Genotyping and virulence factors assessment of bovine mastitis

*Escherichia coli*

Shlomo Blum*, Gabriel Leitner

National Mastitis Center, Department of Bacteriology, Kimron Veterinary Institute, Ministry of Agriculture and Rural Development, P.O. Box 12, Bet Dagan 50250, Israel.

*Corresponding author. POB 12, Bet Dagan, 50250, Israel. Tel.: +972-3-9681680.*

Email: shlomobl@moag.gov.il.
Abstract

*Escherichia coli* is a major agent of bovine mastitis worldwide. However, specific *E. coli* virulence factors associated to pathogenicity during intra-mammary infections are yet unknown and this pathotype remains uncharacterized. The objectives of the present work were to assess the presence of a wide range of known virulence factors in a large set of *E. coli* strains isolated from bovine mastitis (mastitis set) and to study their genotypic distribution in comparison to a set of strains isolated from cows’ environment in dairy farms (environmental set). Virulence factors were assessed by DNA hybridization microarray. The three most prevalent virulence factors found in the mastitis set were *lpfA* (long polar fimbriae), *iss* (increased serum resistance) and *astA* (enteroaggregative *E. coli* heat-stable enterotoxin 1). None, however, characterized the majority of these strains. Genotyping was assessed by phylogenetic grouping, multilocus sequence typing (MLST) and pulsed-field gel electrophoresis (PFGE). Strains in the mastitis and environmental sets were differentially distributed into phylogenetic groups; groups A and B1 being the most prevalent ones. Multiple MLST strain types were found in the two sets of strains, but only a few were common to both, and diversity was higher in the environmental set. A variety of PFGE patterns were found in the mastitis and environmental sets. Two clusters comprising mostly highly similar mastitis strains were identified. The results confirm that mastitis *E. coli* strains lack most of known *E. coli* virulence factors. In addition, it is shown that the genotypic diversity of mastitis strains does not reflect the diversity found in the environmental *E. coli* population.
Key words: *Escherichia coli*, bovine, mastitis, virulence, genotyping, multilocus sequence type.

**Introduction**

*Escherichia coli* is a major agent of bovine mastitis worldwide (Bradley and Green, 2000; Longo et al., 2001; Shpigel et al., 1998). The *E. coli* species encompasses pathogenic and non-pathogenic strains. Pathogenic strains are associated with diverse intestinal and extra-intestinal infections in human and animals. Based on shared pathogenic mechanisms, pathogenic *E. coli* strains are classified into different pathotypes, which are disease- and host-specific and tend to be clonal (Kaper et al., 2004).

Various research studies attempted to characterize the *E. coli* strains causing bovine mastitis. Although these strains mostly lack known virulence genes (Bean et al., 2004; Kaipainen et al., 2002; Lehtolainen et al., 2003; Suojala et al., 2011; Wenz et al., 2006), there is evidence that the genetic and phenotypic variability of bovine mastitis strains does not completely reflect those of the *E. coli* strains population found in cows’ environment (Blum et al., 2008; Bradley et al., 2001), suggesting that *E. coli* may have partially adapted to the intra-mammary milieu (Blum et al., 2008; Bradley and Green, 2000; Dogan et al., 2006). Different pathogenic potentials are found among bovine mastitis *E. coli* strains. While bovine mastitis caused by *E. coli* is mostly acute, specific strains are prone to cause persistent infections (Zadoks et al., 2011). Strains associated with persistent infections invade mammary epithelial cells at higher rates and cause milder immune responses compared to strains causing acute infections (Almeida et al., 2011; Dogan et al., 2006; Kerro Dego et al., 2011).

Genotyping comparison between bovine mastitis strains and strains present in cows' environment could lead to the identification of clonal groups differently present
among mastitis pathogenic *E. coli*, which could be further studied in search of
pathotype-specific markers of virulence. The objectives of the present work were to
compare a set of *E. coli* strains isolated from bovine mastitis to one isolated from
cows' environment genotypically and for the presence of virulence factors.

