



Molecular analysis of an Integrative Conjugative Element, ICEH, present in the chromosome of different strains of *Mycoplasma hyopneumoniae*

Paulo Marcos Pinto^{1*}, Marcos Oliveira de Carvalho^{1*}, Leonardo Alves-Junior¹, Marcelo Brocchi³ and Irene Silveira Schrank^{1,2}

¹Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

²Departamento de Biologia Molecular e Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

³Departamento de Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas, São Paulo, Brazil.

Abstract

Diversification of bacterial species and pathotypes is largely caused by lateral gene transfer (LGT) of diverse mobile DNA elements such as plasmids, phages, transposons and genomic islands. Thus, acquisition of new phenotypes by LGT is very important for bacterial evolution and relationship with hosts. This paper reports a 23 kb region containing fourteen CDSs with similarity to the previous described Integrative Conjugative Element of *Mycoplasma fermentans* (ICEF). This element, named ICEH, is present as one copy at distinct integration sites in the chromosome of 7448 and 232 pathogenic strains and is absent in the type strain J (non-pathogenic). Notable differences in the nucleotide composition of the insertion sites were detected, and could be correlated to a lack of specificity of the ICEH integrase. Although present in strains of the same organism, the ICEH elements are more divergent than the typical similarity between other chromosomal locus of *Mycoplasma hyopneumoniae*, suggesting an accelerated evolution of these constins or an ongoing process of degeneration, while maintaining conservation of the *tra* genes. An extrachromosomal form of this element had been detected in the 7448 strain, suggesting a possible involvement in its mobilization and transference of CDSs to new hosts.

Key words: LGT, *Mycoplasma*, constins.

Received: April 12, 2006; Accepted: November 6, 2006.

Introduction

In recent years, lateral gene transfer (LGT), the intra-species and interspecies exchange of genetic information, has been substantially correlated to changes in genome content and with the evolution of bacteria, disseminating key traits within and among bacterial species (Boucher *et al.*, 2003; Lerat *et al.*, 2005). With the increase in genome sequence information associated with microarray technology, it is now possible to trace the presence or absence of genes in closely related bacterial genomes that could determine a sudden adaptation of bacteria to new environmental conditions. (Calcutt *et al.*, 2002; Porwollik and McClelland, 2003; Ochman *et al.*, 2005).

A number of phenotypes has been associated with this type of dissemination during conjugation of chromo-

somal regions: the sucrose metabolism in *Salmonella enterica* Seftenberg via *cTnscr94* (Hochhut *et al.*, 1997); the symbiosis traits associated with the 500 kb symbiosis island in *Mesorhizobium loti* (Sullivan and Ronson, 1998); the degradation of chlorocatechol mediated by gene products resident on the 105 kb *clc* element of *Pseudomonas putida* (Ravatt *et al.*, 1998); the tetracycline resistant element Tn916 of *Bacteroides* (Salyers *et al.*, 1995); the large STX unit of *Vibrio cholerae* (Hochhut and Waldor, 1999); and the antibiotic resistance gene cluster from *Salmonella* (Doublet *et al.*, 2005). The term constin has been recently coined to describe this diverse group of conjugative, self-transmissible, integrating elements (Hochhut *et al.*, 2001). Despite the existence of a diversity of LGT mechanisms, only evolutionarily relevant events of transfer persist within the bacteria genome (Jain *et al.*, 2003).

Mycoplasma hyopneumoniae is recognized as a potent pathogen causing mycoplasmal pneumonia in swine after colonising the respiratory tract and by inducing an inflammatory response (Maes *et al.*, 1996). This chronic

Send correspondence to Marcos Oliveira de Carvalho. Centro de Biotecnologia, Departamento de Biologia Molecular e Biotecnologia, Av. Bento Gonçalves 9500, Prédio 43421, 91591-970 Porto Alegre, RS, Brazil. E-mail: operon@cbiot.ufrgs.br.

*These authors contributed equally to this work.

worldwide disease causes important economic losses, including those in Brazil (Sobestiansky *et al.*, 2001). With the completion of genome sequence of 10 different species of Mycoplasmas, the search for horizontally acquired genes revealed the presence of constins only in *M. fermentans* (Calcutt *et al.*, 2002) and *M. hyopneumoniae* (Vasconcelos *et al.*, 2005). During a survey of specific sequences, a region containing fourteen CDS (\approx 23 kb) with similarity to the previous described Integrative Conjugal Element of *M. fermentans* (ICEF) (Calcutt *et al.*, 2002) was found in the Brazilian field isolate (7448) strain (Vasconcelos *et al.*, 2005). Recently, subtractive hybridization analysis indicated the presence of coding sequences related to ICEF in *Mycoplasma bovis* and *M. agalactiae* genomes (Marenda *et al.*, 2005).

Scrutiny of the complete genome from three strains (7448, J and 232) of *M. hyopneumoniae* (Vasconcelos *et al.*, 2005; Minion *et al.*, 2004) revealed the presence of a constin element in only two of them. This genetic element is present as one copy at distinct integration sites in the chromosome of 7448 and 232 strains and is absent in the type strain J. An extrachromosomal form of this element can be detected in the 7448 strain by the inverse PCR method (Vasconcelos *et al.*, 2005).

Methods

Sequence analysis

Sequence data from *Mycoplasma hyopneumoniae* strains analyzed in this paper were obtained from GenBank accession no. NC_006360 for strain 232, NC_007295 for strain J and NC_007332 for strain 7448.

