A comparative genomics and immunoinformatics approach to identify epitope-based peptide vaccine candidates against bovine hemoplasmosis

Rosa Estela Quiroz-Castañeda*, Hugo Aguilar-Díaz, Diana Laura Flores-García, Fernando Martínez-Ocampo, Itzel Amaro-Estrada

1Unidad de Anaplasmosis, Centro Nacional de Investigación Disciplinaria en Salud Animal e Inocuidad, INIFAP. Carretera Federal Cuernavaca-Cuautla No. 8534, Col. Progreso, 62550, Jiutepec, Morelos, México.

2Unidad de Artropodología, Centro Nacional de Investigación Disciplinaria en Salud Animal e Inocuidad, INIFAP. Carretera Federal Cuernavaca-Cuautla No. 8534, Col. Progreso, 62550, Jiutepec, Morelos, México.

3Universidad Politécnica del Estado de Morelos, Paseo Cuauhnahuac 566, Lomas del Texcal, C.P. 62574 Jiutepec, Morelos, México.

4Laboratorio de Estudios Ecogenómicos, Centro de Investigación en Biotecnología, Universidad Autónoma del Estado de Morelos, Av. Universidad No. 1001, Col. Chamilpa, 62209, Cuernavaca, Morelos, México.

*Corresponding author

E-mail: requiroz79@yahoo.com.mx
Abstract

Mycoplasma wenyonii and ‘Candidatus Mycoplasma haemobos’ have been described as major hemoplasmas that infect cattle worldwide. Currently, three bovine hemoplasma genomes are known. The aim of this work was to know the main genomic characteristics and the evolutionary relationships between hemoplasmas, as well as to provide a list of epitopes identified by immunoinformatics that could be used as vaccine candidates against bovine hemoplasmosis. So far, there is not a vaccine to prevent this disease that impact economically in cattle production around the world.

In this work, we used comparative genomics to analyze the genomes of the hemoplasmas so far reported. As a result, we confirm that ‘Ca. M haemobos’ INIFAP01 is a divergent species from M. wenyonii INIFAP02 and M. wenyonii Massachusetts. Although both strains of M. wenyonii have genomes with similar characteristics (length, G+C content, tRNAs and position of rRNAs) they have different structures (alignment coverage and identity of 51.58 and 79.37%, respectively).

The correct genomic characterization of bovine hemoplasmas, never studied before, will allow to develop better molecular detection methods, to understand the possible pathogenic mechanisms of these bacteria and to identify epitopes sequences that could be used in the vaccine design.

Introduction

Hemotrophic mycoplasmas (hemoplasmas) are a group of erythrocytic pathogens of the Mollicutes class that infect a wide range of vertebrate animals [1,2]. At first, these small and uncultivable in vitro bacteria were classified as species of the genera Haemobartonella and Eperythrozoon, within the Anaplasmataceae family and Rickettsiales order [1]. However, the
genetic analysis of 16S ribosomal RNA (rRNA) gene and morphologic similarities showed that these bacteria are closely related to the *Mycoplasma* genus [2,3]. In 2001, the formal proposal was presented to transfer these organisms to genus *Mycoplasma*, within *Mycoplasmataceae* family [4]. Currently, 12 hemoplasma genomes have been identified in the GenBank database, including *Mycoplasma wenyonii* strains and *Candidatus* Mycoplasma haemobos [5–7]. These hemoplasma genomes have provided relevant information about possible pathogenic mechanisms, metabolism and divergences when compared to other *Mycoplasma* species [2].

To date, *M. wenyonii* and *Ca. M. haemobos* have been described as major hemoplasmas that infect cattle worldwide [8–10]. In cattle, acute hemoplasma infections are rare but are characterized by anemia, fever, depression and diarrhea [11,12]. Chronic bovine hemoplasma infections have been associated with variable clinical signs, including low-grade bacteremia, weight loss, decreased milk production, reduced calf birth weight, pyrexia, scrotal and hind limb edema, infertility and reproductive inefficiency, and consequently, the bovine hemoplasmas have caused major economic losses worldwide, mainly when they associate with pathogens of genus *Anaplasma* or *Babesia* [13–15]. In addition, latent and asymptomatic infections have also been reported [14]. Single infections or coinfections between *M. wenyonii* and ‘*Ca. M. haemobos*’ are being reported [16–18]. However, reports of the genomic characterization of bovine hemoplasmas are scarce. So far, two genomes of this species were reported: *M. wenyonii* Massachusetts [5] and *M. wenyonii* INIFAP02 [7] and only one genome of this species was reported: ‘*Ca. M. haemobos*’ INIFAP01 [6]. On the other hand, computational *in silico* tools have been used to design vaccines by rational and cost effective manner, this strategy has several advantages, including prolonged immunity, elimination of unspecific responses and cost-and time-effectiveness [19,20]. In this sense, immunoinformatics is an effective tool that helps to
predict and to identify immunogenic sequences and the epitopes that could be recognized by antibodies that induce immune responses [21].

