHYDRATASE (EC 4.2.1.-)
FABI_HAEIN	42%	559	ENOYL-[ACYL-CARRIER-PROTEIN] REDUCTASE (NADH) (EC 1.3.1.9)
DLTA_LACCA	24%	653	D-ALANINE-ACTIVATING ENZYME (EC 6.3.2.-)
ACCC_METJA	48%	662	BIOTIN CARBOXYLASE (EC 6.3.4.14)
FABD_BACSU	45%	781	MALONYL COA-ACYL CARRIER PROTEIN TRANSACYLASE (EC 2.3.1.39)
FABG_BACSU	46%	782	3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100)
FABF_VIBHA	42%	783	3-OXOACYL-[ACYL-CARRIER-PROTEIN] SYNTHASE II (EC 2.3.1.41)
BCCP_ANASP	51%	784	BIOTIN CARBOXYL CARRIER PROTEIN OF ACETYL-COA CARBOXYLASE (EC 6.4.1.2)
FABZ_HAEIN	50%	784	(3R)-HYDROXYMYRISTOYL-[ACYL CARRIER PROTEIN] DEHYDRATASE (EC 4.2.1.-)
ACCC_BACSU	59%	786	BIOTIN CARBOXYLASE (EC 6.3.4.14)
ACCA_HAEIN	49%	788	ACETYL-COENZYME A CARBOXYLASE CARBOXYL TRANSFERASE SUBUNIT ALPHA (EC 6.4.1.2)
ALPHA (EC 6.4.1.2)			
KDGL_STRMU	68%	1093	DIACYLGLYCEROL KINASE (EC 2.7.1.107)
ATOB_HAEIN	44%	1608	ACETYL-COA ACETYLTRANSFERASE (EC 2.3.1.9)
FABG_BACSU	39%	1824	3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100)
PAC_BACSH	45%	1883	PENICILLIN ACYLASE (EC 3.5.1.11)
PGSA_BACSU	43%	2026	CDP-DIACYLGLYCEROL--GLYCEROL-3-PHOSPHATE 3-PHOSPHATIDYLTRANSFERASE (EC 2.7.8.5)
CDSA_PSEAE	32%	2179	PHOSPHATIDATE CYTIDYLTRANSFERASE (EC 2.7.7.41)
BIO08. PURINES, PYRIMIDINES, NUCLEOSIDES AND NUCLEOTIDES			
		47 genes	
BIO08.01 2'-deoxyribonucleotide metabolism			
		0 genes	

Table 3.

BIO08.02 Nucleotide and nucleoside interconversions			4 genes
DGK2_LACAC	62%	491	DEOXYGUANOSINE KINASE (EC 2.7.1.113) / DEOXYADENOSINE KINASE (EC 2.7.1.76) SUBUNIT 2
KITH_STRGC	79%	583	THYMIDINE KINASE (EC 2.7.1.21)
DGK2_LACAC	55%	1168	DEOXYGUANOSINE KINASE (EC 2.7.1.113) / DEOXYADENOSINE KINASE (EC 2.7.1.76) SUBUNIT 2
KCY_BACSU	49%	1740	PROBABLE CYTIDYLATE KINASE (EC 2.7.4.14)
BIO08.03 Purine ribonucleotide biosynthesis			22 genes
IMDH_STRPY	82%	219	INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205)
ADD_ECOLI	33%	285	ADENOSINE DEAMINASE (EC 3.5.4.4)
FOLD_STRTR	65%	874	METHYLENETETRAHYDROFOLATE DEHYDROGENASE (EC 1.5.1.5) / METHENYLtetrahydrofolate cyclohydrolase (EC 3.5.4.9)
FTHS_STRMU	65%	958	FORMATE--TETRAHYDROFOLATE LIGASE (EC 6.3.4.3)
IMDH_HELPY	37%	1156	INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205)
GUAA_BACSU	67%	1514	GMP SYNTHASE (GLUTAMINE-HYDROLYZING) (EC 6.3.5.2)
PURK_SYNY3	33%	1548	PHOSPHORIBOSYLAMINOIMIDAZOLE CARBOXYLASE ATPASE SUBUNIT (EC 4.1.1.21)
PUR6_HAEIN	59%	1549	PHOSPHORIBOSYLAMINOIMIDAZOLE CARBOXYLASE CATALYTIC SUBUNIT (EC 4.1.1.21)
FUR2_BACSU	55%	1550	PHOSPHORIBOSYLAMINE--GLYCINE LIGASE (EC 6.3.4.13)
PUR9_BACSU	55%	1555	PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXAMIDE FORMYLTRANSFERASE (EC 2.1.2.3)
PUR3_BACSU	45%	1561	PHOSPHORIBOSYLGLYCINAMIDE FORMYLTRANSFERASE (EC 2.1.2.2)
PUR5_BACSU	60%	1562	PHOSPHORIBOSYLFORMYLGLYCINAMIDINE CYCLO-LIGASE (EC 6.3.3.1)
PUR1_BACSU	61%	1568	AMIDOPHOSPHORIBOSYLTRANSFERASE PRECURSOR (EC 2.4.2.14)
PURL_BACSU	58%	1571	PHOSPHORIBOSYLFORMYLGLYCINAMIDINE SYNTHASE II (EC 6.3.5.3)
PUR0_BACSU	61%	1573	PHOSPHORIBOSYLFORMYLGLYCINAMIDINE SYNTHASE I (EC 6.3.5.3)
PUR7_STRPN	66%	1574	PHOSPHORIBOSYLAMINOIMIDAZOLE-SUCCINOCARBOXAMIDE SYNTHASE (EC 6.3.2.6)
PUR8_BACSU	76%	1685	ADENYLOSUCCINATE LYASE (EC 4.3.2.2)
PUR8_STRMU	81%	1685	ADENYLOSUCCINATE LYASE (EC 4.3.2.2)
KGUA_SYNY3	48%	1946	GUANYLATE KINASE (EC 2.7.4.8)
PURA_BACSU	59%	2009	ADENYLOSUCCINATE SYNTHETASE (EC 6.3.4.4)
PUR0_BACSU	31%	2107	PHOSPHORIBOSYLFORMYLGLYCINAMIDINE SYNTHASE I (EC 6.3.5.3)
KAD_LACLA	98%	2137	ADENYLATE KINASE (EC 2.7.4.3)
BIO08.04 Pyrimidine ribonucleotide biosynthesis			10 genes
PYRG_BACSU	64%	486	CTP SYNTHASE (EC 6.3.4.2)
PYRE_BACSU	57%	1078	OROTATE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.10)
PYRC_LACLE	42%	1079	DIHYDROOROTASE (EC 3.5.2.3)
DCOP_LACLC	97%	1380	OROTIDINE 5'-PHOSPHATE DECARBOXYLASE (EC 4.1.1.23)
PYDB_LACLC	93%	1381	DIHYDROOROTATE DEHYDROGENASE B (EC 1.3.3.1)
CDD_BACSU	54%	1461	CYTIDINE DEAMINASE (EC 3.5.4.5)
TYSY_LACLA	98%	1579	THYMIDYLATE SYNTHASE (EC 2.1.1.45)
PYDA_LACLC	98%	1589	DIHYDROOROTATE DEHYDROGENASE A (EC 1.3.3.1)
CARA_BACSU	56%	1640	CARBAMOYL-PHOSPHATE SYNTHASE, PYRIMIDINE-SPECIFIC, SMALL CHAIN (EC 6.3.5.5)
PYRB_SYNY3	42%	1641	ASPARTATE CARBAMOYLTRANSFERASE (EC 2.1.3.2)
BIO08.05 Salvage of nucleosides and nucleotides			11 genes
HPRT_LACLA	99%	25	HYPOXANTHINE-GUANINE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.8)
KTHY_BACSU	48%	401	THYMIDYLATE KINASE (EC 2.7.4.9)
APT_ECOLI	54%	621	ADENINE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.7)
KPRS_BACSU	66%	824	RIBOSE-PHOSPHATE PYROPHOSPHOKINASE (EC 2.7.6.1)
DEOB_BACSU	55%	953	PHOSPHOPENTOMUTASE (EC 5.4.2.7)
XPT_BACSU	52%	1157	XANTHINE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.-)
HPRT_BACSU	56%	1557	HYPOXANTHINE-GUANINE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.8)
URK_HAEIN	52%	1705	URIDINE KINASE (EC 2.7.1.48)
KPRS_BACGL	26%	1906	RIBOSE-PHOSPHATE PYROPHOSPHOKINASE (EC 2.7.6.1)
UPP_LACLA	97%	1972	URACIL PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.9)
PURR_BACSU	50%	2331	PUR OPERON REPRESSOR
BIO08.06 Sugar-nucleotide biosynthesis and interconversions			0 genes
BIO09. REGULATORY FUNCTIONS			94 genes
BIO09.01 General			41 genes
YQAE_BACSU	35%	11	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN SPOIIC-CWLA INTERGENIC REGION (ORF7)
XRE_BACSU	50%	44	PUTATIVE PBSX REPRESSOR
RELA_STRBQ	66%	106	PUTATIVE GTP PYROPHOSPHOKINASE (EC 2.7.6.5)
TER8_PASPI	29%	128	TETRACYCLINE REPRESSOR PROTEIN CLASS H
YQAE_BACSU	34%	133	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN SPOIIC-CWLA INTERGENIC REGION (ORF7)
TER2_ECOLI	38%	143	TETRACYCLINE REPRESSOR PROTEIN CLASS B (TRANSPOSON TN10)

Table 3.

CODY_BACSU	49%	162	CODY PROTEIN (VEGETATIVE PROTEIN 286B)
MECB_BACSU	51%	549	NEGATIVE REGULATOR OF GENETIC COMPETENCE MECB
CTSR_BACSU	45%	627	TRANSCRIPTIONAL REGULATOR CTSR
MECB_BACSU	50%	629	NEGATIVE REGULATOR OF GENETIC COMPETENCE MECB
YP20_BACLI	34%	737	HYPOTHETICAL P20 PROTEIN
SLYA_SALTY	26%	747	TRANSCRIPTIONAL REGULATOR SLYA (SALMOLYSIN)
YP20_BACLI	33%	795	HYPOTHETICAL P20 PROTEIN
BLAI_BACLI	31%	843	PENICILLINASE REPRESSOR (REGULATORY PROTEIN BLAI)
DEGU_BACSU	39%	910	TRANSCRIPTIONAL REGULATORY PROTEIN DEGU (PROTEASE PRODUCTION ENHANCER PROTEIN)
HRCA_LACLA	99%	975	HEAT-INDUCIBLE TRANSCRIPTION REPRESSOR HRCA
SORC_KLEPN	21%	1204	SORBITOL (SOR) OPERON REGULATOR
XRE_BACSU	42%	1220	POTENTIAL FBSX REPRESSOR
YQAE_BACSU	35%	1300	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN SPOIIC-CWLA INTERGENIC REGION (ORF7)
Y186_HAEIN	39%	1506	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR HI0186
TER2_ECOLI	29%	1551	TETRACYCLINE REPRESSOR PROTEIN CLASS B (TRANSPOSON TN10)
TPDT_ALCEU	40%	1594	TRANSCRIPTIONAL REGULATORY PROTEIN TPDT
GLNB_FREDI	56%	1631	NITROGEN REGULATORY PROTEIN P-II (PII SIGNAL TRANSDUCING PROTEIN)
PYRR_BACCL	60%	1644	PYRIMIDINE OPERON REGULATORY PROTEIN PYRR
SMTB_SYNYE	34%	1645	TRANSCRIPTIONAL REPRESSOR SMTB HOMOLOG
ARR2_ECOLI	31%	1730	ARSENICAL RESISTANCE OPERON REPRESSOR
GERE_BACSU	41%	1737	GERMINATION PROTEIN GERE
PHOU_ECOLI	36%	1750	PHOSPHATE TRANSPORT SYSTEM REGULATORY PROTEIN
TENA_BACSU	41%	1818	TRANSCRIPTIONAL ACTIVATOR TENA
BIRA_BACSU	34%	1820	BIRA BIFUNCTIONAL PROTEIN (BIOTIN OPERON REPRESSOR)
Y186_HAEIN	31%	1927	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR HI0186
BIRA_BACSU	33%	1953	BIRA BIFUNCTIONAL PROTEIN (BIOTIN OPERON REPRESSOR)
RPC2_BPP22	30%	2003	REPRESSOR PROTEIN C2
Y186_HAEIN	34%	2032	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR HI0186
NADR_ECOLI	30%	2047	TRANSCRIPTIONAL REGULATOR NADR
FLP_LACCA	41%	2062	PROBABLE TRANSCRIPTIONAL REGULATOR FLP
ACKA_BACSU	54%	2069	ACETATE KINASE (EC 2.7.2.1)
GLNR_BACSU	50%	2264	REGULATORY PROTEIN GLNR
YQAE_BACSU	41%	2276	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN SPOIIC-CWLA INTERGENIC REGION (ORF7)
FNR_BACSU	28%	2298	ANAEROBIC REGULATORY PROTEIN
ACRR_ECOLI	32%	2317	POTENTIAL ACRA OPERON REPRESSOR
BIO09.02 Two-component systems			13 genes
RESE_BACSU	33%	400	SENSOR PROTEIN RESE (EC 2.7.3.-)
YYCF_BACSU	69%	401	HYPOTHETICAL 27.2 KD SENSORY TRANSDUCTION PROTEIN IN ROCR-PURA INTERGENIC REGION
DEGS_BACBR	40%	909	SENSOR PROTEIN DEGS (EC 2.7.3.-)
CZCR_ALCEU	39%	1028	TRANSCRIPTIONAL ACTIVATOR PROTEIN CZCR
YEDV_ECOLI	28%	1030	PROBABLE SENSOR PROTEIN YEDV (EC 2.7.3.-)
PHOP_BACSU	40%	1456	ALKALINE PHOSPHATASE SYNTHESIS TRANSCRIPTIONAL REGULATORY PROTEIN PHOP
PHOR_BACSU	32%	1457	ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN PHOR (EC 2.7.3.-)
RESE_BACSU	27%	1633	SENSOR PROTEIN RESE (EC 2.7.3.-)
CZCR_ALCEU	43%	1635	TRANSCRIPTIONAL ACTIVATOR PROTEIN CZCR
RESE_BACSU	33%	1722	SENSOR PROTEIN RESE (EC 2.7.3.-)
COPR_PSESM	34%	1723	TRANSCRIPTIONAL ACTIVATOR PROTEIN COPR
YXDK_BACSU	33%	1784	HYPOTHETICAL SENSOR-LIKE HISTIDINE KINASE IN IDH 3'REGION (EC 2.7.3.-)
YXDJ_BACSU	42%	1785	HYPOTHETICAL 26.6 KD SENSORY TRANSDUCTION PROTEIN IN IDH 3'REGION
BIO09.03 LacI-family regulators			6 genes
CYTR_ECOLI	28%	728	TRANSCRIPTIONAL REPRESSOR CYTR
RBSR_BACSU	31%	1533	RIBOSE OPERON REPRESSOR
RBSR_BACSU	42%	1684	RIBOSE OPERON REPRESSOR
CCPA_STAXY	47%	1692	PROBABLE CATABOLITE CONTROL PROTEIN A
PURR_HAEIN	34%	1724	PURINE NUCLEOTIDE SYNTHESIS REPRESSOR
YYAG_BACSU	36%	2196	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN TETB-EXOA INTERGENIC REGION
BIO09.04 LysR-family regulators			4 genes
YWFK_BACSU	25%	375	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN PTA 3'REGION
MLER_LACLA	100%	893	MALOLACTIC FERMENTATION SYSTEM TRANSCRIPTIONAL ACTIVATOR
YWBI_BACSU	29%	1261	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN EPR-GALK INTERGENIC REGION
YWBI_BACSU	30%	1926	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN EPR-GALK INTERGENIC REGION
BIO09.05 AraC-family regulators			4 genes

Table 3.