Similarity searches were performed using a locally installed NCBI BLAST 2.2.12 (Altschul *et al.*, 1997) against custom built databases of known constins and the *nr* database (downloaded in 05/09/2005) from NCBI. Initial analysis of sequence families was done using the Tribe-MCL software (Enright *et al.*, 2002) and the Bidirectional Best Hits (BBH) (Tatusov *et al.*, 2000) approach was used to find putative orthologues among the analyzed constins. Sequence annotation was manually curated with the help of similarity searches against the PFAM (Sonnhammer *et al.*, 1997), PRODOM (Bru *et al.*, 2005) and INTERPRO (Mulder *et al.*, 2005) databases. Genometrics indexes were obtained with the EMBOSS suite of programs (Rice *et al.*, 2000). Phylogenetic analyses were performed with MEGA 3 (Kumar *et al.*, 2004), and multiple sequence alignments were generated with CLUSTALX (Thompson *et al.*, 1997). JTT distance was employed for both the neighbor-joining and parsimony tree reconstruction analysis, and a value of 1000 replications was used for bootstrap analysis. The software TOPALi (Milne *et al.*, 2004) was used for the recombination analysis of the *traE* genes, and PDM and HMM analysis were done with default settings.

Bacterial strains and culture conditions

M. hyopneumoniae strain J (ATCC 25934), a non-pathogenic strain with reduced capacity to adhere to porcine cilia (Zielinski and Ross 1990; Zielinski and Ross, 1993; Zhang *et al.*, 1995) was acquired from American Type Culture Collection by CNPSA, EMBRAPA (Concórdia, SC, Brazil). *M. hyopneumoniae* strain 7448 was isolated from an infected swine in Lindóia do Sul, Santa Catarina, Brazil, as described by Vasconcelos *et al.* (2005).

PCR analysis and DNA sequencing

PCR amplification of the target sequences was performed in a DNA thermal Cycler (MJ Research). The PCR mixture contained the reaction buffer (50 mM KCl, 10 mM Tris-Cl pH 8.3, 1.5 mM MgCl₂), 200 mM of each dNTPs, 20 pmol of each primer, and 1U of *Taq* polymerase (CenBiot enzymes, Centro de Biotecnologia, UFRGS). Assays were carried out using 50 ng of genomic DNA from both *M. hyopneumoniae* strains (J and 7448) and distilled water for a total of 25 μ L. The reaction mixture was subjected to PCR under the following conditions: heat denaturation at 94 °C for 5 min. and then an additional 30 cycles with heat denaturation at 94 °C for 1 min., annealing at 60.6 °C for 1 min., and DNA extension at 68 °C for 1 min. After the last cycle, samples were maintained at annealing temperature for 5 min. followed by 68 °C for 8 min. PCR products are resolved in a 1.2% gel electrophoresis, stained with ethidium bromide and visualized by UV illumination. PCR amplification was performed with primers ICEH1-F (5'TTTTTCATTGCCTAAATCTTGTTT3') and ICEH2-R (5'AAATCACAACTTAAAAATGCCAAT3') (Invitrogen, USA). Before the sequencing reaction, the amplification products (300 ng) were purified using GFX kit (Amersham Biosciences, GE Healthcare) or treated with 3.3U of exonuclease III and 0.2U shrimp alkaline phosphatase (SAP) (Amersham Biosciences, GE Healthcare) at 37 °C during 30 min. After this period, the enzymes were inactivated by incubation at 80 °C during 15 min. Amplification products were sequenced using the DYEnamic ET dye terminator cycle sequencing (MegaBACE) kit and run on MegaBACE 1000 capillary sequencers (Amersham Biosciences, GE Healthcare). The generated sequences were assembled and a quality score of 20 was considered for the final assembly. The derived sequence was searched using the BLASTn software using the complete *M. hyopneumoniae* strain 7448 genome as the database to identify the element in the chromosome.

Results and Discussion

General characteristics

Although conjugative elements have been described for members of many bacterial genera, indigenous self-transmissible molecules have been reported only in *M. fermentans* so far in Mycoplasma. During an ongoing anal-

ysis of genes encoding surface components of *M. fermentans*, genetic elements designated ICEF were found (Calcutt *et al.*, 2002). Two kinds of elements that differ regarding the presence or absence of a few CDSs were characterized by those authors and called Type I and II. The CDSs present in these elements suggested that they belong to the constin family. The sequence comparisons of three *M. hyopneumoniae* genomes carried out recently (Vasconcelos *et al.*, 2005) also reveal the presence of putative ICE elements in two of these strains that were called ICEH. Interestingly, these elements were present in the two pathogenic strains (7448 and 232) but were absent from the non-pathogenic one (J strain).

Unlike ICEF elements I and II, which consist of twenty one and nineteen CDSs, respectively, the ICEH7448 elements consist of 17 CDSs and ICEH232 of 22 CDSs. However, despite these differences, the organization of these elements is very similar.

