# Nonparametric Analysis of Temperature and Carbon Source Effects on *vfr* Gene Expression and Regulation

by

Shiqi Gao

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
May 7, 2016

Keywords: Nonparametric Analysis, Gene regulation, *Pseudomonas aeruginosa*



Approved by

Asheber Abebe, Chair, Professor of Mathematics and Statistics
Peng Zeng, Associate Professor of Mathematics and Statistics
Bertram Zinner, Associate Professor of Mathematics and Statistics

Abstract

Since temperatures and carbon sources are potential conditions that affect gene expression and regulation, experiments of *vfr* expression and regulation in the organism *Pseudomonas aeruginosa* were designed and performed to address the following questions of interest: i) Do low or high temperature (30℃/42℃) have an effect on *vfr* expression? ii) Do low or high temperature (30℃/42℃) have an effect on *vfr* regulations controlled by tmRNA, GacA or GacS, which were three potential *vfr* regulators found in the previous laboratory work? iii) Do carbon sources like Glycerol, Glucose or Succinate have an effect on *vfr* expression? And iv) Do these carbon sources have an effect on tmRNA, GacA or GacS controlled *vfr* regulation individually?

As our data (both temperatures and carbon sources) did not satisfy the normality assumption, a set of distribution-free nonparametric analyses were applied in this work. For studying the effects of the temperatures and carbon sources on *vfr* expressions, regulations as well as their variances, powerful and robust nonparametric analysis for location, Nemenyi test, and the Anasari-Bradley test for scale comparison were used to make final conclusions.

According to the Nemenyi test, temperatures (30℃ or 42℃) affected *vfr* expression and but not the tmRNA, GacA or GacS controlled *vfr* regulations in *P.aeruginosa*. In addition, Ansari-Bradley test indicated that, the low temperature would not affect the variances until the late-log phase, while the high temperature affected the variances of *vfr* expression in *P. aeruginosa* from mid-log to late-log phase, but not to the end. Furthermore, high temperature affected all of the variances of tmRNA, GacA and GacS controlled *vfr* expression in mid-log, but only GacS

regulated *vfr* expression in late-log, and only tmRNA dependent expression at the stationary phase in *P. aeruginosa*., while low temperature affected the variances of GacA regulated *vfr* expression at the mid-log phase, and affected the variances of tmRNA and GacS regulated expression at the late-log phase in *P. aeruginosa*, and at the stationary phase, none of the variances of these *vfr* regulations were altered by the low temperature.

In order to test the effects of carbon sources on *vfr* expressions and regulations, similar analyses were applied to the carbon source study. According to Nemenyi approach, *vfr* expressions in *P. aetuginosa* were not be affected by any supply of Glycerol, Glucose or Succinate in NCE medias when they served as sole carbon sources. Also, since all of the tmRNA, GacA and GacS controlled *vfr* regulations functioned very well, *vfr* regulations in *P. aetuginosa* were also not affected in Glycerol, Glucose or Succinate supplemented NCE medias. Therefore, we conclude that carbon sources would affect neither *vfr* expressions nor the tmRNA, GacA and GacS controlled regulations in *P. aetuginosa.* According to the Ansari-Bradley test, the variances of *vfr* expressions in *P. aetuginosa* were not affected by the supply of Glycerol, Glucose or Succinate in NCE medias when they were served as sole carbon sources. The variances of tmRN, GacA, GacS controlled *vfr* regulations were not altered too. In conclusion, supply of Glycerol, Glucose or Succinate in NCE medias as the sole carbon sources would not affect the variances of *vfr* expressions in *P. aetuginosa.* Carbon sources also did not affect tmRNA, GacA and GacS controlled *vfr* regulations.

Acknowledgments

In the first place, I would like to give my heartfelt gratitude to my major advisor, Dr. Asheber Abebe for training me in this thrilling research of statistical application in microbiology for my master degree. I really appreciate his invaluable guidance and helpful suggestions in my research work. I also thank him for his endless support and patience, which gave me a great deal of encouragement and take me through hard times. Without his technical support, ideas and advice, this thesis would not be accomplished.