**Material and methods**

**Strains collection**

Two sets of strains were collected for comparison: bovine mastitis strains and
environmental ones. Bovine mastitis strains were isolated from milk samples
aseptically collected from mammary quarters of cows with mastitis. Strains from the
environment were isolated from bacteriological swabs used to sample the manure in
different places of cowsheds. Both sets were isolated from two dairy farms.
Environmental strains were collected at the same periods of time of mastitis sampling.
Strains were isolated and identified to species level using standard bacteriological
techniques, including colony morphology on McConkey agar, oxidase, catalase,
urease and indole tests. Bacteria were stocked frozen in BHI with 30% glycerol at -80°C. Mastitis strains were selected for this study when *E. coli* was isolated in at least
two consecutive samples from the same quarter and only from cultures with rich *E.
coli* growth and with no other bacteria growth, as described before (Blum et al., 2008).
Strains isolated from a third farm used in a previous work were added to the present
study (Blum et al., 2008).

**DNA extraction**

Bacteria were suspended in sterile water and lysed by boiling for 10 min
followed by freezing at -18°C. Lysates were kept frozen until use.

**Phylogenetic grouping**
The method described by Clermont (Clermont et al., 2000) was used to classify strains into ECOR phylogenetic groups, namely A, B1, B2 and D. This method is based on a multiplex PCR for detection of three genetic markers, which are then used to assign the phylogenetic group based on a flowchart. PCR reactions were set in 20 µl including 10 µl of 2× Qiagen Multiplex PCR Master Mix (Qiagen), 0.4 pmoles of each primer (Metabion, Martinsried, Germany), 3 µl of bacterial DNA and PCR-grade water to complete the reaction volume (Biological Industries, Bet Haemek, Israel). PCR amplification was carried out by 15 min activation at 95°C, 30 cycles of 5 sec denaturation at 95°C, 10 sec annealing at 59°C and 30 sec extension at 72°C, followed by a 5 min final extension at 72°C. *E. coli* strain ATCC 25922 reacts positively to the three genetic markers and was used as positive control. Amplicons were separated by electrophoresis in 2% agarose gels and visualized by ethidium bromide staining. Phylogenetic groups were assigned to a total of 98 mastitis and 55 environmental isolates.

*Pulsed-field gel electrophoresis (PFGE)*

A standard PFGE technique was performed as described before (Wolk et al., 2004). Bacteria were lysed in agarose cubes and genomic DNA was digested with XbaI. Cluster analysis of restriction patterns was done by UPGMA in Bionumerics 6.6 (Applied Maths). DICE similarity coefficient was calculated with a 2% tolerance and 2% optimization sets. A total of 46 mastitis and 25 environmental isolates were analysed by PFGE. These isolates were randomly selected from the isolates assigned to phylogenetic groups as described above, including multiple isolates from cases suspected as persistent infections.

*Multilocus sequence typing (MLST)*
MLST was performed using partial sequences of seven house-keeping genes as described before (Wirth et al., 2006). PCR reactions were set in 25 µl including 5 µl of 5x GoTaq Flexi buffer, 1.5 mM of MgCl₂, 0.2 mM of each dNTP, 1.25 units of GoTaq Flexi DNA polymerase (Promega, Madison, WI, US), 10 pmol of each primer (Metabion, Martinsried, Germany), 3 µl of bacterial DNA and PCR-grade water up to reaction volume (Biological Industries, Bet Haemek, Israel). PCR amplification was carried out by 2 min initial denaturation at 95°C, 30 cycles of 1 min denaturation at 95°C, 1 min annealing at primer-specific annealing temperature and 2 min extension at 72°C, followed by a 5 min final extension at 72°C. Positive amplification was tested by gel electrophoresis. Amplicons were purified using Exonuclease Shrimp Alkaline Phosphatase (Exo-Sap IT, HDV Pharmacia). Sanger sequencing was performed for both sense and antisense strands, using the same primer pairs as for PCR amplification, at the Center of Genomic Technologies, The Hebrew University of Jerusalem, Israel. Most of the isolates analysed by PFGE were selected for MLST analysis but a few ones with identical PFGE profiles. Extra isolates among those assigned phylogenetic groups were included in the MLST analysis for a total of 62 mastitis and 31 environmental isolates evaluated by MLST. *E. coli* strain ATCC 25922 was used as a reference strain.