In this work, we analyzed, for the first time, the genomic characteristics and the evolutionary relationships between bovine hemoplasmas. Also, we performed a immunoinformatic analysis to elucidate B-cell epitopes found in several proteins of *M. wenyonii* and *Ca. M. haemobos* that could be used as potential vaccine candidates to prevent bovine hemoplasmosis.

### Materials and methods

#### Genome sequences and annotation

The 12 hemoplasma genomes that infect different hosts, included in this study and reported in the GenBank database ([https://bit.ly/314fOre](https://bit.ly/314fOre)), are listed in Table S1. In Mexico, the Anaplasmosis Unit (CENID-SAI, INIFAP) has reported the draft genomes of two Mexican strains of hemoplasmas that infect cattle, ‘*Ca. Mycoplasma haemobos*’ INIFAP01 [6] and *M. wenyonii* INIFAP02 [7]. The general features of 12 hemoplasma genomes were obtained using the QUAST (Quality Assessment Tool for Genome Assemblies) (v5.0.2) program [22] with default settings.

All genomes were annotated automatically to predict the coding sequences (CDS) using the RAST (Rapid Annotation using Subsystem Technology) (v2.0) server ([https://bit.ly/2XjTTev](https://bit.ly/2XjTTev)) [23] with the Classic RAST algorithm.

The mapping of ribosomal genes (rRNA) was done based on the information reported in NCBI database of genomes of *M. wenyonii* Massachusetts, *M. wenyonii* INIFAP02 and *Ca. M. haemobos* INIFAP01 (NC_018149.1; NZ_QKVO00000000.1; and LWUJ00000000.1, respectively). Transfer (tRNA) RNA genes was carried out using ARAGORN (v1.2.38)
The sequence and length of 16S and 23S rRNA genes was obtained from the RNAmmer (v1.2) server [25].

**Phylogenetic and pan-genomic analysis**

For the phylogenetic reconstruction, the 16S rRNA gene sequences of 15 bovine hemoplasmas were aligned with 22 downloaded rRNA gene sequences of other hemoplasma species and two downloaded rRNA gene sequences of the genus *Ureaplasma*, which were obtained from the GenBank database using the nucleotide BLAST (Blastn) suite [26]. Multiple alignments between 39 16S rRNA gene sequences were made using the MUSCLE (v3.8.31) program [27]. The jModelTest (v2.1.10) program was used to select the best model of nucleotide substitution with the Akaike information criterion. The phylogenetic tree was estimated under the Maximum-Likelihood method using the PhyML (v3.1) program [29] with 1,000 bootstrap replicates. The phylogenetic tree was visualized and edited using the FigTree (v.1.4.4) program.

Two pan-genomic analyzes were performed using the GET_HOMOLOGUES (v3.3.2) software package [30] with the following options: i) among the 12 hemoplasma genomes; and ii) among the three genomes of bovine hemoplasmas. Briefly, the FAA (Fasta Amino Acid) annotation files of hemoplasma genomes were used as input files by the GET_HOMOLOGUES software package. The get_homologues.pl and compare_clusters.pl Perl scripts were used to compute a consensus pan-genome, which resulting from the clustering of the all-against-all protein BLAST (Blastp) results with the COGtriangles and OMCL algorithms. The pan-genomic analysis was performed using the binary (presence-absence) matrix.

**Comparative genomics**
The average nucleotide identity (ANI) values of 12 hemoplasma genomes were calculated using the calculate_ani.py Python script (https://bit.ly/2X96hho) with the BLAST-based ANI (ANIb) algorithm. Ultimately, the level of conserved genomic sequences of bovine hemoplasmas was visualized by alignment the genomes of Mexican strains (‘Ca. M. haemobos’ INIFAP01 and M. wenyonii INIFAP02) against the reference genome of M. wenyonii Massachusetts, using the NUCmer program of MUMmer (v3.0) software package (Kurtz et al., 2004) to get the positions of nucleotides that were aligned; and Circos (v0.69-9) software package [32]. The circular comparative genomic map of bovine hemoplasmas was edited with Adobe Photoshop CC (v14.0 x64) program.

**Prediction of antigenic proteins**

After RAST annotation, we identified several proteins of the subsystems, including virulence, disease and defense, cell division and cell cycle, fatty acids, lipids and isoprenoids, regulation and cellular signaling, stress response, and DNA metabolism. To predict antigenicity, the sequence of 11 proteins of M. wenyonii and 12 proteins of Ca. M. haemobos were submitted to VaxiJen v2.0 server (http://www.ddgpharmfac.net/vaxijen/VaxiJen/VaxiJen.html) with default parameters.

**Prediction of subcellular localization and stability of the proteins**

Predicted antigenic proteins of M. wenyonii and Ca. M. haemobos were submitted to the prediction of secondary structure server Raptor X (http://raptorx.uchicago.edu/).