The *M. hyopneumoniae* elements ICEH7448 and ICEH232 are 22,816 and 21,061 nucleotides in length, respectively. The average length of the ICEH CDSs is 1,218 bp in ICEH7448 and 1,187 bp in ICEH232. Most of the CDSs (14 ICEMH7448 and 12 in ICEMH232) are considered hypothetical since no significant similarity was found with the available databases. Conversely, some CDSs present similarity to *tra* genes that are associated with the conjugative plasmids of bacteria (Snyder and Champness, 2002) such as *traK*, *traI*, *traE* and a CDS encoding for a single strand binding protein (Ssb) that is essential for the transfer process. The ICEF element presents two *tra* (*traG* and *traE*) genes, with an average similarity of 30% to *tra* genes found on ICEH232 (*traG* and *traE*) and 29% to *tra* gene of ICEH7448 (only *traE*). The ICEH232 element has three *tra* genes, with one *traG* and two copies of the *traE* gene. In the *M. hyopneumoniae* 7448 ICEH element, we found only one *traE* gene with an average similarity of 44% to the *traE* genes found in the ICEH232. A characteristic that is shared by both ICEH and ICEF type elements is the lack of a recognizable gene encoding for an integrase function. Calcutt *et al.* (2002) suggested that the integrase function is possibly mediated by the hypothetical CDSs present in both type I and II ICEF elements. This observation may also be true for the ICEH elements because they contain hypothetical CDSs as well (Table I). The AT content of some CDSs in the ICEH element is higher than the overall percentage for the *M. hyopneumoniae* genome. For instance, CDSs ICE_MH7448-ORF09 and ICE_MH7448-ORF017 exhibit 76.13% and 76.06% AT content respectively, suggesting a heterologous origin for these ORFs in the ICEH element.

A complete reannotation analysis of the elements ICEH7448, ICEH232 and ICEF was also conducted. From these analyses, we have attributed functions to seven previously not annotated CDSs, including a CDS in ICEH7448 that was not recognized initially and was identified as a pu-

tative helicase. Table 1 shows the CDSs annotated in this way and they are marked with asterisks. Note the identification of a transposase-like CDS in the element ICEH232, but not in ICEF and ICEH7448. However, other ICEs have already been described which contain CDSs related to transposases, like the ICE element from *Streptococcus thermophilus* (Pavlovic *et al.* 2004). The role of transposases residing inside mobile genomic islands is not clear. However it is possible that such elements could help the transfer process of the ICEs acting as triggers for the activation of the damage-repair machinery found in the host cells. It has already been suggested that the damage-repair machinery is important for the integration of the ICEs in the chromosome of the host cells (Auchtung *et al.*, 2005).

The ICEH7448 element is inserted at the position 517,895 of the genome of *M. hyopneumoniae* 7448, unlike ICEH232 which is found at position 654,116 of the *M. hyopneumoniae* 232 genome from the replication origin. This indicates that these elements may insert themselves in different genomic regions. We were unable to determine with accuracy the presence of direct or inverted repeats flanking both elements. However, ICEH7448 is flanked by a duplication of part of the YP_287818.1 CDS, which is located downstream from the element. Nevertheless, considering that the duplicated gene is not exactly at the terminus of the element, we cannot conclude that this duplication is result of the ICEH7448 insertion.

The difference in number of CDSs present in distinct ICE elements suggests that they are able to acquire or lose CDSs over time. For instance, the CDSs AAN85219.1-AAN85220.1 and AAN85229.1, although present in the ICEF-I element are not present in the ICEF-II element, possibly indicating a very recent acquisition of new sequences by ICEF-I. Thus, it is possible that these elements may work as carriers of CDSs that, when mobilized to a new host, could confer new properties.

Genometric analyses have been conducted on the genomes of *M. hyopneumoniae* 7448 and 232 strains. Interestingly, the ICEH7448 element was found in an island of high AT content, relative to the genome, while the ICEH232 element was located in a high GC island content. The observed deviations on the GC content of the insertion sites indicates that DNA structural features derived from compositional constraints may not be a major determinant for the element insertion site determination. Individually, the CDSs of the ICEH232 element display a slightly higher GC content. The element positions and the graph of GC% cumulative skew are depicted in Figure 1. The variability in base composition is not limited to the element as a whole when compared to the genome. Analysis of the genometric properties of the ICE elements revealed that specific regions contain deviations of sequence properties. The *traE* gene sequence is peculiar in the sense that, in all elements, this gene presents a higher AT content than the other CDSs of the element. This internal variation in base composition

Table 1 - Annotated CDS of the elements ICEH7448, ICEH232 and ICEF.