In addition, I would like to express my special thanks to my other committee members, Dr. Peng Zeng and Dr. Bertram Zinner, for their valuable suggestions and advices. I also would like to thank for my biology advisor Dr. Sang-jin Suh, for my technique training and data collection in the laboratory.

I would like to thank all the professors in Department of Mathematics and Statistics who impart me knowledge in classes. I am also thankful for all the staffs in Auburn who helped me directly and indirectly.

Finally, I would like to express my deepest thanks and gratitude to my family that gave me unconditional love and support. I also thank for their patience throughout my graduate career. I am especially grateful to my husband for support, understanding and love.

Table of Contents

Abstract ..... ii

Acknowledgments ..... iv

List of Tables ..... viii

List of Figures ..... xi

List of Abbreviations ..... xii

Chapter 1: Introduction ..... 1

1.1 Linear regression model ..... 2

1.1.1 General overview ..... 2

1.1.2 Robust linear regression ..... 4

1.2 Nonparametric analysis ..... 5

1.2.1 Tests by scale ..... 6

1.2.1.1 Ansari-Bradley test ..... 6

1.2.2 Tests by location ..... 10

1.2.2.1 Tukey's test ..... 10

1.2.2.2 Nemenyi test ..... 15

1.2.3 Tests by scale and location ..... 18

1.2.3.1 Lepage test ..... 18

1.2.4 Rank Estimation of Linear Models ..... 20

Chapter 2: Experiment design and methods ..... 23

2.1 Literature review ..... 23
2.1.1 *vfr* gene regulation ..... 23
2.2 Temperature and carbon source ..... 24
2.3 Motivation and goal ..... 25
2.4 *vfr* regulators in *Pseudomonas. aeruginosa* ..... 26
2.5 Effects of temperature on *vfr* gene expression and regulation ..... 27
2.6 Effects of carbon source on *vfr* gene expression and regulation ..... 28
2.7 Statistical methods ..... 28
Chapter 3: Results of effects of temperature on *vfr* gene expression and regulation ..... 29
3.1 Overall description ..... 29
3.2 Linear regression ..... 33
3.3 Nonparametric analysis ..... 38
3.3.1 Effects of temperatures on *vfr* gene expression ..... 41
3.3.2 Effects of temperatures on *vfr* gene regulation ..... 47
Chapter 4: Results of effects of carbon source on *vfr* gene expression and regulation ..... 57
4.1 Overall description ..... 57
4.2 Linear regression ..... 65
4.3 Nonparametric analysis ..... 71
4.3.1 Effects of carbon source on *vfr* gene expression ..... 74
4.3.2 Effects of carbon source on *vfr* gene regulation ..... 79
Chapter 5: Discussion and Conclusion ..... 89
5.1 Interpretation ..... 89
5.1.1 Effects of temperature on *vfr* expression ..... 89

5.1.2 Effects of temperature on *vfr* regulation ..... 91

5.1.3 Effects of carbon source on *vfr* expression ..... 94

5.1.4 Effects of carbon source on *vfr* regulation ..... 95

5.2 Limitation and potential problems ..... 99

5.3 Implications of this study ..... 100

5.4 Future work ..... 102

References ..... 102

Appendix 1: Figures ..... 104

Appendix 2: Partial SAS Code ..... 106

List of Tables

Table 1 Descriptive Table for temperature-effect of *vfr* regulation ..... 30

Table 2 ANOVA table for the temperature effect ..... 34

Table 3 R-square of the liner regression for the temperature effect ..... 34

Table 4 The liner regression for the temperature effect ..... 34

Table 5 R-square of the robust regression for the temperature effect ..... 35

Table 6 Parameter Estimates of the robust regression for the temperature effect ..... 36