**Linear B-cell epitope prediction and three-dimensional modelling**

B-Cell epitopes were predicted using BCEpred (http://crdd.osdd.net/raghava/bcepred/), and Predicting Antigenic Peptides tool (http://imed.med.ucm.es/Tools/antigenic.pl).
The PHYRE2 server was used to predict the tridimensional structure of the proteins of both hemoplasmas. Phyre2 PDB files were visualized with Protter tool.

**Results**

**General features of genomes**

The major genomic features of hemoplasmas are shown in Table 1.

**Table 1. General features of 12 hemoplasma genomes.**

<table>
<thead>
<tr>
<th>Organism</th>
<th>Assembly level</th>
<th>Length (bp)*</th>
<th>G+C content (%)*</th>
<th>CDS **</th>
<th>rRNAs#</th>
<th>tRNAs##</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Ca. M. haemobos’ INIFAP01</td>
<td>18 contigs</td>
<td>935,638</td>
<td>30.46</td>
<td>1,180</td>
<td>3</td>
<td>31</td>
</tr>
<tr>
<td><em>M. haemocanis Illinois</em></td>
<td>Chromosome</td>
<td>919,992</td>
<td>35.33</td>
<td>1,234</td>
<td>3</td>
<td>31</td>
</tr>
<tr>
<td><em>M. haemofelis Langford 1</em></td>
<td>Chromosome</td>
<td>1,147,259</td>
<td>38.85</td>
<td>1,595</td>
<td>3</td>
<td>31</td>
</tr>
<tr>
<td><em>M. haemofelis Ohio2</em></td>
<td>Chromosome</td>
<td>1,155,937</td>
<td>38.81</td>
<td>1,650</td>
<td>3</td>
<td>31</td>
</tr>
<tr>
<td>‘Ca. M. haemolamae’ Purdue</td>
<td>Chromosome</td>
<td>756,845</td>
<td>39.27</td>
<td>1,045</td>
<td>3</td>
<td>33</td>
</tr>
<tr>
<td>‘Ca. M. haemominutum’ Birmingham 1</td>
<td>Chromosome</td>
<td>513,880</td>
<td>35.52</td>
<td>587</td>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td><em>M. ovis Michigan</em></td>
<td>Chromosome</td>
<td>702,511</td>
<td>31.69</td>
<td>918</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td><em>M. parvum Indiana</em></td>
<td>Chromosome</td>
<td>564,395</td>
<td>26.98</td>
<td>578</td>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td><em>M. suis Illinois</em></td>
<td>Chromosome</td>
<td>742,431</td>
<td>31.08</td>
<td>914</td>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td><em>M. suis K13806</em></td>
<td>Chromosome</td>
<td>709,270</td>
<td>31.08</td>
<td>856</td>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td><em>M. wenyonii INIFAP02</em></td>
<td>37 contigs</td>
<td>596,665</td>
<td>33.43</td>
<td>678</td>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td><em>M. wenyonii Massachusetts</em></td>
<td>Chromosome</td>
<td>650,228</td>
<td>33.92</td>
<td>727</td>
<td>3</td>
<td>32</td>
</tr>
</tbody>
</table>

CDS: coding sequences; *data obtained with the QUAST program; **data obtained with the RAST server; #data obtained with the RNAmmer server; and ##data obtained with the ARAGORN server.

Of the 12 hemoplasma genomes, ten genomes are assembled in a single chromosome and two draft genomes are assembled in contigs. The genomic features of hemoplasmas allow them to be separated into two groups: the group 1 (previously named *Haemobartonella*) consists of four genomes of ‘Ca. M. haemobos’, *M. haemocanis* and *M. haemofelis* species, that have a length...
from 0.9 to 1.1 Mb, a number of CDS from 1,180 to 1,650, and specifically 31 tRNA genes; and
the group 2 (previously named *Eperythrozoon*) consists of eight genomes of ‘Ca. M.
haemolamae’, ‘Ca. M. haemominutum’, *M. ovis*, *M. parvum*, *M. suis* and *M. wenyonii* species,
that have a length from 0.5 to 0.7 Mb, a number of CDS from 578 to 1,045; and 32 or 33 tRNA
genes.
The mapping of rRNA genes shows that hemoplasmas are also separated into two groups. The
due genomes of group 1 contain one copy of 16S-23S-5S rRNA operon (Fig 1A). The 16S
rRNA gene sequences of group 1 have a length from 1,459 to 1,460 bp. Conversely, seven
genomes of group 2 contain one copy of 16S rRNA gene with a length from 1,479 to 1,499 pb,
which is separated from one copy of 23S-5S rRNA operon (Fig 1B and 1C). Also, the genome of
*M. ovis* Michigan of group 2 contains two copies of 16S rRNA gene with lengths of 1,467 and
3,219 bp, which are separated from each other, and they are separated from the one copy of 23S-
5S rRNA operon (Fig 1D).