SLAP_LACBR	30%	480	S-LAYER PROTEIN PRECURSOR (SURFACE LAYER PROTEIN)
ADAA_BACSU	44%	517	METHYLPHOSPHOTRIESTER-DNA ALKYLTRANSFERASE
LUMQ_PHOLE	32%	1547	PROBABLE TRANSCRIPTIONAL REGULATOR LUMQ
TOXA_CLODI	30%	2284	TOXIN A
BIO09.06 GntR-family regulators 6 genes			
YHHX_ECOLI (F345)	44%	261	HYPOTHETICAL 38.8 KD PROTEIN IN GNTR-GGT INTERGENIC REGION
TRER_BACSU	44%	434	TREHALOSE OPERON TRANSCRIPTIONAL REPRESSOR
PARR_ECOLI	30%	1459	FATTY ACYL RESPONSIVE REGULATOR (F30 PROTEIN)
YINI_STRAM (ORF 1)	30%	1676	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN INSTABLE DNA LOCUS
YXAM_BACSU	24%	2215	HYPOTHETICAL 44.7 KD PROTEIN IN GNTR-HTPG INTERGENIC REGION
YXAT_BACSU	18%	2305	HYPOTHETICAL 44.3 KD PROTEIN IN GNTR-HTPG INTERGENIC REGION
BIO09.07 DeoR-family regulators 8 genes			
YKEP_BACSU	33%	278	HYPOTHETICAL 41.6 KD PROTEIN IN IDH-DEOR INTERGENIC REGION
YXEB_BACSU REGION PRECURSOR	35%	327	PROBABLE ABC TRANSPORTER BINDING PROTEIN IN IDH-DEOR INTERGENIC REGION
YXEA_BACSU PRECURSOR	37%	575	HYPOTHETICAL 13.0 KD PROTEIN IN IDH-DEOR INTERGENIC REGION
GATR_ECOLI	31%	794	GALACTITOL UTILIZATION OPERON REPRESSOR
YKEP_BACSU	43%	980	HYPOTHETICAL 41.6 KD PROTEIN IN IDH-DEOR INTERGENIC REGION
YBJG_ECOLI	32%	1031	HYPOTHETICAL 22.4 KD PROTEIN IN DEOR-CMR INTERGENIC REGION
GLPR_HAEIN	33%	1330	GLYCEROL-3-PHOSPHATE REGULON REPRESSOR
YXEA_BACSU PRECURSOR	39%	1504	HYPOTHETICAL 13.0 KD PROTEIN IN IDH-DEOR INTERGENIC REGION
BIO09.08 MarR-family regulators 2 genes			
YDEA_ECOLI	32%	329	HYPOTHETICAL 42.5 KD PROTEIN IN UXAB-MARR INTERGENIC REGION
PETP_RHOCA	34%	1338	PETP PROTEIN
BIO09.09 BglG-family regulators 1 genes			
LICT_BACSU	41%	1491	TRANSCRIPTION ANTITERMINATOR LICIT
BIO09.10 GTP-binding proteins 9 genes			
YYAF_BACSU INTERGENIC REGION	68%	12	HYPOTHETICAL 40.1 KD GTP-BINDING PROTEIN IN RPSF-SPO0J
HFLX_ECOLI	44%	222	GTP-BINDING PROTEIN HFLX
ERA_STRMU	75%	353	GTP-BINDING PROTEIN ERA HOMOLOG
YPHC_BACSU REGION	62%	760	HYPOTHETICAL 48.8 KD GTP-BINDING PROTEIN IN CMK-GPSA INTERGENIC REGION
LEPA_BACSU	77%	1116	GTP-BINDING PROTEIN LEPA
YSXC_BACSU (ORFX)	66%	1163	HYPOTHETICAL GTP-BINDING PROTEIN IN LONA-HEMA INTERGENIC REGION
OBG_BACSU	64%	1626	SPOOB-ASSOCIATED GTP-BINDING PROTEIN
SR54_SYNY3	51%	1654	SIGNAL RECOGNITION PARTICLE PROTEIN (FIFTY-FOUR HOMOLOG)
THDF_BACSU	60%	2308	POSSIBLE THIOPHENE AND FURAN OXIDATION PROTEIN THDF
BIO10. REPLICATION 77 genes			
BIO10.01 Degradation of DNA 11 genes			
ADDB_BACSU	28%	5	ATP-DEPENDENT NUCLEASE SUBUNIT B
ADDA_BACSU	32%	7	ATP-DEPENDENT NUCLEASE SUBUNIT A
ADDA_BACSU	33%	9	ATP-DEPENDENT NUCLEASE SUBUNIT A
UVRB_HAEIN	40%	555	EXCINUCLEASE ABC SUBUNIT B
RECJ_ERWCH	36%	619	SINGLE-STRANDED-DNA-SPECIFIC EXONUCLEASE RECJ (EC 3.1.-.-)
EXOA_BACSU	40%	796	EXODEOXYRIBONUCLEASE (EC 3.1.11.2)
UVRC_BACSU	54%	855	EXCINUCLEASE ABC SUBUNIT C
EX7L_BACSU	43%	876	PUTATIVE EXODEOXYRIBONUCLEASE LARGE SUBUNIT (EC 3.1.11.6)
END3_BACSU	50%	1081	PROBABLE ENDONUCLEASE III (EC 4.2.99.18)
EX5A_ECOLI	27%	1777	EXODEOXYRIBONUCLEASE V 67 KD POLYPEPTIDE (EC 3.1.11.5)
EX5A_ECOLI	32%	1778	EXODEOXYRIBONUCLEASE V 67 KD POLYPEPTIDE (EC 3.1.11.5)
BIO10.02 DNA replication, restriction, modification, recombination, and repair 66 genes			
DNAA_STAAN	45%	1	CHROMOSOMAL REPLICATION INITIATOR PROTEIN DNAA
DP3B_STAAN	41%	2	DNA POLYMERASE III, BETA CHAIN (EC 2.7.7.7)
MFD_BACSU	50%	19	TRANSCRIPTION-REPAIR COUPLING FACTOR (TRCF)
PRIM_BPP4	21%	38	PUTATIVE P4-SPECIFIC DNA PRIMASE (EC 2.7.7.-)
YQXN_BACSU (ORF3)	31%	58	HYPOTHETICAL 29.3 KD PROTEIN IN BEX-DNAG/DNAE INTERGENIC REGION
YQEU_BACSU	44%	103	HYPOTHETICAL 28.8 KD PROTEIN IN DNAJ-RPSU INTERGENIC REGION
YAAK_BACSU	44%	120	HYPOTHETICAL 11.8 KD PROTEIN IN DNAZ-RECR INTERGENIC REGION
T3RE_BACCE 3.1.21.5)	29%	151	TYPE III RESTRICTION-MODIFICATION SYSTEM ENZYME RES (EC
UNG_STRPN	53%	230	URACIL-DNA GLYCOSYLASE (EC 3.2.2.-)

Table 3.

RECR_STRTR	76%	338	RECOMBINATION PROTEIN RECR
FPG_LACLC	90%	356	FORMAMIDOPYRIMIDINE-DNA GLYCOSYLASE (EC 3.2.2.23)
RECA_LACLA	94%	357	RECA PROTEIN
SSB_BACSU	44%	395	SINGLE-STRAND BINDING PROTEIN (SSB)
HOLB_BACSU	35%	402	DNA POLYMERASE III, DELTA' SUBUNIT (EC 2.7.7.7)
DNLJ_RHOMR	43%	423	DNA LIGASE (EC 6.5.1.2)
DNAC_ECOLI	27%	454	DNA REPLICATION PROTEIN DNAC
DP3A_HELPY	26%	494	DNA POLYMERASE III, ALPHA CHAIN (EC 2.7.7.7)
PRIM_LACLA	100%	542	DNA PRIMASE (EC 2.7.7.-)
DFF_ECOLI	30%	565	DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN
TIR1_ECOLI	32%	640	TYPE I RESTRICTION ENZYME ECOR124II R PROTEIN (EC 3.1.21.3)
YC18_METJA	27%	644	HYPOTHETICAL PROTEIN MJ1218
YABF_BACSU	47%	687	HYPOTHETICAL 20.7 KD PROTEIN IN METS-KSGA INTERGENIC REGION
YAAJ_BACSU	51%	725	HYPOTHETICAL 17.8 KD PROTEIN IN SERS-DNAH INTERGENIC REGION
DNAC_BACSU	56%	752	REPLICATIVE DNA HELICASE (EC 3.6.1.-)
DNAB_BACSU	27%	756	REPLICATION INITIATION AND MEMBRANE ATTACHMENT PROTEIN
DNAI_BACSU	38%	757	PRIMOSOMAL PROTEIN DNAI
LEF3_NPVAC	26%	833	LATE EXPRESSION FACTOR 3
MUTY_ECOLI	43%	857	A/G-SPECIFIC ADENINE GLYCOSYLASE (EC 3.2.2.-)
REC�_BACSU	44%	881	DNA REPAIR PROTEIN REC� (RECOMBINATION PROTEIN N)
GYRB_STRPN	82%	926	DNA GYRASE SUBUNIT B (EC 5.99.1.3)
GRPE_LACLA	99%	975	GRPE PROTEIN
RIR2_MYCTU	48%	999	RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE BETA CHAIN (EC 1.17.4.1)
RIR1_MYCTU	50%	1001	RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE ALPHA CHAIN (EC 1.17.4.1)
PARE_STAAU	70%	1005	TOPOISOMERASE IV SUBUNIT B (EC 5.99.1.-)
DING_BACSU	38%	1007	PROBABLE ATP-DEPENDENT HELICASE DING HOMOLOG
PARC_STAAU	53%	1009	TOPOISOMERASE IV SUBUNIT A (EC 5.99.1.-)
RADC_BACSU	41%	1033	DNA REPAIR PROTEIN RADC HOMOLOG (ORFB)
SSB_BACSU	61%	1041	SINGLE-STRAND BINDING PROTEIN (SSB)
DNAD_BACSU	37%	1081	DNA REPLICATION PROTEIN DNAD
GYRA_STAAU	23%	1120	DNA GYRASE SUBUNIT A (EC 5.99.1.3)
PCRA_BACST	49%	1132	ATP-DEPENDENT HELICASE PCRA (EC 3.6.1.-)
MUTX_STRPN	34%	1134	MUTATOR MUTT PROTEIN (7,8-DIHYDRO-8-OXOGUANINE-TRIPHOSPHATASE)
DYR_LACLA	87%	1161	DIHYDROFOLATE REDUCTASE (EC 1.5.1.3)
DFPA_HAEIN	43%	1251	DNA PROCESSING CHAIN A
TOPI_BACSU	64%	1253	DNA TOPOISOMERASE I (EC 5.99.1.2)
DNAC_ECOLI	27%	1446	DNA REPLICATION PROTEIN DNAC
SSB_BACSU	68%	1447	SINGLE-STRAND BINDING PROTEIN (SSB)
MUTE_AQUPY	34%	1689	DNA MISMATCH REPAIR PROTEIN
RECQ_ECOLI	35%	1854	ATP-DEPENDENT DNA HELICASE RECQ (EC 3.6.1.-)
DING_BACSU	34%	1878	PROBABLE ATP-DEPENDENT HELICASE DING HOMOLOG
DING_BACSU	28%	1880	PROBABLE ATP-DEPENDENT HELICASE DING HOMOLOG
PRIA_BACSU	58%	1944	PRIMOSOMAL PROTEIN N' (REPLICATION FACTOR Y)
YAAA_BACSU	49%	2030	HYPOTHETICAL 7.9 KD PROTEIN IN DNAN-REC F INTERGENIC REGION
REC F_LACLC	99%	2031	REC F PROTEIN
DINP_ECOLI	42%	2085	DNA-DAMAGE-INDUCIBLE PROTEIN P
RADA_BACSU	60%	2129	DNA REPAIR PROTEIN RADA HOMOLOG (DNA REPAIR PROTEIN SMS
HOMOLOG)			
DP3A_BACSU	42%	2171	DNA POLYMERASE III, ALPHA CHAIN (EC 2.7.7.7)
DPO1_STRPN	68%	2192	DNA POLYMERASE I (EC 2.7.7.7)
SSB_BACSU	60%	2254	SINGLE-STRAND BINDING PROTEIN (SSB)
DP3X_BACSU	43%	2259	DNA POLYMERASE III SUBUNITS GAMMA AND TAU (EC 2.7.7.7)
RUVB_PSEAE	59%	2269	HOLLIDAY JUNCTION DNA HELICASE RUVB
RUVA_HAEIN	40%	2269	HOLLIDAY JUNCTION DNA HELICASE RUVA
HEXB_STRPN	57%	2271	DNA MISMATCH REPAIR PROTEIN HEXE
HEXA_STRPN	79%	2274	DNA MISMATCH REPAIR PROTEIN HEXA
DNAJ_LACLA	99%	2287	DNAJ PROTEIN
RECG_ECOLI	36%	2311	ATP-DEPENDENT DNA HELICASE RECG (EC 3.6.1.-)

BIO11. TRANSCRIPTION**27 genes****BIO11.01 Degradation of RNA****18 genes**

TRUA_ECOLI	34%	482	PSEUDOURIDYLATE SYNTHASE I (EC 4.2.1.70)
YACO_BACSU	26%	595	HYPOTHETICAL RRNA METHYLASE IN CYSS 3'REGION
MIAA_MYCLE	32%	613	PROBABLE TRNA DELTA(2)-ISOPENTENYLPIROPHOSPHATE TRANSFERASE (EC 2.5.1.8)
RNC_BACSU	44%	807	RIBONUCLEASE III (EC 3.1.26.3)
VACB_HAEIN	37%	965	VACB PROTEIN HOMOLOG
ALDC_BACBR	31%	1114	ALPHA-ACETOLACTATE DECARBOXYLASE PRECURSOR (EC 4.1.1.5)
TRUB_SYNY3	39%	1138	TRNA PSEUDOURIDINE 55 SYNTHASE (PSI55 SYNTHASE)
YGL3_BACST	57%	1160	HYPOTHETICAL 18.2 KD PROTEIN IN GLDA 3'REGION (ORF3)
ALDC_BACSU	35%	1250	ALPHA-ACETOLACTATE DECARBOXYLASE (EC 4.1.1.5)
RNH2_ECOLI	47%	1326	RIBONUCLEASE HII (EC 3.1.26.4)
SBCD_BACSU	32%	1354	EXONUCLEASE SBCD HOMOLOG (FRAGMENT)
PAPS_BACSU	39%	1598	POLY(A) POLYMERASE (EC 2.7.7.19)
RIMM_HAEIN	26%	1603	16S RRNA PROCESSING PROTEIN RIMM
PNPA_BACSU	58%	1903	POLYRIBONUCLEOTIDE NUCLEOTIDYLTRANSFERASE (EC 2.7.7.8)

Table 3.