Table 7 The model of temperatures without interaction by GLM procedure ..... 39

Table 8 The model of temperatures with interaction by GLM procedure ..... 39

Table 9 The model of temperatures with interaction by rank-based GLM procedure ..... 39

Table 10 The full model of temperature in the Drop test ..... 40

Table 11 Drop in Dispersion Test of temperature data ..... 40

Table 12 Tukey's test for 37℃ versus 30℃ / 42℃ comparisons of *vfr* expression ..... 42

Table 13 Robust Tukey's test for 37℃ versus 30℃ / 42℃ comparisons of *vfr* expression ..... 42

Table 14 Rank-based Tukey's test for 37℃ versus 30℃ / 42℃ comparisons of *vfr* expression ..... 43

Table 15 Nemenyi test for 37℃ versus 30℃ / 42℃ comparisons of *vfr* expression ..... 44

Table 16 Ansari-Bradley test for 37°C versus 30°C / 42°C comparisons of vfr expression ..... 45

Table 17 Lepage test for 37℃ versus 30℃ / 42℃ comparisons of *vfr* expression ..... 46

Table 18 Tukey's test for WT versus mutants' comparisons under three temperatures ..... 48

Table 19 Robust Tukey's test for WT versus mutants' comparisons under three temperature ..... 48

Table 20 Rank-based Tukey's test for WT versus mutants' comparisons under three temperatures ..... 49
Table 21 Nemenyi test for WT versus 3 mutants comparisons of *vfr* expression ..... 52
Table 22 Ansari-Bradley test for WT versus mutants' comparisons of *vfr* expression ..... 53
Table 23 Lepage test for WT versus 3 mutants comparisons of *vfr* expression ..... 56
Table 24 Descriptive Table for carbon-source-effect on *vfr* regulation ..... 58
Table 25 ANOVA table for the carbon sources effect ..... 65
Table 26 R-square of the liner regression for the carbon sources effect ..... 66
Table 27 The liner regression for the carbon sources effect ..... 66
Table 28 R-square of the robust regression for the carbon sources effect ..... 67
Table 29 Parameter Estimates of the robust regression for the carbon sources effect ..... 68
Table 30 The model of carbon sources without interaction by GLM procedure ..... 72
Table 31 The model of carbon sources with interaction by GLM procedure ..... 72
Table 32 The model of carbon sources with interaction by rank-based GLM procedure ..... 73
Table 33 The full model of carbon sources in the Drop test ..... 73
Table 34 Drop in Dispersion Test of carbon source data ..... 73
Table 35 Rank-based Tukey's test for LB versus carbon sources added medias' comparisons ..... 75
Table 36 Ansari-Bradley test for vfr expression in LB versus carbon sources added medias' comparisons ..... 77
Table 37 Lepage test for vfr expression in LB versus carbon sources added medias' comparisons ..... 78
Table 38 Rank-based Tukey's test for WT versus mutant comparisons in ten medias ..... 80
Table 39 Nemenyi test for WT versus mutant comparisons in ten medias ..... 83
Table 40 Ansari-Bradley test for WT versus mutant comparisons of *vfr* expression in ten medias ..... 84
Table 41 Lepage test for WT versus mutant comparisons of *vfr* expression in ten medias ..... 87
Table 42 Summary of all nonparametric analyses for 37℃ versus 30℃ / 42℃ comparisons of *vfr* expression ..... 90

Table 43 Summary of all nonparametric analyses for WT versus mutants' comparisons under 3 temperatures ..... 92
Table 44 Summary of all nonparametric analyses for comparisons of LB versus carbon sources added medias ..... 95
Table 45 Summary of all nonparametric analyses for WT versus mutant comparisons in 10 medias ..... 97

List of Figures

Figures 1 Boxplot for *vfr* gene expression under 3 temperatures ..... 32

Figures 2 Boxplot for *vfr* gene expression under 4 strains under all temperatures..... 32

