Polygenic plague resistance in the great gerbil uncovered by population sequencing

Pernille Nilsson[1], Mark Ravinet[1*], Yujun Cui[2*], Paul R. Berg[1,6*], Yujiang Zhang[3*], Rong Guo[3], Tao Luo[3], Yajun Song[2], Emiliano Trucchi[5], Siv N. K. Hoff[1], Ruichen Lv[2], Boris V. Schmid[1], W. Ryan Easterday[1], Kjetill S. Jakobsen[1,#], Nils Chr. Stenseth[1,4,#], Ruifu Yang[2,#] and Sissel Jentoft[1,#]

[1] Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Norway

[2] State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, China

[3] Xinjiang Center for Disease Control and Prevention, Urumqi, China

[4] Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China

[5] Department of Life and Environmental Sciences, Marche Polytechnic University, Via Brecce Bianche, 60131, Ancona, Italy

[6] Centre for Coastal Research, Department of Natural Sciences, University of Agder, Norway

[*] Shared second authors

[#] Co-corresponding authors

Abstract

Pathogens may elicit a high selective pressure on hosts and can alter genetic diversity over short evolutionary timescales. Intraspecific variation in immune response can be observed as variable survivability from specific infections. The great gerbil (*Rhombomys opimus*) is a rodent plague host with a heterogenic but highly resistant phenotype. Here, we investigate if the most plague-resistant phenotypes are linked to genomic differences between survivors and susceptible individuals by exposure of wild-caught great gerbils from Northwest China to plague (*Yersinia pestis*). Whole genome sequencing of ten survivors and ten moribund individuals revealed a low genome-wide mean divergence, except for a subset of genomic regions that showed elevated differentiation. Gene ontology (GO) analysis of candidate genes within regions of increased differentiation, demonstrated enrichment of pathways involved in transcription and translation and their regulation), as well as genes directly involved in immune functions, cellular metabolism and the regulation of apoptosis. Differential RNA expression analysis revealed that the early activated great gerbil immune response to plague consisted of classical components of the innate immune system. Our approach combining challenge experiments with transcriptomics and population level sequencing, provides new insight into the genetic background of plague-resistance and confirms its complex nature, most likely involving multiple genes and pathways of both the immune system and regulation of basic cellular functions.

Keywords: great gerbil, population genomics, polygenic resistance, plague, challenge experiment, transcriptomics

Introduction

Pathogens are well recognized as one of the strongest selective agents influencing host population genomic diversity through adaptation (Cagliani & Sironi, 2013; Fumagalli et al., 2011; Haldane, 1932). Differences in pathogen pressures within and between host populations can reduce or increase genetic diversity, depending on the lethality of the disease and the type of selection. For instance, highly virulent pathogens and purifying selection at resistance loci generally act to reduce genetic diversity (Barreiro, Laval, Quach, Patin, & Quintana-Murci, 2008; Trudeau, Britten, & Restani, 2004), while balancing selection and low and intermediate virulence can result in higher genetic diversity (Sackett, Collinge, & Martin, 2013; Schulte, Makus, Hasert, Michiels, & Schulenburg, 2010). The underlying genomic basis for host resistance to infectious diseases is predominantly caused by polygenic traits, where many genes or variants collectively contribute to the phenotype (Barreiro & Quintana-Murci, 2009; Casanova & Abel, 2007; Hill, 2001). For instance, resistance to tuberculosis in humans has been linked to multiple genomic loci (Möller et al., 2018). Thus, selection for pathogen resistance is likely to shape diversity at multiple regions in the genome.

The great gerbil (*Rhombomys opimus*) is a well-studied wild, social rodent and serves as reservoir for the plague bacterium (*Yersinia pestis*) (Kausrud et al., 2007; Reijniers, Begon, Ageyev, & Leirs, 2014). Knowledge of their high level of plague resistance has existed since the early 1950's and individual heterogeneity in response to infection has been revealed by challenge experiments (Petrunina, 1951; Y. Zhang et al., 2012). Heterogenic response to plague infection is generally considered a prerequisite for permanent plague reservoirs and has also

been reported within populations of other rodent plague hosts like the black rat (*Rattus rattus*) and Asian house shrew (*Suncus murinus*) in Madagascar (Andrianaivoarimanana et al., 2012; Gage & Kosoy, 2005; Gascuel, Choisy, Duplantier, Débarre, & Brouat, 2013; Rahelinirina et al., 2017). For plague to persist for long times on a small local scale, there has to be a balance between having enough rodents experiencing high-level, but often fatal, bacteremia and enough rodents that persist to avoid local extinction (Anisimov, Lindler, & Pier, 2004; Lowell et al., 2015). One way to achieve such a balance is through heterogenic response to plague infection. Several studies of laboratory and wild rodent species have been used to investigate the genetic basis for plague resistance, and current evidence strongly suggests that it is likely governed by multiple genes (Blanchet et al., 2010; Chevallier et al., 2012; Tencati & Tapping, 2016). Although some of these studies were unable to determine candidate genes due to the low resolution of their method (Chevallier et al., 2012), others identified multiple candidate genes involved in immune related processes as well as genes lacking a known immune function (Blanchet et al., 2010; Tollenaere, Jacquet, et al., 2012b; Tollenaere et al., 2008). Both Tollenaere et al. (2012b) and Blanchet et al. (2010) highlights receptors of interleukin 17 (IL17) as potential important factors in plague resistance in black rats and SEG mice, respectively (Blanchet et al., 2010; Tollenaere, Jacquet, et al., 2012b). In general, our understanding of the genomic basis for plague susceptibility and resistance is still poor and the exact mechanism might differ between species (Demeure et al., 2012; Hubbert & Goldenberg, 1970; Rocke et al., 2012; Tollenaere, Ivanova, et al., 2012a).

Host-pathogen coevolution results in resistance phenotypes on both sides as a consequence of natural selection for adaptation and counter-adaptation (Paterson et al., 2010). *Y. pestis* is a relatively new pathogen viewed on an evolutionary timescale and displays

distinct eco-geographical variations in vector-host transmission modes. For example, it has been demonstrated that in a particular site in China (Guertu) *Y. pestis* transmission is enhanced by bacterial gene selection-directed biofilm formation in the flea (extended phenotype changes) caused by climate fluctuations (Cui et al., 2020). Based on current phylogenetic analysis of *Y. pestis*, it seems plausible that the bacterium gradually evolved from *Yersinia pseudotuberculosis* into its more lethal, flea-borne form approximately 5-6000 years ago, in Central Asia or Western China (Achtman et al., 2004; 1999; Morelli et al., 2010; Rasmussen et al., 2015).

The Gurbantünggüt desert of the Junggar Basin in Northwest China is one of the most recently identified plague foci and constitutes one of the more eastern parts of the great gerbil (hereafter 'gerbil') distribution (Figure 1A) (Y.-J. Zhang et al., 2008). The first occurrences of plague in wildlife in this area were reported in 2005 (Y.-J. Zhang et al., 2008). Plague has been known to reappear in areas after decades of quiescence (Arbaji et al., 2005; Bertherat et al., 2007). As such, plague might have been present in the rodent population prior to initiation of surveillance in the 1950s (Y. Zhang et al., 2018), with ample time for co-evolution. However, if plague indeed only entered the system in 2005, the signatures of an arms race between *Y. pestis* and gerbil could still be detectable in the host due to the short generation times and large populations sizes of this animal group (Sironi, Cagliani, Forni, & Clerici, 2015). Furthermore, the strong selective pressure exerted by a highly virulent pathogen could lead to rapid evolution of host adaptation and resistance to infection. Such rapid evolution to pathogens has been reported in other host populations including wild finches in North America and European rabbits in Australia (Best & Kerr, 2000; Bonneaud et al., 2011; Kerr et al., 2017; Marshall & Fenner, 1958).

Advances in high-throughput sequencing technology over the past decades have resulted in a burst of detailed investigations of genomic adaptation in non-model species (Ellegren, 2014). Recently, Nilsson and co-workers generated a highly contiguous reference genome assembly of the great gerbil (Nilsson et al., 2020). The study revealed a species-specific duplication of a Major Histocompatibility Complex class II (*MHCII*) gene with a predicted high affinity for *Yersinia* epitopes (Nilsson et al., 2020). Here, we aimed at further investigating the genetic basis of disease resistance in gerbil by whole genome sequencing wild captured specimens from a location in the plague focus of the Gurbantünggut desert in Northwest China. During a 22-day experiment, we infected gerbils with *Y. pestis*, and subsequently performed whole genome sequencing (12x coverage) on individuals from the two phenotypic outcomes of the infection, i.e., those that were moribund and those that survived the infection. This data set combined with transcriptome sequencing enabled us to investigate potential adaptive effects on the host by a highly virulent pathogen and locate genomic differences that potentially underlie variation in host response to infection.

Methods

Experimental design – animals and bacterial strain

Sub-adult gerbils were captured from the natural plague focus in the Junggar Basin located in the Gurbatntünggut desert, Xinjiang, China within a 10 x 10 km area. The gerbils were housed in the field laboratory for at least two weeks prior to screening for *Y. pestis* F1 antigens and anti-F1 antibodies using up-converting phosphor technology-based lateral flow strips (P. Zhang et al., 2014) and indirect hemagglutination assays (IHA) respectively (Y. Zhang et al.,

2012; 2015). Gerbils negative for both assays were used for subsequent animal challenges with the fully virulent *Y. pestis* strain 2505. The strain was isolated from a live great gerbil in Minai County of the Junggar Basin, Xinjiang, China in 2005 and has an LD50 of <10 in BALB/c mice (Y. Zhang et al., 2018).

Y. pestis challenge

Forty-five gerbils (18 males and 27 females) were randomly divided into nine housing groups (n = 5 per group). After anesthesia by ether inhalation, all 45 gerbils were subcutaneously injected in the groin with one ml of *Y. pestis* 2505 culture suspended in physiological saline (5.6×10^9 CFU/ml). A subcutaneous injection was chosen as it allows for the required volume containing the large dose of plague to be injected. Another nine gerbils (three males and six females) were injected with one ml of physiological saline to serve as a handling control (making sure the animals did not succumb to stress from the procedure). All the gerbils were raised group-wise in filter-top cages in an air-conditioned room set at 22°C with chow and water *ad libitum*. All challenged individuals were observed twice per day for clinical signs of plague (including increased anal temperature, polydipsia, closed eyes, ruffled fur, hunched posture and lethargy) (Y. Zhang et al., 2012). Animals were classified into three different categories: i) moribund (those that showed signs of plague symptoms and concurrently lack of vitality), ii) surviving (those that recovered with less strong signs of plague symptoms), iii) healthy (those that showed no signs of plague infection) throughout the experimental period. The moribund animals were euthanized as soon as signs of obvious disease appeared. In total, twenty-one challenged, moribund gerbils were euthanized at different days p.i. In addition, one gerbil was found dead at day 13 p.i., and was dissected aseptically to collect the liver.

Surviving animals and the control individuals were euthanized at day 22 post infection (p.i.). Four of the five individuals showing no signs of symptoms during the trial (i.e. the healthy individuals) were euthanized on day three and four p.i. while the fifth was euthanized on day 22 p.i. Immediately after euthanasia the animals were dissected aseptically to collect the liver. See Table S1 for details and metadata of all gerbils.

The liver tissue samples were split in two. Half of the collected liver tissue was snap frozen in liquid nitrogen and then transferred to clean tubes and kept at -20 °C until DNA extraction. The remaining half of the animal's livers were cut into smaller pieces and submerged in five mL RNA*later* (Ambion), incubated at 4 °C overnight and then frozen at -20 °C until RNA extraction. Ten moribund and ten surviving individuals (n = 20, nine males and 11 females) were whole genome sequenced at an average 12x coverage per individual (see details in next section).

Animal challenge experiments and the use of great gerbil tissue in this study were performed abiding by the biosafety and ethical regulations issued by the Ministry of Health, China, and approved by the Committee for Animal Welfares of Xinjiang CDC. Sampling was performed prior to China's signature of the Nagoya Protocol (date of accession September 6th 2016). The sampled species have a "least concern" status in the IUCN Red List of Threatened Species.

DNA extraction and sequencing

DNA from gerbil livers was extracted using Qiagen Blood and Tissue DNeasy® kit (Qiagen Inc. USA) following the manufacture's protocol. DNA extractions were freeze dried prior to shipping to the University of Oslo and upon arrival resuspended in 200 µl E.Z.N.A.® Elution

buffer (Omega Biotek) and placed in a heating block at 37 °C for four hours. The extractions were then analyzed with Qubit™ (Thermo Fisher Scientific), NanoDrop™ (Thermo Fisher Scientific) and Bioanalyzer® (2100, Agilent Technologies) to assess the quality and quantity of DNA. DNA samples from ten survivors and ten moribund individuals (n = 20) from the experiment, were selected for DNA sequencing based on high quality DNA (Table 1). Prior to library prep, an additional 100 µl E.Z.N.A.® Elution buffer was added to the samples due to high DNA yields. Library prep was performed using the Illumina® TruSeq® DNA PCR Free protocol and the samples were sequenced using Illumina® HighSeq 2500 with a 350 bp insert size at the Norwegian Sequencing Centre (NSC), University of Oslo, Norway.

Repeat masking, read alignment and variant calling

Prior to population genomic analyses, all types of repeats were masked from the gerbil reference genome (Nilsson et al., 2020) using RepeatMasker v4.0.6 (Tarailo-Graovac & Chen, 2009) with default settings and "rhombomys opimus" species filter to avoid calling variants in these areas (Briskine & Shimizu, 2017). Thirty-four percent (34 %) of the 2.47 Gb gerbil reference genome was masked by RepeatMasker prior to mapping and included all types of repeats (Table S2). Raw sequence reads were trimmed for Illumina adaptors and low-quality reads using Trimmomatic v0.36 (Bolger, Lohse, & Usadel, 2014). Reads with an average quality of less than 20 across 5 bp step windows were removed as well as reads below 40 bp in length. Trimmed and filtered reads (paired only) were mapped to the repeat masked reference genome using bwa-mem v0.7.8 (Li & Durbin, 2009) with default parameters except adding an -M parameter to enable Picard (http://picard.sourceforge.net) and also specifying read group. As each individual was sequenced on multiple lanes, each sequence lane was mapped

separately and then merged to produce a final bam file for each individual. Bams were then sorted, filtered for duplicates and indexed using Picard v1.72. Bams were realigned around insertion-deletion polymorphisms and variant calls were made for all sites using the GATK HaplotypeCaller of GATK's Genome Analysis Toolkit v3.7 (DePristo et al., 2011; McKenna et al., 2010). The raw vcf was filtered to extract SNPs only and hard filtered according to GATK's recommendations, including filtering on QualByDepth (QD, with thresholds adjusted to fit the data, see Figure S1). In the process, the SNPs in the vcf were annotated with filter thresholds. For downstream analyses, additional filters were applied creating a high-quality dataset, which included only SNPs occurring in all individuals and a minor allele threshold of 0.05 (MAF < 5 %) and DP > 4 (removing DP < 5) (hereafter 'filtered variants').

