nature portfolio

Corresponding author(s): Kirsten Nielsen

Last updated by author(s): Sep 13, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Variant calling for each strain was adapted from the best practices described for the Genome Analysis Toolkit (GATK v3.3.0).

Heat map clusters were generated using Heatmapper (https://pubmed.ncbi.nlm.nih.gov/27190236/).

The first 10 principal components were then calculated from the filtered genotype data with PLINK 1.9.

The ordinal categorical variables were analyzed with a proportional odds logistic mixed model (POLMM) implemented in the 'GRAB' package for R. The quantitative variables were analyzed using a mixed linear model implemented in the GEMMA software package.

For the RNAseq, clean reads were mapped to the annotated genome of C. neoformans H99 (NCBI:txid235443) using HISAT2 software (version 2.0.5). Differential expression analyses were conducted using the DESeq2 package (version 1.20.0).

Data analysis

Principal Component Analysis
The variants used in the principal component analysis were those previously identified as effect variants. A custom Python script was used to transform the genotype data into a PLINK binary format. Variant sites with a minor allele count of less than three were excluded from the dataset. The first 10 principal components were then calculated from the filtered genotype data with PLINK 1.9. A scree plot was used to determine the number of principal components that should be included as covariates to the association analyses: an "elbow" in the proportion of variance explained by each principal component was chosen as the threshold of inclusion. Isolates were clustered using K-means clustering and Euclidean distance to determine the distance between points. Gap statistic and silhouette score were used to compute the optimal number of clusters. PC1 and PC2 plot was generated with the color of the points corresponding to their K=3 assignments.

Genome Wide Association Study
A total of 31 phenotypes were analyzed for genotype to phenotype associations with a genome wide association study (GWAS) framework. Phenotypes were divided into two categories, ordinal categorical and quantitative (Table S1). The ordinal categorical variables were analyzed with a proportional odds logistic mixed model (POLMM) implemented in the 'GRAB' package for R. The quantitative variables were analyzed using a mixed linear model implemented in the GEMMA software package. A total of 562 variants survived the minor allele count filter and were used in the association analyses. The first two principal components of the genotype data were used as covariates to control for relatedness or population structure in the sample. Significant variants were identified by likelihood ratio test P values < 0.05 after a Bonferroni correction, where the number of variants (562) was used as the number of independent hypothesis tests. The bioinformatic pipeline is at GitHub: https://github.com/TomJKono/C_neoformans_Association.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The bioinformatic pipeline scripts are available through GitHub at https://github.com/TomJKono/C_neoformans_Association. The original sequence data is housed in the BioProject ID PRJNA549026 and the RNAseq dataset can be found in the NCBI database under the submission ID SUB14630915 and BioProject ID PRJNA1142147.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.* *Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.* *Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Mouse survival studies (figure 2, 5a-c, 6a-c) were performed with either 5 or 10 mice per isolate.

Sample size	For the studies on clinical isolates (figure 2), 38 strains were tested in mice, so, to avoid unnecessary loss of animal life, we selected 5 mice for as many of the strains as possible. We chose 5 instead of 3 because clinical isolates are inherently variable. We did an initial pilot study using 3 clinical isolates and found that we were unable to perform statistics due to the variability in fungal burden and survival. 5 mice was the minimum number that gave us a clear description of the mouse survival without requiring us to repeat the studies with additional mice. For 29 of the 38 strains, we opted to also measure cytokine response. For those isolates, we wanted a more defined view of survival and so used 10 mice for survival analysis and 5 mice for cytokine analysis (15 mice total). This allowed us to select the most representative isolates possible for the cytokine studies (ie- if some mice earmarked for cytokines died before the 21 day time point, random mice could be selected from the remaining 10). 5 mice were used for each cytokine analysis; again, clinical isolates showed enough variability that when we did a pilot study using only 3 mice, there was too much variability for us to determine statistics or glean useful data. For all cytokines, each sample from the 5 mice were run in triplicate. For mouse studies using deletion mutants (figures 5a-c, 6a-c), we opted to use 15 mice, 10 for survival and 5 for cytokines. We chose to do this because we wanted to make sure we had the most clear picture of survival without needed to repeat any of the mouse studies. Historically, some of the deletion mutants had only minor impacts on survival and so we wanted to ensure we were identifying genes with minor impacts in this study. For the histopathology experiments (figure 3), we infected 3 mice with each strain. Pictures were taken across all samples and stains and representative images were chosen. For the cell and capsule size studies, 150 cells from 5 mice (50/mouse) were measured from the lungs. We tried to use a similar number for cells collected from the brain, but the brain is a challenging organ in which to collect cells and we were not always able to obtain enough yeast cells. 100 is the standard number of cells required for Cryptococcus titan cell and capsule analysis. For the RNAseq and qPCR (figure 7), three biological replicates were used, which is a standard number required for a one-way ANOVA.
Data exclusions	ROUT outlier analysis was performed to remove outliers from the cytokines and fungal burden in the organs. While all the cytokines were run in technical triplicates, if one replicate from the triplicates was substantially different from the other two replicates across all the cytokines, that replicate was excluded. IL-2 results were excluded because IL-2 was not significantly increased in any of the clinical isolates we studied. Mouse data was excluded if there was a subjective reason for euthanasia (ie-mouse in distress), but no fungal burden in lungs or brain. Mouse data was also excluded if the mouse died for reasons unrelated to fungal infection (ie- tumor). Mechanisms of exclusion were determined in advance, with the exception of IL-2. We opted to exclude IL-2 once we saw that it was universally not significant.
Replication	This study addressed the fundamental question (what SNPs and genes in Cryptococcus neoformans are associated with virulence?) using two different methods. In the first, we used four published genomic studies that used diverse methods and diverse isolates to separately identify genes and gene regions associated with virulence. We combined the data of those four studies and identified 7 genes and gene regions that overlapped in at least two studies. Next, we used a closely related set of clinical isolates, infected mice with those isolates, and then performed a genome wide association study. The genes identified following the GWAS included 6 found in the meta-analysis, with an additional one that was identified in a previous genomic study. Overall, this convinces us that the genes and gene regions identified are likely to be correct and associated with virulence. Finally, we identified SNPs in genes associated with increased IFNgamma; we then deleted those genes, infected mice, and found an IFNgamma phenotype in two of the strains. Additionally, data was internally consistent throughout the study. For example, we saw a high IFNgamma phenotype in hypervirulent isolates based on cytokine results; the histopathology also showed signs of inflammation. All studies were performed using technical (three) or biological replicates (5-10). While there were some outliers, in general the replicates grouped and data were significant. Clusters from the principal component analysis were determined using both K-means clustering and gap statistic.
Randomization	Randomization was not relevant to most of this this study. We used a preexisting library of 38 clinical isolates and infected mice with all 38 isolates. Studies using deletion mutants were performed on all genes that had SNPs associated with IFNgamma. We did select 29 isolates to determine the immune response in a subset of the 38 isolates: we determined the cytokine response for all hypervirulent and latent isolates. The remaining 20 were selected at random.
Blinding	Blinding was not relevant to this study. None of the clinical isolates or deletion strains had been in mice previously, so we had no preconceptions about what the data would be. The genome wide association study was performed by an independent investigator who was uninvolved in data generation. Terminal endpoints for mouse studies were objective (ie-weight loss) and when subjective (ie-mouse appears in distress) the reason for euthanizing the mouse was recorded.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