**Fig 1. Mapping of rRNA genes of hemoplasmas.** A) Genomes of ‘Ca. M. haemobos’
INIFAP01, *M. haemocanis* Illinois, *M. haemofelis* Langford 1 and *M. haemofelis* Ohio2 of group
1 contain one copy of 16S-23S-5S rRNA operon. B) Genomes of ‘Ca. M. haemolamae’ Purdue,
INIFAP02 and *M. wenyonii* Massachusetts of group 2 contain one copy of 16S rRNA gene
which is separate from one copy of 23S-5S rRNA operon in different chain. C) Genome of *M.
parvum* Indiana of group 2 contains one copy of 16S rRNA gene which is separate from one
copy of 23S-5S rRNA operon in the same chain. D) Genome of *M. ovis* Michigan of group 2
contains two copies of 16S rRNA gene which are separated from each other, and they are
separated from the one copy of 23S-5S rRNA operon in the same chain. The 16S, 23S and 5S rRNA genes are represented by blue, green and red arrows, respectively.

The 16S rRNA gene sequence of ‘Ca. M. haemobos’ INIFAP01 has alignment coverage of 82-98% and identity of 98.71-99.93% with ‘Ca. M. haemobos’, ‘Ca. M. haemobos’ clone 307, ‘Ca. M. haemobos’ clone 311 and ‘Ca. M. haemobos’ isolate cattle no. 18. Additionally, 16S rRNA gene sequence of ‘Ca. M. haemobos’ INIFAP01 has alignment coverage of 99% and identity of 81.83 and 81.73% with M. wenyonii INIFAP02 and M. wenyonii Massachusetts, respectively.

On the other hand, the genomes of M. wenyonii INIFAP02 and M. wenyonii Massachusetts are very similar in length and G+C content to each other, and they have the same number of tRNA genes and distribution of rRNA genes. However, the 16S rRNA gene sequence of M. wenyonii INIFAP02 has alignment coverage of 100% and identity of 97.57% with M. wenyonii Massachusetts. Additionally, 16S rRNA gene sequence of M. wenyonii INIFAP02 has: i) alignment coverage of 91-98% and identity of 99.24-99.93% with M. wenyonii isolate Fengdu, M. wenyonii clone 1, M. wenyonii isolate ada1 and M. wenyonii isolate C124; and ii) alignment coverage of 90-98% and identity of 97.50-97.87% with M. wenyonii strain CGXD, M. wenyonii isolate B003, M. wenyonii isolate C031 and M. wenyonii strain Langford.

**Phylogenetic and pan-genomic analyzes**

The model of nucleotide substitution of the phylogenetic tree that is based on 16S rRNA gene of hemoplasmas was GTR+I+G. The phylogenetic tree shows that groups 1 and 2 of hemplasma species are separated into different clades (Fig 2). The clade 1 (blue lines) contains two subclades: i) ‘Ca. M. haemobos’ species; and ii) M. haemocanis and M. haemofelis species. The clade 2 (red lines) also contains two subclades: i) ‘Ca. M. haemolamae’, ‘Ca. M.
haemominutum’, *M. ovis* and *M. wenyonii* species; and ii) *M. parvum* and *M. suis* species. The phylogenetic tree topology shows a divergence between two groups of hemoplasmas.

**Fig 2. Phylogenetic relationships based on the 16S rRNA genes of group 1 (blue lines) and 2 (red lines) of hemoplasma species.** The phylogenetic tree was obtained using the PhyML program with the Maximum-Likelihood method and 1,000 bootstrap replicates. Bootstrap values (>50%) are displayed in the nodes. The model of nucleotide substitution was GTR+I+G. The INIFAP01 and INIFAP02 Mexican strains of bovine hemoplasmas are shown in blue and red letters, respectively. GenBank accession numbers are shown in square brackets.

Pan-genomic analysis among the 12 hemoplasmas shows that the core, soft core, shell and cloud genomes are composed of 110, 146, 787 and 3,099 gene clusters, respectively (Figs S1 and S2). Additionally, the core genomes of groups 1 and 2 of hemoplasmas are composed of 236 and 149 gene clusters, respectively.

Pan-genomic analysis among the three genomes of bovine hemoplasmas shows that the core genome is composed of 154 gene clusters. Also, the two genomes of *M. wenyonii* species share 273 gene clusters. Moreover, ‘Ca. *M. haemobos*’ INIFAP01, *M. wenyonii* INIFAP02 and *M. wenyonii* Massachusetts contain 312, 190 and 157 unique gene clusters, respectively.

**Comparative genomics**

ANIb values between different hemoplasma species show that alignment coverage is less than 79% (Fig 3A and Table S2); and identity is less than 83% (Fig 3B and Table S3).