NCBI Ids	CDS	Annotation	NCBI Ids	CDS	Annotation
AAN85213.1	ICEF_ORF-01	Hypothetical	YP_116033.1	ICE_MH232_4-ORF08	hypothetical
AAN85214.1	ICEF_ORF-02	putative relaxase	YP_116034.1	ICE_MH232_4-ORF09	hypothetical
AAN85215.1	ICEF_ORF-02_2	Hypothetical	YP_116035.1	ICE_MH232_4-ORF10	TraG *
-	ICEF_ORF-03	putative TraG (contains a frameshift)	YP_116036.1	ICE_MH232_4-ORF11	hypothetical
AAN85216.1	ICEF_ORF-04	Hypothetical	YP_116037.1	ICE_MH232_4-ORF12	hypothetical
AAN85217.1	ICEF_ORF-05	putative methyltransferase *	YP_116038.1	ICE_MH232_4-ORF13	hypothetical
AAN85218.1	ICEF_ORF-06	hypothetical	YP_116039.1	ICE_MH232_4-ORF14	hypothetical
AAN85219.1	ICEF_ORF-07	hypothetical	YP_116040.1	ICE_MH232_4-ORF15	TraE/TrsE family NTPase*
AAN85220.1	ICEF_ORF-08	hypothetical	YP_116041.1	ICE_MH232_4-ORF16	TraE/TrsE family NTPase
AAN85221.1	ICEF_ORF-08_2	hypothetical	YP_116042.1	ICE_MH232_4-ORF17	hypothetical
AAN85222.1	ICEF_ORF-09	SSB protein	YP_116043.1	ICE_MH232_4-ORF18	hypothetical
AAN85223.1	ICEF_ORF-10	hypothetical	YP_116044.1	ICE_MH232_4-ORF19	hypothetical
AAN85224.1	ICEF_ORF-11	P57 lipoprotein	YP_116045.1	ICE_MH232_4-ORF20	hypothetical
AAN85225.1	ICEF_ORF-12	putative phosphatidate transferase *	YP_116046.1	ICE_MH232_4-ORF21	hypothetical
AAN85226.1	ICEF_ORF-13	hypothetical	YP_116047.1	ICE_MH232_4-ORF22	putative transposase *
AAN85227.1	ICEF_ORF-14	TraE/TrsE family NTPase	MHP0414	ICE_MH7448-ORF02	hypothetical
AAN85228.1	ICEF_ORF-15	hypothetical	MHP0415	ICE_MH7448-ORF03	hypothetical
-	ICEF_ORF-16	hypothetical (contains a frameshift)	MHP0416	ICE_MH7448-ORF04	hypothetical
-	ICEF-ORF16A	hypothetical (contains a frameshift)	MHP0417	ICE_MH7448-ORF05	hypothetical
AAN85229.1	ICEF_ORF-17	hypothetical	MHP0418	ICE_MH7448-ORF06	TraE/TrsE family NTPase
AAN85230.1	ICEF_ORF-18	hypothetical	MHP0419	ICE_MH7448-ORF07	hypothetical
AAN85231.1	ICEF_ORF-19	hypothetical	MHP0420	ICE_MH7448-ORF08	hypothetical
YP_115644.1	ICE_MH232_1-ORF01	putative DNA processing protein *	MHP0421	ICE_MH7448-ORF09	hypothetical
YP_115645.1	ICE_MH232_1-ORF02	hypothetical	MHP0422	ICE_MH7448-ORF10	SSB protein
YP_115646.1	ICE_MH232_1-ORF03	hypothetical	MHP0423	ICE_MH7448-ORF11	hypothetical
YP_115647.1	ICE_MH232_1-ORF04	TraE/TrsE family NTPase	MHP0424	ICE_MH7448-ORF12	hypothetical
YP_115648.1	ICE_MH232_1-ORF05	hypothetical	MHP0425	ICE_MH7448-ORF13	hypothetical
YP_116031.1	ICE_MH232_4-ORF06	hypothetical	MHP0426	ICE_MH7448-ORF14	hypothetical
YP_116032.1	ICE_MH232_4-ORF07	SSB protein *	-	ICE_MH7448-ORF15	putative helicase *
			MHP0427	ICE_MH7448-ORF16	hypothetical
			MHP0428	ICE_MH7448-ORF17	hypothetical
			-	ICE_MH7448-ORF18	hypothetical

could explain a high expression for the *traE* genes. High codon usage bias has already been linked to a high expression profile in genes with distinctive base composition (Wang *et al.*, 2005). This is important as the *traE* genes are fundamental in the processes of element transfer, and an indication that the selection for optimum codon usage is a process that can exist in constins in a marked way.

ICEH elements

It was not possible to determine a single value for similarity among the studied elements, as the sequences are heterogeneous regarding their local similarities scores. The similarity rates range from regions with none (even be-

tween ICEH232 and ICEH7448) to regions with 67% of similarity (the *traE* genes for instance). Figures 2 and 3 demonstrate the similarity between the ICEH232 and ICEH7448 elements. Figure 2 exhibits a nucleic acid word match, while Figure 3 is a TBLASTX comparison based on the 3-frame translated amino acid sequence of all CDSs in the elements. The notable divergence between the ICEH elements, present in different strains of the same organism, could suggest an ongoing process of degeneration or accelerated evolution. These questions can be resolved when more sequences of ICEH constins become available.

While searching for CDSs related to ICEs in the chromosomes of *M. hypopneumoniae*, strains J, 7448 and 232, a

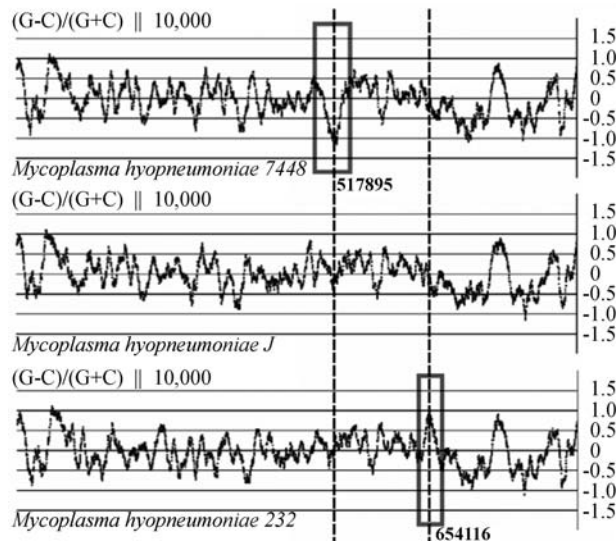


Figure 1 - GC% Skew graph showing the differential regions between the strains 7448, J and 232 of *Mycoplasma hyopneumoniae*. The dotted lines indicate the position of the ICEH7448 and ICEH232, respectively.

