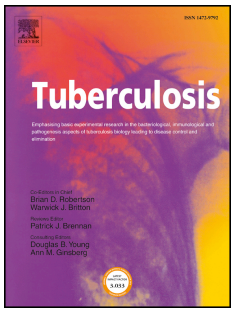
Journal Pre-proof



Genomic comparison of two strains of *Mycobacterium avium* subsp. *paratuberculosis* with contrasting pathogenic phenotype

M.A. Colombatti Olivieri, P. Fresia, M. Graña, M.X. Cuerda, A. Nagel, F. Alvarado Pinedo, M.I. Romano, K. Caimi, L. Berná, M.P. Santangelo

PII: S1472-9792(22)00136-6

DOI: https://doi.org/10.1016/j.tube.2022.102299

Reference: YTUBE 102299

To appear in: *Tuberculosis*

Received Date: 30 September 2022

Revised Date: 28 November 2022

Accepted Date: 19 December 2022

Please cite this article as: Colombatti Olivieri MA, Fresia P, Graña M, Cuerda MX, Nagel A, Alvarado Pinedo F, Romano MI, Caimi K, Berná L, Santangelo MP, Genomic comparison of two strains of *Mycobacterium avium* subsp. *paratuberculosis* with contrasting pathogenic phenotype, *Tuberculosis* (2023), doi: https://doi.org/10.1016/j.tube.2022.102299.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier Ltd.

Journal Pre-proof

Author Contributions: Genome assembly and annotation: LB and PF. bMDM infections, acquisition of data, analysis and interpretation of data: MAC, AN and MXC. Variant identification: LB. Phylogenetic analysis PF. 3D structure MG. Strains isolation FAP. All authors contributed to the analysis and interpretation of data. Conceived the idea and provided financial support: MPS and MIR. Designed the experiments and performed the analysis: MPS and LB. Wrote the draft of manuscript: MPS, LB and KC. All authors critically revised the manuscript.

Journal Pre-proof

Journal Pre-proof

Genomic comparison of two strains of *Mycobacterium avium* subsp. *paratuberculosis* with contrasting pathogenic phenotype

Colombatti Olivieri, MA[a]; Fresia, P[b]; Graña, M[c]; Cuerda, MX[a]; Nagel, A[a]; Alvarado Pinedo, F[d]; Romano, MI[a]; Caimi, K[a]; Berná, L[e] **[§]and Santangelo MP[a]*[§]

a. Instituto de Agrobiotecnología y Biología Molecular (IABIMO). INTA-CONICET. Dr. Nicolás Repetto y De Los Reseros S/Nº B1686IGC Hurlingham-Buenos Aires-Argentina

b. Unidad Mixta Pasteur+INIA, Institut Pasteur de Montevideo. Mataojo 2020 CP11400. Montevideo-Uruguay.

c. Unidad de Bioinformática, Institut Pasteur de Montevideo. Mataojo 2020 CP11400. Montevideo-Uruguay

d. Centro de Diagnóstico e Investigaciones Veterinarias (CEDIVE), Facultad de Ciencias Veterinarias - Universidad de La Plata (UNLP), Chascomus-Buenos Aires-Argentina.

e. Unidad de Biología Molecular, Institut Pasteur de Montevideo. Mataojo 2020 CP 11400. Montevideo-Uruguay.

*corresponding author: santangelo.maria@inta.gob.ar

**corresponding author: lberna@pasteur.edu.uy

§both authors contributed equally to the work

Dr. Nicolás Repetto y De Los Reseros S/Nº

B1686IGC Hurlingham-Buenos Aires-Argentina

Tel: +54 11 4621 1447 ext.3578

colombatti.alejandra@inta.gob.ar; pfresia@pasteur.edu.uy; mgrana@pasteur.edu.uy; cuerda.maria@inta.gob.ar; nagel.ariel@inta.gob.ar ; fiorella.alvaradopinedo@gmail.com; romano.mariaisabel@inta.gob.ar; caimi.karina@inta.gob.ar

Journal Pre-proof

Journal Pre-proof

Abstract

In a previous study, we evaluated the degree of virulence of *Mycobacterium avium* subsp. *paratuberculosis* (Map) strains isolated from cattle in Argentina in a murine model. This assay allowed us to differentiate between high-virulent MapARG1347 and low-virulent MapARG1543 strains. To corroborate whether the differences in virulence could be attributed to genetic differences between the strains, we performed Whole Genome Sequencing and compared the genomes and gene content between them and determined the differences related to the reference strain MapK10. We found 233 SNPs/INDELS in one or both strains relative to Map K10. The two strains share most of the variations, but we found 15 mutations present in only one of the strains. Considering NS-SNP/INDELS that produced a severe effect in the coding sequence, we focus the analysis on four predicted proteins, putatively related to virulence. Survival of MapARG1347 strain in bMDM was higher than MapARG1543 and was more resistant to acidic pH and H_2O_2 stresses than MapK10. The genomic differences between the two strains found in genes MAP1203 (a putative peptidoglycan hydrolase), MAP0403 (a putative serine protease) MAP1003c (a member of the PE-PPE family) and MAP4152 (a putative mycofactocin binding protein) could contribute to explain the contrasting phenotype previously observed in mice models.

Key words: Genomics, WGS, virulence, oxidative stress, phenotype

Journal Pre-proof

Journal Pre-proof

1. Introduction

Mycobacterium avium subsp. *paratuberculosis* (Map) is a member of the Non Tuberculous Mycobacteria group, belonging to *Mycobacterium avium* Complex. Map is the causative agent of paratuberculosis (PTB) or Johne's disease, a chronic enteritis of ruminants, such as cows, goats, deer and sheep. Efforts to control the spread of the disease in domestic livestock have been largely ineffective and the disease is now recognized worldwide. Economic losses due to paratuberculosis vary among different countries, while in the US dairy herds was estimated to be an annual loss of US$200 million ± US$ 160 million [1]. In Argentina data of PTB prevalence are scarce [2, 3]. The economic losses in the province of Buenos Aires due to paratuberculosis were estimated to be 22 million dollars for beef cattle production and 6.3 million for dairy cattle [4, 5]. Some regions of the country showed seroprevalence of Map ranging from 7 to 19.6% in breeding herds, according to the Instituto Nacional de Tecnología Agropecuaria. Map infections mainly affect dairy productions reducing milk production and fertility or increasing susceptibility to other diseases, mortality and culling [6].

In a previous study, we evaluated the degree of virulence, immune response, and protection efficacy of Map strains isolated from cattle in Argentina with different MIRU-VNTR and SSR genotype in a murine model [7]. This assay allowed us to differentiate between high-virulent and low-virulent Map strains. The low-virulent strain MapARG1543 failed to induce a significant production of antibodies and proinflammatory cytokines, such as IFNγ and TNF, but elicited a significant increase of the anti-inflammatory cytokine IL-10. On the other hand, the high-virulent strain MapARG1347 induced the highest levels of antibodies and the production of proinflammatory cytokines. This strain was efficient in establishing a persistent infection, with the highest levels of CFU counts in spleen and lesions in the liver (number of granulomas and presence of acid-fast bacilli) [7].

Considering the contrasting pathogenic phenotype of the MapARG1347 strain versus the attenuated phenotype of the MapARG1543 strain observed in mice, we hypothesized that the differences in virulence could be attributable to differences in the genomic sequences. The complete genome sequence of the MapK10 strain, was first reported in 2005 by Li and collaborators, and further re-examined in 2010 [8]. The updated MapK10 genome is 4,829,781 bp encoding 4,350 Open Reading Frames (ORFs) with a 69.3% GC content. About 60% of the ORFs have similar sequences in microbial genetic databases, but only approximately 35% had well predicted or identified functions. Although about 75% of the Map genes have counterparts in *Mycobacterium tuberculosis*, and 39 proteins have been predicted to be unique to Map, still the mechanisms of pathogenesis of Map are not fully understood [9]. Moreover, clinical Map strains may conserve virulence factors that the K10 laboratory-adapted strain may have lost.

Map genomics has many applications, such as the development of molecular detection tools, determining the molecular evolution of Map as a pathogen, identifying virulence determinants, drug targets,

Journal Pre-proof

attenuation targets for vaccine development and/or diagnostic antigens [10]. With this in mind, we performed Whole Genome Sequencing (WGS) in order to evaluate genomic variability of the strains circulating in Argentina and the region, and to identify nonsynonymous single nucleotide polymorphisms (NS-SNP) and insertions/deletions (INDELS) by comparative genomic analysis. We selected those mutations that produced a drastic effect in the coding sequences of the MapARG1347 and MapARG1543 strains.

2. Materials and methods

2.1 Map strains, culture and DNA extraction

Map isolates (Cattle-type or C-Type) were obtained from naturally infected cattle from the same herd of Buenos Aires province, Argentina between 2010 and 2011, and were selected based on their different degree of virulence in mice models [7]. The strains were originally isolated in Herrold's Egg Yolk medium from samples obtained from bowel mucosa at the necropsy (MapArg1543) and fecal matter (MapARG1347). In both cases, the isolates were from adult cows (Argentine Holstein breed), in the clinical stage of the disease. Then they were grown for 12 to 16 weeks in Middlebrook 7H9 medium (Difco) with 0.05% Tween 80, supplemented with 0.5% albumin, 0.4% dextrose, 0.5% glycerol and 2 mg/mL mycobactin J (Allied Monitor Inc., Fayette,USA). Tubes with 50 mL of the Map cultures were centrifuged, washed and suspended in sterile phosphate buffer saline (PBS). Genomic DNA extraction was performed as previously described [11].

2.2 Whole genome sequencing

Whole genome sequencing was performed using HiSeq, Illumina's integrated next generation sequencing instrument, using reversible-terminator sequencing-by-synthesis technology at the Genomic Unit from INTA. Paired-end runs were adjusted to read lengths of 2 × 250 base pairs. Libraries were prepared using Nextera XT library preparation kit. Quality analysis was done using FastQC, and filtering and trimming using Trimmomatic (TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:60).

2.3 Genome assembly and annotation

All processed reads were assembled using Spades (SPAdes version: 3.8.0)[12] with different kmers, ranging from 55 to 127, and kmer 127 performed best. Descriptive genomics statistics and quality evaluation was performed with Quast (version 5.0.2) [13] and compared with reference strains (Supplementary Table 1). The structural and functional annotation was performed with Prokka (version1.12)[14]. Predicted proteins were annotated for the identification with the Clusters of Orthologous Groups (COG) by eggNOG Mapper v2.1.7 [15]. MapARG1347 and MapARG1543 genomes were deposited in the BioSample database under

Journal Pre-proof

the accession numbers SAMN16063091 1347_498_INMV2 and SAMN16063092 1543_481_INMV2 respectively (PRJNA661552 project, SRR12593860 and SRR12593859, respectively).

2.4 Variant identification

The reads filtered for each strain were mapped using Burrow Wheeler Aligner MEM algorithm v. 0.7.17 [16] to the reference genome Map K10 (GenBank: AE016958.1) to determine the variants (SNPs and InDels) between these strains. Sequence alignment files were sorted and indexed with Samtools v1.9 [17]. To perform variant calling (SNPs and InDels) we first used samtools mpileup [18]**,** and then Varscan v2.3.9 [19] was applied using pileup2snp and pileup2indels with parameters --min-reads2 20 - -min-var-freq 0.2, allowing heterozygous mutations. Detected variants were annotated using SnpEff v4.3 [20]**.**

We selected those variants that produce a change in the coding sequences, changing the amino acid or the open reading frame, which can produce premature stop codons. These variants were manually curated through visual inspection of the mapping of reads and reference annotations using IGV [21].

2.5 Phylogenetic analysis

We downloaded 119 Map whole genome sequences from GenBank (Supplementary Table 2) including different types (B-type, S-type and C-type) for the purpose of selecting the world's representative isolates. Of these, a subset was obtained with those belonging to the C-type.

For both sets of Map isolates phylogenetic relationships were inferred with the Snippy pipeline (https://github.com/tseemann/snippy) [22], in which a core SNP alignment is obtained by first doing variant calling analysis, and then filtering recombinant regions with Gubbins (https://github.com/nickjcroucher/gubbins) [23]. SNP-sites v2.4.0 (https://github.com/sanger-pathogens/snp-sites) [24] was used to create the core SNP alignment and RAxML v8.2.11 [25] was used to reconstruct a maximum-likelihood phylogenetic tree based on the generalized time-reversible model. The tree was visualized using the R package ggtree [26]. Node support was evaluated with 1,000 bootstraps.

2.6 Structural modeling

As an approach to qualitatively evaluate the effect of the mutation, the three-dimensional structural information of the MAP1203 WT form was used and mutated forms were predicted based on the effect of the amino acid changes in the mutant forms using computational methods. WT models were obtained from the AlphaFold Protein Structure Database [27]. Just as explorations, mutant models were calculated using ColabFold [28], aware that all methods are unsuited for variant prediction.

2.7 Bovine monocyte-derived macrophages (bMDM) infection

Healthy cows from the Instituto Nacional de Tecnología Agropecuaria (INTA) experimental herd, negative for bovine tuberculosis and paratuberculosis infection, were selected for peripheral blood mononuclear cells (PBMC) isolation. Samples (60 mL) of blood were taken from each animal under sterile conditions according to the instructions of the Committee for Institutional Care and Use of Animal Experimentation (CICUAE-CICVyA). PBMC were separated from whole blood by centrifugation with Ficoll- Histopaque (Sigma) according to the manufacturer's instructions. The mononuclear cells were collected from the interface and washed three times with PBS solution. The cells were resuspended in complete RPMI 1640-HEPES medium (Invitrogen™) supplemented with 4 mM glutamine, 1 mM pyruvate, non-essential amino acids (Sigma), 10% heat inactivated fetal calf serum (Internegocios S.A., Buenos Aires, Argentina), antibiotics (100 UI/mL, penicillin and 100 µg/mL streptomycin) and antimycotic (0.25 µg/mL amphotericin B). PBMC were resuspended in RPMI-1640 medium (2.5×10^6 cells/mL) and distributed into wells (2 mL per well) in an ultra-low attachment 6-well plate (Corning®Costar®). The monocytes were allowed to adhere for 4 h at 37 ◦C in 5% CO2 and non-adherent cells were removed by washing with warm RPMI 1640. Adherent monocytes were cultured in complete RPMI-1640 with 10% fetal bovine serum (FBS) for 7–10 days to obtain MDM. The macrophages were detached with 2 mM EDTA in ice for 10 min. Cell viability was confirmed by trypan blue exclusion assay. Cell concentration was adjusted to 10^5 cells/mL and placed in 96-well plates at a concentration of 2×10^4 cells/well and cultured in RPMI-1640 with 10% FBS without antibiotics. Macrophages were infected with a multiplicity of infection (MOI) of 10. Infected cells were incubated for 4 h (T = 0), then treated with 50 µg/mL of gentamicin for 2 h, and finally washed three times with fresh RPMI medium to eliminate the extracellular bacteria. At 0, 2, 4 and 6 days post-infection (dpi), cells were scraped and lysed with 100 µL of 0.1% Triton X-100. The dilutions of lysed cells were plated to assess colony forming units (CFUs). The percentage of bacteria ingested was calculated as (the number of bacilli recovered from the monolayer lysates/the number of bacilli added to the wells) × 100 [29]. Three independent experiments with three technical replicates were performed. The percentage of phagocytosis (inoculum ingested) was calculated as = (the number of bacilli recovered from the monolayer lysates 4h post-infection (T0)/the number of bacilli added between to the wells) ×100. The data are shown as the Median ± Interquartile range (IQR) from three independent experiments (each time in triplicate). The statistical analysis was performed by using Kruskal-Wallis and Dunn post-test. (#) Significant differences between MapARG1347 and Map K-10 at 2 days post-infection (dpi) (* $p<0.05$). (a) Significant differences between 0 and 6 days post-infection in Map K-10 (** $p<0.01$). (b) Significant differences between 2 and 6 dpi in MapARG1347 (**** $p<0.0001$).

2.8 Nitric Oxide production

The culture supernatant of the infected bMDM was collected at 1, 2 and 4 dpi, and frozen at −70 ◦C before the measurement of nitrite. Nitric Oxide (NO) quantification of the culture medium was measured using the

Griess Reagent kit (Invitrogen™) according to the manufacturer's instructions. The absorbance at 548 nm was measured in a microplate reader (Multiskan, Thermo-Fisher). The NO concentration was calculated from a standard curve prepared using sodium nitrite as a reference. As a positive control an inoculum of *E. coli* was used to infect cells with a MOI 1:1. For the negative control, supernatant was taken from uninfected cells, and the NO ratio between infected vs. uninfected cells was performed. This experiment was performed in triplicate.

2.9 Sensitivity to hydrogen peroxide and pH.

Early-log-phase cultures were centrifuged at 3,000 X g for 8 min and washed with enriched 7H9 medium containing 0.02% tyloxapol at a pH of 6.5. They were centrifuged at 120 X g for 10 min to remove clumps. Single-cell suspensions were adjusted to 5 X 10^6 CFU/ml in enriched 7H9 medium containing 0.02% tyloxapol (Ty) at a pH of 6.5. The bacteria were incubated at 37°C for 2 or 4 h with 5, 20 or 50 mM H_2O_2 (Sigma-Aldrich). To measure sensitivity to pH, early–log-phase cultures were washed with 7H9-Ty-pH4.5 medium and centrifuged at 120 X g for 10 min. We adjusted single-cell suspensions to 5 X10^6 CFU ml–1 in 7H9-Ty-pH4.5 or 7Hp Ty-pH 6.5 medium and incubated them at 37 ºC. CFU were determined by plating serial dilutions of the suspensions on 7H10 agar plates. Percent of survival was calculated as (the number of output CFU/ the mean number of input CFU) x100. *In vitro* growth kinetics were performed in 7H9-Ty-pH 6.5 medium and estimated by optical density (OD). Aliquots were removed from each culture at 0, 12, 24, 30, 36, 42, 48 and 56 days, and OD was subsequently determined at 600 nm. These experiments were performed in triplicate.

3. Results

3.1 Whole genome sequencing analysis compared to Map K10 reference strain

To identify genetic bases that may be associated with differences in virulence in mice of previously reported Map strains [7], we sequenced the entire genomes using Illumina technology. After filtering and trimming the reads, we obtained a total of 3.1 and 1.3 Mbp of reads for MapARG1347 and MapARG1543 with a depth of 165X and 68X respectively (Table 1). The assembly was performed using different kmers and the best result was obtained with kmer 127. A total of 85 and 97 contigs were recovered for each strain, with an N50 of 106Kb and 100Kb for MapARG1347 and MapARG1543 respectively. Similar genome sizes were obtained for both strains around 4.78 Mbs with a GC content of 69.3% (Table 1).

Both genomic assemblies showed very good and similar quality, despite the differences in their initial coverage (number of reads used). The L50 reflects that half of the genome is distributed in about only 15 contigs (Table 1). This result shows the potential of 250 bp reads to obtain greater contiguity in genome assembly compared to the use of shorter reads (100 or 150 bp). On the other hand, the limit of resolution (saturation) in the assembly allowed by the 250 bp reads is evident, as the coverage of MapARG1347 was

Journal Pre-proof

more than double that of MapARG1543, obtaining equivalent assemblies in terms of continuity. A circos plot showing GC content and GC-skew and the coverage of the sequenced strains versus the MapK10 reference is shown in Figure 1. It can be seen that the coverage increases at sites with lower GC values, in some cases coincident with the location of the PE-PPE multigene proteins.

