Statistical end-to-end analysis of large-scale microbial growth data with `DGrowthR`

Medina Feldl[1,2,*], Roberto Olayo-Alarcon[1,2,*], Martin K. Amstalden[3], Annamaria Zannoni[6], Stefanie Peschel[1], Cynthia M. Sharma[6], Ana Rita Brochado[3,4,5], and Christian L. Müller[1,2,7]

[1]Department of Statistics, Ludwig-Maximilians-Universität München
[2]Institute of Computational Biology, Helmholtz Munich
[3]Department of Microbiology, Biocenter, Julius-Maximilians-Universität Würzburg
[4]Interfaculty Institute of Microbiology and Infection Medicine Tübingen (IMIT), University of Tübingen
[5]Cluster of Excellence 'Controlling Microbes to Fight Infections' (CMFI), University of Tübingen
[6]Department of Molecular Infection Biology II, Institute of Molecular Infection Biology (IMIB), Julius-Maximilians-Universität Würzburg
[7]Center for Computational Mathematics, Flatiron Institute, New York
[*]Authors contributed equally

Abstract

Quantitative analysis of microbial growth curves is essential for understanding how bacterial populations respond to environmental cues. Traditional analysis approaches make parametric assumptions about the functional form of these curves, limiting their usefulness for studying conditions that distort standard growth curves. In addition, modern robotics platforms enable the high-throughput collection of large volumes of growth data, thus requiring strategies that can analyze large-scale growth data in a flexible and efficient manner. Here, we introduce `DGrowthR`, a statistical R and standalone app framework for the integrative analysis of large growth experiments. `DGrowthR` comprises methods for data pre-processing and standardization, exploratory functional data analysis, and non-parametric modeling of growth curves using Gaussian Process regression. Importantly, `DGrowthR` includes a rigorous statistical testing framework for differential growth analysis. To illustrate the range of application scenarios of `DGrowthR`, we analyzed three large-scale bacterial growth datasets that tackle distinct scientific questions. On an in-house large-scale growth dataset comprising two pathogens that were subjected to a large chemical perturbation screen, `DGrowthR` enabled the discovery of compounds with *significant* growth inhibitory effects as well as compounds that induce non-canonical growth dynamics. We also re-analyzed two publicly available datasets and recovered reported adjuvants and antagonists of antibiotic activity, as well as bacterial genetic factors that determine susceptibility to specific antibiotic treatments. We anticipate that `DGrowthR` will streamline the analysis of modern high-volume growth experiments, enabling researchers to gain novel biological insights in a standardized and reproducible manner.

Introduction

The quantitative analysis of microbial growth is essential for characterizing bacterial populations. Experiments where measurements of bacterial growth (typically reported as optical density (OD)) are taken at multiple time points provide valuable information about the dynamics of bacterial populations over time [1, 2]. Current automatic platforms enable researchers to collect large volumes of these growth curves with ease by gathering data from multiple samples simultaneously [3]. The readouts from these experiments have the potential to provide valuable insights into the mechanisms underlying the bacterial response to various environmental and chemical cues. Appropriate statistical approaches and computational tools are necessary in order to gain valuable insights from these high-throughput data.

Traditional approaches to model time-series growth data, such as the Logistic model, make parametric assumptions about the functional form of these curves [4]. Multiple popular software packages, including `Sicegar` [5] and `GrowthcurveR` [6], provide options for modeling single growth curves with sigmoidal or double sigmoidal fits and extracting relevant microbial growth parameters. However, growth curves gathered under various environment challenges can deviate significantly from these canonical forms and therefore violate the parametric assumptions of these models. While more flexible approaches for growth curve modeling exist, including `grofit` [7], `gcplyr` [8], and `QurvE` [9], they often rely on techniques such as splines and are sensitive to outliers and technical variations. Furthermore, these tools lack principled statistical testing capabilities for comparing growth curve characteristics across different experimental conditions. Tools such as `NeckaR` [3] can compare growth curves based on area under the curve (AUC) values, but lack the ability to extract relevant growth parameters.

To model and analyze more complex bacterial growth dynamics, Gaussian Process (GP) regression has recently emerged as a valuable alternative [10–12]. GP regression is a non-parametric approach that does not make explicit assumptions about the functional form of the growth curves and can jointly model multiple (replicate) growth curves, thus reducing the influence of outliers. The outcome of GP regression can be used to determine robust growth parameters, such as growth rates, carrying capacity, and AUC values, as well as determine different phases of growth [10, 12]. Additionally, Tonner et al. [11] showed that GP regression can be leveraged to perform hypothesis testing, comparing complete growth dynamics instead of relying on singular growth parameters [11]. The software package `AMiGA` [12] provides a comprehensive framework for the analysis of microbial growth curves using GP regression. However, `AMiGA` is implemented in Python and requires the use of command-line instructions, leaving a gap in the R ecosystem. Furthermore, none of the tools provide functionality to perform exploratory data analysis in form of functional data analysis [13], clustering, or popular dimensionality reduction techniques.

In this contribution, we present `DGrowthR`, an R package that enables statistical end-to-end analysis of high-throughput microbial growth datasets. Firstly, `DGrowthR` offers functionality for data pre-processing, normalization, and standardization. Secondly, `DGrowthR` supports a wide range of visualizations for large-scale datasets, including plotting tools for individual curves as well as low-dimensional embeddings via Functional Principal Component Analysis (FPCA) and Uniform Manifold Approximation and Projection (UMAP) [14–16]. Moreover, these embeddings facilitate the identification of distinct growth patterns (clusters) within the data by employing advanced methods from functional data clustering [15]. Thirdly, `DGrowthR` can flexibly model collections of growth curves using GP regression and automatically extract all relevant growth curve characteristics. Finally, `DGrowthR` includes a computationally efficient permutation-based statistical testing framework for comparing growth curves across different experimental conditions, providing measurements of evidence in the form of Gamma-approximated permutation p-values [17]. The modular software structure of `DGrowthR` also facilitates storing of all relevant analysis results in structured objects that can be conveniently accessed for custom downstream analysis.

We illustrate the broad applicability and the unique features of `DGrowthR` by analyzing three large-scale datasets that comprise different bacterial species and their response to various forms of genetic and chemical perturbations. The first dataset consists of in-house chemical screens of two pathogens, *Salmonella enterica* and *Campylobacter jejuni*, against a diverse chemical library of 2,419 compounds. The second dataset is from Brenzinger et al. [18], where changes in the susceptibility of *Vibrio cholerae* to various chemical challenges is measured upon deletion of the cyclic-oligonucleotide-based anti-phage signaling system (CBASS). The third dataset from Brochado et al. [19] comprises measurements to study the effect of pairwise drug combinations on the growth of gram-negative pathogens. Using `DGrowthR`, we not only recapitulate the results of the original studies in an automated and reproducible fashion but also uncover additional insights into the growth dynamics of the bacteria in the different scenarios. We posit that `DGrowthR` provides a general flexible statistical end-to-end framework, enabling microbiologists to streamline their analysis of large-scale microbial growth datasets in a computationally efficent and reproducible fashion.

