

SCIENTIFIC REPORTS



OPEN

Changes in pathogenicity and immunogenicity of *Mycoplasma mycoides* subsp. *mycoides* strains revealed by comparative genomics analysis

Received: 19 June 2015
Accepted: 04 December 2015
Published: 11 January 2016

Yuan Li^{1,*}, Yang Wang^{1,2,*}, Rui Wang³, Yongqiang Zhu³, Suli Liu¹, Qi Wang⁴, Jiari Shao⁵, Ying Chen⁶, Liping Gao¹, Changping Zhou¹, Henggui Liu¹, Xiumei Wang¹, Huajun Zheng^{3,7} & Jiuqing Xin¹

Mycoplasma mycoides subsp. *mycoides* is the causative agent of contagious bovine pleuropneumonia. A pathogenic strain BEN-1 was isolated from bovine lung and underwent continuous passages in rabbits for 468 generations. During this process, the strain's strong virulence became weak and, gradually, it lost the ability to confer protective immunity in cattle but developed virulence in rabbits. In order to gain insight into the mechanisms behind the reduction in virulence and the loss of immunogenicity, we sequenced five representative strains of the BEN series, including the original strain (BEN-1), the strain generation that first acquired virulence in rabbits (BEN-50), the two vaccine strain generations (BEN-181 and BEN-326), and the strain generation showing the greatest loss of immunogenicity (BEN-468). The gene mutation rate in the four different propagation stages varied greatly, and over half of variations observed in each generation were removed during the propagation process. However, the variation maintained in the BEN-468 generation might contribute to its changes in virulence and immunogenicity. We thus identified 18 genes associated with host adaptation, six genes contributing to virulence in cattle, and 35 genes participating in conferring immunity in cattle. These findings might help us optimize the vaccine to obtain more effective immunization results.

Contagious bovine pleuropneumonia (CBPP) is caused by *Mycoplasma mycoides* subsp. *mycoides* (Mmm). Mmm is included in the classical *M. mycoides* cluster: *Mycoplasma capricolum* subsp. *capripneumoniae* (Mccp), *Mycoplasma capricolum* subsp. *capricolum* (Mcc), *Mycoplasma mycoides* subsp. *capri* (Mmc), *Mycoplasma mycoides* subsp. *mycoides* (Mmm), and *Mycoplasma leachii* (Ml). Infection results in severe pathological changes in the lungs and respiratory tract, leading to significant mortality and morbidity and greatly reducing animal welfare and livestock production¹⁻³.

CBPP is the only bacterial disease included on List A of communicable animal diseases issued by the Office International Des Epizooties (OIE). In the late 19th and early 20th centuries, CBPP was transmitted from Europe to Australia and then to other countries, including New Zealand, India, China, Korea, and Japan⁴. Accordingly, veterinary authorities have spent a great amount of effort trying to control and eradicate the disease by improving

¹National Contagious Bovine Pleuropneumonia Reference Laboratory, Division of Bacterial Diseases, State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, CAAS, Harbin, China. ²Key Laboratory of Fermentation Engineering (Ministry of Education), Hubei Provincial Cooperative Innovation Center of Industrial Fermentation, College of Bioengineering, Hubei University of Technology, Wuhan, China. ³Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, China. ⁴College of Resources and Environmental, Northeast Agricultural University, Harbin, China. ⁵College of Animal Science and Technology, Jilin Agricultural University, Changchun, China. ⁶College of Veterinary Medicine, Northeast Agricultural University, Harbin, China. ⁷Laboratory of Medical Foods, Shanghai Institute of Planned Parenthood Research, Shanghai, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to H.Z. (email: zhenghj@chgc.sh.cn) or J.X. (email: xinjiuqing2001@126.com)

diagnostic techniques and developing vaccines. CBPP was eradicated from the USA and Britain in the 19th century without any knowledge of the causative agent but it persisted in other European countries, particularly those in Southern Europe, until the end of the last century⁵. In Europe, the last reported outbreak was in northern Portugal in 1999. Although a number of countries in sub-Saharan Africa including South Africa, Botswana, and Zimbabwe became free of CBPP in the 1930s, most remained infected, and others like Zambia and Tanzania became re-infected in the second half of the 20th century⁶. To date, Australia, India, Switzerland, Botswana, Portugal, the USA, and China have been officially certified as CBPP-free (<http://www.oie.int/en/animal-health-in-the-world/official-disease-status/cbpp/list-cbbp-free-members/>).

The T1/44 vaccine approved by the OIE was originally developed from a virulent Mmm strain isolated in Tanzania, which had been passaged 44 times in chicken embryos in an effort to reduce its virulence. A vaccine was then prepared by culturing the strain in Gourlay's medium⁷. Although the T1/44 vaccine has been administered for nearly 60 years, the protection it provides in cattle is highly variable and transient and it has therefore been unable to control the spread of CBPP in Africa⁸.

In the early 1950s, China embarked on measures to control CBPP, i.e., vaccination using low-virulence Mmm strains passaged in culture medium, surveillance and quarantine of infected cattle, and slaughtering of CBPP-positive cattle and those suspected to be diseased⁹. Although vaccination showed some success, CBPP was far from controlled due to adverse vaccine reactions and instability of the vaccine, which had a short shelf life because of the shortage of lyophilization equipment¹⁰. In 1956, the virulent Mmm strain (BEN-1) was isolated from the lung of a CBPP-infected cow in China¹¹. To attenuate the virulence of the BEN-1 strain and preserve its protective ability, the isolate was re-isolated after inoculation into the testicles of rabbits and into the rabbit thorax. As a result, after the subsequent isolates were continuously passaged 468 times in rabbits, the strain's virulence in cattle was weakened (from BEN-1 to BEN-468), and it gradually lost the ability to confer protective immunity in cattle (from BEN-327 to BEN-468), but developed virulence in rabbits (from BEN-50 to BEN-468). The strain showed reduced virulence but retained its ability to confer protective immunity in cattle (from BEN-181 to BEN-326)¹². By using the BEN-181 vaccine, China has eradicated CBPP and has been officially certified as a CBPP-free country by OIE. The molecular mechanisms leading to the attenuation of the Mmm strain BEN-1, the adaptation to rabbits, and the loss of immunogenicity remain unknown. Some factors related to the pathogenicity of Mmm have been demonstrated in recent studies⁴. However, their functions and influences on the Mmm life cycle and on its pathogenicity require further definitive investigation.