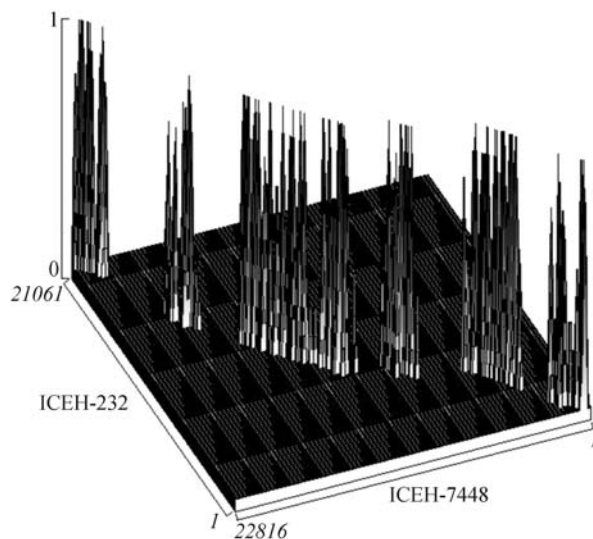


Figure 2 - Word density 3D plot of the ICEH-232 and ICEH-7448 sequences. The peaks correspond to the word density match (number of nucleotide local sequence similarities in a region) between the sequences of the elements, where more elevated peaks denote higher similarity. Location of the peaks in the base plan is determined by the coordinates of the word matches between the two sequences, represented in the X-Z axis.

specific region of about 10 kb that has significant similarity to the genes *traE* and two other hypothetical proteins have been found. We named these regions ICEH-like MHJ, ICEH-like 7448 and ICEH-like 232. Figure 3 A depicts the ICEH-232 region organization. It is interesting to note that this ICE-like region does not only have sequence similarity, but the gene order is also conserved, indicating a common origin with the ICEH elements. Additionally these regions are flanked by inverted repeats of 45 nucleotides in length and show a conservation of 71% between them. This ele-

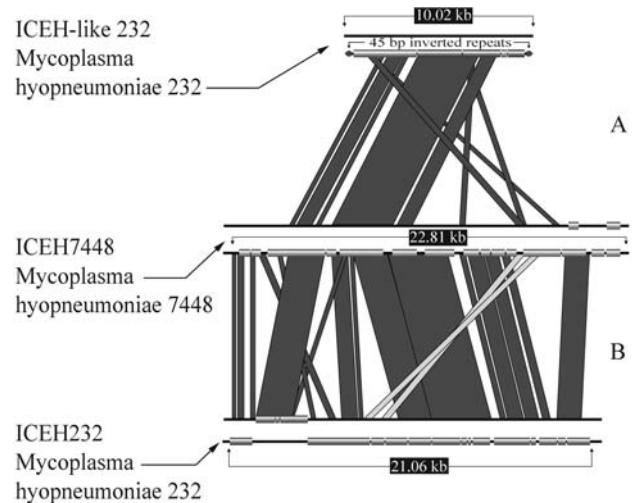


Figure 3 - TBLASTX (translated search of nucleotide sequences) similarity of the elements ICEH-like MH232, ICEH7448 and ICEH232. Dark gray bars denote direct sequence similarity between elements; light gray bars connect reverse sequence similarity between elements. CDS are represented by boxes. Inverted repeats of the ICEH-like elements are represented by arrows. The CDS represented are listed in Table 1 in order of appearance in the figure, except for the ICEH232 element that have the matches flipped for better visualization and the CDS are in inverted order as represented in the table. A: Comparison between the ICEH232-like element and ICEH232 element. B: Comparison between the ICEH232 element and ICEH7448 element.

ment is present in the three *M. hyopneumoniae* strains sequenced to date with a highly conserved gene content, gene order and structure, except for the ICEH-like MH232 which has an additional CDS located upstream from the SMF DNA processing protein without any similarity to known genes. It is possible that these ICEH-like elements could be the product of an ancestral ICE integration event that suffered an irregular excision leaving parts of them inserted in the chromosome. However, phylogeny analyses of the *traE* genes suggest that these regions are maintained at approximately the same evolutionary rate as the other genes in the chromosome, with lower divergence between the *traE* genes found in the ICEH-like region or in the ICEH232 and ICEH7448 elements (Figure 4 B). These might be an indication that the ICEH-like elements have acquired new functions during evolution, which differ from those assigned to the ICEH elements that assume an independent behavior in the host.

Comparisons between mycoplasma ICE elements

Two clustering experiments have been also conducted, using the CDSs available from six elements (ICEH-like MHJ, ICEH7448, ICEH232, pSKU146, pBJS-O and ICES1) sequenced so far, one with the BBH strategy and another employing the MCL algorithm. Both methods found a small number of putative orthologous genes among the analysed sequences, and the *traE* gene was the most conserved of all. It is interesting to note that although present in the same species, the two ICEH elements are less

conserved in relation to each other than the pSKU146 and pBJS-O elements. These two latter elements, although present in different species, are more conserved and syntenic. This could be explained by the fact that the ICEH elements diverged before the pSKU146 and pBJS-O elements, as evidenced by the phylogenetic analysis of 3 conserved orthologous CDSs of these elements (Figure 4 A). Although the phylogenetic tree of the elements is congruent with their host species, it disagrees for the *traE* gene, the most conserved putative orthologous gene. This result might indicate a process of CDS exchange by horizontal gene transfer between these elements or that this gene is under more stringent selective pressures. These analyses also in-

dicating that the *traE* gene of ICE clearly diverged before the same genes found in the more recent ICEs such as pSKU146 and pBJS-O.

