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

Rosa Estela Quiroz-Castañeda^{1*}, Hugo Aguilar-Díaz², Diana Laura Flores-García³, Fernando Martínez-Ocampo⁴, 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>), 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 QCAST (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/2XjTTey>) [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_QKVO000000000.1; and LWUJ000000000.1, respectively). Transfer (tRNA) RNA genes was carried out using ARAGORN (v1.2.38)

(<https://bit.ly/3k1R2QT>) server [24]. The sequence and length of 16S and 23S rRNA genes was obtained from the RNAmmer (v1.2) (<https://bit.ly/3glQKCj>) 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 (<https://bit.ly/314fOre>) using the nucleotide BLAST (Blastn) suite (<https://bit.ly/3k2Wkvs>) [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 [28] 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 (v1.4.4) program (<https://bit.ly/39ROMXV>). 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.

Organism	Assembly level	Length (bp)*	G+C content (%)*	CDS **	rRNAs [#]	tRNAs ^{##}
‘Ca. M. haemobos’ INIFAP01	18 contigs	935,638	30.46	1,180	3	31
<i>M. haemocanis</i> Illinois	Chromosome	919,992	35.33	1,234	3	31
<i>M. haemofelis</i> Langford 1	Chromosome	1,147,259	38.85	1,595	3	31
<i>M. haemofelis</i> Ohio2	Chromosome	1,155,937	38.81	1,650	3	31
‘Ca. M. haemolamae’ Purdue	Chromosome	756,845	39.27	1,045	3	33
‘Ca. M. haemominutum’ Birmingham 1	Chromosome	513,880	35.52	587	3	32
<i>M. ovis</i> Michigan	Chromosome	702,511	31.69	918	4	32
<i>M. parvum</i> Indiana	Chromosome	564,395	26.98	578	3	32
<i>M. suis</i> Illinois	Chromosome	742,431	31.08	914	3	32
<i>M. suis</i> KI3806	Chromosome	709,270	31.08	856	3	32
<i>M. wenyonii</i> INIFAP02	37 contigs	596,665	33.43	678	3	32
<i>M. wenyonii</i> Massachusetts	Chromosome	650,228	33.92	727	3	32

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 four 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, ‘*Ca. M. haemominutum*’ Birmingham 1, *M. suis* Illinois, *M. suis* KI3806, *M. wenyonii* 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 hemoplasma species are separated into different clades (Fig 2). The clade 1 (blue lines) contains two sub-clades: 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.

Table 2. Prediction of antigenicity of proteins of *Ca. M. haemobos* and *M. wenyonii*.

Candidatus Mycoplasma haemobos INIFAP01

Classification	Prediction score as antigen (Vaxijen)
----------------	---------------------------------------

RAST Category: Virulence, disease, and defense

DNA gyrase subunit B	0.5352 (Antigen)
DNA gyrase subunit A	0.4373 (Antigen)
SSU ribosomal protein S7p	0.518 (Antigen)
Translation Elongation factor G	0.5399 (Antigen)
Translation Elongation factor Thermo unstable (Tu)	0.4268 (Antigen)
SSU ribosomal protein S12p	0.7537 (Antigen)
DNA-directed RNA polymerase beta subunit	0.3781 (Antigen)
DNA-directed RNA polymerase	0.4488 (Antigen)

RAST Category: Division and Cell Cycle

ProteinTsaD/Kae1/Qri7	0.3846 (Non Antigen)
RNA polymerase sigma factor RpoD	0.3857 (Non Antigen)
DNA primase	0.3009 (Non Antigen)

RAST Category: Fatty acids, Lipids and Isoprenoids

Cardiolipin synthase	0.3463 (Non Antigen)
----------------------	----------------------

RAST Category: Stress Response

Manganase superoxide dismutase	0.3487 (Non Antigen)
--------------------------------	----------------------

Mycoplasma wenyonii INIFAP02

RAST Category: Virulence, disease, and defense

Ribosomal protein SSU S7p	0.5435 (Antigen)
Translation Elongation factor G	0.5650 (Antigen)
Translation Elongation factor Thermo unstable (Tu)	0.4359 (Antigen)
Ribosomal protein SSU S12p	0.7774 (Antigen)
Ribosomal protein LSU L35p	0.5491 (Antigen)