Population structure estimation

Linkage disequilibrium (LD) decay was calculated across each of the 20 largest scaffolds (10% of the assembled genome) using the SNP set of filtered variants. Pairwise r^2 values were calculated between all SNPs within a 500 kb window in Plink v1.90b3b (Purcell et al., 2007). Decay plots (Figure S2) were created by binning the distance between SNPs in increments of 1 kb and averaging the r^2 values within each bin (Elgvin et al., 2017). To investigate population structure a second dataset (hereafter 'LD pruned') was created by further performing linkage pruning on the filtered variants set using Plink2 v2.00a2LM (Chang et al., 2015) filtering for all loci within 100 kb windows with an r^2 exceeding 0.1. Genome window size was then determined based on the rate of LD decay from the 20 largest scaffolds (Figure S2). PCA was then performed on the LD pruned variants using Plink2 v2.00a2LM. As the gerbils were captured within a relatively small area, we tested the individuals for familial relationship by

running identity by descent (IBD) calculations on the LD pruned dataset using Plink v 1.90b3b. This allowed us to see if relatedness had an impact on the clustering pattern observed in the PCA and to evaluate if any individuals should be excluded from downstream analysis.

Phasing and demographic analyses

A separate filtering approach was used on the raw, unfiltered dataset prior to phasing and demographic analyses. First, scaffolds under 1 Mb were excluded from the dataset and the remaining SNPs and genotypes were filtered on low quality using vcflib (DP – summed across individuals – between 40 and 600, FS < 40, QUAL > 19.99). To test for robustness of the demographic reconstruction to individual genotype, minimum coverage and to missing data per locus, we repeated the analyses filtering for individual genotype DP > 2 and DP > 4 as well as filtering out loci with more than 20 % missing individuals (Figure S3).

Phasing was performed on these two variant datasets (hereafter 'phased variants') using beagle v4.1 with default parameters and no imputation. We wanted to investigate the recent demographic history of the gerbil population to see if any population decline could be linked to the emergence of plague. Since plague is a young disease, this required a sensitive method with good ability to model the recent past of the gerbil population (Schiffels & Durbin, 2014). Demographic analyses using MSMC2 (Schiffels & Durbin, 2014), were performed on the phased variants of the 21 longest scaffolds only, all over 1 Mb in size. Five randomly generated MSMC2 input files were created, each containing 3 of the 20 individuals, and used in replicated runs. The results were plotted in R with a 1-year generation time and a mutation rate of 5×10^{-9} (sub/site/gen) (Milholland et al., 2017; Uchimura et al., 2015).

Genome scans for signatures of selection

To identify genomic regions and potential candidate genes involved in plague resistance, we looked for signs of selective sweeps at two levels; first in the populations as a whole and secondly looking for differences between survivors and moribund individuals by treating the two groups as two separate populations in population analyses.

Integrated haplotype score (iHS) and cross-population haplotype homozygosity (xpEHH)

We calculated integrated haplotype score (iHS) and cross-population haplotype homozygosity (xpEHH) from the phased variants (DP > 4 only) to look for signs of recent positive selection in the gerbil population as a whole (Voight, Kudaravalli, Wen, & Pritchard, 2006) using the R package *rehh* 2.0 (Gautier, Klassmann, & Vitalis, 2017). Whereas iHS is designed to identify signatures of selective sweeps within populations, xpEHH is a comparative statistic that identifies divergent haplotype structure between pairwise populations. In this case, we used xpEHH to investigate differences in haplotype homozygosity between gerbils classified as moribund or survivors after the challenge experiment. Scripts for converting the phased vcf into *rehh* input files were modified from (Ravinet et al., 2018). Due to the discovered relatedness among the individuals in the experiment, we conducted sensitivity analysis on a single scaffold, scaffold00080 as it has the most significant xpEHH peak, by dropping one individual at a time and recalculating xpEHH to see if the results changed significantly (Figure S4).

F_{ST} and nucleotide diversity analyses

Genome-wide F_{ST} was calculated with the R package PopGenome v2.6.1 (Pfeifer, Wittelsbürger, Ramos-Onsins, & Lercher, 2014) on the variant filtered set, which had been divided into scaffolds using SnpSift v4.0 (Cingolani et al., 2012). Sliding-window F_{ST} was calculated on a 50 SNP basis (width = 50, jump = 25) using the sliding.windows.transform flag.

The same parameters per SNP were calculated on a concatenated dataset. All 50-SNP windows with F_{ST} values > 0.2 were extracted and the corresponding scaffolds were identified (Table S3). Per SNP F_{ST} values for the identified scaffolds were plotted to visualize the F_{ST} distribution within these scaffolds (Figure 2). PCA, DAPC plots and compoplots were constructed for all scaffolds with F_{ST} 50-SNP windows > 0.2 to visualize clustering between moribund and survivors, using the R packages VcfR 1.8.0 (Knaus & Grünwald, 2017)) and Adegenet 2.1.1 (Jombart, 2008). Due to the discovered relatedness and since the F_{ST} values are based on calculations from only 20 individuals (10 in each "population"), the robustness of the results was evaluated by calculating F_{ST} values after a leave-one-out procedure on one of the scaffolds (scaffold00102) (Figure S5). This scaffold was chosen as it had a large peak containing the highest calculated F_{ST} value (F_{ST} = 0.57). Nucleotide diversity (π) was calculated in sliding windows of 100 kb in 25 kb steps using vcftools v0.1.14 (Danecek et al., 2011), for the moribund and survivors separately.

Gene enrichment analysis of candidate genes

We extracted candidate genes surrounding the outlier peaks from the iHS and xpEHH analyses at a 250 kb distance using scripts modified from (Ravinet et al., 2018). For the iHS and xpEHH outliers, we set the threshold at a log10 p-value of 6 (this is equivalent to a $p = 1 \times 10^{-6}$) prior to candidate gene extraction. We also extracted candidate genes within 250 kb of all F_{ST} > 0.2 windows found on the 30 scaffolds with high F_{ST} regions. The gene lists were filtered to contain only a unique set of gene IDs and names were matched to human orthologs before running ontology analyses using clueGO in Cytoscape v3.7.0 (Bindea et al., 2009). All analyses were run with medium network specificity, in a right-sided hypergeometric test with

289 Benjamini-Hochberg FDR correction, only reporting pathways with a $p < 0.05$. For enriched
290 pathways highlighted in the results, all *p*-values are after FDR correction. The identity and
291 function of candidate genes was also examined manually through web-based searches.

292 **RNA extraction, sequencing and differential expression analysis**

293 RNA was extracted from the liver samples preserved in RNAlater™ using a standard
294 chloroform procedure (Chomczynski & Sacchi, 2006). In brief, this involved homogenizing the
295 liver tissue in 1 mL TRIzol™ Reagent (Invitrogen) with glass beads, then adding 0.2 mL
296 chloroform, shake samples and incubate at room temperature (RT) before centrifuging. The
297 supernatant was gently mixed with equal amounts of isopropanol and afterward centrifuged,
298 this time discarding the supernatant. 1 mL 75 % ethanol was added to the pellet and the
299 samples stored at -80 °C. Library prep and sequencing were conducted at the Beijing Genomics
300 Institute (BGI, https://www.bgi.com/us/sequencing-services/dna-sequencing/) using
301 Illumina® TruSeq® RNA Sample Prep Kit and PE sequencing on the HiSeq4000 instrument (150
302 bp read length).

303 The raw sequence reads were trimmed using trimmomatic v0.36 and mapped to the
304 gerbil reference genome using STAR v2.5.2a (Dobin et al., 2013) with default parameters. A
305 less stringently filtered annotation file for the reference genome than presented in Nilsson et
306 al. (2020) was used in the mapping process. A raw count matrix was created using htseq v0.7.2
307 with strandedness option set to 'no' and otherwise default parameters to extract raw counts
308 from the mapped files. The annotation file was then used to extract each feature count. Prior
309 to normalization and differential expression (DE), the count matrix was filtered by requiring
310 expression in minimum 2 libraries and excluding genes with read counts < 1 across all samples.

Normalization and dispersion were calculated using rmFactors, estimateDisp, estimateCommonDisp and estimateTaqwiseDisp using default parameters. One individual was excluded from further analysis due to high levels of individual variation that would reduce the power of downstream differential expression (Supplementary Note 1 and Figure S6). Differences in expression between moribund gerbils and survivors were analyzed using the edgeR package in R (Robinson, McCarthy, & Smyth, 2010). Differential expression was calculated using exacttest (function exactTest) between dead and survivor. Resulting differential expressed genes were filtered using $p < 0.05$. The generated list of genes with adjusted *p* values (FDR < 0.05) were further analyzed for significantly enriched pathways in ClueGO in Cytoskape v3.7.0. For the enrichment analyses the complete list of genes was analyzed as well as lists separated by genes that were upregulated and downregulated in the moribund individuals. The three enrichment analyses were run with medium network specificity in a right-sided hypergeometric test with Benjamini-Hochberg FDR correction and $p < 0.05$. Finally, the identity and function of DE genes was also examined manually through web-based searches.

Results

Plague challenge experiment

After plague infection, 21 of 45 challenged gerbils (47 %) were deemed moribund and euthanized, all except one during the first five days post infection (p.i) (Table S1). The majority of gerbils responded quickly to the disease with the onset of disease symptoms on day one p.i. For ten individuals the symptoms appeared on day two p.i., while five gerbils did not show

any signs of disease prior to sacrifice. The surviving animals recovered quickly from the infection after being sick for only a few days, and always less than a week. Seventeen of the 21 moribund gerbils were females while the majority of the surviving individuals were male (14 out of 19), yet all the surviving gerbils with no symptoms, were female. The results of the q-square test indicate a sex-bias with females more likely dying of the infection and/or not being visibly sick (X-squared = 16.157, df = 2, p-value = 0.00031) (Table S4).

Population structure

We whole genome sequenced 20 specimens from the moribund (n = 10) and surviving (n = 10) groups at an average 12x coverage per individual (Table 1). Furthermore, eight individuals from each group were RNA sequenced (liver), of which all were included in the population genome sequencing listed in Table 1, except one survivor. After mapping to the repeat-masked reference genome, calling and filtering variants, we retained a set of 1,120,260 high-quality SNPs and used a subset of 32,816 linkage disequilibrium (LD) pruned biallelic SNPs for non-parametric (PCA) inference of population structure. The individuals do not separate based on outcome of the challenge experiment, but some separation is seen along the first principal component (8.9 % PVE), with the majority of individuals from both groups forming a relatively tight cluster (Figure 1B). A small number of individuals are also separated from the rest along the second principal component (6.71 % PVE) and consists of four moribund individuals and one survivor (Figure 1B). Estimation of relatedness (by calculating IBD), using the LD pruned SNP set, revealed some degree of relatedness between several of the individuals. This matched the clustering pattern seen in the PCA where they form the more spread out groups separated from the main cluster on both PC1 and PC2 (Figure S7; Table S5). These IBD calculations

suggest that the individuals clustering in the upper right corner might be a mix of first- and second-degree relatives and also suggest a first-degree relation for the two individuals at the bottom of the PCA plot (Table S5).

Genome-wide estimates of mean F_{ST} and nucleotide diversity (π) revealed that both the relative differentiation between moribund and survivors as well as the (overall) nucleotide diversity is extremely low (F_{ST} = -0.0084 ± 0.0578, π = 0.000152 ± 0.000136). Low nucleotide diversity might have been affected by the decline in effective population size seen for the population over the last 200K years, which reached a minimum around 4-5K years ago followed by a recovery around 1-2K years ago (Figure S3).

Signatures of selective sweeps

We found signs of recent positive selection identifying 234 iHS and 122 xpEHH outlier peaks from across the phased genome at a set threshold for significance ($\log_{10}$ 1 x 10^{-6}) (Tables S6, S7; Figure S8). This is a conservative threshold representing a low false discovery rate of 0.001 and hence, only identifies SNPs with the strongest signatures of apparent selection. Sensitivity analysis performed on the scaffold with the highest xpEHH peak (scaffold00080; Table S7) where one individual at the time was excluded from the calculations, demonstrated that the signal was robust (Figure S4). Searching for regions with high differentiation between moribund gerbils and survivors ($F_{ST} > 0.2$) in 50-SNP windows revealed 30 scaffolds containing areas with elevated mean F_{ST} (0.0059 ± 0.077). Several of these scaffolds contain very high values of differentiation upwards of F_{ST} = 0.57. For one scaffold (scaffold00043), high F_{ST} windows span several Mb (Figure 2; Table S3). In total we identified 2,706 SNPs with elevated F_{ST} values (0.24 % of 1,120,260 total SNPs analyzed) on these 30 scaffolds. We took one of the

scaffolds with highest F_{ST} values (scaffold00102; F_{ST} = 0.57 at positions 1,971,662; 1,972,081 and 1,980,357) and repeated the analysis multiple times, leaving out one individual each time, to ensure the signal was not dependent on the inclusion of any one individual and to account for the relatedness detected in our data (Figure S5). PCAs generated for these 30 scaffolds did not reveal any separation between moribund and survivors, however when extracting only SNPs from the regions of high F_{ST}, there is moderate to strong separation between moribund and survivors in the PCA plots for several scaffolds (Figure S9). The groups display near complete separation along PC1 for scaffolds 22 (45.89 % PVE), 43 (64.17 % PVE), 59 (72.07% PVE and 102 (51.65% PVE) (Figures S9) which is reflected in the DAPC and compoplots (Figure S10).

