A multitrait locus regulates sarbecovirus pathogenesis

Alexandra Schäfer  
*University of North Carolina at Chapel Hill*

Emma S Winkler  
*Washington University School of Medicine in St. Louis*

Michael S Diamond  
*Washington University School of Medicine in St. Louis*

et al

Follow this and additional works at: https://digitalcommons.wustl.edu/oa_4

Part of the Medicine and Health Sciences Commons

Please let us know how this document benefits you.

**Recommended Citation**

https://digitalcommons.wustl.edu/oa_4/203

This Open Access Publication is brought to you for free and open access by the Open Access Publications at Digital Commons@Becker. It has been accepted for inclusion in 2020-Current year OA Pubs by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
A Multitrait Locus Regulates Sarbecovirus Pathogenesis


ABSTRACT Infectious diseases have shaped the human population genetic structure, and genetic variation influences the susceptibility to many viral diseases. However, a variety of challenges have made the implementation of traditional human Genome-wide Association Studies (GWAS) approaches to study these infectious outcomes challenging. In contrast, mouse models of infectious diseases provide an experimental control and precision, which facilitates analyses and mechanistic studies of the role of genetic variation on infection. Here we use a genetic mapping cross between two distinct Collaborative Cross mouse strains with respect to severe acute respiratory syndrome coronavirus (SARS-CoV) disease outcomes. We find several loci control differential disease outcome for a variety of traits in the context of SARS-CoV infection. Importantly, we identify a locus on mouse chromosome 9 that shows conserved synteny with a human GWAS locus for SARS-CoV-2 severe disease. We follow-up and confirm a role for this locus, and identify two candidate genes, CCR9 and CXCR6, that both play a key role in regulating the severity of SARS-CoV, SARS-CoV-2, and a distantly related bat sarbecovirus disease outcomes. As such we provide a template for using experimental mouse crosses to identify and characterize multitrait loci that regulate pathogenic infectious outcomes across species.

IMPORTANCE Host genetic variation is an important determinant that predicts disease outcomes following infection. In the setting of highly pathogenic coronavirus infections, genetic determinants underlying host susceptibility and mortality remain unclear. To elucidate the role of host genetic variation on sarbecovirus pathogenesis and disease outcomes, we utilized the Collaborative Cross (CC) mouse genetic reference population as a model to identify susceptibility alleles to SARS-CoV and SARS-CoV-2 infections. Our findings reveal that a multitrait loci found in chromosome 9 is an important regulator of sarbecovirus pathogenesis in mice. Within this locus, we identified and validated CCR9 and CXCR6 as important regulators of host disease outcomes. Specifically, both CCR9 and CXCR6 are protective against severe SARS-CoV, SARS-CoV-2, and SARS-related HKU3 virus disease in mice. This chromosome 9 multitrait locus may be important to help identify genes that regulate coronavirus disease outcomes in humans.

Editor John A. Lednicky, University of Florida
Copyright © 2022 Schäfer et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.
Address correspondence to Alexandra Schäfer, aschaefe@email.unc.edu, Martin T. Ferris, mtferris@email.unc.edu, or Ralph S. Baric, rbaric@email.unc.edu.
The authors declare no conflict of interest.
This article is a direct contribution from Ralph S. Baric, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Luis Enjuanes, Centro Nacional de Biotecnologia, CNB-CSIC, and David Threadgill, Texas A&M University.
Received 31 May 2022
Accepted 17 June 2022
Published 12 July 2022

Downloaded from https://journals.asm.org/journal/mbio on 12 September 2022 by 128.252.79.225.
KEYWORDS collaborative cross, host response, pathogenesis, SARS-CoV-2, sarbecoviruses

Studies support the hypothesis that natural host genetic variation contributes substantially to microbial susceptibility and disease severity (1, 2). However, the specific genes and alleles that regulate differential disease outcomes to respiratory virus infections remain largely unknown. The viral family Coronaviridae is comprised of several human and animal pathogens, and at least five zoonotic coronaviruses (CoV) have emerged or rapidly expanded their geographic range in mammals in the 21st century (3, 4). To date, the most significant human pathogens include the group 2B sarbecoviruses (severe acute respiratory syndrome coronavirus [SARS-CoV] and SARS-CoV-2), which likely emerged from bats to cause human epidemic or pandemic outbreaks of severe acute respiratory infections, leading to substantial morbidity and mortality (5, 6). Many other high-risk group 2B SARS-like viruses including bat sarbecoviruses (BtCoV), group 2C Middle East respiratory coronavirus (MERS-CoV), and MERS-like bat CoVs also appear poised to cause future human epidemics or pandemics (7–9). Emerging sarbecoviruses vary widely in their ability to cause human and animal disease (10). The 2003 SARS-CoV strain caused ~8,000 infections and ~800 deaths leading to a 10% mortality rate, whereas SARS-CoV-2 has infected >500 million humans, leading to over 6.2 million deaths in an ongoing pandemic (11, 12). SARS-CoV and SARS-CoV-2 demonstrate a broad host range with the ability to cause variable disease outcomes that range from asymptomatic infection to death in other mammals, raising questions whether common underlying host genes contribute to disease across species (13). While the 2003 SARS-CoV had a higher mortality rate compared to SARS-CoV-2, COVID-19 outcomes can vary from asymptomatic to life-threatening acute respiratory disease syndrome and death in humans, supporting a documented role for viral and interhost genetic control of disease severity (14–17). Mapping the underlying natural host gene variation that regulates susceptibility and disease severity after diverse sarbecovirus infections and across multiple species is expected to reveal common genetic loci that regulate pathogenic outcomes, inform threat and risk potential, lead to refined small animal models of human disease, and reveal novel targets for therapeutic interventions.