Results

The DGrowthR framework

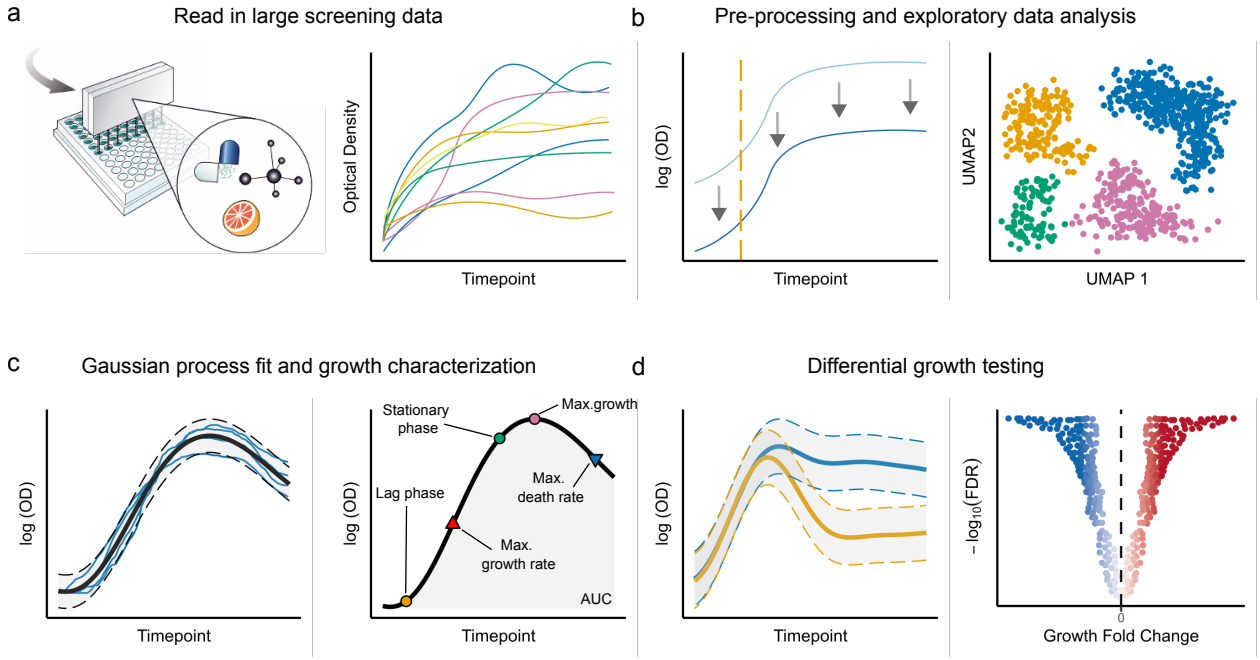

Figure 1: Schematic overview of the DGrowthR software framework. **a**. DGrowthR can read-in data from automatic plate-readers that gather data for multiple growth curves simultaneously. These curves can be visually inspected to decide on the next steps. **b**. Raw growth curve pre-processing includes baseline correction, log-transformation of OD measurements, and determination of measurement starting points. Exploratory data analysis via low dimensional embeddings includes FPCA and UMAP, which can also be used to cluster different growth dynamics. **c**. GP regression is used to flexibly model growth curves and extract relevant growth parameters, including maximum growth and death rates and Area Under the Curve (AUC) values. Replicate growth curves can be pooled and modeled jointly. **d**. GP regression is combined with permutation-based testing to compare growth curves from different experimental conditions, providing statistical evidence for differential growth.

DGrowthR is a comprehensive R package designed for the analysis of high-throughput growth experiments. As such, it offers functionality for reading multiple data files from automatic plate readers as well as any metadata associated, such as experimental conditions. A commonplace example of such high-volume experiments include chemical library screenings (Figure 1a). The resulting data is stored in a custom object, which conveniently stores all subsequent analysis results. The raw OD measurements can be visually inspected, in order to determine appropriate pre-processing steps (Figure 1a). The DGrowthR package includes common pre-processing functions including the removal of initial timepoints (typically due to noisy measurements), log-transformation of OD measurements, and baseline correction such that every growth curve's initial measurements start at zero. The pre-processed growth curves can be visually inspected (Figure 1b). To get a grasp on the different growth dynamics present in this large volume of data, DGrowthR offers functionality for embedding growth curves into a low-dimensional space using Functional Principal Component Analysis (FPCA) or Uniform Manifold Approximation and Projection (UMAP). Density-based clustering of these embeddings (such as with DBSCAN) clusters growth curves with similar dynamics, and identifies potential outliers (Figure 1b) [14, 15]. The quantitative analysis of growth curves is performed using Gaussian Process (GP) regression, a non-parametric approach that can model complex growth dynamics without making assumptions about the functional form of the growth curves. GP regression is used to model each growth curve, individually or by pooling replicates. The resulting model can directly be used to quantify growth parameters such as maximum growth, area under the curve (AUC), and growth loss. The first derivative of

the model is leveraged to identify maximum growth and death rates as well as doubling time. Finally, the second derivative is used to identify the moments of greatest increase and decrease in the growth rate which, in the case of sigmoidal shaped growth curves, correspond to the end of the lag phase and the start of the stationary phase of growth, respectively (Figure 1c). These growth parameters can also be determined by sampling from the posterior of the resulting growth model. Finally, `DGrowthR` provides a statistical testing framework for comparing growth curves from different experimental conditions, as described by Tonner et al. [11]. This framework is based on permutation testing, where the labels of the growth curves are permuted and the likelihood ratio between the alternative and null models is calculated. To enable accurate and computationally efficient approximation of small p-values, we can limit the number of necessary permutations by approximating the distribution of the test statistics with a Gamma distribution. We adjust the resulting p- values for multiple comparisons using the Benjamini-Hochberg procedure, thus controlling the False Discovery Rate (FDR) to determine the statistical significance of the observed differences in growth dynamics (Figure 1d). This provides an evidence-based approach to compare growth curves that is agnostic to individual growth parameters.

In summary, `DGrowthR` provides a comprehensive toolbox for the end-to-end analysis of high-throughput growth experiments, enabling researchers to gain novel biological insights in a standardized and reproducible manner. We next showcase the capabilities of `DGrowthR` by analyzing three independent datasets, dealing with different bacterial species and their response to various forms of chemical perturbations.

Exploratory data analysis of a large chemical screen with `DGrowthR`

In recent years, there has been a growing body of evidence showing that the growth of bacteria can be significantly altered by exposure to various types of chemical compounds, including non-antibiotic drugs [20–23]. The susceptibility of bacteria to these compounds varies depending, among many other factors, on the species in question [20, 23, 24]. Here, we use an in-house screen of two gram-negative pathogens, *Salmonella enterica* Typhimurium and *Campylobacter jejuni*, against a diverse chemical library of 2,415 compounds from MedChemExpress [23]. This library comprises FDA-approved drugs, metabolites, and food homologous compounds. The growth of bacterial species was monitored using a high-throughput plate reader, capturing OD measurements at regular intervals (Methods). The resulting dataset contains 16,128 growth curves for *S. enterica* and 4,686 for *C. jejuni*, with 17 time points per growth curve in each instance.

Figure 2 summarizes the exploratory analysis of the complete dataset. Both the FPCA and UMAP embeddings immediately highlight the presence of different growth dynamics for both species. In the case of *S. enterica*, `DGrowthR`'s clustering routine identifies three clusters that can be broadly described as growing (cluster 1), non-growing (cluster 2) and intermediate (cluster 0) growth curves (Figure 2c). Accordingly, the non-growing and growing clusters are found at opposite extremes on the axis of the first principal component (FPC-1), with intermediate growers placed between these two groups (Figure 2a). A similar finding is made for *C. jejuni*, where `DGrowthR` identifies five clusters. As before, the main mode of variation separates clusters from non-growing (cluster 4) to growing (cluster 1, 2, and 3), with intermediate curves found between these two groups in the FPCA space (Figure 2d-f).

These findings indicate that exposure to chemical stress gives rise to a variety of the growth of dynamics for *S. enterica* and *C. jejuni*, beyond a simple binary classification of growth or no growth. The effect that these compounds have on the growth of these bacteria is complex and can be further investigated by modeling the growth curves and obtaining relevant growth parameters.

Estimating growth parameters with `DGrowthR`

We next applied `DGrowthR` to model the underlying growth response of *S. enterica* and *C. jejuni* to each chemical challenge. GP regression is leveraged to pool all replicate growth curves that were gathered for the same treatment. After GP model estimation of all pooled growth curves, `DGrowthR` enables the global visualization of the data using the derived growth parameters. For example, Figure 3a and c show the estimated changes in AUC and maximum growth rate with respect to the values obtained for control DMSO treatments for each compound and species, respectively. In the case of *S. enterica*, it can be appreciated