Candidate genes and gene ontology

We identified 904 unique genes within 250 kb on either side of the 234 iHS outlier peaks from across the phased genome (scaffolds > 1Mb). Gene ontology (GO) analysis identified 36 enriched gene pathways among the outlier gene set (Table S8). These included pathways involved in regulation of gene expression and protein translation, cellular metabolism and intracellular transport and regulation of cell death (Table S8).

Some of the genes associated with significant iHS peaks are well investigated immune genes, like the gene encoding the pro inflammatory cytokine *IL17A* (Interleukin 17A) and *NLRP1B* (NLR Family Pyrin Domain Containing 1) which is involved in innate immunity and inflammation as a component of the NLR1 inflammasome (Mitchell, Sandstrom, & Vance, 2019; van de Veerdonk, Netea, Dinarello, & Joosten, 2011) (Figure 3). Moreover, several of the significant iHS peaks were found within genes (n = 18), including NLRP1B (Table S9). One of the top 10 most significant iHS peaks is located on scaffold00144 just upstream (34.3kb) of the

399 *ARHGEF25* (Rho guanine nucleotide exchange factor 25) gene (Figure 3). This gene works as
400 a guanine nucleotide exchange factor for Rho family of small GTPases (Rossman, Der, &
401 Sondek, 2005).

402 We conducted two complementary genome scans aiming at locating differences
403 between moribund and surviving gerbils, i.e. the xpEHH and F_{ST} pairwise analyses. From the
404 xpEHH analysis, we identified 385 unique genes within 250 kb on either side of the 122 outlier
405 peaks across the phased genome (scaffolds > 1 Mb). GO analysis yielded 52 enriched pathways
406 among the gene outlier set and included several immune related pathways (Table S10). In
407 addition, pathways involved in chromatin structure and gene regulation, transcription and
408 translation were also identified, as well as glucose and energy metabolism (Table S10). Lastly,
409 the GO analysis identified enriched pathways involved in cytoskeletal dynamics and
410 apoptosis (Table S10).

411 Several significant xpEHH peaks are found within genes including *CXCL14* (C-X-C
412 Motif Chemokine Ligand 14), *OR13A1* (olfactory receptor 13A1) and *BPGM*
413 (bisphosphoglycerate mutase) on scaffold00061, scaffold00376 and scaffold00548, encoding a
414 chemokine, an olfactory receptor and a trifunctional enzyme with mutase, synthase and
415 phosphatase activity, respectively (Table S9) (Fothergill-Gilmore & Watson, 1989; Hromas et
416 al., 1999; Shurin et al., 2005). Two of the top 10 most significant xpEHH peaks are found on
417 scaffold00056 and scaffold00136 (Table S7). Within scaffold00056 the peak overlaps with the
418 *ABCG3* (ATP-binding cassette sub-family G member 3) gene and *GBP6* (Guanylate-binding
419 protein 6) is located downstream of the peak, closely associated with other outlier values (10
420 kb downstream of closest significant outlier), while *CYCS* (Cytochrome C, Somatic) is located
421 just downstream (41 kb) of the peak on scaffold00136 (Figure 4B and C). In addition, among

the candidate genes are *CD180* (previously known as *RP105*), encoding a cell surface molecule present on immune cells, and located just upstream (42.5 kb) of a highly significant xpEHH peak on scaffold00091 (Figure S11B).

We further identified 565 unique genes falling within 250 kb on either side of the elevated F_{ST} regions (F_{ST} > 0.2) on the 30 identified scaffolds with higher divergence. GO analysis revealed 11 enriched pathways among the gene set (Table S11) and again included pathways involved in innate immunity, intracellular transport and translation; *Macrophage activation involved in immune response* (p = 0.031), *SRP-dependent cotranslational protein targeting to membrane* ($p = 1.68 \times 10^{-11}$) and *ribosome biogenesis* ($p = 4.72 \times 10^{-5}$). We additionally identified 32 genes that had one or more significantly elevated SNP located within their annotated boundaries (Table S9). A gene of particular interest from this subset is *MLF1* (myeloid leukemia factor 1) on scaffold00022 due to its involvement in lineage commitment of primary hemopoietic progenitor cells (Figure 4A). No known genes fall within the cluster of large F_{ST} values containing the highest peak on scaffold00102, but two genes were located nearby (Figure 4C). These were *FTSJ2* (also known as *MRM2* - mitochondrial rRNA methyltransferase 2) located upstream (185 kb) of the peak and *ZFAT* (zinc finger and AT-hook domain containing) located downstream of the peak (177 kb downstream of closest F_{ST} > 0.2).

Of the total 950 genes identified in the xpEHH and F_{ST} analyses, 24 candidate genes were shown to be overlapping (Table 2). One of these genes is the *VDAC1* (Voltage dependent anion channel 1) gene, which is associated with an F_{ST} peak identified on scaffold00043, as well as signatures for selection in the surviving gerbils in xpEHH and nucleotide diversity analyses (Figure S11A).

Gene expression during infection

Eight individuals from each group (eight moribund and eight surviving, $n = 16$, seven males and nine females) were RNA sequenced, comprising of 15 of the DNA sequenced individuals and one additional individual (Table 1 and supplementary table S1). The mean total unique mapping against the reference genome was high (79.52 %) with low degree of multimapping (1.3 %). Differential expression (DE) analysis between moribund and surviving animals with edgeR resulted in 146 significantly DE expressed genes (FDR $p < 0.05$) (Table S12). GO analyses revealed 22 significantly enriched pathways in the full gene set and showed that the gerbil immune system is highly activated during infection with plague. Several prominent inflammatory pathways are enriched and include both cellular and humoral immunity such as *acute inflammatory response* ($p = 1.58 \times 10^{-9}$), *acute-phase response* ($p = 4.93 \times 10^{-6}$), *response to lipopolysaccharide* ($p = 1.14 \times 10^{-5}$) *leukocyte mediated immunity* ($p = 4.93 \times 10^{-12}$) and *humoral immune response* ($p = 3.50 \times 10^{-5}$) (Table S13). Of the 146 DE genes, 124 were upregulated and 22 were downregulated in moribund individuals (Table S12). Most of the enriched pathways reported for all 146 genes are those reported for just the upregulated genes (Table S14). Three significantly enriched pathways were reported for the 22 downregulated genes; *positive regulation of IL-6 production* ($p = 3.29 \times 10^{-4}$), *antibiotic catabolic process* ($p = 2.29 \times 10^{-4}$) and *positive regulation of G1/S transition of mitotic cell cycle* ($p = 1.03\ 10^{-3}$) (Table S15).

Twenty-two of the genes identified in the DE analysis overlapped with candidate genes identified in one or more of the genome-wide scans for recent positive selection (iHS and xpEHH) and differentiation (F_{ST}) (Table 3). For instance, the genes *PSD4* (PH and SEC7 domain-containing protein 4) and *SERPINA* (Serpin Family A Member 1) were found to be upregulated in moribund gerbils while two ribosomal genes *RPL27* and *RPSA* were downregulated.

Discussion

Whole genome sequencing of wild gerbils that either survived or succumbed to a plague challenge experiment showed no overall genomic divergence between infection response phenotypes. However, when scrutinizing the data, smaller genomic regions with high differentiation between survivors and moribund individuals reveal evidence of genetic divergence that might underlie the outcomes of plague infection in great gerbils. These regions encompass factors involved in transcription and translation, and the regulation of these processes. Notably, several genes are directly involved in immune functions, indirectly associated with immune function such as the regulation of apoptosis, as well as cellular metabolism. Hence, what determines the outcome of a plague infection in great gerbils might be a complex combination of differences in regulation of certain key immune and (intracellular) metabolic pathways affecting the ability to regulate and mobilize resources to fight the infection.

A genetic basis for differences in plague resistance in gerbils

Examining a disease trait in a single population is complicated, particularly if the disease resistance has a polygenic basis, then the causative variants are difficult to detect as they likely explain only a small proportion of the variation (Visscher et al., 2017). Despite an overall low genetic diversity of the population, we detected smaller genomic regions showing genetic differentiation between survivors and moribund gerbils. In particular, the genomic regions on scaffolds 22, 43, 59 and 102 identified in the F_{ST} analysis, are responsible for near complete and complete separation between the two groups with 46, 64, 72 and 52 % of the variance in the

490 structure explained, respectively. The three candidate genes that overlap or are found in
491 proximity to these regions of elevated differentiation on scaffold 22 and 102 are all involved
492 in regulation of transcription or believed to modify components involved in translation. Two
493 of them (*ZFAT* and *MLF1*) are implicated in regulating cell lineage commitment of various cell
494 populations of the immune system. *ZFAT* is located downstream one of the largest identified
495 peaks, on scaffold00102, and encodes a DNA binding protein thought to be involved in
496 transcriptional regulation. In mice, the *ZFAT* gene is strongly expressed in T cells and B cells
497 of immune-related tissues such as the thymus, spleen and lymph nodes (Koyanagi et al., 2008),
498 it plays a role in thymic T-cell development and peripheral T-cell homeostasis (Doi et al., 2012)
499 and has also been identified as an antiapoptotic molecule in a human leukemia cell line
500 (Fujimoto et al., 2009; Ishikura et al., 2015). Several diseases have been linked to genetic
501 variants of *ZFAT*, including autoimmune thyroid disease and severity of Hashimoto disease
502 in humans and susceptibility to enterotoxigenic *Escherichia coli* infection in pigs (Inoue et al.,
503 2012; Ji et al., 2016; Shirasawa et al., 2004). *MLF1*, overlapping with an F_{ST} peak on
504 scaffold00022, is involved in lineage commitment of primary hemopoietic progenitor cells
505 restricting erythroid formation and enhancing myeloid formation (Williams et al., 1999). In
506 humans, *MLF1* is also known as an oncogene where overexpression in hematopoietic cells is
507 associated with acute myeloid leukemia (Matsumoto et al., 2000). It is tempting to speculate
508 that there could be differences in critical cell populations of the immune system that tips the
509 scale in the gerbil's favor for the surviving individuals.

510 *Y. pestis* is known to interfere with numerous proteins encoded by many genes and
511 immune pathways manipulating the inflammatory response (Cornelis, 2002; Mukherjee, 2006;
512 Shao, 2008; Sweet, Conlon, Golenbock, Goguen, & Silverman, 2007). One of the strategies used

513 by the bacterium, is to induce immune cell death in neutrophils, macrophages and dendritic
514 cells by apoptosis as opposed to the highly inflammatory pyroptosis, as well as globally
515 depleting natural killer cells, effectively hampering the innate immune systems' ability to
516 properly respond to and alert the adaptive immune system about the ongoing infection
517 (Cornelis, 2002; Kerschen, Cohen, Kaplan, & Straley, 2004). In our study, we find evidence of
518 positive selection on gene(s) responsible for the regulation of apoptosis, i.e. the *CYCS* gene in
519 the surviving gerbils. The *CYCS* gene identified on scaffold00136 by the xpEHH analysis
520 (Figure 4D) encodes the small heme-containing protein, cytochrome *c*, which is central in
521 mitochondrial respiration by transferring electrons to the respiratory chain to maintain ATP
522 function (Kulikov et al., 2012). However, it is also involved in the apoptotic pathway when
523 released into the cytosol where it can trigger the activation of the caspase cascade in intrinsic
524 apoptotic pathway (Kulikov et al., 2012). Its release occurs through a pore in the outer
525 mitochondrial membrane generated by the oligomerization of the protein product of the
526 *VDAC1* (Voltage dependent anion channel 1) gene, which is associated with an F_{ST} peak
527 identified on scaffold00043, as well as signatures for selection in xpEHH and nucleotide
528 diversity analyses (Figure S11A) (Shoshan-Barmatz, Maldonado, & Krelin, 2017). It is possible
529 that the immune cells of surviving individuals are more capable of resisting bacterial induced
530 apoptosis due to genetic differences in these key apoptosis regulating genes.
531 We also discovered genetic differences associated with genes more directly involved
532 in regulation of innate and adaptive immune responses between surviving and dying gerbils.
533 The *GBP6* gene identified on scaffold00056 in association with an xpEHH peak encodes a
534 member of the family of guanylate binding proteins (GBPs) which are highly induced by
535 interferon gamma and other inflammatory cytokines and are documented to be involved in

innate immune functions mainly towards intracellular pathogens (Praefcke, 2018). Other GBP family members have been shown to control bacterial or protozoan infections in studies of *Listeria monocytogenes* and *Toxoplasma gondii,* conferring some resistance in murine models (Pilla-Moffett, Barber, Taylor, & Coers, 2016). Furthermore, we uncovered genetic divergence between the two gerbil groups regarding the cell surface molecule CD180. The pattern recognition receptor CD180 dimerizes with MD1 forming a cell surface complex working in concert with TLR4 on antigen presenting cells like macrophages, B cells and dendritic cells, controlling the recognition and signaling of lipopolysaccharide (LPS) found in gram-negative bacteria (Kimoto, Nagasawa, & Miyake, 2003; Ogata et al., 2000). In studies investigating LPS-induced systemic inflammation and sepsis, CD180 appeared to be involved in dampening inflammatory responses (Divanovic et al., 2005). CD180 has also been shown to have a potential role in host resistance to infections caused by other pathogens like *Staphylococcus aureus* (B. Liu et al., 2013). CD180 could potentially be an important regulator of immune responses during plague infection in a way that increase survival, as the signatures of selection associated with CD180 appeared to have occurred in the survivors (Figure S11B).