Translation Initiation factor 3	0.4488 (Antigen)
Ribosomal protein LSU L20p	0.3668 (Antigen)

RAST Category: Division and Cell Cycle

Protein TsaD/Kae1/Qri7	0.3954 (Non Antigen)
RNA polymerase sigma factor RpoD	0.3979 (Non Antigen)
DNA primase	0.3929 (Non Antigen)

RAST Category: Fatty acids, lipids and isoprenoids

Cardiolipin synthase	0.3305 (Non Antigen)
----------------------	----------------------

The collection of B-cell epitopes (linear antigens) predicted with Predicting Antigenic Peptides tool and BCEPred server is shown in Table 3.

Table 3.B-Cell epitopes of *Ca. M. haemobos* and *M. wenyonii* predicted by immunoinformatics

<i>Ca. M. haemobos</i> INIFAP01		<i>M. wenyonii</i> INIFAP02	
DNA gyrase subunit B		Ribosomal protein SSU S7p	
Position (aa)	Epitope sequence	Position (aa)	Epitope sequence
23-33	SSIRVLEGLEA	8-16	LARRIVYNA
48-59	KGLHHLIWEVLD	35-43	AIRNVAPSI
77-87	LKKGHVISVSD	54-63	NYQVPVESSK
131-138	GVGSTCVN	67-76	EALALRWLIK
140-150	LSSFLEVNVR	Translation Elongation factor G	
216-225	GKKFVVFVNEI	Position (aa)	Epitope sequence
251-261	SIHDVIFIHSE	26-33	ERILYYTG
289-297	SSIVHSFCN	70-77	WKGVLNL
313-321	DGLLSCIRE	122-132	NKYKVPRIIFC
392-401	QRKVILQRVD	154-161	NIKFSPIQ
442-451	FSELYVVEGD	212-220	LLNEVLVYD
546-554	LLEHGYVYI	237-244	EIKMCIRK
565-574	NKEVVYLFDD	Translation Elongation factor Thermo unstable (Tu)	
DNA gyrase subunit A		Position (aa)	Epitope sequence
97-105	SIYDALVRM	167-178	NTPVIRGSALKA
107-116	QDFSLRYPLI	Ribosomal protein SSU S12p	
139-148	RLSKLGKYFL		

SSU ribosomal protein S7p		Position (aa)	Epitope sequence
Position (aa)	Epitope sequence	10-18	LRKFAKVRV
83-91	SNYQVPVEA	20-29	NNHEVLAYIP
97-106	ETLSLRWLIN	35-42	LQEHHVVM
116-125	MVEALAHEII	Ribosomal protein LSU L35p	
Translation Elongation factor G		Position (aa)	Epitope sequence
Position (aa)	Epitope sequence	15-24	LSKRVVVLGS
82-96	GHVDFTVEVERSLRV	59-67	QYKITAHL
142-151	YFKSVQSLRE	Translation Initiation factor 3	
143-149	YFKSVQSL	Position (aa)	Epitope sequence
153-162	LKVNAVLIQL	8-18	RQLDLLLVNNK
170-178	FIGIIDLIT	20-29	NPPVVKLLNF
211-219	ELLDEVLT	84-94	EKVQLIIRTPY
235-244	VEEIKHCIRI	Ribosomal protein LSU L20p	
272-282	VIDYLPSPIDL	Position (aa)	Epitope sequence
387-397	VKGNQVVLESM	87-94	FVLNRKVL
420-428	SMVLSRLSE	107-114	KIVDTSVK
451-459	ELHLEILID		
518-525	EFVDKIVG		
Translation Elongation factor Thermo unstable (Tu)			
Position (aa)	Epitope sequence		
97-107	AQIDAAILVVS		
SSU ribosomal protein S12p			
Position (aa)	Epitope sequence		
20-27	KSQALLKS		
31-39	LDKKVTFQS		
42-49	FKRGVCTR		
61-68	ALRKYAKV		
72-81	NNYEVLAYIP		
87-97	LQEHHVVMVRG		
DNA-directed RNA polymerase beta subunit			
Position (aa)	Epitope sequence		
61-71	NGLFCESIFGP		
111-120	HIELACPVAH		
168-177	IFSELEIIDV		
DNA-directed RNA polymerase			
Position (aa)	Epitope sequence		
7-17	RVNSFSPIVNR		
32-42	DLKGVQLGAYT		