Mouse-adapted SARS-CoV and SARS-CoV-2 cause disease in mice by inducing acute respiratory distress syndrome (ARDS) and recapitulating several aspects of COVID-19 pathogenesis including age-dependent severe disease (18, 19). Mouse genetic reference populations (GRPs) also have been employed as highly relevant models of human disease and used to identify host susceptibility loci and genes, as well as gene networks and higher level interactions that regulate phenotypic variation and disease severity. Among mouse GRPs, the Collaborative Cross (CC) has over 44 million single nucleotide polymorphisms (SNPs) and 4 million insertions and deletions (InDels), which segregate between the eight founder strains (20). In addition, several thousand novel variants (SNPs and small indels, as well as large deletions) exist within individual CC strains (21–24). As such, the CC population models human genetic diversity with both common and rare variants poised to impact a variety of phenotypes, including disease severity after acute viral infections, like influenza virus, flaviviruses, filoviruses, and coronaviruses (20, 25–31).

In this study, we leveraged the CC model to characterize the genetic susceptibility landscape of sarbecovirus infections in mice. We demonstrated that multiple loci regulate the host disease responses to this subgroup of coronaviruses. Specifically, we identified two genes Ccr9 and Cxcr6 within a multitrait QTL located on chromosome 9 (Chr9), which have natural polymorphisms driving altered expression levels that correlate with SARS-CoV and SARS-CoV-2 disease severity in mice. This QTL shows conserved synteny with a locus in humans on Chr3 (which includes Ccr9 and Cxcr6) that was identified in COVID-19 human Genome-wide Association Studies (GWAS) predicting severe outcome and hospitalization. In total, several GWAS in humans have identified a common locus on Chr3 (3p21.31), that identified six genes associated with COVID-19 severity, susceptibility, respiratory failure, and risk of hospitalization and included CCR9 and CXCR6 (14, 15, 32–34). In addition, across these human studies there...
was a high amount of genetic variation as well as different lead SNPs in this region, suggesting that multiple genes may regulate disease associations and severity. Our concordant susceptibility profiles of the CC011 and CC074 parent mouse lines, as well as Ccr9- and Cxcr6-deficient mice infected with SARS-CoV MA15, SARS-CoV-2 MA10 or BtCoV HKU3-SRBD MA, demonstrate involvement of this locus in severe COVID-19 disease susceptibility across species and across different sarbecoviruses, as well as highlighting the utility of preemergence disease models. Our results also demonstrate that the CC mouse panel is well suited to identify and validate relevant susceptibility regions for other human infectious and chronic diseases, while providing focus for a deeper understanding of emerging sarbecovirus disease patterns in animal and human populations.

RESULTS

Variable disease phenotypes after CC-RI strain infection. To extend our previous understanding of how genetic variation contributes to differential SARS-CoV outcomes, we studied groups of female mice from five different CC strains (CC005/TauUnc, CC011/Unc, CC020/TauUnc, CC068/TauUnc, and CC074/Unc) infected with 2003 SARS-CoV using our SARS-CoV MA15 model, which recapitulates many of the disease outcomes reported in humans (35). As these strains did not segregate alleles at previously identified SARS QTL, differences in disease would be due to uncharacterized genetic differences (27, 28). Infected mice were followed for weight loss and possible mortality through 4 days postinfection (Fig. 1A). We observed a wide variation of weight loss across these strains. Due to divergent disease susceptibilities to SARS-CoV MA15 and narrow variation within a given strain, we focused on the disease-resistant CC011/Unc (hereafter CC011, <5% weight loss) and the highly susceptible CC074/Unc strain (hereafter CC074, >15% weight loss). To assess disease progression in these strains in greater detail, we challenged additional groups of CC011 and CC74 mice (of both sexes) with SARS-CoV MA15. CC074 mice showed severe disease by 4 days postinfection (dpi) marked by ~80% mortality of ~80% in 2 separate experiments, and a >15% loss of body weight. In contrast, CC011 had a maximal weight loss of 5% body weight and 100% survival (Fig. 1B and C). Despite differences in weight loss, mortality, and congestion score, both strains had high levels of virus detectable in lungs at 2 and 4 dpi (Fig. 1D and E, respectively; mean ~5 × 10^6 PFU/lobe, CC011 n = 6; CC074 n = 1, 5 mice died). While confirming the original observation that these two strains exhibit distinct disease susceptibilities, these results also suggested that differences in host immune responses, rather than viral burden, drive the different disease outcomes in these strains.

SARS-CoV MA15 CC-F2 screen. To identify the genetic basis of the observed disease outcomes and SARS-CoV pathogenesis, we generated a large F2 intercross between CC011 and CC074 (n = 403 F2 mice, both males and females). F2 mice were inoculated intranasally at 9 to 12 weeks of age with 1 × 10^4 PFU of SARS-CoV MA15. To expand our understanding of response to SARS-CoV infection, in addition to standard SARS-CoV-associated phenotypes including weight loss, viral burden, mortality, and lung congestion (Fig. 2A to D), we also examined circulating immune cells as well as lung function (Fig. 2E to G). Across these phenotypes, we observed an expanded range of disease responses in the F2 cross relative to their parent CC strains, (Fig. 2 and Fig. S1). While a significant fraction of the F2 population (approximately 25%) died or were euthanized, this was diminished compared to CC074 (~80% mortality).

Concurrent with viral challenge, the F2 animals were genotyped with the MiniMUGA array and QTL mapping in R/QTL conducted (36, 37). We identified a total of 2,750 markers segregating between CC011 and CC074 among the autosomes and X-chromosome. These markers were well distributed across the genome (23 to 48,228,264 nucleotides between markers, median = 509,769 nucleotides between markers). Using this genetic map, we identified a total of ten genetic loci segregating in this F2 population for a variety of traits (host response to SARS HrS24-28 at genomewide, P < 0.05; HrS29-33, P < 0.1) (Fig. S1 and Table 1). Significant loci included a
Multitrait locus (Hrs26, Chr 9), as well as single loci associated with overall mortality (Hrs24, Chr 4), weight loss in males at 4 dpi (HsR25, Chr 4), peripheral lymphocyte and neutrophil levels at 4 dpi (Hrs27, Chr 11), and finally, altered pulmonary function (PenH) at 2 dpi (Hrs28, Chr 15). The suggestive loci were identified in contributing to altered weight loss, mortality, and altered lung function phenotypes (Table 1).