Signatures of positive selection

Our analysis of selection in the population as a whole (i.e. the iHS analysis) identified several immune related genes which could indicate an increased general resistance to pathogens in this species. A proper pro-inflammatory environment is required during an infection for an effective activation and execution of adaptive cellular responses while avoiding hyperactivation that can lead to excessive collateral damage. One of the identified candidate

genes, IL17A, belongs to a family of cytokines that are strong inducers of inflammation, signaling through a pathway that depends on the adaptor molecule ACT1 ultimately activating proinflammatory mediators such as the transcription factor nuclear factor-κB (NF-κB) (Gaffen, 2009). IL17A is expressed on several subsets of lymphocytes such as CD8+ T cells, natural killer (NK) cells and NKT cells. Its main function has been shown to be neutrophil recruitment, and IL17A is considered an important driver of inflammation and immunity to extracellular pathogens due to the highly pro-inflammatory program of gene expression it induces (Gaffen, 2009). Notably, several IL17 receptors have previously been highlighted as potential players in plague resistance in other rodent species (Blanchet et al., 2010; Tollenaere, Jacquet, et al., 2012b). Furthermore, the candidate gene NLRP1B is also a contributor to innate immunity and inflammation as a sensor component of the NLR1 inflammasome (Moayeri et al., 2010). Inflammasomes are multimolecular complexes in the cell cytosol consisting of members of the nucleotide-binding domain-leucine-rich repeat (NLR) family proteins and inactive pro-caspase-1 (CASP1) (Latz, 2010). Once activated, CASP1 triggers downstream inflammatory responses resulting in pyroptotic cell death. Inflammasomes are known to be activated by directly detecting pathogen ligands or their activities such as bacterial effector proteins or toxins (Mitchell et al., 2019). In mice, NLRP1B was found to be activated in response to *Bacillus anthracis* lethal toxin and certain Nlrp1b alleles conferred resistance in mice to anthrax spore challenge through successful activation of the inflammasome and release of cytokines orchestrating a potent neutrophil response (Moayeri et al., 2010; Terra et al., 2010). Investigations into great gerbil NLRP1B alleles might be an interesting avenue for further research into plague resistance in this species.

Some of the identified signatures of selection may be connected to the previously reported *MHCII DRB* gene duplication in gerbils (Nilsson et al., 2020), like the signal associated with the *PSD4* gene. Upregulation of *PSD4* is in this context of special interest as it controls the movement of *MHCII*-containing vesicles along the actin cytoskeleton in dendritic cells through the activation of the GTPase *ARL14* (ADP ribosylation factor like GTPase 14) (Paul et al., 2011). This might indicate that the gerbils have modified parts of the MHCII signaling system/pathway and how (effectively) bacterial antigens are presented to the adaptive immune system.

Several candidate genes identified in our genome scans has functions that can be connected to cytoskeletal functions. The connection to actin cytoskeleton regulation is particularly interesting as *Y. pestis* is known to impair the cytoskeletal dynamics of phagocytes, blunting phagocytosis and paralyzing dendritic cell movement in the host (Aepfelbacher & Heesemann, 2001; Velan et al., 2006). The seventh strongest signal of selection in the iHS analysis was located upstream of *ARHGEF25*. As a guanine nucleotide exchange factor for Rho family of small GTPases, it is reportedly involved in multiple physiologic functions such as vascular smooth muscle contractility and cell movement (Hall, 1998; Lutz et al., 2007; Momotani et al., 2011; Swenson-Fields et al., 2008).

Plague resistance in great gerbil impacted by sex-bias?

Our results could indicate that the great gerbil display a sex-bias in its plague resistance, as we see that females are more prone to die from plague, while males are better survivors. Differences in immune functions between males and females are well known in mammals and

602 other animals, where it has been theoretically and empirically proven that females have a more
603 sensitive whereas males a more specific immune system (Klein & Flanagan, 2016; Metcalf &
604 Graham, 2018) meaning that females are better at detecting a broad specter of pathogens while
605 males mount a more efficient immune response. Recent studies in mice suggests that immune
606 responses to microbial infections are influenced not only by mouse genotype but also sex as
607 females had lower survival rates than males with the same genotype (Shutinoski et al., 2019).
608 Further scrutiny into specific alleles of candidate genes identified in this study should
609 therefore also factor in sex. However, it should also be noted that the females had generally a
610 lower body weight than the males used in this study, and thus it is possible that the sex-bias
611 in plague resistance observed here could be caused by a relative higher dose of *Y. pestis* per g
612 bodyweight in females compared to males.
613

614 **Gene expression differences in plague resistance**

615 The upregulated pathways in moribund gerbils during plague infection are those commonly
616 involved in fighting gram-negative bacteria, such as *humoral immune response*, *blood coagulation*,
617 *response to lipopolysaccharide*, *acute inflammatory response* and *myeloid leukocyte mediated immunity*
618 (Table S14). In effect, some of the significantly enriched pathways and genes might promote
619 plague infection by reducing the proinflammatory response and the differentiation and
620 proliferation of host immune cells (Gage & Ben Beard, 2017; Scheller, Chalaris, Schmidt-Arras,
621 & Rose-John, 2011; Tanaka, Narazaki, & Kishimoto, 2014). However, determining which
622 upregulated and downregulated genes and pathways are driven by and are beneficial to the
623 host and which are regulated due to manipulation by bacterial virulence genes, is hard to

disentangle. The differential expression in plague-challenged gerbils analyzed here was a comparison between moribund individuals in the midst of infection and survivors sampled at the end of the experiment, effectively done fighting the bacterium. This set up is a clear drawback for looking at differences in expression among gerbils dying from, or surviving a plague infection. However, design constraints limited the number of gerbils that could be included in the experiment. A stronger design would have been to sacrifice *n* survivors for every moribund individual, but would also require some form of biomarker for survival and/or death to make sure sampled individuals labelled as survivors are indeed going to survive the infection. This was and is currently not possible but would be important to pursue in future experiments.

Signatures of past population decline

Our demographic analysis of the gerbils shows a general declining trend in effective population size over the last 200 K years that appears to plateau around 20-50 K years ago (Figure S7). This coincides with expansion of deserts and sand fields in China occurring during the onset of the Last Glacial Maximum approximately 26 Kya characterized by a cold-dry climate. From around 10 Kya there is a strong decline with a minimum in population size by around 4-5 Kya. This is consistent with a reduction in available desert habitat during the Holocene Optimum around 5-9 Kya, when a transition to a warm-wet climate led to increased vegetation growth that nearly completely covered the sand fields (Lu et al., 2013). Effective population size reaches its minimum around 5 Kya, which likely explains the extremely low degree of nucleotide diversity observed. When effective population size becomes low, genetic

646 drift will also be stronger than selection further reducing the diversity (Star & Spencer, 2013).
647 The gerbils were captured in a relatively small area and in reality, are highly likely to be part
648 of the same (meta)population. Hence, the low divergence we observe most likely also reflects
649 the relatively small geographic range of the sampling area in combination with the social
650 structure and patterns of dispersal for gerbils (Randall, 2005; Y. Wang, Liu, Wang, Zhong, &
651 Wan, 2011). As the individuals in this experiment were all of same age (i.e. sub-adults) the
652 potential relationship would be full and half-siblings. Notably, even though there are some
653 degree of relatedness, we do detect differences in plague resistance. We believe the low level
654 of genetic differentiation strengthens our ability to detect differences between individuals that
655 can be connected to plague survival. For instance, in several pairs of individuals that appear
656 to be full siblings, one survived the infection while the other was deemed moribund. This
657 might suggest that specific alleles (or combinations of alleles) could confer resistance.
658 Whether plague (or a similarly lethal disease) might have impacted the gerbil
659 population during the last 6 Ky is difficult to determine due to the recent evolutionary history
660 of the pathogen, but is a prediction that could be tested in the future with a more detailed
661 demographic analysis.

662 ## Concluding remarks

663 The outcome of an infection is determined by the combined effects of collateral damage caused
664 by the host immune response and the direct (tissue) damage caused by the pathogen. Survival
665 in this context might depend on the balance and timing of host responses. Here, we reveal
666 evidence for selection on genes related to the innate and adaptive immune systems, as well as
667 in basic cellular functions such as regulation of transcription, translation and cellular

metabolism in the great gerbil. The involvement of the innate immune system mirrors that of Busch et al. 2011, 2013 who found the innate immune system to be involved in plague resistance in prairie dogs (Busch et al., 2011; 2013). Furthermore, we see evidence for selection on genes related to regulation of apoptosis which may be associated with an increased ability to survive a plague infection. Our findings suggest plague resistance is polygenic and that genetic differences among moribund gerbils and survivors are likely responsible for the outcome of the infection. Future work should aim to establish how these genetic differences confer functional differences in the regulation of cellular and immune functions during plague infection.

Acknowledgements

DNA sequencing was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no) and RNA sequencing by the Beijing Genomics Institute (BGI, https://www.bgi.com). We thank Cassandra N. Trier & Marine S. Brieuc, and Monica H. Solbakken for helpful discussions and advice regarding population analyses and transcriptomics, respectively. Lastly, we thank the anonymous reviewers for important suggestions for improving our analyses, manuscript and bringing to our attention the potential sex-bias in the outcome of infection and the statistics to show it. This project was funded by University of Oslo Molecular Life Science (MLS, allocation #152950), the Research Council of Norway (RCN grant #179569) and the European Research Council (ERC-2012-AdG No. 324249 -MedPlag).

References

Achtman, M., Morelli, G., Zhu, P., Wirth, T., Diehl, I., Kusecek, B., et al. (2004). Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proceedings of the National Academy of Sciences*, *101*(51), 17837–17842. http://doi.org/10.1073/pnas.0408026101

Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A., & Carniel, E. (1999). *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proceedings of the National Academy of Sciences*, *96*(24), 14043–14048. http://doi.org/10.1073/pnas.96.24.14043

Aepfelbacher, M., & Heesemann, J. (2001). Modulation of Rho GTPases and the actin cytoskeleton by Yersinia outer proteins (Yops). *International Journal of Medical Microbiology: IJMM*, *291*(4), 269–276. http://doi.org/10.1078/1438-4221-00130

Andrianaivoarimanana, V., Telfer, S., Rajerison, M., Ranjalahy, M. A., Andriamiarimanana, F., Rahaingosoamamitiana, C., et al. (2012). Immune responses to plague infection in wild *Rattus rattus*, in Madagascar: a role in foci persistence? *PLoS ONE*, *7*(6), e38630. http://doi.org/10.1371/journal.pone.0038630

Anisimov, A. P., Lindler, L. E., & Pier, G. B. (2004). Intraspecific Diversity of *Yersinia pestis*. *Clinical Microbiology Reviews*, *17*(2), 434–464. http://doi.org/10.1128/CMR.17.2.434-464.2004

Arbaji, A., Kharabsheh, S., Al-Azab, S., Al-Kayed, M., Amr, Z. S., Abu Baker, M., & Chu, M. C. (2005). A 12-case outbreak of pharyngeal plague following the consumption of camel meat, in north-eastern Jordan. *Annals of Tropical Medicine and Parasitology*, *99*(8), 789–793. http://doi.org/10.1179/136485905X65161

Barreiro, L. B., & Quintana-Murci, L. (2009). From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nature Reviews Genetics*, *11*(1), 17–30. http://doi.org/10.1038/nrg2698

Barreiro, L. B., Laval, G., Quach, H., Patin, E., & Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. *Nature Genetics*, *40*(3), 340–345. http://doi.org/10.1038/ng.78

Bertherat, E., Bekhoucha, S., Chougrani, S., Razik, F., Duchemin, J. B., Houti, L., et al. (2007). Plague reappearance in Algeria after 50 years, 2003. *Emerging Infectious Diseases*, *13*(10), 1459–1462. http://doi.org/10.3201/eid1310.070284

Best, S. M., & Kerr, P. J. (2000). Coevolution of host and virus: the pathogenesis of virulent and attenuated strains of myxoma virus in resistant and susceptible European rabbits. *Virology*, *267*(1), 36–48. http://doi.org/10.1006/viro.1999.0104

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, *25*(8), 1091–1093. http://doi.org/10.1093/bioinformatics/btp101

Blanchet, C., Jaubert, J., Carniel, E., Fayolle, C., Milon, G., Szatanik, M., et al. (2010). Mus spretus SEG/Pas mice resist virulent *Yersinia pestis*, under multigenic control. *Genes and Immunity*, *12*(1), 23–30. http://doi.org/10.1038/gene.2010.45

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. http://doi.org/10.1093/bioinformatics/btu170

Bonneaud, C., Balenger, S. L., Russell, A. F., Zhang, J., Hill, G. E., & Edwards, S. V. (2011). Rapid evolution of disease resistance is accompanied by functional changes in gene expression in a wild bird. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(19), 7866–7871. http://doi.org/10.1073/pnas.1018580108

Briskine, R. V., & Shimizu, K. K. (2017). Positional bias in variant calls against draft reference assemblies. *BMC Genomics*, *18*(1), 263. http://doi.org/10.1186/s12864-017-3637-2

Busch, J. D., Van Andel, R., Cordova, J., Colman, R. E., Keim, P., Rocke, T. E., et al. (2011). Population differences in host immune factors may influence survival of Gunnison's prairie dogs (Cynomys gunnisoni) during plague outbreaks. *Journal of Wildlife Diseases*, *47*(4), 968–973. http://doi.org/10.7589/0090-3558-47.4.968

Busch, J. D., Van Andel, R., Stone, N. E., Cobble, K. R., Nottingham, R., Lee, J., et al. (2013). The innate immune response may be important for surviving plague in wild Gunnison's prairie dogs. Journal of Wildlife Diseases, 49(4), 920–931. http://doi.org/10.7589/2012-08-209

Cagliani, R., & Sironi, M. (2013). Pathogen-driven selection in the human genome. *International Journal of Evolutionary Biology*, *2013*(2), 204240–6. http://doi.org/10.1155/2013/204240

Casanova, J.-L., & Abel, L. (2007). Human genetics of infectious diseases: a unified theory. *The EMBO Journal*, *26*(4), 915–922. http://doi.org/10.1038/sj.emboj.7601558

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. http://doi.org/10.1186/s13742-015-0047-8

Chevallier, L., Blanchet, C., Jaubert, J., Pachulec, E., Demeure, C., Carniel, E., et al. (2012). Resistance to plague of *Mus spretus* SEG/Pas mice requires the combined action of at least four genetic factors. *Genes and Immunity*, *14*(1), 35–41. http://doi.org/10.1038/gene.2012.50

Chomczynski, P., & Sacchi, N. (2006). The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nature Protocols*, *1*(2), 581–585. http://doi.org/10.1038/nprot.2006.83

Cingolani, P., Patel, V. M., Coon, M., Nguyen, T., Land, S. J., Ruden, D. M., & Lu, X. (2012). Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Frontiers in Genetics*, *3*, 35. http://doi.org/10.3389/fgene.2012.00035

Cornelis, G. R. (2002). Yersinia type III secretion: send in the effectors. *The Journal of Cell Biology*, *158*(3), 401–408. http://doi.org/10.1083/jcb.200205077

Cui, Y., Schmid, B. V., Cao, H., Dai, X., Du, Z., Easterday, W.R., et al. (2020). Evolutionary selection of biofilm-mediated extended phenotypes in *Yersinia pestis* in response to a fluctuating environment. *Nature Communications*, *11*(1), 281–8. http://doi.org/10.1038/s41467-019-14099-w

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. http://doi.org/10.1093/bioinformatics/btr330

Demeure, C. E., Blanchet, C., Fitting, C., Fayolle, C., Khun, H., Szatanik, M., et al. (2012). Early systemic bacterial dissemination and a rapid innate immune response characterize genetic resistance to plague of SEG mice. *The Journal of Infectious Diseases*, *205*(1), 134–143. http://doi.org/10.1093/infdis/jir696

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. http://doi.org/10.1038/ng.806

Divanovic, S., Trompette, A., Atabani, S. F., Madan, R., Golenbock, D. T., Visintin, A., et al. (2005). Negative regulation of Toll-like receptor 4 signaling by the Toll-like receptor homolog RP105. *Nature Immunology*, *6*(6), 571–578. http://doi.org/10.1038/ni1198

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. http://doi.org/10.1093/bioinformatics/bts635

Doi, K., Fujimoto, T., Okamura, T., Ogawa, M., Tanaka, Y., Mototani, Y., et al. (2012). ZFAT plays critical roles in peripheral T cell homeostasis and its T cell receptor-mediated response. *Biochemical and Biophysical Research Communications*, *425*(1), 107–112. http://doi.org/10.1016/j.bbrc.2012.07.065

Elgvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J., et al. (2017). The genomic mosaicism of hybrid speciation. *Science Advances*, *3*(6), e1602996. http://doi.org/10.1126/sciadv.1602996

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, *29*(1), 51–63. http://doi.org/10.1016/j.tree.2013.09.008

Fothergill-Gilmore, L. A., & Watson, H. C. (1989). The phosphoglycerate mutases. *Advances in Enzymology and Related Areas of Molecular Biology*, *62*, 227–313.