In this study, we sequenced the whole genomes of five representative strains of the BEN series, which were passaged and maintained in our lab. This includes the original virulent strain (BEN-1), the strain that initially gained virulence in rabbits (BEN-50), the vaccine strains (BEN-181 and BEN-326), and the strain showing the greatest loss in the capacity to confer protective immunity (BEN-468, the key strain during continuous passaging). Furthermore, we compared the entire genomes of Mmm BEN-1, BEN-50, BEN-181, BEN-326, and BEN-468. The results indicate that there are 18 genes responsible as the key factors in the adaptation to rabbits, six genes harboring mutations associated with the weakened virulence in cattle, and 35 genes that were lost in the BEN-468 strain generation that contribute to conferring protective immunity in cattle. The identified genes associated with virulence and immunogenicity will deepen our understanding of the vaccine and help us optimize the vaccine through genetic engineering.

Results and Discussion

Genome features of the BEN series. *General genomic features.* The Mmm BEN-1 genome contains a single circular chromosome of 1,145,921 bp with a GC content of 23.9% (Fig. 1 and Table 1). We identified 1,043 genes in the genome with an average length of 939 bp and a mean GC content of 24.2% that occupied 85.5% of the genome.

Among the 1,043 genes, 629 genes (60.3%) could be assigned into Clusters of Orthologous Groups (COG) comprising 20 functional categories (Table 2). Biological functions were assigned to 695 genes (66.6%), and 335 genes were homologous with conserved proteins of unknown function in other organisms. Only 13 genes were unique to BEN-1.

We identified 114 insertion sequence (IS) elements in BEN-1 (Supplementary Table 1), occupying 10% of the genome and reflecting high genome plasticity. IS elements could also explain the variation in genome size of the five strains, as BEN-181 exhibited the largest number of IS elements (143) while BEN-326 had the fewest (101; Supplementary Fig. 1). Correspondingly, only three small clustered regularly interspaced short palindromic repeat (CRISPR) loci having just three direct repeats (DRs) were found in BEN-1 (Supplementary Table 2), which was postulated to limit the intracellular spread of mobile genetic elements such as IS elements and transposons¹³. Although integrative conjugative elements (ICE) had been revealed in the genome of *Mycoplasma mycoides* subsp. *capri* 95010³, no ICEs were observed in the Mmm BEN genomes, indicating that the structure of the BEN genome is relatively conserved.

Virulence genes in BEN-1. Pathogen virulence is usually associated with exposed cellular proteins. Forty-seven secreted proteins, 39 lipoproteins, and 170 transmembrane proteins were identified in the BEN-1 genome, occupying 25% of the proteome (Supplementary Fig. 2). Lipoproteins of Mmm have previously been characterized as strong major antigens, e.g., LppB (*Ben1_0562*)¹⁴, LppQ (*Ben1_1077*)^{15,16}, and variable surface protein Vmm (*Ben1_0422*)¹⁷. In addition to these, we identified LppA (*Ben1_0014*), LppC (*Ben1_1059*), LppD (*Ben1_0685*), and seven variable surface protein (VSP) genes in the genome (Supplementary Table 3). The seven VSP genes formed a cluster, and four of them (*Ben1_0851–Ben1_0854*) shared 100% identity with each other and 96% identity with *Ben1_0855*, hinting that they originated from a recent gene duplication event (Supplementary Fig. 3). It is worth noting that the lipoprotein genes were usually located adjacent to IS elements, such as *lppA*, *lppB*, *lppD*, and the VSP gene cluster, which increases the possibility of virulence gene rearrangement or duplication.