114-124	SVENVFLGNLP		
140-147	FIISQIVR		
205-214	LIQFPLTTLL		
320-327	EIIKVTLE		
468-477	IEELLIYRTH		
532-540	LKLVDELLQ		
587-597	KPFQIVVKEFF		
677-685	FFVTPYYRV		
747-754	CHQLVSVS		
756-763	SLIPFLEH		
849-859	CKNQYPLVRPK		
901-908	IILSSRLI		
966-975	DILVGKVSPL		
1096-1103	QVLETHLG		
1340-1347	CLKINVQY		

Discussion

The hemoplasmas have underwent 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

Davidson BL, Allen ED, Kozarsky KF, Wilson JM, and Roessler BJ (1993). A model system for *in vivo* gene transfer into the central nervous system using an adenoviral vector. Nat Genet 3, 219–223.

References

1. Messick JB. Hemotrophic mycoplasmas (hemoplasmas): a review and new insights into pathogenic potential. Vet Clin Pathol [Internet]. 2004 Mar [cited 2017 Mar 2];33(1):2–13. Available from: <http://doi.wiley.com/10.1111/j.1939-165X.2004.tb00342.x>
2. Guimaraes AMS, Santos AP, do Nascimento NC, Timenetsky J, Messick JB. Comparative genomics and phylogenomics of hemotrophic mycoplasmas. Balish MF, editor. PLoS One [Internet]. 2014 Mar 18;9(3):e91445. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3958358/>
3. Rikihisa Y, Kawahara M, Wen B, Kociba G, Fuerst P, Kawamori F, et al. Western immunoblot analysis of *Haemobartonella muris* and comparison of 16S rRNA gene sequences of *H. muris*, *H. felis*, and *Eperythrozoon suis*. J Clin Microbiol. 1997 Apr;35(4):823–9.
4. Neimark H, Johansson KE, Rikihisa Y, Tully JG. Proposal to transfer some members of the genera *Haemobartonella* and *Eperythrozoon* to the genus *Mycoplasma* with descriptions of ‘*Candidatus Mycoplasma haemofelis*’, ‘*Candidatus Mycoplasma haemomuris*’, ‘*Candidatus Mycoplasma haemosui*. Int J Syst Evol Microbiol [Internet]. 2001;51(3):891–9. Available from: <http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-51-3-891>
5. dos Santos AP, Guimaraes AMS, do Nascimento NC, SanMiguel PJ, Messick JB. Complete genome sequence of *Mycoplasma wenyonii* strain Massachusetts. J Bacteriol.

2012 Oct;194(19):5458–9.

6. Martínez-Ocampo F, Rodríguez-Camarillo SD, Amaro-Estrada I, Quiroz-Castañeda RE. Draft genome sequence of “Candidatus Mycoplasma haemobos,” a hemotropic mycoplasma identified in cattle in Mexico. *Genome Announc* [Internet]. 2016 Aug 25;4(4):1–2. Available from: <http://genomea.asm.org/content/4/4/e00656-16.abstract>
7. Quiroz-Castañeda RE, Martínez-Ocampo F, Dantán-González E. Draft Genome Sequence of *Mycoplasma wenyonii*, a Second Hemotropic *Mycoplasma* Species Identified in Mexican Bovine Cattle. *Microbiol Resour Announc*. 2018;7(9):1–2.
8. Hoelzle K, Hofmann-Lehmann R, Hoelzle LE. ‘Candidatus *Mycoplasma haemobos*’, a new bovine haemotrophic *Mycoplasma* species? *Vet Microbiol*. 2010;144(3–4):525–6.
9. Tagawa M, Matsumoto K, Inokuma H. Molecular detection of *Mycoplasma wenyonii* and ‘Candidatus *Mycoplasma haemobos*’ in cattle in Hokkaido, Japan. *Vet Microbiol*. 2008;132(1–2):177–80.
10. Niethammer FM, Ade J, Hoelzle LE, Schade B. Hemotropic mycoplasma in Simmental cattle in Bavaria: Prevalence, blood parameters, and transplacental transmission of ‘Candidatus *Mycoplasma haemobos*’ and *Mycoplasma wenyonii*. *Acta Vet Scand*. 2018;60(1):1–8.
11. Genova SG, Streeter RN, Velguth KE, Snider TA, Kocan KM, Simpson KM. Severe anemia associated with *Mycoplasma wenyonii* infection in a mature cow. *Can Vet J* [Internet]. 2011 Sep;52(9):1018–21. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3157061/>
12. Ade J, Niethammer F, Schade B, Schilling T, Hoelzle K, Hoelzle LE. Quantitative analysis of *Mycoplasma wenyonii* and ‘Candidatus *Mycoplasma haemobos*’ infections in