Fujimoto, T., Doi, K., Koyanagi, M., Tsunoda, T., Takashima, Y., Yoshida, Y., et al. (2009). ZFAT is an antiapoptotic molecule and critical for cell survival in MOLT-4 cells. *FEBS Letters*, *583*(3), 568–572. http://doi.org/10.1016/j.febslet.2008.12.063

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admetlla, A., Pattini, L., & Nielsen, R. (2011). Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. *PLOS Genetics*, *7*(11), e1002355. http://doi.org/10.1371/journal.pgen.1002355

Gaffen, S. L. (2009). Structure and signalling in the IL-17 receptor family. *Nature Reviews Immunology*, *9*(8), 556–567. http://doi.org/10.1038/nri2586

Gage, K. L., & Ben Beard, C. (2017). Plague. In J. Cohen, W. G. Powderly, & S. M. Opal (Eds.), *Infectious Diseases* (Fourth Edition, Vol. 2, pp. 1078–1084).

Gage, K. L., & Kosoy, M. Y. (2005). Natural history of plague: perspectives from more than a century of research. *Annual Review of Entomology*, *50*(1), 505–528. http://doi.org/10.1146/annurev.ento.50.071803.130337

Gascuel, F., Choisy, M., Duplantier, J.-M., Débarre, F., & Brouat, C. (2013). Host resistance, population structure and the long-term persistence of bubonic plague: contributions of a modelling approach in the Malagasy focus. *PLoS Computational Biology*, *9*(5), e1003039. http://doi.org/10.1371/journal.pcbi.1003039

Gautier, M., Klassmann, A., & Vitalis, R. (2017). rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources*, *17*(1), 78–90. http://doi.org/10.1111/1755-0998.12634

Haldane, J. B. S. (1932). The causes of evolution.
Hall, A. (1998). Rho GTPases and the actin cytoskeleton. *Science (New York, N.Y.)*, *279*(5350), 509–514.
Hill, A. V. (2001). The genomics and genetics of human infectious disease susceptibility. *Annual Review of Genomics and Human Genetics*, *2*(1), 373–400. http://doi.org/10.1146/annurev.genom.2.1.373
Hromas, R., Broxmeyer, H. E., Kim, C., Nakshatri, H., Christopherson, K., Azam, M., & Hou, Y. H. (1999). Cloning of BRAK, a novel divergent CXC chemokine preferentially expressed in normal versus malignant cells. *Biochemical and Biophysical Research Communications*, *255*(3), 703–706. http://doi.org/10.1006/bbrc.1999.0257
Hubbert, W. T., & Goldenberg, M. I. (1970). Natural resistance to plague: genetic basis in the vole (*Microtus californicus*). *The American Journal of Tropical Medicine and Hygiene*, *19*(6), 1015–1019.
Inoue, N., Watanabe, M., Yamada, H., Takemura, K., Hayashi, F., Yamakawa, N., et al. (2012). Associations between autoimmune thyroid disease prognosis and functional polymorphisms of susceptibility genes, CTLA4, PTPN22, CD40, FCRL3, and ZFAT, previously revealed in genome-wide association studies. *Journal of Clinical Immunology*, *32*(6), 1243–1252. http://doi.org/10.1007/s10875-012-9721-0
Ishikura, S., Ogawa, M., Doi, K., Matsuzaki, H., Iwaihara, Y., Tanaka, Y., et al. (2015). Zfat-deficient $CD4^+$ $CD8^+$ double-positive thymocytes are susceptible to apoptosis with deregulated activation of p38 and JNK. *Journal of Cellular Biochemistry*, *116*(1), 149–157. http://doi.org/10.1002/jcb.24954
Ji, H. Y., Yang, B., Zhang, Z. Y., Ouyang, J., Yang, M., Zhang, X. F., et al. (2016). A genome-wide association analysis for susceptibility of pigs to enterotoxigenic *Escherichia coli* F41. *Animal : an International Journal of Animal Bioscience*, *10*(10), 1602–1608. http://doi.org/10.1017/S1751731116000306
Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. http://doi.org/10.1093/bioinformatics/btn129
Kausrud, K. L., Viljugrein, H., Frigessi, A., Begon, M., Davis, S., Leirs, H., et al. (2007). Climatically driven synchrony of gerbil populations allows large-scale plague outbreaks. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1621), 1963–1969. http://doi.org/10.1098/rspb.2007.0568
Kerr, P. J., Cattadori, I. M., Liu, J., Sim, D. G., Dodds, J. W., Brooks, J. W., et al. (2017). Next step in the ongoing arms race between myxoma virus and wild rabbits in Australia is a novel disease phenotype. *Proceedings of the National Academy of Sciences*, *114*(35), 9397–9402. http://doi.org/10.1073/pnas.1710336114
Kerschen, E. J., Cohen, D. A., Kaplan, A. M., & Straley, S. C. (2004). The Plague Virulence Protein YopM Targets the Innate Immune Response by Causing a Global Depletion of NK Cells. *Infection and Immunity*, *72*(8), 4589–4602. http://doi.org/10.1128/IAI.72.8.4589-4602.2004
Kimoto, M., Nagasawa, K., & Miyake, K. (2003). Role of TLR4/MD-2 and RP105/MD-1 in innate recognition of lipopolysaccharide. *Scandinavian Journal of Infectious Diseases*, *35*(9), 568–572. http://doi.org/10.1080/00365540310015700
Klein, S. L., & Flanagan, K. L. (2016). Sex differences in immune responses. *Nature Reviews Immunology*, *16*(10), 626–638. http://doi.org/10.1038/nri.2016.90

867 Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant
868 call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53.
869 http://doi.org/10.1111/1755-0998.12549
870 Koyanagi, M., Nakabayashi, K., Fujimoto, T., Gu, N., Baba, I., Takashima, Y., et al. (2008).
871 ZFAT expression in B and T lymphocytes and identification of ZFAT-regulated genes.
872 *Genomics*, *91*(5), 451–457. http://doi.org/10.1016/j.ygeno.2008.01.009
873 Kulikov, A. V., Shilov, E. S., Mufazalov, I. A., Gogvadze, V., Nedospasov, S. A., &
874 Zhivotovsky, B. (2012). Cytochrome c: the Achilles' heel in apoptosis. *Cellular and*
875 *Molecular Life Sciences*, *69*(11), 1787–1797. http://doi.org/10.1007/s00018-011-0895-z
876 Latz, E. (2010). The inflammasomes: mechanisms of activation and function. *Current Opinion*
877 *in Immunology*, *22*(1), 28–33. http://doi.org/10.1016/j.coi.2009.12.004
878 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
879 transform. *Bioinformatics*, *25*(14), 1754–1760. http://doi.org/10.1093/bioinformatics/btp324
880 Liu, B., Fu, Y., Feng, S., Zhang, X., Liu, Z., Cao, Y., et al. (2013). Involvement of RP105 and
881 toll-like receptors in the activation of mouse peritoneal macrophages by *Staphylococcus*
882 *aureus*. *Scandinavian Journal of Immunology*, *78*(1), 8–16. http://doi.org/10.1111/sji.12050
883 Lowell, J. L., Antolin, M. F., Andersen, G. L., Hu, P., Stokowski, R. P., & Gage, K. L. (2015).
884 Single-Nucleotide Polymorphisms Reveal Spatial Diversity Among Clones of *Yersinia*
885 *pestis* During Plague Outbreaks in Colorado and the Western United States. *Vector Borne*
886 *and Zoonotic Diseases (Larchmont, N.Y.)*, *15*(5), 291–302.
887 http://doi.org/10.1089/vbz.2014.1714
888 Lu, H., Yi, S., Xu, Z., Zhou, Y., Zeng, L., Zhu, F., et al. (2013). Chinese deserts and sand fields
889 in Last Glacial Maximum and Holocene Optimum. *Chinese Science Bulletin*, *58*(23), 2775–
890 2783. http://doi.org/10.1007/s11434-013-5919-7
891 Lutz, S., Shankaranarayanan, A., Coco, C., Ridilla, M., Nance, M. R., Vettel, C., et al. (2007).
892 Structure of Galphaq-p63RhoGEF-RhoA complex reveals a pathway for the activation of
893 RhoA by GPCRs. *Science (New York, N.Y.)*, *318*(5858), 1923–1927.
894 http://doi.org/10.1126/science.1147554
895 Marshall, I. D., & Fenner, F. (1958). Studies in the epidemiology of infectious myxomatosis of
896 rabbits. V. Changes in the innate resistance of Australian wild rabbits exposed to
897 myxomatosis. *Journal of Hygiene*, *56*, 288–302.
898 Matsumoto, N., Yoneda-Kato, N., Iguchi, T., Kishimoto, Y., Kyo, T., Sawada, H., et al. (2000).
899 Elevated MLF1 expression correlates with malignant progression from myelodysplastic
900 syndrome. *Leukemia*, *14*(10), 1757–1765.
901 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010).
902 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation
903 DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.
904 http://doi.org/10.1101/gr.107524.110
905 Metcalf, C. J. E., & Graham, A. L. (2018). Schedule and magnitude of reproductive
906 investment under immune trade-offs explains sex differences in immunity. *Nature*
907 *Communications*, *9*(1), 4391–9. http://doi.org/10.1038/s41467-018-06793-y
908 Milholland, B., Dong, X., Zhang, L., Hao, X., Suh, Y., & Vijg, J. (2017). Differences between
909 germline and somatic mutation rates in humans and mice. *Nature Communications*, *8*,
910 15183. http://doi.org/10.1038/ncomms15183

911 Mitchell, P. S., Sandstrom, A., & Vance, R. E. (2019). The NLRP1 inflammasome: new
912 mechanistic insights and unresolved mysteries. *Current Opinion in Immunology*, *60*, 37–45.
913 http://doi.org/10.1016/j.coi.2019.04.015
914 Moayeri, M., Crown, D., Newman, Z. L., Okugawa, S., Eckhaus, M., Cataisson, C., et al.
915 (2010). Inflammasome sensor Nlrp1b-dependent resistance to anthrax is mediated by
916 caspase-1, IL-1 signaling and neutrophil recruitment. *PLoS Pathogens*, *6*(12), e1001222.
917 http://doi.org/10.1371/journal.ppat.1001222
918 Momotani, K., Artamonov, M. V., Utepbergenov, D., Derewenda, U., Derewenda, Z. S., &
919 Somlyo, A. V. (2011). p63RhoGEF Couples G αq/11-Mediated Signaling to Ca 2+
920 Sensitization of Vascular Smooth Muscle Contractility. *Circulation Research*, *109*(9), 993–
921 1002. http://doi.org/10.1161/CIRCRESAHA.111.248898
922 Morelli, G., Song, Y., Mazzoni, C. J., Eppinger, M., Roumagnac, P., Wagner, D. M., et al.
923 (2010). *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic
924 diversity. *Nature Reviews Genetics*, *42*(12), 1140–1143. http://doi.org/10.1038/ng.705
925 Möller, M., Kinnear, C. J., Orlova, M., Kroon, E. E., van Helden, P. D., Schurr, E., & Hoal, E.
926 G. (2018). Genetic Resistance to *Mycobacterium tuberculosis* Infection and Disease.
927 *Frontiers in Immunology*, *9*, 2219. http://doi.org/10.3389/fimmu.2018.02219
928 Mukherjee, S. (2006). Yersinia YopJ Acetylates and Inhibits Kinase Activation by Blocking
929 Phosphorylation. *Science (New York, N.Y.)*, *312*(5777), 1211–1214.
930 http://doi.org/10.1126/science.1126867
931 Nilsson, P., Solbakken, M. H., Schmid, B. V., Orr, R. J. S., Lv, R., Cui, Y., et al. (2020). The
932 Genome of the Great Gerbil Reveals Species-Specific Duplication of an MHCII Gene.
933 *Genome Biology and Evolution*, *12*(2), 3832–3849. http://doi.org/10.1093/gbe/evaa008
934 Ogata, H., Su, I., Miyake, K., Nagai, Y., Akashi, S., Mecklenbräuker, I., et al. (2000). The toll-
935 like receptor protein RP105 regulates lipopolysaccharide signaling in B cells. *The Journal*
936 *of Experimental Medicine*, *192*(1), 23–29. http://doi.org/10.1084/jem.192.1.23
937 Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., et al.
938 (2010). Antagonistic coevolution accelerates molecular evolution. *Nature*, *464*(7286), 275–
939 278. http://doi.org/10.1038/nature08798
940 Paul, P., van den Hoorn, T., Jongsma, M. L. M., Bakker, M. J., Hengeveld, R., Janssen, L., et
941 al. (2011). A Genome-wide multidimensional RNAi screen reveals pathways controlling
942 MHC class II antigen presentation. *Cell*, *145*(2), 268–283.
943 http://doi.org/10.1016/j.cell.2011.03.023
944 Petrunina, O. M. (1951). A course of plague infection in great gerbils (*Rhombomys opimus*
945 Licht.) under experimental inoculation. *Trudy Sredneaziatskogo Nauchno-Issledovatelskogo*
946 *Protivochumnogo Instituta, Alma-Ata. Monogr.*, *1*, 17–25.
947 Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: an
948 efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and*
949 *Evolution*, *31*(7), 1929–1936. http://doi.org/10.1093/molbev/msu136
950 Pilla-Moffett, D., Barber, M. F., Taylor, G. A., & Coers, J. (2016). Interferon-Inducible GTPases
951 in Host Resistance, Inflammation and Disease. *Journal of Molecular Biology*, *428*(17), 3495–
952 3513. http://doi.org/10.1016/j.jmb.2016.04.032
953 Praefcke, G. J. K. (2018). Regulation of innate immune functions by guanylate-binding
954 proteins. *International Journal of Medical Microbiology : IJMM*, *308*(1), 237–245.
955 http://doi.org/10.1016/j.ijmm.2017.10.013