**Fig 3. Heatmaps of BLAST-based average nucleotide identity (ANIb) values of 12 hemoplasma genomes.** A) Heatmap of ANIb values of alignment coverage. B) Heatmap of
ANIb values of identity. Color intensity increases from white to deep blue when ANIb values approach from 0.0 to 1.0 (0 - 100%), respectively.

Also, ANIb values show that ‘Ca. M. haemobos’ INIFAP01 has an alignment coverage of 0.46 and 0.31%; and identity of 74.12 and 74.16% with M. wenyonii INIFAP02 and M. wenyonii Massachusetts, respectively. Moreover, ANI values between the same species show that: i) M. haemofelis genomes have an alignment coverage and identity of 97.65 and 97.41%, respectively; ii) M. suis genomes have an alignment coverage and identity of 95.13 and 97.63%, respectively; and iii) M. wenyonii genomes have an alignment coverage and identity of 51.58 and 79.37%, respectively.

The circular map (Fig 4) shows that ‘Ca. M. haemobos’ INIFAP01 genome only has three small regions (red lines highlighted with green marker in inner track) greater than 78% identity that were aligned with M. wenyonii Massachusetts genome (black circle in outer track). Also, the circular map shows that M. wenyonii INIFAP02 has few regions (blue lines in intermediate track) greater than 78% identity that were aligned with M. wenyonii Massachusetts genome.

Fig 4. Circular map of M. wenyonii Massachusetts and ‘Ca. M. haemobos’ INIFAP01 genomes.

Selection and prediction of B-cell epitopes in proteins

The sequences of 11 proteins of M. wenyonii and 12 proteins of Ca. M. haemobos were submitted to VaxiJen. The prediction of Antigen/Non Antigen for each selected protein is shown in Table 2.
**Candidatus Mycoplasma haemobos INIFAP01**

<table>
<thead>
<tr>
<th>Classification</th>
<th>Prediction score as antigen (Vaxijen)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAST Category: Virulence, disease, and defense</td>
<td></td>
</tr>
<tr>
<td>DNA gyrase subunit B</td>
<td>0.5352 (Antigen)</td>
</tr>
<tr>
<td>DNA gyrase subunit A</td>
<td>0.4373 (Antigen)</td>
</tr>
<tr>
<td>SSU ribosomal protein S7p</td>
<td>0.518 (Antigen)</td>
</tr>
<tr>
<td>Translation Elongation factor G</td>
<td>0.5399 (Antigen)</td>
</tr>
<tr>
<td>Translation Elongation factor Thermo unstable (Tu)</td>
<td>0.4268 (Antigen)</td>
</tr>
<tr>
<td>SSU ribosomal protein S12p</td>
<td>0.7537 (Antigen)</td>
</tr>
<tr>
<td>DNA-directed RNA polymerase beta subunit</td>
<td>0.3781 (Antigen)</td>
</tr>
<tr>
<td>DNA-directed RNA polymerase</td>
<td>0.4488 (Antigen)</td>
</tr>
<tr>
<td>RAST Category: Division and Cell Cycle</td>
<td></td>
</tr>
<tr>
<td>ProteinTsaD/Kae1/Qri7</td>
<td>0.3846 (Non Antigen)</td>
</tr>
<tr>
<td>RNA polymerase sigma factor RpoD</td>
<td>0.3857 (Non Antigen)</td>
</tr>
<tr>
<td>DNA primase</td>
<td>0.3009 (Non Antigen)</td>
</tr>
<tr>
<td>RAST Category: Fatty acids, Lipids and Isoprenoids</td>
<td></td>
</tr>
<tr>
<td>Cardiolipin synthase</td>
<td>0.3463 (Non Antigen)</td>
</tr>
<tr>
<td>RAST Category: Stress Response</td>
<td></td>
</tr>
<tr>
<td>Manganase superoxide dismutase</td>
<td>0.3487 (Non Antigen)</td>
</tr>
</tbody>
</table>