Sequence Types (ST) were assigned online (http://mlst.ucc.ie/mlst/dbs/Ecoli).

A minimum spanning tree (MST) of the results was generated in PHILOVIZ 1.0 (Francisco et al., 2012) using the goeBURST algorithm (Francisco et al., 2009). For phylogenetic tree construction, the concatenated sequences of the seven house-keeping genes analysed (including MLST allelic sequences and flanking regions) were aligned by MUSCLE in MEGA 5.0 (Tamura et al., 2011). A phylogenetic tree
was built using the Maximum Parsimony method and was analysed by 1,000 replicates in the bootstrap test.

**Virulence genotyping**

The presence of known *E. coli* virulence genes was assessed using a commercial DNA hybridization microarray by the manufacturer (Clondiag GmbH, Jena, Germany) as described in details before (Geue et al., 2010). The presence of 78 virulence factors was evaluated. A complete list of genes and alleles may be found at the company's website (http://identibac.com/fileadmin/Media/Downloads/Genelist__E.coli_03_m.pdf) and is provided in the Supplemental material. This array was used also for O and H genotyping. All mastitis isolates analysed by MLST but two were examined by microarray. Twenty four environmental isolates from those analysed previously by MLST were examined. Environmental isolated were selected such that diverse PFGE types and MLST STs were represented.

**Statistical analysis**

Chi-square tests were used to compare the mastitis and the environmental sets of isolates for the features assessed. Isolates which were identified as being of the same strain based on isolation from the same mammary quarter and being of the same phylogenetic group, PFGE or MLST type were included only once in the statistical analysis in order to prevent bias.

**Results**

**Phylogenetic grouping**

Most of isolates in both sets were assigned to groups A or B1 (Fig. 1). However, the distribution of phylogenetic groups was significantly different between
In the environmental set, 71% of isolates were assigned to group B1 and only 16% were assigned to group A. In the mastitis set, in contrast, 38% of strains were assigned to group A and 51% to group B.

Multilocus sequence typing

Isolates were assigned Sequence Types (STs) according to the allelic profile of seven house-keeping genes. Twenty three different STs found were unique to the mastitis set, 15 were unique to the environmental set and only six STs were common to both sets. Five novel STs were found in this study; ST2970 in the mastitis set and ST2517, ST2971, ST2972 and ST3013 in the environmental set. Fig. 2 shows the MST generated with the results. ST10 was the most prevalent ST in the mastitis set (27%), followed by ST58 (11%). ST10 was significantly more prevalent in the mastitis set then in the environmental set ($P < 0.05$); 80% of ST10 isolates were mastitis ones. In the environmental set, ST10 and ST58 were assigned to 12% and 16% of the strains, respectively. The phylogenetic tree inferred from the concatenated sequences of the seven house-keeping genes analysed is shown in Fig. 3. There were a total of 4,741 positions in the final concatenated sequences, including MLST allelic and flanking regions.

Pulsed-field gel electrophoresis

The PFGE patterns of XbaI DNA digestion and UPGMA cluster analysis are shown in Fig 4. Twelve mastitis isolates obtained from the same mammary quarters in four different cows were found to have identical restriction pattern (data not shown). Thus, only one isolate from each of these groups of identical isolates was included in the cluster analysis. These were VL2767 (two identical isolates, seven days apart), AF7504 (four identical isolates, during 33 days), VL2729 (two isolates, 52 days apart) and VL2732 (four identical isolates, during 186 days). The last three groups of
isolates were clearly persistent infections (signed with asterisks in Fig. 4). Mastitis isolates showed a wide variety of PFGE patterns. A clear separation between mastitis and environmental strains into separate clusters was not observed. However, two clusters of highly similar strains (above 83% similarity) were identified (outlined I and II in Fig. 4) consisting mostly of mastitis isolates. Some level of geographic variability was observed between these clusters: VL and AF indicate different farms; cluster I comprises mostly AF mastitis isolates, whereas cluster II comprises mostly VL mastitis ones.