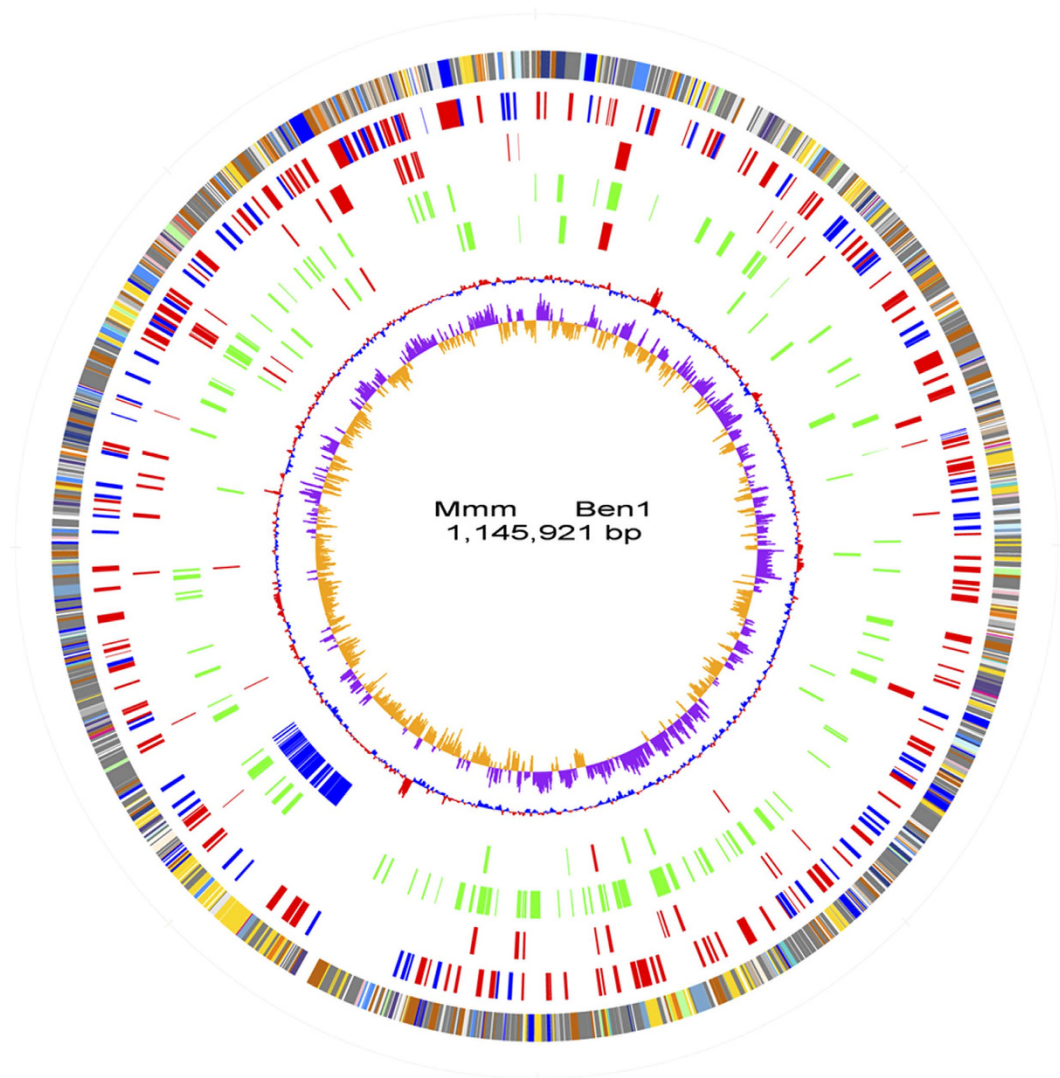


Figure 1. Genome atlas of MmmBEN-1 and evolution variations. Each concentric circle represents genomic data for MmmBEN-1 and its comparison with BEN-50–BEN-468. The outer circle illustrates predicted coding sequences in the BEN-1 genome, colored by functional categories according to COG classification. The second circle represents locations of virulence genes (red) and IS elements (green). The third circle displays genes associated with virulence in rabbits (red). The fourth circle shows genes with virulence in cattle (green). The fifth circle presents genes contributing to protective immunity to cattle, with green representing BEN-468 genes under positive selection, red genes indicating fixed genes under positive selection from BEN-50 to BEN-468, and blue represents the 50 kb region of lost genes in BEN-468. The sixth and seventh circles (innermost) separately represent GC content and GC skew $(G-C)/(G+C)$, calculated using a 1-kb window.

Genome	Size (bp)	GC Content	Gene Number	Average Gene Length (bp)	Coding Region (bp)	Coding Ratio
BEN-1	1,145,921	23.90%	1,043	939	979,530	85.48%
BEN-50	1,161,055	24%	1,049	950	996,537	85.83%
BEN-181	1,149,615	24%	1,095	868	950,745	82.70%
BEN-326	1,109,375	24%	1,019	921	938,883	84.63%
BEN-468	1,091,780	23.90%	1,020	917	935,235	85.66%

Table 1. Genome characters of five Mmm genomes.

In addition, 194 BEN-1 proteins matched known virulence genes of different organisms (Supplementary Table 4). Among them, 174 (89.7%) were cytoplasmic proteins that exhibited a significant enrichment in COG class of “Cell envelope biogenesis, outer membrane” and “Carbohydrate transport and metabolism” (Supplementary

COG	BEN-1	BEN-50	BEN-181	BEN-326	BEN-468
Amino acid transport and metabolism	40	41	42	38	39
Carbohydrate transport and metabolism	54	54	56	56	51
Cell division and chromosome partitioning	10	10	9	9	10
Cell envelope biogenesis, outer membrane	27	25	25	17	26
Coenzyme metabolism	15	15	15	15	15
Defense mechanisms	10	10	10	10	10
DNA replication, recombination, and repair	121	129	124	118	120
Energy production and conversion	27	27	29	27	27
Function unknown	33	31	33	31	28
General function prediction only	59	61	60	59	58
Inorganic ion transport and metabolism	22	21	24	23	20
Intracellular trafficking and secretion	7	7	8	7	7
Lipid metabolism	9	9	9	9	9
Nucleotide transport and metabolism	31	31	32	31	31
Posttranslational modification, protein turnover, chaperones	18	19	19	18	18
Secondary metabolites biosynthesis, transport, and catabolism	2	2	2	2	2
Signal transduction mechanisms	5	5	5	5	5
Transcription	29	28	28	29	28
Translation, ribosomal structure and biogenesis	110	110	109	109	110
Not in COG	414	414	456	406	406
Total	1043	1049	1095	1019	1020

Table 2. COG Distribution in five Mmm genomes.

Table 5). This indicated that cytoplasmic proteins were important for the pathogenesis of Mmm during strain growth. Metabolic pathway enzymes have been characterized as primary virulence factors, for example, the alpha-glycerophosphate oxidase (GlpO) of Mmm was shown to release H₂O₂ from glycerol and cause host cell injury¹⁸. In the BEN-1 genome, *Ben1_0284* encodes GlpO and, together with an ABC transport system for glycerol (*Ben1_0558–Ben1_0561*) and a glycerol uptake facilitator (*Ben1_0282*), might contribute to its pathogenicity.

The T cell component of the immune system is essential for host recognition and control of pathogens and, therefore, depends on the binding of specific antigen epitopes¹⁹. Thus, mutations in T cell epitopes affect their capacity for infection and immune protection²⁰. We found 40 known T cell epitopes in 15 BEN-1 proteins (Supplementary Table 6), and eight proteins matched known virulence genes (Supplementary Table 4). These epitopes remained unchanged in the whole propagation process.

To further investigate variations in protein family (PF) numbers among Mmm species, a comprehensive comparative analysis was performed. A total of 736 PFs were identified among all of the Mmm species (Supplementary Table 7), including 679 core families common to all five strains, 51 dispensable families between any two species, and six unique families found in only one species. We then searched for gene families that contained expansion or shrinkage events relative to BEN-1. Within BEN-1, 725 domains were revealed, and of these, 139 domains showed an increase or loss of family members during the propagation from BEN-50 to BEN-468 (Supplementary Table 7), with PF13346 (ABC-2 family transporter protein) completely lost in the other continually propagated strains (Supplementary Table 8). In BEN-181, a total of 31 domains were lost. Five domains were lost from BEN-326, and eight were lost from BEN-468. Meanwhile, two domains were gained in BEN-50 and stably existed throughout the propagation process (PF02367P-loop hydrolase and PF10298-Sporulation Regulator WhiA N terminal; Supplementary Table 9).