Curiously, a duplication of the *traE* genes in the MH232 ICE element has been identified. From the phylogenetic analysis (Figure 4 B), it is possible to assume that the CDS15 originated by duplication of CDS16, which is the putative ortholog of the unique *traE* gene from ICEH7448. This observation is corroborated by the best bidirectional hit observed among these genes. Apparently, CDS15 is under a process of degeneration as it is the more distinct *traE* gene among the other homologous genes found in the ICEH elements. An analysis of recombination done on the most conserved *traE* genes, from the elements ICEH-Like MH232, ICEH-like MH7448, ICEH MH232 and ICEH MH7448, showed that a probable recombination event has occurred in an ancestral form of these genes, originating a mosaic structure comprising two phylogenetically distinct domains (Figure 5).

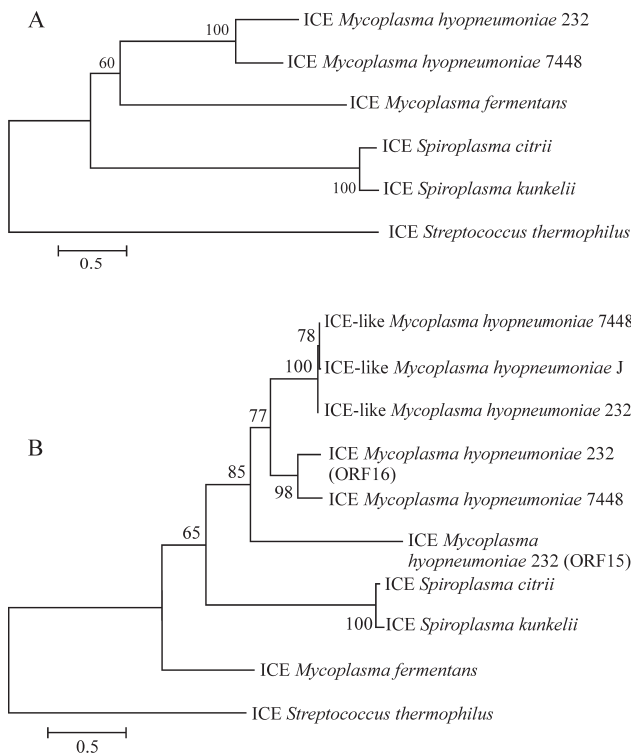


Figure 4 - Phylogenetic analysis of the ICE elements and TraE genes A: Neighbor-Joining tree calculated using the JTT distance of three concatenated orthologous CDS between the six analyzed elements. (The parsimony tree had the same topology of the Neighbor-joining analysis and was omitted from the figure). B: Neighbor-Joining tree calculated using the JTT distance of 10 TraE genes found in the elements ICEH-Like MH232, ICEH-Like MH7448, ICEH-Like MHJ, ICEH7448, ICEH232, pkstu, pksc and ICESt1.

Evidence for a circular intermediate ICEH

Extrachromosomal forms of the integrative conjugative element have been shown in *M. fermentans* (Calcutt *et al.*, 2002) and *M. agalactiae* (Marenda *et al.*, 2006). A model has been proposed where the element is excised, at low frequency, from the chromosome, circularized as a nonreplicative intermediate, and transferred by conjugation to a recipient. Finally, in the recipient cell, the element is integrated into the chromosome (Calcutt *et al.*, 2002).

To investigate the possible presence of extrachromosomal forms of ICEH in *M. hyopneumoniae*, two outwardly facing PCR primers have been designed, each one annealing at regions located near the ends of the element. The presence of a circular form should yield a fragment of about 600 bp after amplification by PCR. PCR reactions were carried out using 50 ng of genomic DNA from both *M. hyopneumoniae* strains (J and 7448). Figure 6 shows that a product of 633 bp was observed in *M. hyopneumoniae* strain 7448 and as expected, no product was observed in strain J. The PCR product was sequenced and perfectly aligned to the extremities of the ICEH, as expected if this element has a circular form (data not shown). This data indicates the presence of an ICEH circular intermediate in *M. hyopneumoniae* strain 7448, possibly gener-

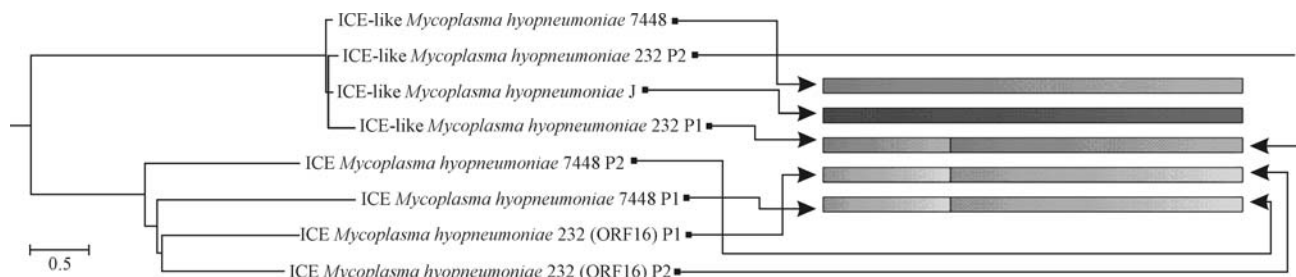


Figure 5 - Recombination tree based on the *traE* sequences on the left is based on the representation of the *traE* mosaic structure.