YACO_BACSU	47%	1921	HYPOTHETICAL RRNA METHYLASE IN CYSS 3'REGION
SPLR_NPVAC	30%	2008	SPHEROIDIN-LIKE PROTEIN PRECURSOR (SPINDOLIN-LIKE PROTEIN)
SAGF_STRFY	47%	2094	STREPTOCOCCAL ACID GLYCOPROTEIN
NUSG_STRGR	43%	2154	TRANSCRIPTION ANTI-TERMINATION PROTEIN NUSG
BIO11.02 RNA synthesis, modification, and DNA transcription			8 genes
RPSD_NEIGO	24%	466	RNA POLYMERASE SIGMA FACTOR RPOD (SIGMA-70)
RPSD_LACLA	100%	544	RNA POLYMERASE SIGMA FACTOR RPOD
RPOE_BACSU	33%	621	DNA-DIRECTED RNA POLYMERASE DELTA SUBUNIT (EC 2.7.7.6)
GREA_RICPR	47%	624	TRANSCRIPTION ELONGATION FACTOR GREA (TRANSCRIPT CLEAVAGE FACTOR GREA)
RPOC_MYCGE	46%	1838	DNA-DIRECTED RNA POLYMERASE BETA' CHAIN (EC 2.7.7.6)
RPOB_SPICT	31%	1842	DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6)
RPOA_BACSU	57%	2132	DNA-DIRECTED RNA POLYMERASE ALPHA CHAIN (EC 2.7.7.6)
SIGV_BACSU	23%	2222	RNA POLYMERASE SIGMA FACTOR SIGV
BIO11.03 RNA processing			1 genes
YQFR_BACSU	48%	414	PROBABLE RNA HELICASE IN CCCA-SODA INTERGENIC REGION
BIO12. TRANSLATION			138 genes
BIO12.01 Amino acyl tRNA synthetases			26 genes
SYW_CLOLO	63%	68	TRYPTOPHANYL-TRNA SYNTHETASE (EC 6.1.1.2)
TGT_HAEIN	54%	153	QUEUINE TRNA-RIBOSYLTRANSFERASE (EC 2.4.2.29)
SYK_BACSU	58%	374	LYSYL-TRNA SYNTHETASE (EC 6.1.1.6)
SYI_BACSU	55%	388	TYROSYL-TRNA SYNTHETASE 1 (EC 6.1.1.1)
Y449_MYCFN	34%	393	HYPOTHETICAL PROTEIN MG449 HOMOLOG
SYM_BACSU	62%	798	METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10)
SVL_MYCTU	43%	826	LEUCYL-TRNA SYNTHETASE (EC 6.1.1.4)
SYGB_BACSU	41%	1101	GLYCYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.14)
SYK_BACSU	32%	1129	LYSYL-TRNA SYNTHETASE (EC 6.1.1.6)
YHI3_LACLA	97%	1228	HYPOTHETICAL 38.0 KD PROTEIN IN HISC-HISC INTERGENIC REGION (ORF3)
SYS_BACSU	57%	1747	SERYL-TRNA SYNTHETASE (EC 6.1.1.11)
SYA_THETH	41%	1759	ALANYL-TRNA SYNTHETASE (EC 6.1.1.7)
SYN_BACSU	59%	1875	ASPARAGINYL-TRNA SYNTHETASE (EC 6.1.1.22)
SYC_HELPY	43%	1898	CYSTEINYL-TRNA SYNTHETASE (EC 6.1.1.16)
SYI_STAUA	28%	1911	ISOLEUCYL-TRNA SYNTHETASE (EC 6.1.1.5)
SYI_CAMJE	43%	1912	ISOLEUCYL-TRNA SYNTHETASE (EC 6.1.1.5)
SYT_MYCLE	42%	1968	THREONYL-TRNA SYNTHETASE (EC 6.1.1.3)
SYFB_BACSU	45%	1987	PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20)
SYFA_BACSU	62%	1989	PHENYLALANYL-TRNA SYNTHETASE ALPHA CHAIN (EC 6.1.1.20)
SYD_SYNY3	51%	2021	ASPARTYL-TRNA SYNTHETASE (EC 6.1.1.12)
SYH_STRQ	63%	2023	HISTIDYL-TRNA SYNTHETASE (EC 6.1.1.21)
SYR_SYNY3	36%	2096	ARGINYL-TRNA SYNTHETASE (EC 6.1.1.19)
SYE_HELPY	45%	2120	GLUTAMYL-TRNA SYNTHETASE (EC 6.1.1.17)
SYP_HAEIN	45%	2177	PROLYL-TRNA SYNTHETASE (EC 6.1.1.15)
SYV_THETH	43%	2229	VALYL-TRNA SYNTHETASE (EC 6.1.1.9)
SYFB_MYCGE	29%	2233	PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20)
BIO12.02 Degradation of proteins, peptides, and glycopeptides			22 genes
GCP_HAEIN	41%	291	PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57)
AMPN_LACLC	98%	301	AMINOPEPTIDASE N (EC 3.4.11.2)
AMP1_SYNY3	45%	599	PUTATIVE METHIONINE AMINOPEPTIDASE A (EC 3.4.11.18)
CLPP_STRSL	87%	670	ATP-DEPENDENT CLP PROTEASE PROTEOLYTIC SUBUNIT (EC 3.4.21.92)
YQHT_BACSU	41%	688	PUTATIVE PEPTIDASE IN GCVT-SPOIIIA INTERGENIC REGION (EC 3.4.-)
PEPV_LACDL	47%	858	XAA-HIS DIPEPTIDASE (EC 3.4.13.3)
P54_ENTPC	40%	959	P54 PROTEIN PRECURSOR
CLPX_BACSU	67%	1162	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPX
CLPP_SYNP7	29%	1433	ATP-DEPENDENT CLP PROTEASE PROTEOLYTIC SUBUNIT (EC 3.4.21.92)
CLPB_HAEIN	56%	1564	CLPB PROTEIN
PEPQ_LACDL	50%	1693	XAA-PRO DIPEPTIDASE (EC 3.4.13.9)
PRTM_LACPA	32%	1761	PROTEASE MATURATION PROTEIN PRECURSOR
PEPF_LACLA	97%	1763	OLIGOENDOPEPTIDASE F (EC 3.4.24.-)
PEPO_LACLC	99%	1847	NEUTRAL ENDOPEPTIDASE (EC 3.4.24.-)
PEPC_LACLC	98%	1927	AMINOPEPTIDASE C (EC 3.4.22.-)
PTR_ECOLI	22%	2028	PROTEASE III PRECURSOR (EC 3.4.24.55)
YMXG_BACSU	23%	2029	HYPOTHETICAL PROCESSING PROTEASE (EC 3.4.-.-)
YECQ_ECOLI	37%	2049	PUTATIVE PROTEASE IN BAER-OGRK INTERGENIC REGION (EC 3.4.-.-)
LEP3_BACSU	34%	2084	TYPE 4 PREPILIN-LIKE PROTEIN SPECIFIC LEADER PEPTIDASE (EC 3.4.99.-)
PEPX_LACLC	79%	2115	XAA-PRO DIPEPTIDYL-PEPTIDASE (EC 3.4.14.11)
YYXA_BACSU	47%	2185	HYPOTHETICAL PROTEASE IN ROCR-PURA INTERGENIC REGION (EC 3.4.21.-)
HTOA_HAEIN	33%	2218	PROBABLE PERIPLASMIC SERINE PROTEASE DO/HHOA-LIKE PRECURSOR

Table 3.

BIO12.03 Nucleoproteins			2 genes
DBH_BACST	67%	500	DNA-BINDING PROTEIN II (HB)
DBH_VIBFR	37%	900	DNA-BINDING PROTEIN HU
BIO12.04 Protein modification			2 genes
CYPH_MYCTU	31%	366	PROBABLE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE (EC 5.2.1.8)
CYPH_MYCTU	37%	912	PROBABLE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE (EC 5.2.1.8)
BIO12.05 Ribosomal proteins: synthesis and modification			55 genes
PRMA_BACSU	43%	102	PROBABLE METHYLTRANSFERASE (EC 2.1.1.-)
RL34_COXBU	70%	131	50S RIBOSOMAL PROTEIN L34
RL28_BACSU	56%	193	50S RIBOSOMAL PROTEIN L28
RS21_BACSU	87%	235	30S RIBOSOMAL PROTEIN S21 (BS-B)
RS4_BACSU	62%	281	30S RIBOSOMAL PROTEIN S4 (BS4)
R14H_BACSH	35%	306	30S RIBOSOMAL PROTEIN S14 HOMOLOG
DEAD_KLEPN	45%	351	ATP-DEPENDENT RNA HELICASE DEAD
YVYD_BACSU	51%	630	HYPOTHETICAL 22.0 KD PROTEIN IN FLIT-SECA INTERGENIC REGION
RL33_LACLA	100%	659	50S RIBOSOMAL PROTEIN L33
RL9_BACST	52%	751	50S RIBOSOMAL PROTEIN L9 (BL17)
YLXQ_ENTFC	48%	773	PROBABLE RIBOSOMAL PROTEIN IN INF6 5'REGION
RS1_BUCAP	32%	852	30S RIBOSOMAL PROTEIN S1
RL19_BACST	73%	895	50S RIBOSOMAL PROTEIN L19
RS14_SYNY3	41%	908	30S RIBOSOMAL PROTEIN S14
RL21_BACSU	67%	1088	50S RIBOSOMAL PROTEIN L21 (BL20)
RL27_BACSU	74%	1089	50S RIBOSOMAL PROTEIN L27 (BL30)
RL7_BACSU	61%	1298	50S RIBOSOMAL PROTEIN L7/L12 (BL9)
RL7_STRFN	76%	1299	50S RIBOSOMAL PROTEIN L7/L12 (FRAGMENT)
RL10_BACSU	63%	1299	50S RIBOSOMAL PROTEIN L10 (BL5)
RS16_BACSU	59%	1606	30S RIBOSOMAL PROTEIN S16 (BS17)
YKGM_ECOLI	45%	1636	HYPOTHETICAL 9.9 KD PROTEIN IN INTF-EA6H INTERGENIC REGION
RL20_BACSU	69%	1891	50S RIBOSOMAL PROTEIN L20
RS15_BACSU	62%	1934	30S RIBOSOMAL PROTEIN S15 (BS18)
RL1_BACSU	67%	2058	50S RIBOSOMAL PROTEIN L1 (BL1)
RL11_STACA	80%	2059	50S RIBOSOMAL PROTEIN L11
R14H_BACSH	30%	2125	30S RIBOSOMAL PROTEIN S14 HOMOLOG
RL17_BACSU	72%	2131	50S RIBOSOMAL PROTEIN L17 (BL15)
RS11_BACSU	82%	2133	30S RIBOSOMAL PROTEIN S11 (BS11)
RS13_BACSU	70%	2133	30S RIBOSOMAL PROTEIN S13 (BS14)
RL15_STAAU	67%	2139	50S RIBOSOMAL PROTEIN L15
RS5_BACSU	77%	2140	30S RIBOSOMAL PROTEIN S5 (BS5)
RL6_BACST	63%	2141	50S RIBOSOMAL PROTEIN L6 (BL10)
RS8_BACSU	72%	2142	30S RIBOSOMAL PROTEIN S8 (BS8)
RL5_BACST	80%	2144	50S RIBOSOMAL PROTEIN L5
RL24_BACST	62%	2144	50S RIBOSOMAL PROTEIN L24
RL14_BACSU	80%	2144	50S RIBOSOMAL PROTEIN L14
RS17_BACSU	77%	2145	30S RIBOSOMAL PROTEIN S17 (BS16)
RL16_BACSU	76%	2145	50S RIBOSOMAL PROTEIN L16
RS3_BACSU	73%	2146	30S RIBOSOMAL PROTEIN S3 (BS3)
RL22_BACSU	62%	2146	50S RIBOSOMAL PROTEIN L22
RS19_BACST	77%	2147	30S RIBOSOMAL PROTEIN S19 (BS19)
RL2_BACSU	76%	2147	50S RIBOSOMAL PROTEIN L2 (BL2)
RL4_BACSU	60%	2149	50S RIBOSOMAL PROTEIN L4
RL3_BACSU	69%	2149	50S RIBOSOMAL PROTEIN L3 (BL3)
RS10_STRMU	81%	2150	30S RIBOSOMAL PROTEIN S10
RL33_MYCCA	57%	2155	50S RIBOSOMAL PROTEIN L33
RS2_PEDAC	71%	2207	30S RIBOSOMAL PROTEIN S2
RS2_BACSU	100%	2208	30S RIBOSOMAL PROTEIN S2 (BS1)
RS18_BACST	63%	2254	30S RIBOSOMAL PROTEIN S18 (BS21)
RS6_BACSU	58%	2254	30S RIBOSOMAL PROTEIN S6 (BS9)
RS9_BACSU	65%	2326	30S RIBOSOMAL PROTEIN S9 (BS10)
RL13_STACA	63%	2327	50S RIBOSOMAL PROTEIN L13
RS7_ECOLI	64%	2334	30S RIBOSOMAL PROTEIN S7
RS7_BACSU	71%	2334	30S RIBOSOMAL PROTEIN S7 (BS7)
RS12_SYNY3	67%	2335	30S RIBOSOMAL PROTEIN S12
BIO12.06 tRNA modification			8 genes
SP5C_BACSU	55%	17	PROBABLE PEPTIDYL-TRNA HYDROLASE (EC 3.1.1.29)
RNPA_BACSU	47%	129	RIBONUCLEASE P PROTEIN COMPONENT (EC 3.1.26.5)
KSGA_BACSU	52%	687	DIMETHYLADENOSINE TRANSFERASE (EC 2.1.1.-)
YGCA_HAEIN	30%	1192	HYPOTHETICAL RNA METHYLTRANSFERASE HI0333 (EC 2.1.1.-)
TRMD_HAEIN	50%	1603	TRNA (GUANINE-N1)-METHYLTRANSFERASE (EC 2.1.1.31)
MDMC_STRMY	31%	1762	O-METHYLTRANSFERASE (EC 2.1.1.-)
FMT_BACSU	46%	1941	METHIONYL-TRNA FORMYLTRANSFERASE (EC 2.1.2.9)
YGCA_HAEIN	27%	2235	HYPOTHETICAL RNA METHYLTRANSFERASE HI0333 (EC 2.1.1.-)
BIO12.07 Translation factors			23 genes
YDGI_BACSU	30%	183	PUTATIVE NAD(P)H NITROREDUCTASE (EC 1.-.-.-)

Table 3.

RF3_SYNY3	47%	349	PEPTIDE CHAIN RELEASE FACTOR 3 (RF-3)
RF2_HELPY	25%	451	PEPTIDE CHAIN RELEASE FACTOR 2 (RF-2)
IF2_ENTFC	36%	543	TRANSLATION INITIATION FACTOR IF-2
RF1_BACSU	54%	584	PEPTIDE CHAIN RELEASE FACTOR 1 (RF-1)
FLAV_DESVH	38%	595	FLAVODOXIN
EFP_BACSU	48%	689	ELONGATION FACTOR P (EF-P)
IF2_ENTFC	69%	775	TRANSLATION INITIATION FACTOR IF-2
RBFA_BACSU	37%	777	RIBOSOME-BINDING FACTOR A (P15B PROTEIN)
RF2_BACSU	53%	996	PEPTIDE CHAIN RELEASE FACTOR 2 (RF-2)
YDGI_BACSU	49%	1500	PUTATIVE NAD(P)H NITROREDUCTASE (EC 1.-.-.-)
MVIM_ECOLI	34%	1872	VIRULENCE FACTOR MVIM HOMOLOG
IF3_BACSU	61%	1892	TRANSLATION INITIATION FACTOR IF-3
EFTU_STROR	80%	1909	ELONGATION FACTOR TU (EF-TU)
RRF_HAEIN	43%	2066	RIBOSOME RECYCLING FACTOR (RIBOSOME RELEASING FACTOR)
YIHK_ECOLI	55%	2073	65.4 KD GTP-BINDING PROTEIN IN GLNA-RBN INTERGENIC REGION
IF1_LACLA	100%	2134	TRANSLATION INITIATION FACTOR IF-1
FMNB_DESVM	39%	2161	FMN-BINDING PROTEIN
EFTS_HELPY	43%	2206	ELONGATION FACTOR TS (EF-TS)
EFTS_STRPN	71%	2207	ELONGATION FACTOR TS (EF-TS)
EFTS_BACSU	77%	2207	ELONGATION FACTOR TS (EF-TS)
EFG_BACSU	72%	2333	ELONGATION FACTOR G (EF-G)
EFG_THEMA	55%	2333	ELONGATION FACTOR G (EF-G)
BIO13. TRANSPORT AND BINDING PROTEINS			185 genes
BIO13.01 General			35 genes
LCNC_LACLA	94%	84	LACTOCOCCIN A TRANSPORT ATP-BINDING PROTEIN LCNC
RFBB_MYXXA	40%	202	O-ANTIGEN EXPORT SYSTEM ATP-BINDING PROTEIN RFBB
CHVA_AGRTU	35%	260	BETA-(1-->2)GLUCAN EXPORT ATP-BINDING PROTEIN CHVA
P69_MYCHR	26%	296	TRANSPORT SYSTEM PERMEASE PROTEIN P69
ABC_ECOLI	42%	319	ATP-BINDING PROTEIN ABC
LCN3_LACLA	34%	517	LACTICIN 481/LACTOCOCCIN TRANSPORT ATP-BINDING PROTEIN LCNDR3
UUP1_HAEIN	29%	560	ABC TRANSPORTER ATP-BINDING PROTEIN UUP-1
LMRA_STRLN	34%	597	LINCOMYCIN RESISTANCE PROTEIN
KUP_ECOLI	36%	608	KUP SYSTEM POTASSIUM UPTAKE PROTEIN
KUP_ECOLI	29%	610	KUP SYSTEM POTASSIUM UPTAKE PROTEIN
CYDC_BACSU	48%	709	TRANSPORT ATP-BINDING PROTEIN CYDC
CYDD_BACSU	45%	711	TRANSPORT ATP-BINDING PROTEIN CYDD
NRDH_LACLC	99%	1003	GLUTAREDOXIN-LIKE PROTEIN NRDH
PERM_HAEIN	26%	1196	PUTATIVE PERMEASE PERM HOMOLOG
ATMB_SALTY	51%	1284	MG(2+) TRANSPORT ATPASE, P-TYPE 2 (EC 3.6.1.-)
MGLA_HAEIN	23%	1352	GALACTOSIDE TRANSPORT ATP-BINDING PROTEIN MGLA
MGLA_MYCGE	32%	1364	GALACTOSIDE TRANSPORT ATP-BINDING PROTEIN MGLA HOMOLOG
AC22_STRCO	26%	1383	PROBABLE ACTINORHODIN TRANSPORTER
LPLC_BACSU	42%	1527	LPLC PROTEIN
LPLB_BACSU	27%	1528	LPLB PROTEIN
LCN3_LACLA	41%	1559	LACTICIN 481/LACTOCOCCIN TRANSPORT ATP-BINDING PROTEIN LCNDR3
COMA_STRPN	24%	1560	TRANSPORT ATP-BINDING PROTEIN COMA
AC22_STRCO	28%	1584	PROBABLE ACTINORHODIN TRANSPORTER
UUP_ECOLI	31%	1592	ABC TRANSPORTER ATP-BINDING PROTEIN UUP
YJJK_HAEIN	38%	1593	ABC TRANSPORTER ATP-BINDING PROTEIN HI1252
MVAA_PSEMV	42%	1607	3-HYDROXY-3-METHYLGLUTARYL-COENZYME A REDUCTASE (EC 1.1.1.88)
YQGI_BACSU	54%	1753	PROBABLE ABC TRANSPORTER PERMEASE PROTEIN IN SODA-COMGA
INTERGENIC REGION (ORF73)			
YQGH_BACSU	32%	1754	PROBABLE ABC TRANSPORTER PERMEASE PROTEIN IN SODA-COMGA
INTERGENIC REGION (ORF72)			
YQGG_BACSU	47%	1755	PROBABLE ABC TRANSPORTER BINDING PROTEIN IN SODA-COMGA
INTERGENIC REGION PRECURSOR (ORF108)			
NOCQ_AGRT5	32%	1835	NOPALINE TRANSPORT SYSTEM PERMEASE PROTEIN NOCQ
PERM_ECOLI	27%	1863	PUTATIVE PERMEASE PERM
UVRA_NEIGO	59%	1866	EXCINUCLEASE ABC SUBUNIT A
ECSB_BACSU	28%	2054	PROTEIN ECSB
ECSA_BACSU	60%	2055	ABC-TYPE TRANSPORTER ATP-BINDING PROTEIN ECSA
FLIP_BORBU	27%	2113	FLAGELLAR BIOSYNTHETIC PROTEIN FLIP
BIO13.02 Amino acids, peptides and amines			52 genes
YW40_MYCTU	25%	70	HYPOTHETICAL TRANSPORT PROTEIN CY39.40 (FRAGMENT)
GABP_ECOLI	23%	70	GABA PERMEASE (4-AMINO BUTYRATE TRANSPORT CARRIER)
LYSP_ECOLI	27%	111	LYSINE-SPECIFIC PERMEASE
PTHP_STRSL	81%	118	PHOSPHOCARRIER PROTEIN HPR (HISTIDINE-CONTAINING PROTEIN)
PTI_STRSL	82%	119	PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (EC 2.7.3.9)
ALYS_ENTFA	49%	267	AUTOLYSIN (EC 3.5.1.28)
YQIZ_BACSU	34%	322	PROBABLE AMINO-ACID ABC TRANSPORTER ATP-BINDING PROTEIN IN
BMRU-ANSR INTERGENIC REGION			
AMIA_STRPN	31%	341	OLIGOPEPTIDE-BINDING PROTEIN AMIA PRECURSOR
OPPA_ECOLI	28%	343	PERIPLASMIC OLIGOPEPTIDE-BINDING PROTEIN PRECURSOR
OPPB_BACSU	46%	345	OLIGOPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPPB

Table 3.

DPPC_BACSU	43%	346	DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPC
OPPF_BACSU	61%	348	OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN OPPF
CYCA_ECOLI	28%	358	D-SERINE/D-ALANINE/GLYCINE TRANSPORTER
AAPA_BACSU	47%	360	AMINO ACID PERMEASE AAPA
LYSP_ECOLI	48%	368	LYSINE-SPECIFIC PERMEASE
OPPA_BACSU	30%	523	OLIGOPEPTIDE-BINDING PROTEIN OPPIA PRECURSOR
ALYS_ENTFA	35%	525	AUTOLYSIN (EC 3.5.1.28)
DEF_HELPHY	39%	553	POLYPEPTIDE DEFORMYLASE (EC 3.5.1.31)
YQIX_BACSU	36%	556	PROBABLE AMINO-ACID ABC TRANSPORTER BINDING PROTEIN IN BMRU-ANSR INTERGENIC REGION PRECURSOR
BRNQ_LACDL	47%	683	BRANCHED-CHAIN AMINO ACID TRANSPORT SYSTEM CARRIER PROTEIN (BRANCHED-CHAIN AMINO ACID UPTAKE CARRIER)
DTPPT_LACLA	97%	703	DI-/TRIPETIDE TRANSPORTER
YDGF_BACSU	33%	817	HYPOTHETICAL TRANSPORT PROTEIN IN EXPZ-DINB INTERGENIC REGION
OPUA_BACSU	34%	863	GLYCINE BETAINE TRANSPORT ATP-BINDING PROTEIN OPUAA
PROW_BACSU	42%	863	GLYCINE BETAINE/L-PROLINE TRANSPORT SYSTEM PERMEASE PROTEIN PROW
YCKK_BACSU	28%	960	PROBABLE AMINO-ACID ABC TRANSPORTER BINDING PROTEIN PRECURSOR
Y4TG_RHISN	32%	961	PROBABLE AMINO-ACID ABC TRANSPORTER PERMEASE PROTEIN Y4TG
GLTL_HAEIN	59%	962	GLUTAMATE/ASPARTATE TRANSPORT ATP-BINDING PROTEIN GLTL
POTA_HAEIN	45%	1173	SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POTA
POTC_HAEIN	35%	1175	SPERMIDINE/PUTRESCINE TRANSPORT SYSTEM PERMEASE PROTEIN POTC
POTD_ECOLI	43%	1176	SPERMIDINE/PUTRESCINE-BINDING PERIPLASMIC PROTEIN PRECURSOR (SPBP)
GRSB_BACBR	30%	1290	GRAMICIDIN S SYNTHETASE II (GRAMICIDIN S BIOSYNTHESIS GRSB PROTEIN) (EC 6.-.-.-)
XASA_ECOLI	50%	1323	AMINO ACID ANTIPOINTER (EXTREME ACID SENSITIVITY PROTEIN)
OPUC_BACSU	47%	1471	GLYCINE BETAINE-BINDING PROTEIN PRECURSOR
OPUB_BACSU	51%	1471	GLYCINE BETAINE TRANSPORT SYSTEM PERMEASE PROTEIN OPUAB
OPUA_BACSU	58%	1473	GLYCINE BETAINE TRANSPORT ATP-BINDING PROTEIN OPUAA
PMSR_BACSU	54%	1590	PUTATIVE PEPTIDE METHIONINE SULFOXIDE REDUCTASE (PEPTIDE MET
GLNH_ECOLI	36%	1797	GLUTAMINE-BINDING PERIPLASMIC PROTEIN PRECURSOR (GLNBP)
YA79_HAEIN	38%	1798	HYPOTHETICAL AMINO-ACID ABC TRANSPORTER PERMEASE PROTEIN HI0179
YQIZ_BACSU	64%	1799	PROBABLE AMINO-ACID ABC TRANSPORTER ATP-BINDING PROTEIN IN BMRU-ANSR INTERGENIC REGION
GLNQ_ECOLI	56%	1833	GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLNQ
HISP_ECOLI	51%	1835	HISTIDINE TRANSPORT ATP-BINDING PROTEIN HISP
PEPT_LACLC	95%	1857	PEPTIDASE T (EC 3.4.11.-)
YDGF_BACSU	35%	1870	HYPOTHETICAL TRANSPORT PROTEIN IN EXPZ-DINB INTERGENIC REGION
OPPIA_LACLA	90%	1885	OLIGOPEPTIDE-BINDING PROTEIN OPPIA PRECURSOR
OPPC_LACLA	100%	1887	OLIGOPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPPIA
OPPB_LACLC	100%	1888	OLIGOPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPPIB
OPPF_LACLA	100%	1889	OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN OPIA
OPPD_LACLC	100%	1890	OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN OPPID
PMSR_BACSU	53%	2064	PUTATIVE PEPTIDE METHIONINE SULFOXIDE REDUCTASE (PEPTIDE MET
ARCD_PSEAE	35%	2092	ARGININE/ORNITHINE ANTIPOINTER
YXEM_BACSU	33%	2118	PROBABLE AMINO-ACID ABC TRANSPORTER BINDING PROTEIN IN IDH-DEOR INTERGENIC REGION PRECURSOR
LYSP_ECOLI	37%	2257	LYSINE-SPECIFIC PERMEASE
BIO13.03 Anions			13 genes
MEGL_PSEPU	38%	77	METHIONINE GAMMA-LYASE (EC 4.4.1.11)
ATZN_ECOLI	28%	95	ZINC-TRANSPORTING ATPASE (EC 3.6.1.-)
SNTD_HAEIN	25%	180	PROBABLE 5'-NUCLEOTIDASE PRECURSOR (EC 3.1.3.5)
DHAK_CITFR	35%	243	DIHYDROXYACETONE KINASE (EC 2.7.1.29)
CN16_YEREN	31%	298	2',3'-CYCLIC-NUCLEOTIDE 2'-PHOSPHODIESTERASE PRECURSOR (EC 3.1.4.16)
GLPT_BACSU	34%	546	GLYCEROL-3-PHOSPHATE TRANSPORTER (G-3-P TRANSPORTER)
MANA_STRMU	65%	777	MANNOSE-6-PHOSPHATE ISOMERASE (EC 5.3.1.8)
ATKA_ENTFA	41%	845	POTASSIUM/COPPER-TRANSPORTING ATPASE A (EC 3.6.1.36)
CADA_BACFI	28%	869	PROBABLE CADMIUM-TRANSPORTING ATPASE (EC 3.6.1.-)
SNTD_HAEIN	33%	1098	PROBABLE 5'-NUCLEOTIDASE PRECURSOR (EC 3.1.3.5)
SCRK_STRMU	62%	1516	FRUCTOKINASE (EC 2.7.1.4)
PTA_BACSU	61%	1704	PROBABLE PHOSPHATE ACETYLTRANSFERASE (EC 2.3.1.8)
SPHX_SYNP7	28%	1756	SPHX PROTEIN PRECURSOR
BIO13.04 Carbohydrates, organic alcohols and acids			13 genes
GLPF_ECOLI	35%	245	GLYCEROL UPTAKE FACILITATOR PROTEIN
FECE_ECOLI	42%	324	IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
FECD_ECOLI	35%	326	IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD
CITN_LACLA	48%	921	CITRATE-SODIUM SYMPORT (CITRATE TRANSPORTER)
GLPF_STRPN	58%	1267	GLYCEROL UPTAKE FACILITATOR PROTEIN
RBSB_HAEIN	45%	1679	D-RIBOSE-BINDING PERIPLASMIC PROTEIN PRECURSOR
RBSC_BACSU	60%	1680	RIBOSE TRANSPORT SYSTEM PERMEASE PROTEIN RBSC
RBSA_BACSU	42%	1681	RIBOSE TRANSPORT ATP-BINDING PROTEIN RBSA
RBSD_BACSU	56%	1682	HIGH AFFINITY RIBOSE TRANSPORT PROTEIN RBSD
LACY_STRTR	43%	2043	LACTOSE PERMEASE (LACTOSE-PROTON SYMPORT)

Table 3.

RBSA_HAEIN	24%	2163	RIBOSE TRANSPORT ATP-BINDING PROTEIN RBSA
EBA2_FLAME	27%	2238	ENDO-BETA-N-ACETYLGALACTOSAMINIDASE F2 PRECURSOR (EC 3.2.1.96)
GNTP_BACLI	59%	2246	GLUCONATE PERMEASE
BIO13.05 Cations			
FE0B_ECOLI	39%	188	FERROUS IRON TRANSPORT PROTEIN B
FE0A_ECOLI	38%	190	FERROUS IRON TRANSPORT PROTEIN A
TCR1_ECOLI	41%	303	TETRACYCLINE RESISTANCE PROTEIN, CLASS A (TRANSPOSON TN1721)
FHUB_BACSU	37%	325	FERRIC IRON TRANSPORT PERMEASE PROTEIN FHUB
Y1D2_MYCTU	36%	567	PUTATIVE CATION-TRANSPORTING ATPASE CY21C12.02 (EC 3.6.1.-)
ATA1_SYNY3	40%	675	CATION-TRANSPORTING ATPASE PMA1 (EC 3.6.1.-)
NHAA_ECOLI	42%	1148	NA(+)/H
PBUX_BACSU	48%	1158	XANTHINE PERMEASE
Y1D2_MYCTU	35%	1188	PUTATIVE CATION-TRANSPORTING ATPASE CY21C12.02 (EC 3.6.1.-)
ATCL_SYNP7	37%	1402	CATION-TRANSPORTING ATPASE PACL (EC 3.6.1.-)
FUR_CAMGE	30%	1503	FERRIC UPTAKE REGULATION PROTEIN
UIDB_ECOLI	33%	1541	GLUCURONIDE CARRIER PROTEIN (GLUCURONIDE PERMEASE)
Y058_METJA	31%	1632	PUTATIVE AMMONIUM TRANSPORTER MJG058
URAA_HAEIN	36%	1643	PROBABLE URACIL PERMEASE (URACIL TRANSPORTER)
FER_CLOTH	38%	1741	FERREDOXIN
NAPA_ENTHR	48%	1973	NA(+)/H
ATCS_SYNY3	36%	2061	CATION-TRANSPORTING ATPASE PACS (EC 3.6.1.-)
ATA1_SYNY3	36%	2315	CATION-TRANSPORTING ATPASE PMA1 (EC 3.6.1.-)
BIO13.06 Nucleosides, purines and pyrimidines			
PDP_BACST	56%	1463	PYRIMIDINE-NUCLEOSIDE PHOSPHORYLASE (EC 2.4.2.2)
BIO13.07 PTS system			
PTMA_ECOLI	46%	31	PTS SYSTEM, MANNITOL-SPECIFIC IIABC COMPONENT (EIIABC-MTL)
PTCC_BACSU	34%	175	PTS SYSTEM, CELLOBIOSE-SPECIFIC IIC COMPONENT (EIIIC-CEL)
PTCB_BACSU	52%	416	PTS SYSTEM, CELLOBIOSE-SPECIFIC IIB COMPONENT (EIIIB-CEL)
PTCA_BACSU	48%	417	PTS SYSTEM, CELLOBIOSE-SPECIFIC IIA COMPONENT (EIIIA-CEL)
PTGA_STRMU	48%	435	PTS SYSTEM, SUCROSE-SPECIFIC IIABC COMPONENT (EIIABC-SCR)
PTSA_STRMU	31%	436	PTS SYSTEM, SUCROSE-SPECIFIC IIABC COMPONENT (EIIABC-SCR)
PTCC_BACSU	31%	830	PTS SYSTEM, CELLOBIOSE-SPECIFIC IIC COMPONENT (EIIIC-CEL)
PTFB_RHOCA	52%	984	PTS SYSTEM, FRUCTOSE-SPECIFIC IIBC COMPONENT (EIIIBC-FRU)
PTBA_BACSU	37%	1144	PTS SYSTEM, BETA-GLUCOSIDES-SPECIFIC IIABC COMPONENT (EIIABC-BGL)
PTBA_ERWCH	41%	1489	PTS SYSTEM, BETA-GLUCOSIDES-SPECIFIC IIABC COMPONENT (EIIABC-BGL)
PTNA_ECOLI	41%	1742	PTS SYSTEM, MANNANOSE-SPECIFIC IIAB COMPONENT (EIIAB-MAN)
PTNC_ECOLI	46%	1743	PTS SYSTEM, MANNANOSE-SPECIFIC IIC COMPONENT (EIIIC-MAN)
PTND_ECOLI	57%	1744	PTS SYSTEM, MANNANOSE-SPECIFIC IID COMPONENT (EIIID-MAN)
BIO13.08 Multidrug resistance			
BMR2_BACSU	46%	123	MULTIDRUG RESISTANCE PROTEIN 2 (MULTIDRUG-EFFLUX TRANSPORTER 2)
EMRB_ECOLI	26%	930	MULTIDRUG RESISTANCE PROTEIN B
MDL_ECOLI	36%	1302	MULTIDRUG RESISTANCE-LIKE ATP-BINDING PROTEIN MDL
BIO13.09 Other			
YBJZ_ECOLI	34%	15	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ
YCHM_ECOLI	37%	53	HYPOTHETICAL 58.4 KD PROTEIN IN PTH-PRSA INTERGENIC REGION
YBIT_ECOLI	32%	74	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBIT
YHES_ECOLI	35%	100	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YHES
YHIH_ECOLI	50%	248	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YHIH
YFIB_BACSU	39%	257	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN 1 IN GLVBC
3'REGION			
Y179_MYCGE	44%	273	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN MG179
Y180_MYCGE	47%	274	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN MG180
Y412_METJA	37%	295	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN MJ0412
YFIB_BACSU	33%	307	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN 1 IN GLVBC
3'REGION			
YFIC_BACSU	41%	309	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN 2 IN GLVBC
3'REGION			
YARE_ECOLI	44%	320	HYPOTHETICAL ABC TRANSPORTER PERMEASE PROTEIN YARE
YHIH_ECOLI	27%	321	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YHIH
Y081_MYCTU	44%	347	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY50.01
YCVJ_ECOLI	60%	426	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YCVJ
YCFV_ECOLI	41%	650	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YCFV
YVJA_BACSU	35%	715	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN IN ACDA
5'REGION			
YBJZ_ECOLI	40%	718	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ
Y796_METJA	35%	892	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN MJ0796
Y415_SYNY3	35%	990	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN SLL0415
YRHG_BACSU	43%	993	HYPOTHETICAL 28.5 KD PROTEIN IN SIGV-GREA INTERGENIC REGION
YBJZ_ECOLI	25%	998	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ
YDIF_BACSU	52%	1072	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YDIF

Table 3.