956 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007).
957 PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage
958 Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575.
959 http://doi.org/10.1086/519795
960 Rahelinirina, S., Rajerison, M., Telfer, S., Savin, C., Carniel, E., & Duplantier, J.-M. (2017). The
961 Asian house shrew *Suncus murinus* as a reservoir and source of human outbreaks of
962 plague in Madagascar. *PLoS Neglected Tropical Diseases*, *11*(11), e0006072.
963 http://doi.org/10.1371/journal.pntd.0006072
964 Randall, J. A. (2005). Flexible social structure of a desert rodent, *Rhombomys opimus*:
965 philopatry, kinship, and ecological constraints. *Behavioral Ecology*, *16*(6), 961–973.
966 http://doi.org/10.1093/beheco/ari078
967 Rasmussen, S., Allentoft, M. E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.-G., et al.
968 (2015). Early Divergent Strains of *Yersinia pestis* in Eurasia 5,000 Years Ago. *Cell*, *163*(3),
969 571–582. http://doi.org/10.1016/j.cell.2015.10.009
970 Ravinet, M., Elgvin, T. O., Trier, C., Aliabadian, M., Gavrilov, A., & Sætre, G.-P. (2018).
971 Signatures of human-commensalism in the house sparrow genome. *Proceedings. Biological*
972 *Sciences / the Royal Society*, *285*(1884), 20181246. http://doi.org/10.1098/rspb.2018.1246
973 Reijniers, J., Begon, M., Ageyev, V. S., & Leirs, H. (2014). Plague epizootic cycles in Central
974 Asia. *Biology Letters*, *10*(6), 20140302–20140302. http://doi.org/10.1098/rsbl.2014.0302
975 Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for
976 differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–
977 140. http://doi.org/10.1093/bioinformatics/btp616
978 Rocke, T. E., Williamson, J., Cobble, K. R., Busch, J. D., Antolin, M. F., & Wagner, D. M.
979 (2012). Resistance to Plague Among Black-Tailed Prairie Dog Populations. *Vector-Borne*
980 *and Zoonotic Diseases*, *12*(2), 111–116. http://doi.org/10.1089/vbz.2011.0602
981 Rossman, K. L., Der, C. J., & Sondek, J. (2005). GEF means go: turning on RHO GTPases with
982 guanine nucleotide-exchange factors. *Nature Reviews. Molecular Cell Biology*, *6*(2), 167–180.
983 http://doi.org/10.1038/nrm1587
984 Sackett, L. C., Collinge, S. K., & Martin, A. P. (2013). Do pathogens reduce genetic diversity
985 of their hosts? Variable effects of sylvatic plague in black-tailed prairie dogs. *Molecular*
986 *Ecology*, *22*(9), 2441–2455. http://doi.org/10.1111/mec.12270
987 Scheller, J., Chalaris, A., Schmidt-Arras, D., & Rose-John, S. (2011). The pro- and anti-
988 inflammatory properties of the cytokine interleukin-6. *Biochimica Et Biophysica Acta*,
989 *1813*(5), 878–888. http://doi.org/10.1016/j.bbamcr.2011.01.034
990 Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history
991 from multiple genome sequences. *Nature Genetics*, *46*(8), 919–925.
992 http://doi.org/10.1038/ng.3015
993 Schulte, R. D., Makus, C., Hasert, B., Michiels, N. K., & Schulenburg, H. (2010). Multiple
994 reciprocal adaptations and rapid genetic change upon experimental coevolution of an
995 animal host and its microbial parasite. *Proceedings of the National Academy of Sciences of the*
996 *United States of America*, *107*(16), 7359–7364. http://doi.org/10.1073/pnas.1003113107
997 Shao, F. (2008). Biochemical functions of Yersinia type III effectors. *Current Opinion in*
998 *Microbiology*, *11*(1), 21–29. http://doi.org/10.1016/j.mib.2008.01.005
999 Shirasawa, S., Harada, H., Furugaki, K., Akamizu, T., Ishikawa, N., Ito, K., et al. (2004). SNPs
1000 in the promoter of a B cell-specific antisense transcript, SAS-ZFAT, determine

susceptibility to autoimmune thyroid disease. *Human Molecular Genetics*, *13*(19), 2221–2231. http://doi.org/10.1093/hmg/ddh245

Shoshan-Barmatz, V., Maldonado, E. N., & Krelin, Y. (2017). VDAC1 at the crossroads of cell metabolism, apoptosis and cell stress. *Cell Stress*, *1*(1), 11–36. http://doi.org/10.15698/cst2017.10.104

Shurin, G. V., Ferris, R. L., Ferris, R., Tourkova, I. L., Perez, L., Lokshin, A., et al. (2005). Loss of new chemokine CXCL14 in tumor tissue is associated with low infiltration by dendritic cells (DC), while restoration of human CXCL14 expression in tumor cells causes attraction of DC both in vitro and in vivo. *Journal of Immunology (Baltimore, Md. : 1950)*, *174*(9), 5490–5498. http://doi.org/10.4049/jimmunol.174.9.5490

Shutinoski, B., Hakimi, M., Harmsen, I. E., Lunn, M., Rocha, J., Lengacher, N., et al. (2019). Lrrk2 alleles modulate inflammation during microbial infection of mice in a sex-dependent manner. *Science Translational Medicine*, *11*(511), eaas9292. http://doi.org/10.1126/scitranslmed.aas9292

Sironi, M., Cagliani, R., Forni, D., & Clerici, M. (2015). Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nature Reviews Genetics*, *16*(4), 224–236. http://doi.org/10.1038/nrg3905

Star, B., & Spencer, H. G. (2013). Effects of genetic drift and gene flow on the selective maintenance of genetic variation. *Genetics*, *194*(1), 235–244. http://doi.org/10.1534/genetics.113.149781

Sweet, C. R., Conlon, J., Golenbock, D. T., Goguen, J., & Silverman, N. (2007). YopJ targets TRAF proteins to inhibit TLR-mediated NF-kappaB, MAPK and IRF3 signal transduction. *Cellular Microbiology*, *9*(11), 2700–2715. http://doi.org/10.1111/j.1462-5822.2007.00990.x

Swenson-Fields, K. I., Sandquist, J. C., Rossol-Allison, J., Blat, I. C., Wennerberg, K., Burridge, K., & Means, A. R. (2008). MLK3 limits activated Galphaq signaling to Rho by binding to p63RhoGEF. *Molecular Cell*, *32*(1), 43–56. http://doi.org/10.1016/j.molcel.2008.09.007

Tanaka, T., Narazaki, M., & Kishimoto, T. (2014). IL-6 in inflammation, immunity, and disease. *Cold Spring Harbor Perspectives in Biology*, *6*(10), a016295–a016295. http://doi.org/10.1101/cshperspect.a016295

Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, *Chapter 4*, Unit 4.10. http://doi.org/10.1002/0471250953.bi0410s25

Tencati, M., & Tapping, R. I. (2016). Resistance of mice of the 129 background to *Yersinia pestis* maps to multiple loci on chromosome 1. *Infection and Immunity*, IAI.00488–16. http://doi.org/10.1128/IAI.00488-16

Terra, J. K., Cote, C. K., France, B., Jenkins, A. L., Bozue, J. A., Welkos, S. L., et al. (2010). Cutting edge: resistance to *Bacillus anthracis* infection mediated by a lethal toxin sensitive allele of Nalp1b/Nlrp1b. *Journal of Immunology (Baltimore, Md. : 1950)*, *184*(1), 17–20. http://doi.org/10.4049/jimmunol.0903114

Tollenaere, C., Ivanova, S., Duplantier, J.-M., Loiseau, A., Rahalison, L., Rahelinirina, S., & Brouat, C. (2012a). Contrasted Patterns of Selection on MHC-Linked Microsatellites in Natural Populations of the Malagasy Plague Reservoir. *PLoS ONE*, *7*(3), e32814. http://doi.org/10.1371/journal.pone.0032814

Tollenaere, C., Jacquet, S., Ivanova, S., Loiseau, A., Duplantier, J. M., Streiff, R., & Brouat, C. (2012b). Beyond an AFLP genome scan towards the identification of immune genes involved in plague resistance in *Rattus rattus* from Madagascar. *Molecular Ecology*, *22*(2), 354–367. http://doi.org/10.1111/mec.12115

Tollenaere, C., Rahalison, L., Ranjalahy, M., Rahelinirina, S., Duplantier, J. M., & Brouat, C. (2008). CCR5 polymorphism and plague resistance in natural populations of the black rat in Madagascar. *Infection, Genetics and Evolution*, *8*(6), 891–897. http://doi.org/10.1016/j.meegid.2008.07.005

Trudeau, K. M., Britten, H. B., & Restani, M. (2004). Sylvatic plague reduces genetic variability in black-tailed prairie dogs. *Journal of Wildlife Diseases*, *40*(2), 205–211. http://doi.org/10.7589/0090-3558-40.2.205

Uchimura, A., Higuchi, M., Minakuchi, Y., Ohno, M., Toyoda, A., Fujiyama, A., et al. (2015). Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Research*, *25*(8), 1125–1134. http://doi.org/10.1101/gr.186148.114

van de Veerdonk, F. L., Netea, M. G., Dinarello, C. A., & Joosten, L. A. B. (2011). Inflammasome activation and IL-1β and IL-18 processing during infection. *Trends in Immunology*, *32*(3), 110–116. http://doi.org/10.1016/j.it.2011.01.003

Velan, B., Bar-Haim, E., Zauberman, A., Mamroud, E., Shafferman, A., & Cohen, S. (2006). Discordance in the effects of *Yersinia pestis* on the dendritic cell functions manifested by induction of maturation and paralysis of migration. *Infection and Immunity*, *74*(11), 6365–6376. http://doi.org/10.1128/IAI.00974-06

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, *101*(1), 5–22. http://doi.org/10.1016/j.ajhg.2017.06.005

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, *4*(3), e72. http://doi.org/10.1371/journal.pbio.0040072

Wang, Y., Liu, W., Wang, G. M., Zhong, W., & Wan, X. (2011). Genetic consequences of group living in Mongolian gerbils. *The Journal of Heredity*, *102*(5), 554–561. http://doi.org/10.1093/jhered/esr069

Williams, J. H., Daly, L. N., Ingley, E., Beaumont, J. G., Tilbrook, P. A., Lalonde, J. P., et al. (1999). HLS7, a hemopoietic lineage switch gene homologous to the leukemia-inducing gene MLF1. *The EMBO Journal*, *18*(20), 5559–5566. http://doi.org/10.1093/emboj/18.20.5559

Zhang, P., Liu, X., Wang, C., Zhao, Y., Hua, F., Li, C., et al. (2014). Evaluation of up-converting phosphor technology-based lateral flow strips for rapid detection of *Bacillus anthracis* Spore, *Brucella spp*., and *Yersinia pestis*. *PLoS ONE*, *9*(8), e105305. http://doi.org/10.1371/journal.pone.0105305

Zhang, Y., Dai, X., Wang, Q., Chen, H., Meng, W., Wu, K., et al. (2015). Transmission efficiency of the plague pathogen (*Y. pestis*) by the flea, *Xenopsylla skrjabini*, to mice and great gerbils. *Parasites & Vectors*, *8*(1), 256. http://doi.org/10.1186/s13071-015-0852-z

Zhang, Y., Dai, X., Wang, X., Maituohuti, A., Cui, Y., Rehemu, A., et al. (2012). Dynamics of *Yersinia pestis* and its antibody response in great gerbils (*Rhombomys opimus*) by

subcutaneous infection. *PLoS ONE*, *7*(10), e46820. http://doi.org/10.1371/journal.pone.0046820

Zhang, Y., Luo, T., Yang, C., Yue, X., Guo, R., Wang, X., et al. (2018). Phenotypic and Molecular Genetic Characteristics of *Yersinia pestis* at an Emerging Natural Plague Focus, Junggar Basin, China. *The American Journal of Tropical Medicine and Hygiene*, *98*(1), 231–237. http://doi.org/10.4269/ajtmh.17-0195

Zhang, Y.-J., Dai, X., Abulimiti, Jiang, W., Abulikemu, Wang, X.-H., et al. (2008). [Study on the situation of plague in Junggar Basin of China]. *Zhonghua Liu Xing Bing Xue Za Zhi = Zhonghua Liuxingbingxue Zazhi*, *29*(2), 136–144.

Data accessibility

The version of the reference genome used in this paper as well as the two differentially filtered annotation files are available at Figshare: https://figshare.com/s/9035ed40f970d0545d06 The genome assembly used here is a previous version of the assembly reported in Nilsson et al. 2020 (Nilsson et al., 2020), but prior to NCBI removal of a small duplicated gene and masking of minor contamination. Whole genome sequencing and RNA sequencing data will be available at the European Nucleotide Archive.