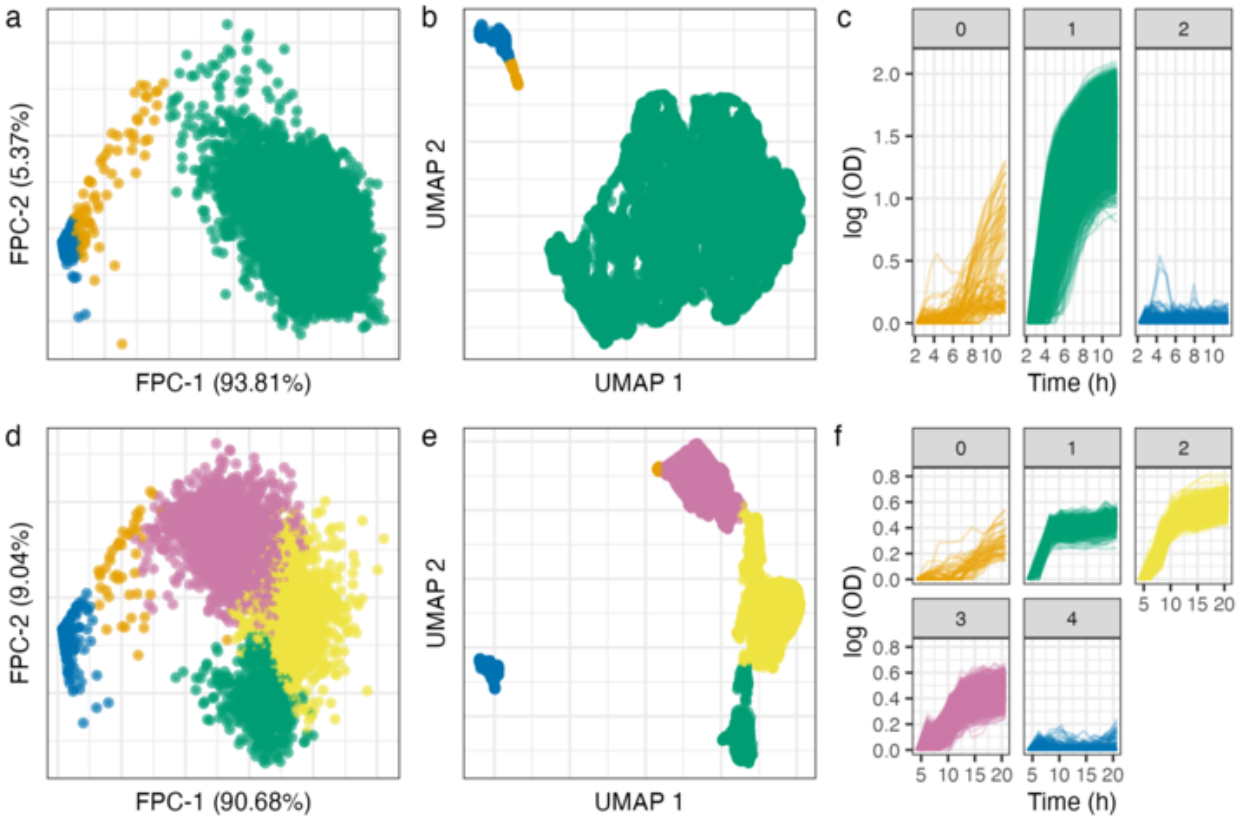

Figure 2: Exploratory analysis of a large chemical screen. **a**. *S. enterica* growth curves embedded in the first two FPC's. The percentage of the total variance explained by each FPC is show in parenthesis. Points are colored based on cluster membership (see panel **c**). **b**. UMAP embedding of *S. enterica* growth curves, colored by cluster membership. UMAP coordinates are used to cluster growth curves with DBSCAN. **c**. Growth curves for *S. enterica* colored by cluster membership. **d**. *C. jejuni* growth curves embedded in the first two FPC's. Cluster membership shown in panel **f**. **e**. UMAP embedding of *C. jejuni* growth curves, colored by cluster membership. UMAP coordinates are used to cluster growth curves with DBSCAN. **f**. Growth curves for *C. jejuni* colored by cluster membership.

that the majority of treatments result in a decrease of AUC and growth rate (Figure 3a). The observed changes correspond to the dynamics shown for each cluster in Figure 2. This can be seen in the case of the intermediate growth dynamics in the presence of the compound Paromomycin, an anti-parasitic drug (Figure 3b).

In contrast, *C. jejuni* shows a more diverse response to the chemical challenges, with some treatments resulting in an increase in AUC and growth rate (Figure 3c). This is exemplified by the antipsychotic drug Droperidol, which resulted in an increased AUC yet a lower growth rate with respect to DMSO (Figure 3d).

`DGrowthR`'s ability to flexibly model the resulting growth dynamics facilitates a holistic interpretable comparison of treatments within and across datasets. In the present case, the analysis of individual growth parameters reveals that the same set of chemical treatments results in pathogen-specific growth parameter profiles.

Statistical comparison and differential analysis of growth curves with `DGrowthR`

While investigating individual growth characteristics and parameters is informative, summarizing and comparing growth curves across multiple growth metrics can be cumbersome for large-scale screens. Moreover,

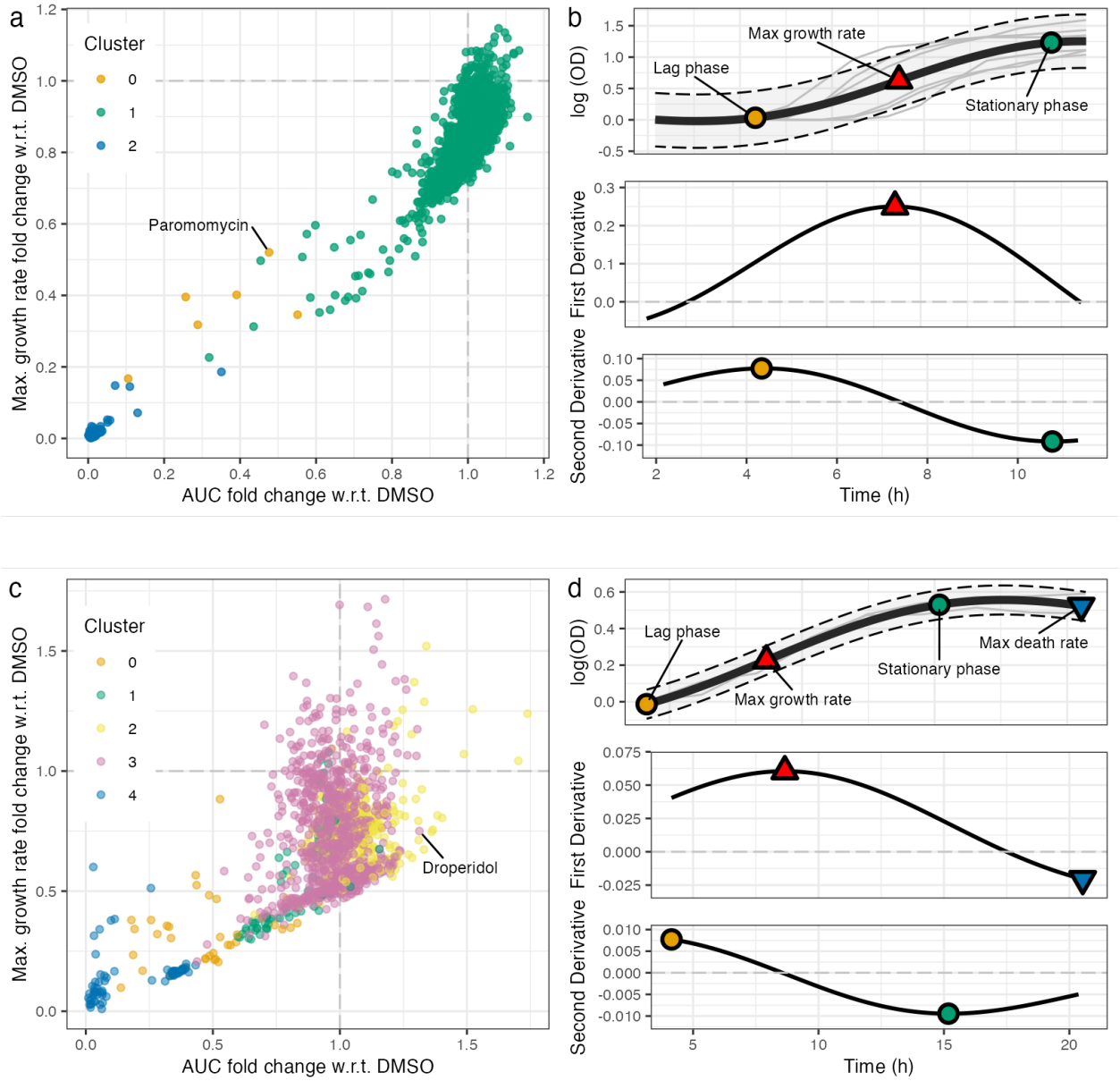

Figure 3: Growth parameters estimated with `DGrowthR`. **a**. All replicate growth curves are pooled and modeled with GP regression. Growth parameters are determined from the resulting underlying model. Fold changes in AUC and maximum growth rate with respect to the values observed for control wells (containing DMSO), are shown for *S. enterica*. Points are colored according to cluster membership from Figure 2. **b**. Growth curve for *S. enterica* treated with Paromomycin. The determination of growth parameters using the first and second derivative is shown. **c**. Fold changes in AUC and maximum growth rate with respect to the values observed for control wells (containing DMSO), are shown for *C. jejuni*. **d**. Growth curve for *C. jejuni* treated with Droperidol.

since the estimation of growth parameters can be sensitive to the experimental setup, the magnitude and scale of the modeled growth curves can wildly differ across experiments [11, 12]. To detect differential growth phenotypes arising due to chemical stress in *S. enterica* and *C. jejuni*, we next illustrate `DGrowthR`'s statistical testing framework that uses Bayes factors to compare growth curves [11]. Briefly, we compare growth curves gathered from exposure to a given compound to those obtained from control conditions (DMSO).

