

Streptomycin resistance in bacteria: insights into evolution and evolvability

Summaries of analyses

2025-05-28

Summary of reported mutations

- All mutations:
 - Number of studies: 98
 - Number of species: 40
 - Total number of mutations: 311
 - Amino acid positions with mutations: 31
 - Number of unique amino acid mutations: 66
- Filtered mutations:
 - Number of studies: 95
 - Number of species: 39
 - Total number of mutations: 219
 - Amino acid positions with mutations: 5
 - Number of unique amino acid mutations: 8
- Mutations in *Mycobacterium tuberculosis* (MTB) clinical isolates:
 - Number of filtered mutations also reported in MTB: 4
 - Number of filtered mutations collated without MTB data: 8
 - Number of filtered mutations collated without MTB data also reported in MTB: 4

Summary of screened target rpsL sequences

- Target sequence statistics before filtering:
 - Number of genomes in search results: 20866
 - Number of genomes with extracted rpsL sequence(s): 20808

- Number of genomes where no rpsL sequence could be extracted: 58
- Number of species with downloaded genomes: 20804
- Total number of rpsL sequences: 20872
- Number of genomes with more than one rpsL sequence: 59
- Maximum number of rpsL sequences per genome: 4
- Filtering statistics:
 - Minimum sequence length: 63
 - Median sequence length: 375.00
 - Mean sequence length: 384.1486
 - Maximum sequence length: 468
 - Minimum aligning score: -614.3201
 - Maximum aligning score: 537.959
 - Minimum distance from reference rpsL core: 0
 - Maximum distance from reference rpsL core: 115
- Filters applied:
 - Minimum sequence length: 300
 - Minimum alignment score to E. coli rpsL: -614.320068359375
 - Maximum core distance to E. coli rpsL: 40
- Target sequence statistics after filtering:
 - Number of genomes with extracted rpsL sequence(s): 15725
 - Number of species with downloaded genomes: 15721
 - Total number of rpsL sequences: 15740
 - Number of genomes with more than one rpsL sequence: 15
 - Maximum number of rpsL sequences per genome: 2

Summary of the mutation screen

- Screened mutations:
 - Number of distinct mutations screened: 8
 - Number of amino acid sites: 5
- Predicted intrinsic resistance:
 - Total number of resistance mutations present across species: 1193
 - Number of unique mutations present across species: 4
 - Number of resistant species: 1190
 - Percentage of resistant species: 7.57%
 - Number of species with multiple resistance mutations: 3
 - Number of species where one rpsL copy confers resistance and one does not: 0

- Evolvability I (number of AA mutations that a species can mutate to):
 - Range: 2...8
 - 95% inter-quantile range: 6...8
 - Median: 8
 - Theoretical range: 0...8
- Evolvability II (number of mutations that can produce a resistance mutation):
 - Range: 5...18
 - 95% inter-quantile range: 7...10
 - Median: 9
 - Theoretical range: 0...17
- Associations between evolvabilities (phylogenetically uncontrolled):
 - Evolvability I vs. evolvability II: $r=0.559$
 - Intrinsic resistance vs. evolvability I: $p=0$ (Welch Two Sample t-test, $t=-115$)
 - Intrinsic resistance vs. evolvability II: $p=5.41e-17$ (Welch Two Sample t-test, $t=8.5$)

Summary of the phylogenetic analyses

- Phylogenetic tree:
 - Number of species: 12162
 - Number of species predicted to be resistant in tree: 744
- Number of species sampled from tree when calculating phylogenetic signals: 5000
- Phylogenetic signal in predicted resistance:
 - Test: permutation test of mean phylogenetic distance of resistant species
 - Number of permutations: 1000
 - p-value: 0
- Phylogenetic signal in evolvability I (number of evolvable AA mutations):
 - Pagel's lambda: 0.994
 - $p(\text{lambda})$: 0
 - Blomberg's K: 0.362
 - $p(K)$: 0.001
- Phylogenetic signal in evolvability II (number of nt mutations producing AA resistance mutations):
 - Pagel's lambda: 0.938
 - $p(\text{lambda})$: 0
 - Blomberg's K: 0.153
 - $p(K)$: 0.001

Summary of the amino acid conservation analyses

- Mean Hamming distance to E. coli across all sequences:
 - Number of sequences: 15740
 - Mean across AA positions: 0.226062118293233
 - Max across AA positions: 0.943202033036849
 - Min across AA positions: 0
- Mean Hamming distance across randomly sampled pairs of frpsL sequences:
 - Number of sequence pairs: 100000
 - Mean across AA positions: 0.231670434571266
 - Max across AA positions: 0.88367714376493
 - Min across AA positions: 0
- Mean Grantham distance to E. coli across all sequences:
 - Number of sequences: 15740
 - Mean across AA positions: 15.7975369179425
 - Max across AA positions: 161.740199504416
 - Min across AA positions: 0
- Mean Grantham distance across randomly sampled pairs of frpsL sequences:
 - Number of sequence pairs: 100000
 - Mean across AA positions: 15.4150348878232
 - Max across AA positions: 77.8605684017521
 - Min across AA positions: 0

Session Info

R version 4.4.3 (2025-02-28)