3.2 Genome Annotation and Variant Identification

Prokka was used to determine the structure and function of genes. A total of 4481 and 4482 genes were obtained for MapARG1347 and MapARG1543, respectively, in addition to 46 tRNAS with two copies for 11 tRNAs, and three copies for tRNA-Met and one tmRNA (Table 1 and Supplementary Table 1). Using COG annotations a function was assigned for 4481 genes for both strains. These annotated genes were classified into 21 metabolic pathways within four major categories including: Metabolism (45%); Information, storage and processing (12%); Cellular processes and signaling (15%) and poorly characterized or without functional annotation (28%) (Supplementary Figure 1A). Specifically, more genes were found related to the functional categories: Lipid metabolism (461 genes), Secondary structure (396 genes) and Transcription (346 genes) (Supplementary Figure 1A).

Both strains were compared to the MapK10 reference strains to determine genomic differences between them and the reference and 233 SNP/INDELS were found in one or both strains compared to the MapK10 reference strain. The annotation of the SNP/INDELS found shows that most of them are within genes involved in metabolic processes, mainly in lipid metabolism and cell wall biosynthesis (Supplementary Figure 1C). Interestingly, most of the variations were conserved between both strains 219/233 (94%), indicating that they are more related among each other than they are with MapK10. This was also evident when we performed a phylogenetic tree comparing the local strains with the genomes available in the GenBank (Figure 2 and Supplementary Figure 2). For this study, a dataset of 119 strains from different origins including sheep strains (S strains, Type I and Type III sub-lineages) and cattle strains (C strains or Type II, with sub-lineages Bison strains). They were initially used to be globally representative (Supplementary Figure 2), and then only 90 C-type strains were evaluated to better visualize the closest relationships to these new Argentinian strains (Figure 2). In the phylogenetic tree, it can be observed that the Argentinian strains are grouped with strains belonging to Ireland and the United Kingdom, and are closer related to other European origin strains, while the reference strain belongs to a phylogenetic group containing samples from USA and Canada.

Considering the differences present between MapARG1347 and MapARG17543 strains, we found only 15/233 (6,44%) of the mutations that were exclusive to one of the two strains. 8 SNP/ INDELS (3,43%) were exclusive to the low-virulent strain MapARG1543 (Supplementary Table 3). Of these, we were able to manually curate 7 (Table 2). On the other hand, 6 SNP (2,6 %) were exclusive to the high-virulent strain MapARG1347 (one SNP was excluded because it was in heterozygous, being a synonymous SNP) (Table 2 and Supplementary Figure 1B).

Moreover, if we consider the effect of the mutations in the CDS, most of the SNP produce synonymous variations or missense variations in only one amino acid (9/13). These variations will probably have a low or moderate effect in the protein function. We thus focus in those variations that will have a higher effect in the coding sequence of the protein i.e. SNP/INDELS that will cause a frameshift variant, a premature STOP codon or an in frame deletion (Table 2 and Supplementary Figure 1B).

3.3 Comparative analysis of predicted proteins in Map strains

3.3.1 MAP1203

The wt protein MAP1203 encoded in strains MapARG1347 and K10 (470 amino acids: aa) has a predicted signal peptide at its N-terminal (1-39 aa) and a catalytic domain of the NLPC/P60 family at its C-terminus (353-470aa) (Figure 3A). The insertion of a G at the position 1337 of MAP1203 (1413 nucleotides total length), produces a frameshift mutation that affects 24 amino acids (aa) of the C-terminal domain of the protein encoded in the low-virulent strain MapARG1543 (Figure 3B). MAP1203 encodes for a putative peptidoglycan hydrolase with a NLPC/P60 conserved domain in the C-terminal of the protein. The alignment of the aa sequence of the wild type version of MAP1203 with the orthologous protein in *M. tuberculosis* RipA (Rv1477) shows a similarity of 79%. However, the similarity of the NLPC/P60 domain is 94%. The similarity of the mutated version of the NLPC/P60 domain encoded in the low-virulent strain, decreased to 80%. The frameshift mutation results in a C-terminal tail that could in turn affect the function of the protein.

AlphaFold strongly suggests an interaction of the independently folded C-terminal domain and the N-terminal domain, which adopts a coil-coil conformation shown in cartoons (Figure 4A). This is actually supported by recent work describing the crystal structure of RipA from *Corynebacterium glutamicum*, where the N-terminal coil-coil domain covers the C-terminal enzymatic domain, keeping the protein in an auto-inhibited state. The modified C-terminal end of the mutant MAP1203 in MapARG1543 would change the electrostatic properties of the globular C-terminal domain, potentially affecting the interactions with the coil-coil (Figure 4B), which rely on specific hydrogen bonds, including the catalytic C-terminal cysteine (Cys 381) and the two adjacent residues Asp380 and Ser 382 (Cys513, Asp512 and Ser514 in *C. glutamicum*), with two amino acid residues from the coiled-coil helix, Glu72 and Asn75 (Glu 69 and Asn72 in *C. glutamicum)*, are conserved in the whole RipA family [30]. The catalytic triad reported for Rv1477 (RipA) in *M. tuberculosis* (Cys383, His432 and Glu444) [20] is also conserved in both wild type and mutated versions of MAP1203 (Cys381, His430 and Glu442).

3.3.2 MAP4152

A 12bp deletion was found at the position 17-28 of MAP4152 in the low-virulent strain MapARG1543. It was confirmed by PCR using primers flanking the deleted region (data not shown) that amplify a fragment

of 160bp in the strain carrying the deletion, while in MapK10 and MapARG1347 amplify a fragment of 172bp. The deletion produces the lack of aa 7 **VPAP** 10, restoring the frame downstream the deletion. This gene encodes for a putative Mycofactocin binding protein with a MftB domain at the position 23-102 Lasso Peptide Biosynthesis Pathway. AlphaFold predicts a reliable model whose best structural match within crystallographic structures is the Lasso peptide synthetase B1 from *Thermobifida fusca* (PDB ID 6JX3). Lasso peptides fall within a broad class of ribosomally synthesized and post-translationally modified peptides (RiPPs), which fulfill diverse roles including bacterial defense. N-terminal residues 1-18 are highly variable and presumably disordered. The 12bp deletion within this highly variable sequence stretch does not seem to affect the globular domain (Figure 3 C and D). Examining the AlphaFold model of the WT form (https://www.uniprot.org/uniprotkb/Q73SC2/entry#structure), strongly suggests that the deletion would simply shorten the N-terminal tail, outside the globular region.

3.3.3 MAP1003c

The insertion of a T at the position 623 of Map1003c, produces a frameshift mutation that affects the C-terminal of the protein encoded in the high-virulent strain MapARG1347. This gene in Map K10 encodes for a protein of 696aa carrying PE/PPE family domains in the intervals 4-94 and 321-471 aa. The function of this protein family is uncertain, but a role has been suggested that they may be related to antigenic variation of *Mycobacterium tuberculosis*. The mutant version of the protein in strain MapARG1347, in case being properly synthesized and folded, would be less than half-sized protein carrying an anomalous C-terminal 90 residues long, starting at Ala207. No structural modeling was undertaken for this chimeric product (Figure 3 E and F).

3.3.4 MAP0403

A 12bp deletion was found at the position 1064-1076 of MAP0403 in the high-virulent strain MapARG1347. It was confirmed by PCR using primers flanking the deleted region (data not shown) that amplify a fragment of 160bp in the strain carrying the deletion while in MapK10 and MapARG1543 amplify a fragment of 172bp. The deletion removes amino acids 357 VVFG 360, restoring the frame downstream the deletion. This gene encodes for a putative MarP serine-protease with a trypsin-like peptidase domain at the position 220- 357, orthologous to Rv3671c in *M. tuberculosis*. The catalytic residues His235, Asp264 and Ser343, observed in the crystal structure of both the active and inactive forms of for Rv3671c are conserved in the serine protease family [31] and strictly conserved in Corynebacteriales. Sequences from both WT strain MapK10 and mutant strain MapARG1347 also feature this triad (Figure 3 G and H). However, in the mutant strain, synthesis of the protein and/or its activity might be compromised. Regarding the synthesis of the protein, the four missing residues define the C-ter extreme of beta strand 5 of the protease domain, and their absence might disrupt the Beta strand and prevent the protein from folding correctly. As for activity, if the protein were to fold correctly, the active site would exist within a very different local environment.

Indeed, aliphatic V357-V358, aromatic F360 and hydrophobic G360 define a rather hydrophobic volume that isolates one side of the active siteA (Figure 5).

3.4 Characterization of the MapARG isolates *in vitro* conditions and in macrophages infections

3.4.1 *In vitro* growth of MapARG1347 and MapARG1543

We first evaluated the growth rate of the two isolates and the reference strain MapK10 in liquid medium, and evaluated the bacterial growth by optical density at 600 nm at days 0, 12, 24, 30, 36, 42, 48 and 56 (Supplementary Figure 3). The growth kinetics of the three Map strains is similar at the exponential phase, while significant differences were observed at the stationary phase (after 48 days of culture). This is probably due to the Map tendency to aggregate, making optical density and CFU counting difficult [32]. Map K10 strain is more adapted to develop in culture media and is less clumpy than the MapARG isolates.

3.4.2 bMDM infections

Since the polymorphisms were located in genes coding for proteins putatively involved in the fitness of Map during the infection, we further characterized the survival of the two strains in macrophage infections and in *in vitro* conditions that mimic the stress conditions that the bacterium encounters inside macrophages.

In order to evaluate the survival of the Map strains inside macrophages, we performed bMDM infections at a MOI 10:1 and evaluated CFUs recovery along the time of the infection (2, 4 and 6 dpi) as previously described [33].

The percentage of bacilli phagocytized were similar between the strains (12.5±2.2%). The results are expressed as CFU/mL (Supplementary Table 4) and as the percentage of survival of the different strains calculated related to the initial inoculum (T0) (Figure 6A). Significant differences were observed between the high-virulent strain MapARG1347 and the reference strain MapK10 at all the time points evaluated, while significant differences between the low-virulent strain MapARG1543 and the reference strain were observed only at 2dpi (Figure 6A). Regarding NO production, the low-virulent strain MapARG1547 induced significantly lower values than MapK10 and the high-virulent strain MapARG1347 at 4 dpi (Figure 6B). The cells infected with *E. coli* (positive control) had a significant NO production at all the evaluated time points (data not shown).

3.4.3 Susceptibility to oxidative and low pH stress

Inside macrophages, Map is exposed to oxidative stress produced by reactive oxygen that becomes more potent at low pH. We thus mimic these conditions by exposure of Map cultures to different concentrations of H_2O_2 and low pH in axenic cultures. Sensitivity to peroxide was evaluated by incubating the cells at different concentrations of H_2O_2 during 3 hours. The high-virulent strain MapARG1347 was more resistant

to peroxide at 20 and 50mM than the low-virulent strain MapARG1543 and the reference MapK10 (Figure 6C).

Sensitivity to pH was evaluated incubating the cells at pH 4,5 during 12, 24 and 36 days of culture. Both strains were more resistant to low pH than the reference strain MapK10 (Figure 6D) with statistical significant differences only at 36 days.

4. DISCUSSION

Understanding the mechanisms of virulence that enable Map to succeed during the infection is key to improve diagnostic tests, the development of protective vaccine candidates and the identification of potential targets for therapeutic agents.

With this purpose, in this study we used a genomic approach to predict the molecular bases that can explain the contrasting phenotypes in terms of virulence in mice of two Map strains isolated from cattle in Argentina. 4481 and 4482 ORFs were predicted for MapARG1347 and MapARG1543, respectively. While 4350 were recorded in the annotated genome of the reference strain MapK10 [9]. An analysis of the genes in our strains not present in the MapK10 annotation revealed that they are in fact differences in annotation but the gene sequences are present in all strains but not annotated. These differences could be attributable to differences in annotation pipelines and errors in single nucleotide assignment, resulting in frame shifts or ORF additions or deletions [34]. On the other hand, 120 genes present in the reference strain, were not annotated in the Argentinian isolates. Most of them were hypothetical proteins, while only 10 had predicted functions. However, when looking at the read alignments, many of these genes are present with coverage, and therefore present in the assembled genomes but not annotated. The fact that our genomes are fragmented into about 100 contigs could result in genes at the ends being interrupted and therefore not annotated or, conversely, in redundant annotations of the same gene or gene family (e.g. PE/PPE family genes, pks, transposases) [35]. Looking for variations at the genomic level, the comparison of the whole genomes of one high-virulent strain MapARG1347 and one low-virulent strain MapARG1543 in mice, versus the reference strain MapK10, retrieved a total of 233 SNPs/INDELs, 94% of which were conserved between the strains. A phylogenetic tree comparing the local strains with the genomes available in the GenBank indicated that these strains were highly similar and grouped with strains from Ireland and UK. These findings support the previous hypothesis that mycobacterial pathogens were introduced in South America with the trade of British cattle breeds imported from the United Kingdom for the last 200 years [36-38].

Only 6% of the identified mutations correspond to differences between the two strains and among these, only 1,7% could have a high effect in the coding sequence of the protein, since the SNP/INDEL produced a frameshift variant, a premature STOP codon or an in frame deletion. Moreover, these mutations mapped in

proteins putatively related to virulence factors. More functional experiments are necessary to definitively demonstrate that the genomic variations identified are responsible for the pathogenic differences observed between the two strains. However, Map Knockout experiments and complementation, requires several months to achieve. We thus performed macrophages infections and *in vitro* conditions that mimic the oxidative stress that Map encounters during infection, to characterize more deeply the fitness of the two strains. The macrophages are the first line of defense against mycobacteria, and play an important role in Map clearance. However, Map possesses different strategies to survive and multiply within the macrophages, such as modulation of the immune response, inhibition of phagosome maturation and phagosome-lysosome fusion, acidic resistance, nitric oxide (NO) and reactive oxygen species (ROS) production, among others [39]. The high-virulent strain MapARG1347 showed a higher percentage of survival in bMDM, maintaining the same level of NO production by macrophages. In fact, NO production is used as indicative of macrophages activation since it may not have an important role in eliminating intracellular Map because the macrophages cannot produce the sufficient levels of NO levels to kill these mycobacteria [40]. In addition, this strain was more resistant to H_2O_2 *in-vitro*, than MapK10 and the low-virulent MapARG1543 strains. In contrast, the low-virulent strain MapARG1543 displayed lower survival, similar to the MapK10 strain along with lower NO production and a higher susceptibility to H_2O_2. The resistance to acid pH was similar in both isolates and significantly higher than the MapK10.

Throughout the comparative genomic analysis, we identified variations in two genes putatively related to virulence, MAP0403 and MAP1003c in the high-virulent strain MapARG1347.

MAP0403 encodes for a putative membrane-associated serine protease, homologous to *M. tuberculosis* H37 Rv3671c, also called acid resistance serine protease or MarP. In the Map genome, there are 43 protease-coding genes, 38 of which constitute a core of protease genes conserved among *Mycobacterium leprae*, *Mycobacterium bovis, Mycobacterium avium* subsp*. paratuberculosis* and *M. tuberculosis* [41] with homology ranging from 63% to 97%. The function of Rv3671c was well characterized by Biswas et al 2010 [31]. The authors demonstrated that Rv3671c is a functional serine protease by structural and biochemical analyses. The activity of Rv3671c protects *M. tuberculosis* against both acidic and oxidative stress encountered in the macrophage phagosome [31, 42], and is particularly important to establish infection when the bacterium fails to avoid phagolysosomal maturation [43, 44]. More recently, Botella and collaborators [45] demonstrated that MarP cleaves and activates RipA during acid stress. RipA is a peptidoglycan hydrolase involved in cell wall homeostasis and is also important for the survival of *M. tuberculosis* in acidic conditions [45]. Surprisingly, MAP1203, the orthologous of RipA in Map, was also found in this study with a variation in the low virulent strain (see below). In addition, MarP may also have a modulatory effect on the immune response since the recombinant protein induced cytokine secretion of TNF-α and IL-10 from THP-1 cells [46].

In our experiments, we observed that MapARG1347 is more resistant to oxidative stress *in vitro* and survives longer in bMDM than the strains carrying the wild type version of the protein (MapARG1543 and MapK10). Although we still need to evaluate if the deletion of 357VVFG360 in MAP0403 affects the serine protease activity, it could be speculated that the deletion results in a more stable protein due to disulfide bonds formation between two conformational closed Cys residues. Biswas and collaborators proposed that the redox-sensitive disulfide bond increases the proteolytic activity of Rv3671c by stabilizing the protease in the conformation in which the active site residues are properly positioned for substrate binding and catalysis [31]. Even though the mechanism by which MAP0403 might protect Map against acid and oxidative damage remains to be identified, it is a good candidate to evaluate it´s role in the virulence of Map.

Another variation was found in MAP1003c, which encodes for a PE/PPE family protein (PE4/PPE15) close to the *M. tuberculosis* orthologous Rv1040c or PE8 with 75.55% similarity [47]. This gene family is restricted to pathogenic *Mycobacterium* [48] and takes up to approximately 7% of coding capacity in the *M.tuberculosis* genome (169 pe/ppe genes) [49]. However, it comprises 1% of the Map genome with only 10 PE homologous and 37 PPE homologous [50]. These proteins are acidic and glycine-rich proteins, and are identified by their specific domains (Pro-Pro-Glu and Pro-Glu, respectively), that are thought to play an important role in mycobacterial infection from both an antigenic as well as an immunologic standpoint.

The mechanism of action of PE/PPE proteins is still unclear, but it has been reported that they are selectively expressed during infection. They can modulate the host immune response and help mycobacteria to tolerate oxidative stress and low pH inside host macrophages, cell death pathways and inhibit autophagy [51-54]. In particular, MAP1003c gene is up-regulated in acid-nitrosative stress conditions [55] and in bMDM *in-vitro* infection [56] suggesting the importance of this protein in Map virulence.

In several studies, it has been reported by genome comparison, that regions of the genome encoding PE/PPE proteins have a higher SNP frequency. These results support the hypothesis that they are recognized by the immune system since they are thought to be expressed on the cell surface and provide the antigenic variation [48, 52, 57, 58] and can explain the polymorphism encountered in MAP1003c. Although the SNP present in the high-virulent strain yielded a truncated protein, it conserves the first 207 aa containing the PE domain and a new PPE domain (Figure 3 E and F).

On the other side, our analysis showed polymorphisms in two genes that could contribute to the attenuation of the low virulent strain MapARG1543. This strain was more susceptible to the stress conditions encountered inside macrophages than the reference strain and MapARG1347. One of the variations consists of a deletion of 12 nucleotides in MAP4152 that result in an in-frame deletion of 4 aa in the putative protein orthologous to MtfB. Mycofactocin (MFT) is a class of ribosomal synthesized and post-translationally modified peptides conserved in many Actinobacteria. MFT biosynthesis is predicted to

comprise a six-gene ensemble (mft genes): mftA and mftB encode the precursor peptide and its chaperone, respectively, and mftC-F encode products that modify MftA or are functionally associated with the MftA-derived end product(s) or both [59]. The role of MftB in this system is not known. However, it may be playing a role by binding the peptide substrate MFT to facilitate binding to MftC for further modifications [60]. This small protein family may serve as a scaffolding protein during mycofactocin maturation or as a carrier protein for the mature product, a putative novel redox carrier. Therefore, it is proposed that mature mycofactocin mediates electron transfer, which is vital for *M. tuberculosis* survival [61]. Although the role of MFT in *M. tuberculosis* survival, persistence and pathogenesis is not clearly understood, MFT function appears to be essential for *M. tuberculosis in vitro* growth in cholesterol, a substrate available during infection under nutrient/oxygen-deprived conditions [62]. MAP4152 could play a similar role in the virulence and survival of Map in macrophages since this gene is up-regulated after 24h post-infection in THP-1 cells [55].