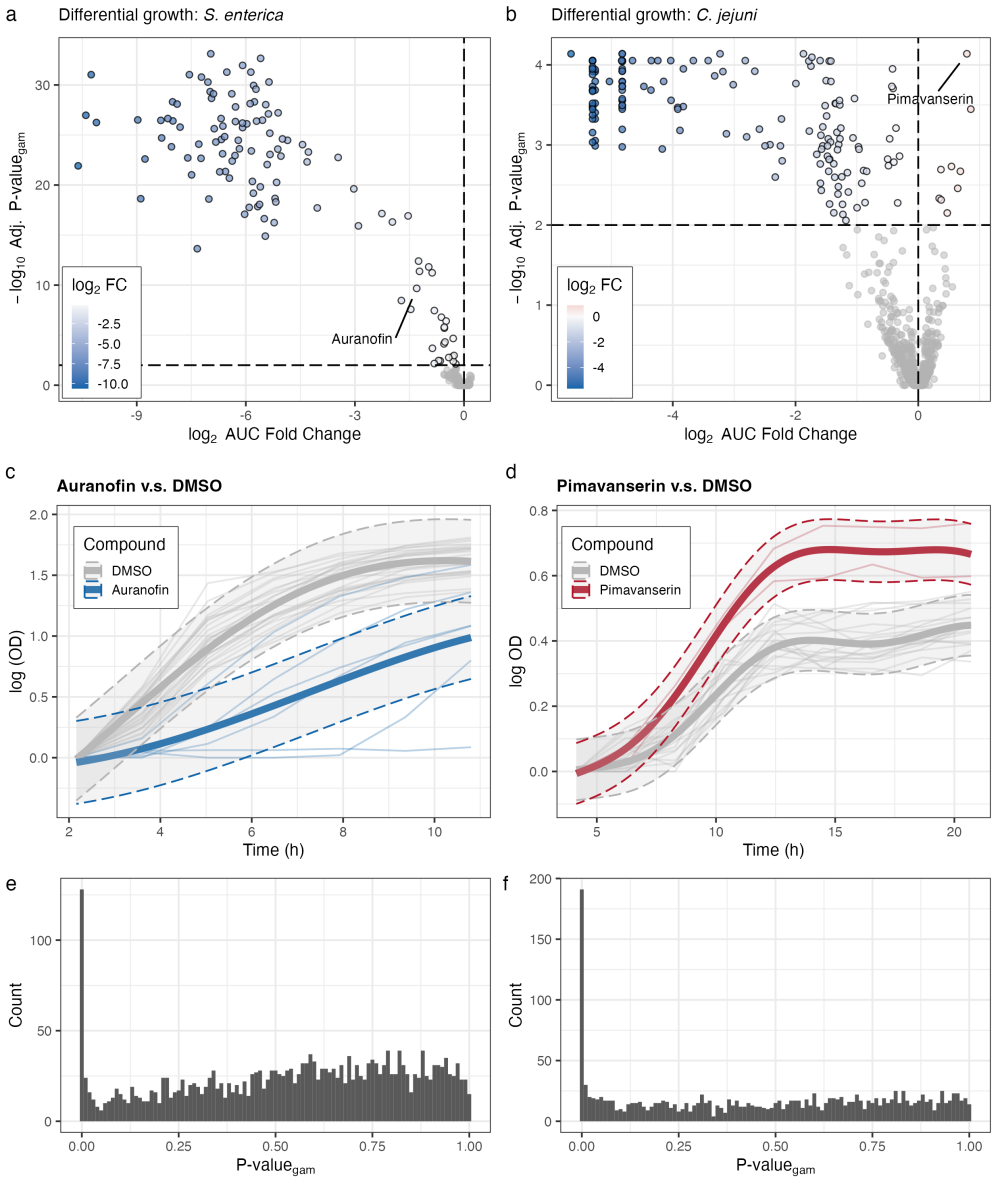

Figure 4: Differential growth analysis with `DGrowthR`. **a**. Volcano plot showing the results of the differential growth analysis for *S. enterica* against 2,415. Corrected p-values are compared to overall changes in the amount of growth (log fold change in AUC). The dashed vertical line indicates a p-value threshold of 0.05, while horizontal lines indicate an absolute AUC log fold change of 0.5. **b**. Volcano plot showing the results of the differential growth analysis for *C. jejuni* against 1,745 compounds. **c**. Alternative model of *S. enterica* growth treated with Auranofin. Shaded area indicates 95% confidence interval of the fitted model. Replicate growth curves are shown. **d**. Alternative model of *C. jejuni* growth treated with Pimavanserin. **e**. Gamma-approximated p-values for *S. enterica*. **f**. Gamma-approximated p-values for *C. jejuni*.

Growth is modeled using two GP regression models: the null model where all growth curves are pooled and modeled only as a function of time, and the alternative model where growth curves are modeled as a function of time and treatment. The likelihood of the data under each model is calculated, and the ratio of the log-likelihoods is the observed Bayes factor. In order to determine the statistical significance of the observed Bayes factor, we permute the treatment labels of the growth curves to obtain a null distribution of the Bayes factor test statistic. A Gamma-approximated p-value is then calculated as the proportion of

permutations that result in a Bayes factor greater than the observed value (see **Methods** for details). In this way, we can determine the statistical significance of the observed differences in growth dynamics by examining complete growth curves, instead of relying on singular growth parameters.

Figure 4a and b illustrate the results of `DGrowthR`'s differential growth analysis for *S. enterica* and *C. jejuni*, respectively, using Volcano plots. In the case of *S. enterica*, 1000 permutations were performed for each treatment. Based on the test statistics obtained, the Gamma-approximated p-values were adjusted for multiple tests using the Benjamini-Hochberg procedure.

Considering a significance level of 0.01 (dashed line in Figure 4a), a total of 117 compounds were found to have a significant effect on the growth of *S. enterica*, all showing decreased overall growth with respect to DMSO. Figure 4c shows the growth curves and the associated GP model for the compound Auranofin, an antirheumatic drug, in comparison to DMSO. While Auranofin does not completely inhibit the growth of *S. enterica*, it gives rise to a growth curve that `DGrowthR`'s testing procedure deems significantly different from the control, with a lower maximum growth rate, extended lag phase, and a lower AUC. For completeness and to show the statistical validity of `DGrowthR`'s testing framework, we show the Gamma-approximated p-value distribution for the *S. enterica* screen in Figure 4e. The p-value distribution shows the typical near-uniform distribution across the $[0, 1]$ interval, with enrichment for small p-values.

We next performed the same analysis for the *C. jejuni* screen using 1000 permutations for each comparison. Notably, while the majority of treatments resulted in a decrease of overall growth, `DGrowthR` identified nine drugs that increase the overall growth of *C. jejuni* (Figure 4b). One such compound is Pimavanserin, an antipsychotic drug. Indeed, Figure 4d, shows the growth curve gathered for *C. jejuni* treated with Pimavanserin shows a higher maximum growth rate and AUC with respect to the control. Finally, the Gamma-approximated p-value distribution for the *C. jejuni* screen is shown in Figure 4f, confirming the validity of the testing procedure for this screen. These findings add to the growing body of evidence that specific drugs can have a significant effect on the growth of pathogenic bacteria. In our in-house large-scale data screen, we found *C. jejuni* to be sensitive to a greater number of these compounds. We observed that the effects on growth can have a complex pattern, beyond simple growth inhibition, yet can be reliably detected using `DGrowthR`. Further studies are required to investigate the mechanisms underlying these growth phenotypes and to determine their clinical relevance.

Exploring the effects of genetic factors on bacterial growth with `DGrowthR`

The variety of growth dynamics present in our chemical screen suggests multiple potential mechanisms of action for the compounds in question. The activity of compounds can be influenced by a variety of factors, including the genetic background of the bacteria in question. Recently, Brenzinger et al. [18] have shown that the activity of antifolate antibiotics can be influenced by the presence of the cyclic-oligonucleotide-based anti-phage signaling system (CBASS) in *Vibrio cholerae*. To showcase the flexibility of `DGrowthR`, we re-analyzed the dataset from Brenzinger et al. [18], where growth in response to chemical stress is measured for wild-type and a CBASS-operon deleted (ΔCBASS) strain of *V. cholerae*. The two strains were screened against a diverse chemical library of 94 compounds at multiple concentrations. In the end, the screen produced 2,304 growth curves. `DGrowthR`'s clustering of the growth curves revealed a variety of growth dynamics present in the screen (see Figure 5a). We pooled all replicate growth curves gathered for the same genotype-compound-concentration combination and obtained growth parameters from the resulting GP models. By comparing the AUC obtained for the wild-type and ΔCBASS strains, we were able to recapitulate the findings made by Brenzinger et al. where the ΔCBASS strain showed decreased susceptibility to Sulfamethoxazole and Trimethoprim, two antifolate antibiotics (Figure 5b) [18]. Additionally, we observe that the ΔCBASS was less susceptible to Amoxicillin, an antibiotic that targets cell wall synthesis.

To showcase `DGrowthR`'s ability to statistically analyze non-standard growth characteristics, we investigated the differences between the wild-type and ΔCBASS strains in terms of *maximum death rates*, as obtained in response to the same chemical challenge. We found that the ΔCBASS strain showed a lower maximum death rate in response to Amoxicillin, as well as other inhibitors of cell wall synthesis in Meropenem and Penicillin G (Figure 5c). These differences in growth parameters can be well observed in the growth curves

obtained for the wild-type and ΔCBASS strains treated with Amoxicillin (Figure 5d).