The multitrait Hrs26 had a major impact on clinical disease (odds ratio [OR] of 3.15 for mortality), as well as lung pathology 10.24% of overall population variation) including lung function (7.76% of respiratory resistance (PenH) and immune cell inflammation.

**FIG 1** CC strains demonstrate different susceptibility to SARS-CoV infection and disease. Age-matched female mice (n = 4) of 5 different CC strains (CC005, CC011, CC020, CC068, and CC074) were infected with 1 × 10^6 PFU and monitored for weight loss until 4 dpi. An additional group of age-matched CC011 and CC074 (both sexes) were infected with 1 × 10^6 PFU, and mice were monitored for disease progression until 4 dpi. (A) Weight loss for CC strains CC005, CC011, CC020, CC068, and CC074. (B) Weight loss of the two parental CC strains CC011 and CC074. (C) Percentage survival of the two parental CC strains CC011 and CC074. (D) Lung viral titer of the two parental CC strains CC011 and CC074 on 2 and 4 dpi. (E) Lung congestion score of the two parental CC strains CC011 and CC074 (female 10- to 12-week-old mice were infected with 1 × 10^6 PFU SARS-CoV MA15; CC011: n = 4 for 2 dpi and n = 6 for 4 dpi; CC074: n = 4 for 2 dpi and n = 6 for 4 dpi, respectively). Data were analyzed using Mann-Whitney test (lung titer and congestion scores); *, P < 0.05.
As such, it represents a locus of major effect driving disease differences between these strains following SARS-CoV MA15 challenge.

**CC011 and CC074 disease outcomes following SARS-CoV-2 MA10 infection.**
Concurrent with our ongoing analyses to phenotype SARS-CoV MA15 infection outcomes in CC parental and F2 strains, SARS-CoV-2 suddenly emerged. Although SARS-CoV-2 was genetically distinct with about 22% nucleotide differences across the genome compared to SARS-CoV (6), the two viruses provided an opportunity to test whether similar host genes regulated pathogenesis across distinct sarbecoviruses. Accordingly, we infected CC011 and CC074 with the SARS-CoV-2 MA10 strain, which shares many of the COVID-19 disease phenotypes seen in human patients (Fig. 3) (18). Importantly, the parental strains showed similar clinical disease phenotypes after SARS-CoV-2 MA10 infection (Fig. 3A to D) and SARS-CoV MA15 (Fig. 1B to E). Despite a small difference in weight loss and lung viral load between both virus strains, CC074 showed the same severe disease phenotypes with severe weight loss, lung discoloration scores, and 100% mortality by 4 dpi after infection with SARS-CoV-2 MA10 (Fig. 3). These results suggested that the QTL identified for SARS-CoV susceptibility might also be relevant for SARS-CoV-2-induced disease.
Identification of a multitrait QTL on Chr9. Recently, several reports have identified a common locus in humans on Chr3 (3p21.31) that is associated with COVID-19 severity, susceptibility, and risk of hospitalization (14, 15, 32–34). This locus, encompassing genes such as SLC6A20, LZTFL1, Fyco1, XCR1, CXCR6, and CCR9, shows conserved synteny with the proximal region of the multitrait locus mapped on Chr9 in mice (HsR26 and HsR30) (Fig. 4A and Table 1). In addition, the studies in humans confirmed a high degree of polymorphism in this area on Chr9. We examined the sequence differences between CC011 and CC074 in this region of conserved synteny, where CC011 possessed a PWK/PhJ haplotype and CC074 had an A/J haplotype, and detected several regions and all other SNPs were in regulatory regions such as the 5′-flanking region, at splice sites, and within introns, pointing toward differential expression of these target genes (Table 2). Similar findings and altered gene expression levels have been reported in the human GWAS, transcriptome-wide association study, and epigenetic studies involving these gene sets (15, 16, 39–41).

Consequently, gene expression profiles of Ccr9, Cxcr6, Xcr1, Lztf1, Fyco1, and Slc6a20a/b were measured in the lungs of infected CC011 and CC074. Significantly reduced expression of Ccr9, Cxcr6, and Lztf1 was observed in highly susceptible CC074 mice (P < 0.005), compared to the resistant CC011 mice. In contrast, expression of Fyco1 was significantly upregulated (P < 0.05) in CC074 compared to the CC011 mice, while levels of Xcr1 and Slc6a20a/b expression were not significantly different (Fig. 4B). We took advantage of RNAseq data from uninfected CC mice that contained either the protective (PWK) or susceptible (A/J) haplotype at this locus to determine the expression of these six genes (Table S1). We found significant differences in expression of Lztf1 (increased in A/J), Fyco1
(decreased in A/J), and Slc6a20a/b (increased in A/J), as well as a trend for Xcr1 (increased in A/J). Together, these data suggest a complex cis-regulatory architecture of these genes at baseline along with induction in the context of infection.

**Ccr9 regulates sarbecovirus infection and pathogenesis in vivo.** The sarbecoviruses are divided into three phylogenetic clades designated clade Ia (SARS-CoV), clade II (HKU3) and clade Ib (SARS-CoV-2) (Fig. 5A). To initially examine the impact of differences in gene expression have on virus disease phenotypes, we focused on Ccr9 and Cxcr6, which were differentially expressed by the respective mouse strains (Fig. 4B).