**Mycoplasma wenyonii INIFAP02**

<table>
<thead>
<tr>
<th>Classification</th>
<th>Prediction score as antigen (Vaxijen)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAST Category: Virulence, disease, and defense</td>
<td></td>
</tr>
<tr>
<td>Ribosomal protein SSU S7p</td>
<td>0.5435 (Antigen)</td>
</tr>
<tr>
<td>Translation Elongation factor G</td>
<td>0.5650 (Antigen)</td>
</tr>
<tr>
<td>Translation Elongation factor Thermo unstable (Tu)</td>
<td>0.4359 (Antigen)</td>
</tr>
<tr>
<td>Ribosomal protein SSU S12p</td>
<td>0.7774 (Antigen)</td>
</tr>
<tr>
<td>Ribosomal protein LSU L35p</td>
<td>0.5491 (Antigen)</td>
</tr>
</tbody>
</table>
The collection of B-cell epitopes (linear antigens) predicted with Predicting Antigenic Peptides tool and BCEPred server is shown in Table 3.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Position (aa)</th>
<th>Epitope sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ca. M. haemobos INIFAP01 DNA gyrase subunit B</td>
<td>23-33</td>
<td>SSIRVLEGLEA</td>
</tr>
<tr>
<td>M. wenyonii INIFAP02 DNA gyrase subunit A</td>
<td>97-105</td>
<td>SIYDALVRM</td>
</tr>
<tr>
<td></td>
<td>107-116</td>
<td>QDFSLRYPPI</td>
</tr>
<tr>
<td></td>
<td>139-148</td>
<td>RLSKLGKYGFL</td>
</tr>
<tr>
<td>DNA gyrase subunit B</td>
<td>48-59</td>
<td>KGLHHLIWEVLD</td>
</tr>
<tr>
<td></td>
<td>77-87</td>
<td>LKKGHIISVSVD</td>
</tr>
<tr>
<td></td>
<td>131-138</td>
<td>GVGSTCVN</td>
</tr>
<tr>
<td></td>
<td>140-150</td>
<td>LSSFLEVNVYR</td>
</tr>
<tr>
<td></td>
<td>216-225</td>
<td>GKKFVFVNEI</td>
</tr>
<tr>
<td></td>
<td>251-261</td>
<td>SIHDVIFHSE</td>
</tr>
<tr>
<td></td>
<td>289-297</td>
<td>SSIVHSFCN</td>
</tr>
<tr>
<td></td>
<td>313-321</td>
<td>DGLLSCIRE</td>
</tr>
<tr>
<td></td>
<td>392-401</td>
<td>QKRVILQRVD</td>
</tr>
<tr>
<td></td>
<td>442-451</td>
<td>FSELYVVEGD</td>
</tr>
<tr>
<td></td>
<td>546-554</td>
<td>LLEHGYVYI</td>
</tr>
<tr>
<td></td>
<td>565-574</td>
<td>NKEVYVYLDD</td>
</tr>
<tr>
<td>RAST Category: Division and Cell Cycle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein Tsad/Kae1/Qri7</td>
<td>97-105</td>
<td>SIYDALVRM</td>
</tr>
<tr>
<td></td>
<td>107-116</td>
<td>QDFSLRYPPI</td>
</tr>
<tr>
<td></td>
<td>139-148</td>
<td>RLSKLGKYGFL</td>
</tr>
<tr>
<td></td>
<td>167-178</td>
<td>NTPVIRGSLAKA</td>
</tr>
<tr>
<td>DNA primase</td>
<td>54-63</td>
<td>NYQVPVESSK</td>
</tr>
<tr>
<td></td>
<td>67-76</td>
<td>EALALRWLKI</td>
</tr>
<tr>
<td>RAST Category: Fatty acids, lipids and isoprenoids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiolipin synthase</td>
<td>26-33</td>
<td>ERIILYTG</td>
</tr>
<tr>
<td></td>
<td>70-77</td>
<td>WKGVVNLNL</td>
</tr>
<tr>
<td></td>
<td>122-132</td>
<td>NKYKVPRIIFC</td>
</tr>
<tr>
<td></td>
<td>154-161</td>
<td>NIKFSPIQ</td>
</tr>
<tr>
<td></td>
<td>212-220</td>
<td>LLNEVLVYD</td>
</tr>
<tr>
<td></td>
<td>237-244</td>
<td>EIKMCIRK</td>
</tr>
</tbody>
</table>