Pan-genomic analysis of the BEN series. From BEN-1 to BEN-468, 679 genes remained unchanged (Supplementary Table 10), constituting the core genes of the Ben series. Half of the virulence genes were included in the core genes, such as *lppA*, *lppQ*, *glpO*, and *vmm*, indicating that these genes had no contribution to the weakened virulence and immunogenicity of the BEN series in cattle. In other words, these genes might be essential to maintain the fundamental virulence of Mmm but have no role in host adaptation.

After excluding IS elements, we found that only 55 genes (5.3%) in BEN-50 varied, with 15 genes affected by indels and 40 genes by single nucleotide polymorphisms (SNPs). These varied genes showed enrichment of the transferase activity of glycosyl groups or phosphorus-containing groups (Supplementary Table 11) and might explain the virulence in rabbits acquired by BEN-50. Among them, *Ben1_0988* encodes HPr kinase/phosphorylase, which phosphorylates/dephosphorylates the HPr protein of the bacterial phosphotransferase system in response to the nutrient status of the cell²¹. Phosphorylated HPr plays distinct roles in the regulation of virulence²². The Asn251Asp mutation in the C-terminal domain of *Ben1_0988* might affect the kinase activity.

In BEN-181, 112 genes were affected by indels and 48 genes affected by SNPs, comprising 15.3% of the total genes. These genes were enriched for ATPase activity and ribosome biogenesis processes. Mutation of genes

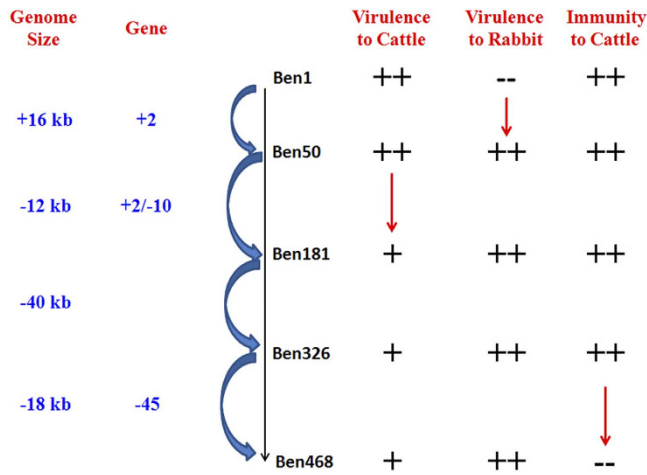


Figure 2. Virulence changes associated with genome variation. ‘+ +’ represents high virulence or immunity, ‘+’ represents virulence or immunity, ‘--’ represents no virulence or immunity. The number below ‘Gene’ indicates the number of genes gained or lost during the propagation process.

involved in ribosome biogenesis might change the strain’s growth rate within the host and thus facilitate adaptation to changing hosts. ATPase activity had roles in energy-driven reactions, coupled with ABC-type transporters²³, and even cytoadherence²⁴. This large proportion of genetic variations coincided with the weakening of BEN-181’s virulence in cattle.

In BEN-326, 92 genes were affected by indels, and 49 were affected by SNPs, with most them being similar to those found in the BEN-181 strain. However, enrichment for nucleotidyltransferase activity was observed. In BEN-468, 53 genes were affected by SNPs, and 56 genes were affected by indels, with 35 genes being completely lost from the genome, including *lppD*, which may have contributed to the loss of protective immunity in cattle. Correspondingly, enrichment for plasma membrane genes was observed, indicating that transmembrane genes were mutated or lost.

Comparison with *Mmm* PG1^T. The genome of *Mmm* type strain PG1^T is characterized by a single circular chromosome of 1,211,703 bp²⁵. PG1^T shares 99.88% identity with BEN-1 but is 65 kb larger (Supplementary Fig. 4). The genome size variation was mainly caused by an IS element as no significant functional gain or loss of genes was observed in either strain (Supplementary Fig. 5). Indeed, the 38/30 unique genes existing in BEN-1/PG1^T were almost all hypothetical proteins and evenly distributed across the genome.

Comparative genomics revealed the genetic basis of changes in virulence and immunogenicity. After series propagation in the rabbit, we found three distinct phenomenon changes (Fig. 2). One was the development of virulence in rabbit, from BEN-1 to BEN-50; the second is the weaker cattle virulence, from BEN-50 to BEN-181; the third is the loss of protective immunity in cattle, from BEN-326 to BEN-468.

Coincidentally, although no structural genomic changes were observed, the genomic sequences of these strains showed variance (Supplementary Fig. 6). The genome size of BEN-50 was the largest among the five representative genomes, suggesting that BEN-50 obtained external genes to adapt to the host environment of the rabbit. Comparing the genome of BEN-50 with that of BEN-1 might help us identify the genes associated with host adaptation.

The genome size of BEN-468 was the smallest, reflecting the gene loss during propagation. In comparison with BEN-326, the genes missing from BEN-468 may have contributed to conferring of protective immunity in cattle. The protein coding genes affected by these variations might explain the changes in virulence and immunogenicity.

Gain of Virulence in Rabbits. We hypothesized that during the propagation in rabbit, *Mmm* generated adaptive mutations that facilitated survival in the rabbit environment. To characterize genetic variations that evolved during serial propagation, we analyzed SNPs and indels occurring in BEN-50 relative to BEN-1.

A total of 155 SNPs and 231 indels occurred in BEN-50 (Table 3). Among them, 118 SNPs occurred in a coding DNA sequence (CDS) and caused non-synonymous variations in 49 genes (Supplementary Table 12), with 37 of the genes encoding transposase. Of the remaining 12 genes, five were known to be associated with virulence, including two polipoproteins and a ferric uptake regulator.

The majority of the indels were single-base insertions or deletions, and 97 of them caused frameshifts in 70 genes (Supplementary Table 13). Like the SNPs, 43 genes affected by indels encodes transposase, indicating that the IS elements had a high rate of mutation. Of the other functional genes affected by the frameshifts, *Ben1_0100*, encoding hemolysin A, is a known virulence factor.