Another mutation is the insertion of a G that produces a frameshift variation in the C-terminal domain of MAP1203. AlphaFold [27] consistently predicts that the globular C-terminal domain, which has been experimentally characterized (e.g. through X-ray crystallography: http://doi.org/10.2210/pdb2XIV/pdb), interacts with the more elusive N-terminal domain, which would form an elongated coil-coil. It is important to note that this prediction stands on co-evolutionary occurrences and ignores protein chemistry. In the mutated form, a significant charge anisotropy would be introduced by the C-terminal sequence (Figure 4 B), and we surmise this would affect the interaction with its molecular partner, namely the N-terminal coil-coil. More functional studies should be performed to evaluate if the mutation affects the activity of the protein in the low-virulent strain MapARG1543. This protein has a conserved NLPC/P60- peptidoglycan hydrolase domain, characteristic of the peptidoglycan hydrolase superfamily with endopeptidase activity. The orthologous in *M. tuberculosis* is RipA (Rv1477) is an important factor for remodeling the cell wall [63] and essential for cell separation during acidic stress [64]. In this condition, it is proteolytic activated by MarP (Rv3671c) [45]. In addition, RipA can modulate the immune response by inhibiting autophagy by mTOR signaling, allowing intracellular survival of *M. tuberculosis* [54]. The expression of MAP1203 in *M. smegmatis* increases the invasion capability of these naturally non-invasive mycobacteria in bovine MDBK epithelial cells and survival in RAW 264.7 macrophages. Overexpression of MAP1203 in Map also increases its capability of binding and invasion of epithelial cells. Also, may influence the growth rate or ability of Map to grow *in-vitro* culture [65]. Similar results were observed in *M. marinum* that overexpresses lipA (MMAR_2284), the homologue of RipA [66]. Strong increased expression of MAP1203, was also observed when exposed to milk and in a strain that overexpresses LuxR, indicating that it's regulated by the LuxR regulator. In that work, the authors postulate that LuxR modulates the lipid content and physical properties of the membrane that will affect the ability of the bacteria to interact with the host cells [67].

By combining crystallographic and mutational studies with functional assays and molecular modeling, it was shown that the catalytic activity of RipA relies on a Cys383, His432 and Glu444 triad [68]. The substitution of the Cys383 residue by Ala, prevented peptidoglycan hydrolysis [69]. These amino acids are conserved in both wild type and mutated versions of MAP1203 (Cys381, His430 and Glu442), encoded in the Argentinian strains. The MAP genome sequence encodes five NlpC/P60 domain-containing proteins. Bannantine and collaborators examined the structure and catalytic properties of two of these proteins, and demonstrated that the Cys-to-Ser substitution in MAP1204 (ortholog of RipB in *M. tuberculosis*) also renders the NlpC/P60 domain enzymatically inactive but does not affect binding to peptidoglycan. Members of this family of enzymes are considered attractive targets for next-generation antibiotic development [70]. These combined results prompted us to speculate that the variation in MAP1203 could contribute to explain the lower virulence of MapARG1543.

Overall, in this study we have identified by genomic comparison, variations in four genes putative involved in virulence that are good candidates to explain the differences in phenotype observed in mice between the two strains. Further research is needed to obtain more precise results, regarding the role of the mutations in the activity of the proteins and their role in virulence. While mutations in MAP1203 and MAP4152 should affect the protein functionality, most intriguing are the mutations in the virulent strain, which should result in an eventual gain of function residing in some pleiotropic effect. Also, we cannot discard the possibility that the contrasting phenotype observed between the two strains could also be explained by the effect of 9 other NS SNP (Table 2) identified that were excluded from our analysis because they produced a missense variation in one unique aa. Besides, because these are draft genomes which are not closed, the search for differences remains incomplete.

Funding: This work was supported by the National Institute of Agriculture Research (grant number PNSA PDI105); National Agency for Science and Technology Promotion, Argentina (grant number PICT 2016-0287); CONICET-Institut Pasteur Bilateral Cooperation Program.

MAC, MXC, KC, AN, MIR and MPS are CONICET fellows.

Conflict of interest: All authors declare that they have approved the submission of this manuscript for publication in Tuberculosis. This manuscript contains unpublished original work that is not under consideration for publication to any other journals and all authors have no conflict of interests.

Declarations of interest: none

Author Contributions: Genome assembly and annotation: LB and PF. bMDM infections, acquisition of data, analysis and interpretation of data: MAC, AN and MXC. Variant identification: LB. Phylogenetic analysis PF. 3D structure MG. Strains isolation FAP. All authors contributed to the analysis and interpretation of data.

Journal Pre-proof

Conceived the idea and provided financial support: MPS and MIR. Designed the experiments and performed the analysis: MPS and LB. Wrote the draft of manuscript: MPS, LB and KC. All authors critically revised the manuscript.

Journal Pre-proof

Journal Pre-proof

Figure Captions

Figure 1: Genome sequence comparison of two Argentinian Map strains versus MapK10 reference drawn using BRIG. From center: 1 GC content, 2 GC-Skew, 3 MapARG1347 (mean coverage 165X, scale max 300X), 4 MapARG1543 (mean coverage 65X, scale max 150X coverage)

Figure 2. Maximum likelihood phylogenetic tree of *Mycobacterium avium* subsp. *paratuberculosis* TypeII-C obtained based on core SNPs.

Figure 3: Schematic representation of the proteins and domains predicted by PFAM http://pfam.xfam.org (images are out of scale). A. MAP1203 wt. N-terminal Signal peptide (SS), C-terminal catalytic domain of the NLPC/P60 family (yellow). B. MAP1203 Mut: variation in the 24aa of the C-terminal of the protein (indicated with stripes) within the NLPC/P60 domain (yellow). C. MAP4152 wt. Mycofactocin MftB superfamily domain (red). D. MAP4152 Mut: deletion of 4aa 7VPAP10 (triangle). E. MAP1003c wt. Two predicted PE and PE/PPE domains (4-94aa (purple) and 321-471 (grey)). F. MAP1003c Mut. PE domain conserved at the position 4-94 (purple), new PE/PPE domain 213-294 aa (indicated with stripes). G. MAP0403 wt. MarP family serine protease with a Trypsin-like peptidase domain at the position 220-357. h. MAP0403 Mut. Deletion of 4aa 357GVVF360 (triangle).

Figure 4. Electrostatic surface representations for MAP1203 C-terminal domains interacting with the N-terminal coil-coil. The scale goes from positive blue (+5kT/e) to negative red (-5kT/e). (A) Wild-type and (B) mutated forms of MAP1203 with a circled modified zone. (C) Pairwise alignment displaying residue differences between extreme C-terminal of WT and mutated forms of MAP1203. The ellipse corresponds to the zone circled in panel B. WT model obtained from the AlphaFold Protein Structure Database [71] and the mutated form calculated using ColabFold [28]. Alignment represented using ESPript [72].

Figure 5. AlphaFold model for MAP0403 [UniProt entry Q744D8_MYCPA] Color codes indicate model confidence. A. Cartoon view of the entire protein. N-terminal helical domain shown in transparent cartoons; high confidence C-terminal peptidase domain in bold cartoons, B. Close view of the catalytic triad, which in the Corynebacteriales Order always 'sits on top of' an apolar volume defined residues V357-G360, missing in MAP0403 mutant form from virulent strain MapARG1347. Within the transparent volume, the extreme of Beta strand 5 is sketched, disrupted in the mutant form.

Figure 6. bMDM infections, NO production, H_2O_2 and pH susceptibility. (A) Percentage of survival in bMDM at 2, 4 and 6 days post-infection was calculated as the Median (CFU/mL Tx/ CFU/mL T0) * 100) ± interquartile range. Significant differences among average from three independent experiments were reported. The statistical analysis was performed using Kruskal-Wallis Test and Dunn Test. (B) Nitric Oxide production in the infected bMDM at 1, 2 and 4 days post-infecction was calculated as the Media ([$_u$M]

infected cells / [$_u$M] non-infected cells) ± SEM from three independent experiments. The statistical analysis was performed using two-way ANOVA, bonferroni post-test. (C) Susceptibility to H_2O_2. The results are shown as % of survival calculated as the Media ((CFU/mL $_x$[mM] / CFU/mL [0 mM]) *100) +/- SEM. The statistical analysis was performed using two-way ANOVA, bonferroni post-test. (D) Susceptibility to acid pH (pH 4.5). The results are shown as % of survival calculated as the Media (CFU/mL pH4.5 / CFU/mL pH6.5)*100) +/- SEM. The statistical analysis was performed using two-way ANOVA, bonferroni post-test. (*p < 0.05, **p < 0.01, *** p < 0.001, **** p < 0.0001)

Table 1 Genome assembly and annotation of Argentinian *Mycobacterium avium* paratuberculosis *strains*

Table 2 Specific single nucleotide polymorphisms and insertions and deletions in MapARG1347 and MapARG1543 in comparison with MapK10

Supplementary Table 1 Genome annotation compared to MapK10 reference strain

Supplementary Table 2 119 Map whole genome sequences downloaded from GenBank

Supplementary Table 3 Total Variants single nucleotide polymorphisms and insertions/deletions

Supplementary Figure 1 (A) COG annotations of all coding genes identified for strain MapARG1347, almost same result were identified for MapARG1543 (B) Venn diagram showing the mutations identified in MapARG1347 and MapARG1543 versus MapK10. The majority of variations are present in both strains, and only 13 correspond to differences between Argentinean strains. (C) Molecular Function Gene Ontology of genes with SNP/indels common MapARG1347 and MapARG1543 versus MapK10.

Supplementary Figure 2. Maximum likelihood phylogenetic tree of *Mycobacterium avium* subsp. *paratuberculosis* obtained based on core SNPs including Type I-S, Type II B and C, Type III-S.

Supplementary Figure 3. In vitro growth curves of Map strains: Map K10 (green), MapARG1347 (red), MapARG1543 (yellow). The bacterial growth was estimated by optical density at 600 nm at days 0, 12, 24, 30, 36, 42, 48 and 56. The results from three independent experiments are shown as the average Mean OD ± Standard Error of the Mean (SEM). One-way ANOVA and Bonferroni post-test. Significant differences were observed at 48 and 56 days. a= significant differences of MapARG1347 with Map K-10 and MapARG1547 (* p<0.05). b= significant differences between MapARG1543 and Map K-10 (**** p<0.0001). c= significant differences between MapARG1347 and Map K-10 (**** p<0.0001)

579 References

580 [1] Whittington R, Donat K, Weber MF, Kelton D, Nielsen SS, Eisenberg S, et al. Control of paratuberculosis:
581 who, why and how. A review of 48 countries. BMC veterinary research 2019;15:198.
582 [2] Espeschit IF, Schwarz DGG, Faria ACS, Souza MCC, Paolicchi FA, Juste RA, et al. Paratuberculosis in Latin
583 America: a systematic review. Tropical animal health and production 2017;49:1557-76.
584 [3] Paolicchi F SE, P. SM, Traveria GE, Alonso B, Mundo S. . Paratuberculosis in Argentina: Current status of
585 disease, control and application of diagnostic tools., International Dairy Federation, Bulletin of the
586 International Dairy Federation, 2018, p. p. 108.
587 [4] Moreira MA SE, Morsella C. ., St John´s College Cambridge United Kingdom. Seroprevalence of Johne´s
588 disease in eleven districts of Buenos Aires Argentina. Proceedings of the fourth International Colloquium on
589 Paratuberculsis, 1994.
590 [5] Moreira MA TJ. Paratuberculosis bovina: Importancia de la enfermedad en la región. INTA, Informe
591 Tecnico INTA Cerbas, 1995, p. 14.
592 [6] Bannantine JP, Etienne G, Laval F, Stabel JR, Lemassu A, Daffe M, et al. Cell wall peptidolipids of
593 Mycobacterium avium: from genetic prediction to exact structure of a nonribosomal peptide. Molecular
594 microbiology 2017;105:525-39.
595 [7] Colombatti Olivieri MA, Moyano RD, Traveria GE, Alvarado Pinedo MF, Mon ML, Gravisaco MJ, et al.
596 Protection efficacy of Argentinian isolates of Mycobacterium avium subsp. paratuberculosis with different
597 genotypes and virulence in a murine model. Research in veterinary science 2018;121:4-11.
598 [8] Wynne JW, Seemann T, Bulach DM, Coutts SA, Talaat AM, Michalski WP. Resequencing the
599 Mycobacterium avium subsp. paratuberculosis K10 genome: improved annotation and revised genome
600 sequence. Journal of bacteriology 2010;192:6319-20.
601 [9] Li L, Bannantine JP, Zhang Q, Amonsin A, May BJ, Alt D, et al. The complete genome sequence of
602 Mycobacterium avium subspecies paratuberculosis. Proceedings of the National Academy of Sciences of
603 the United States of America 2005;102:12344-9.
604 [10] Rathnaiah G, Zinniel DK, Bannantine JP, Stabel JR, Grohn YT, Collins MT, et al. Pathogenesis, Molecular
605 Genetics, and Genomics of Mycobacterium avium subsp. paratuberculosis, the Etiologic Agent of Johne's
606 Disease. Frontiers in veterinary science 2017;4:187.
607 [11] Zumarraga MJ, Meikle V, Bernardelli A, Abdala A, Tarabla H, Romano MI, et al. Use of touch-down
608 polymerase chain reaction to enhance the sensitivity of Mycobacterium bovis detection. Journal of
609 veterinary diagnostic investigation : official publication of the American Association of Veterinary
610 Laboratory Diagnosticians, Inc 2005;17:232-8.
611 [12] Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler.
612 Current protocols in bioinformatics 2020;70:e102.
613 [13] Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies.
614 Bioinformatics 2013;29:1072-5.
615 [14] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068-9.
616 [15] Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional
617 Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Molecular biology
618 and evolution 2021;38:5825-9.
619 [16] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Genomics (q-
620 bio.GN) 2013;3.
621 [17] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format
622 and SAMtools. Bioinformatics 2009;25:2078-9.
623 [18] Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population
624 genetical parameter estimation from sequencing data. Bioinformatics 2011;27:2987-93.
625 [19] Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and
626 copy number alteration discovery in cancer by exome sequencing. Genome research 2012;22:568-76.
627 [20] Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and
628 predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila
629 melanogaster strain w1118; iso-2; iso-3. Fly 2012;6:80-92.
630 [21] Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics
631 viewer. Nature biotechnology 2011;29:24-6.

[22] Seemann T. Snippy: fast bacterial varian calling from NGS reads. 2015.
[23] Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic acids research 2015;43:e15.
[24] Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microbial genomics 2016;2:e000056.
[25] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312-3.
[26] Yu G SD, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 2017;28-36.
[27] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583-9.
[28] Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nature methods 2022;19:679-82.
[29] Feola RP, Collins MT, Czuprynski CJ. Hormonal modulation of phagocytosis and intracellular growth of Mycobacterium avium ss. paratuberculosis In bovine peripheral blood monocytes. Microbial pathogenesis 1999;26:1-11.
[30] Quentin Gaday Daniela M, Giacomo Carloni , Mariano Martinez , Bohdana Sokolova, Mathilde Ben Assaya, Pierre Legrand , Sebastien Brûlé, Ahmed Haouz , Anne Marie Wehenkel , and Pedro M. Alzari. FtsEX-independent control of RipA-mediated cell separation in Corynebacteriales. PNAS 2022;119:1-10.
[31] Biswas T, Small J, Vandal O, Odaira T, Deng H, Ehrt S, et al. Structural insight into serine protease Rv3671c that Protects M. tuberculosis from oxidative and acidic stress. Structure 2010;18:1353-63.
[32] Elguezabal N, Bastida F, Sevilla IA, Gonzalez N, Molina E, Garrido JM, et al. Estimation of Mycobacterium avium subsp. paratuberculosis growth parameters: strain characterization and comparison of methods. Applied and environmental microbiology 2011;77:8615-24.
[33] Viale MN, Colombatti Olivieri MA, Alonso N, Moyano RD, Imperiale B, Morcillo N, et al. Effect of the deletion of lprG and p55 genes in the K10 strain of Mycobacterium avium subspecies paratuberculosis. Research in veterinary science 2021;138:1-10.
[34] Deshayes C, Perrodou E, Gallien S, Euphrasie D, Schaeffer C, Van-Dorsselaer A, et al. Interrupted coding sequences in Mycobacterium smegmatis: authentic mutations or sequencing errors? Genome biology 2007;8:R20.
[35] Ceres KM, Stanhope MJ, Grohn YT. A critical evaluation of Mycobacterium bovis pangenomics, with reference to its utility in outbreak investigation. Microbial genomics 2022;8.
[36] Cataldi AA, Gioffre A, Santtangelo MP, Alito A, Caimi K, Bigi F, et al. [The genotype of the principal Mycobacterium bovis in Argentina is also that of the British Isles: did bovine tuberculosis come from Great Britain?]. Revista Argentina de microbiologia 2002;34:1-6.
[37] Smith NH. The global distribution and phylogeography of Mycobacterium bovis clonal complexes. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 2012;12:857-65.
[38] Lasserre M, Fresia P, Greif G, Iraola G, Castro-Ramos M, Juambeltz A, et al. Whole genome sequencing of the monomorphic pathogen Mycobacterium bovis reveals local differentiation of cattle clinical isolates. BMC genomics 2018;19:2.
[39] Arsenault RJ, Maattanen P, Daigle J, Potter A, Griebel P, Napper S. From mouth to macrophage: mechanisms of innate immune subversion by Mycobacterium avium subsp. paratuberculosis. Veterinary research 2014;45:54.
[40] Simutis FJ, Jones DE, Hostetter JM. Failure of antigen-stimulated gammadelta T cells and CD4+ T cells from sensitized cattle to upregulate nitric oxide and mycobactericidal activity of autologous Mycobacterium avium subsp. paratuberculosis-infected macrophages. Veterinary immunology and immunopathology 2007;116:1-12.
[41] Ribeiro-Guimaraes ML, Pessolani MC. Comparative genomics of mycobacterial proteases. Microbial pathogenesis 2007;43:173-8.