In summary, our re-analysis of this dataset with `DGrowthR` adds further evidence to the noted influence of the CBASS system on the susceptibility to cell-wall synthesis inhibitor antibiotics, thus providing testable hypotheses for experimental determination of the underlying mechanisms.

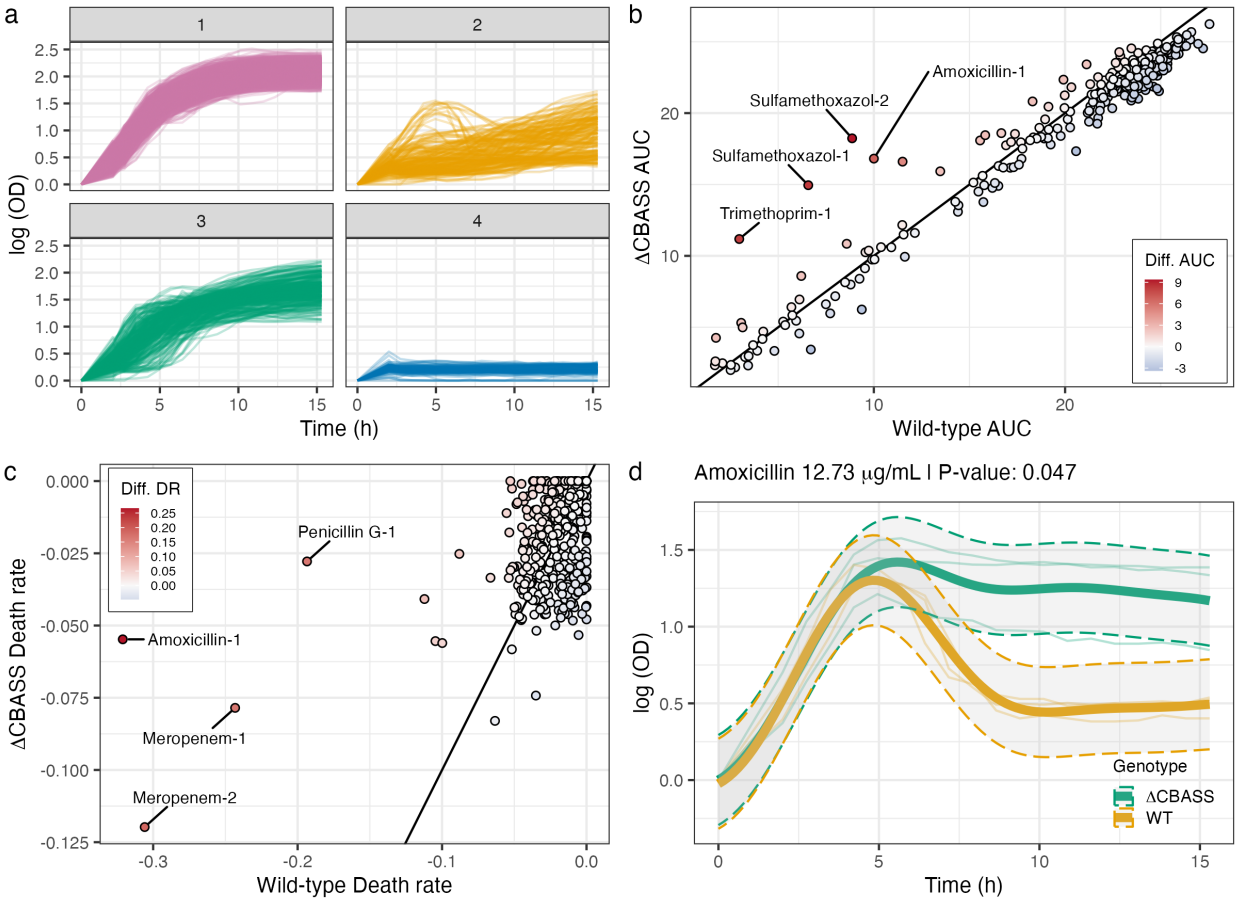

Figure 5: Analyzing the influence of the CBASS anti-phage defense system on antibiotic sensitivity with `DGrowthR`. **a**. Growth curves from the chemical screen done by Brenzinger et al. [18] was embedded and clustered into 4 groups. **b**. Comparison of AUC values determined by `DGrowthR`, for the wild-type and ΔCBASS strains of *V. cholerae* treated with the same compound-concentration combination. **c**. Comparison of maximum death rates determined by `DGrowthR`, for the wild-type and ΔCBASS strains of *V. cholerae*. **d**. Alternative model of *V. cholerae* growth treated with Amoxicillin, comparing the influence of the CBASS systen. Shaded area indicates 95% confidence interval.

Determining combinatorial treatment effects on bacterial growth with `DGrowthR`

Another common application of chemical screenings is to determine the effects of drug combinations on bacterial growth. Non-antibiotic compounds have the potential to act as adjuvants or antagonists of antibiotic activity. In a study by Brochado et al. [19] the effects of several pairwise drug combinations on the growth of gram-negative pathogens was measured. We re-analyzed this dataset using `DGrowthR` to investigate the effects of drug combinations vs. individual treatments on the growth dynamics of *Escherichia coli* BW.

In this study, 4,977 drug combinations were evaluated at multiple concentrations against *E. coli* BW, resulting in a total of 111,668 growth curves. Figure 6a and b show `DGrowthR`'s FPCA and UMAP embeddings of this dataset, respectively, where each growth curve is colored by its maximum OD. Both representations

reveal a strong stratification by maximum OD along the first axis. By exploring the location of treatments in these embeddings we were able to re-discover the synergistic effect of Vanillin with Spectinomycin. Individual treatments of Spectinomycin and Vanillin led to a higher maximum growth compared to the combination of both compounds (Figure 6c in FPCA representation). This is further confirmed by the statistical testing framework provided by `DGrowthR` where the combination of Vanillin and Spectinomycin resulted in a significant decrease in growth compared to Spectinomycin alone (Figure 6d).

Following the same analysis strategy, we were able to detect the antagonistic effect of Caffeine on treatment with Amoxicillin. The combination of Caffeine and Amoxicillin resulted in a higher maximum growth compared to Amoxicillin alone, though not as high as Caffeine alone (Figure 6e). This antagonistic effect was confirmed by the statistical testing framework (Figure 6f). Indeed, the antagonistic effect of Caffeine against antibiotic treatment in *E. coli* was validated and investigated further in an independent follow-up study [25].

In summary, our re-analysis of this dataset demonstrates the utility of `DGrowthR` in detecting complex interactions between antibiotics and secondary compounds, providing valuable testable hypotheses for further investigation.

Discussion

Here, we presented `DGrowthR`, an end-to-end framework for visualizing, modeling, and analyzing large-scale bacterial growth curve data. `DGrowthR` combines state-of-the-art statistical methods, including functional data analysis, Gaussian Process regression, and computationally efficient permutation-based inference schemes, into a comprehensive data analysis framework, enabling robust and reproducible growth data analysis. Using hundreds of thousands of bacterial growth curves from in-house chemical screens and publicly available datasets [18, 19] we illustrated several `DGrowthR` analysis workflows that can serve as templates for future end-to-end analysis of bacterial growth data.

In our in-house chemical screen, `DGrowthR` identified complex growth dynamics to *S. enterica* and *C. jejuni* exposed to different compounds. These findings are consistent with previous studies that have shown the potential of compounds to alter bacterial growth [20–23], and add new information for the studied pathogens. Among the compounds identified, Auranofin was found to significantly reduce the growth of *S. enterica*, despite not leading to complete growth inhibition. This finding is consistent with previous studies that have shown the potential of Auranofin as an antimicrobial agent, with reduced potency against Gram-negative species [26, 27]. In contrast, Pimavanserin was found to increase the growth of *C. jejuni*, which contrasts the reported antibacterial activities of other antipsychotic drugs [21]. The re-analysis of the dataset from Brenzinger et al. [18] with `DGrowthR` confirmed the influence of the CBASS anti-phage defense system on the susceptibility of *V. cholerae* to antifolate antibiotics. With `DGrowthR`, we further identified a decreased death rate in response to cell wall synthesis inhibitors in the ΔCBASS strain, a relationship that can be further investigated to determine the underlying mechanisms. Finally, our `DGrowthR` analysis also reproduced one of the main findings from Brochado et al. [19], uncovering the synergistic effect between Vanillin and Spectinomycin, resulting in increased inhibition of *E. coli* growth. Furthermore, we identified the antagonistic effect of Caffeine against the activity of Amoxicillin, leading to increased growth of *E. coli*. These findings were further validated in an independent follow-up study [25]. The ability of `DGrowthR` to detect complex interactions between antibiotics and secondary compounds is a valuable feature for antimicrobial studies, providing testable hypotheses for further investigation.