Knockout mice have the engineered construct introduced in a cell line (typically derived from a 129 mouse) before being bred to a relevant control strain. We genotyped our CCR9 and CXCR6ko mice with the genome-wide MiniMUGA genotyping array. We confirmed that both strains were largely (all ~99.5%) of the relevant genetic background (C57BL/6NJ for CCR9 and C57BL/6J for CXCR6), with a small cluster of SNPs differing at the distal chromosome 9 locus where the knocked-out genes are located. These SNPs were consistent with a 129 genetic background, thus confirming how these knockout strains were created.

To ensure that the presence of a 129 haplotype at the locus itself was not causal for disease differences, we identified three CC strains which contained a 129 haplotype at this locus (CC039, CC041, and CC065) and assessed the response to SARS-CoV2 in these three strains relative to CC011 and CC074 mice (Fig. S2). While these three strains...
showed a range of disease responses from mild weight loss to mortality, the fact that there was not a consistent disease response in these strains suggest that the 129 haplotype at HrS26 is not informative for SARS-CoV2 disease. As such, when contrasting knockouts to relevant controls, we can be confident that the results are due to the targeted mutation itself and not an artifact of the haplotype the mutation was generated in. Therefore, we infected Ccr9 deficient mice (CCR9^2/2) of both sexes as well as the relevant age-matched C57BL/6NJ controls with SARS-CoV MA15. CCR9^2/2 mice showed enhanced susceptibility relative to control mice. The latter was marked by significant weight loss of 20% (HrS30), high viral load in the lungs and severe lung pathology, prolonged respiratory dysfunction as measured by whole body plethysmography (HrS26), increased lung congestion scores (HrS26), and elevated mortality (HrS30) (Fig. 5B to F). In addition to SARS-CoV, we also studied the role of CCR9 in clinical disease outcomes using a mouse adapted HKU3-SRBD MA (Fig. 5G to I) (10). We infected CCR9^2/2 mice and C57BL/6NJ control mice with 1 x 10^4 PFU SARS-CoV-2 MA10; CC011: n = 5, CC074: n = 5). Data were analyzed using Mann-Whitney test; *, P < 0.05; **, P < 0.005.

### FIG 4
Identification of a Multitrait QTL on Chr9. Each of the individual F2 mice were genotyped using MiniMuga and QTL mapping conducted by testing the strength of association between each F2 mouse’s phenotype and their genotypes at each marker. (A) A quantitative multitrait locus with major effect was identified on Chr9 (74.9–124 Mb), which affected mortality, weight loss, lung congestion, lung function, and peripheral hematology. CC011/Unc has a C57BL/6/PWK haplotype and CC074/Unc has an A/J/PWK haplotype in this QTL region. (B) Expression levels of Ccr9, Cxcr6, Xcr1, Lztf1, Fyco1, and Slc6a20a/b of SARS-CoV-2 MA10-infected CC011 and CC074 mice at 4 dpi, determined by quantitative RT-PCR, normalized to Gapdh (female 10- to 12-week-old mice were infected with 1 x 10^4 PFU SARS-CoV-2 MA10; CC011: n = 5, CC074: n = 5). Data were analyzed using Mann-Whitney test; *, P < 0.05; **, P < 0.005.

### TABLE 2
SNPs in CC011 and CC074 possible target genes in the Chr9 multitrait QTL^a (38)

<table>
<thead>
<tr>
<th>Variants</th>
<th>SCL6A20A/B</th>
<th>LZTF1L</th>
<th>CCR9</th>
<th>FYCO1</th>
<th>CCR6</th>
<th>XCR1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missense</td>
<td>8</td>
<td>3</td>
<td>20</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5’ UTR</td>
<td>1</td>
<td>12</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>3’ UTR</td>
<td>7</td>
<td>33</td>
<td>57</td>
<td>729</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Intron</td>
<td>139</td>
<td>699</td>
<td>424</td>
<td>34</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>Splicing</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

^aCC011, PWK; CC074, A/J; UTR, untranslated region.
The hypothesis that null or low-level expression of CCR9 likely contributes to the multitrait QTL disease phenotypes noted following clade Ia and clade II sarbecovirus infections.

Clade Ib SARS-CoV-2 MA10 infection in CCR9<sup>2/2</sup> mice also developed more severe clinical disease including greater weight loss (<i>P</i> < 0.0001 to <i>P</i> < 0.05) (Fig. 6A) and
Ccr9 Regulates SARS-CoV-2 infection and pathogenesis in vivo. To study the effect of Ccr9 on the susceptibility to SARS-CoV-2 infection, groups of age-matched CCR9^−/− mice were infected with 1 x 10^5 PFU SARS-CoV-2 MA10 and followed for several days for disease outcomes. (A) Weight loss of CCR9^−/− mice and C57BL/6NJ control mice. (B) Lung viral titer of CCR9^−/− mice and C57BL/6NJ control mice on 2, 4, 6, and 14 dpi. (C) PenH of CCR9^−/− mice and C57BL/6NJ control mice. (D) Cytokine/chemokine distribution in the lung of CCR9^−/− mice and C57BL/6NJ control mice on 2, 4, and 6 dpi. (E) Composition of lung infiltrating immune cells in the lung of CCR9^−/− mice and C57BL/6NJ control mice on 4 and 6 dpi (15- to 18-week-old mice were infected with 1 x 10^5 PFU SARS-CoV-2 MA10; C57BL/6NJ: n = 29, n = 4 for 2 dpi, n = 7 for 4 dpi, n = 15 for 6 dpi, and n = 3 for 14 dpi; CCR9^−/−: n = 24 total, n = 4 for 2 dpi, n = 7 for 4 dpi, n = 10 for 6 dpi, and n = 3 for 14 dpi, respectively, for weight loss and viral titer, C57BL/6NJ: n = 4 and CCR9^−/−: n = 4 lung function analysis, 57BL/6NJ: n = 4 for 2 dpi, n = 4 for 4 dpi, n = 15 for 6 dpi; CCR9^−/−: n = 4 for 2 dpi, n = 4 for 4 dpi, n = 11 for 6 dpi for (Continued on next page)
virus titers ($P < 0.0005$) (Fig. 6B) than C57BL/6NJ mice. Furthermore, infected CCR9$^{−/−}$ mice displayed increased PenH respiratory dysfunction ($P < 0.0005$ to $P < 0.0005$) (Fig. 6C), lung congestion ($P < 0.05$) (Fig. S3B), and mortality (Fig. S3A). Finally, CCR9$^{−/−}$ mice also demonstrated severe lung pathology including diffuse alveolar damage and prolonged loss of surfactant protein C (SFTPC) expression in AT2 cells, comparable to the wild-type controls (Fig. S3C to G). Analysis of the cytokine profile in lungs by multiplex immune assay showed increased subsets of cytokines and chemokines involved in promoting allergic airway inflammation, including IL-6, CCL3, G-CSF, CCL2, IL-13, CXCL1, and CCL11 (Eotaxin) (Fig. 6D). Flow cytometric analysis showed that in CCR9$^{−/−}$ mice there was a significant increase in CD4$^{+}$ T cells, CD8$^{+}$ effector T cells, CD11b$^{+}$ dendritic cells (DCs), and eosinophils at 6 dpi, consistent with an airway inflammatory response (Fig. 6D to E). In summary, these data support an important role of CCR9 in protection against severe sarbecovirus disease.