- cattle using novel gapN-based realtime PCR assays. *Vet Microbiol.* 2018;220(December 2017):1–6.
13. Hoelzle K, Winkler M, Kramer MM, Wittenbrink MM, Dieckmann SM, Hoelzle LE. Detection of *Candidatus Mycoplasma haemobos* in cattle with anaemia. *Vet J* [Internet]. 2011 Mar;187(3):408–10. Available from: <http://www.sciencedirect.com/science/article/pii/S1090023310000328>
14. Tagawa M, Yamakawa K, Aoki T, Matsumoto K, Ishii M, Inokuma H. Effect of Chronic Hemoplasma Infection on Cattle Productivity. *J Vet Med Sci* [Internet]. 2013 Oct 15;75(10):1271–5. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3942926/>
15. McFadden A, Ha HJ, Donald JJ, Bueno IM, van Andel M, Thompson JC, et al. Investigation of bovine haemoplasmas and their association with anaemia in New Zealand cattle. *N Z Vet J.* 2016 Jan;64(1):65–8.
16. Fujihara Y, Sasaoka F, Suzuki J, Watanabe Y, Fujihara M, OOSHITA K, et al. Prevalence of Hemoplasma Infection among Cattle in the Western Part of Japan. *J Vet Med Sci.* 2011;73(12):1653–5.
17. Girotto A, Zangirolamo AF, Bogado ALG, Souza ASLE, da Silva GCF, Garcia JL, et al. Molecular detection and occurrence of ‘*Candidatus Mycoplasma haemobos*’ in dairy cattle of Southern Brazil. *Rev Bras Parasitol Vet.* 2012;21(3):342–4.
18. Tagawa M, Ybanez AP, Matsumoto K, Yokoyama N, Inokuma H. Prevalence and Risk Factor Analysis of Bovine Hemoplasma Infection by Direct PCR in Eastern Hokkaido, Japan. *J Vet Med Sci.* 2012;74(9):1171–6.
19. Parvizpour S, Pourseif MM, Razmara J, Rafi MA, Omidi Y. Epitope-based vaccine

- design: a comprehensive overview of bioinformatics approaches. Drug Discov Today
[Internet]. 2020; Available from: <https://doi.org/10.1016/j.drudis.2020.03.006>
20. Ahmad TA, Eweida AE, Sheweita SA. B-cell epitope mapping for the design of vaccines and effective diagnostics. Trials Vaccinol [Internet]. 2016;5:71–83. Available from: <http://www.sciencedirect.com/science/article/pii/S1879437816300079>
21. Ranjbar MH, Ebrahimi MM, Shahsavandi MM, Farhadi S, Mirjalili T, Tebianian A, et al. Novel applications of immuno-bioinformatics in vaccine and bio-product developments at research institutes. Arch Razi Inst. 2019;74(3):219–33.
22. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics [Internet]. 2013 Apr 15;29(8):1072–5. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624806/>
23. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008 Feb;9:75.
24. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004 Jan;32(1):11–6.
25. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35(9):3100–8.
26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct;215(3):403–10.
27. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.
28. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics

and high-performance computing Europe PMC Funders Group. Nat Methods.

2012;9(8):772.

29. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010 May;59(3):307–21.
30. Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol. 2013 Dec;79(24):7696–701.
31. Martin Shumway CA and SLSSKAPALDMS, Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5(2):12.
32. Connors J, Krzywinski M, Schein J, Gascoyne R, Horsman D, Jones SJ, et al. Circos : An information aesthetic for comparative genomics. Genome Res. 2009;19(604):1639–45.
33. Peters IR, Helps CR, McAuliffe L, Neimark H, Lappin MR, Gruffydd-Jones TJ, et al. RNase P RNA gene (rnpB) phylogeny of hemoplasmas and other mycoplasma species. J Clin Microbiol [Internet]. 2008 May 1;46(5):1873–7. Available from: <http://jcm.asm.org/content/46/5/1873.abstract>
34. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 2007;57(1):81–91.
35. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 2009;106(45):19126–31.
36. Figueras MJ, Beaz-Hidalgo R, Hossain MJ, Liles MR. Taxonomic affiliation of new

genomes should be verified using average nucleotide identity and multilocus phylogenetic analysis. *Genome Announc.* 2014;2(6):6–7.

37. Michel-Todó L, Reche PA, Bigey P, Pinazo MJ, Gascón J, Alonso-Padilla J. In silico Design of an Epitope-Based Vaccine Ensemble for Chagas Disease. *Front Immunol.* 2019;10(November):1–16.

38. Hajissa K, Zakaria R, Suppian R, Mohamed Z. Epitope-based vaccine as a universal vaccination strategy against *Toxoplasma gondii* infection: A mini-review. *J Adv Vet Anim Res* [Internet]. 2019 Mar 24;6(2):174–82. Available from: <https://pubmed.ncbi.nlm.nih.gov/31453188>

39. Manzano-Román R, Díaz-Martín V, Oleaga A, Pérez-Sánchez R. Identification of protective linear B-cell epitopes on the subolesin/akirin orthologues of *Ornithodoros* spp. soft ticks. *Vaccine.* 2015 Feb;33(8):1046–55.

40. de la Fuente J, Moreno-Cid JA, Canales M, Villar M, de la Lastra JMP, Kocan KM, et al. Targeting arthropod subolesin/akirin for the development of a universal vaccine for control of vector infestations and pathogen transmission. *Vet Parasitol.* 2011 Sep;181(1):17–22.

41. Liu X, Marrakchi M, Xu D, Dong H, Andreescu S. Biosensors based on modularly designed synthetic peptides for recognition, detection and live/dead differentiation of pathogenic bacteria. *Biosens Bioelectron* [Internet]. 2016;80:9–16. Available from: <http://www.sciencedirect.com/science/article/pii/S0956566316300410>

42. Quiroz-Castañeda R, Tapia-Uriza TR, Mujica-Valencia C, Rodríguez-Camarillo SD, Preciado-De la Torre JF, Amaro-Estrada I, et al. Synthetic peptides-based indirect ELISA for the diagnosis of bovine anaplasmosis. *Int J Appl Res Vet Med.* 2019;17(2):65–70.

43. Molina AIB, Bayúgar RC, Gutiérrez-Pabello JA, López ATT, De La Torre JFP, Camarillo SDR. Immunolocalization of VirB11 protein in the *Anaplasma marginale* outer membrane and its reaction with bovine immune sera. *Rev Mex Ciencias Pecu.* 2018;9(4):769–91.
44. Wang CY, Chang TY, Walfield AM, Ye J, Shen M, Chen SP, et al. Effective synthetic peptide vaccine for foot-and-mouth disease in swine. *Vaccine* [Internet]. 2002;20(19):2603–10. Available from: <http://www.sciencedirect.com/science/article/pii/S0264410X02001482>
45. Droppa-Almeida D, Franceschi E, Padilha FF. Immune-Informatic Analysis and Design of Peptide Vaccine From Multi-epitopes Against *Corynebacterium pseudotuberculosis*. *Bioinform Biol Insights* [Internet]. 2018;12:1177932218755337. Available from: <https://europepmc.org/articles/PMC5954444>
46. Bashir S, Abd-elrahman KA, Hassan MA, Almofti YA. Multi Epitope Based Peptide Vaccine against Mareks Disease Virus Serotype 1 Glycoprotein H and B. *Am J Microbiol Res.* 2018;6:124–39.

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 (ANIb) values of alignment coverage of 12 hemoplasma genomes.

S3_Table. BLAST-based average nucleotide identity (ANIb) values of identity of 12 hemoplasma genomes.

A) Clade I

B) Clade II

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

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

Fig1

Fig3

***Mycoplasma wenyonii* Massachusetts**
650,228 bp

Fig2