In summary, our results demonstrate the utility of `DGrowthR` in analyzing large-scale growth experiments in a diverse set of scenarios, where different research questions are prioritized. In all cases, the `DGrowthR` framework was able to provide novel insights into the growth dynamics of the studied bacteria and to suggest testable hypotheses for further investigation. The successful use of GP regression in `DGrowthR` is consistent with previous work in microbial growth analysis, where GP models have been shown to offer flexibility beyond traditional parametric approaches [10–12]. By providing a GP-based approach, `DGrowthR` also fills a critical gap in R software for microbial growth analysis. Existing R packages such as `gcplyr` and `NeckaR`

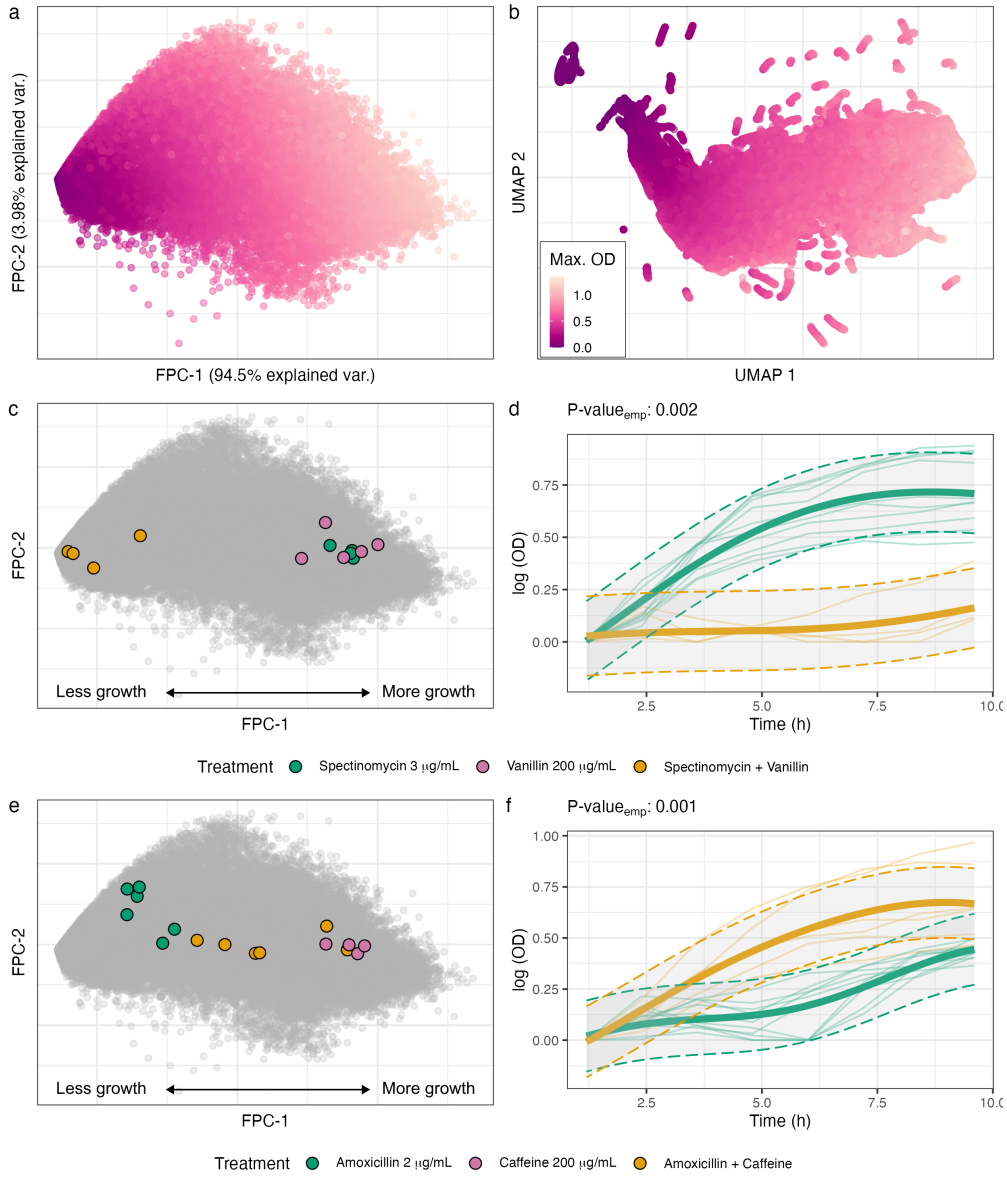

Figure 6: Analyzing the effects of pairwise drug combinations on the growth of *E. coli* BW with `DGrowthR`. **a**. FPCA embedding of the growth curves from the chemical screen done by Brochado et al. [19] against *E. coli* BW. Points are colored by maximum measured OD value. **b**. UMAP embedding of the growth curves from the chemical screen done by Brochado et al. [19] against *E. coli* BW. Points are colored by maximum measured OD value. **c**. Position of replicate growth curves in the UMAP embedding, gather from treatment with Vanillin ($200\mu g/mL$) and Spectinomycin ($3\mu g/mL$) individually individually and in combination. **d**. Alternative model of *E. coli* BW growth treated with Spectinomycin alone and in combination with Vanillin. Shaded area indicates 95% confidence interval. Empirical p-values after 1,000 permutations is shown. **e**. Position of replicate growth curves in the UMAP embedding, gather from treatment with Amoxicillin ($2\mu g/mL$) and Caffeine ($200\mu g/mL$) individually and in combination. **f**. Alternative model of *E. coli* BW growth treated with Amoxicillin alone and in combination with Caffeine. Shaded area indicates 95% confidence interval. Empirical p-values after 1,000 permutations is shown.

are limited in their statistical tests, and packages such as `grofit` can struggle with non-sigmoidal growth patterns. `DGrowthR` introduces GP growth modeling to R, extending the research capabilities available to

the R community and improving the modeling of microbial growth under stress conditions.

While DGrowthR addresses several existing limitations in microbial growth modeling, several challenges remain. Large data sets require extensive permutation testing, which can be computationally intensive and may limit accessibility to users without high performance computing resources. Our approach to include Gamma approximations of p-values are a first step toward reducing the computational burden for the user. Nevertheless, the robustness of permutation tests could be compromised if there is a significant imbalance or only a small sample size available in the condition and control groups. This implies that DGrowthR's hypothesis testing framework, like all permutation-based approaches, requires sufficiently many replicates to be powerful.

In summary, we believe that DGrowthR provides a valuable statistical framework for the analysis of large-scale growth curve data. With the rapidly growing availability of robotics platform-based chemical screen protocols [3], we anticipate DGrowthR to play a substantial role in the robust and reproducible analysis of future large-scale microbial growth curve data.

Methods

Overview of the DGrowthR Framework

DGrowthR is an R package designed for the comprehensive analysis of bacterial growth curves. It efficiently handles high-dimensional data, integrating both growth measurements and metadata. By leveraging advanced statistical techniques, the package enables modeling and interpretation of growth dynamics under diverse experimental conditions.

Data Input and Pre-processing

Data Structuring. DGrowthR supports direct import of data typically output by automated plate readers. The package accommodates multiple input files, aligning them along a common time axis for consistency. All imported data are stored in a structured object that also retains associated experimental metadata. To streamline analysis, DGrowthR utilizes an object-oriented approach, where a specialized object is created to hold both growth curve data and metadata. This structure consists of multiple slots that store intermediate analysis results, minimizing redundant computations and optimizing efficiency.

Pre-processing. Pre-processing is crucial for ensuring data quality and accurate modeling, particularly in large datasets derived from multiple experiments. DGrowthR provides several pre-processing options after data import, including the removal of initial time points, logarithmic transformation, and baseline correction. Early time points in a growth curve often contain technical noise and may not be informative; therefore, DGrowthR allows users to specify the number of initial time points to remove. Logarithmic transformation is applied to optical density (OD) measurements to better model exponential growth, facilitating the estimation of maximum growth rates and the identification of transition phases. Baseline correction addresses instrument fluctuations and non-biological variations in initial OD readings. Since the initial measured population size can vary across experiments, baseline correction ensures consistency by subtracting the initial population size from all subsequent measurements.

Low-Dimensional Embedding and Visualization

DGrowthR is designed for detailed analysis of large-scale data sets, these often contain subtle growth patterns that may not be immediately apparent. To reveal these patterns, we implemented two low-dimensional embedding techniques: Functional Principal Component Analysis (FPCA) and Uniform Manifold Approximation and Projection (UMAP).

Functional Principal Component Analysis. Like Principal Component Analysis (PCA), Functional Principal Component Analysis (FPCA) aims to capture the primary sources of variation within the dataset. This

approach allows users to identify dominant growth patterns within their data [28, 29]. In `DGrowthR`, FPCA calculations for growth curves are implemented using the `fdapace` package [30].