Y796_METJA	48%	1108	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN MJ0796
YLUD_LACLA 3'REGION	100%	1242	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN IN LEUD
YSC1_STRGC 5'REGION (ORF1)	62%	1349	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN IN SCAA
YXBC_BACSU INTERGENIC REGION	51%	1538	HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN HTPG-IOLR
YJEM_ECOLI (O514)	22%	1736	HYPOTHETICAL 56.3 KD PROTEIN IN GENX-PSD INTERGENIC REGION
YEJF_ECOLI	27%	1751	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YEJF
YQGJ_BACSU INTERGENIC REGION (ORF74)	60%	1751	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN IN SODA-COMGA
YQGJ_BACSU INTERGENIC REGION (ORF74)	59%	1752	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN IN SODA-COMGA
YXDL_BACSU 3'REGION	49%	1789	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN IN IDH
YATR_BACFI	43%	1815	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN
YJEM_ECOLI (O514)	39%	1855	HYPOTHETICAL 56.3 KD PROTEIN IN GENX-PSD INTERGENIC REGION
YXDL_BACSU 3'REGION	35%	2076	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN IN IDH
YYBT_BACSU INTERGENIC REGION	28%	2265	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN IN COTF-TETB
Y796_METJA	42%	2329	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN MJ0796

BIO14. OTHER CATEGORIES

94 genes

BIO14.01 Adaptations and atypical conditions

13 genes

CSP2_LACPL	82%	170	COLD SHOCK PROTEIN 2
TCMA_STRGA	28%	233	TETRACENOMYCIN C RESISTANCE AND EXPORT PROTEIN
MMR_BACSU	30%	382	METHYLENOMYCIN A RESISTANCE PROTEIN (MMR PEPTIDE)
CSPC_BACSU	75%	515	COLD SHOCK PROTEIN CSPC
TCMA_STRGA	25%	596	TETRACENOMYCIN C RESISTANCE AND EXPORT PROTEIN
VANZ_ENTPC	42%	1097	VANZ PROTEIN
FSR_ECOLI	25%	1182	FORMIDOMYCIN RESISTANCE PROTEIN
ARSC_BACSU	51%	1409	PUTATIVE ARSENATE REDUCTASE
TCMA_STRGA	29%	1553	TETRACENOMYCIN C RESISTANCE AND EXPORT PROTEIN
TCMA_STRGA	27%	1678	TETRACENOMYCIN C RESISTANCE AND EXPORT PROTEIN
CMCT_NOCIA	34%	1925	CEPHAMYCIN EXPORT PROTEIN CMCT
FRI_LISIN	44%	2082	NON-HEME IRON-CONTAINING FERRITIN
MSCL_ECOLI	44%	2150	LARGE CONDUCTANCE MECHANOSENSITIVE CHANNEL

BIO14.02 Drug and analog sensitivity

1 genes

KKH_LACLA	100%	1232	PROBABLE AMINOGLYCOSIDE 3'-PHOSPHOTRANSFERASE (EC 2.7.1.95)
-----------	------	------	---

BIO14.03 Phage related functions and prophages

21 genes

TERS_BPSP6	37%	36	TERMINASE SMALL SUBUNIT (GIP)
Y367_METJA	26%	49	PROBABLE INTEGRASE/RECOMBINASE PROTEIN MJ0367
VINT_BPL54	25%	445	INTEGRASE
VG64_BPML5	45%	460	GENE 64 PROTEIN (GP64)
VP3_BPHK7	21%	464	PORTAL PROTEIN (GP3)
VP4_BPHK7	31%	465	PUTATIVE PROHEAD PROTEASE (GP4)
VINT_BPL54	31%	501	INTEGRASE
XERC_SALTY	32%	633	INTEGRASE/RECOMBINASE XERC
YHGE_BACSU (ORFB)	31%	902	HYPOTHETICAL 84.1 KD PROTEIN IN HEMY-GL/TT INTERGENIC REGION
TNPA_STAAU	45%	1034	TRANSPOSASE A (TRANSPOSON TN554)
VP4_BPHK7	27%	1053	PUTATIVE PROHEAD PROTEASE (GP4)
VG13_BPPH2	33%	1064	MORPHOGENESIS PROTEIN 1 (LATE PROTEIN GP13)
VINT_BPL54	32%	1210	INTEGRASE
CODV_BACSU	31%	1256	PROBABLE INTEGRASE/RECOMBINASE CODV
XERC_HABIN	37%	1257	INTEGRASE/RECOMBINASE XERC
SBCC_ECOLI	44%	1350	EXONUCLEASE SBCC
MCP_BPF41	48%	1358	MAJOR CAPSID PROTEIN (MCP)
VINT_BPSP6	26%	1454	INTEGRASE
PIP_LACLA	96%	1716	PHAGE INFECTION PROTEIN
VG37_BFML5	48%	1873	GENE 37 PROTEIN (GP37)
VINT_BPL54	25%	2005	INTEGRASE

BIO14.04 Transposon related functions

59 genes

YI3A_LACJO	30%	53	INSERTION ELEMENT IS1223 HYPOTHETICAL 20.7 KD PROTEIN (ORFA)
T150_ECOLI	41%	53	PUTATIVE TRANSPOSASE FOR INSERTION SEQUENCE IS150
TRA4_LACLA	100%	54	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
YI91_SHIDY	25%	55	INSERTION ELEMENT IS911 HYPOTHETICAL 12.7 KD PROTEIN
TRA9_MYCTU (ORFB)	30%	89	PUTATIVE TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS986
TRA4_LACLA	39%	91	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904

Table 3.

YI91_SHIDY	25%	135	INSERTION ELEMENT IS911 HYPOTHETICAL 12.7 KD PROTEIN
TRA4_LACLA	89%	135	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
YI6A_MYCMY	31%	136	INSERTION ELEMENT IS1296 HYPOTHETICAL 21.4 KD PROTEIN (ORFA)
T150_ECOLI	41%	137	PUTATIVE TRANSPOSASE FOR INSERTION SEQUENCE IS150
TRA4_LACLA	99%	138	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
YI91_SHIDY	25%	138	INSERTION ELEMENT IS911 HYPOTHETICAL 12.7 KD PROTEIN
TRA4_LACLA	98%	371	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
YI3A_LACJO	30%	372	INSERTION ELEMENT IS1223 HYPOTHETICAL 20.7 KD PROTEIN (ORFA)
T150_ECOLI	41%	373	PUTATIVE TRANSPOSASE FOR INSERTION SEQUENCE IS150
TRA4_LACLA	100%	624	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
YI91_SHIDY	26%	625	INSERTION ELEMENT IS911 HYPOTHETICAL 12.7 KD PROTEIN
T150_ECOLI	41%	626	PUTATIVE TRANSPOSASE FOR INSERTION SEQUENCE IS150
YI3A_LACJO	30%	626	INSERTION ELEMENT IS1223 HYPOTHETICAL 20.7 KD PROTEIN (ORFA)
TRA4_LACLA	39%	648	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
TRAI_STRSL	38%	680	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
TRAI_STRSL	38%	705	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
TRA4_LACLA	39%	726	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
YI91_SHISO	28%	727	INSERTION ELEMENT IS600 (ISO-S3) HYPOTHETICAL 11 KD PROTEIN
TRA4_LACLA	98%	834	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
YI91_SHIDY	25%	834	INSERTION ELEMENT IS911 HYPOTHETICAL 12.7 KD PROTEIN
T150_ECOLI	39%	835	PUTATIVE TRANSPOSASE FOR INSERTION SEQUENCE IS150
YI3A_LACJO	30%	836	INSERTION ELEMENT IS1223 HYPOTHETICAL 20.7 KD PROTEIN (ORFA)
YI91_SHIDY	25%	836	INSERTION ELEMENT IS911 HYPOTHETICAL 12.7 KD PROTEIN
TRA4_LACLA	99%	837	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
TRAI_STRSL	38%	955	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
TRA4_LACLA	39%	1214	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
YI91_SHISO	28%	1215	INSERTION ELEMENT IS600 (ISO-S3) HYPOTHETICAL 11 KD PROTEIN
YI91_SHISO	28%	1219	INSERTION ELEMENT IS600 (ISO-S3) HYPOTHETICAL 11 KD PROTEIN
TRA4_LACLA	39%	1220	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
TRA5_LACLA	97%	1223	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS905
TRA4_LACLA	39%	1273	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
YI91_SHISO	28%	1274	INSERTION ELEMENT IS600 (ISO-S3) HYPOTHETICAL 11 KD PROTEIN
TRAI_STRSL	38%	1336	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
TRAI_STRSL	38%	1394	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
TRAI_STRSL	38%	1552	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
TRA4_LACLA	39%	1582	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
TRA4_LACLA	39%	1727	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
YI91_SHISO	28%	1728	INSERTION ELEMENT IS600 (ISO-S3) HYPOTHETICAL 11 KD PROTEIN
TRAI_STRSL	38%	1734	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
TRAI_STRSL	38%	1933	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
TRAI_STRSL	38%	1958	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
TRAI_STRSL	38%	1997	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
TRAI_STRSL	42%	2063	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
TRA4_LACLA	39%	2083	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
TRAI_STRSL	38%	2127	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
TRA4_LACLA	98%	2135	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
T150_ECOLI	41%	2136	PUTATIVE TRANSPOSASE FOR INSERTION SEQUENCE IS150
TRAI_STRSL	38%	2183	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
TRA4_LACLA	100%	2195	TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS904
YI91_SHIDY	25%	2196	INSERTION ELEMENT IS911 HYPOTHETICAL 12.7 KD PROTEIN
YH21_HAEIN	47%	2196	HYPOTHETICAL TRANSPOSASE-LIKE PROTEIN HI1721
YI3A_LACJO	31%	2197	INSERTION ELEMENT IS1223 HYPOTHETICAL 20.7 KD PROTEIN (ORFA)
TRAI_STRSL	38%	2248	PROBABLE TRANSPOSASE FOR INSERTION-LIKE SEQUENCE ELEMENT IS1161
BIO14.05 Other			0 genes
BIO15. HYPOTHETICAL			315 genes
BIO15.01 General			315 genes
YABO_BACSU	63%	21	HYPOTHETICAL 9.7 KD PROTEIN IN MPD-DIVIC INTERGENIC REGION
YACA_BACSU	32%	24	HYPOTHETICAL 55.1 KD PROTEIN IN SPOIIE-HPT INTERGENIC REGION
YE12_HAEIN	35%	43	HYPOTHETICAL PROTEIN HI1412
YAFP_ECOLI	36%	50	HYPOTHETICAL 17.6 KD PROTEIN IN DIMP-PRFH INTERGENIC REGION
YJHQ_ECOLI (F181)	29%	56	HYPOTHETICAL 20.0 KD PROTEIN IN PECI-FIMB INTERGENIC REGION
Y293_MYCGE	41%	66	HYPOTHETICAL PROTEIN MG293
YPGQ_BACSU	35%	67	HYPOTHETICAL 23.1 KD PROTEIN IN BSAA-ILVD INTERGENIC REGION
YCSE_BACSU	29%	73	HYPOTHETICAL 28.1 KD PROTEIN IN SIPU 3'REGION
Y531_METJA	34%	79	HYPOTHETICAL PROTEIN MJ0531
Y0A9_MYCTU	42%	97	HYPOTHETICAL 47.5 KD PROTEIN CY9C4.09
YADE_ECOLI PRECURSOR	29%	104	HYPOTHETICAL 46.3 KD PROTEIN IN HPT-PAND INTERGENIC REGION
YIHZ_HAEIN	52%	108	HYPOTHETICAL PROTEIN HI0670
YIHX_ECOLI (O206)	29%	109	HYPOTHETICAL 23.5 KD PROTEIN IN GLNA-RBN INTERGENIC REGION
YVMB_BACSU	26%	112	HYPOTHETICAL 19.7 KD PROTEIN IN CYPX-UVRA INTERGENIC REGION
YKXH_BACSU	58%	118	HYPOTHETICAL 7.4 KD PROTEIN IN PTSX OPERON (PROTEIN K)

Table 3.