[42] Small JL, O'Donoghue AJ, Boritsch EC, Tsodikov OV, Knudsen GM, Vandal O, et al. Substrate specificity of MarP, a periplasmic protease required for resistance to acid and oxidative stress in Mycobacterium tuberculosis. The Journal of biological chemistry 2013;288:12489-99.
[43] Vandal OH, Pierini LM, Schnappinger D, Nathan CF, Ehrt S. A membrane protein preserves intrabacterial pH in intraphagosomal Mycobacterium tuberculosis. Nature medicine 2008;14:849-54.
[44] Levitte S, Adams KN, Berg RD, Cosma CL, Urdahl KB, Ramakrishnan L. Mycobacterial Acid Tolerance Enables Phagolysosomal Survival and Establishment of Tuberculous Infection In Vivo. Cell host & microbe 2016;20:250-8.
[45] Botella H, Vaubourgeix J, Lee MH, Song N, Xu W, Makinoshima H, et al. Mycobacterium tuberculosis protease MarP activates a peptidoglycan hydrolase during acid stress. The EMBO journal 2017;36:536-48.
[46] Garcia-Gonzalez G, Ascacio-Martinez JA, Hernandez-Bello R, Gonzalez GM, Palma-Nicolas JP. Expression of recombinant protease MarP from Mycobacterium tuberculosis in Pichia pastoris and its effect on human monocytes. Biotechnology letters 2021;43:1787-98.
[47] Mackenzie N, Alexander DC, Turenne CY, Behr MA, De Buck JM. Genomic comparison of PE and PPE genes in the Mycobacterium avium complex. Journal of clinical microbiology 2009;47:1002-11.
[48] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 1998;393:537-44.
[49] McGuire AM, Weiner B, Park ST, Wapinski I, Raman S, Dolganov G, et al. Comparative analysis of Mycobacterium and related Actinomycetes yields insight into the evolution of Mycobacterium tuberculosis pathogenesis. BMC genomics 2012;13:120.
[50] Gey van Pittius NC, Sampson SL, Lee H, Kim Y, van Helden PD, Warren RM. Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC evolutionary biology 2006;6:95.
[51] Yu X, Feng J, Huang L, Gao H, Liu J, Bai S, et al. Molecular Basis Underlying Host Immunity Subversion by Mycobacterium tuberculosis PE/PPE Family Molecules. DNA and cell biology 2019;38:1178-87.
[52] Medha, Sharma S, Sharma M. Proline-Glutamate/Proline-Proline-Glutamate (PE/PPE) proteins of Mycobacterium tuberculosis: The multifaceted immune-modulators. Acta tropica 2021;222:106035.
[53] Dou Y, Xie Y, Zhang L, Liu S, Xu D, Wei Y, et al. Host MKRN1-Mediated Mycobacterial PPE Protein Ubiquitination Suppresses Innate Immune Response. Frontiers in immunology 2022;13:880315.
[54] Shariq M, Quadir N, Alam A, Zarin S, Sheikh JA, Sharma N, et al. The exploitation of host autophagy and ubiquitin machinery by Mycobacterium tuberculosis in shaping immune responses and host defense during infection. Autophagy 2022;1-21.
[55] Cossu A, Sechi LA, Zanetti S, Rosu V. Gene expression profiling of Mycobacterium avium subsp. paratuberculosis in simulated multi-stress conditions and within THP-1 cells reveals a new kind of interactive intramacrophage behaviour. BMC microbiology 2012;12:87.
[56] Zhu X, Tu ZJ, Coussens PM, Kapur V, Janagama H, Naser S, et al. Transcriptional analysis of diverse strains Mycobacterium avium subspecies paratuberculosis in primary bovine monocyte derived macrophages. Microbes and infection / Institut Pasteur 2008;10:1274-82.
[57] Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, et al. Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. Journal of bacteriology 2002;184:5479-90.
[58] Forrellad MA, Klepp LI, Gioffre A, Sabio y Garcia J, Morbidoni HR, de la Paz Santangelo M, et al. Virulence factors of the Mycobacterium tuberculosis complex. Virulence 2013;4:3-66.
[59] Haft DH. Bioinformatic evidence for a widely distributed, ribosomally produced electron carrier precursor, its maturation proteins, and its nicotinoprotein redox partners. BMC genomics 2011;12:21.
[60] Bruender NA, Bandarian V. The Radical S-Adenosyl-l-methionine Enzyme MftC Catalyzes an Oxidative Decarboxylation of the C-Terminus of the MftA Peptide. Biochemistry 2016;55:2813-6.
[61] Khaliullin B, Aggarwal P, Bubas M, Eaton GR, Eaton SS, Latham JA. Mycofactocin biosynthesis: modification of the peptide MftA by the radical S-adenosylmethionine protein MftC. FEBS letters 2016;590:2538-48.
[62] Krishnamoorthy G, Kaiser P, Lozza L, Hahnke K, Mollenkopf HJ, Kaufmann SHE. Mycofactocin Is Associated with Ethanol Metabolism in Mycobacteria. mBio 2019;10.

[63] Chao MC, Kieser KJ, Minami S, Mavrici D, Aldridge BB, Fortune SM, et al. Protein complexes and proteolytic activation of the cell wall hydrolase RipA regulate septal resolution in mycobacteria. PLoS pathogens 2013;9:e1003197.
[64] Hett EC, Chao MC, Deng LL, Rubin EJ. A mycobacterial enzyme essential for cell division synergizes with resuscitation-promoting factor. PLoS pathogens 2008;4:e1000001.
[65] Everman JL, Danelishvili L, Flores LG, Bermudez LE. MAP1203 Promotes Mycobacterium avium Subspecies paratuberculosis Binding and Invasion to Bovine Epithelial Cells. Frontiers in cellular and infection microbiology 2018;8:217.
[66] Gao LY, Pak M, Kish R, Kajihara K, Brown EJ. A mycobacterial operon essential for virulence in vivo and invasion and intracellular persistence in macrophages. Infection and immunity 2006;74:1757-67.
[67] Alonso-Hearn M, Eckstein TM, Sommer S, Bermudez LE. A Mycobacterium avium subsp. paratuberculosis LuxR regulates cell envelope and virulence. Innate immunity 2010;16:235-47.
[68] Squeglia F, Ruggiero A, Romano M, Vitagliano L, Berisio R. Mutational and structural study of RipA, a key enzyme in Mycobacterium tuberculosis cell division: evidence for the L-to-D inversion of configuration of the catalytic cysteine. Acta crystallographica. Section D, Biological crystallography 2014;70:2295-300.
[69] Both D, Schneider G, Schnell R. Peptidoglycan remodeling in Mycobacterium tuberculosis: comparison of structures and catalytic activities of RipA and RipB. Journal of molecular biology 2011;413:247-60.
[70] Bannantine JP, Lingle CK, Adam PR, Ramyar KX, McWhorter WJ, Stabel JR, et al. NlpC/P60 domain-containing proteins of Mycobacterium avium subspecies paratuberculosis that differentially bind and hydrolyze peptidoglycan. Protein science : a publication of the Protein Society 2016;25:840-51.
[71] Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic acids research 2022;50:D439-D44.
[72] Gouet P, Robert X, Courcelle E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic acids research 2003;31:3320-3.

Journal Pre-proof

Table 1. Genome assembly and annotation of Argentinian *Mycobacterium avium paratuberculosis* strains

Assembly	**MapARG1347**	**MapARG1543**
# raw reads	4838566	2119676
# filtered reads	3184239	1308390
Coverage	165X	68X
# contigs	89	100
Total length (bp)	4788364	4788052
Largest contig (bp)	397688	236982
GC (%)	69.33	69.33
L50	15	17
Annotation		
CDS	4482	4481
tRNA	46*	46*
tmRNA	1	1

* 11 tRNAs have two copies, and tRNA-Met has three.

Journal Pre-proof

Journal Pre-proof

Table 2. Specific single nucleotide polymorphisms and insertions and deletions in MapARG1347 and MapARG1543 in comparison with MapK10

Gene	Function	REF	VAR	MapARG1347	MapARG1543	SNP	Effect	AA (ref)	AA (var)
MAP_1003c	**PE**	**C**	**CA**	**VAR**	**REF**	**frameshift_variant**	**HIGH**	**-**	**90 aa Cter VAR**
MAP_0403	**hypothetical protein**	**CTCGGCGTGGTGT**	**C**	**VAR**	**REF**	**disruptive_inframe_deletio**	**MODERATE**	**GVVF**	**in frame deletion**
MAP_1966c	GlnA2	A	C	VAR	REF	missense_variant	MODERATE	Phe	Ser
MAP_0953	hypothetical protein	C	T	VAR	REF	missense_variant	MODERATE	Ala	Val
MAP_1432	hypothetical protein	G	A	VAR	REF	missense_variant	MODERATE	Leu	Glu
MAP_2258c	hypothetical protein	G	C	VAR	REF	missense_variant	MODERATE	-	-
MAP_1203	**hypothetical protein**	**C**	**CG**	**REF**	**VAR**	**frameshift_variant**	**HIGH**	**-**	**24 aa Cter VAR**
MAP_4152	**hypothetical protein**	**ACCGTGCCGGCG**	**A**	**REF**	**VAR**	**disruptive_inframe_deletio**	**MODERATE**	**VPAP**	**in frame deletion**
MAP_0322c	DnaZX	C	T	REF	VAR	upstream_gene_variant	MODIFIER	non coding regior	non coding region
MAP_0001	DnaA	G	C	REF	VAR	missense_variant	MODERATE	Lys	Asn
MAP_1965c	GlnE	G	A	REF	VAR	missense_variant	MODERATE	Thr	Lys
MAP_3177	hypothetical protein	G	C	REF	VAR	missense_variant	MODERATE	Ala	Pro
MAP_1422	hypothetical protein	C	T	REF	VAR	missense_variant	MODERATE	Arg	Cys

Gene	Function	REF	VAR	Strain		SNP	Effect	Aminoacid	
				MapARG1347	MapARG154			REF	VAR
MAP_1203	**hypothetical protein**	C	CG	REF	VAR	**frameshift_variant**	HIGH		24 aa Cter VAR
MAP_3177	**hypothetical protein**	G	C	REF	VAR	**missense_variant**	MODERATE	Ala	Pro
MAP_0322c	**DnaZX**	C	T	REF	VAR	**upstream_gene_variant**	MODIFIER	non coding region	
MAP_1422	**hypothetical protein**	C	T	REF	VAR	**missense_variant**	MODERATE	arg	cys
MAP_4152	**hypothetical protein**	ACCGTGCCGGCG	A	REF	VAR	**disruptive_inframe_deletio**	MODERATE	VPAP	in frame deletion
MAP_0001	**DnaA**	G	C	REF	VAR	**missense_variant**	MODERATE	lys	asn
MAP_1965c	**GlnE**	G	A	REF	VAR	**missense_variant**	MODERATE	thr	lys
MAP_0953	**hypothetical protein**	C	T	VAR	REF	**missense_variant**	MODERATE	ala	val
MAP_1003c	**PE**	C	CA	VAR	REF	**frameshift_variant**	HIGH		90 aa Cter VAR
MAP_1432	**hypothetical protein**	G	A	VAR	REF	**missense_variant**	MODERATE	Leu	Glu
MAP_0403	**hypothetical protein**	CTCGGCGTGGTGT	C	VAR	REF	**disruptive_inframe_deletio**	MODERATE	GVVF	in frame deletion
MAP_1966c	**GlnA2**	A	C	VAR	REF	**missense_variant**	MODERATE	Phe	Ser

Journal Pre-proof

MAP_2258c	**hypothetical protein**	G	C	VAR	REF	**missense_variant**	MODERATE	

Journal Pre-proof

Genome annotation compared to MapK10 reference strain

strains	**K10**	**MapARG1347**	**MapARG1543**
Genome size, bp	48,29,781	47,88,364	47,88,052
G+C content, %	69.3	69.3	69.3
ORFs	4,350	4482	4481
tRNAs	45	46*	46*
rRNA operon	1	1	1

* 11 tRNAs have two copies, and tRNA-Met has three.

Journal Pre-proof

Journal Pre-proof

119 MAP whole genome sequences downloa

Sequence ID	Type	Average Coverage
ERR038025	Type II/C	187x
ERR248986	Type II/C	46x
SRR1793761	Type II/C	79x
ERR026344	Type II/C	65x
ERR038023	Type II/C	284x
ERR037990	Type II/C	278x
ERR037948	Type II/C	385x
ERR038022	Type II/C	311x
ERR040075	Type II/C	185x
ERR038622	Type II/C	144x
ERR037393	Type II/C	120x
ERR037956	Type II/C	237x
ERR037982	Type II/C	402x
ERR037394	Type II/C	348x
ERR037395	Type II/C	178x
ERR040085	Type II/C	199x
ERR037389	Type II/C	166x
SRR3050045	Type II/C	37x
ERR026350	Type II/C	50x
SRR3050026	Type II/C	76x
SRR1793772	Type II/C	62x
ERR038026	Type II/C	293x
ERR040082	Type II/C	236x
ERR037952	Type II/C	199x
ERR040077	Type II/C	223x
ERR038016	Type II/C	402x
ERR038015	Type II/C	210x
sar02	Type II/C	84x
sar03	Type II/C	98x
sar01	Type II/C	204x
SRR1793770	Type II/C	284x
ERR026342	Type II/C	395x

Journal Pre-proof

Journal Pre-proof

ERR040080	Type II/C	37x
ERR037980	Type II/C	458x
ERR037994	Type II/C	225x
ERR248982	Type II/C	55x
ERR037991	Type II/C	22x
ERR040076	Type II/C	32x
ERR026347	Type II/C	48x
SRR1793745	Type II/C	111x
SRR3050038	Type II/C	171x
ERR026345	Type II/C	23x
ERR037391	Type II/C	53x
ERR038006	Type II/C	52x
SRR3050011	Type II/C	20x
SRR3050019	Type II/C	49x
SRR1793758	Type II/C	66x
SRR3050006	Type II/C	46x
SRR3050060	Type II/C	215x
SRR1793771	Type II/C	100x
ERR248984	Type II/C	361x
SRR198002	Type II/C	716x
SRR201790	Type II/C	226x
SRR197999	Type II/C	1758x
ERR045299	Type II/C	84x
SRR5413272	**Type II/C**	232x
SRR641389	Type II/C	212x
SRR1793691	Type II/C	138x
ERR038024	Type II/C	145x
ERR037953	Type II/C	114x
ERR037385	Type II/C	220x
ERR038013	Type II/C	355x
ERR038007	Type II/C	207x
ERR040078	Type II/C	351x
ERR038017	Type II/C	107x
ERR037392	Type II/C	43x

Journal Pre-proof

Journal Pre-proof

ERR038626	Type II/C	48x
SRR1793688	Type II/C	52x
SRR3050052	Type II/C	233x
ERR248985	Type II/C	209x
SRR1793754	Type II/C	41x
SRR5204549	Type II/C	228x
SRR201791	Type II/C	181x
SRR1793696	Type II/C	84x
ERR038003	Type II/C	67x
SRR5204548	Type II/C	283x
ERR026341	Type II/C	153x
ERR026349	Type II/C	146x
ERR038027	Type II/C	121x
ERR037989	Type II/C	104x
ERR037390	Type II/C	255x
ERR037388	Type II/C	466x
ERR038630	Type II/C	254x
ERR038029	Type II/C	479x
ERR037981	Type II/C	137x
ERR037384	Type II/C	10x
ERR038020	Type II/C	138x
ERR038621	Type II/C	126x
SRR3050058	Type II/C	50x
ERR038624	Type II/C	69x
SRR3050004	Type II/B	63x
ERR026348	Type II/B	62x
ERR026346	Type II/B	64x
SRR3050064	Type II/B	37x
SRR3050010	Type II/B	48x
SRR3050048	Type II/B	109x
SRR3050020	Type II/B	109x
SRR3050021	Type II/B	248x
SRR1793685	Type II/B	248x
ERR026340	Type II/B	70x

Journal Pre-proof

Journal Pre-proof

ERR040083	Type II/B	87x
ERR040084	Type II/B	62x
ERR026339	Type II/B	209x
ERR026343	Type II/B	209x
ERR037950	Type II/B	243x
ERR266518	Type I/S	153x
ERR037996	Type I/S	185x
ERR037995	Type I/S	164x
ERR038000	Type I/S	233x
ERR038001	Type I/S	189x
ERR037999	Type I/S	32x
ERR037997	Type I/S	152x
ERR037998	Type I/S	418x
ERR038002	Type I/S	427x
SRR3050018	Type III/S	209x
ERR038625	Type III/S	52x
ERR038028	Type III/S	129x
ERR038019	Type III/S	307x
ERR038627	Type III/S	249x
ERR266515	Type III/S	
ERR038629	Type III/S	
ERR038021	Type III/S	

Journal Pre-proof

Journal Pre-proof

Total variants single nucleotide polymorphisms and insertions/deletions:

Total Variants single nucleotide polymorphisms and insertions/deletions. Variant calling was done using samtools mpileup and then Varscanv2.3.9 was applied using pileup2snp and pileup2indels with parameters --min-reads2 20 --min-var-freq 0.2, allowing heterozygous mutations. Detected variants were annotated usingSnpEff v4.3.

Localization	#CHROM	POS	REF	ALT	FILTER	INFO	FORMAT	MapARG1347	MapARG1543	MapARG1347	MapARG1543	SNPEFF annot Variant	Modifier	Gene Name	Gene Id	Protein Id	Description
gen	Chromosome	1265279	C	CG	PASS	ADP=51;WT=2	GT:GQ:SDP:D	0/0:115:62:61:(	1/1:118:29:29:(	0/0	1/1	ADP=51;WT=2	frameshift_varia	HIGH	MAP_1203	MAP_1203	AAS03520.1	peptidoglycan DL-endopeptidase RipA
5' utr	Chromosome	2450214	A	G	PASS	ADP=53;WT=2	GT:GQ:SDP:D	0/0:127:70:67:(	1/1:127:29:29:(	0/0	1/1	ADP=53;WT=2	upstream_gene	MODIFIER	MAP_2202c	MAP_2202c	AAS04519.1	no KO assigned | (GenBank) hypothetical protein
gen	Chromosome	3528714	G	C	PASS	ADP=60;WT=2	GT:GQ:SDP:D	0/0:152:86:82:(	1/1:173:39:39:(	0/0	1/1	ADP=60;WT=2	missense_varia	MODERATE	MAP_3177	MAP_3177	AAS05725.1	no KO assigned | (GenBank) hypothetical protein
5' utr	Chromosome	348992	C	T	PASS	ADP=44;WT=2	GT:GQ:SDP:D	0/0:122:67:66:(	./.:.:11	0/0	1/1	ADP=44;WT=2	upstream_gene	MODIFIER	dnaZX	MAP_0322c	AAS02639.1	DNA polymerase III subunit gamma/tau- purine metabolism
gen	Chromosome	1554884	C	T	PASS	ADP=46;WT=2	GT:GQ:SDP:D	0/0:119:63:63:(	./.:.:18	0/0	1/1	ADP=46;WT=2	missense_varia	MODERATE	MAP_1422	MAP_1422	AAS03739.1	no KO assigned | (GenBank) hypothetical protein
gen	Chromosome	4630320	ACCGTGCCG(	A	PASS	ADP=47;WT=2	GT:GQ:SDP:D	0/0:134:74:71:(	./.:.:16	0/0	1/1	ADP=47;WT=2	disruptive_infra	MODERATE	MAP_4152	MAP_4152	AAS06702.1	no KO assigned | (GenBank) hypothetical protein
gen	Chromosome	783	G	C	PASS	ADP=58;WT=1	GT:GQ:SDP:D	0/0:138:75:73:(	1/1:146:28:26:(	0/0	1/1	ADP=58;WT=1	missense_varia	MODERATE	dnaA	MAP_0001	AAS02318.1	K02313 chromosomal replication initiator protein | (GenBank) dnaA; DnaA
gen	Chromosome	2170122	G	A	PASS	ADP=43;WT=1	GT:GQ:SDP:D	0/0:107:61:58:(	1/1:127:30:29:(	0/0	1/1	ADP=43;WT=1	missense_varia	MODERATE	glnE	MAP_1965c	AAS04282.1	K00982 [glutamine synthetase] adenylyltransferase / [glutamine synthetase]-adenylyl-L-tyrosine phosphorylase
gen	Chromosome	4481811	C	T	PASS	ADP=43;WT=2	GT:GQ:SDP:D	0/1:72:64:62:41	./.:.:17	0/1	0/0	ADP=43;WT=2	synonymous_va	LOW	hemL	MAP_4020	AAS06570.1	K01845 glutamate-1-semialdehyde 2,1-aminomutase [EC:5.4.3.8] | (GenBank) hemL; HemL
gen	Chromosome	986655	C	T	PASS	ADP=52;WT=2	GT:GQ:SDP:D	1/1:255:74:73:(	./.:.:22	1/1	0/0	ADP=52;WT=2	missense_varia	MODERATE	MAP_0953	MAP_0953	AAS03270.1	no KO assigned | (GenBank) hypothetical protein
gen	Chromosome	1048303	C	CA	PASS	ADP=34;WT=2	GT:GQ:SDP:D	1/1:201:45:44:(	./.:.:15	1/1	0/0	ADP=34;WT=2	frameshift_varia	HIGH	PE	MAP_1003c	AAS03320.1	no KO assigned | (GenBank) PE; PE
gen	Chromosome	1565698	G	A	PASS	ADP=39;WT=2	GT:GQ:SDP:D	1/1:254:46:44:(	./.:.:18	1/1	0/0	ADP=39;WT=2	missense_varia	MODERATE	MAP_1432	MAP_1432	AAS03749.1	no KO assigned | (GenBank) hypothetical protein
gen	Chromosome	427039	CTCGGCGTG(	C	PASS	ADP=58;WT=2	GT:GQ:SDP:D	1/1:255:78:75:(	0/0:59:31:31:3	1/1	0/0	ADP=58;WT=2	disruptive_infra	MODERATE	MAP_0403	MAP_0403	AAS02720.1	no KO assigned | (GenBank) hypothetical protein
gen	Chromosome	2172671	A	C	PASS	ADP=48;WT=2	GT:GQ:SDP:D	1/1:255:56:52:(	0/0:55:31:30:3	1/1	0/0	ADP=48;WT=2	missense_varia	MODERATE	glnA2	MAP_1966c	AAS04283.1	K01915 glutamine synthetase [EC:6.3.1.2] | (GenBank) glnA2; GlnA2
gen	Chromosome	2523394	G	C	PASS	ADP=47;WT=2	GT:GQ:SDP:D	1/1:249:61:58:(	0/0:40:23:22:2	1/1	0/0	ADP=47;WT=2	missense_varia	MODERATE	MAP_2258c	MAP_2258c	AAS04575.1	K01950 NAD+ synthase (glutamine-hydrolysing) [EC:6.3.5.1] | (GenBank) hypothetical protein