**Virulence genotyping**

The presence of a total of 78 *E. coli* virulence genes was assessed using a DNA hybridization microarray. The array included virulence factors of intestinal and extra-intestinal *E. coli* pathotypes. The number of positive genes per isolate was low in both sets (Fig. 5) and no patterns of virulence genes were observed in any of the two sets of isolates. The most prevalent virulence genes observed in the mastitis set were *lpfA*, *astA* and *iss* (Table 1). The only gene more significantly prevalent in the mastitis set compared to the environmental set was *astA*. From the 33 *lpfA* positive mastitis isolates, nine were also positive for *astA*, nine for *iss* and only two were positive for the three genes combined. In addition to virulence factors, the array also assessed O and H antigen genotypes. Most of isolates could not be assigned an O-type by the array probes. However, all but one isolate could be assigned an H-type. Multiple H-types were found in both sets. The most prevalent one in the mastitis set was H19 (8 isolates, 15%), which was found in only one isolate from the environment. The eight H19 mastitis isolates were *lpfA* and *astA* positive, whereas the environmental H19 isolate was *lpfA* positive but *astA* negative. The full results of virulence genotyping are available in the Supplemental material.
No apparent correlation was found between virulence genes, number of virulence genes or H-type and ST. However, lpfA was significantly more prevalent in B1 phylogenetic group isolates in the mastitis and the environmental sets (100% and 90% of group B1 isolates were lpfA positive in the mastitis and the environmental sets, respectively) than in group A isolates (10% and 30% of group A isolates were lpfA positive in the mastitis and the environmental sets, respectively) \((P < 0.05)\). No such correlation was found regarding astA and iss.

A full list of strains analysed by each assay and results obtained is provided in the Supplemental material.

**Discussion**

Though there is little doubt that host-related factors play an important role in the pathogenesis of \(E.\ coli\) mastitis (Burvenich et al., 2003), it has become accepted that particular \(E.\ coli\) strains are associated with specific clinical outcomes of bovine intra-mammary infections. In light of the fact that there is apparently not a single model strain for research representing the whole pathogenic potentials that might be found in \(E.\ coli\) mastitis strains, the characterization of a mastitis pathotype should follow two steps. First, assessment of genotypes distribution of \(E.\ coli\) strains causing bovine mastitis and second, the search for shared traits, preferably pathogenic ones, in the most prevalent lineages. In the present study, a collection of \(E.\ coli\) strains isolated from bovine mastitis was compared to a set of strains isolated from cows' environment in dairy farms using three different genotyping methods: phylogenetic grouping, MLST and PFGE. In addition, a wide range of known \(E.\ coli\) virulence genes were assessed using a DNA hybridization microarray.
In spite of the large number of genes assessed, results show that *E. coli* mastitis isolates mostly lack known genes of virulence. These results are in accordance to previous works, although fewer genes have been assessed thus far. In the present study, only three virulence genes were found to be considerably prevalent in mastitis strains, but none actually characterized the majority of isolates in the mastitis set. Gene *lpfA* was first identified in *Salmonella Typhimurium* (Baumler et al., 1996), where it promotes adhesion to host cells, and later in different diarrheagenic *E. coli* pathotypes and non-pathogenic ECOR strains (Toma et al., 2006). Dogan et al. reported a similar prevalence of *lpfA* in *E. coli* mastitis isolates as found here (Dogan et al., 2012). It is suggested that *lpfA* might improve virulence in the mammary gland by mediating epithelial adhesion (Dogan et al., 2012). Gene *iss* is found in 75% to 91% of strains of extra-intestinal *E. coli* pathotypes, notably APEC, NMEC and UPEC and in 50% of bovine necrotoxigenic *E. coli* (Johnson et al., 2008). Serum resistance is the only trait repeatedly reported as a common finding in *E. coli* mastitis isolates. Nevertheless, published results vary widely from 16.7% up to 99.5% of mastitis strains being serum-resistant (Kaipainen et al., 2002; Nemeth et al., 1994; Sanchez-Carlo et al., 1984; Suojala et al., 2011). A high percentage of faecal isolates are serum-resistant as well (Nemeth et al., 1994). In a previous study, serum resistance characterized environmental, but not mastitis *E. coli* isolates (Blum et al., 2008). Several different methods have been used to assess this trait phenotypically and different genes have been assessed molecularly (*TraT* and *iss*, for instance). Here, similar rates of *iss* were found in the mastitis and the environmental sets. Serum resistance is found in other mastitis pathogen species as well, thus it may probably be an unspecific factor that improves, but is not necessary for virulence in the mammary gland (Rainard, 2003). Gene *astA* was the only gene found at a significantly higher
rate in the mastitis set in comparison to the environmental one, although it too does not characterize the majority of mastitis isolates. astA encodes the enteroaggregative E. coli heat-stable enterotoxin 1 (EAST1), found in various human E. coli diarrheagenic strains and other human enteric pathogens, such as Salmonella (Ménard et al., 2004), and in bovine E. coli isolates (Veilleux and Dubreuil, 2005). The exact role of this enterotoxin in pathogenicity is still uncertain (Ménard et al., 2004). By sequence homology between EAST1 and the enterotoxigenic domain of the STa enterotoxin and similar modes of action, it is suggested that the receptor for EAST1 and STa is the same (Veilleux and Dubreuil, 2005). Expression of this receptor is limited to the intestinal epithelium (Veilleux and Dubreuil, 2005), making the virulence significance of astA in the context of the mammary gland, if any at all, even more intriguing.