YQKD_BACSU	32%	121	HYPOTHETICAL	34.6	KD	PROTEIN	IN	GLNQ-ANSR	INTERGENIC REGION
YCEE_ECOLI (ORF1)	44%	127	HYPOTHETICAL	43.9	KD	PROTEIN	IN	MSYB-WTRB	INTERGENIC REGION
YCSN_BACSU	38%	132	HYPOTHETICAL	34.1	KD	PROTEIN	IN	PBPC 3'REGION	
YPH1_MYCCA	29%	133	HYPOTHETICAL	37.9	KD	PROTEIN	IN	PTSH 3'REGION (ORF1)	
YQGF_ECOLI (O180)	35%	139	HYPOTHETICAL	15.2	KD	PROTEIN	IN	GSHB-ANSB	INTERGENIC REGION
YWFA_BACSU	22%	155	HYPOTHETICAL	44.5	KD	PROTEIN	IN	ROCC-PTA	INTERGENIC REGION
YXIO_BACSU	27%	156	HYPOTHETICAL	47.3	KD	PROTEIN	IN	WAPA-LICT	INTERGENIC REGION
YPNP_BACSU	30%	169	HYPOTHETICAL	48.5	KD	PROTEIN	IN	ILVA 3'REGION	
YQGN_BACSU	36%	170	HYPOTHETICAL	21.4	KD	PROTEIN	IN	SODA-COMGA	INTERGENIC REGION
YQGP_BACSU	33%	171	HYPOTHETICAL	56.4	KD	PROTEIN	IN	SODA-COMGA	INTERGENIC REGION
YFGB_ECOLI	36%	182	HYPOTHETICAL	43.1	KD	PROTEIN	IN	NDK-GCPE	INTERGENIC REGION
YWAB_BACSU	27%	185	HYPOTHETICAL	33.8	KD	PROTEIN	IN	DAE-TYRZ	INTERGENIC REGION
Y448_MYCPN	57%	190	HYPOTHETICAL			PROTEIN	MG448	HOMOLOG	
Y369_MYCGE	31%	192	HYPOTHETICAL			PROTEIN	MG369		
YQHY_BACSU	32%	193	HYPOTHETICAL	14.7	KD	PROTEIN	IN	ACCC-FOLD	INTERGENIC REGION
Y320_MYCPN	26%	194	HYPOTHETICAL			PROTEIN	MG320	HOMOLOG	
YIBD_ECOLI	30%	204	HYPOTHETICAL	40.5	KD	PROTEIN	IN	SECB-TDH	INTERGENIC REGION
Y4GN_RHISN	28%	205	HYPOTHETICAL	45.0	KD	PROTEIN	Y4GN		
YC22_METJA	28%	209	HYPOTHETICAL			PROTEIN	MJ1222		
YC11_KLEPN	20%	211	HYPOTHETICAL	55.0	KD	PROTEIN	IN	CPS REGION (ORF11)	
YIBD_ECOLI	34%	214	HYPOTHETICAL	40.5	KD	PROTEIN	IN	SECB-TDH	INTERGENIC REGION
Y060_MYCPN	33%	215	HYPOTHETICAL			PROTEIN	MG060	HOMOLOG	
YQEH_BACSU	55%	221	HYPOTHETICAL	41.0	KD	PROTEIN	IN	NUCB-AROD	INTERGENIC REGION
YQET_BACSU	45%	223	HYPOTHETICAL	10.8	KD	PROTEIN	IN	AROD-COMER	INTERGENIC REGION
YQEU_BACSU	35%	224	HYPOTHETICAL	22.2	KD	PROTEIN	IN	AROD-COMER	INTERGENIC REGION
YQEK_BACSU	37%	225	HYPOTHETICAL	21.3	KD	PROTEIN	IN	AROD-COMER	INTERGENIC REGION
YQEL_BACSU	46%	226	HYPOTHETICAL	13.3	KD	PROTEIN	IN	AROD-COMER	INTERGENIC REGION
YQEM_BACSU	42%	226	HYPOTHETICAL	28.3	KD	PROTEIN	IN	AROD-COMER	INTERGENIC REGION
YEBE_ECOLI	40%	229	HYPOTHETICAL	26.4	KD	PROTEIN	IN	RUVV-ASPS	INTERGENIC REGION
Y003_HAEIN	33%	239	HYPOTHETICAL			PROTEIN	H10003		
YBHS_ECOLI	22%	249	HYPOTHETICAL	42.1	KD	PROTEIN	IN	MOAE-RHLE	INTERGENIC REGION
YWBH_BACSU	50%	250	HYPOTHETICAL	44.4	KD	PROTEIN	IN	EPR-GALK	INTERGENIC REGION
YHIN_ECOLI	32%	254	HYPOTHETICAL	43.8	KD	PROTEIN	IN	RHSB-PTT	INTERGENIC REGION
YFOL_STRTR (ORF1091)	47%	255	HYPOTHETICAL	29.9	KD	PROTEIN	IN	FOLD-PBP2B	INTERGENIC REGION
YW05_MYCTU	49%	256	HYPOTHETICAL	18.2	KD	PROTEIN	CY39.05C		
YQHE_ECOLI	45%	264	HYPOTHETICAL			OXIDOREDUCTASE	IN	METC-SUFI	INTERGENIC REGION
Y181_MYCPN	35%	275	HYPOTHETICAL			PROTEIN	MG181	HOMOLOG (GT9_ORF434)	
YHGF_NEIME	47%	283	HYPOTHETICAL	83.1	KD	PROTEIN	IN	REGION E	
YKQC_BACSU	61%	287	HYPOTHETICAL	61.5	KD	PROTEIN	IN	ADEC-PDHA	INTERGENIC REGION
YEAB_BACSU (ORFX)	46%	305	HYPOTHETICAL	31.8	KD	PROTEIN	IN	GABP-GUAA	INTERGENIC REGION
YQHE_ECOLI	45%	330	HYPOTHETICAL			OXIDOREDUCTASE	IN	METC-SUFI	INTERGENIC REGION
YYBR_BACSU	49%	331	HYPOTHETICAL	14.7	KD	PROTEIN	IN	COTF-TETB	INTERGENIC REGION
YWKD_BACSU	50%	331	HYPOTHETICAL	14.8	KD	PROTEIN	IN	TDK-PRFA	INTERGENIC REGION
YWAC_BACSU	38%	364	HYPOTHETICAL	24.6	KD	PROTEIN	IN	DAE-TYRZ	INTERGENIC REGION
Y128_MYCGE	30%	364	HYPOTHETICAL			PROTEIN	MG128		
YHCT_BACSU	40%	365	HYPOTHETICAL	33.7	KD	PROTEIN	IN	CSPB-GLPP	INTERGENIC REGION
YIEF_ECOLI	33%	369	HYPOTHETICAL	20.4	KD	PROTEIN	IN	TNAB-BGLB	INTERGENIC REGION
YEIH_ECOLI	28%	376	HYPOTHETICAL	36.9	KD	PROTEIN	IN	LYSP-NFO	INTERGENIC REGION
Y372_MYCGE	32%	381	HYPOTHETICAL			PROTEIN	MG372		
YJCE_ECOLI (O549)	37%	386	HYPOTHETICAL	60.5	KD	PROTEIN	IN	SOXR-ACS	INTERGENIC REGION
YAAP_BACSU	44%	403	HYPOTHETICAL	31.2	KD	PROTEIN	IN	XPAC-ABRB	INTERGENIC REGION
Y552_HELPY	35%	404	HYPOTHETICAL			PROTEIN	HP0552		
YF56_SYNY3	37%	409	HYPOTHETICAL	37.5	KD	PROTEIN	SLL1556		
YWBA_BACSU	40%	419	HYPOTHETICAL	47.6	KD	PROTEIN	IN	EPR-GALK	INTERGENIC REGION
Y105_MYCPN	35%	431	HYPOTHETICAL			PROTEIN	MG105	HOMOLOG	
YW11_MYCTU	38%	438	HYPOTHETICAL	145.8	KD	PROTEIN	CY39.11C		
Y097_MYCTU	39%	442	HYPOTHETICAL	18.2	KD	PROTEIN	CY78.07		
LYTR_BACSU	31%	443				MEMBRANE-BOUND	PROTEIN	LYTR	
YQAK_BACSU	46%	451	HYPOTHETICAL	32.2	KD	PROTEIN	IN	SPOIIC-CWLA	INTERGENIC REGION
YMCP_BPF41	27%	471	HYPOTHETICAL			PROTEIN	IN	MCP 3' REGION (FRAGMENT)	
YQBO_BACSU	35%	473	HYPOTHETICAL	171.0	KD	PROTEIN	IN	SPOIIC-CWLA	INTERGENIC REGION
YWLG_BACSU	48%	484	HYPOTHETICAL	19.4	KD	PROTEIN	IN	SPOIIR-GLYC	INTERGENIC REGION
DEGV_BACSU	34%	496				DEGV	PROTEIN		
YFMR_BACSU	29%	497	HYPOTHETICAL	28.9	KD	PROTEIN	IN	ILVA 3'REGION	
YBBD_BACSU	25%	499	HYPOTHETICAL	70.6	KD	PROTEIN	IN	PEUA 5'REGION PRECURSOR (ORF1)	
YQGX_BACSU	44%	526	HYPOTHETICAL	23.2	KD	PROTEIN	IN	SODA-COMGA	INTERGENIC REGION
YPOA_STROR	73%	529	HYPOTHETICAL	23.1	KD	PROTEIN	IN	PONA 5'REGION	
YPSA_BACSU	40%	530	HYPOTHETICAL	21.1	KD	PROTEIN	IN	COTD-KDUD	INTERGENIC REGION
YQGZ_BACSU	42%	530	HYPOTHETICAL	14.8	KD	PROTEIN	IN	SODA-COMGA	INTERGENIC REGION
YNGE_CLOPE	25%	545	HYPOTHETICAL	31.2	KD	PROTEIN	IN	NAGH 5'REGION (ORPB)	
YJER_ECOLI	28%	551	HYPOTHETICAL	15.6	KD	PROTEIN	IN	PURA-VACB	INTERGENIC REGION
YGDP_ECOLI	33%	553	HYPOTHETICAL	20.8	KD	PROTEIN	IN	PTPH-MUTH	INTERGENIC REGION

Table 3.

YJAB_ECOLI (F147)	41%	569	HYPOTHETICAL	16.4	KD	PROTEIN	IN	RRPE-META	INTERGENIC REGION
Y369_MYCGE	28%	572	HYPOTHETICAL			PROTEIN		MG369	
DEGV_BACSU	30%	573	DEGV			PROTEIN			
YQJW_BACSU	35%	578	HYPOTHETICAL	45.9	KD	PROTEIN	IN	GLNQ-ANSR	INTERGENIC REGION
YF63_METJA	31%	581	HYPOTHETICAL			PROTEIN		MJ1563	
YWCG_BACSU	41%	585	HYPOTHETICAL	28.3	KD	PROTEIN	IN	QOXD-VPR	INTERGENIC REGION
YAT3_STAAU	28%	587	HYPOTHETICAL	16.6	KD	PROTEIN	IN	ATL 5'REGION	(ORF3)
YWLC_BACSU	33%	588	HYPOTHETICAL	37.0	KD	PROTEIN	IN	SPOIIR-GLYC	INTERGENIC REGION
Y553_METJA	40%	594	HYPOTHETICAL			PROTEIN		MJ0553	
YWCD_BACSU	33%	601	HYPOTHETICAL	14.3	KD	PROTEIN	IN	EPR-GALK	INTERGENIC REGION
Y085_MYCGE	33%	603	HYPOTHETICAL			PROTEIN		MG085	
YQFU_BACSU	29%	612	HYPOTHETICAL	32.5	KD	PROTEIN	IN	CCCA-SODA	INTERGENIC REGION
Y119_MYCTU	25%	615	HYPOTHETICAL	41.2	KD	PROTEIN		CY277.09	
YQJK_BACSU	51%	617	HYPOTHETICAL	34.0	KD	PROTEIN	IN	GLNQ-ANSR	INTERGENIC REGION
YCBG_ECOLI	30%	623	HYPOTHETICAL	38.2	KD	PROTEIN	IN	PABC-HOLB	INTERGENIC REGION
Y242_METJA	32%	642	HYPOTHETICAL			PROTEIN		MJECL42	
Y0A5_MYCTU	22%	651	HYPOTHETICAL	40.2	KD	PROTEIN		CY9C4.05C	
Y553_SYNY3	47%	658	HYPOTHETICAL	22.5	KD	PROTEIN		SLR0553	
YCEE_ECOLI (ORF1)	42%	659	HYPOTHETICAL	43.9	KD	PROTEIN	IN	MSYB-HTRB	INTERGENIC REGION
YRC2_LACLA	95%	660	HYPOTHETICAL			PROTEIN	IN	RPMG 3'REGION	(ORF2)
YQGZ_BACSU	38%	672	HYPOTHETICAL	14.8	KD	PROTEIN	IN	SODA-COMGA	INTERGENIC REGION
YCDQ_ECOLI	35%	678	HYPOTHETICAL	50.8	KD	PROTEIN	IN	PHOH-CSGG	INTERGENIC REGION
YADE_ECOLI PRECURSOR	28%	680	HYPOTHETICAL	46.3	KD	PROTEIN	IN	HPT-PAND	INTERGENIC REGION
YJDI_ECOLI (O90A)	35%	684	HYPOTHETICAL	10.5	KD	PROTEIN	IN	DCUB-LYSU	INTERGENIC REGION
YJDI_ECOLI (O76)	56%	684	HYPOTHETICAL	8.6	KD	PROTEIN	IN	DCUB-LYSU	INTERGENIC REGION
YABD_BACSU	43%	685	HYPOTHETICAL	29.2	KD	PROTEIN	IN	METS-KSGA	INTERGENIC REGION
YQHY_BACSU	39%	690	HYPOTHETICAL	14.7	KD	PROTEIN	IN	ACCC-FOLD	INTERGENIC REGION
YXJH_BACSU	43%	722	HYPOTHETICAL	38.3	KD	PROTEIN	IN	PEPT-KATB	INTERGENIC REGION
YICI_ECOLI	33%	727	HYPOTHETICAL	88.1	KD	PROTEIN	IN	GLTS-SELC	INTERGENIC REGION
YFBB_HAEIN	26%	734	HYPOTHETICAL			PROTEIN		HI0282	
Y188_METJA	24%	739	HYPOTHETICAL			PROTEIN		MJ0188	
Y371_MYCGE	28%	740	HYPOTHETICAL			PROTEIN		MG371	
YQSE_BACSU	40%	742	HYPOTHETICAL	28.1	KD	PROTEIN	IN	SIPU 3'REGION	
YYBT_BACSU	38%	750	HYPOTHETICAL	74.3	KD	PROTEIN	IN	RPLI-COTF	INTERGENIC REGION
Y347_MYCPN	26%	754	HYPOTHETICAL			PROTEIN		MG347	HOMOLOG
YEND_BACSU	36%	758	HYPOTHETICAL	27.9	KD	PROTEIN	IN	PHRC-GDH	INTERGENIC REGION
YHCS_BACSU	31%	768	HYPOTHETICAL	22.0	KD	PROTEIN	IN	CSPB-GLPP	INTERGENIC REGION
YWBG_BACSU	36%	770	HYPOTHETICAL	25.8	KD	PROTEIN	IN	EPR-GALK	INTERGENIC REGION
YWBH_BACSU	27%	770	HYPOTHETICAL	14.3	KD	PROTEIN	IN	EPR-GALK	INTERGENIC REGION
YLXS_BACSU	38%	771	HYPOTHETICAL	17.6	KD	PROTEIN	IN	NUSA 5'REGION	(P15A)
YLXR_BACSU (ORF3)	51%	773	HYPOTHETICAL	10.4	KD	PROTEIN	IN	NUSA-INF8	INTERGENIC REGION
YHI3_RHOCA	44%	779	HYPOTHETICAL	33.7	KD	PROTEIN	IN	HIMA 5'REGION	(ORF3)
Y06B_MYCTU	31%	809	HYPOTHETICAL	139.6	KD	PROTEIN		CY338.11C	PRECURSOR
YQGS_BACSU	34%	813	HYPOTHETICAL	73.2	KD	PROTEIN	IN	SODA-COMGA	INTERGENIC REGION
YIDA_ECOLI	38%	821	HYPOTHETICAL	29.7	KD	PROTEIN	IN	IBPA-GYRB	INTERGENIC REGION
Y143_HAEIN	25%	828	HYPOTHETICAL			PROTEIN		HI0143	
Y456_METJA	24%	848	HYPOTHETICAL			PROTEIN		MJ0456	
YCFB_ECOLI (ORF-15)	58%	850	HYPOTHETICAL	42.6	KD	PROTEIN	IN	PURB-ICDA	INTERGENIC REGION
YEH2_ECOLI PRECURSOR	31%	864	HYPOTHETICAL	32.6	KD	PROTEIN	IN	MOLR-BGLX	INTERGENIC REGION
YHAD_HAEIN	50%	877	HYPOTHETICAL			PROTEIN		HI0091	
YLXA_BACSU	64%	885	HYPOTHETICAL	35.3	KD	PROTEIN	IN	PTSL 5'REGION	(ORFB)
YWAC_BACSU	45%	894	HYPOTHETICAL	24.6	KD	PROTEIN	IN	DAB-TYRZ	INTERGENIC REGION
YCDQ_ECOLI	28%	897	HYPOTHETICAL	50.8	KD	PROTEIN	IN	PHOH-CSGG	INTERGENIC REGION
Y531_METJA	37%	900	HYPOTHETICAL			PROTEIN		MJ0531	
LYTR_BACSU	25%	907	MEMBRANE-BOUND			PROTEIN		LYTR	
YQSE_BACSU	27%	911	HYPOTHETICAL	28.1	KD	PROTEIN	IN	SIPU 3'REGION	
YABR_BACSU	31%	912	HYPOTHETICAL	14.2	KD	PROTEIN	IN	DIVIC-SPOIIE	INTERGENIC REGION
YICL_ECOLI	34%	923	HYPOTHETICAL	33.1	KD	PROTEIN	IN	SELC-NLPA	INTERGENIC REGION
Y912_HAEIN	35%	928	HYPOTHETICAL			PROTEIN		HI0912	
YIBD_ECOLI	33%	939	HYPOTHETICAL	40.5	KD	PROTEIN	IN	SECB-TDH	INTERGENIC REGION
YECD_ECOLI	40%	967	HYPOTHETICAL	21.7	KD	PROTEIN	IN	ASPS-BISZ	INTERGENIC REGION
YVCI_BACSU	52%	986	HYPOTHETICAL	33.9	KD	PROTEIN	IN	CRH-TRXB	INTERGENIC REGION
YVCK_BACSU	46%	987	HYPOTHETICAL	34.7	KD	PROTEIN	IN	CRH-TRXB	INTERGENIC REGION
Y247_MYCGE	36%	1003	HYPOTHETICAL			PROTEIN		MG247	
YECD_ECOLI	26%	1006	HYPOTHETICAL	21.7	KD	PROTEIN	IN	ASPS-BISZ	INTERGENIC REGION
YLYB_BACSU (ORF-X)	59%	1024	HYPOTHETICAL	33.7	KD	PROTEIN	IN	LSP-PYRR	INTERGENIC REGION
Y501_SYNY3	42%	1027	HYPOTHETICAL	36.7	KD	PROTEIN		SL0501	
YQAN_BACSU	41%	1043	HYPOTHETICAL	16.1	KD	PROTEIN	IN	SPOIIC-CWLA	INTERGENIC REGION

Table 3.