Author Contributions

PN, MR and PRB performed data analyses and wrote the first draft of the manuscript. ET generated the phased data and performed the demographic analysis. SNKH resuspended DNA samples, performed quality control and generated the sequencing libraries. YZ, RG and TL did field work, tended to the animals in the lab, performed the challenge experiment and extracted DNA. YC, YS, and RL helped design and assisted in the challenge experiment and DNA extractions, performed the RNA extraction and RNA library prep. PN, BVS, NCS, RY,

WRE, KSJ and SJ helped design the challenge experiment and oversaw the project. All authors contributed to the writing of the paper.

Figures

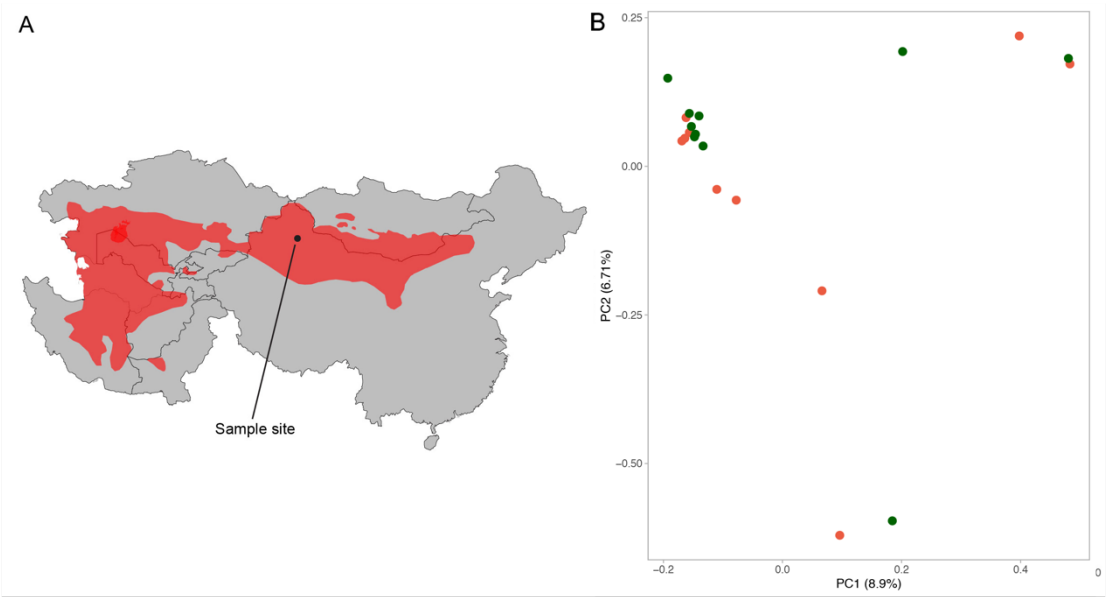


Figure 1. ***Gerbil distribution and population structure at the Xinjiang sampling site.*** (A) Distribution of gerbils in Central Asia, ranging from the Caspian Sea in the west to deep within China to the east. The sampling site in Xinjiang, China is marked by a black point and line. (B) Principal component analysis (PCA) of high-quality, LD pruned SNPs does not separate according to disease outcome. Dark green circles represent survivors and dark orange circles depict moribund individuals.

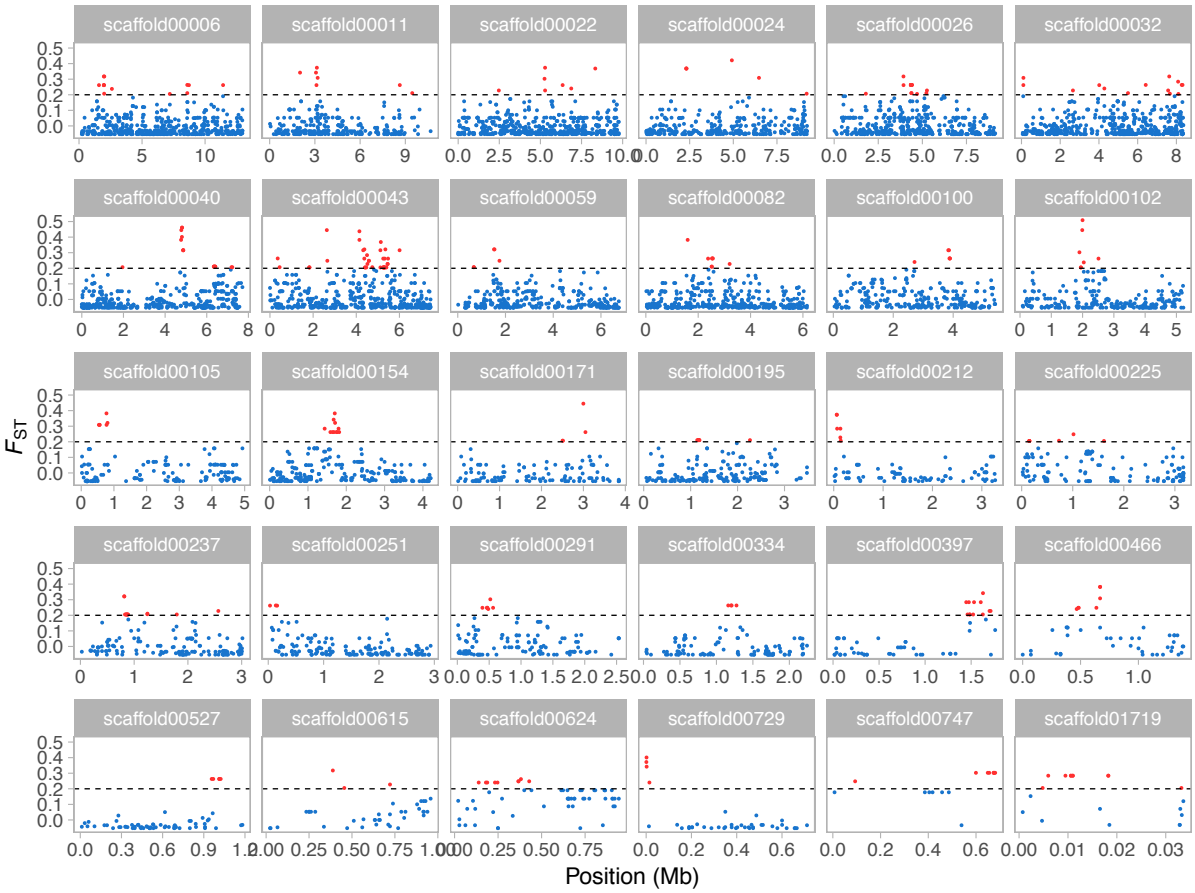


Figure 2. ***FST patterns for the 30 scaffolds containing elevated FST values.***
Plots of pairwise *FST* along the scaffolds shows clear peaks of differentiation (elevated *FST*), with particularly large peaks on scaffolds 22, 40 and 102 and an extended peak on scaffold00043. Horizontal dashed line represents the threshold of high *FST* values for outlier SNPs (*FST* > 0.2), also indicated by red points. The data has been subsampled to 0.3 using the sample_frac() command of the dplyr R package for visualization purposes only.

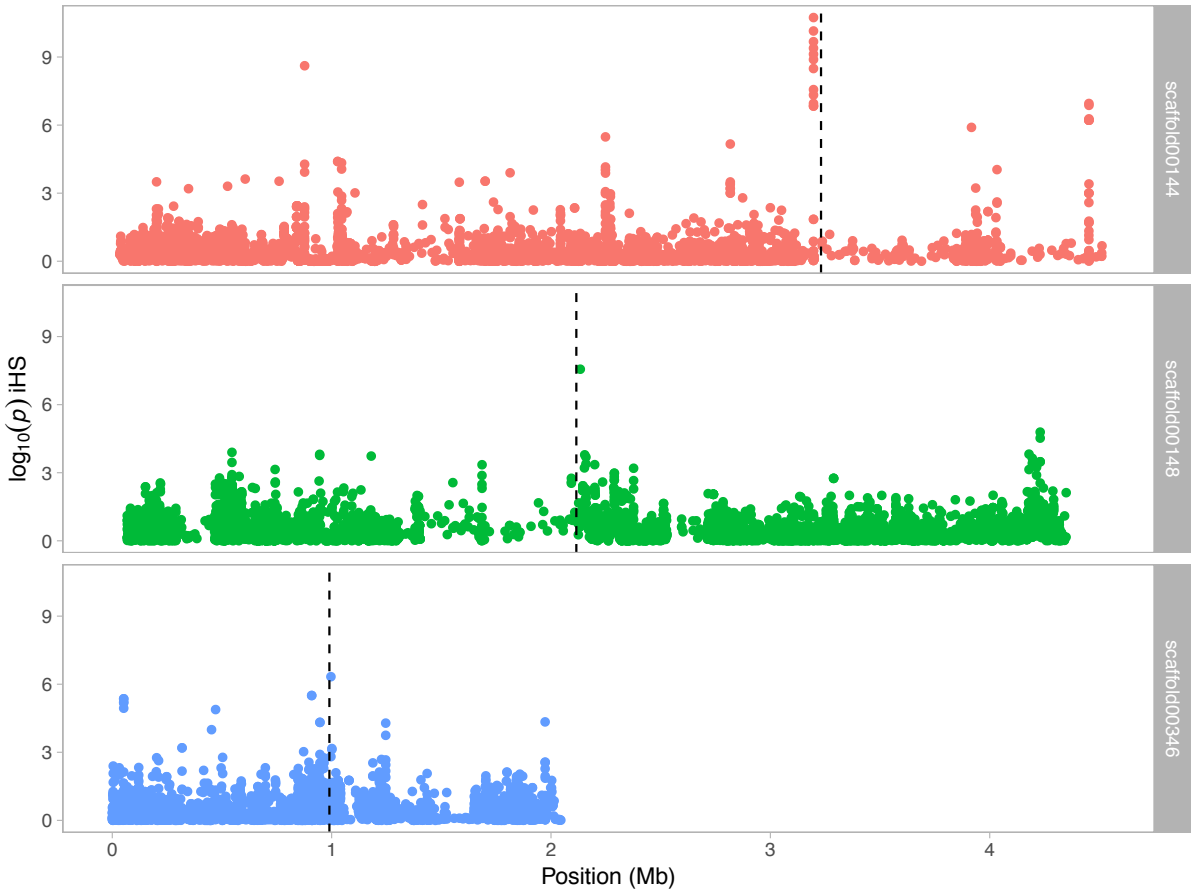
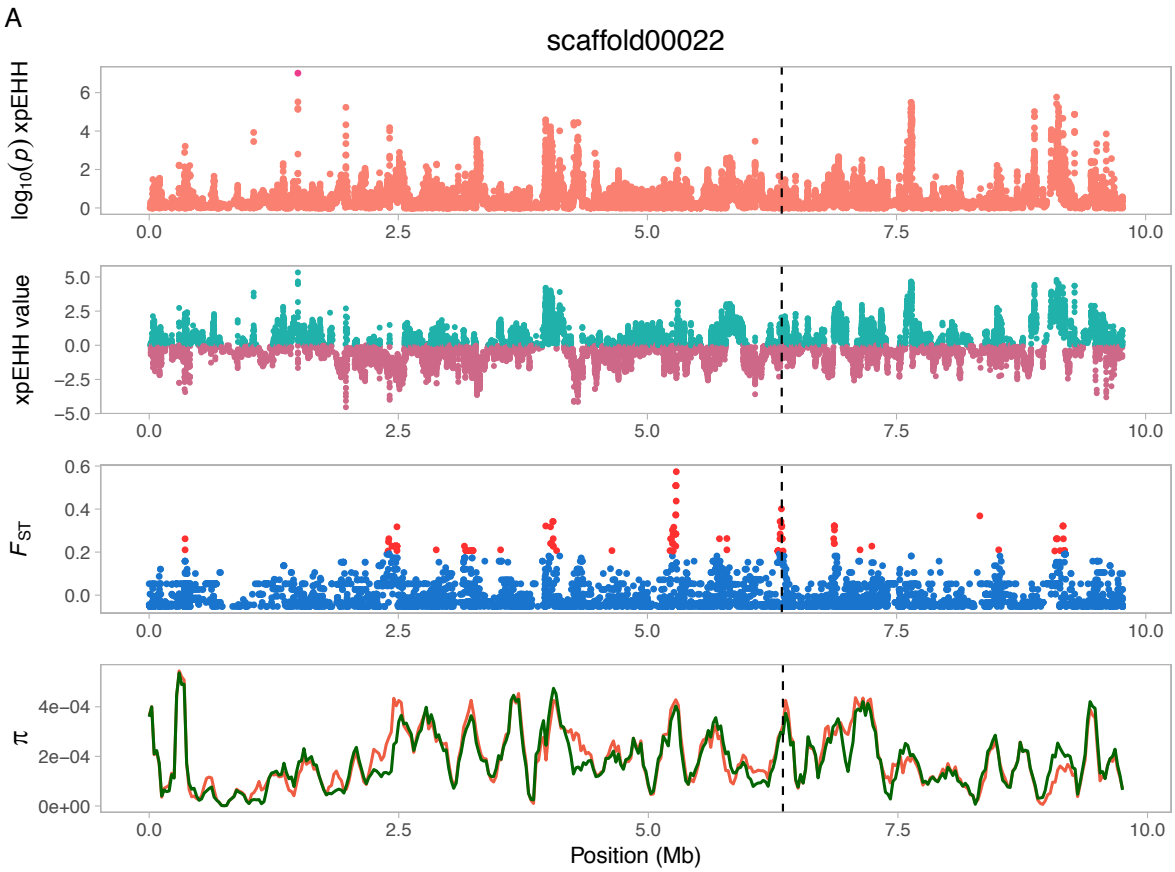


Figure 3. ***Signatures of recent selection associated with candidate genes on scaffolds 144, 148 and 346.***

The plot show $\log_{10}$ of the iHS analysis *p* value, where high values indicate strong signals of selection (threshold of significance set at $\log_{10}$ 1 x 10^{-6}). The scaffold numbers are displayed in the facet (on the right) and the vertical dashed line indicates the location of candidate genes. From top panel to bottom; ARHGEF25 (gene start), IL17A (gene midpoint) and NLRP1B (gene midpoint).