**Cxcr6 regulates SARS-CoV MA15 and SARS-CoV-2 MA10 pathogenesis in mice.**

As the Chr9 locus (HsR26 and HsR30) also contained variants segregating in Cxcr6 that might contribute to reduced gene expression and severe disease, we also infected Cxcr6-deficient (Cxcr6$^{−/−}$) and aged-matched controls of both sexes with SARS-CoV-MA15 and followed the mice for 7 days. We detected no significant differences in weight loss (Fig. S4B) but significant higher viral load, congestion scores, prolonged pulmonary dysfunction (as measured by whole body plethysmography), and increased mortality (Fig. S4C to F). The mortality rate in Cxcr6$^{−/−}$ was significantly increased with 100% mortality by 7 dpi compared to 70% in the control mice and a median survival rate of 4 dpi compared to 6 dpi in the control mice (Fig. S4F). We then evaluated whether CXCR6$^{−/−}$ also serves as a susceptibility gene for SARS-CoV-2 MA10 infection. CXCR6$^{−/−}$ mice showed more severe disease progression associated with significant weight loss (>20%) ($P < 0.0001$ to $P < 0.0005$), prolonged airflow restriction, and increased mortality (50%, not shown) compared to wild-type C57BL/6 controls (Fig. 7A and B) Cxcr6$^{−/−}$ lungs also had significantly higher viral loads and elevated congestion scores ($P < 0.005$ for 4 dpi) (Fig. 7C and D), compared to wild-type controls. Lung cytokine and chemokine analysis showed levels of IL-6, CCL2, G-CSF, and CXCL1 that were significantly elevated in CXCR6$^{−/−}$ compared to wild-type infected mice (Fig. 7E). The composition of infiltrating cells into the lungs (Fig. 7F and G) of Cxcr6$^{−/−}$ mice greater numbers of CD4$^{+}$ T cells (Fig. 7F), alveolar macrophages, inflammation-promoting CD11b$^{+}$ DCs, and monocytes at 4 dpi (Fig. 7G), consistent with an increased inflammatory airway response.

**DISCUSSION**

Emerging viruses represent an ongoing threat to human health and economic stability. As many emergence events are characterized by small outbreaks that are not amenable to GWAS, new strategies are needed to predict and understand how host genetic variation might regulate disease severity. Infectious diseases have shaped the natural genetic variation in the genomes of mammals. Although highly penetrant monogenic traits have been described that regulate disease severity in humans, most viral diseases likely are regulated by complex genetic traits (42–45). Novel viruses regularly emerge from zoonotic reservoirs to replicate and colonize one or more new mammalian hosts. Consequently, evolution may select for common susceptibility loci and genes that regulate disease severity across species, although this hypothesis remains unproven (42, 46–49). The successive emergence of genetically distinct sarbecoviruses in 2002 and 2019, which replicate in and cause severe disease in multiple mammalian hosts, provide an opportunity to determine the extent to which common or unique susceptibility loci regulate disease severity across species. During the first SARS-CoV epidemic, no consensus was reached on the human loci and genes that influence disease severity (50).

**FIG 6** Legend (Continued)

chemokine/cytokine analysis; $n = 12$ CCR9$^{−/−}$ and $n = 13$ C57BL/6NJ with $n = 2$ each for mock, $n = 6–7$ for 4 dpi, and $n = 5–6$ for 6 dpi for analysis of infiltrating cells). Data were analyzed using 2-way ANOVA with multiple comparison (weight and PenH) and Mann-Whitney test (viral titer, congestion score, cytokine/chemokine, and infiltrating cells); *, $P < 0.05$; **, $P < 0.005$; ***, $P < 0.0005$; ****, $P < 0.0001$. 

July/August 2022 Volume 13 Issue 4 10.1128/mbio.01454-22
Recent GWAS in humans identified a locus 3p21.31, associated with susceptibility to respiratory failure, severe disease outcome, and hospitalizations after SARS-CoV-2 infection. This locus encompasses six genes, all of which showed strong signals in multiple GWAS studies: SLC6A20, LZTFL1, CCR9, FYCO1, CXCR6, and XCR1, with the causative signals between studies spread across these genes (14, 15, 32–34, 51). These genes are differentially regulated in mild and severe human disease, suggesting the presence of genetic variants that regulate gene expression (15, 16, 39–41). To date, none of these genes have been definitively linked to disease severity after SARS-CoV-2 infection (15, 16, 39–41). This human locus shows conserved synteny with the proximal region of our Chr9 multitrait QTL, HsR26 in mice, and shows similar genetic order and orientation of all six candidate genes. Furthermore, in addition to nonsynonymous differences in LZTFL1, CCR9, FYCO1, and SLC6A20A/B, regulatory differences were seen for genes in this locus at baseline as well as in the context of infection. As such, it is likely that complex interactions between multiple candidates in this locus give rise to the observed disease differences at HsR26 and present the intriguing possibility that some of these regulatory differences underlie common disease mechanisms between mice and humans.