Relative to BEN-1, only two new genes (*Ben50_0329* and *Ben50_0330*) encoding hypothetical proteins emerged. As the two novel genes cannot explain the virulence in rabbit, we then studied the genes under positive

	BEN-50 vs BEN-1	BEN-181 vs BEN-50	BEN-468 vs BEN-326
SNPs	155	262	205
Insertions	120	114	363
Deletions	101	64	226
Emerged Genes	2	2	0
Lost Genes	0	10	45
CDS affected by SNPs	49	40	37
CDS affected by Indels	70	93	151

Table 3. Main variation in the evolution to BEN-468.

selection during the evolution from BEN-1 to BEN-50. Our results showed that 19 genes were under positive selection ($Ka/Ks > 1$; Supplementary Table 14).

Taken together, 47 genes (excluding transposase genes) were mutated in the BEN-50 generation, including genes encoding variable surface proteins and hemolysin A (Supplementary Table 15). Only 10 genes were involved in the metabolic pathway, with three participating in amino sugar and nucleotide sugar metabolism ($p < 0.01$).

As the virulence in rabbits was retained throughout propagation to BEN-468, we focused on genes remaining mutations from BEN-50 to BEN-468, including nine genes affected by indels and nine genes affected by non-synonymous variation (Supplementary Table 10). In addition to hemolysin A, these mutated genes included a ferric uptake regulator and prolipoprotein. C-terminal truncation in hemolysin A in BEN-50 may be associated with the emergence of virulence in rabbits, as hemolysin-mediated hemolysis is a potential virulence factor among Mollicutes²⁶. In addition, ferric uptake regulator was key in the regulation of intracellular iron homeostasis and virulence^{27,28}, and the acquisition of the K138E mutation as a result of positive selection may be a contributing factor to the pathogenicity in rabbits.

Toll-like receptors (TLRs) have been shown to be a critical component in host innate immunity, which plays a key role in host defense against bacterial infection²⁹. Lipoproteins and/or lipopeptides from mycoplasmas have been shown to induce host innate immune responses, predominantly through the TLR2 signaling pathways^{30,31}. In rabbit (*Oryctolagus cuniculus*), TLR2 proteins share only 70% identity to cattle (*Bos taurus*). This suggests that escape from the rabbit innate immune system by Mmm from cattle would require several changes in the ligands of TLR2. The nine genes affected by indels, especially the prolipoprotein and three hypothetical proteins, may therefore be pathogen-associated molecular patterns (PAMPs) recognized by the host.

Essentially all the lipids of mycoplasmas are located in the cell membrane³². As a part of lipoprotein, the mycoplasma lipid moiety was the main component recognized by the innate immune system through Toll-like receptors (TLR) 2 and 6^{33,34}. So we further investigated whether continuous passage caused gene variation affected biosynthetic pathways of Mycoplasma lipids. Mmm genome did not encode proteins involved in fatty acid synthesis, elongation or degradation. But several genes involved in glycerolipid metabolism and glycerophospholipid metabolism was revealed, including a complete pathway synthesizing cardiolipin (CL) from glycerone (Supplementary Table 16). Among them, *Ben1_0446*, encoding glycerol-3-phosphate dehydrogenase, was affected by a frameshift from BEN-50 to BEN-468 (Supplementary Table 10). As *Ben1_0446* catalyzed the first step in cardiolipin synthesis, we speculated that its mutation together with three other varied genes (*Ben1_0382*, *Ben1_0401*, and *Ben1_0856*) in BEN-50 might affect the production of glycerophospholipid either in amount or structure and contribute to the acquired virulence to rabbit.

MHC-I binding prediction further revealed three mutations of BEN-50 proteins that alter the MHC binding affinity in rabbit (Supplementary Table 17), indicating an adaptation to rabbits.

The *LppB* gene *Ben1_0562* had a non-synonymous variation in BEN-50, but this mutation was not observed in the latter strains of the propagation (from BEN-181 to BEN-468). This indicated that the strain was subject to a rapid mutation rate during the first period of adaptation to rabbits (Fig. 3) but the mutations that would not help the strain adapt to a new host would not be maintained.

Attenuated Virulence of BEN-181 in Cattle. With further propagation in rabbit, the virulence of Mmm for cattle was attenuated in BEN-181. A total of 262 SNPs and 178 indels occurred in BEN-181 (Table 3). However, 55 of the SNPs in BEN-181 reverted to the base nucleotides of BEN-1, indicating that these SNPs were under strong purifying selection as the effect of the amino acid substitutions was not favorable for the survival of Mmm in the host (Supplementary Table 12). The remaining 207 SNPs caused non-synonymous variation in 40 genes (Supplementary Table 18), with half of them encoding transposase. These genes included a gene encoding a variable surface prolipoprotein that might be associated with virulence.

Of the 178 indels, 129 of them caused frameshift mutations in 93 genes (Supplementary Table 19), while only 19 of the genes encoded transposase. The functions of the remaining genes might be associated with virulence in cattle. The glycerol ABC transporter and Na^+ ABC transporter are two examples. Their import and export of nutrients and toxic substances across the membrane might indirectly cause virulence³⁵, and the functional damage caused by a frameshift may be the reason for the attenuation of virulence.

Relative to BEN-50, two additional new genes emerged, *Ben181_0274* and *Ben181_1034*, which encoded hypothetical proteins. Ten BEN-50 genes were lost in BEN-181, including *Ben50_0551*, which encodes glycerol

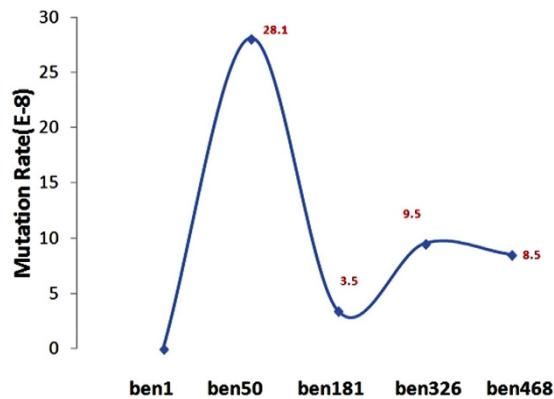


Figure 3. Mutation rate during the evolution from BEN-1 to BEN-468. The mutation rate was calculated based on the ds in four propagation stages. The number corresponding to each strain represents the mutation rate evolving to this generation.