Platform: x86_64-apple-darwin20

locale: en_US.UTF-8|en_US.UTF-8|en_US.UTF-8|C|en_US.UTF-8|en_US.UTF-8

attached base packages: *grid*, *parallel*, *stats4*, *stats*, *graphics*, *grDevices*, *utils*, *datasets*, *methods* and *base*

other attached packages: *igraph*(v.2.1.4), *tidygraph*(v.1.3.1), *ggraph*(v.2.2.1), *cowplot*(v.1.1.3), *ggrepel*(v.0.9.6), *wesanderson*(v.0.3.7), *htmlwidgets*(v.1.6.4), *NGLViewR*(v.1.4.0), *quarto*(v.1.4.4), *pander*(v.0.6.6), *ggh4x*(v.0.3.0), *GGally*(v.2.2.1), *RColorBrewer*(v.1.1-3), *plotrix*(v.3.8-4), *ggpubr*(v.0.6.0), *patchwork*(v.1.3.0), *scales*(v.1.3.0), *colorspace*(v.2.1-1), *ggnewscale*(v.0.5.1), *bio3d*(v.2.4-5), *treeio*(v.1.30.0), *tidytrees*(v.0.4.6), *ggtree*(v.3.14.0), *phangorn*(v.2.12.1), *phytools*(v.2.4-4), *maps*(v.3.4.2.1), *ape*(v.5.8-1), *castor*(v.1.8.3),

Rcpp(v.1.0.14), *MSA2dist*(v.1.10.1), *palign*(v.1.2.0), *rentrez*(v.1.2.3), *varhandle*(v.2.0.6), *future.apply*(v.1.11.3), *future*(v.1.40.0), *openssl*(v.2.3.2), *lubridate*(v.1.9.4), *forcats*(v.1.0.0), *stringr*(v.1.5.1), *dplyr*(v.1.1.4), *purrr*(v.1.0.4), *readr*(v.2.1.5), *tidyr*(v.1.3.1), *tibble*(v.3.2.1), *ggplot2*(v.3.5.2), *tidyverse*(v.2.0.0), *ALJEbinf*(v.0.1.0), *Biostrings*(v.2.74.1), *GenomeInfoDb*(v.1.42.3), *XVector*(v.0.46.0), *IRanges*(v.2.40.1), *S4Vectors*(v.0.44.0) and *BiocGenerics*(v.0.52.0)

loaded via a namespace (and not attached): *later*(v.1.4.2), *ggplotify*(v.0.1.2), *polyclip*(v.1.10-7), *XML*(v.3.99-0.18), *lifecycle*(v.1.0.4), *rstatix*(v.0.7.2), *doParallel*(v.1.0.17), *globals*(v.0.16.3), *processx*(v.3.8.6), *lattice*(v.0.22-7), *vroom*(v.1.6.5), *MASS*(v.7.3-65), *backports*(v.1.5.0), *magrittr*(v.2.0.3), *rmarkdown*(v.2.29), *yaml*(v.2.3.10), *httpuv*(v.1.6.15), *askpass*(v.1.2.1), *ade4*(v.1.7-23), *abind*(v.1.4-8), *zlibbioc*(v.1.52.0), *expm*(v.1.0-0), *quadprog*(v.1.5-8), *GenomicRanges*(v.1.58.0), *yulab.utils*(v.0.2.0), *tweenr*(v.2.0.3), *GenomeInfoDbData*(v.1.2.13), *listenv*(v.0.9.1), *RSpectra*(v.0.16-2), *parallelly*(v.1.43.0), *codetools*(v.0.2-20), *ggforce*(v.0.4.2), *tidyselect*(v.1.2.1), *aplot*(v.0.2.5), *UCSC.utils*(v.1.2.0), *farver*(v.2.1.2), *viridis*(v.0.6.5), *jsonlite*(v.2.0.0), *Formula*(v.1.2-5), *iterators*(v.1.0.14), *systemfonts*(v.1.2.2), *foreach*(v.1.5.2), *tools*(v.4.4.3), *ragg*(v.1.4.0), *glue*(v.1.8.0), *mnormt*(v.2.1.1), *gridExtra*(v.2.3), *xfun*(v.0.52), *withr*(v.3.0.2), *numDeriv*(v.2016.8-1.1), *combinat*(v.0.0-8), *fastmap*(v.1.2.0), *digest*(v.0.6.37), *timechange*(v.0.3.0), *R6*(v.2.6.1), *mime*(v.0.13), *gridGraphics*(v.0.5-1), *textshaping*(v.1.0.0), *utf8*(v.1.2.4), *generics*(v.0.1.3), *clusterGeneration*(v.1.3.8), *graphlayouts*(v.1.2.2), *httr*(v.1.4.7), *scatterplot3d*(v.0.3-44), *ggstats*(v.0.9.0), *pkgconfig*(v.2.0.3), *gtable*(v.0.3.6), *htmltools*(v.0.5.8.1), *carData*(v.3.0-5), *ggfun*(v.0.1.8), *knitr*(v.1.50), *rstudioapi*(v.0.17.1), *tzdb*(v.0.5.0), *coda*(v.0.19-4.1), *nlme*(v.3.1-168), *cachem*(v.1.1.0), *DEoptim*(v.2.2-8), *pillar*(v.1.10.2), *vctrs*(v.0.6.5), *promises*(v.1.3.2), *car*(v.3.1-3), *xtable*(v.1.8-4), *evaluate*(v.1.0.3), *cli*(v.3.6.4), *compiler*(v.4.4.3), *rlang*(v.1.1.6), *crayon*(v.1.5.3), *ggsignif*(v.0.6.4), *labeling*(v.0.4.3), *ps*(v.1.9.1), *natural sort*(v.0.1.3), *plyr*(v.1.8.9), *fs*(v.1.6.6), *stringi*(v.1.8.7), *viridisLite*(v.0.4.2), *munsell*(v.0.5.1), *lazyeval*(v.0.2.2), *optimParallel*(v.1.0-2), *Matrix*(v.1.7-3), *hms*(v.1.1.3), *bit64*(v.4.6.0-1), *seqinr*(v.4.2-36), *shiny*(v.1.10.0), *broom*(v.1.0.8), *memoise*(v.2.0.1), *fastmatch*(v.1.1-6) and *bit*(v.4.6.0)