Uniform Manifold Approximation and Projection. UMAP provides non-linear dimensionality reduction, preserving the intrinsic structure of complex datasets while reducing them to two dimensions. This method facilitates visualization and highlights patterns or discrepancies in growth curves that might otherwise remain undetected in high-dimensional data [16].

Clustering. To further explore dataset structure, we implemented Density-Based Spatial Clustering of Applications with Noise (`DBSCAN`) on the UMAP embeddings. `DBSCAN` identifies clusters based on local density and is particularly effective in handling noise and outliers [31, 32]. An adaptation of `DBSCAN` from Herrmann et al. [15] enhances cluster analysis by inferring UMAP's topological structure. In the context of growth curves, clustering can be used to identify groups of curves that exhibit similar growth dynamics. This strategy is implemented in `DGrowthR` using the `dbscan` package [33].

FPCA and UMAP visualizations can be annotated with cluster assignments or metadata categories, enhancing the interpretability of growth patterns under different conditions.

Gaussian Process Regression Modeling

Gaussian Process Regression (GPR) is well-suited for modeling the complex, non-linear, and non-canonical behaviors typical of bacterial growth data [34]. GPR accounts for variability arising from intrinsic biological characteristics and experimental conditions, providing a robust framework for predictive modeling and analysis. Specifically, it defines a probability distribution over functions, enabling flexible modeling without assuming a fixed functional form. This non-parametric Bayesian framework describes a prior distribution over functions, which is updated with observed data to yield posterior predictions.

For a given growth curve, the GP prior is constructed using input time points $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$ and corresponding optical density (OD) values $\mathbf{Y}$. The posterior predictive distribution of OD values at a new input $\mathbf{x}$ is given by:

$$OD(\mathbf{x}) \sim \mathcal{N}\big(\boldsymbol{\mu}(\mathbf{x}), \mathbf{K}(\mathbf{x})\big).$$

Here, $\boldsymbol{\mu}(\mathbf{x})$ represents the predicted OD value at a new input $\mathbf{x}$, while $\mathbf{K}(\mathbf{x})$ provides confidence intervals around the prediction. Specifically, $\mathbf{K}(\mathbf{x})$ is defined by the Gaussian kernel:

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\sum_{m=1}^{M} \frac{(x_{i,m} - x_{j,m})^2}{\theta_m}\right) + \eta \mathbb{I}_{i=j}$$

In `DGrowthR`, GPR is used to model growth curves. The implementation is based on the `laGP` package [35]. Model fitting involves hyperparameter optimization, covariance matrix construction, and OD measurement prediction. Hyperparameters are optimized using gradient-based methods to maximize the log marginal likelihood. The optimized parameters are then used to construct the covariance matrix $\mathbf{K}$, capturing relationships among data points and enabling the model to predict growth at new time points, yielding both mean estimates and uncertainty intervals.

Growth Parameter Estimation and Statistical Inference

Estimation of Growth Parameters. In `DGrowthR`, the modeled response for $OD(x)$ is used to estimate key growth parameters. From the predicted growth values, the maximum optical density ($OD_{\max}$), the area under the curve (AUC), and growth loss are directly derived. Growth rate parameters, including the maximum growth rate ($\alpha_{\max}$), maximum death rate (α_{death}), and doubling time, are estimated using an approximation of the first derivative of the Gaussian Process:

$$OD(\mathbf{x}_t)' \approx \frac{OD(\mathbf{x}_{t+\Delta}) - OD(\mathbf{x}_t)}{\Delta}.$$

Similarly, phase-transition parameters are approximated using the second derivative:

$$OD(\mathbf{x}_t)'' \approx \frac{OD(\mathbf{x}_{t+\Delta})' - OD(\mathbf{x}_t)'}{\Delta}.$$

This second derivative approximation enables the estimation of the time of greatest increase in growth rate (t_{lag}) as well as the time of greatest decrease in growth rate ($t_{\text{stationary}}$).

Differential Growth Testing. A central capability of `DGrowthR` is testing for significant differences in growth curve dynamics across two distinct conditions. This approach builds on the framework introduced by Tonner et al. [11], which employs a Bayes Factor test statistic to compare the posterior likelihood of the data under two competing hypotheses.

The Bayes Factor is defined as:

$$BF = \frac{p(\mathbf{y} \mid H_a)}{p(\mathbf{y} \mid H_0)}$$

where the null hypothesis (H_0) assumes that all growth curves are identical, modeling OD as a function of time. In this case, each input $\mathbf{x}_i \in \mathbf{X}$ is one-dimensional, containing only the time at which the measurement was taken. In contrast, the alternative hypothesis (H_a) accounts for differences between conditions by modeling growth curves as a function of both time and an additional covariate. Here, each input $\mathbf{x}_i$ becomes two-dimensional, including both the time of measurement and a binary variable indicating the experimental condition [36].

In the context of high-throughput testing, variations in prior probabilities and potential deviations from modeling assumptions motivate the use of an empirical p-value for the observed Bayes Factor, which can be obtained through permutations [37]. In this approach, artificial datasets are generated by randomly reassigning covariate labels to the growth curves while preserving the original distribution of the covariate. Importantly, label shuffling is performed at the level of entire growth curves, ensuring that the temporal structure of the data remains intact. The empirical p-value is then computed as the proportion of permuted Bayes Factors that are greater than or equal to the observed Bayes Factor:

$$\text{P-value}_{emp} = \frac{|\mathbf{BF}_{perm} \geq BF_{obs}| + 1}{|\mathbf{BF}_{perm}| + 1}.$$

The addition of 1 to both the numerator and denominator prevents the p-value from being exactly zero when none of the permuted Bayes Factors exceed the observed one. This adjustment yields a conservative and unbiased estimate of the p-value, particularly when the number of permutations is limited [38]. In our case, the number of permutations is restricted to a maximum of 1000 due to computational constraints, resulting in a minimum achievable p-value of approximately 0.001. This resolution is often insufficient to differentiate between highly significant results, as even strong signals will "bottom out" at the minimum possible p-value.

To overcome this limitation, we approximate the distribution of $\mathbf{BF}_{\text{perm}}$ using a Gamma distribution, following the approach proposed by Winkler et al. [17]:

$$\mathbf{BF}_{perm} \sim \Gamma_{perm}(\alpha, \beta),$$

where the rate parameter α and shape the parameter β are estimated via maximum likelihood estimation. The p-value for the observed Bayes Factor is then derived from the fitted Gamma distribution as follows:

$$\text{P-value}_{\text{gam}} = 1 - F_{\Gamma}(BF_{\text{obs}}; \alpha, \beta),$$

where $F_\Gamma(\cdot;\alpha,\beta)$ denotes the cumulative distribution function of the Gamma distribution. To save computation time, this approximation is only applied if the empirical p-value P-value$_{emp}$ is below 0.2, as our primary interest lies in approximating small p-values as accurately as possible.

Computational Framework and Implementation

Predicted Data Access. In addition to performing analyses, `DGrowthR` stores analysis results, including predicted values, permuted Bayes Factors, and OD differences, within the `DGrowthR` object. These values are accessible for further analysis or integration into external workflows, allowing flexibility for custom analyses.

Parallelization and Summary statistics. `DGrowthR` leverages parallel computing to enhance efficiency in processing, result compilation, and parameter extraction. With multicore processors and distributed computing environments, `DGrowthR` can concurrently fit and predict models across multiple compounds, significantly reducing computation time. Analytical results are systematically organized and saved as CSV files. Additionally, `DGrowthR` extracts critical hyperparameters from the Gaussian Process Regression (GPR) fits, such as length scale and nugget parameters, which are essential for assessing model performance and understanding biological implications.

Implementation Details and Software Dependencies. `DGrowthR` is implemented in the R programming language (version 4.2.3).

Data Availability.

A complete workflow for applying `DGrowthR`, including all steps from data pre-processing to analysis and visualization, is available at `https://github.com/bio-datascience/DGrowthR`. This repository provides a detailed tutorial, a comprehensive vignette, and a sample dataset specifically created for `DGrowthR`. All data used in this study, including the results presented here, are openly accessible, ensuring reproducibility and ease of use. A visualization of the entire workflow is also included to facilitate understanding and implementation of the methods. Publicly available growth data from [19] and [18] were obtained from the respective publications.

Code Availability

The open-source code for `DGrowthR` is accessible at `https://github.com/bio-datascience/DGrowthR`. This repository contains the full source code, installation instructions, example scripts, and a visualization of the workflow to assist users in getting started with `DGrowthR`. The code is provided under the MIT License, allowing for broad use and adaptation by the research community.

References

[1] John A Myers, Brandon S Curtis, and Wayne R Curtis. Improving accuracy of cell and chromophore concentration measurements using optical density. *BMC biophysics*, 6:1–16, 2013.