ABC transporter permease (Supplementary Table 20). These lost genes might contribute to the attenuation of virulence seen in BEN-181.

Loss of protective immunity in cattle. In the process of evolving to BEN-468 from BEN-326, a large region of around 50kb in BEN-326 was lost. This region contained 45 genes (*Ben326_0674–Ben326_0721*), including seven transposase genes, four prolipoprotein genes, and two mycoplasma virulence signal regions (Supplementary Table 21). As BEN-468 lost the capacity to confer protective immunity in cattle, the lost region might contribute to the ability to convey immunity, which was supported by the fact that only two mycoplasma virulence signal region genes existed in Mmm. Lipoproteins, including LppD, are known as strong major antigens and play a crucial role in interactions with their hosts. It has been reported that the N-terminal peptides of lipoproteins contributed to the specificity of recognition by TLR2 heteromers³⁶. Thus the loss of lipoprotein encoding genes would affect the activation of early host responses after infection. Hence, the loss of four prolipoprotein genes of BEN-468 might contribute greatly to the lost capacity to confer immunity.

MHC-I binding prediction also found that more T-cell epitopes of the 45 genes were recognized by cattle than by rabbit (Supplementary Table 22). In addition to the large fragment loss, 205 SNPs and 589 indels were found in BEN-468, with 37 genes showing non-synonymous variation (Supplementary Table 23) and 151 genes being affected by frameshift mutations (Supplementary Table 24). All of the genes mutated in BEN-468 might be associated with its loss of ability to confer protective immunity in cattle.

Dynamic changes of positive selection genes. In the process of propagation, 679 genes (65.2%) remained unchanged, showing a stringent selection pressure and possible contribution to essential functions of the Mmm strain.

Estimation of mutation rate. The mutation rate of Mmm was calculated using the following formula: $\text{Age} = \text{ds} / (\text{rate} \times 2)$, where ds is the mean number of synonymous substitutions per site, calculated from the five genomes. After Jukes-Cantor correction, the four propagation stages showed a varying ds (Fig. 3). From BEN-1 to BEN-50, the ds = 0.000027578, corresponding to a mutation rate of 2.81E-07 per base per generation. From BEN-50 to BEN-181, the ds = 0.000009193, corresponding to a mutation rate of 3.51E-08. From BEN-181 to BEN-326, the ds = 0.000027579, corresponding to a mutation rate of 9.51E-08. From BEN-326 to BEN-468, the ds = 0.000027579, corresponding to a mutation rate of 8.51E-8.

We can conclude that during the first 50 generations following transfer from the natural host (cattle) to a new host (rabbit), the strain endured severe selection pressure, and more mutations were maintained to adapt to the new environment. A mutation rate of 2.81E-07 is much higher than the spontaneous mutation rate of 8.9E-11 in *E.coli*³⁷. It has been long known that mycoplasmas have a high mutation rate due to the lack of a mismatch repair system as well as 3' to 5' proofreading exonuclease activity from the single DNA polymerase³⁸, which allowed mutations to occur at an accelerated rate. Here, we calculated the detailed mutation rate and validated this postulation.

From BEN-50 to BEN-181, which corresponded to the reduction in cattle virulence, the mutation rate decreased to 3.51E-08. This might indicate that the Mmm strain had adapted to the rabbit host and a purifying selection was occurring with the aim of removing deleterious mutations. Beginning with BEN-181 and continuing to BEN-468, the mutation rate increased to 8.51~9.51E-08, a relatively stable mutation rate, throughout the final 280 generations.

Positive selection genes. Changing host environments from cattle to rabbits should create strong selection pressure for the strain. We therefore calculated the Ka/Ks among the five genomes and revealed 77 genes under positive selection ($Ka/Ks > 1$; Supplementary Tables 14 and 25). After 468 generations of propagation in rabbits, many mutations were finally removed by purifying selection. In BEN-50, 19 genes showed variation under

positive selection, and the variation in nine genes were maintained in the subsequent propagation until BEN-468 (Supplementary Table 25). From the functional analysis, we can see that these nine genes participated in DNA replication and repair, carbohydrate metabolism, and amino acid metabolism. We also found seven genes showing positive selection only in BEN-326 and BEN-468; these genes might be directly associated with the loss of protective immunity in cattle.

Conclusion

In this study, we sequenced the whole genomes of Mmm BEN-1, BEN-50, BEN-181, BEN-326, and BEN-468, which were the key strains during continuous passaging. Furthermore, we compared the entire genomes of Mmm BEN-1, BEN-50, BEN-181, BEN-326, and BEN-468. The results indicated that the gene mutation rate in the four different propagation stages varied greatly. Eighteen genes were identified as key factors in adapting to the rabbit host environment, six genes carried mutations associated with a reduction in virulence in cattle, and 35 genes were missing from BEN-468 that are thought to contribute to the ability to confer protective immunity in cattle. The success of the vaccine strain was the result of both weakened virulence and the retention of immunogenicity. Further experimental validation of these genes' function were needed to identify the key factors that have led to the vaccine strain. The identification of these genes associated with virulence and immunogenicity will not only deepen the understanding of Mmm strain evolution, but also accelerate the development of a new vaccine strain to obtain more effective immunization results.

Materials and Methods

Bacterial growth and DNA extraction. For *Mycoplasma* growth, we used a modified ATCC 1699 broth (0.8 g glucose, 20% pig serum, 100 units of penicillin, and 0.05% acetic acid thallium). A commercial tissue genomic DNA extraction kit (Axygen, Inc., USA) was used to purify DNA.

High-density pyrosequencing and sequence assembly of the genome. The complete genomic sequencing was conducted using a Roche GS FLX system³⁹. For each strain, an average of 47,934 reads totaling 31,092,954 bases (average read length: 648 bp), was obtained resulting in 27-fold genome coverage. Assembly was performed using the GS *de novo* Assembler software (<http://www.454.com/>). The relationship of the contigs was determined by multiplex PCR⁴⁰. Gaps were then filled in by sequencing the PCR products by using ABI 3730xl capillary sequencers. Phred, Phrap, and Consed software packages (<http://www.genome.washington.edu>) were used for final assembly and editing, and low-quality regions of the genome were resequenced. The final sequencing accuracy was 99.99%.