Chromosome	19072	A	AC	PASS	ADP=42;WT=(	GT:GQ:SDP:D	1/1:255:66:62:(	1/1:141:25:24:(	1/1	1/1							
Chromosome	104639	CG	C	PASS	ADP=94;WT=(	GT:GQ:SDP:D	1/1:255:134:12	1/1:231:43:42::	1/1	1/1							
Chromosome	112809	AGCACTATCC	A	PASS	ADP=92;WT=(	GT:GQ:SDP:D	1/1:255:137:13	1/1:147:37:36:(	1/1	1/1							
Chromosome	131778	C	CCGG	PASS	ADP=29;WT=(	GT:GQ:SDP:D	1/1:191:48:44:(	1/1:103:21:18::	1/1	1/1							
Chromosome	148110	TC	T	PASS	ADP=56;WT=(	GT:GQ:SDP:D	1/1:255:61:59:(	1/1:133:30:29:(	1/1	1/1							
Chromosome	485939	CG	C	PASS	ADP=50;WT=(	GT:GQ:SDP:D	1/1:255:74:70:(	1/1:173:33:32::	1/1	1/1							
Chromosome	499159	G	GGCGACTCA(	PASS	ADP=40;WT=(	GT:GQ:SDP:D	0/1:118:49:44:(	0/1:97:30:30:6:	0/1	0/1							
Chromosome	683510	GGTGTGTTCC	G	PASS	ADP=73;WT=(	GT:GQ:SDP:D	1/1:255:88:87:(	1/1:110:29:29:(	1/1	1/1							
Chromosome	850668	GC	G	PASS	ADP=46;WT=(	GT:GQ:SDP:D	1/1:255:68:60:(	1/1:116:27:27::	1/1	1/1							
Chromosome	944996	T	TGAG	PASS	ADP=44;WT=(	GT:GQ:SDP:D	1/1:204:58:55:(	1/1:95:24:21:3:	1/1	1/1							
Chromosome	989250	G	GC	PASS	ADP=42;WT=(	GT:GQ:SDP:D	1/1:255:60:55:(	1/1:103:22:21::	1/1	1/1							
Chromosome	1028127	GGGT	G	PASS	ADP=51;WT=(	GT:GQ:SDP:D	1/1:255:69:66:(	1/1:105:28:26:(	1/1	1/1							
Chromosome	1116253	AT	A	PASS	ADP=35;WT=(	GT:GQ:SDP:D	1/1:255:49:46:(	1/1:117:21:18:(	1/1	1/1							
Chromosome	1308431	G	GCC	PASS	ADP=90;WT=(	GT:GQ:SDP:D	1/1:255:131:12	1/1:158:40:38:(	1/1	1/1							
Chromosome	1318322	GA	G	PASS	ADP=40;WT=(	GT:GQ:SDP:D	1/1:255:52:51:(	1/1:111:20:20:(	1/1	1/1							
Chromosome	1457853	CG	C	PASS	ADP=52;WT=(	GT:GQ:SDP:D	1/1:255:66:65:(	1/1:135:24:23:(	1/1	1/1							
Chromosome	1793089	GCCCCCCCC(	G	PASS	ADP=68;WT=(	GT:GQ:SDP:D	1/1:255:87:84:(	1/1:214:41:39::	1/1	1/1							
Chromosome	1848410	AC	A	PASS	ADP=44;WT=(	GT:GQ:SDP:D	1/1:235:51:50:(	1/1:158:28:28:(	1/1	1/1							
Chromosome	1864323	AC	A	PASS	ADP=76;WT=(	GT:GQ:SDP:D	1/1:255:107:10	1/1:248:43:43:(	1/1	1/1							
Chromosome	1886935	GC	G	PASS	ADP=107;WT=	GT:GQ:SDP:D	1/1:255:145:13	1/1:255:53:50::	1/1	1/1							
Chromosome	1890290	AC	A	PASS	ADP=80;WT=(	GT:GQ:SDP:D	1/1:255:126:12	1/1:141:25:23:(	1/1	1/1							
Chromosome	1914846	TG	T	PASS	ADP=51;WT=(	GT:GQ:SDP:D	1/1:255:60:59:(	1/1:111:20:20:(	1/1	1/1							
Chromosome	2300420	GC	G	PASS	ADP=43;WT=(	GT:GQ:SDP:D	1/1:255:61:58:(	1/1:111:20:20:(	1/1	1/1							
Chromosome	2757198	GC	G	PASS	ADP=86;WT=(	GT:GQ:SDP:D	1/1:255:107:10	1/1:255:51:50::	1/1	1/1							
Chromosome	2874398	G	GC	PASS	ADP=53;WT=(	GT:GQ:SDP:D	1/1:255:75:75:(	1/1:141:25:25:(	1/1	1/1							
Chromosome	2974559	C	CG	PASS	ADP=43;WT=(	GT:GQ:SDP:D	1/1:255:52:52:(	1/1:109:22:22::	1/1	1/1							
Chromosome	3142424	C	CG	PASS	ADP=52;WT=(	GT:GQ:SDP:D	1/1:255:72:72:(	1/1:129:23:22:(	1/1	1/1							
Chromosome	3225725	TC	T	PASS	ADP=47;WT=(	GT:GQ:SDP:D	1/1:255:64:60:(	1/1:138:31:29:(	1/1	1/1							
Chromosome	3300014	TC	T	PASS	ADP=57;WT=(	GT:GQ:SDP:D	1/1:255:76:71:(	1/1:126:28:25:(	1/1	1/1							
Chromosome	3410746	G	GC	PASS	ADP=47;WT=(	GT:GQ:SDP:D	1/1:255:60:56:(	1/1:135:24:23:(	1/1	1/1							
Chromosome	3417490	GC	G	PASS	ADP=46;WT=(	GT:GQ:SDP:D	1/1:242:60:58:(	1/1:98:22:20:2:	1/1	1/1							
Chromosome	3558719	AC	A	PASS	ADP=45;WT=(	GT:GQ:SDP:D	1/1:255:67:62:(	1/1:126:25:23:(	1/1	1/1							
Chromosome	3558726	TG	T	PASS	ADP=46;WT=(	GT:GQ:SDP:D	1/1:255:67:65:(	1/1:135:24:24:(	1/1	1/1							
Chromosome	3558824	AC	A	PASS	ADP=45;WT=(	GT:GQ:SDP:D	1/1:255:58:55:(	1/1:138:29:28:(	1/1	1/1							
Chromosome	3607486	C	CT	PASS	ADP=47;WT=(	GT:GQ:SDP:D	1/1:255:61:59:(	1/1:109:22:21:(	1/1	1/1							
Chromosome	3994447	GC	G	PASS	ADP=49;WT=(	GT:GQ:SDP:D	1/1:255:62:58:(	1/1:123:23:22:(	1/1	1/1							
Chromosome	3999115	GA	G	PASS	ADP=58;WT=(	GT:GQ:SDP:D	1/1:255:93:89:(	1/1:158:28:28:(	1/1	1/1							
Chromosome	3999142	TG	T	PASS	ADP=56;WT=(	GT:GQ:SDP:D	1/1:255:86:79:(	1/1:152:27:26:(	1/1	1/1							
Chromosome	4140707	GC	G	PASS	ADP=68;WT=(	GT:GQ:SDP:D	1/1:255:102:10	1/1:126:25:25::	1/1	1/1							
Chromosome	4409626	TGG	T	PASS	ADP=65;WT=(	GT:GQ:SDP:D	1/1:255:89:86:(	1/1:138:29:28::	1/1	1/1							
Chromosome	4417540	G	GA	PASS	ADP=64;WT=(	GT:GQ:SDP:D	1/1:255:95:91:(	1/1:173:33:32::	1/1	1/1							
Chromosome	4422698	C	CG	PASS	ADP=78;WT=(	GT:GQ:SDP:D	1/1:255:102:98	1/1:255:47:46::	1/1	1/1							
Chromosome	4553473	AC	A	PASS	ADP=56;WT=(	GT:GQ:SDP:D	1/1:255:70:68:(	1/1:141:25:25:(	1/1	1/1							
Chromosome	4577679	AC	A	PASS	ADP=65;WT=(	GT:GQ:SDP:D	1/1:255:114:10	1/1:123:22:21:(	1/1	1/1							
Chromosome	4634485	CG	C	PASS	ADP=42;WT=(	GT:GQ:SDP:D	1/1:255:61:59:(	1/1:132:26:26::	1/1	1/1							
Chromosome	4691109	CG	C	PASS	ADP=58;WT=(	GT:GQ:SDP:D	1/1:255:84:78:(	1/1:255:45:43:(	1/1	1/1							
Chromosome	4713463	TG	T	PASS	ADP=49;WT=(	GT:GQ:SDP:D	1/1:255:74:70:(	1/1:111:26:25::	1/1	1/1							
Chromosome	4827081	AC	A	PASS	ADP=49;WT=(	GT:GQ:SDP:D	1/1:255:56:54:(	1/1:146:26:24:(	1/1	1/1							
Chromosome	11156	C	A	PASS	ADP=106;WT=	GT:GQ:SDP:D	0/1:255:143:14	0/1:108:43:42::	0/1	0/1							
Chromosome	11180	G	T	PASS	ADP=60;WT=(	GT:GQ:SDP:D	1/1:255:81:79:(	1/1:170:31:30:(	1/1	1/1							
Chromosome	110114	C	T	PASS	ADP=84;WT=(	GT:GQ:SDP:D	1/1:255:122:12	1/1:236:41:41:(	1/1	1/1							
Chromosome	114907	G	A	PASS	ADP=85;WT=(	GT:GQ:SDP:D	1/1:255:98:93:(	1/1:255:47:46:(	1/1	1/1							
Chromosome	165498	T	C	PASS	ADP=70;WT=(	GT:GQ:SDP:D	1/1:255:94:91:(	1/1:164:30:29:(	1/1	1/1							
Chromosome	223371	A	C	PASS	ADP=47;WT=(	GT:GQ:SDP:D	1/1:255:56:54:(	1/1:146:26:26:(	1/1	1/1							
Chromosome	300301	T	C	PASS	ADP=83;WT=(	GT:GQ:SDP:D	1/1:255:111:10	1/1:170:30:30:(	1/1	1/1							
Chromosome	300672	G	A	PASS	ADP=97;WT=(	GT:GQ:SDP:D	1/1:255:151:14	1/1:224:40:39:(	1/1	1/1							
Chromosome	322326	T	G	PASS	ADP=56;WT=(	GT:GQ:SDP:D	1/1:255:74:70:(	1/1:164:30:29:(	1/1	1/1							
Chromosome	338241	T	C	PASS	ADP=49;WT=(	GT:GQ:SDP:D	1/1:255:62:61:(	1/1:117:23:21:(	1/1	1/1							
Chromosome	449903	C	T	PASS	ADP=56;WT=(	GT:GQ:SDP:D	1/1:255:68:65:(	1/1:152:27:27:(	1/1	1/1							
Chromosome	470098	G	A	PASS	ADP=36;WT=(	GT:GQ:SDP:D	1/1:255:47:45:(	1/1:158:29:28:(	1/1	1/1							
Chromosome	470099	G	C	PASS	ADP=37;WT=(	GT:GQ:SDP:D	1/1:254:47:44:(	1/1:164:29:29:(	1/1	1/1							
Chromosome	585994	G	T	PASS	ADP=61;WT=(	GT:GQ:SDP:D	1/1:255:83:81:(	1/1:200:37:35:(	1/1	1/1							
Chromosome	618652	C	G	PASS	ADP=66;WT=(	GT:GQ:SDP:D	1/1:255:93:92:(	1/1:152:29:27:(	1/1	1/1							
Chromosome	631842	T	C	PASS	ADP=49;WT=(	GT:GQ:SDP:D	1/1:255:81:77:(	1/1:123:24:22:(	1/1	1/1							
Chromosome	644663	A	G	PASS	ADP=33;WT=(	GT:GQ:SDP:D	1/1:255:48:46:(	1/1:111:21:20:(	1/1	1/1							
Chromosome	660167	G	T	PASS	ADP=42;WT=(	GT:GQ:SDP:D	1/1:255:57:54:(	1/1:135:27:24:(	1/1	1/1							
Chromosome	680826	T	C	PASS	ADP=44;WT=(	GT:GQ:SDP:D	1/1:255:66:65:(	1/1:111:20:20:(	1/1	1/1							
Chromosome	685570	A	C	PASS	ADP=76;WT=(	GT:GQ:SDP:D	1/1:255:103:98	1/1:218:38:38:(	1/1	1/1							
Chromosome	698421	C	T	PASS	ADP=37;WT=(	GT:GQ:SDP:D	1/1:255:46:46:(	1/1:111:20:20:(	1/1	1/1							
Chromosome	700060	G	A	PASS	ADP=67;WT=(	GT:GQ:SDP:D	1/1:255:102:10	1/1:194:35:34:(	1/1	1/1							
Chromosome	722272	G	C	PASS	ADP=56;WT=(	GT:GQ:SDP:D	1/1:255:83:80:(	1/1:152:30:28:(	1/1	1/1							
Chromosome	730808	A	G	PASS	ADP=49;WT=(	GT:GQ:SDP:D	1/1:255:61:59:(	1/1:164:29:29:(	1/1	1/1							
Chromosome	736129	A	C	PASS	ADP=59;WT=(	GT:GQ:SDP:D	1/1:255:73:73:(	1/1:152:27:27:(	1/1	1/1							
Chromosome	784351	C	T	PASS	ADP=94;WT=(	GT:GQ:SDP:D	1/1:255:129:12	1/1:212:37:37:(	1/1	1/1							
Chromosome	811828	G	T	PASS	ADP=37;WT=(	GT:GQ:SDP:D	1/1:254:45:44:(	1/1:152:27:27:(	1/1	1/1							
Chromosome	837573	C	A	PASS	ADP=55;WT=(	GT:GQ:SDP:D	1/1:255:71:70:(	1/1:164:29:29:(	1/1	1/1							
Chromosome	893214	C	T	PASS	ADP=63;WT=(	GT:GQ:SDP:D	1/1:255:89:82:(	1/1:188:34:33:(	1/1	1/1							
Chromosome	899149	A	G	PASS	ADP=59;WT=(	GT:GQ:SDP:D	1/1:255:69:66:(	1/1:170:30:30:(	1/1	1/1							
Chromosome	932096	T	A	PASS	ADP=48;WT=(	GT:GQ:SDP:D	1/1:255:60:58:(	1/1:135:26:24:(	1/1	1/1							
Chromosome	999683	T	C	PASS	ADP=49;WT=(	GT:GQ:SDP:D	1/1:255:61:59:(	1/1:182:32:32:(	1/1	1/1							
Chromosome	1018152	T	C	PASS	ADP=57;WT=(	GT:GQ:SDP:D	1/1:255:85:82:(	1/1:146:27:26:(	1/1	1/1							
Chromosome	1053678	G	A	PASS	ADP=45;WT=(	GT:GQ:SDP:D	1/1:255:78:74:(	1/1:111:22:20:(	1/1	1/1							
Chromosome	1113270	T	C	PASS	ADP=55;WT=(	GT:GQ:SDP:D	1/1:255:71:69:(	1/1:200:36:35:(	1/1	1/1							
Chromosome	1144837	G	A	PASS	ADP=48;WT=(	GT:GQ:SDP:D	1/1:255:68:67:(	1/1:158:28:28:(	1/1	1/1							
Chromosome	1276473	A	G	PASS	ADP=65;WT=(	GT:GQ:SDP:D	1/1:255:107:96	1/1:152:27:27:(	1/1	1/1							
Chromosome	1337904	C	T	PASS	ADP=57;WT=(	GT:GQ:SDP:D	1/1:255:66:66:(	1/1:158:28:28:(	1/1	1/1							