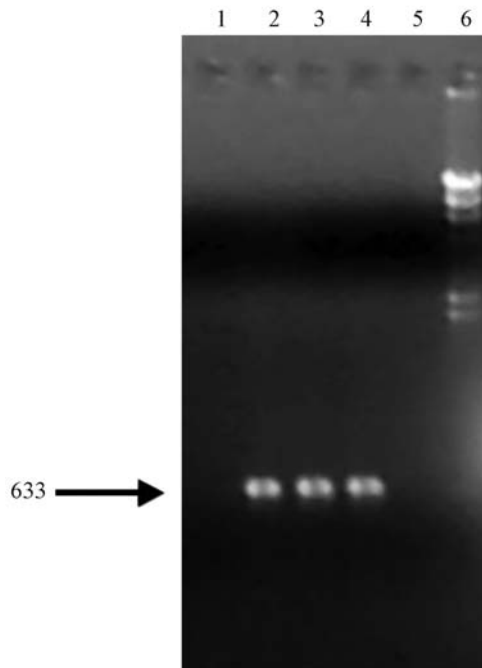


Figure 6 - Identification of extrachromosomal copy of ICEH. Genomic DNA from *M. hyopneumoniae* strains J (5) and 7448 (2-4) were used in PCR reactions with primers ICEH1-F and ICEH2-R. (1) negative control and (6) Lambda DNA digested with *Hind*III. Arrow shows the 633 amplicom from different DNA samples of *M. hyopneumoniae* strain 7448.

ated by the excision from genome, like occurred in *M. fermentans* (Calcutt *et al.*, 2002) and *M. agalactiae* (Marenda *et al.*, 2006).

The circular conformation of the excised ICEH has important implications for the biology of the element, since in this form, the constin can remain more stable, increasing the probability of horizontal transfers. As constins do not rely on transposases, which remain associated with the insertion sequence DNA molecule during the course of the excision/insertion process, a linear form would be quickly degraded by exonucleases, compromising the relocation to a different host or locus. Additionally, circular DNA forms are structurally more stable than linear ones, helping the element to maintain its structural integrity over the transfer process.

Concluding Remarks

Our analyses indicate the presence of putative integrative conjugal elements (ICE) in pathogenic strains of *M. hyopneumoniae*. It was also demonstrated that this element exists in an extra-chromosomal form. These results suggest that ICE is a mobile DNA that is probably involved in genomic recombination events and in pathogenicity. Nevertheless, the transference of ICEH between cells remains to be demonstrated experimentally. Genomic sequence analysis also indicated the existence of ICE-like elements in *M. hyopneumoniae* which are probably derived from an ancestral ICEH integration event that suffered an irregular excision. The ICEs are not confined to *M. hyopneumoniae*.

Sequences related to these elements were also found in other species such as *M. fermentans*, *M. bovis* and *M. agalactiae*. Interestingly, some CDSs of these elements have similarities to Type IV secretion machinery present in different bacterial species. This system is involved in the secretion of DNA and proteins to the extra-cellular milieu or into eukaryotic cells which have a central role in pathogenicity (reviewed by Christie *et al.*, 2005). This observation reinforces the role of ICEs in pathogenicity. However, there is no evidence that links them to protein secretion.

Acknowledgments

This work was undertaken by the Brazilian National Genome Program (Southern Network for Genome Analysis and Brazilian National Genome Project Consortium). PMP is the recipient of a CNPq pre-doctoral fellowship; MOC is the recipient of a DTI scholarship from CNPq. This work was supported by MCT/CNPq and SCT/FAPERGS (RS).

References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402.
- Auchtung JM, Lee CA, Monson RE, Lehman AP and Grossman AD (2005) Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proc Natl Acad Sci USA* 102:12554-12559.
- Boucher Y, Douady CJ, Papke RT, Walsh DA, Boudreau MER, Nesbø CL, Case RJ and Doolittle WF (2003) Lateral gene transfer and the origins of prokaryotic groups. *Annu Rev Genet* 37:283-328.
- Bru C, Courcelle E, Carrere S, Beausse Y, Dalmar S and Kahn D (2005) The ProDom database of protein domain families: More emphasis on 3D. *Nucleic Acids Res* 33:212-215.
- Calcutt MJ, Lewis MS and Wise KS (2002) Molecular genetic analysis of ICEF, an integrative conjugal element that is present as a repetitive sequence in the chromosome of *Mycoplasma fermentans* PG18. *J Bacteriol* 184:6929-6941.
- Christie PJ, Atmakuri K, Krishnamoorthy V, Jakubowski S and Cascales E (2005) Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu Rev Microbiol* 59:451-485.
- Doublet B, Boyd D, Mulvey MR and Cloeckaert A (2005) The Salmonella genomic island 1 is an integrative mobilizable element. *Mol Microbiol* 55:1911-1924.
- Enright AJ, Van Dongen S and Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30:1575-1584.
- Hochhut B and Waldor MK (1999) Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. *Mol Microbiol* 32:99-110.
- Hochhut B, Jahreis K, Lengeler JW and Schmid K (1997) CTnscr94, a conjugal transposon found in enterobacteria. *J Bacteriol* 179:2097-3102.
- Hochhut B, Lotfi Y, Mazel D, Faruque SM, Woodgate R and Waldor MK (2001) Molecular analysis of antibiotic resis-