Table 3. B-Cell epitopes of Ca. M. haemobos and M. wenyonii predicted by immunoinformatics.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Position (aa)</th>
<th>Epitope sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSU ribosomal protein S7p</td>
<td>10-18</td>
<td>LRKFAKVVRV</td>
</tr>
<tr>
<td></td>
<td>20-29</td>
<td>NNHEVLAYIP</td>
</tr>
<tr>
<td></td>
<td>35-42</td>
<td>LQEHHVVM</td>
</tr>
<tr>
<td>Ribosomal protein LSU L35p</td>
<td>15-24</td>
<td>LSKRVVVLGS</td>
</tr>
<tr>
<td></td>
<td>59-67</td>
<td>QYKITAHLL</td>
</tr>
<tr>
<td>Translation Elongation factor G</td>
<td>82-96</td>
<td>GHVDFTEVERSLRV</td>
</tr>
<tr>
<td></td>
<td>142-151</td>
<td>YFKSVQSLRE</td>
</tr>
<tr>
<td></td>
<td>143-149</td>
<td>YFKSVQSL</td>
</tr>
<tr>
<td></td>
<td>153-162</td>
<td>LKVNAVLQL</td>
</tr>
<tr>
<td></td>
<td>170-178</td>
<td>FIGIIDLT</td>
</tr>
<tr>
<td></td>
<td>211-219</td>
<td>ELLDEVLY</td>
</tr>
<tr>
<td></td>
<td>235-244</td>
<td>VEEIKHCIRI</td>
</tr>
<tr>
<td></td>
<td>272-282</td>
<td>VIDYLPSPIDL</td>
</tr>
<tr>
<td></td>
<td>387-397</td>
<td>VKGNQVVLSEM</td>
</tr>
<tr>
<td></td>
<td>420-428</td>
<td>SMVLSRLSE</td>
</tr>
<tr>
<td></td>
<td>451-459</td>
<td>ELHLEILID</td>
</tr>
<tr>
<td></td>
<td>518-525</td>
<td>EFVDKIVG</td>
</tr>
<tr>
<td>Translation Elongation factor Thermo</td>
<td>87-94</td>
<td>FVLNRKVL</td>
</tr>
<tr>
<td>unstable (Tu)</td>
<td>107-114</td>
<td>KIVDTSVK</td>
</tr>
<tr>
<td>SSU ribosomal protein S12p</td>
<td>20-27</td>
<td>KSQALLKS</td>
</tr>
<tr>
<td></td>
<td>31-39</td>
<td>LDKKVTFQS</td>
</tr>
<tr>
<td></td>
<td>42-49</td>
<td>FKRGVCCTR</td>
</tr>
<tr>
<td></td>
<td>61-68</td>
<td>ALRKYAKV</td>
</tr>
<tr>
<td></td>
<td>72-81</td>
<td>NNYEVLAYIP</td>
</tr>
<tr>
<td></td>
<td>87-97</td>
<td>LQEHHVVMVRG</td>
</tr>
<tr>
<td>DNA-directed RNA polymerase beta</td>
<td>61-71</td>
<td>NGLFCESIFGP</td>
</tr>
<tr>
<td>subunit</td>
<td>111-120</td>
<td>HIELACPVAH</td>
</tr>
<tr>
<td></td>
<td>168-177</td>
<td>IFSELEIIDV</td>
</tr>
<tr>
<td>DNA-directed RNA polymerase</td>
<td>7-17</td>
<td>RVNSFSPIVNR</td>
</tr>
<tr>
<td></td>
<td>32-42</td>
<td>DLKGVQQLGAYT</td>
</tr>
</tbody>
</table>
Discussion

The hemoplasmas have undergone phylogenetic reclassification after several studies based on molecular markers [33]. Their genome size variation, positional shuffling of genes and poorly conserved gene synteny are evidence of the high dynamic of their genomes [2]. In this work, we found that 12 genomes of hemoplasmas are classified in two groups, and have a different number of CDS and number of tRNAs, additionally, the G+C content vary from 30.46 to 39.27% between the members of the two groups. Thus, G+C content is not specific to each group.

Specifically, in bovine hemoplasmas we found differences in genomic features between species. The genome of ‘Ca. M. haemobos’ INIFAP01 is significantly longer in length than two genomes of M. wenyonii species, but ‘Ca. M. haemobos’ INIFAP01 has a lower G+C content. Also, the number of tRNA genes and distribution of rRNA genes are specific to each species. The phylogenetic tree shows that ‘Ca. M. haemobos’ INIFAP01 is phylogenetically distant through evolution with M. wenyonii INIFAP02 and M. wenyonii Massachusetts. Also, the INIFAP02 and Massachusetts strains are closely related through the evolution of group 2.
The number of genes in the core, soft, shell and cloud genomes revealed in the pan-genomic analysis suggest that there is considerable loss/gain of genes through evolution of the 12 hemoplasmas genomes. Also, the genomes of group 1 (four genomes of three species) are more conserved than group 2 (eight genomes of six species); however, to confirm the previous result it is necessary to use a greater number of genomes of different species of group 1.

In regard to pan-genomic analysis of bovine hemoplasmas, due to the low number of gene clusters in the core genome it confirms that ‘Ca. M. haemobos’ INIFAP01 is a divergent species from *M. wenyonii* INIFAP02 and *M. wenyonii* Massachusetts. In comparative genomics of bovine hemoplasmas, the low percentages of alignment coverage and identity between *Ca. M. haemobos*, *M. wenyonii* INIFAP02 and *M. wenyonii* Massachusetts in the ANI values suggest that *Ca. M. haemobos* INIFAP01 genome has a different structure than genomes of *M. wenyonii* INIFAP02 and *M. wenyonii* Massachusetts.

Surprisingly, the alignment coverage and identity percentages among both genomes of *M. wenyonii* (strains INIFAP02 and Massachusetts) suggest that these strains may not belong to the same species because the ANI values were <95%, the species ANI cutoff value [34–36].

A visual evaluation of circular map suggests that genomes of three bovine hemoplasmas are not conserved between them, in fact, this data confirms that ‘*Ca. M. haemobos*’ INIFAP01 genome has a highly different structure than genomes of *M. wenyonii* INIFAP02 and *M. wenyonii* Massachusetts.