Journal Pre-proof

Chromosome	1367709	A	G	PASS	ADP=55;WT=0	GT:GQ:SDP:D	1/1:255:64:63:(	1/1:123:22:22:(1/1	1/1
Chromosome	1374360	G	A	PASS	ADP=50;WT=0	GT:GQ:SDP:D	1/1:255:71:71:(	1/1:135:25:24:(1/1	1/1
Chromosome	1478080	C	T	PASS	ADP=81;WT=0	GT:GQ:SDP:D	0/1:185:108:10	0/1:76:48:46:2! 0/1	0/1
Chromosome	1478087	T	C	PASS	ADP=88;WT=0	GT:GQ:SDP:D	0/1:220:113:11	0/1:88:51:50:2(0/1	0/1
Chromosome	1478099	T	C	PASS	ADP=89;WT=0	GT:GQ:SDP:D	0/1:214:116:11	0/1:89:51:48:2- 0/1	0/1
Chromosome	1478117	C	G	PASS	ADP=88;WT=0	GT:GQ:SDP:D	0/1:228:118:11	0/1:98:50:48:2; 0/1	0/1
Chromosome	1478125	C	T	PASS	ADP=85;WT=0	GT:GQ:SDP:D	0/1:217:112:10	0/1:94:49:47:2; 0/1	0/1
Chromosome	1478699	T	C	PASS	ADP=86;WT=0	GT:GQ:SDP:D	0/1:161:114:10	0/1:75:52:49:2(0/1	0/1
Chromosome	1478747	G	C	PASS	ADP=86;WT=0	GT:GQ:SDP:D	0/1:145:106:10	0/1:74:53:51:3(0/1	0/1
Chromosome	1478750	C	T	PASS	ADP=85;WT=0	GT:GQ:SDP:D	0/1:141:105:10	0/1:79:52:50:2(0/1	0/1
Chromosome	1478751	A	G	PASS	ADP=86;WT=0	GT:GQ:SDP:D	0/1:149:105:10	0/1:74:52:50:2(0/1	0/1
Chromosome	1478756	C	G	PASS	ADP=89;WT=0	GT:GQ:SDP:D	0/1:148:107:10	0/1:78:51:51:2(0/1	0/1
Chromosome	1478764	C	G	PASS	ADP=86;WT=0	GT:GQ:SDP:D	0/1:140:109:10	0/1:74:52:51:3(0/1	0/1
Chromosome	1478765	C	G	PASS	ADP=86;WT=0	GT:GQ:SDP:D	0/1:144:110:10	0/1:75:52:49:2(0/1	0/1
Chromosome	1478797	G	C	PASS	ADP=82;WT=0	GT:GQ:SDP:D	0/1:144:111:10	0/1:70:54:51:3(0/1	0/1
Chromosome	1500019	C	A	PASS	ADP=43;WT=0	GT:GQ:SDP:D	1/1:255:66:62:(	1/1:176:31:31:(1/1	1/1
Chromosome	1501466	A	G	PASS	ADP=48;WT=0	GT:GQ:SDP:D	1/1:255:61:59:(	1/1:135:24:24:(1/1	1/1
Chromosome	1512466	G	A	PASS	ADP=40;WT=0	GT:GQ:SDP:D	1/1:255:54:47:(	1/1:146:27:26:(1/1	1/1
Chromosome	1564977	G	C	PASS	ADP=65;WT=0	GT:GQ:SDP:D	0/1:185:78:78:3	0/1:93:38:35:1(0/1	0/1
Chromosome	1564980	A	G	PASS	ADP=65;WT=0	GT:GQ:SDP:D	0/1:172:78:74:3	0/1:96:38:37:1(0/1	0/1
Chromosome	1577472	T	C	PASS	ADP=69;WT=0	GT:GQ:SDP:D	1/1:255:91:88:(	1/1:170:31:30:(1/1	1/1
Chromosome	1590491	A	G	PASS	ADP=53;WT=0	GT:GQ:SDP:D	1/1:255:67:65:(	1/1:212:37:37:(1/1	1/1
Chromosome	1597139	C	T	PASS	ADP=31;WT=0	GT:GQ:SDP:D	1/1:206:39:36:(	1/1:111:21:20:(1/1	1/1
Chromosome	1601193	T	C	PASS	ADP=55;WT=0	GT:GQ:SDP:D	1/1:255:68:66:(	1/1:123:22:22:(1/1	1/1
Chromosome	1624240	A	T	PASS	ADP=57;WT=0	GT:GQ:SDP:D	1/1:255:74:72:(	1/1:111:20:20:(1/1	1/1
Chromosome	1633620	G	A	PASS	ADP=58;WT=0	GT:GQ:SDP:D	1/1:255:77:76:(	1/1:135:25:24:(1/1	1/1
Chromosome	1635694	T	C	PASS	ADP=30;WT=0	GT:GQ:SDP:D	1/1:188:36:33:(	1/1:117:21:21:(1/1	1/1
Chromosome	1652200	T	C	PASS	ADP=47;WT=0	GT:GQ:SDP:D	1/1:255:59:58:(	1/1:123:23:22:(1/1	1/1
Chromosome	1660589	G	A	PASS	ADP=51;WT=0	GT:GQ:SDP:D	1/1:255:68:66:(	1/1:117:21:21:(1/1	1/1
Chromosome	1680930	G	A	PASS	ADP=78;WT=0	GT:GQ:SDP:D	0/1:129:92:84:4	0/1:72:46:44:2(0/1	0/1
Chromosome	1710661	C	A	PASS	ADP=55;WT=0	GT:GQ:SDP:D	1/1:255:80:78:(	1/1:135:26:24:(1/1	1/1
Chromosome	1740290	G	A	PASS	ADP=44;WT=0	GT:GQ:SDP:D	1/1:255:53:49:(	1/1:129:24:23:(1/1	1/1
Chromosome	1764683	G	C	PASS	ADP=56;WT=0	GT:GQ:SDP:D	1/1:255:73:70:(	1/1:212:37:37:(1/1	1/1
Chromosome	1855963	T	C	PASS	ADP=38;WT=0	GT:GQ:SDP:D	1/1:255:49:45:(	1/1:141:27:25:(1/1	1/1
Chromosome	1872321	T	C	PASS	ADP=89;WT=0	GT:GQ:SDP:D	1/1:255:121:11	1/1:255:48:47:(1/1	1/1
Chromosome	1875266	G	A	PASS	ADP=74;WT=0	GT:GQ:SDP:D	1/1:255:90:88:(	1/1:194:35:34:(1/1	1/1
Chromosome	1948228	C	T	PASS	ADP=74;WT=0	GT:GQ:SDP:D	1/1:255:106:99	1/1:188:36:33:(1/1	1/1
Chromosome	1971639	C	A	PASS	ADP=37;WT=0	GT:GQ:SDP:D	1/1:255:48:47:3	1/1:141:25:25:(1/1	1/1
Chromosome	1974756	G	A	PASS	ADP=48;WT=0	GT:GQ:SDP:D	1/1:255:68:65:(	1/1:123:23:22:(1/1	1/1
Chromosome	1989123	C	T	PASS	ADP=45;WT=0	GT:GQ:SDP:D	1/1:255:59:53:(	1/1:146:26:26:(1/1	1/1
Chromosome	2038456	G	C	PASS	ADP=68;WT=0	GT:GQ:SDP:D	1/1:255:94:89:(	1/1:176:32:31:(1/1	1/1
Chromosome	2042896	G	A	PASS	ADP=52;WT=0	GT:GQ:SDP:D	1/1:255:72:66:(	1/1:188:34:33:(1/1	1/1
Chromosome	2127817	A	G	PASS	ADP=65;WT=0	GT:GQ:SDP:D	1/1:255:92:89:(	1/1:188:35:33:(1/1	1/1
Chromosome	2148194	A	C	PASS	ADP=45;WT=0	GT:GQ:SDP:D	1/1:255:57:56:(	1/1:182:34:32:(1/1	1/1
Chromosome	2151343	A	G	PASS	ADP=38;WT=0	GT:GQ:SDP:D	1/1:248:45:43:(	1/1:141:25:25:(1/1	1/1
Chromosome	2184050	G	A	PASS	ADP=32;WT=0	GT:GQ:SDP:D	1/1:230:40:40:(	1/1:117:21:21:(1/1	1/1
Chromosome	2246909	T	C	PASS	ADP=33;WT=0	GT:GQ:SDP:D	1/1:194:34:34:(	1/1:111:22:20:(1/1	1/1
Chromosome	2254759	A	G	PASS	ADP=68;WT=0	GT:GQ:SDP:D	1/1:255:91:89:(	1/1:158:31:28:(1/1	1/1
Chromosome	2255698	A	C	PASS	ADP=49;WT=0	GT:GQ:SDP:D	1/1:255:70:67:(	1/1:141:25:25:(1/1	1/1
Chromosome	2270313	C	T	PASS	ADP=70;WT=0	GT:GQ:SDP:D	1/1:255:87:84:(	1/1:182:32:32:(1/1	1/1
Chromosome	2276371	A	G	PASS	ADP=61;WT=0	GT:GQ:SDP:D	1/1:255:75:71:(	1/1:224:39:39:(1/1	1/1
Chromosome	2298328	C	G	PASS	ADP=100;WT=	GT:GQ:SDP:D	1/1:255:139:13	1/1:255:52:47:(1/1	1/1
Chromosome	2300526	A	G	PASS	ADP=40;WT=0	GT:GQ:SDP:D	1/1:255:59:55:(	1/1:123:22:22:(1/1	1/1
Chromosome	2305231	A	C	PASS	ADP=82;WT=0	GT:GQ:SDP:D	1/1:255:120:11	1/1:158:31:28:(1/1	1/1
Chromosome	2305608	G	C	PASS	ADP=66;WT=0	GT:GQ:SDP:D	1/1:255:96:91:(	1/1:152:30:27:(1/1	1/1
Chromosome	2316683	A	G	PASS	ADP=49;WT=0	GT:GQ:SDP:D	1/1:255:76:73:(	1/1:129:24:23:(1/1	1/1
Chromosome	2324244	C	T	PASS	ADP=52;WT=0	GT:GQ:SDP:D	1/1:255:69:66:(	1/1:135:26:24:(1/1	1/1
Chromosome	2339449	G	A	PASS	ADP=59;WT=0	GT:GQ:SDP:D	1/1:255:77:74:(	1/1:135:24:24:(1/1	1/1
Chromosome	2414957	G	T	PASS	ADP=44;WT=0	GT:GQ:SDP:D	1/1:255:56:53:(	1/1:146:28:26:(1/1	1/1
Chromosome	2436501	G	T	PASS	ADP=52;WT=0	GT:GQ:SDP:D	1/1:255:67:60:(	1/1:117:22:21:(1/1	1/1
Chromosome	2437484	C	G	PASS	ADP=70;WT=0	GT:GQ:SDP:D	1/1:255:99:92:(	1/1:212:38:37:(1/1	1/1
Chromosome	2446645	T	C	PASS	ADP=72;WT=0	GT:GQ:SDP:D	1/1:255:90:90:(	1/1:152:27:27:(1/1	1/1
Chromosome	2447624	C	G	PASS	ADP=67;WT=0	GT:GQ:SDP:D	1/1:255:93:88:(	1/1:123:23:22:(1/1	1/1
Chromosome	2483684	G	A	PASS	ADP=52;WT=0	GT:GQ:SDP:D	1/1:255:67:62:(	1/1:141:25:25:(1/1	1/1
Chromosome	2511188	G	C	PASS	ADP=57;WT=0	GT:GQ:SDP:D	1/1:255:72:71:(	1/1:152:29:27:(1/1	1/1
Chromosome	2524801	A	G	PASS	ADP=42;WT=0	GT:GQ:SDP:D	1/1:255:54:51:3	1/1:135:24:24:(1/1	1/1
Chromosome	2550325	G	C	PASS	ADP=57;WT=0	GT:GQ:SDP:D	1/1:255:84:80:3	1/1:170:31:30:(1/1	1/1
Chromosome	2573496	A	G	PASS	ADP=42;WT=0	GT:GQ:SDP:D	1/1:255:62:61:(	1/1:129:23:23:(1/1	1/1
Chromosome	2617404	G	C	PASS	ADP=52;WT=0	GT:GQ:SDP:D	1/1:255:73:69:(	1/1:170:32:30:(1/1	1/1
Chromosome	2650339	G	C	PASS	ADP=76;WT=0	GT:GQ:SDP:D	1/1:255:102:99	1/1:141:27:25:(1/1	1/1
Chromosome	2652703	G	A	PASS	ADP=57;WT=0	GT:GQ:SDP:D	1/1:255:76:76:(	1/1:135:24:24:(1/1	1/1
Chromosome	2687631	C	G	PASS	ADP=71;WT=0	GT:GQ:SDP:D	1/1:255:107:10	1/1:206:37:36:(1/1	1/1
Chromosome	2687632	T	G	PASS	ADP=71;WT=0	GT:GQ:SDP:D	1/1:255:107:10	1/1:206:37:36:(1/1	1/1
Chromosome	2687633	C	T	PASS	ADP=69;WT=0	GT:GQ:SDP:D	1/1:255:107:10	1/1:200:37:35:(1/1	1/1
Chromosome	2696802	T	C	PASS	ADP=62;WT=0	GT:GQ:SDP:D	1/1:255:91:86:(	1/1:188:35:33:(1/1	1/1
Chromosome	2704641	A	G	PASS	ADP=52;WT=0	GT:GQ:SDP:D	1/1:255:74:70:(	1/1:170:31:30:(1/1	1/1
Chromosome	2779463	G	C	PASS	ADP=65;WT=0	GT:GQ:SDP:D	1/1:255:100:99	1/1:200:37:35:(1/1	1/1
Chromosome	2798291	G	C	PASS	ADP=61;WT=0	GT:GQ:SDP:D	1/1:255:100:98	1/1:141:26:25:(1/1	1/1
Chromosome	2807928	T	C	PASS	ADP=78;WT=0	GT:GQ:SDP:D	1/1:255:111:10	1/1:188:36:33:(1/1	1/1
Chromosome	2820410	A	G	PASS	ADP=47;WT=0	GT:GQ:SDP:D	1/1:255:66:63:(	1/1:111:20:20:(1/1	1/1
Chromosome	2832009	T	G	PASS	ADP=61;WT=0	GT:GQ:SDP:D	1/1:255:85:83:(	1/1:236:41:41:(1/1	1/1
Chromosome	2909852	G	A	PASS	ADP=66;WT=0	GT:GQ:SDP:D	1/1:255:96:92:(	1/1:141:25:25:(1/1	1/1
Chromosome	2923499	G	A	PASS	ADP=61;WT=0	GT:GQ:SDP:D	1/1:255:78:74:(	1/1:188:34:33:(1/1	1/1
Chromosome	2927919	G	T	PASS	ADP=33;WT=0	GT:GQ:SDP:D	1/1:218:40:38:(	1/1:109:22:22:3	1/1	1/1
Chromosome	2977087	A	C	PASS	ADP=52;WT=0	GT:GQ:SDP:D	1/1:255:65:64:(	1/1:141:25:25:(1/1	1/1
Chromosome	2983418	T	C	PASS	ADP=37;WT=0	GT:GQ:SDP:D	1/1:255:55:53:(	1/1:135:25:24:(1/1	1/1
Chromosome	3009785	G	C	PASS	ADP=55;WT=0	GT:GQ:SDP:D	1/1:255:84:82:(	1/1:141:26:25:(1/1	1/1
Chromosome	3017644	A	G	PASS	ADP=52;WT=0	GT:GQ:SDP:D	1/1:255:58:57:(	1/1:242:43:42:(1/1	1/1
Chromosome	3055743	T	A	PASS	ADP=38;WT=0	GT:GQ:SDP:D	1/1:236:45:41:(	1/1:176:33:31:(1/1	1/1
Chromosome	3082376	G	A	PASS	ADP=58;WT=0	GT:GQ:SDP:D	1/1:255:82:78:(	1/1:135:26:24:(1/1	1/1
Chromosome	3110789	C	T	PASS	ADP=50;WT=0	GT:GQ:SDP:D	1/1:255:67:67:(	1/1:152:27:27:(1/1	1/1
Chromosome	3110922	G	A	PASS	ADP=61;WT=0	GT:GQ:SDP:D	1/1:255:93:85:(	1/1:206:37:36:(1/1	1/1
Chromosome	3124158	C	T	PASS	ADP=54;WT=0	GT:GQ:SDP:D	1/1:255:65:59:(	1/1:146:26:26:(1/1	1/1
Chromosome	3149744	C	T	PASS	ADP=55;WT=0	GT:GQ:SDP:D	1/1:255:70:69:(	1/1:152:30:27:(1/1	1/1
Chromosome	3156400	G	A	PASS	ADP=51;WT=0	GT:GQ:SDP:D	1/1:255:65:61:(	1/1:194:35:34:(1/1	1/1
Chromosome	3189889	C	T	PASS	ADP=54;WT=0	GT:GQ:SDP:D	1/1:255:71:69:(	1/1:164:31:29:(1/1	1/1
Chromosome	3230371	G	A	PASS	ADP=55;WT=0	GT:GQ:SDP:D	1/1:255:77:73:(	1/1:158:30:28:(1/1	1/1
Chromosome	3312429	G	C	PASS	ADP=54;WT=0	GT:GQ:SDP:D	1/1:255:81:77:(	1/1:123:22:22:(1/1	1/1
Chromosome	3349253	G	A	PASS	ADP=58;WT=0	GT:GQ:SDP:D	1/1:255:76:73:(	1/1:212:40:37:(1/1	1/1
Chromosome	3390128	G	A	PASS	ADP=55;WT=0	GT:GQ:SDP:D	1/1:255:69:68:(	1/1:176:32:31:(1/1	1/1
Chromosome	3417220	C	T	PASS	ADP=54;WT=0	GT:GQ:SDP:D	1/1:255:75:71:(	1/1:129:25:23:(1/1	1/1
Chromosome	3440365	C	T	PASS	ADP=45;WT=0	GT:GQ:SDP:D	1/1:255:61:58:(	1/1:146:29:26:(1/1	1/1
Chromosome	3487960	T	C	PASS	ADP=54;WT=0	GT:GQ:SDP:D	1/1:255:85:83:(	1/1:164:29:29:(1/1	1/1
Chromosome	3517668	C	G	PASS	ADP=64;WT=0	GT:GQ:SDP:D	1/1:255:84:82:(	1/1:176:31:31:(1/1	1/1
Chromosome	3653206	C	T	PASS	ADP=57;WT=0	GT:GQ:SDP:D	1/1:255:83:83:(	1/1:117:22:21:(1/1	1/1
Chromosome	3670614	C	G	PASS	ADP=48;WT=0	GT:GQ:SDP:D	1/1:255:67:63:3	1/1:212:37:37:(1/1	1/1
Chromosome	3670615	G	C	PASS	ADP=48;WT=0	GT:GQ:SDP:D	1/1:255:67:64:(	1/1:200:37:35:(1/1	1/1