- tance gene clusters in *Vibrio cholerae* O139 and O1 SXT constins. *Antimicrob Agents Chemother* 45:2991-3000.
- Jain R, Rivera MC, Moore JE and Lake JA (2003) Horizontal gene transfer accelerates genome innovation and evolution *Mol Biol Evol* 20:1598-1602.
- Kumar S, Tamura K and Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics analysis and sequence alignment. *Brief Bioinform* 5:150-163.
- Lerat E, Daubin V, Ochman H and Moran NA (2005) Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol* 3:130.
- Maes D, Verdonck M, Deluyker H and de Kruif A (1996) Enzootic pneumonia in pigs. *Vet Q* 18:104-109.
- Marenda MS, Sagne E, Poumarat F and Citti C (2005) Suppression subtractive hybridization as a basis to assess *Mycoplasma agalactiae* and *Mycoplasma bovis* genomic diversity and species-specific sequences. *Microbiol* 151:475-489.
- Marenda M, Barbe V, Gourgues G, Mangenot S, Sagne E and Citti C (2006) A new integrative conjugative element occurs in *Mycoplasma agalactiae* as chromosomal and free circular forms. *J Bacteriol* 188:4137-4141.
- Milne I, Wright F, Rowe G, Marshall DF, Husmeier D and McGuire G (2004) TOPALi: Software for automatic identification of recombinant sequences within DNA multiple alignments. *Bioinformatics* 20:1806-1807.
- Minion FC, Lefkowitz EJ, Madsen ML, Cleary BJ, Swartzell SM and Mahairas GG (2004) The genome sequence of *Mycoplasma hyopneumoniae* strain 232, the agent of swine mycoplasmosis. *J Bacteriol* 186:7123-7133.
- Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bradley P, Bork P, Bucher P, Cerutti L, *et al.* (2005) InterPro, progress and status in 2005. *Nucleic Acids Res* 33:201-205.
- Ochman H, Lerat E and Daubin V (2005) Examining bacterial species under the specter of gene transfer and exchange. *Proc Natl Acad Sci USA* 102:6595-6599.
- Pavlovic G, Burrus V, Gintz B, Decaris B and Guedon G (2004) Evolution of genomic islands by deletion and tandem accretion by site-specific recombination: ICES1-related elements from *Streptococcus thermophilus*. *Microbiology* 4:759-74.
- Porwollik S and McClelland M (2003) Lateral gene transfer in *Salmonella*. *Microbes Infect* 5:977-989.
- Ravatn R, Studer S, Springael D, Zehnder AJ and van der Meer JR (1998) Chromosomal integration, tandem amplification, and deamplification in *Pseudomonas putida* F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from *Pseudomonas* sp. Strain B13. *J Bacteriol* 180:4360-4369.
- Rice P, Longden I and Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* 16:276-277.
- Salyers AA, Shoemaker NB, Stevens AM and Li LY (1995) Conjugative transposons: An unusual and diverse set of integrated gene transfer elements. *Microbiol Rev* 59:579-590.
- Snyder L and Champness W (2002) *Molecular Genetics of Bacteria*, 2nd ed. ASM Press, Washington DC, 566 pp.
- Sobestiansky J, Matos MPC and Hirose F (2001) *Pneumonia Enzoótica Suína: Prevalência, Impacto Econômico, Fatores de Risco e Estratégia de Controle*. 1ra ed. Goiânia, 44 pp.
- Sonnhammer EL, Eddy SR and Durbin R (1997) Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins* 28:405-420.
- Sullivan JT and Ronson CW (1998) Evolution of *Rhizobia* by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci USA* 95:5145-5149.
- Tatusov RL, Galperin MY, Natale DA and Koonin EV (2000) The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28:33-36.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876-4882.
- Vasconcelos AT, Ferreira HB, Bizarro CV, Bonatto SL, Carvalho MO, Pinto PM, Almeida DF, Almeida LG, Almeida R, Alves-Filho L, *et al.* (2005) Swine and poultry pathogens: The complete genome sequences of two strains of *Mycoplasma hyopneumoniae* and a strain of *Mycoplasma synoviae*. *J Bacteriol* 187:5568-5577.
- Wang R, Prince JT and Marcotte EM (2005) Mass spectrometry of the *M. smegmatis* proteome: Protein expression levels correlate with function, operons, and codon bias. *Genome Res* 8:1118-26.
- Zhang Q, Young TF and Ross RF (1995) Identification and characterization of a *Mycoplasma hyopneumoniae* adhesin. *Infect Immun* 63:1013-1019.
- Zielinski GC and Ross RF (1990) Effect of growth in cell cultures and strain on virulence of *Mycoplasma hyopneumoniae* for swine. *Am J Vet Res* 51:344-348.
- Zielinski GC and Ross RF (1993) Adherence of *Mycoplasma hyopneumoniae* to porcine ciliated respiratory tract cells. *Am J Vet Res* 54:1262-1269.
- Internet Resources
 ProDom, <http://prodom.prabi.fr/prodom/current/html/home.php> (November 15, 2005).
 PFAM, <http://www.sanger.ac.uk/Software/Pfam/> (November 16, 2006).
 InterPro - The Integrated Resource of Protein Domains and Protein Families, <http://www.ebi.ac.uk/interpro/> (November 12, 2005).
 National Center for Biotechnology Information (NCBI), <http://www.ncbi.nlm.nih.gov>.
 National Center for Biotechnology Information (NCBI) Microbial Genomes, <http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi> (September 05, 2006).

Associate Editor: Ana Tereza Vasconcelos