The complete genome sequences of Mmm BEN-1-BEN-468 have been uploaded to the database of Genbank with the accession number CP011260-CP011264.

Genome annotation. Putative CDS were identified by Glimmer 3.02⁴¹ and ZCURVE 1.02⁴², and peptides shorter than 30 amino acids were eliminated. Insert sequences were first detected using the IS Finder database (<http://www-is.biotoul.fr/is.html>) using the default parameters and manual selection. Transfer RNA genes were predicted by tRNA Scan-SE⁴³. Functional annotation of CDS was performed by searching against GenBank's non-redundant (nr) protein database using BLASTP⁴⁴ and the CDD databases⁴⁵ by RPS-BLAST. Pseudogenes were identified by BLASTP with alignment amino acid (aa) length less than 20% of the referenced proteins' aa lengths.

The metabolic pathways were constructed using the KEGG database⁴⁶. Subcellular localization of the proteins was predicted by PSORTb (v2.0.1)⁴⁷, and lipoproteins were identified with LipoP 1.0⁴⁸. Pathogenicity islands were detected by IslandViewer⁴⁹. Genome comparisons were performed using MAUVE⁵⁰. The genome atlas was drawn using GenomeViz1.1⁵¹.

A new pan-genome analysis pipeline (PGAP)⁵² was used to identify the orthologs among the five genomes. The non-synonymous (Ka) and synonymous (Ks) substitution rates of orthologous genes between the four pairs of strains were calculated by KaKs_Calculator through model averaging⁵³. The mutation rate of Mmm was calculated using the following formula: $\text{Age} = \text{ds}/(\text{rate} \times 2)$, where ds is the mean number of synonymous substitutions per site calculated from the five genomes after Jukes-Cantor correction using the SNAP program⁵⁴.

T cell epitope prediction. All assay validated linear epitope sequences were downloaded from the IEDB (<http://www.iedb.org/>) and searched against Mmm Ben series proteome using BlastP. An epitope with at least 90% length coverage and 90% identity with Ben series proteome was identified as an epitope in the Mmm Ben series.

A cow MHC-I binding prediction (<http://tools.immuneepitope.org/mhci/>) was performed using IEDB analysis for all mutated genes during Ben series passages. As rabbit was not available in host selection, we used rat instead. The prediction with a low percentile rank (< 1%) and IC50 < 500 was taken as the T-cell epitope recognized by cow or rabbit.

References

- Weisburg, W. G. *et al.* A phylogenetic analysis of the mycoplasmas: basis for their classification. *J Bacteriol* **171**, 6455–6467 (1989).
- Manso-Silvan, L. *et al.* *Mycoplasma leachii* sp. nov. as a new species designation for *Mycoplasma* sp. bovine group 7 of Leach, and reclassification of *Mycoplasma mycoides* subsp. *mycoides* LC as a serovar of *Mycoplasma mycoides* subsp. *capri*. *Int J Syst Evol Microbiol* **59**, 1353–1358 (2009).
- Thiaucourt, F. *et al.* *Mycoplasma mycoides*, from “*mycoides* Small Colony” to “*capri*”. A microevolutionary perspective. *BMC genomics* **12**, 114 (2011).
- ter Laak, E. A. Contagious bovine pleuropneumonia. A review. *Vet Q* **14**, 104–110 (1992).
- Nicholas, R. A. *et al.* A comparison of serological tests and gross lung pathology for detecting contagious bovine pleuropneumonia in two groups of Italian cattle. *Vet Rec* **139**, 89–93 (1996).