Journal Pre-proof

Chromosome	3704178	T	C	PASS	ADP=53;WT=0	GT:GQ:SDP:D	1/1:255:73:70:(	1/1:188:34:33:(	1/1	1/1
Chromosome	3747086	G	T	PASS	ADP=57;WT=0	GT:GQ:SDP:D	1/1:255:85:82:(	1/1:158:30:28:(	1/1	1/1
Chromosome	3767913	T	G	PASS	ADP=54;WT=0	GT:GQ:SDP:D	1/1:255:70:67:(	1/1:135:24:24:(	1/1	1/1
Chromosome	3767939	G	C	PASS	ADP=55;WT=0	GT:GQ:SDP:D	1/1:255:75:70:(	1/1:123:23:22:(	1/1	1/1
Chromosome	3809323	C	T	PASS	ADP=52;WT=0	GT:GQ:SDP:D	1/1:255:73:64:(	1/1:200:38:35:(	1/1	1/1
Chromosome	3902477	T	C	PASS	ADP=55;WT=0	GT:GQ:SDP:D	1/1:255:63:62:(	1/1:158:29:28:(	1/1	1/1
Chromosome	3946781	C	G	PASS	ADP=52;WT=0	GT:GQ:SDP:D	1/1:255:64:62:(	1/1:164:34:29:(	1/1	1/1
Chromosome	3965578	G	C	PASS	ADP=52;WT=0	GT:GQ:SDP:D	1/1:255:60:59:(	1/1:182:32:32:(	1/1	1/1
Chromosome	3968313	C	T	PASS	ADP=55;WT=0	GT:GQ:SDP:D	1/1:255:70:68:(	1/1:111:20:20:(	1/1	1/1
Chromosome	3975317	G	A	PASS	ADP=49;WT=0	GT:GQ:SDP:D	1/1:255:71:65:(	1/1:117:22:21:(	1/1	1/1
Chromosome	4047091	T	C	PASS	ADP=41;WT=0	GT:GQ:SDP:D	1/1:255:51:50:(	1/1:176:33:31:(	1/1	1/1
Chromosome	4060110	C	G	PASS	ADP=57;WT=0	GT:GQ:SDP:D	1/1:255:87:83:(	1/1:164:30:29:(	1/1	1/1
Chromosome	4068132	C	G	PASS	ADP=37;WT=0	GT:GQ:SDP:D	1/1:248:45:43:(	1/1:117:23:21:(	1/1	1/1
Chromosome	4083393	T	A	PASS	ADP=49;WT=0	GT:GQ:SDP:D	1/1:255:70:67:(	1/1:123:22:22:(	1/1	1/1
Chromosome	4087849	T	C	PASS	ADP=55;WT=0	GT:GQ:SDP:D	1/1:255:79:76:(	1/1:182:33:32:(	1/1	1/1
Chromosome	4110302	G	A	PASS	ADP=56;WT=0	GT:GQ:SDP:D	1/1:255:73:70:(	1/1:176:31:31:(	1/1	1/1
Chromosome	4119657	T	C	PASS	ADP=62;WT=0	GT:GQ:SDP:D	1/1:255:84:81:(	1/1:236:43:41:(	1/1	1/1
Chromosome	4123525	T	C	PASS	ADP=48;WT=0	GT:GQ:SDP:D	1/1:255:62:58:(	1/1:141:26:25:(	1/1	1/1
Chromosome	4123912	G	A	PASS	ADP=44;WT=0	GT:GQ:SDP:D	1/1:255:68:62:(	1/1:146:26:26:(	1/1	1/1
Chromosome	4137756	A	G	PASS	ADP=73;WT=0	GT:GQ:SDP:D	1/1:255:112:10	1/1:176:34:31:(	1/1	1/1
Chromosome	4203770	C	T	PASS	ADP=100;WT=	GT:GQ:SDP:D	1/1:255:128:12	1/1:218:40:38:(	1/1	1/1
Chromosome	4240799	A	C	PASS	ADP=57;WT=0	GT:GQ:SDP:D	1/1:255:66:64:(	1/1:152:28:27:(	1/1	1/1
Chromosome	4258322	T	G	PASS	ADP=66;WT=0	GT:GQ:SDP:D	1/1:255:88:82:(	1/1:158:28:28:(	1/1	1/1
Chromosome	4258358	T	G	PASS	ADP=62;WT=0	GT:GQ:SDP:D	1/1:255:86:79:(	1/1:158:29:28:(	1/1	1/1
Chromosome	4294179	A	G	PASS	ADP=58;WT=0	GT:GQ:SDP:D	1/1:255:77:76:(	1/1:158:29:28:(	1/1	1/1
Chromosome	4346606	C	G	PASS	ADP=56;WT=0	GT:GQ:SDP:D	1/1:255:69:67:(	1/1:188:33:33:(	1/1	1/1
Chromosome	4370612	G	A	PASS	ADP=54;WT=0	GT:GQ:SDP:D	1/1:255:74:71:(	1/1:152:29:27:(	1/1	1/1
Chromosome	4422706	C	G	PASS	ADP=77;WT=0	GT:GQ:SDP:D	1/1:255:96:92:(	1/1:255:47:47:(	1/1	1/1
Chromosome	4422711	A	G	PASS	ADP=76;WT=0	GT:GQ:SDP:D	1/1:255:94:91:(	1/1:249:48:47:1	1/1	1/1
Chromosome	4422713	G	C	PASS	ADP=76;WT=0	GT:GQ:SDP:D	1/1:255:94:89:(	1/1:255:48:47:(	1/1	1/1
Chromosome	4449793	G	A	PASS	ADP=57;WT=0	GT:GQ:SDP:D	1/1:255:67:64:(	1/1:200:37:35:(	1/1	1/1
Chromosome	4453282	T	C	PASS	ADP=66;WT=0	GT:GQ:SDP:D	1/1:255:83:79:(	1/1:182:32:32:(	1/1	1/1
Chromosome	4507971	G	A	PASS	ADP=49;WT=0	GT:GQ:SDP:D	1/1:255:69:67:(	1/1:117:23:21:(	1/1	1/1
Chromosome	4557531	T	C	PASS	ADP=53;WT=0	GT:GQ:SDP:D	1/1:255:79:72:(	1/1:135:24:24:(	1/1	1/1
Chromosome	4578997	G	C	PASS	ADP=54;WT=0	GT:GQ:SDP:D	1/1:255:64:61:(	1/1:111:21:20:(	1/1	1/1
Chromosome	4586112	C	G	PASS	ADP=62;WT=0	GT:GQ:SDP:D	1/1:255:77:76:(	1/1:182:33:32:(	1/1	1/1
Chromosome	4617550	C	T	PASS	ADP=57;WT=0	GT:GQ:SDP:D	1/1:255:80:68:(	1/1:129:26:23:(	1/1	1/1
Chromosome	4734998	G	T	PASS	ADP=71;WT=0	GT:GQ:SDP:D	1/1:255:100:95	1/1:158:28:28:(	1/1	1/1
Chromosome	4736447	G	A	PASS	ADP=91;WT=0	GT:GQ:SDP:D	1/1:255:117:11	1/1:230:42:40:(	1/1	1/1
Chromosome	4771441	A	T	PASS	ADP=54;WT=0	GT:GQ:SDP:D	1/1:255:77:70:(	1/1:146:28:26:(	1/1	1/1
Chromosome	4771588	G	A	PASS	ADP=51;WT=0	GT:GQ:SDP:D	1/1:255:84:77:(	1/1:123:24:22:(	1/1	1/1
Chromosome	4794819	C	G	PASS	ADP=56;WT=0	GT:GQ:SDP:D	1/1:255:73:70:(	1/1:135:25:24:(	1/1	1/1
Chromosome	4817430	T	C	PASS	ADP=53;WT=0	GT:GQ:SDP:D	1/1:255:65:63:(	1/1:152:28:27:(	1/1	1/1
Chromosome	326891	C	CG	PASS	ADP=40;WT=1	GT:GQ:SDP:D	1/1:255:53:52:2	./.:.:21	1/1	./.
Chromosome	1348418	CG	C	PASS	ADP=40;WT=1	GT:GQ:SDP:D	1/1:255:56:54:2	./.:.:16	1/1	./.
Chromosome	1531282	CG	C	PASS	ADP=40;WT=1	GT:GQ:SDP:D	1/1:231:43:43:1	./.:.:16	1/1	./.
Chromosome	1561642	AG	A	PASS	ADP=38;WT=1	GT:GQ:SDP:D	1/1:255:50:45:2	./.:.:14	1/1	./.
Chromosome	1649604	AC	A	PASS	ADP=40;WT=1	GT:GQ:SDP:D	1/1:255:50:50:2	./.:.:15	1/1	./.
Chromosome	1847089	TC	T	PASS	ADP=59;WT=1	GT:GQ:SDP:D	1/1:255:67:64:2	./.:.:19	1/1	./.
Chromosome	2156148	CCGCCGCGC	C	PASS	ADP=33;WT=1	GT:GQ:SDP:D	0/1:151:51:51:1	./.:.:14	0/1	./.
Chromosome	2436509	G	GTATCCCTCC	PASS	ADP=29;WT=1	GT:GQ:SDP:D	0/1:91:45:42:15	./.:.:10	0/1	./.
Chromosome	2784767	CG	C	PASS	ADP=42;WT=1	GT:GQ:SDP:D	1/1:255:53:50:1	./.:.:20	1/1	./.
Chromosome	2803339	C	CCCT	PASS	ADP=26;WT=0	GT:GQ:SDP:D	1/1:148:45:45:1	./.:.:5	1/1	./.
Chromosome	2893583	GC	G	PASS	ADP=39;WT=1	GT:GQ:SDP:D	1/1:255:52:52:2	./.:.:20	1/1	./.
Chromosome	2955360	ACAG	A	PASS	ADP=42;WT=1	GT:GQ:SDP:D	1/1:244:63:57:1	./.:.:15	1/1	./.
Chromosome	3412848	TG	T	PASS	ADP=36;WT=1	GT:GQ:SDP:D	1/1:255:53:49:2	./.:.:15	1/1	./.
Chromosome	3586166	GC	G	PASS	ADP=29;WT=1	GT:GQ:SDP:D	1/1:248:43:39:(	./.:.:13	1/1	./.
Chromosome	3842517	C	CAGG	PASS	ADP=33;WT=1	GT:GQ:SDP:D	1/1:167:38:37:4	./.:.:15	1/1	./.
Chromosome	3936457	CGCCCGCAG	C	PASS	ADP=44;WT=1	GT:GQ:SDP:D	0/1:81:56:54:32	./.:.:28	0/1	./.
Chromosome	4235461	GA	G	PASS	ADP=34;WT=1	GT:GQ:SDP:D	1/1:249:47:42:1	./.:.:13	1/1	./.
Chromosome	4356702	GTACGGCTTT	G	PASS	ADP=43;WT=1	GT:GQ:SDP:D	0/1:205:70:63:1	./.:.:19	0/1	./.
Chromosome	4534497	CG	C	PASS	ADP=44;WT=1	GT:GQ:SDP:D	1/1:255:64:64:2	./.:.:14	1/1	./.
Chromosome	35266	C	T	PASS	ADP=49;WT=1	GT:GQ:SDP:D	1/1:255:68:66:(	./.:.:18	1/1	./.
Chromosome	130051	G	A	PASS	ADP=38;WT=1	GT:GQ:SDP:D	1/1:255:52:51:(	./.:.:18	1/1	./.
Chromosome	187935	T	A	PASS	ADP=45;WT=1	GT:GQ:SDP:D	1/1:255:61:60:(	./.:.:18	1/1	./.
Chromosome	229297	C	T	PASS	ADP=45;WT=1	GT:GQ:SDP:D	1/1:255:72:67:(	./.:.:17	1/1	./.
Chromosome	290616	T	C	PASS	ADP=21;WT=1	GT:GQ:SDP:D	1/1:188:33:33:(	./.:.:8	1/1	./.
Chromosome	374739	T	C	PASS	ADP=51;WT=1	GT:GQ:SDP:D	1/1:255:71:70:(	./.:.:16	1/1	./.
Chromosome	405941	A	T	PASS	ADP=37;WT=1	GT:GQ:SDP:D	1/1:255:59:55:(	./.:.:9	1/1	./.
Chromosome	421411	A	G	PASS	ADP=31;WT=1	GT:GQ:SDP:D	1/1:206:36:36:(	./.:.:16	1/1	./.
Chromosome	444953	T	C	PASS	ADP=39;WT=1	GT:GQ:SDP:D	1/1:255:53:47:(	./.:.:15	1/1	./.
Chromosome	529627	C	G	PASS	ADP=52;WT=1	GT:GQ:SDP:D	1/1:255:81:78:(	./.:.:14	1/1	./.
Chromosome	645951	T	C	PASS	ADP=42;WT=1	GT:GQ:SDP:D	1/1:255:54:51:(	./.:.:19	1/1	./.
Chromosome	816666	A	T	PASS	ADP=36;WT=1	GT:GQ:SDP:D	1/1:255:50:46:(	./.:.:11	1/1	./.
Chromosome	951086	C	T	PASS	ADP=31;WT=1	GT:GQ:SDP:D	1/1:236:43:41:(	./.:.:16	1/1	./.
Chromosome	986004	G	A	PASS	ADP=35;WT=1	GT:GQ:SDP:D	1/1:255:53:48:(	./.:.:13	1/1	./.
Chromosome	988820	G	A	PASS	ADP=38;WT=1	GT:GQ:SDP:D	1/1:236:47:41:(	./.:.:19	1/1	./.
Chromosome	1074253	G	C	PASS	ADP=34;WT=1	GT:GQ:SDP:D	1/1:255:45:45:(	./.:.:16	1/1	./.
Chromosome	1077116	C	T	PASS	ADP=27;WT=1	GT:GQ:SDP:D	1/1:242:50:42:(	./.:.:19	1/1	./.
Chromosome	1181287	G	A	PASS	ADP=52;WT=1	GT:GQ:SDP:D	1/1:255:81:79:(	./.:.:19	1/1	./.
Chromosome	1190426	T	C	PASS	ADP=33;WT=1	GT:GQ:SDP:D	1/1:248:45:43:(	./.:.:14	1/1	./.
Chromosome	1219096	T	G	PASS	ADP=30;WT=1	GT:GQ:SDP:D	1/1:255:48:45:(	./.:.:14	1/1	./.
Chromosome	1332448	A	C	PASS	ADP=49;WT=1	GT:GQ:SDP:D	1/1:255:73:70:(	./.:.:18	1/1	./.
Chromosome	1476641	T	C	PASS	ADP=34;WT=1	GT:GQ:SDP:D	1/1:255:54:51:(	./.:.:10	1/1	./.
Chromosome	1477695	C	T	PASS	ADP=51;WT=1	GT:GQ:SDP:D	0/1:131:71:67:3	./.:.:23	0/1	./.
Chromosome	1477696	T	A	PASS	ADP=49;WT=1	GT:GQ:SDP:D	0/1:134:70:63:2	./.:.:23	0/1	./.
Chromosome	1477697	G	T	PASS	ADP=55;WT=1	GT:GQ:SDP:D	0/1:138:73:72:3	./.:.:23	0/1	./.
Chromosome	1477704	T	C	PASS	ADP=56;WT=1	GT:GQ:SDP:D	0/1:158:78:73:3	./.:.:25	0/1	./.
Chromosome	1477728	C	G	PASS	ADP=54;WT=1	GT:GQ:SDP:D	0/1:121:72:69:3	./.:.:23	0/1	./.
Chromosome	1477733	C	G	PASS	ADP=52;WT=1	GT:GQ:SDP:D	0/1:122:71:66:3	./.:.:23	0/1	./.
Chromosome	1477734	G	C	PASS	ADP=53;WT=1	GT:GQ:SDP:D	0/1:130:72:70:3	./.:.:23	0/1	./.
Chromosome	1477752	A	G	PASS	ADP=54;WT=1	GT:GQ:SDP:D	0/1:123:74:74:4	./.:.:24	0/1	./.
Chromosome	1477766	A	G	PASS	ADP=52;WT=1	GT:GQ:SDP:D	0/1:125:77:70:3	./.:.:27	0/1	./.
Chromosome	1478271	G	A	PASS	ADP=65;WT=1	GT:GQ:SDP:D	0/1:72:86:86:64	./.:.:28	0/1	./.
Chromosome	1481004	C	G	PASS	ADP=62;WT=1	GT:GQ:SDP:D	0/1:107:71:70:4	./.:.:31	0/1	./.
Chromosome	1481005	A	G	PASS	ADP=58;WT=1	GT:GQ:SDP:D	0/1:104:70:66:3	./.:.:31	0/1	./.
Chromosome	1575470	T	C	PASS	ADP=51;WT=1	GT:GQ:SDP:D	1/1:255:80:78:(	./.:.:16	1/1	./.
Chromosome	1605283	T	C	PASS	ADP=62;WT=1	GT:GQ:SDP:D	0/1:159:92:87:4	./.:.:26	0/1	./.
Chromosome	1605286	C	G	PASS	ADP=65;WT=1	GT:GQ:SDP:D	0/1:157:92:92:4	./.:.:29	0/1	./.
Chromosome	1605619	A	T	PASS	ADP=47;WT=1	GT:GQ:SDP:D	0/1:162:59:57:1	./.:.:22	0/1	./.
Chromosome	1699964	G	A	PASS	ADP=37;WT=1	GT:GQ:SDP:D	1/1:255:55:53:(	./.:.:15	1/1	./.
Chromosome	1850132	C	T	PASS	ADP=37;WT=1	GT:GQ:SDP:D	1/1:255:49:47:(	./.:.:15	1/1	./.
Chromosome	1856742	T	C	PASS	ADP=48;WT=1	GT:GQ:SDP:D	1/1:255:69:66:(	./.:.:19	1/1	./.
Chromosome	2041780	T	C	PASS	ADP=20;WT=1	GT:GQ:SDP:D	1/1:135:26:24:(	./.:.:11	1/1	./.
Chromosome	2123720	C	A	PASS	ADP=41;WT=1	GT:GQ:SDP:D	1/1:255:58:56:(	./.:.:15	1/1	./.
Chromosome	2156119	A	G	PASS	ADP=35;WT=1	GT:GQ:SDP:D	1/1:255:53:48:(	./.:.:14	1/1	./.

Journal Pre-proof

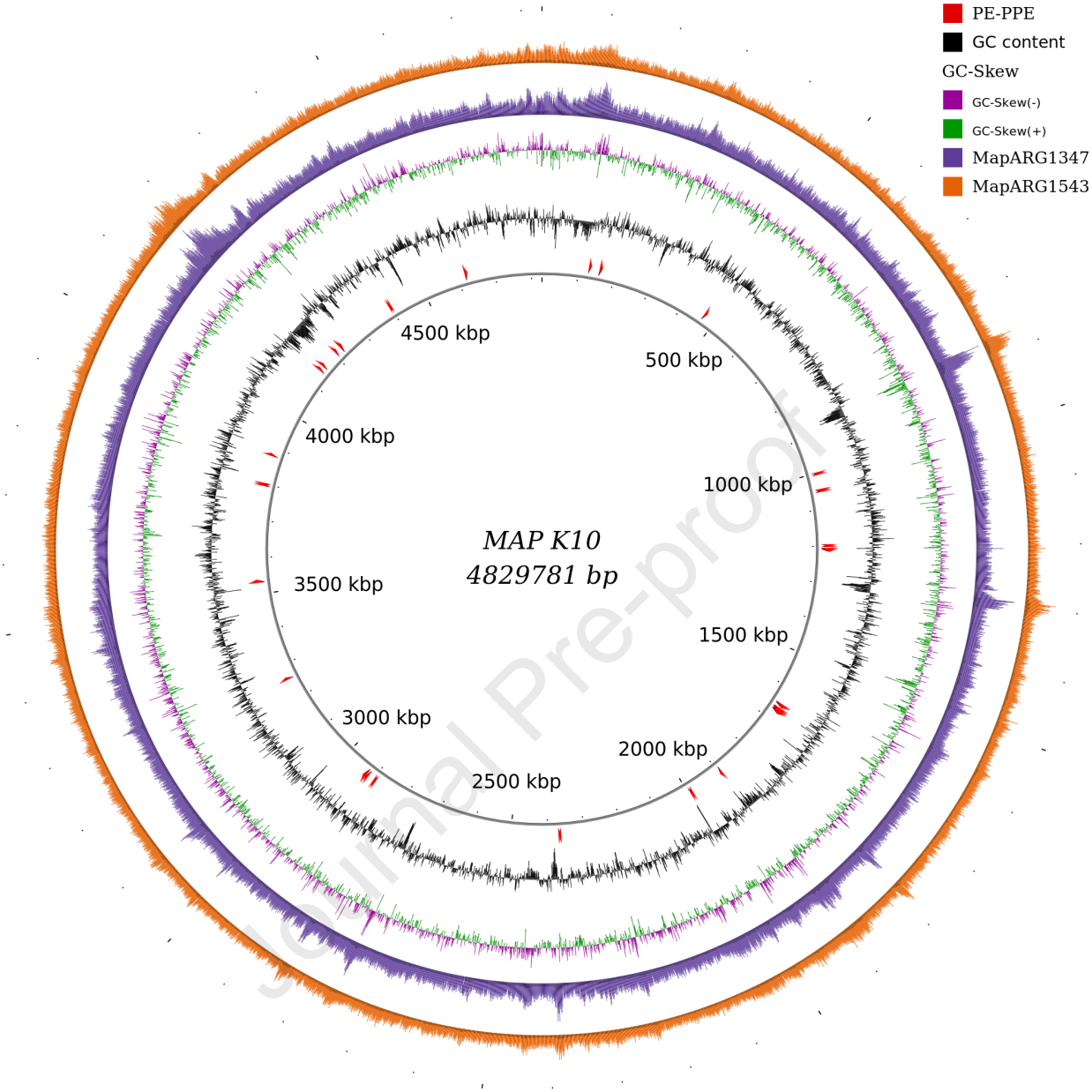
Chromosome	2204818	G	C	PASS	ADP=38;WT=1	GT:GQ:SDP:D	1/1:254:49:44:(	./.::17
Chromosome	2262972	T	C	PASS	ADP=24;WT=(	GT:GQ:SDP:D	1/1:176:32:31:(	./.::6
Chromosome	2279549	T	C	PASS	ADP=45;WT=1	GT:GQ:SDP:D	1/1:255:50:48:(	./.::18
Chromosome	2802761	A	G	PASS	ADP=19;WT=1	GT:GQ:SDP:D	1/1:123:24:22:(	./.::12
Chromosome	2802764	C	G	PASS	ADP=19;WT=1	GT:GQ:SDP:D	1/1:129:24:23:(	./.::12
Chromosome	2802770	A	G	PASS	ADP=19;WT=1	GT:GQ:SDP:D	1/1:135:24:24:(	./.::13
Chromosome	2802884	A	G	PASS	ADP=26;WT=1	GT:GQ:SDP:D	1/1:200:35:35:(	./.::16
Chromosome	2900072	G	C	PASS	ADP=36;WT=1	GT:GQ:SDP:D	1/1:255:55:55:(	./.::13
Chromosome	2900076	A	T	PASS	ADP=35;WT=1	GT:GQ:SDP:D	1/1:255:55:55:(	./.::13
Chromosome	2904668	A	G	PASS	ADP=43;WT=1	GT:GQ:SDP:D	1/1:255:71:68:(	./.::17
Chromosome	2992754	C	G	PASS	ADP=47;WT=1	GT:GQ:SDP:D	1/1:255:64:64:(	./.::18
Chromosome	3006223	A	G	PASS	ADP=44;WT=1	GT:GQ:SDP:D	1/1:255:49:46:(	./.::19
Chromosome	3040982	T	G	PASS	ADP=48;WT=1	GT:GQ:SDP:D	1/1:255:70:67:(	./.::16
Chromosome	3052144	C	A	PASS	ADP=47;WT=1	GT:GQ:SDP:D	1/1:255:69:66:(	./.::19
Chromosome	3081309	C	A	PASS	ADP=47;WT=1	GT:GQ:SDP:D	1/1:255:64:62:(	./.::16
Chromosome	3178581	T	C	PASS	ADP=39;WT=1	GT:GQ:SDP:D	1/1:255:57:57:(	./.::15
Chromosome	3197385	C	G	PASS	ADP=23;WT=1	GT:GQ:SDP:D	1/1:135:27:24:(	./.::9
Chromosome	3278537	A	C	PASS	ADP=41;WT=1	GT:GQ:SDP:D	1/1:255:54:53:(	./.::16
Chromosome	3326744	T	G	PASS	ADP=32;WT=1	GT:GQ:SDP:D	1/1:255:56:55:(	./.::9
Chromosome	3387967	G	A	PASS	ADP=44;WT=1	GT:GQ:SDP:D	1/1:255:64:60:(	./.::14
Chromosome	3422425	T	C	PASS	ADP=38;WT=1	GT:GQ:SDP:D	1/1:255:53:49:(	./.::19
Chromosome	3581372	G	A	PASS	ADP=24;WT=1	GT:GQ:SDP:D	1/1:212:38:37:(	./.::15
Chromosome	3756534	A	C	PASS	ADP=43;WT=1	GT:GQ:SDP:D	1/1:255:69:65:(	./.::19
Chromosome	3757120	G	A	PASS	ADP=39;WT=1	GT:GQ:SDP:D	1/1:255:51:50:(	./.::19
Chromosome	3842730	G	C	PASS	ADP=27;WT=1	GT:GQ:SDP:D	1/1:188:35:33:(	./.::8
Chromosome	3842808	T	C	PASS	ADP=36;WT=1	GT:GQ:SDP:D	1/1:255:47:46:(	./.::17
Chromosome	3890009	G	A	PASS	ADP=46;WT=1	GT:GQ:SDP:D	1/1:255:72:69:(	./.::19
Chromosome	3975214	A	C	PASS	ADP=46;WT=1	GT:GQ:SDP:D	1/1:255:67:66:(	./.::18
Chromosome	4012463	G	A	PASS	ADP=44;WT=1	GT:GQ:SDP:D	1/1:255:60:58:(	./.::18
Chromosome	4130905	G	A	PASS	ADP=54;WT=1	GT:GQ:SDP:D	1/1:255:77:75:(	./.::19
Chromosome	4254951	G	A	PASS	ADP=39;WT=1	GT:GQ:SDP:D	1/1:255:58:52:(	./.::18
Chromosome	4307523	A	G	PASS	ADP=44;WT=1	GT:GQ:SDP:D	1/1:255:61:61:(	./.::19
Chromosome	4368391	A	G	PASS	ADP=47;WT=1	GT:GQ:SDP:D	1/1:255:70:70:(	./.::9
Chromosome	4391467	A	G	PASS	ADP=34;WT=1	GT:GQ:SDP:D	1/1:255:47:46:(	./.::18
Chromosome	4427408	G	A	PASS	ADP=42;WT=1	GT:GQ:SDP:D	1/1:255:55:53:(	./.::17
Chromosome	4430919	G	T	PASS	ADP=46;WT=1	GT:GQ:SDP:D	1/1:255:69:68:(	./.::11
Chromosome	4485387	G	A	PASS	ADP=32;WT=1	GT:GQ:SDP:D	1/1:255:55:54:(	./.::15
Chromosome	4638025	A	G	PASS	ADP=32;WT=1	GT:GQ:SDP:D	1/1:254:45:44:(	./.::11