YMFN_ECOLI	23%	1051	HYPOTHETICAL	50.9	KD	PROTEIN	IN	INTE-PIN	INTERGENIC	REGION
YAF2_BACLI	40%	1068	HYPOTHETICAL	9.7	KD	PROTEIN	IN	CWLL	5'REGION	
YQGZ_BACSU	42%	1070	HYPOTHETICAL	14.8	KD	PROTEIN	IN	SODA-COMGA	INTERGENIC	REGION
YQFN_BACSU	44%	1082	HYPOTHETICAL	23.7	KD	PROTEIN	IN	CCCA-SODA	INTERGENIC	REGION
YQKB_BACSU	35%	1083	HYPOTHETICAL	11.8	KD	PROTEIN	IN	GLNQ-ANSR	INTERGENIC	REGION
YIDA_ECOLI	44%	1084	HYPOTHETICAL	29.7	KD	PROTEIN	IN	IBPA-GVVB	INTERGENIC	REGION
YWFO_BACSU	52%	1085	HYPOTHETICAL	51.0	KD	PROTEIN	IN	PTA	3'REGION	
YSXB_BACSU (ORF X)	38%	1088	HYPOTHETICAL	12.3	KD	PROTEIN	IN	RPLU-RPMA	INTERGENIC	REGION
YDAR_BACSU	40%	1090	HYPOTHETICAL	45.7	KD	PROTEIN	IN	MUTT-GSIB	INTERGENIC	REGION
YQFG_BACSU	46%	1092	HYPOTHETICAL	17.8	KD	PROTEIN	IN	PHOH-DGKA	INTERGENIC	REGION
YVYE_BACSU	54%	1096	HYPOTHETICAL			PROTEIN	IN	DEGS-TAGO	INTERGENIC	REGION (FRAGMENT)
YXIQ_BACSU	25%	1098	HYPOTHETICAL	45.5	KD	PROTEIN	IN	BGLS-KATB	INTERGENIC	REGION
YADQ_ECOLI	29%	1103	HYPOTHETICAL	46.0	KD	PROTEIN	IN	HEML-PFS	INTERGENIC	REGION
YQBJ_BACSU	46%	1104	HYPOTHETICAL	22.2	KD	PROTEIN	IN	AROD-COMER	INTERGENIC	REGION
Y03F_MYCTU	36%	1106	HYPOTHETICAL	54.3	KD	PROTEIN		CY130.15C		
YFBU_HAEIN	52%	1142	HYPOTHETICAL			PROTEIN		HI0754		
YFIA_BACSU	26%	1145	HYPOTHETICAL	29.3	KD	PROTEIN	IN	GLVG-GLVBC	INTERGENIC	REGION
YBDN_ECOLI	33%	1149	HYPOTHETICAL	47.8	KD	PROTEIN	IN	CSTA-DSBG	INTERGENIC	REGION
YBDM_ECOLI	46%	1150	HYPOTHETICAL	23.9	KD	PROTEIN	IN	CSTA-DSBG	INTERGENIC	REGION
YHCJ_HAEIN	36%	1189	HYPOTHETICAL			PROTEIN		HI0145		
YNGB_CLOPE	33%	1190	HYPOTHETICAL	31.2	KD	PROTEIN	IN	NAGH	5'REGION (ORFB)	
YPTB_CLOAB	51%	1194	HYPOTHETICAL			PROTEIN	IN	PTB	5'REGION (ORF1)	
Y912_HAEIN	28%	1195	HYPOTHETICAL			PROTEIN		HI0912		
YPIA_BACSU	27%	1197	HYPOTHETICAL	48.3	KD	PROTEIN	IN	QCRA	5'REGION	
YWKB_BACSU	26%	1213	HYPOTHETICAL	33.6	KD	PROTEIN	IN	TDK-PRFA	INTERGENIC	REGION
YH11_LACLA	92%	1225	HYPOTHETICAL			PROTEIN	IN	HISC	5'REGION (ORF1)	
YH16_LACLA (ORF6)	98%	1230	HYPOTHETICAL	30.7	KD	PROTEIN	IN	HISD-HISB	INTERGENIC	REGION
YH1C_LACLA	97%	1235	HYPOTHETICAL	31.3	KD	PROTEIN	IN	HISIE	3'REGION (ORF13)	
YABL_BACSU (ORF2)	53%	1250	HYPOTHETICAL	13.7	KD	PROTEIN	IN	PURR-SPOVG	INTERGENIC	REGION
YGT1_STRMU	65%	1255	HYPOTHETICAL			PROTEIN	1	IN	GTFB	REGION (FRAGMENT)
Y0D8_MYCTU	32%	1259	HYPOTHETICAL	69.9	KD	PROTEIN		CY1A11.08		
YQHY_BACSU	24%	1263	HYPOTHETICAL	14.7	KD	PROTEIN	IN	ACCC-FOLD	INTERGENIC	REGION
Y568_METJA	35%	1283	HYPOTHETICAL			PROTEIN		MJ0568		
YCDQ_ECOLI	25%	1294	HYPOTHETICAL	50.8	KD	PROTEIN	IN	PHOH-CSGG	INTERGENIC	REGION
YPA_A_BACSU	31%	1305	HYPOTHETICAL	20.5	KD	PROTEIN	IN	SERA-FER	INTERGENIC	REGION
YFUL_BACSU (ORFX13)	54%	1306	HYPOTHETICAL	26.0	KD	PROTEIN	IN	SPMB-AROC	INTERGENIC	REGION
YPUH_BACSU (ORFX8)	39%	1306	HYPOTHETICAL	22.0	KD	PROTEIN	IN	RIBT-DACB	INTERGENIC	REGION
YPUG_BACSU (ORFX7)	37%	1307	HYPOTHETICAL	29.6	KD	PROTEIN	IN	RIBT-DACB	INTERGENIC	REGION
YSNB_BACSU	39%	1309	HYPOTHETICAL	19.2	KD	PROTEIN	IN	RPH-ILVB	INTERGENIC	REGION
YGGV_ECOLI (O197)	48%	1309	HYPOTHETICAL	21.0	KD	PROTEIN	IN	GSHB-ANSB	INTERGENIC	REGION
YBIO_ECOLI	32%	1321	HYPOTHETICAL	86.8	KD	PROTEIN	IN	DING-GLNQ	INTERGENIC	REGION
Y442_MYCPN	36%	1327	HYPOTHETICAL			PROTEIN		MG442	HOMOLOG	
YQGZ_BACSU	37%	1356	HYPOTHETICAL	14.8	KD	PROTEIN	IN	SODA-COMGA	INTERGENIC	REGION
YEBH_ECOLI (ORF30)	34%	1362	HYPOTHETICAL	30.4	KD	PROTEIN	IN	MANZ-CSPC	INTERGENIC	REGION
Y120_MYCGE	30%	1365	HYPOTHETICAL			PROTEIN		MG120		
Y121_MYCGE	48%	1366	HYPOTHETICAL			PROTEIN		MG121		
YAA_N_BACSU	50%	1373	HYPOTHETICAL	43.8	KD	PROTEIN	IN	XPAC-ABRB	INTERGENIC	REGION
YLXD_BACCL (ORF2)	46%	1382	HYPOTHETICAL	27.6	KD	PROTEIN	IN	PYRAB-PYRD	INTERGENIC	REGION
YMFC_HAEIN	62%	1388	HYPOTHETICAL			PROTEIN		HI0694		
Y882_HAEIN	35%	1404	HYPOTHETICAL			PROTEIN		HI0882		
Y531_METJA	34%	1409	HYPOTHETICAL			PROTEIN		MJ0531		
YQBO_BACSU	27%	1424	HYPOTHETICAL	171.0	KD	PROTEIN	IN	SPOIIC-CWLA	INTERGENIC	REGION
YQBO_BACSU	29%	1426	HYPOTHETICAL	171.0	KD	PROTEIN	IN	SPOIIC-CWLA	INTERGENIC	REGION
YE22_HAEIN	39%	1451	HYPOTHETICAL			PROTEIN		HI1422		
YLXM_BACSU	53%	1455	HYPOTHETICAL	13.2	KD	PROTEIN	IN	FPH	5'REGION	
YBXB_BACSU (P23)	44%	1464	HYPOTHETICAL	22.5	KD	PROTEIN	IN	RPLL-RPOB	INTERGENIC	REGION
YBDM_ECOLI	47%	1467	HYPOTHETICAL	23.9	KD	PROTEIN	IN	CSTA-DSBG	INTERGENIC	REGION
YBDN_ECOLI	32%	1468	HYPOTHETICAL	47.8	KD	PROTEIN	IN	CSTA-DSBG	INTERGENIC	REGION
YTRP_LACLA	99%	1500	HYPOTHETICAL	13.3	KD	PROTEIN	IN	TRPE	5'REGION	
YICL_ECOLI	34%	1505	HYPOTHETICAL	33.1	KD	PROTEIN	IN	SELC-NLPA	INTERGENIC	REGION
Y709_METJA	26%	1511	HYPOTHETICAL			PROTEIN		MJ0709		
YBGG_ECOLI	29%	1530	HYPOTHETICAL	100.0	KD	PROTEIN	IN	HRSA-CYDA	INTERGENIC	REGION
YYAI_BACSU (ORFG)	33%	1539	HYPOTHETICAL	20.2	KD	PROTEIN	IN	TETB-EXOA	INTERGENIC	REGION
YJJG_ECOLI	27%	1554	HYPOTHETICAL	25.3	KD	PROTEIN	IN	RIMI-PRFC	INTERGENIC	REGION
YEXA_BACSU	37%	1573	HYPOTHETICAL	9.7	KD	PROTEIN	IN	PUR	OPERON	
YEGB_ECOLI	27%	1577	HYPOTHETICAL	50.9	KD	PROTEIN	IN	ALKA-BAES	INTERGENIC	REGION

Table 3.

YMO1_RHIME (ORF334)	30%	1591	HYPOTHETICAL 36.4 KD PROTEIN IN MOCC-MOCA INTERGENIC REGION
DEGV_BACSU	35%	1601	DEGV PROTEIN
YQKD_BACSU	25%	1612	HYPOTHETICAL 34.6 KD PROTEIN IN GLNQ-ANSR INTERGENIC REGION
YPBG_BACSU PRECURSOR	30%	1639	HYPOTHETICAL 28.6 KD PROTEIN IN RECQ-CMK INTERGENIC REGION
YH38_HAEIN	30%	1651	HYPOTHETICAL PROTEIN HI1738
YKQC_BACSU	41%	1659	HYPOTHETICAL 61.5 KD PROTEIN IN ADEC-PDHA INTERGENIC REGION
YHID_ECOLI (ORF-C)	42%	1660	HYPOTHETICAL 23.2 KD PROTEIN IN SLP-HDEB INTERGENIC REGION
YDP1_LACLC	50%	1664	HYPOTHETICAL 30.9 KD PROTEIN IN PEPX 5'REGION (ORF1)
YAGG_ECOLI	24%	1672	HYPOTHETICAL 50.6 KD PROTEIN IN PERR-ARGF INTERGENIC REGION
YPSC_BACSU PRECURSOR	46%	1697	HYPOTHETICAL 43.5 KD PROTEIN IN COTD-KDUD INTERGENIC REGION
YQEN_BACSU	32%	1699	HYPOTHETICAL 40.5 KD PROTEIN IN COMEC-RPST INTERGENIC REGION
YGJR_ECOLI	36%	1706	HYPOTHETICAL 37.0 KD PROTEIN IN EBGC-UKAA INTERGENIC REGION
YADQ_ECOLI	37%	1710	HYPOTHETICAL 46.0 KD PROTEIN IN HEML-PFS INTERGENIC REGION
YIDD_PROMI	46%	1719	HYPOTHETICAL 9.6 KD PROTEIN IN RNPA 3'REGION
YVDK_BACSU	60%	1726	HYPOTHETICAL 88.3 KD PROTEIN IN CLPP-CRH INTERGENIC REGION
LYTR_BACSU	31%	1739	MEMBRANE-BOUND PROTEIN LYTR
YFGA_ECOLI	23%	1741	HYPOTHETICAL 36.2 KD PROTEIN IN NDK-GCPE INTERGENIC REGION
YQFU_BACSU	28%	1745	HYPOTHETICAL 32.5 KD PROTEIN IN CCCA-SODA INTERGENIC REGION
YICO_ECOLI	39%	1768	HYPOTHETICAL 49.9 KD PROTEIN IN NLPA-UHPT INTERGENIC REGION
YUPP_MYCHO	32%	1770	HYPOTHETICAL 31.6 KD PROTEIN IN UPP 3'REGION
YKGE_ECOLI	43%	1771	HYPOTHETICAL 26.0 KD PROTEIN IN EAEH-BETA INTERGENIC REGION
YKGF_ECOLI	32%	1772	HYPOTHETICAL 53.1 KD PROTEIN IN EAEH-BETA INTERGENIC REGION
YKGG_ECOLI	31%	1773	HYPOTHETICAL 31.1 KD PROTEIN IN EAEH-BETA INTERGENIC REGION
YBHL_ECOLI	35%	1774	HYPOTHETICAL 25.9 KD PROTEIN IN MOAE-RHLE INTERGENIC REGION
YHAP_BACSU	27%	1814	HYPOTHETICAL 45.4 KD PROTEIN IN SSPB-PRSA INTERGENIC REGION
Y456_METJA	27%	1817	HYPOTHETICAL PROTEIN MJ0456
YNHE_ECOLI	39%	1826	HYPOTHETICAL 56.3 KD PROTEIN IN LPP-AROD INTERGENIC REGION
Y154_MYCLE	27%	1830	HYPOTHETICAL 42.2 KD PROTEIN U1496A/B1496_C1_154
YDQ_BACSU	52%	1859	HYPOTHETICAL 34.0 KD PROTEIN IN COTF-TETB INTERGENIC REGION
Y260_SYNY3	35%	1862	HYPOTHETICAL 49.2 KD PROTEIN SLL0260
YAT3_STAAU	34%	1864	HYPOTHETICAL 16.6 KD PROTEIN IN ATL 5'REGION (ORF3)
YHBQ_ECOLI (O100)	35%	1869	HYPOTHETICAL 11.3 KD PROTEIN IN SOHA-MTR INTERGENIC REGION
YRKA_BACSU	48%	1882	HYPOTHETICAL 49.0 KD PROTEIN IN BLTD-TRKA INTERGENIC REGION
Y531_METJA	26%	1907	HYPOTHETICAL PROTEIN MJ0531
Y556_SYNY3	37%	1917	HYPOTHETICAL 24.0 KD PROTEIN SLR0556
DEGV_BACSU	28%	1920	DEGV PROTEIN
YQHE_ECOLI	37%	1924	HYPOTHETICAL OXIDOREDUCTASE IN METC-SUFI INTERGENIC REGION
YPEC_LACLC	88%	1928	HYPOTHETICAL PROTEIN IN PEPX 5'REGION (ORF 2)
YQKG_BACSU	54%	1930	HYPOTHETICAL 21.0 KD PROTEIN IN GLNQ-ANSR INTERGENIC REGION
YQHY_BACSU	24%	1939	HYPOTHETICAL 14.7 KD PROTEIN IN ACCC-POLD INTERGENIC REGION
YQKG_BACSU	34%	1942	HYPOTHETICAL 21.0 KD PROTEIN IN GLNQ-ANSR INTERGENIC REGION
YTEK_ECOLI	40%	1949	HYPOTHETICAL 23.3 KD PROTEIN IN TNAB-BGLE INTERGENIC REGION
YLP3_PSEPU	40%	1951	HYPOTHETICAL 44.7 KD PROTEIN IN LPD-3 5'REGION (ORF3)
YADS_ECOLI	33%	1966	HYPOTHETICAL 22.1 KD PROTEIN IN HEML-PFS INTERGENIC REGION
YDAR_BACSU	30%	1970	HYPOTHETICAL 45.7 KD PROTEIN IN MUTT-GSIB INTERGENIC REGION
YB37_METJA	28%	1980	HYPOTHETICAL PROTEIN MJ0837
YIGN_ECOLI PRECURSOR	22%	1982	HYPOTHETICAL 54.7 KD PROTEIN IN UDP-UBIE INTERGENIC REGION
YJEQ_HAEIN	29%	1984	HYPOTHETICAL PROTEIN HI1714
YABE_BACSU	39%	1985	HYPOTHETICAL 47.7 KD PROTEIN IN METS-KSGA INTERGENIC REGION
YB12_HAEIN	31%	2002	HYPOTHETICAL PROTEIN HI1412
YF56_METJA	38%	2013	HYPOTHETICAL PROTEIN MJ1556
YACC_BACSU	45%	2014	HYPOTHETICAL 31.8 KD PROTEIN IN FTSH-CYSK INTERGENIC REGION
YWJB_BACSU	40%	2015	HYPOTHETICAL 19.6 KD PROTEIN IN ACDA 5'REGION
YACF_BACSU	57%	2016	HYPOTHETICAL 37.1 KD PROTEIN IN FOLK-LYSS INTERGENIC REGION
YPJC_BACSU	45%	2017	HYPOTHETICAL 23.6 KD PROTEIN IN QCRC-DAPB INTERGENIC REGION
YQFU_BACSU	29%	2019	HYPOTHETICAL 32.5 KD PROTEIN IN CCCA-SODA INTERGENIC REGION
YHIT_MYCPK	37%	2022	HYPOTHETICAL 16.1 KD PROTEIN HIT-LIKE PROTEIN
Y143_HAEIN	24%	2044	HYPOTHETICAL PROTEIN HI0143
YU12_MYCTU	30%	2045	HYPOTHETICAL 34.9 KD PROTEIN CY339.12C
YXJB_BACSU	40%	2052	HYPOTHETICAL 31.5 KD PROTEIN IN KATB 3'REGION
YQHL_BACSU	38%	2081	HYPOTHETICAL 14.6 KD PROTEIN IN GCVT-SPOIIIAB INTERGENIC REGION
YJFG_HAEIN	29%	2099	HYPOTHETICAL PROTEIN HI0121
Y018_MYCGE	36%	2101	HYPOTHETICAL HELICASE MG018
YDP1_LACLC	95%	2114	HYPOTHETICAL 30.9 KD PROTEIN IN PEPX 5'REGION (ORF1)
YDP3_LACLA	88%	2117	HYPOTHETICAL 18.7 KD PROTEIN IN PEPX 3'REGION (ORF3)
YAEL_BACSU	42%	2126	HYPOTHETICAL 40.9 KD PROTEIN IN MECB-GLTX INTERGENIC REGION
YEEO_ECOLI	28%	2151	HYPOTHETICAL 60.1 KD PROTEIN IN AMN-CBL INTERGENIC REGION
YCEC_ECOLI (ORFX)	38%	2161	HYPOTHETICAL 36.0 KD PROTEIN IN RNE-RPMF INTERGENIC REGION
YEBI_HAEIN	28%	2162	HYPOTHETICAL PROTEIN HI0407
YAEL_HAEIN	27%	2178	HYPOTHETICAL PROTEIN HI0918

