Evolution and evolvability of rifampicin resistance across the bacterial tree of life

Summaries of analyses

2024-09-05

Summary of reported mutations

- Number of studies: 116
- Number of species: 75
- Total number of mutations: 1426
- Amino acid positions with mutations: 146
- Number of unique amino acid mutations: 369

Summary of screened target rpoB sequences

- Target sequence statistics before filtering:

    - Number of genomes in search results: 18636
    - Number of genomes with extracted rpoB sequence(s): 18469
    - Number of genomes where no rpoB sequence could be extracted: 167
    - Number of species with downloaded genomes: 18450
    - Total number of rpoB sequences: 18682
    - Number of genomes with more than one rpoB sequence: 208
    - Maximum number of rpoB sequences per genome: 3

- Filtering statistics:

    - Minimum sequence length: 312
    - Median sequence length: 3813.00
    - Mean sequence length: 3865.628
    - Maximum sequence length: 8703
    - Minimum aligning score: -8559.608
    - Maximum aligning score: 5779.872

- – Minimum distance from reference rpoB core: 0
- – Maximum distance from reference rpoB core: 171

- • Filters applied:

  - – Minimum sequence length: 3000
  - – Minimum alignment score to E. coli rpoB: -8559.6083984375
  - – Maximum core distance to E. coli rpoB: 35

- • Target sequence statistics after filtering:

  - – Number of genomes with extracted rpoB sequence(s): 18146
  - – Number of species with downloaded genomes: 18127
  - – Total number of rpoB sequences: 18238
  - – Number of genomes with more than one rpoB sequence: 91
  - – Maximum number of rpoB sequences per genome: 3

Summary of the mutation screen

- • Screened mutations:

  - – Number of distinct mutations screened: 60
  - – Number of amino acid sites: 19

- • Predicted intrinsic resistance:

  - – Total number of resistance mutations present across species: 1567
  - – Number of unique mutations present across species: 37
  - – Number of resistant species: 1453
  - – Percentage of resistant species: 8.02%
  - – Number of species with multiple resistance mutations: 104
  - – Number of species where one rpoB copy confers resistance and one does not: 68

- • Evolvability I (number of AA mutations that a species can mutate to):

  - – Range: 27...49
  - – 95% inter-quantile range: 35...47
  - – Median: 42
  - – Theoretical range: 2...53

- • Evolvability II (number of mutations that can produce a resistance mutation):

  - – Range: 33...123
  - – 95% inter-quantile range: 40...53
  - – Median: 46
  - – Theoretical range: 3...86

- • Associations between evolvabilities (phylogenetically uncontrolled):

- Evolvability I vs. evolvability II: r=0.69
- Intrinsic resistance vs. evolvability I: p=7.25e-80 (Welch Two Sample t-test, t=-20.1)
- Intrinsic resistance vs. evolvability II: p=0.0163 (Welch Two Sample t-test, t=2.41)

Summary of the phylogenetic analyses

- Phylogenetic tree:

    - Number of species: 15521
    - Number of species predicted to be resistant in tree: 1228

- 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: 0.954
    - p(lambda): 0
    - Blomberg's K: 0.204
    - p(K): 0.001

- Phylogenetic signal in evolvability II (number of nt mutations producing AA resistance mutations):

    - Pagel's lambda: 0.74
    - p(lambda): 0
    - Blomberg's K: 0.102
    - p(K): 0.001

Summary of the mutation screen (additional mutations from Yang et al. (2023))

- Screened mutations:

    - Number of distinct mutations screened: 85
    - Number of amino acid sites: 22

- Predicted intrinsic resistance:

    - Total number of resistance mutations present across species: 171
    - Number of unique mutations present across species: 11
    - Number of resistant species: 169

- – Percentage of resistant species: 0.93%
- – Number of species with multiple resistance mutations: 2
- – Number of species where one rpoB copy confers resistance and one does not: 8

- Evolvability I (number of AA mutations that a species can mutate to):

  - – Range: 4...17
  - – 95% inter-quantile range: 5...11
  - – Median: 7
  - – Theoretical range: 0...53

- Evolvability II (number of mutations that can produce a resistance mutation):

  - – Range: 4...35
  - – 95% inter-quantile range: 5...13
  - – Median: 8
  - – Theoretical range: 0...89

- Associations between evolvabilities (phylogenetically uncontrolled):

  - – Evolvability I vs. evolvability II: r=0.913
  - – Intrinsic resistance vs. evolvability I: p=2.85e-22 (Welch Two Sample t-test, t=11.2)
  - – Intrinsic resistance vs. evolvability II: p=8.72e-42 (Welch Two Sample t-test, t=18.2)

Summary of the mutation screen (reported mutations and mutations from Yang et al. (2023) pooled)

- Screened mutations:

  - – Number of distinct mutations screened: 145
  - – Number of amino acid sites: 29

- Predicted intrinsic resistance:

  - – Total number of resistance mutations present across species: 1738
  - – Number of unique mutations present across species: 48
  - – Number of resistant species: 1578
  - – Percentage of resistant species: 8.71%
  - – Number of species with multiple resistance mutations: 132
  - – Number of species where one rpoB copy confers resistance and one does not: 68

- Evolvability I (number of AA mutations that a species can mutate to):

  - – Range: 38...62
  - – 95% inter-quantile range: 43...54
  - – Median: 49

- Theoretical range: 21...77

- Evolvability II (number of mutations that can produce a resistance mutation):

  - Range: 43...158
  - 95% inter-quantile range: 49...60
  - Median: 54
  - Theoretical range: 21...134