Since bovine hemoplasmas show significant differences at genomic level and they impact in cattle health causing economic losses, we decided to perform an immune-informatic analysis to identify B-cell epitopes that could be used in the design of potential vaccines to prevent bovine hemoplasmosis. The development of vaccines based on this strategy has been successfully used
to prevent some diseases in human and animals, including the cytoplasmic protein subolesin used
to prevent infestations of tick *Rhipicephalus microplus* [37–40]. The immune-informatics
analysis predicts several B-cell epitopes (Table 3), which could be used in the design of
molecular detection methods and vaccines. Peptides that contains epitopes have been applied
successfully for pathogen detection by serological methods [41,42], immunolocalization of
pathogen proteins [43] and vaccines against animal diseases [44–46]. The epitope collection
generated in this work will help to design molecular tools that contribute to prevent or diagnose
bovine hemoplasmosis.

**Conclusions**

In this work, we described the main genomic characteristics and the evolutionary relationships
between three bovine hemoplasmas. This is the first study to report the genomic characteristics
of bovine hemoplasma species. Also, the data presented here about antigenic peptides of *M.
wenyonii* INIFAP02 and *Ca. M. haemobos* INIFAP01 identified by immune-informatics have
potential uses to detect and/or prevent hemoplasmosis.

**Acknowledgements**

To Maria Gabriela Guerrero Ruiz (Bioinformatics Analysis Unit, CCG, UNAM) for kindly
collaborating with the comparative genomics analysis when using the CIRCOS program. This
work was partially granted by Consejo Nacional de Ciencia y Tecnología (CONACyT) Project
PN-CONACyT 248855 and CONACyT scholarship 293552.

**References**

**References**


12. Ade J, Niethammer F, Schade B, Schilling T, Hoelzle K, Hoelzle LE. Quantitative analysis of Mycoplasma wenyonii and ‘Candidatus Mycoplasma haemobos’ infections in


http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3942926/


design: a comprehensive overview of bioinformatics approaches. Drug Discov Today [Internet]. 2020; Available from: https://doi.org/10.1016/j.drudis.2020.03.006


425 genomes should be verified using average nucleotide identity and multilocus phylogenetic
427
428 Design of an Epitope-Based Vaccine Ensemble for Chagas Disease. Front Immunol.
430
431 vaccination strategy against Toxoplasma gondii infection: A mini-review. J Adv Vet
432 Anim Res [Internet]. 2019 Mar 24;6(2):174–82. Available from:
434
435 protective linear B-cell epitopes on the subolesin/akirin orthologues of Ornithodoros spp.
437
438 Targeting arthropod subolesin/akirin for the development of a universal vaccine for
439 control of vector infestations and pathogen transmission. Vet Parasitol. 2011
440 Sep;181(1):17–22.
441
41. Liu X, Marrakchi M, Xu D, Dong H, Andreescu S. Biosensors based on modularly
442 designed synthetic peptides for recognition, detection and live/dead differentiation of
445
42. Quiroz-Castañeda R, Tapia-Uriza TR, Mujica-Valencia C, Rodríguez-Camarillo SD,
446 Preciado-De la Torre JF, Amaro-Estrada I, et al. Synthetic peptides-based indirect ELISA


Supporting information

S1_Fig. Pan-genomic analysis of 12 hemoplasmas.

S2_Fig. Core, soft core, shell and cloud genomes of hemoplasmas.

S1_Table. 12 hemoplasma genomes reported in the GenBank database.

S2_Table. BLAST-based average nucleotide identity (AN Ib) values of alignment coverage of 12 hemoplasma genomes.

S3_Table. BLAST-based average nucleotide identity (AN Ib) values of identity of 12 hemoplasma genomes.
**A) Clade I**

(+)  

16S  

(1,429 - 1,486 bp)  

23S  

(2,833 - 2,899 bp)  

5S  

(103 - 114 bp)  

ITS 1  

(175 - 226 bp)  

ITS 2  

(0 - 43 bp)  

(-)

**B) Clade II**

(+)

16S  

(1,469 - 1,508 bp)  

23S  

(2,914 - 2,941 bp)  

5S  

(104 - 110 bp)  

Distance 1  

(131,991 - 231,797 bp)  

ITS 2  

(0 - 61 bp)  

(-)

**C) Clade II (M. parvum Indiana)**

(+)

23S  

(2,945 bp)  

5S  

(110 bp)  

Distance 1  

(70,713 bp)  

16S  

(1,508 bp)  

ITS 2  

(49 bp)  

(-)

**D) Clade II (M. ovis Michigan)**

(+)

5S  

(110 bp)  

23S  

(2,937 bp)  

16S (1)  

(1,479 bp)  

16S (2)  

(1,493 bp)  

ITS 2  

(65 bp)  

Distance 1  

(461,325 bp)  

Distance 2  

(17,077 bp)  

(-)
Fig 3
Fig4