Within the locus, we focused on two genes, Ccr9 and Cxcr6, as both are expressed...
at reduced levels in our parental susceptible parental strain and in vulnerable human cases (16, 39, 40). Gene deletions represent extreme reductions in expression and loss of function, which differ from the subtle impact that allele variation in noncoding or coding regions might have. However, these studies represent an important first step toward elucidating the genetic drivers of sarbecovirus disease severity and highlight a rationale for future study of CCR9 and CXCR6 allele variants. CCR9 is expressed by T cells, DCs, monocytes/macrophages, and eosinophils and is a receptor for CCL25 (52). The CCR9/CCL25 axis functions in homeostasis, inflammation, and inflammation-associated diseases including early respiratory allergic inflammation, asthma, chronic inflammatory bowel diseases, and organ fibrosis (53). CXCR6 is also expressed by T cells and serves as a homing mediator for resident memory T cells (T_{RM}). Previous studies have shown that CXCR6^{−/−} mice have decreased lung resident T_{RM} cells, which is associated with increased host control of Mycobacterium tuberculosis and influenza virus infections (54). Both CCR9 and CXCR6 are involved in recruiting T cells and regulating inflammatory responses during lung infection. Importantly, SARS-CoV-2 infection stimulated distinct yet overlapping infiltrating cell populations and cytokine and chemokine signatures in CCR9^{−/−} and CXCR6^{−/−} mice. SARS-CoV-2-infected CCR9^{−/−} mice showed severe disease progression, with an increased influx of inflammatory macrophage/monocyte cells, T cells, and eosinophils, coupled with a chemokine/cytokine profile that supports an allergic inflammatory response in the lung. SARS-CoV-2-infected CXCR6^{−/−} mice showed an influx of inflammatory cells with expression of proinflammatory Th1-cytokines/chemokines in the lungs. As both genes are downregulated in highly susceptible CC074 mice and in humans with severe respiratory failure, it is possible that both genes, as well as other genes under the QTL, contribute to vulnerability to SARS-CoV-2 infection in mice. Detailed genetic mapping studies in human cohorts have implicated a 5′ synonymous mutation in the noncoding region of the LZTFL1 gene, which contributes to the downregulation in expression, although it still remains uncertain if this gene directly modulates severe human or mouse disease outcomes (41).

Here, we demonstrate that mice lacking Ccr9 or Cxcr6 expression developed severe disease most likely driven by excessive inflammation. The increased numbers of proinflammatory monocytes in both the Ccr9- and Cxcr6-deficient mice suggest a potential mechanism for increased disease severity, as previous studies have implicated the levels of proinflammatory monocytes cells in lethal SARS-CoV infection (55). Lack of CCR9 enhances disease severity, this gene likely has an important protective function during SARS-CoV, HKU3-SRBD, and SARS-CoV-2 infection in mice. Our data also indicates that lack of CXCR6 gene enhances disease severity, implicating a role in protecting against lethal infection as well. In future studies, we plan to define the exact SNP(s) that regulate the differential expression levels and induction of Ccr9 and Cxcr6 and identify the causal mechanisms underlying the observed disease outcomes associated with the human GWAS results, as well as our HrS26 results.

Rich and complex data sets such as described herein enable comparisons with human GWAS studies mapping QTL after SARS-CoV-2 infection. Our data support the idea that there are shared disease susceptibility loci in both human and mice. We provide evidence from our mouse studies that the HrS26 locus is involved in disease outcomes across multiple genetically distinct sarbecoviruses. Our study highlights the power of using animal GRPs to understand the role of host genetic variation on infectious diseases, generate new models of differential disease, probe the role of individual genes in disease progression, and provide mechanistic insight into the role of specific host genes and viral strains in regulating pathogenesis across species. In appropriately selected large population screens such as the entire CC resource, highly penetrant genetic variants that impact disease outcome can be readily identified. In contrast, targeted mapping crosses between highly discordant CC RI strains can identify more complex genetic networks such as variants that are penetrant only in the context of specific genetic backgrounds, or epistatic (gene-gene) interaction networks. In addition, the multiple loci we identified, as well as unexplained heritability in this cross,
suggest that CoV disease and immunity are complex polygenic traits, with the accumulation of variants across many loci driving disease susceptibility. Our studies represent a comprehensive comparison of susceptibility loci for sarbecoviruses in two different mammalian hosts, identify a large collection of susceptibility loci and candidate genes and alleles that regulate multiple aspects type-specific and cross-CoV pathogenesis, validate a potential role for the CCR9 and/or CXCR6 in regulating SARS-CoV-2 disease severity, and provide a resource for community-wide studies. The evidence supporting a common-susceptibility loci harbored in CC074 mice and in highly vulnerable human populations provides new opportunities for studying sarbecovirus-host interactions, inflammation, and immunity in genetically relevant mammalian models, potentially revealing new insights into acute and chronic disease progression and providing a potentially more relevant model for evaluating immunotherapeutic control of life-threatening sarbecovirus infection outcomes.

MATERIALS AND METHODS

Cells and viruses. Recombinant mouse-adapted SARS-CoV MA15, HKU3-SRBD-MA (HKU3-SRBD MA), and SARS-CoV-2 MA10 virus were generated as described previously (10, 18, 35). For virus titration, the caudal lobe of the right lung was homogenized in PBS, resulting homogenate was serial-diluted and inoculated onto confluent monolayers of Vero E6 cells (ATCC CCL-81), followed by agarose overlay. Plaques were visualized with overlay of Neutral Red dye on day 2 (SARS-CoV MA15, HKU3-SRBD MA) or day 3 (SARS-CoV-2 MA10) postinfection. All virological studies were conducted under BSL3 conditions, and personnel wore appropriate personal protective gear.