Phylogenetic groups A and B1 are commonly found in bovine faeces. Group B1 is also widely present in the environment (Higgins et al., 2007). The present results show that groups A and B1 distribution in the mastitis set does not exactly represent that of the environmental set. This may be partially explained by the fact that the most prevalent MLST strain time in the mastitis set, ST10, comprises only group A strains. Phylogenetic group A was assigned to 82.6% (Suojala et al., 2011) and 45% (Ghanarpour and Oswald, 2010) of E. coli mastitis strains in previous studies. The discrepancy between studies may result from different sampling techniques and study strain selection stringency. Phylogenetic group B2 is mostly associated to extra-intestinal infections and invasive strains (Clermont et al., 2000). Although intra-mammary infections are anatomically extra-intestinal, mastitis strains are not highly invasive, possibly explaining the low rates of B2 strains identified.
To the best of our knowledge, this is the first report of the use of MLST in the study of bovine mastitis *E. coli*. MLST-analysed loci are under low, but continuous selective pressure, being thus suitable to phylogenetic study of bacteria core genomes.

There are three *E. coli* MLST schemes publicly available. The Achtman scheme (Wirth et al., 2006) was chosen because it comprised the most varied strains. Due to the possibility of lineage independent acquisition of virulence factors, MLST classification may not completely correlate to pathotype classification (Wirth et al., 2006), explaining the wide variety of STs found in the mastitis set. Interestingly, however, few common STs were found between environmental and mastitis sets, thus mastitis causing strains may be possibly under-represented in the environment. ST10 was the most common ST found, comprising mostly mastitis isolates. By the time this article was written, ST10 was the largest one in the MSLT database (164 isolates), including mainly human strains (130) and only 13 bovine ones.

A clear distinction between mastitis and environmental strains by PFGE was not observed, unlike our previous study (Blum et al., 2008). Factors behind these differences are the addition of a large number of isolates from additional farms and the use of different restriction enzymes (XbaI in the present work, SpeI in the previous one). Two clusters (I and II) were delineated, each including strains from distinct farms, illustrating how farm source may have influenced the results. Also, inconsistent similarity coefficients were found between strains tested in both studies (strains from Blum et al., 2008 are designated here by "K"), showing how the use of different enzymes influenced the results. The large variety of PFGE genotypes found in *E. coli* mastitis isolates was shown by others before (Srinivasan et al., 2007). But differences between mastitis and environmental isolates were also identified (Bradley et al., 2001; Blum et al., 2008). It is interesting to note that strain P4 was assigned to
an out-group cluster consisting mostly environmental isolates, with a very low similarity degree to other mastitis strains.