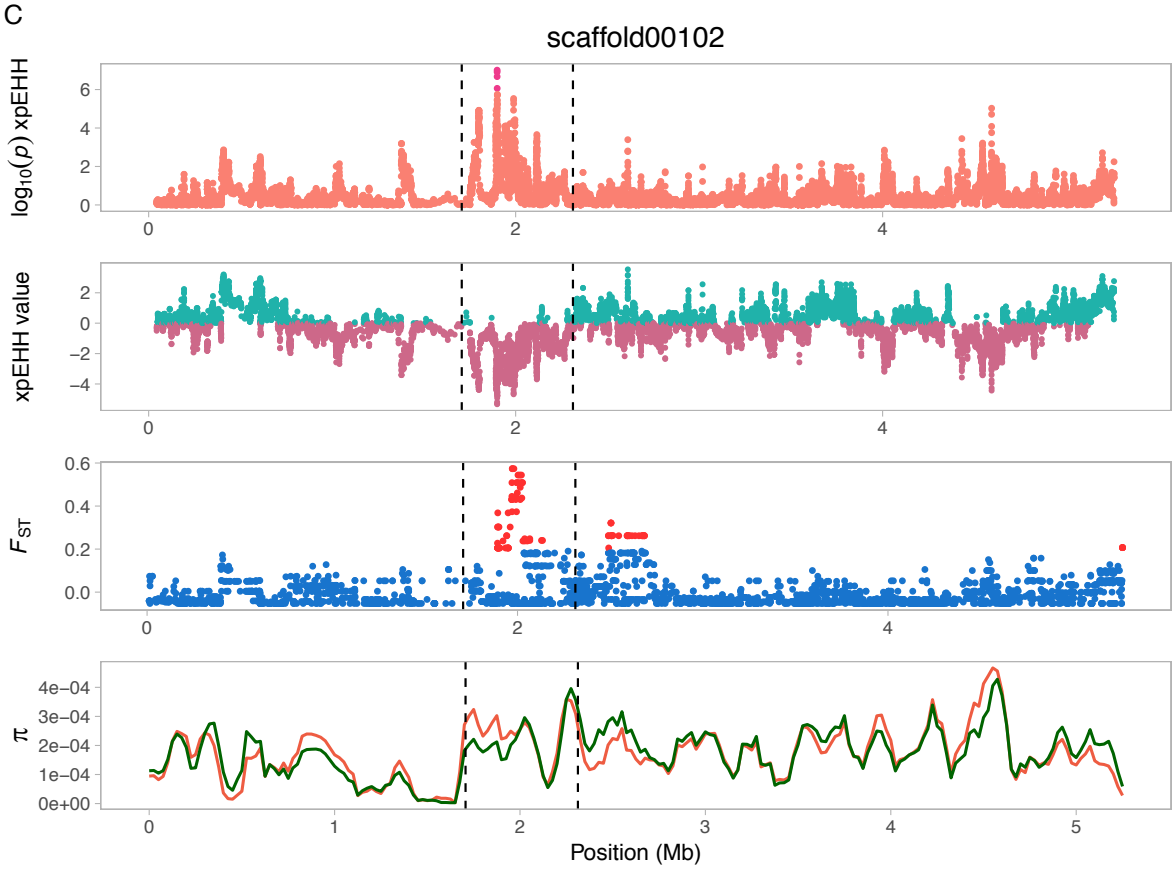

A
scaffold00022
log10(p) xpEHH
xpEHH value
FST
π
4e−04
2e−04
0e+00
Position (Mb)

1149


B
scaffold00056
log10(p) xpEHH
xpEHH value
FST
π
0.00075
0.00050
0.00025
0.00000
Position (Mb)

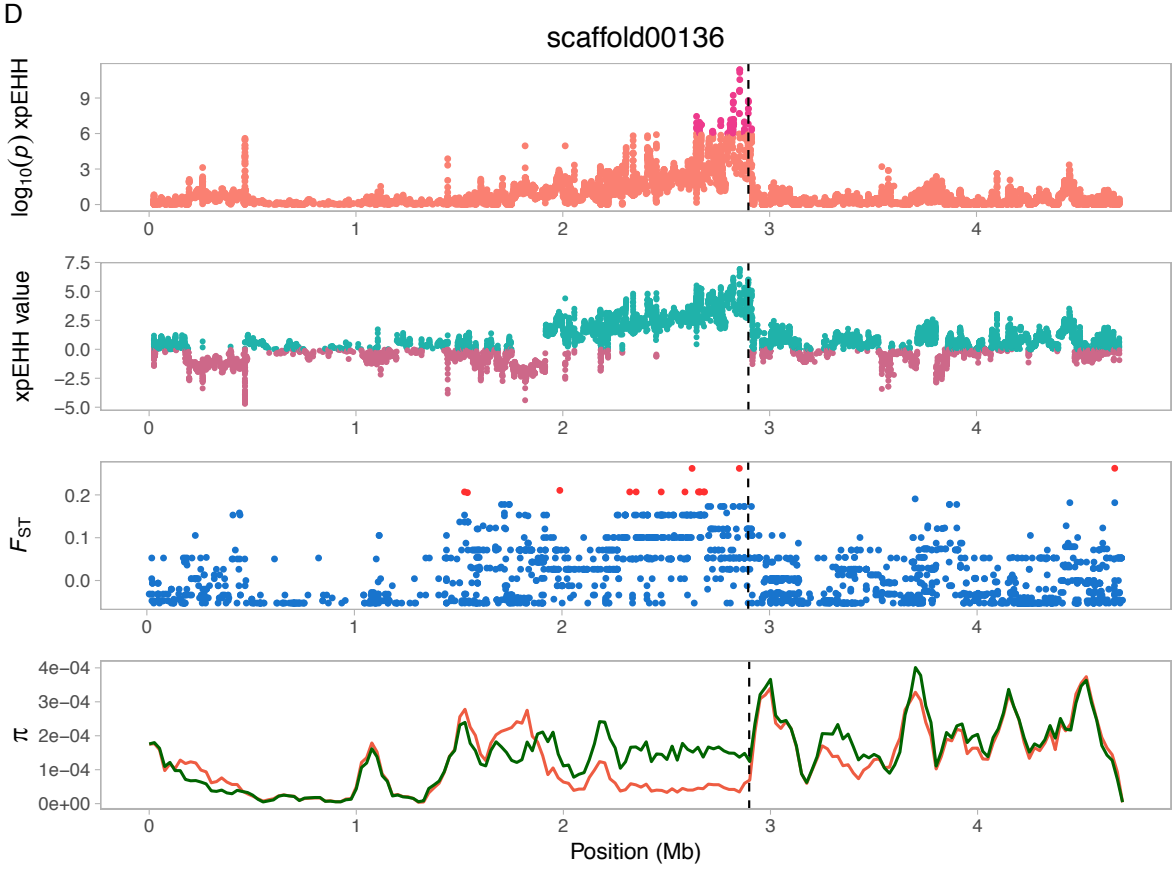
1150

C


scaffold00102
log10(p) xpEHH
xpEHH value
FST
π
Position (Mb)

1151

D


scaffold00136
log10(p) xpEHH
xpEHH value
FST
π
Position (Mb)

1152

Figure 4. ***Signatures of recent selection, differentiation and nucleotide diversity along scaffolds 22 (A), 56 (B), 102 (C) and 136 (D).*** For each; Top panel – Plot of log10 (*p*) xpEHH, where high values indicate strong signals of selection (threshold of significance set at $\log_{10} 1 \times 10^{-6}$) with outlier SNPs in hot pink, second panel; Plot of xpEHH value where positive values (teal) indicate selection has occurred in surviving gerbils, while negative values (dark violet) indicate selection in dying gerbils, third panel; relative differentiation (F_{ST}) between moribund and survivors (red = SNPs with elevated values, $F_{ST} > 0.2$), fourth panel – nucleotide diversity (π) of moribund (orange) and survivors (green). Vertical dashed lines represent in (A) the location of the *MLF1* gene within a significant F_{ST} peak, in (B) the location of *ABCG3* and *GBP6* (beginning) genes, in (C) the location of *FTSJ2* (end) and *ZFAT* (beginning) genes and in (D) the location of the CYCS gene. For the gene position in (A) and (D) the midpoint is used.

Tables

Table 1. ***Metadata on the 20 DNA sequenced gerbils used for population analyses.***
Information on sex, body weight and days of disease onset, recovery and sampling are listed. Those individuals who also have RNA sequenced are marked. The animals' status at sampling is separated into those who died of the infection ('Moribund') and those that survived ('Recovered'). Individuals denoted as 'Healthy' had no clinical symptoms observed. Sample names are composed of outcome (Dead or Survived, animal group and animal number in that group.

Sample name	*Sex*	*Body weight (g)*	*Day of onset of disease (p.i.)*	*Day of recovery (p.i.)*	*Day of sampling (p.i.)*	*Animal status at sampling*
D2-1[a]	Male	176	1		5	Moribund
D2-3	Female	187	1		3	Moribund
D2-4[a]	Male	176	1		5	Moribund
D2-5[a]	Male	224	1		5	Moribund
D4-3[a]	Female	132	2		5	Moribund
D5-1[a]	Female	144	1		4	Moribund
D8-3[a]	Female	123	1		5	Moribund
D9-5	Female	126	1		5	Moribund
D10-3[a]	Male	214	1		3	Moribund
D11-5[a]	Female	145	1		4	Moribund
S3-3[a]	Male	173	2	5	22	Recovered
S5-3	Male	161	2	6	22	Recovered
S6-4[a]	Female	141	1	7	22	Recovered
S8-1[a]	Female	126	1	4	22	Recovered
S9-1[a]	Male	202	1	6	22	Recovered
S9-3	Male	111	1	5	22	Recovered
S10-4[a]	Female	116	2	6	22	Recovered
S11-1[a]	Male	133	1	5	22	Recovered
S5-2[a]	Female	167	-	-	3	Healthy
S10-5	Female	118	-	-	3	Healthy

[a] RNA sequencing performed for individual

Table 2. ***Genes found in both xpEHH and F_{ST} genome scans.***
The table lists the 24 candidate genes identified in both analyses.

Gene	*Full name*	*Scaffold*
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	scaffold00006
PPIA	Peptidyl-prolyl cis-trans isomerase A	scaffold00022
RPL21	60S ribosomal protein L21	scaffold00022
RPL29	60S ribosomal protein L29	scaffold00022
RPS6	40S ribosomal protein S6	scaffold00022
HMGB2	High mobility group protein B2	scaffold00032

AIMP1	Aminoacyl tRNA synthase complex-interacting multifunctional protein 1	scaffold00043
ECHDC1	Ethylmalonyl-CoA decarboxylase	scaffold00043
KIAA0408	Uncharacterized protein KIAA0408	scaffold00043
LEG1	Protein LEG1 homolog	scaffold00043
RPL7A	60S ribosomal protein L7a	scaffold00043
SOGA3	SOGA Family Member 3	scaffold00043
THEMIS	Thymocyte Selection Associated	scaffold00043
TMEM200A	Transmembrane Protein 200A	scaffold00043
UBE2E1	Ubiquitin-conjugating enzyme E2 E1	scaffold00043
VDAC1	Voltage-dependent anion-selective channel protein 1	scaffold00043
RPSA	40S ribosomal protein SA	scaffold00059
HNRNPH2	Heterogeneous nuclear ribonucleoprotein H2	scaffold00100
FTSJ2 (MRM2)	Mitochondrial rRNA methyltransferase 2	scaffold00102
RPS23	40S ribosomal protein S23	scaffold00171
BCL2L13	Bcl-2-like protein 13	scaffold00225
BID	BH3-interacting domain death agonist	scaffold00225
MICAL3	Microtubule Associated Monooxygenase, Calponin And LIM Domain Containing 3	scaffold00225
RPL28	60S ribosomal protein L28	scaffold00237

Table 3. ***Genes identified in both DE and genome scan analyses.***
The table lists the 22 genes identified as differentially expressed between survivors and moribund that was also identified in one or more genome scan analyses.

Gene	*Full name*	*Scaffold*	*Regulation in moribund*	*Genome scan*
SYCP3	Synaptonemal complex protein 3	scaffold00003	Downregulated	iHS
SDC4	Syndecan-4	scaffold00006	Upregulated	fst
MMP9	Matrix metalloproteinase-9	scaffold00006	Upregulated	fst
FGF21	Fibroblast growth factor 21	scaffold00011	Upregulated	fst
RASIP1	Ras-interacting protein 1	scaffold00011	Upregulated	fst
PRSS29	Serine protease 29	scaffold00024	Upregulated	fst, iHS
MASTIN	Mastin	scaffold00024	Upregulated	fst
NPW	Neuropeptide W	scaffold00024	Upregulated	fst
NAA60	N-Alpha-Acetyltransferase 60, NatF Catalytic Subunit	scaffold00024	Upregulated	iHS
SERPINA1	1-antitrypsin	scaffold00030	Upregulated	xpEHH, iHS
HMGB2	High mobility group protein B2	scaffold00032	Downregulated	fst
CLK4	CDC Like Kinase 4	scaffold00039	Downregulated	xpEHH
SPP1	Secreted Phosphoprotein 1	scaffold00056	Upregulated	xpEHH
PLAC8	Placenta Associated 8	scaffold00056	Upregulated	iHS

RPSA	40S ribosomal protein SA	scaffold00059	Downregulated	fst, xpEHH, iHS
MRC2	Mannose Receptor C Type 2	scaffold00104	Upregulated	iHS
BST1	Bone Marrow Stromal Cell Antigen 1	scaffold00109	Upregulated	xpEHH
RPL27	60S ribosomal protein L27	scaffold00180	Downregulated	xpEHH, iHS
GRN	Granulin Precursor	scaffold00209	Upregulated	iHS
IGFBP2	Insulin-like growth factor-binding protein 2	scaffold00245	Upregulated	xpEHH
PCP4	Purkinje cell protein 4	scaffold00271	Upregulated	iHS
PSD4	PH and SEC7 domain-containing protein 4	scaffold00279	Upregulated	iHS

1192