Phagocytosis and survival of Map strains within the bMDM by counting Colony Forming Units (CFU)

		Map strain		
		Map K10	**MapARG1347**	**MapARG1543**
% phagocytosis Median % (IQR)		41 (26.36-52.66)	34.78 (24.41-40.31)	43.19 (35.59-48.75)
Days post-infection Median CFU/mL (IQR)	**0**	105 (2.64 x105-1.68 x	105 (2.43 x1054.84	6.29 x105 (1.49 x105-1.11 x106)
	2	105 (1.93 x105-4.14 x	5 (#) (5.11 x105-8.8	7.44 x105 (2.08 x105-1.39 x106)
	4	105 (5.66 x104-4.63 x	105 (3.4 x105-4.63 x	2.12 x105 (1.53 x105-3.73 x105)
	6	04 (a) (4.34 x104- 1.93	5 (b) (1.92 x105- 2.1	1.59 x105 (4.98 x104- 2.91 x105)

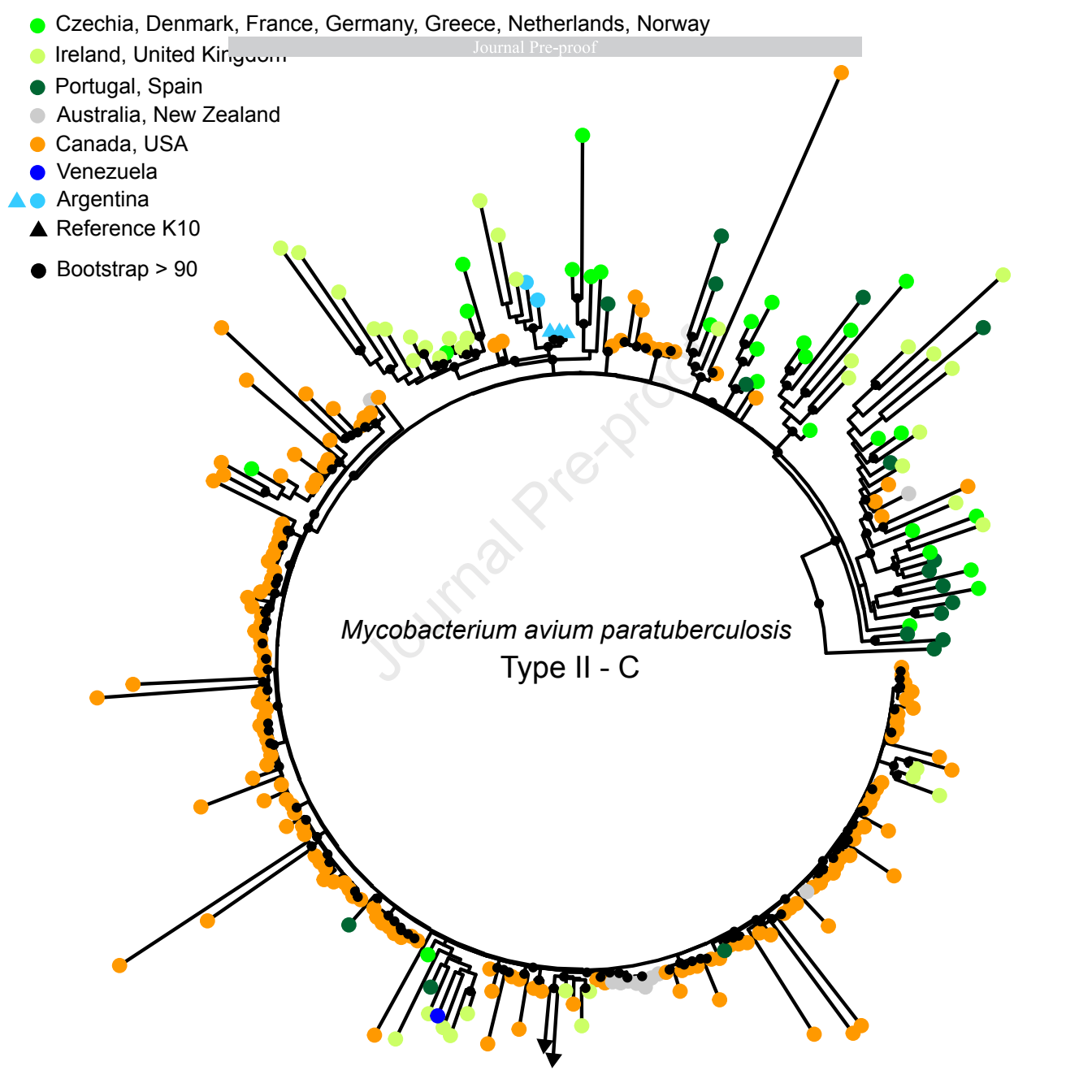
Journal Pre-proof

Journal Pre-proof



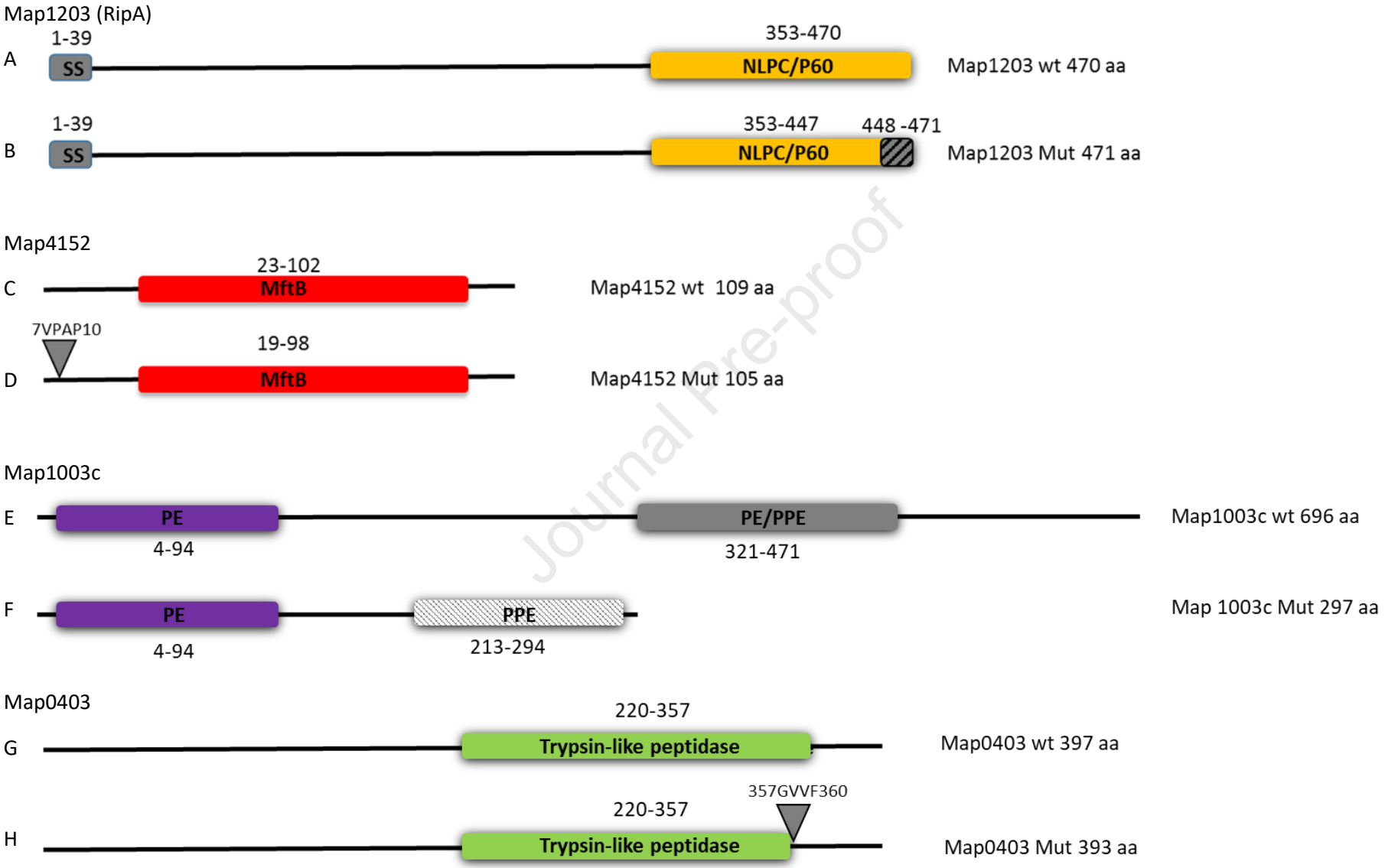
Journal Pre-proof

Journal Pre-proof



Journal Pre-proof

Journal Pre-proof



Journal Pre-proof

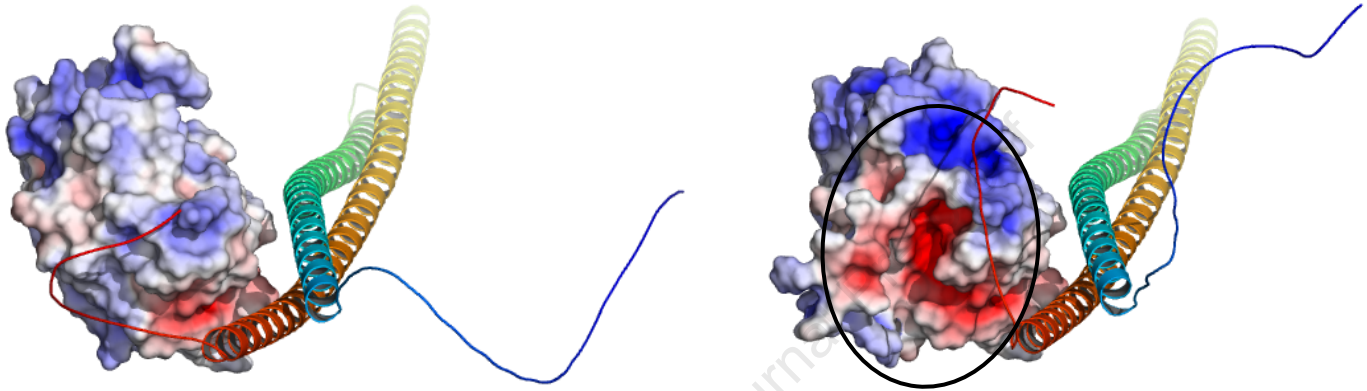
Journal Pre-proof

A

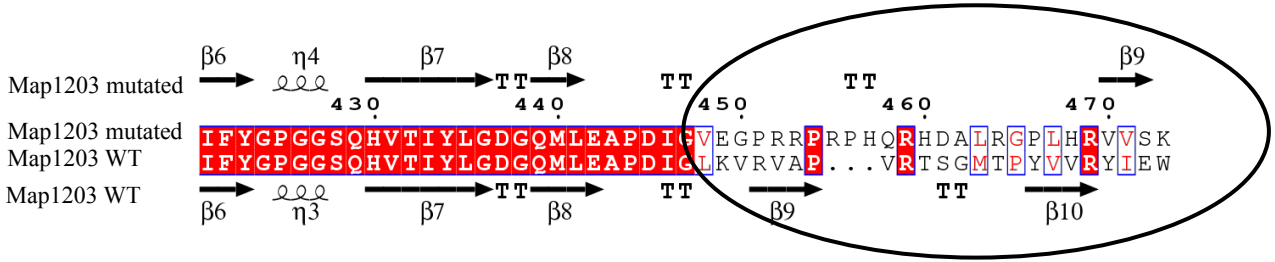
Map1203 WT

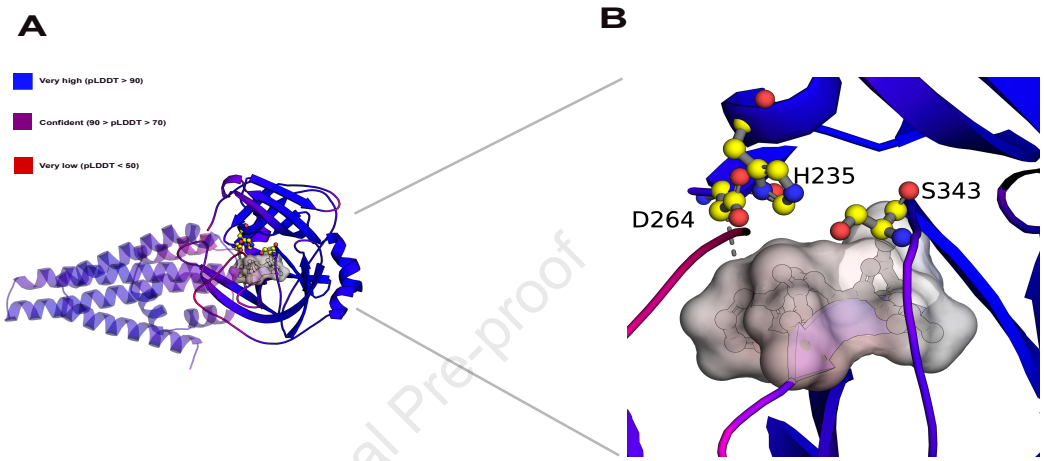
B

Map1203 mutated

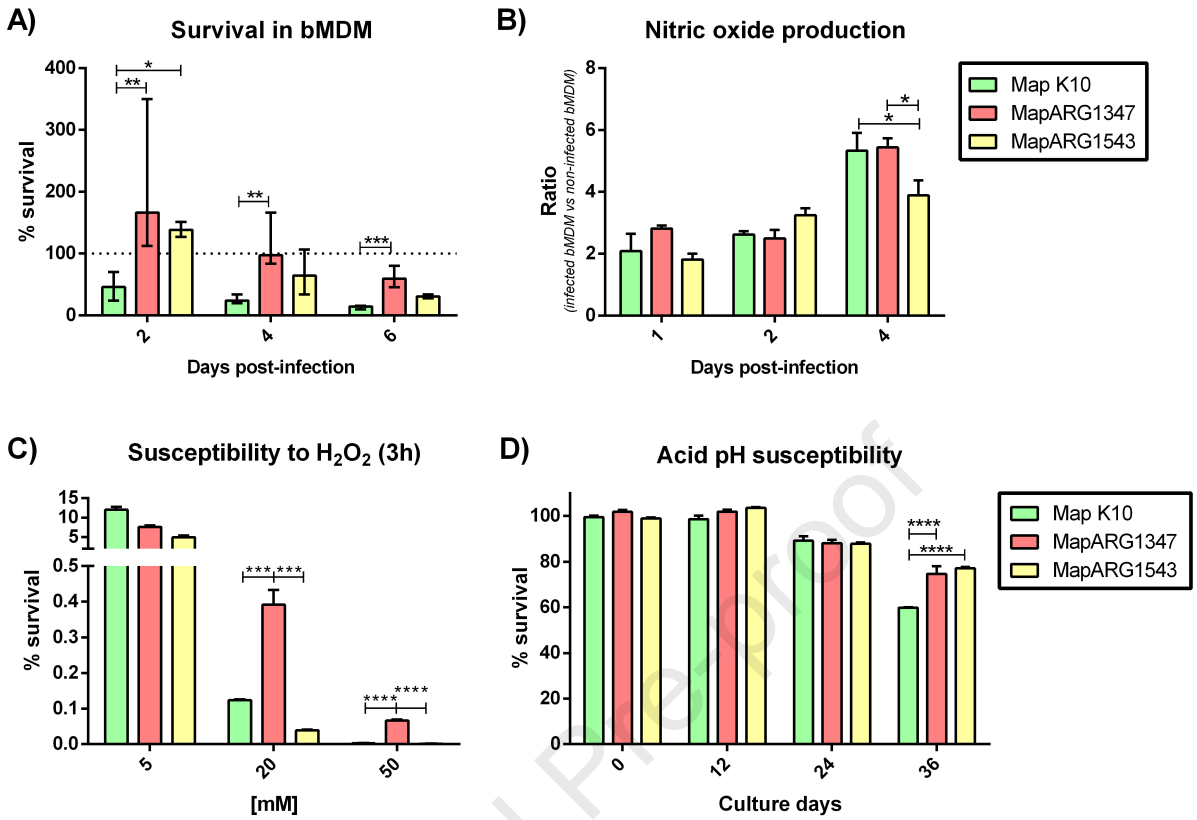


C





Journal Pre-proof



Journal Pre-proof

Journal Pre-proof

Highlights

WGS of Argentinian isolates with contrasting pathogenic phenotypes

Phylogenetic analysis with whole genome sequences available from GenBank

Identification of NS-SNP and INDELS by comparative genomic analysis

Identification of genes candidates that could explain the differences in phenotype between the two strains

Journal Pre-proof