In conclusion, this study confirmed that E. coli mastitis strains generally lack known markers of virulence. These strains are partially clonal, but not completely genotypically uniform. Yet, their genotypic distribution is not identical to that of the E. coli population found in the environment in dairy farms, suggesting that some level of selection takes place. An interesting set of mastitis strains was identified, sharing a common genomic backbone (ST10, phylogenetic group A, PFGE cluster II). In spite of sharing similar genotypes, this set of strains originated from distinct mastitis clinical presentations, which could be reproduced in a murine model of intra-mammary challenge (data not shown). Whole genome comparison between these strains is on-going in order to reveal virulence genes that may be associated with bovine mastitis. As part of this effort, the whole genome sequence of strain P4 was recently published by our group (Blum et al., 2012).

**Competing interests**

The authors declare no competing interests.

**Aknowledgments**

This work was partially supported by research grants from the Israeli Dairy Board. We acknowledge Dr. Assaf Rokney from the E. coli reference laboratory in the Israel Ministry of Health, who performed PFGE analysis, and Dr. Gila Kahila Bar-Gal from the Koret School of Veterinary Medicine in the Hebrew University of Jerusalem, for assistance in MLST analysis.
References


**Figure legends**

Figure 1. Distribution of phylogenetic groups in mastitis and environmental sets.

Mastitis (n = 90) and environmental (n = 55) isolates were assigned to phylogenetic groups A, B1, B2 and D. The distribution of phylogenetic groups in the two sets of isolates was statistically different (*P* < 0.005).

Figure 2. Minimum spanning tree of MLST results generated by goeBURST. Node sizes are proportional to the number of isolates found for that ST. The proportion of mastitis (n = 62) and environmental (n = 31) isolates for a given ST is shown in grey and white, respectively. Group former nodes are circumvented by a black line. Nodes differing by a single locus are linked by complete lines; nodes differing by two loci are linked by dashed lines. ATCC 25922 reference strain is shown in dark grey (ST73).

Figure 3. Phylogenetic tree of MLST sequences. The unrooted tree was inferred by the Maximum Parsimony method using the concatenated sequences of the allelic and flanking regions of seven house-keeping genes (4,741 positions). Tree #1 out of 167
is shown. Bootstrap values are shown beside the branches (1,000 replicates; values
under 50 were omitted) and STs are indicated on the left. Mastitis and environmental
isolates are marked by black and grey squares, respectively. The ATCC 25922
reference strain is marked by a white square.

Figure 4. UPGMA cluster analysis of PFGE DNA restriction patterns. Phylogenetic
group (PG), MLST Strain Type (ST), and Source (E, environment; M, mastitis) are
indicated along Isolates identification. Two clusters bearing similar mastitis strains
are outlined (I, II). The position of the model strain P4 is marked with an arrow.
Persistent strains are marked with an asterisk.

Figure 5. Number of positive genes per isolate in mastitis and environmental sets. The
graphs shows the frequency of the number of positive genes found per isolate in the
mastitis (n = 63, black) and environmental (n = 24, grey) sets.

Table 1. Prevalence of the three main virulence markers found in the mastitis and
environmental sets.

Supplemental material (Additional file 1)
Additional file 1.xlsx
Excel workbook (xlsx)
IDENTIBAC E. coli genotyping array
Full results of IDENTIBAC genotyping array for *E. coli* isolates in the mastitis and environmental sets. Positive probes are marked in blue. A 0.3 threshold was used. Antigen O and H genotyping are shown below.

Full list of strains examined with each assay and results.

Figure 1.
Figure 2.
Figure 5.