Table 3.

Y506_SYNY3	46%	2180	HYPOTHETICAL 28.8 KD PROTEIN SLL0506
Y0BD_MYCTU	31%	2181	HYPOTHETICAL 13.0 KD PROTEIN CY227.13 PRECURSOR
YBEA_ECOLI	31%	2182	HYPOTHETICAL 17.3 KD PROTEIN IN MRDA-PHPB INTERGENIC REGION
YABB_BACSU	47%	2190	HYPOTHETICAL 28.3 KD PROTEIN IN XPAC-ABRB INTERGENIC REGION
Y003_HAEIN	28%	2194	HYPOTHETICAL PROTEIN HI0003
YBHE_ECOLI	28%	2209	HYPOTHETICAL 36.3 KD PROTEIN IN MODC-BIOA INTERGENIC REGION
YQJV_BACSU	25%	2221	HYPOTHETICAL 44.7 KD PROTEIN IN GLNQ-ANSR INTERGENIC REGION
Y661_METUA	31%	2226	HYPOTHETICAL PROTEIN MJ0661
YTFG_ECOLI (F286)	45%	2227	HYPOTHETICAL 29.7 KD PROTEIN IN RPLI-CPDB INTERGENIC REGION
YYBR_BACSU	42%	2227	HYPOTHETICAL 14.7 KD PROTEIN IN COTF-TETB INTERGENIC REGION
YJBR_ECOLI	40%	2232	HYPOTHETICAL 13.5 KD PROTEIN IN APHA-UVRA INTERGENIC REGION
Y143_HAEIN	30%	2252	HYPOTHETICAL PROTEIN HI0143
YPBG_BACSU PRECURSOR	26%	2253	HYPOTHETICAL 28.6 KD PROTEIN IN RECO-CMK INTERGENIC REGION
YCEE_ECOLI (ORF1)	25%	2277	HYPOTHETICAL 43.9 KD PROTEIN IN MSYB-HTRB INTERGENIC REGION
Y441_METUA	38%	2280	HYPOTHETICAL PROTEIN MJ0441
YQEG_BACSU	42%	2283	HYPOTHETICAL 20.1 KD PROTEIN IN NUCB-AR0D INTERGENIC REGION
YWFA_BACSU	28%	2299	HYPOTHETICAL 44.5 KD PROTEIN IN ROCC-PTA INTERGENIC REGION
Y531_METUA	32%	2300	HYPOTHETICAL PROTEIN MJ0531
Y800_METUA	28%	2304	HYPOTHETICAL PROTEIN MJ0800
YJBC_ECOLI	30%	2306	HYPOTHETICAL 32.5 KD PROTEIN IN PEPE-LYSC INTERGENIC REGION
Y0A5_MYCTU	24%	2328	HYPOTHETICAL 40.2 KD PROTEIN CY9C4.05C
YHA2_EIKCO	29%	2333	HYPOTHETICAL 66.3 KD PROTEIN IN HAG2 5'REGION
YCKE_BACME	41%	2337	HYPOTHETICAL 30.5 KD PROTEIN IN GDHI 5'REGION (ORF 2)
YHCC_ECOLI	37%	2339	HYPOTHETICAL 34.6 KD PROTEIN IN ARCB-GLTB INTERGENIC REGION
BIO16. UNKNOWN			83 genes
BIO16.01 General			5 genes
NIFS_LACDE	53%	520	NIFS PROTEIN HOMOLOG (FRAGMENT)
NIFS_ANASP	29%	1829	NIFS PROTEIN
NIFS_ANASP	39%	1908	NIFS PROTEIN
US45_LACLA	39%	1956	SECRETED 45 KD PROTEIN PRECURSOR
US45_LACLA	99%	2293	SECRETED 45 KD PROTEIN PRECURSOR
BIO16.02 Unclassified			78 genes
LCIA_LACLA	33%	48	LACTOCOCCIN A IMMUNITY PROTEIN
OSMC_ECOLI	48%	69	OSMOTICALLY INDUCIBLE PROTEIN C
LCND_LACLA	97%	95	LACTOCOCCIN A SECRETION PROTEIN LCND
CSBB_BACSU	45%	107	CSBB PROTEIN
ACPD_ECOLI	33%	113	ACYL CARRIER PROTEIN PHOSPHODIESTERASE (ACP PHOSPHODIESTERASE)
JAG_BACSU	31%	131	JAG PROTEIN (SPOI1J ASSOCIATED PROTEIN)
AMD_MORCA	48%	164	PUTATIVE AMIDASE (EC 3.5.1.4)
YZDD_BACSU	62%	166	PET112-LIKE PROTEIN
HYIN_AGRVI	42%	168	INDOLEACETAMIDE HYDROLASE (EC 3.5.1.-)
PHNO_ECOLI	40%	225	PHNO PROTEIN
TPX_MYCTU	35%	299	PROBABLE THIOL PEROXIDASE (EC 1.11.1.-)
ACPD_ECOLI	39%	384	ACYL CARRIER PROTEIN PHOSPHODIESTERASE (ACP PHOSPHODIESTERASE)
KIME_METUA	31%	406	MEVALONATE KINASE (EC 2.7.1.36)
KIME_METUA	29%	408	MEVALONATE KINASE (EC 2.7.1.36)
BMRU_BACSU	27%	424	BMRU PROTEIN
MRSA_HAEIN	46%	433	MRSA PROTEIN HOMOLOG
COLL_HSVS7	37%	478	COLLAGEN-LIKE PROTEIN
FAS6_RHOPA	30%	491	HYPOTHETICAL PROTEIN IN AZU 5'REGION (ORF1)
EPUA_STRPN	40%	533	EPUA PROTEIN
XYLH_PSEPU	32%	566	4-OXALOCROTONATE TAUTOMERASE (EC 5.3.2.-)
STA_ECOLI	31%	580	STREPTOTHRICIN ACETYLTTRANSFERASE (EC 2.3.-.-)
HPRK_ENTFA	43%	602	HPR-KINASE (EC 2.7.1.-)
NUSB_BACSU	46%	691	N UTILIZATION SUBSTANCE PROTEIN B HOMOLOG (NUSB PROTEIN)
CLCB_PSEPU	26%	730	CHLOROMUCONATE CYCLOISOMERASE (EC 5.5.1.7)
VEG_BACSU	30%	753	VEG PROTEIN
YORV_CLOAB	52%	755	HYPOTHETICAL PROTEIN HI0943
NUSA_BACSU	48%	772	N UTILIZATION SUBSTANCE PROTEIN A HOMOLOG (NUSA PROTEIN)
LRPR_STRERQ	31%	825	LEUCINE RICH PROTEIN
EXPZ_BACSU	29%	918	NUCLEOTIDE BINDING PROTEIN EXPZ
NRDI_LACLC	96%	1002	NRDI PROTEIN
ELAA_ECOLI	38%	1008	ELAA PROTEIN
RGG_STRGC	25%	1075	RGG PROTEIN
YRP2_LISMO	34%	1083	HYPOTHETICAL PROTEIN MJ0927
PHOL_BACSU	57%	1091	PHOH-LIKE PROTEIN
PHNB_ECOLI	33%	1135	PHNB PROTEIN
RGG_STRGC	29%	1178	RGG PROTEIN
CITG_HAEIN	35%	1209	CITG PROTEIN
Y002_RICFR (ORF39.9)	33%	1236	HYPOTHETICAL 40.0 KD PROTEIN IN HTRB-DINI INTERGENIC REGION

Table 3.

GID_BACSU	69%	1255	GID PROTEIN (FRAGMENT)
2DKG_CORSP	46%	1266	2,5-DIKETO-D-GLUCONIC ACID REDUCTASE (EC 1.1.1.-)
FIBP_ADEB3	29%	1393	FIBER PROTEIN
FMCH_BACNO	31%	1408	POSSIBLE FIMBRIAL ASSEMBLY PROTEIN FIMC (SEROGROUP H1)
P60_LISSE	31%	1414	PROTEIN P60 PRECURSOR (INVASION-ASSOCIATED PROTEIN)
XHLB_BACSU	40%	1414	HOLIN
DFRA_SYNY3	30%	1558	PUTATIVE DIHYDROFLAVONOL-4-REDUCTASE (EC 1.1.1.219)
DC4C_ACICA	34%	1575	4-CARBOXYMUCONOLACTONE DECARBOXYLASE (EC 4.1.1.44)
ESL2_MYCPN	27%	1604	PUTATIVE ESTERASE/LIPASE 2 (EC 3.1.-.-)
PKSG_BACSU	33%	1610	PUTATIVE POLYKETIDE BIOSYNTHESIS PROTEIN PKSG
F4RE_METJA	34%	1652	PUTATIVE F420-DEPENDENT NADP REDUCTASE (EC 1.-.-.-)
LAXC_LACLA	29%	1671	LACK PROTEIN, CHROMOSOMAL
RGG_STRGC	27%	1713	RGG PROTEIN
ARCL_ECOLI	40%	1732	CARBAMATE KINASE-LIKE PROTEIN 1
SMPB_ENTPA	63%	1757	SMALL PROTEIN B HOMOLOG
BMRU_BACSU	28%	1775	BMRU PROTEIN
DEDA_ECOLI	39%	1796	DEDA PROTEIN (DSG-1 PROTEIN)
XYNC_CALSA	33%	1809	ACETYL ESTERASE (EC 3.1.-.-)
THL_CLOAB	33%	1823	ACETYL-COA ACETYLTRANSFERASE (EC 2.3.1.9)
PHPB_ECOLI	34%	1869	PHPB PROTEIN
PFS_HAEIN	43%	1929	PFS PROTEIN HOMOLOG
PROC_ECOLI	38%	1933	PYRROLINE-5-CARBOXYLATE REDUCTASE (EC 1.5.1.2)
SUN_BACSU	33%	1938	SUN PROTEIN (FRAGMENT)
YFP1_MYCMY	44%	1947	HYPOTHETICAL PROTEIN MGL30
LAXC_LACLA	36%	1979	LACK PROTEIN, CHROMOSOMAL
BLTD_BACSU	33%	1996	TRANSACETYLASE BLTD (EC 2.3.1.-)
RGG_STRGC	19%	2024	RGG PROTEIN
FIB1_ADE41	30%	2027	FIBER PROTEIN 1
DC4C_ACICA	35%	2032	4-CARBOXYMUCONOLACTONE DECARBOXYLASE (EC 4.1.1.44)
HIT_BACSU	41%	2057	HIT PROTEIN
ARCL_ECOLI	45%	2090	CARBAMATE KINASE-LIKE PROTEIN 1
STA_ECOLI	31%	2123	STREPTOTHRICIN ACETYLTRANSFERASE (EC 2.3.-.-)
NARA_BACSU	27%	2213	NARA PROTEIN
PQQE_ACICA	30%	2216	COENZYME PQQ SYNTHESIS PROTEIN E (COENZYME PQQ SYNTHESIS PROTEIN III)
RGG_STRGC	31%	2217	RGG PROTEIN
YHHF_ECOLI	30%	2219	21.7 KD PROTEIN IN FTSY-NIKA INTERGENIC REGION
GSPA_ECOLI	28%	2239	PUTATIVE GENERAL SECRETION PATHWAY PROTEIN B
RGG_STRGC	19%	2268	RGG PROTEIN
TOXB_CLODI	27%	2285	TOXIN B
PHNA_ECOLI	47%	2312	PHNA PROTEIN

1495 genes

754

Table 4. Ability of *L. lactis* for growth on different carbon sources

Carbon source	Catabolic genes (position on genome map)	Transport genes (position on genome map)	Experimental growth (reference)
glucose	glycolysis genes		+(Bardowski et al. 1994) +(de Vos 1996) [#]
fructose	fructokinase (1519 kb), glucosamine-fructoso-6P aminotransferase (2230 kb)	PTS (984 kb and 1747 kb)	+(Bardowski et al. 1994) +(de Vos 1996) [#]
N-acetyl glucosamine (1032 kb)			+(Bardowski et al. 1994) +(de Vos 1996) [#]
sucrose		PTS (435 kb) [*]	-(de Vos 1996) [#]
mannose	mannose-6P isomerase (777 kb)	PTS (1748 kb)	+(Bardowski et al. 1994) +(de Vos 1996) [#]
maltose	maltodextrin glucosidase (700 kb and 1526 kb), 4- α -glucanotransferase (692 kb)		+(Bardowski et al. 1994) +(de Vos 1996) [#]
trehalose		PTS (435 kb) [*]	+(Bardowski et al. 1994) +(de Vos 1996) [#]
lactose	β -galactosidase (2040 kb)	lactose permease (2048 kb)	+(de Vos 1996) [#]
galactose	gal operon (2045 kb)		+(Bardowski et al. 1994) +(de Vos 1996) [#]
ribose	ribs operon (1685 kb)	ribose transport protein (1686 kb)	+(Bardowski et al. 1994) -(de Vos 1996) [#]
xylose	xyl operon (1550 kb)		
cellobiose	glucosidase (186 kb and 1521 kb), 6-P β -glucosidase (419 kb, 830 kb and 1490 kb)	PTS (175 kb, 416 kb and 830 kb)	+(Bardowski et al. 1994) +(de Vos 1996) [#]
salicin	6-P β -glucosidase (1722 kb)		+(Bardowski et al. 1994) +(de Vos 1996) [#]
arbutin	See cellobiose	See cellobiose	+(Bardowski et al. 1994)
esculin	See cellobiose	See cellobiose	+(Bardowski et al. 1994)
gluconate	6-P-gluconate gehydrogenase (606 kb), gluconate kinase (2253 kb)	gluconate permease (2246 kb)	
mannitol	mannitol-1P 5-dehydrogenase (33 kb)	PTS (29 kb and 32 kb)	

[#]For the strain mg1363 with the plasmid pMG820.

^{*}Presumably a trehalose specific system.

755

756