Mouse studies and in vivo infections. Mouse studies were performed at the University of North Carolina (UNC) (animal welfare assurance no. A3410-01) using protocols approved by the UNC Institutional Animal Care and Use Committee (IACUC). Animal studies at Washington University were carried out in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocols were approved by the IACUC at the Washington University School of Medicine (assurance no. A3381-01).

Mouse studies were divided into three major classes: F2 intercross mice, CC mice, and inbred wild-type or gene-edited mice. CC mice were purchased directly from the UNC Systems Genetics Core Facility at 3 to 5 weeks of age and acclimated for 1 week in the BSL3 before challenge. We contracted with the Systems Genetics Core Facility at UNC to generate the F2 mice used in this study. First, F1 mice between CC011 and CC074 were generated by cross males and females in both directions, and then the F2 mice were bred in all 4 possible F1 x F1 combinations, to ensure appropriately balanced sex Chromosome and parent-of-origin effects. F2 mice (226 males, 177 females) were weaned such that littermates were randomized to different experimental cages to reduce litter or batch effects on the study, and mice were transferred at 5 to 6 weeks of age to the laboratory for infection between 9 and 12 weeks of age. For studies in genetically relevant wild-type or gene-edited mice. CC mice were purchased directly from the UNC Systems Genetics Core Facility at 3 to 5 weeks of age and acclimated for 1 week in the BSL3 before challenge. We contracted with the Systems Genetics Core Facility at UNC to generate the F2 mice used in this study. First, F1 mice between CC011 and CC074 were generated by cross males and females in both directions, and then the F2 mice were bred in all 4 possible F1 x F1 combinations, to ensure appropriately balanced sex Chromosome and parent-of-origin effects. F2 mice (226 males, 177 females) were weaned such that littermates were randomized to different experimental cages to reduce litter or batch effects on the study, and mice were transferred at 5 to 6 weeks of age to the laboratory for infection between 9 and 12 weeks of age. For studies in genetically defined knockout mice, 15-18 - week-old CCR9−/− mice (strain 027041), CXCR6−/− mice (strain 005693) C57BL/6NJ mice (strain 000664), and C57BL/6J (strain 000664) were purchased from Jackson Laboratory, and the genotype of the mutant mice were confirmed via genotyping on the MiniMUGA array (Neogen, Inc.). CC mice and CC-F2 mice were inoculated with 1 × 104 PFU (CC-F2 with SARS-CoV MA15 and CC mice with either SARS-CoV MA15 or SARS-CoV-2 MA10, respectively). For infection, CXCR6−/−, CCR9−/−, and appropriate C57BL/6 control mice were anesthetized with a mixture of ketamine/xylazine and inoculated intranasally (in both nostrils) with 1 × 104 PFU of either SARS-CoV MA15, SARS-CoV-2 MA10, or HKU3-SRBD MA in 50 μL of PBS. Body weight, mortality, and pulmonary function by whole body plethysmography (56) were monitored daily as indicated. At the designated time points, mice were euthanized and gross pathology (congestion score) of the lung was assessed and scored on a scale from 0 (no lung congestion) to 4 (severe congestion affecting all lung lobes). Then, lung tissue was harvested for histopathology analysis, and blood samples were harvested for analysis of immune cells. Samples were stored at −80°C until homogenized and titrated by plaque assay as described above. Peripheral blood was diluted 1:5 in PBS/EDTA and analyzed with the VetScan HSM as previously described (57). Histopathology samples were fixed in 10% phosphate-buffered formalin for 7 days before paraffin embedding, section stained with hematoxylin and eosin, and scored for DAD and acute lung injury as previously described by our group (58).

Flow cytometry analysis of immune cell infiltrates. For analysis of lung tissues, mice were perfused with sterile PBS, and the right inferior lung lobes were digested at 37°C with 630 μg/mL collagenase D (Roche) and 75 U/mL of DNase I (Sigma-Aldrich) for 2 h. Single cell suspensions were preincubated with Fc Block antibody (BD Pharmingen) in PBS + 2% heat-inactivated FBS for 10 min at room temperature before staining. Cells were incubated with antibodies against the following markers: efluor506 Viability Dye (65-0866-14; Thermo Fisher), BUV395 anti-CD45 (Clone 30-F11; BD Biosciences), BV711 anti-CD11b (Clone M1/ 70; Biolegend), APC-Cy7 anti-CD11c (Clone HL3; BD Biosciences), BV650 anti-Ly6G (Clone 1A8; Biolegend), Pacific Blue anti-Ly6C (Clone HK3A; Biolegend) FITC anti-CD24 (Clone M1/69; Biolegend), PE anti-Siglec F (Clone E50-2440; Biolegend), PE-Cy7 anti-CD64 (Clone XS54-5/7.1; Biolegend), AF700 anti-MHCII (Clone M5/ 114.15.2; Biolegend), BV421 anti-CD3 (Clone 17A2; Biolegend), BV785 anti-CD4 (Clone GK1.5; Biolegend), APC anti-CD8a (Clone 53-6.7; Biolegend), BV421 anti-B220 (Clone RA3-682; Biolegend), APC-Cy7 anti-CD44 (Clone IM7; Biolegend), and BV605 anti-CD62L (Clone MEL-14; Biolegend). All antibodies were used at a dilution of 1:2000. Cells were stained for 20 min at 4°C, washed, fixed, and permeabilized for intracellular staining with
Fox3/Transcription Factor Staining Buffer Set (eBioscience) according to manufacturer's instructions. Cells were incubated overnight at 4°C with BV421 anti-Fox3 (Clone MF-14; BioLegend) washed, fixed with 4% PFA (EMS) for 20 min, and resuspended in permeabilization buffer. Absolute cell counts were determined using Trucount beads (BD). Data were acquired on a flow cytometer (BD-X20; BD Biosciences) and analyzed using FlowJo software (Tree Star) (Fig. S5).