InVivoMAb anti-mouse IFNγ; Bio X Cell; Catalog number #BE0054; clone name R4-6A2 Lot number: 749322J1
InVivoPlus rat IgG1 isotype control, anti-horseradish peroxidase; Bio X Cell; Catalog number #BP0088; clone name HRPN; Lot number: 908123J3B

Validation

Manufacture general validation statements:

Purity
Each lot of each antibody is checked for purity and integrity using SDS-PAGE analysis to ensure the purity level is ≥95%.

Binding Validation
Each lot of each applicable antibody is screened for binding specificity against it's antigen and a negative control antigen via immunoblot. For select products we also ensure binding using flow cytometry.

Antibody Aggregation Analysis
Each lot of each applicable antibody is analyzed using size exclusion chromatography (SEC) to determine the aggregate level and ensure a monomer content of ≥95%.

Murine Pathogen Screening
Each lot of each applicable antibody is screened for an extensive panel of murine viruses as well as mycoplasma using a highly-specific ultra-sensitive real-time PCR based assay.Murine pathogen screening includes:
Mycoplasma including M. pulmonis, M. arginini, M. fermentans, M. hominis, M. hyorhinis, M. orale, M. pirum, M. salivarium, M. agassizii, M. cynos and others.Murine norovirus (MNV), murine parvovirus (MPV) 1, 2, 3, 4, and 5, murine minute virus (MMV/MVM), murine hepatitis virus (MHV), murine reovirus (REO) type 1, 2, and 3, lymphocytic choriomeningitis virus (LCMV), lactate dehydrogenase elevating virus (LDV), murine rotavirus MRV/EDIM), Theiler's murine encephalomyelitis virus (TMEV), ectromelia virus (ECTRO), hantavirus (HANTA), polyoma virus (POLY), murine adenovirus (detects and differentiates MAD1 and MAD2), sendai virus (SEND), pneumonia virus of mice (PVM), murine cytomegalovirus (MCMV) 1 and 2 (MCMV), and K virus.

Isotype Confirmation
Each lot of each antibody is isotyped using a rapid lateral flow antibody isotyping assay ensuring that the antibody is the correct host species, isotype class, subtype class, and light-chain identity.

Endotoxin Testing
Each lot of each antibody is screened and ensured to be ultra-low in bacterial endotoxin using the Limulus Amebocyte Lysate (LAL) gel clotting assay.

Specific manufacture validation:
Manufacture validation information: https://bioxcell.com/pub/media/tds/BE0054-TDS.pdf
Manufacture validation information: https://bioxcell.com/pub/media/tds/BP0088-TDS.pdf

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals

A/J mice from Jackson laboratories

Wild animals

Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Reporting on sex

Female mice were used for all experiments due to the length of the studies (survival was analyzed through 100 days post-infection). Hormone fluctuations in older male mice are known to influence Cryptococcus pathogenesis (Lortholary et al., 2002). This information was added to the methods section.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight	Animal experiments were done in accordance with the Animal Welfare Act, United States federal law, and NIH guidelines. Mice were handled in accordance with guidelines defined by the University of Minnesota Animal Care and Use Committee (IACUC) under protocols 1908A37344, 2207A40205, and 2104A39016.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*