- Associations between evolvabilities (phylogenetically uncontrolled):

  - Evolvability I vs. evolvability II: r=0.638
  - Intrinsic resistance vs. evolvability I: p=0.00165 (Welch Two Sample t-test, t=3.15)
  - Intrinsic resistance vs. evolvability II: p=1.45e-39 (Welch Two Sample t-test, t=13.5)

Session Info

R version 4.4.0 (2024-04-24)

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

other attached packages: *quarto(v.1.4.4)*, *pander(v.0.6.5)*, *ggh4x(v.0.2.8)*, *GGally(v.2.2.1)*, *RColorBrewer(v.1.1-3)*, *plotrix(v.3.8-4)*, *ggpubr(v.0.6.0)*, *patchwork(v.1.2.0)*, *scales(v.1.3.0)*, *colorspace(v.2.1-1)*, *ggnewscale(v.0.4.10)*, *treeio(v.1.28.0)*, *tidytree(v.0.4.6)*, *ggtree(v.3.12.0)*, *phangorn(v.2.11.1)*, *phytools(v.2.1-1)*, *maps(v.3.4.2)*, *ape(v.5.8)*, *castor(v.1.8.0)*, *Rcpp(v.1.0.13)*, *pwalign(v.1.0.0)*, *Biostrings(v.2.72.0)*, *GenomeInfoDb(v.1.40.1)*, *XVector(v.0.44.0)*, *IRanges(v.2.38.0)*, *S4Vectors(v.0.42.0)*, *BiocGenerics(v.0.50.0)*, *varhandle(v.2.0.6)*, *future.apply(v.1.11.2)*, *future(v.1.33.2)*, *openssl(v.2.2.0)*, *lubridate(v.1.9.3)*, *forcats(v.1.0.0)*, *stringr(v.1.5.1)*, *dplyr(v.1.1.4)*, *purrr(v.1.0.2)*, *readr(v.2.1.5)*, *tidyr(v.1.3.1)*, *tibble(v.3.2.1)*, *ggplot2(v.3.5.1)*, *tidyverse(v.2.0.0)* and *ALJEbinf(v.0.1.0)*

loaded via a namespace (and not attached): *rstudioapi(v.0.16.0)*, *jsonlite(v.1.8.8)*, *magrittr(v.2.0.3)*, *farver(v.2.1.2)*, *rmarkdown(v.2.27)*, *fs(v.1.6.4)*, *zlibbioc(v.1.50.0)*, *ragg(v.1.3.2)*, *vctrs(v.0.6.5)*, *memoise(v.2.0.1)*, *askpass(v.1.2.0)*, *rstatix(v.0.7.2)*, *htmltools(v.0.5.8.1)*, *broom(v.1.0.6)*, *gridGraphics(v.0.5-1)*, *parallelly(v.1.37.1)*, *naturalsort(v.0.1.3)*, *plyr(v.1.8.9)*, *cachem(v.1.1.0)*, *igraph(v.2.0.3)*, *lifecycle(v.1.0.4)*, *iterators(v.1.0.14)*, *pkgconfig(v.2.0.3)*, *Matrix(v.1.7-0)*, *R6(v.2.5.1)*, *fastmap(v.1.2.0)*, *GenomeInfoDbData(v.1.2.12)*, *digest(v.0.6.36)*, *numDeriv(v.2016.8-1.1)*, *aplot(v.0.2.2)*, *ps(v.1.7.6)*, *RSpectra(v.0.16-1)*, *textshaping(v.0.4.0)*, *labeling(v.0.4.3)*, *clusterGeneration(v.1.3.8)*, *fansi(v.1.0.6)*, *timechange(v.0.3.0)*, *httr(v.1.4.7)*,

abind(v.1.4-5), compiler(v.4.4.0), bit64(v.4.0.5), withr(v.3.0.1), doParallel(v.1.0.17), backports(v.1.5.0), optimParallel(v.1.0-2), carData(v.3.0-5), ggstats(v.0.6.0), ggsignif(v.0.6.4), MASS(v.7.3-60.2), scatterplot3d(v.0.3-44), tools(v.4.4.0), glue(v.1.7.0), quadprog(v.1.5-8), nlme(v.3.1-164), grid(v.4.4.0), generics(v.0.1.3), gtable(v.0.3.5), tzdb(v.0.4.0), hms(v.1.1.3), car(v.3.1-2), utf8(v.1.2.4), foreach(v.1.5.2), pillar(v.1.9.0), yulab.utils(v.0.1.4), vroom(v.1.6.5), later(v.1.3.2), lattice(v.0.22-6), bit(v.4.0.5), tidyselect(v.1.2.1), knitr(v.1.48), gridExtra(v.2.3), xfun(v.0.46), expm(v.0.999-9), stringi(v.1.8.4), UCSC.utils(v.1.0.0), lazyeval(v.0.2.2), ggfun(v.0.1.4), yaml(v.2.3.10), evaluate(v.0.24.0), codetools(v.0.2-20), ggplotify(v.0.1.2), cli(v.3.6.3), systemfonts(v.1.1.0), munsell(v.0.5.1), processx(v.3.8.4), globals(v.0.16.3), coda(v.0.19-4.1), listenv(v.0.9.1), viridisLite(v.0.4.2), crayon(v.1.5.3), combinat(v.0.0-8), rlang(v.1.1.4), cowplot(v.1.1.3), fastmatch(v.1.1-4) and mnormt(v.2.1.1)