**Cytokine and chemokine protein analysis.** The small center lung lobe of each mouse was homogenized in 1 mL of PBS and briefly centrifuged to remove debris. Fifty microliters of homogenate was used to measure cytokine and chemokine protein abundance using a Bio-Plex Pro mouse cytokine 23-plex assay (Bio-Rad) according to the manufacturer's instructions.

**Lung pathology scoring and RNA in situ hybridization/quantification.** Two separate lung pathology scoring scales, Matute-Bello and Diffuse Alveolar Damage (DAD), were used to quantify acute lung injury (58). For Matute-Bello scoring samples were blinded and three random fields of lung tissue were chosen and scored for the following: A: neutrophils in alveolar space (none = 0, 1 to 5 cells = 1, >5 cells = 2); B: neutrophils in interstitial space (none = 0, 1 to 5 cells = 1, >5 cells = 2); C: hyaline membranes (none = 0, one membrane = 1, >1 membrane = 2), (D) Proteinaceous debris in air spaces (none = 0, one instance = 1, >1 instance = 2); and E: alveolar septal thickening (<2 Å̅ mock thickness = 0, 2 to 4 Å̅ mock thickness = 1, >4 Å̅ mock thickness = 2). Scores from A to E were put into the following formula score = [(20 × A) + (14 × B) + (7 × C) + (7 × D) + (2 × E)]/100 to obtain a lung injury score per field and then averaged for the final score for that sample.

In a similar manner, DAD scoring was obtained by evaluating three random fields of lung tissue that were scored for the in a blinded manner for the following: 1 = absence of cellular sloughing and necrosis, 2 = uncommon solitary cell sloughing and necrosis (1 to 2 foci/field), 3 = multifocal (3 + foci) cellular sloughing and necrosis with uncommon septal wall hyalinization, or 4 = multifocal (>75% of field); cellular sloughing and alveolar collapse with common and/or prominent hyaline membranes. To obtain the final DAD score per mouse, the scores for the three fields per mouse were averaged.

RNA-ISH was performed on paraffin-embedded 5-μm tissue sections using the RNAscope Multiplex Fluorescent Assay v2 according to the manufacturer’s instructions (Advanced Cell Diagnostics). Briefly, tissue sections were deparaffinized with xylene and 100% ethanol twice for 5 min and 1 min, respectively, incubated with hydrogen peroxide for 10 min and in boiling water for 15 min, and then incubated with Protease Plus (Advanced Cell Diagnostics) for 15 min at 40°C. Slides were hybridized with custom probe (RNAscope Probe-Mm-Sftpc-C2; cat no. 314101-C2) at 40°C for 2 h, and signals were amplified according to the manufacturer’s instructions. Stained sections were scanned and digitized by using an Olympus VS200 fluorescence microscope with a 40 × 1.35 NA objective. Lung tissue and Sftpc signal were quantified using a customized analysis protocol package to (i) detect lung tissue using a decision forest classifier, and (ii) probe and analyze signal based on the intensity of the signal in the channel corresponding to the relevant probe (Sftpc). All slides were analyzed using the same conditions. Results were expressed as the area of the probe relative to total lung tissue area.

**Mouse DNA genotyping.** CC011, CC074, their F1 progeny, the F2 cross, CXCXR6−/−, CCR9−/− mice, and appropriate controls were genotyped on the MiniMUGA genotyping array (28, 37). Genomic DNA was isolated from tail-clips of animals using the Qiagen (Hilden, Germany) DNeasy blood and tissue kit. Then, 1.5 μg was sent to Neogen (Lincoln, NE) for processing. We filtered the genotypes upon return for informativeness within this cross. To be considered as valid, the marker had to have one homozygous allele in all CC011 mice genotyped, the alternate homozygous allele in all CC074 mice genotyped, and the relevant homoygous allele in male F1s. This filtering reduced the ~10,800 SNPs on the MiniMUGA array to 2,821 informative markers.

**QTL mapping and statistical analyses.** We used R/QTL for genetic mapping (59, 60). Briefly, after appropriate data transformations, we used the scanone function to determine the strength of the regression of phenotypes on genotypes at each of the informative markers in the cross. Significance thresholds were determined by running 1,000 permutations, scrambling the relationship between phenotypes and genotypes, providing an appropriate threshold of significance that is robust to the phenotype distribution and allele frequencies. For phenotypes where we identified multiple QTL, we ensured that long-range linkage disequilibrium was not driving these observations by fitting multifactor ANOVAs with single QTL and with sets of loci. Only loci for which there was a statistically significant improvement in fit for the full model were kept.

**Data availability.** All data supporting the finding of this study are available within the article and are available from the corresponding authors upon request.

**Statistical analysis.** The number of independent experiments and technical replicates are indicated in the relevant figure legends. Two-way ANOVA with multiple comparison were used for weight loss and lung function comparisons, log-rank was used for survival studies, and Mann-Whitney test was used for viral titer, congestion score, okine/chemokine, infiltrating cells, histopathology score, and gene expression comparisons.

**Code use.** We used R/QTL for genetic mapping and the scanone function to determine the strength of the regression of phenotypes on genotypes at each of the informative markers in the cross.

**SUPPLEMENTAL MATERIAL**

Supplemental material is available online only.

**FIG S1, TIF file, 2.3 MB.**
REFERENCES


Dr. Shelton JF, Dr. Shastri AJ, Dr. Ye C, Dr. Weldon CH, Dr. Filshtein-Sonmez T, Dr. Coker D, Dr. Graham JB, Dr. Swarts JL, Dr. Menachery VD, Dr. Gralinski LE, Dr. Schäfer A, Dr. Plante KS, Dr. Schafer A, Dr. Baric RS, Dr. Ferris MT.