6. Egwu G. O., N. R., Ameh J. A., Bashiruddin & J. B. Contagious bovine pleuropneumonia: an update. *Veterinary Bulletin* **66**, 875–888 (1996).
7. Provost A., P. P., Breard A., Le Goff C., Martel J. L., Cottew & G. S. Contagious bovine pleuropneumonia. *Revue Scientifique et Technique (Office International des Epizooties)* **6**, 625–679 (1987).
8. Thiaucourt, F. *et al.* Contagious bovine pleuropneumonia vaccines, historic highlights, present situation and hopes. *Dev Biol (Basel)* **114**, 147–160 (2003).
9. Wu T. X., W. Y. & Xu, Z. X. Research of CBPP lapinized sheep-adapted vaccine (1956–1959). *Collections of Research Paper* **7**, 5–18 (1981).
10. Zhao X. J. & J. S. Field production and observation of lapinized or sheep-adapted vaccine. *Animal Husbandry and Veterinary Medicine* **9**, 23–25 (1962).
11. Li, Y. *et al.* Strains of *Mycoplasma mycoides* subsp. *mycoides* small colony type isolated in China between 1953 and 1960 show close similarity to strains of the Africa/Australia cluster. *Vet J* **179**, 137–141 (2009).
12. Xin, J. *et al.* A history of the prevalence and control of contagious bovine pleuropneumonia in China. *Vet J* **191**, 166–170 (2012).
13. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170 (2010).
14. Vilei, E. M. *et al.* Genomic and antigenic differences between the European and African/Australian clusters of *Mycoplasma mycoides* subsp. *mycoides* SC. *Microbiology* **146** (Pt 2), 477–486 (2000).
15. Abdo, E. M., Nicolet, J. & Frey, J. Antigenic and genetic characterization of lipoprotein LppQ from *Mycoplasma mycoides* subsp. *mycoides* SC. *Clin Diagn Lab Immunol* **7**, 588–595 (2000).
16. Janis, C. *et al.* Unmarked insertional mutagenesis in the bovine pathogen *Mycoplasma mycoides* subsp. *mycoides* SC: characterization of a lppQ mutant. *Microbiology* **154**, 2427–2436 (2008).
17. Persson, A., Jacobsson, K., Frykberg, L., Johansson, K. E. & Poumarat, F. Variable surface protein Vmm of *Mycoplasma mycoides* subsp. *mycoides* small colony type. *J Bacteriol* **184**, 3712–3722 (2002).
18. Pilo, P. *et al.* A metabolic enzyme as a primary virulence factor of *Mycoplasma mycoides* subsp. *mycoides* small colony. *J Bacteriol* **187**, 6824–6831 (2005).
19. Comas, I. *et al.* Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. *Nat Genet* **42**, 498–503 (2010).
20. Zhang, W. *et al.* Genome sequencing and analysis of BCG vaccine strains. *PLoS one* **8**, e71243 (2013).
21. Merzbacher, M., Detsch, C., Hillen, W. & Stulke, J. Mycoplasma pneumoniae HPr kinase/phosphorylase. *Eur J Biochem* **271**, 367–374 (2004).
22. Halbedel, S., Busse, J., Schmidl, S. R. & Stulke, J. Regulatory protein phosphorylation in Mycoplasma pneumoniae. A PP2C-type phosphatase serves to dephosphorylate HPr(Ser-P). *J Biol Chem* **281**, 26253–26259 (2006).
23. Hopfe, M. & Henrich, B. OppA, the substrate-binding subunit of the oligopeptide permease, is the major Ecto-ATPase of Mycoplasma hominis. *J Bacteriol* **186**, 1021–1028 (2004).
24. Hopfe, M., Dahlmanns, T. & Henrich, B. In Mycoplasma hominis the OppA-mediated cytoadhesion depends on its ATPase activity. *BMC microbiol* **11**, 185 (2011).
25. Westberg, J. *et al.* The genome sequence of *Mycoplasma mycoides* subsp. *mycoides* SC type strain PG1T, the causative agent of contagious bovine pleuropneumonia (CBPP). *Genome Res* **14**, 221–227 (2004).
26. Chambaud, I. *et al.* The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis. *Nucleic Acids Res* **29**, 2145–2153 (2001).
27. Jittawuttipoka, T., Sallabhan, R., Vattanaviboon, P., Fuangthong, M. & Mongkolsuk, S. Mutations of ferric uptake regulator (fur) impair iron homeostasis, growth, oxidative stress survival, and virulence of Xanthomonas campestris pv. campestris. *Arch Microbiol* **192**, 331–339 (2010).
28. Jacobsen, I. *et al.* Deletion of the ferric uptake regulator Fur impairs the *in vitro* growth and virulence of Actinobacillus pleuropneumoniae. *Infect Immun* **73**, 3740–3744 (2005).
29. Hashimoto, M. *et al.* Lipoprotein is a predominant Toll-like receptor 2 ligand in Staphylococcus aureus cell wall components. *Int Immunol* **18**, 355–362 (2006).
30. Chu, H. W. *et al.* TLR2 signaling is critical for Mycoplasma pneumoniae-induced airway mucin expression. *J Immunol* **174**, 5713–5719 (2005).
31. Ozinsky, A. *et al.* The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors. *Proc Natl Acad Sci USA* **97**, 13766–13771 (2000).
32. Kornspan, J. D. & Rottem, S. The phospholipid profile of mycoplasmas. *J Lipids* **2012**, 640762 (2012).
33. Shimizu, T., Kida, Y. & Kuwano, K. A dipalmitoylated lipoprotein from Mycoplasma pneumoniae activates NF- κ B through TLR1, TLR2, and TLR6. *J Immunol* **175**, 4641–4646 (2005).
34. Lugade, A. A. *et al.* Lipid motif of a bacterial antigen mediates immune responses via TLR2 signaling. *PLoS one* **6**, e19781 (2011).
35. Vilei, E. M. & Frey, J. Genetic and biochemical characterization of glycerol uptake in *Mycoplasma mycoides* subsp. *mycoides* SC: its impact on H(2)O(2) production and virulence. *Clin Diagn Lab Immunol* **8**, 85–92 (2001).
36. Buwitt-Beckmann, U. *et al.* Toll-like receptor 6-independent signaling by diacylated lipopeptides. *Eur J Immunol* **35**, 282–289 (2005).
37. Wielgoss, S. *et al.* Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution Experiment With Escherichia coli. *G3* **1**, 183–186 (2011).
38. Razin, S., Yogeve, D. & Naot, Y. Molecular biology and pathogenicity of mycoplasmas. *Microbiol Mol Biol Rev* **62**, 1094–1156 (1998).
39. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
40. Tettelin, H., Radune, D., Kasif, S., Khouri, H. & Salzberg, S. L. Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. *Genomics* **62**, 500–507 (1999).
41. Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**, 4636–4641 (1999).
42. Guo, F. B., Ou, H. Y. & Zhang, C. T. ZCURVE: a new system for recognizing protein-coding genes in bacterial and archaeal genomes. *Nucleic Acids Res* **31**, 1780–1789 (2003).
43. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955–964 (1997).
44. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402 (1997).
45. Marchler-Bauer, A. *et al.* CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* **35**, D237–240 (2007).
46. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32**, D277–280 (2004).
47. Gardy, J. L. *et al.* PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* **21**, 617–623 (2005).
48. Juncker, A. S. *et al.* Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* **12**, 1652–1662 (2003).
49. Langille, M. G. & Brinkman, F. S. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* **25**, 664–665 (2009).

50. Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**, 1394–1403 (2004).
51. Ghai, R., Hain, T. & Chakraborty, T. GenomeViz: visualizing microbial genomes. *BMC bioinformatics* **5**, 198 (2004).
52. Zhao, Y. *et al.* PGAP: pan-genomes analysis pipeline. *Bioinformatics* **28**, 416–418 (2012).
53. Zhang, Z. *et al.* KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* **4**, 259–263 (2006).
54. Ota, T. & Nei, M. Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site. *Mol Biol Evol* **11**, 613–619 (1994).

Author Contributions

Conceived and designed the experiments: J.X. and Z.H. Performed the experiments: Y.L., Y.W. and H.Z. Analyzed the data: H.Z., Y.W., Y.L., H.L., X.W., R.W., Y.Z., S.L., Q.W., J.S., Y.C., L.G. and C.Z. Contributed reagents/materials/analysis tools: H.L., X.W. Wrote the paper: J.X., Z.H., Y.L. and Y.W.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Li, Y. *et al.* Changes in pathogenicity and immunogenicity of *Mycoplasma mycoides* subsp. *mycoides* strains revealed by comparative genomics analysis. *Sci. Rep.* **6**, 19081; doi: 10.1038/srep19081 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>