[2] Jacob Beal, Natalie G Farny, Traci Haddock-Angelli, Vinoo Selvarajah, Geoff S Baldwin, Russell Buckley-Taylor, Markus Gershater, Daisuke Kiga, John Marken, Vishal Sanchania, et al. Robust estimation of bacterial cell count from optical density. *Communications biology*, 3(1):512, 2020.

[3] Patrick Müller, Jacobo de la Cuesta-Zuluaga, Michael Kuhn, Maral Baghai Arassi, Tim Treis, Sonja Blasche, Michael Zimmermann, Peer Bork, Kiran Raosaheb Patil, Athanasios Typas, et al. High-throughput anaerobic screening for identifying compounds acting against gut bacteria in monocultures or communities. *Nature Protocols*, 19(3):668–699, 2024.

[4] M. H. Zwietering, I. Jongenburger, F. M. Rombouts, and K. van 't Riet. Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, 56(6):1875–1881, June 1990. ISSN 0099-2240.

[5] M Umut Caglar, Ashley I Teufel, and Claus O Wilke. Sicegar: R package for sigmoidal and double-sigmoidal curve fitting. *PeerJ*, 6:e4251, 2018.

[6] Kathleen Sprouffske and Andreas Wagner. Growthcurver: an r package for obtaining interpretable metrics from microbial growth curves. *BMC bioinformatics*, 17:1–4, 2016.

[7] Matthias Kahm, Guido Hasenbrink, Hella Lichtenberg-Fraté, Jost Ludwig, and Maik Kschischo. Grofit: Fitting biological growth curves. *Nature Precedings*, pages 1–1, 2010.

[8] Michael Blazanin. gcplyr: an r package for microbial growth curve data analysis. *BMC bioinformatics*, 25(1):1–10, 2024.

[9] Nicolas T. Wirth, Jonathan Funk, Stefano Donati, and Pablo I. Nikel. Qurve: user-friendly software for the analysis of biological growth and fluorescence data. *Nature Protocols*, page 1–3, June 2023. ISSN 1750-2799. doi: 10.1038/s41596-023-00850-7.

[10] Peter S Swain, Keiran Stevenson, Allen Leary, Luis F Montano-Gutierrez, Ivan BN Clark, Jackie Vogel, and Teuta Pilizota. Inferring time derivatives including cell growth rates using gaussian processes. *Nature communications*, 7(1):13766, 2016.

[11] Peter D Tonner, Cynthia L Darnell, Barbara E Engelhardt, and Amy K Schmid. Detecting differential growth of microbial populations with gaussian process regression. *Genome research*, 27(2):320–333, 2017.

[12] Firas S Midani, James Collins, and Robert A Britton. Amiga: software for automated analysis of microbial growth assays. *Msystems*, 6(4):10–1128, 2021.

[13] Han Lin Shang. A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98:121–142, 2014.

[14] Moritz Herrmann and Fabian Scheipl. Unsupervised functional data analysis via nonlinear dimension reduction. *arXiv*, December 2020. doi: 10.48550/arXiv.2012.11987. URL `http://arxiv.org/abs/2012.11987`. arXiv:2012.11987 [stat].

[15] Moritz Herrmann and Fabian Scheipl. A geometric perspective on functional outlier detection. *Stats*, 4 (44):971–1011, December 2021. ISSN 2571-905X. doi: 10.3390/stats4040057.

[16] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[17] Anderson M Winkler, Gerard R Ridgway, Gwenaëlle Douaud, Thomas E Nichols, and Stephen M Smith. Faster permutation inference in brain imaging. *Neuroimage*, 141:502–516, 2016.

[18] Susanne Brenzinger, Martina Airoldi, Adewale Joseph Ogunleye, Karl Jugovic, Martin Krähenbühl Amstalden, and Ana Rita Brochado. The vibrio cholerae cbass phage defence system modulates resistance and killing by antifolate antibiotics. *Nature Microbiology*, 9(1):251–262, January 2024. ISSN 2058-5276. doi: 10.1038/s41564-023-01556-y.

[19] Ana Rita Brochado, Anja Telzerow, Jacob Bobonis, Manuel Banzhaf, André Mateus, Joel Selkrig, Emily Huth, Stefan Bassler, Jordi Zamarreño Beas, Matylda Zietek, et al. Species-specific activity of antibacterial drug combinations. *Nature*, 559(7713):259–263, 2018.

[20] Lisa Maier, Mihaela Pruteanu, Michael Kuhn, Georg Zeller, Anja Telzerow, Exene Erin Anderson, Ana Rita Brochado, Keith Conrad Fernandez, Hitomi Dose, Hirotada Mori, et al. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature*, 555(7698):623–628, 2018.

[21] Hassan Nehme, Patrick Saulnier, Alyaa A Ramadan, Viviane Cassisa, Catherine Guillet, Matthieu Eveillard, and Anita Umerska. Antibacterial activity of antipsychotic agents, their association with lipid nanocapsules and its impact on the properties of the nanocarriers and on antibacterial activity. *PloS one*, 13(1):e0189950, 2018.

[22] Veronica J Wallace, Eric G Sakowski, Sarah P Preheim, and Carsten Prasse. Bacteria exposed to antiviral drugs develop antibiotic cross-resistance and unique resistance profiles. *Communications Biology*, 6(1):837, 2023.

[23] Roberto Olayo-Alarcon, Martin K Amstalden, Annamaria Zannoni, Medina Bajramovic, Cynthia M Sharma, Ana Rita Brochado, Mina Rezaei, and Christian L Müller. Pre-trained molecular representations enable antimicrobial discovery. *bioRxiv*, pages 2024–03, 2024.

[24] Yadid M Algavi and Elhanan Borenstein. A data-driven approach for predicting the impact of drugs on the human microbiome. *Nature Communications*, 14(1):3614, 2023.

[25] Christoph Binsfeld, Roberto Olayo-Alarcon, Morgane Wartel, Mara Stadler, Christian Mueller, and Ana Rita Brochado. Systematic characterization of transport regulation in escherichia coli across defined environmental cues. *bioRxiv*, pages 2024–08, 2024.

[26] Shankar Thangamani, Haroon Mohammad, Mostafa FN Abushahba, Tiago JP Sobreira, Victoria E Hedrick, Lake N Paul, and Mohamed N Seleem. Antibacterial activity and mechanism of action of auranofin against multi-drug resistant bacterial pathogens. *Scientific reports*, 6(1):22571, 2016.

[27] Michael B Harbut, Catherine Vilchèze, Xiaozhou Luo, Mary E Hensler, Hui Guo, Baiyuan Yang, Arnab K Chatterjee, Victor Nizet, William R Jacobs Jr, Peter G Schultz, et al. Auranofin exerts broad-spectrum bactericidal activities by targeting thiol-redox homeostasis. *Proceedings of the National Academy of Sciences*, 112(14):4453–4458, 2015.

[28] Michel Loève. Fonctions aléatoires à décomposition orthogonale exponentielle. *La Revue Scientifique*, 84:159–162, 1946.

[29] Kari Karhunen. Zur spektraltheorie stochastischer prozesse. *Ann. Acad. Sci. Fennicae, AI*, 34, 1946.

[30] Cody Carroll, Alvaro Gajardo, Yaqing Chen, Xiongtao Dai, Jianing Fan, Pantelis Z Hadjipantelis, K Han, H Ji, HG Mueller, and Jane-Ling Wang. fdapace: Functional data analysis and empirical dynamics. *R package version 0.5*, 6, 2021.

[31] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.

[32] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. Dbscan revisited, revisited: why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, 42(3):1–21, 2017.

[33] Michael Hahsler, Matthew Piekenbrock, and Derek Doran. dbscan: Fast density-based clustering with r. *Journal of Statistical Software*, 91:1–30, 2019.

[34] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[35] Robert B. Gramacy. laGP: Large-scale spatial modeling via local approximate gaussian processes in R. *Journal of Statistical Software*, 72(1):1–46, 2016. doi: 10.18637/jss.v072.i01.

[36] Harold Jeffreys. *The theory of probability*. OuP Oxford, 1998.

[37] Bertrand Servin and Matthew Stephens. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS genetics*, 3(7):e114, 2007.

[38] Belinda Phipson and Gordon K Smyth. Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *Statistical applications in genetics and molecular biology*, 9(1), 2010.

Acknowledgements

This work was funded by a grant awarded to C.L.M., A.R.B. and C.M.S. for the StressRegNet consortium within the Bavarian research network bayresq.net funded through the Bavarian State Ministry of Science and Arts, Germany.

Author Contributions

M.F., R.O.A. and C.L.M. conceived the overall objectives and design of the project. M.F., R.O.A., C.L.M., and S.P. contributed to the development of the framework. M.F. and R.O.A. analyzed data from experimental validations and implemented all computational methods. A.Z. and M.K.A. performed the experiments. M.F. and R.O.A. drafted the manuscript. All authors revised and approved the final version of the article.

Competing Interests

The authors declare no competing interests.