Figures 3 Boxplot for *vfr* gene expression under 3 growth levels under all temperatures ..... 33

Figures 4 Outlier and leverage diagnostics for response variable of the temperature effect..... 36

Figures 5 Leverage diagnostics for response variable of the temperature effect..... 37

Figures 6 Distribution of residuals for response variable of the temperature effect..... 37

Figures 7 QQ-plot of residuals for response variable of the temperature effect..... 38

Figures 8 Boxplot for *vfr* gene expression fed with 10 carbon sources..... 63

Figures 9 Boxplot for *vfr* gene expression of 4 strains in all med ..... 64

Figures 10 Boxplot for *vfr* gene expression under 4 strains in all medias..... 64

Figures 11 Outlier and leverage diagnostics for response variable of the carbon sources effect69

Figures 12 Leverage diagnostics for response variable of the carbon sources effect..... 70

Figures 13 Distribution of residuals for response variable of the carbon sources effect..... 70

Figures 14 QQ-plot of residuals for response variable of the carbon sources effect..... 71

List of Abbreviations

| | |
|---|---|
| LB | L broth |
| *P. aetuginosa* | *Pseudomonas. aeruginosa* |
| GU | University of Georgia |
| HU | University of Hawaii |
| ISU | Idaho State University |

# Chapter 1: Introduction

This is an applied thesis in which novel statistical analysis methods are applied on problems from microbiology. In particular, it uses nonparametric analyses to understand how temperature- and carbon-source- affect *vfr* gene expression and regulation. Since *gacS*, *gacA* and *ssrA* were previously discovered in my work in microbiology to serve as *vfr* gene regulators, a series of experiments were conducted to further the earlier study and address the following questions of interest: i) Does temperature affect the *vfr* gene regulation? ii) If so, which temperature affects gene regulation the most? iii) Does the carbon source have an effect on the *vfr* gene regulation? iv) If so, which sort of the carbon source and concentration has the greatest effect?

Similarly, if we switch these biological questions into statistics, the concerned questions would be: i) Does each population differ among three different temperatures? ii) How about the treatments among three different types of carbon sources with three different concentrations?

In order to attain the answers above, a collection of statistical approaches was applied to model the data and explore the differences. The multivariate data analysis was not constructed in this study. Alternatively, I discuss the effects by temperature and carbon source on *vfr* gene expression and regulation individually. Separating these two effects is the greatest option for the response variable, not only because of the perspective on the study of *vfr* gene regulation in *Pseudomonas aeruginosa*, but also because it is the key for evaluating the effects of temperatures and carbon sources on molecular level respectively.

Since the assumption of normality is not fully satisfied, linear regression analysis is not very reliable in this study. Hence, a batch of nonparametric analysis characterized as distribution-free tests, were imposed in this study. We primarily roughly viewed the relationship between the dependent variable and independent variables via the linear regression. Since the data contains outliers, the robust regression method was subsequently utilized to validate the regression model. Then I applied a batch of nonparametric tests for further analyses, which are divided by scale first, location next, and scale-location last, including the Ansari-Bradley test, Tukey's test, Nemenyi test and Lepage test.

This introduction chapter concentrates on the methodology background of nonparametric analysis, biological context of *vfr* gene regulation, as well as the motivation and goals of this study.

## 1.1 The linear regression model

In statistics, according to various contexts, the term "Linear Model" can be applied to different situations. Generally, it is associated with regression models, whose synonymous form is Linear Regression Model. However, robust regression was developed, since outliers and effective observations were appeared in the dataset. Therefore, Linear Regression model no longer hold the efficiency. Thereby, a brief introduction of general linear regression and robust regression models were given at the first place.

### 1.1.1 General overview

Linear regression analysis is a universal statistical technique that models the functional relationship between variables that have independently and identically normally distributed errors. $Y$ is referred to as the dependent variable that we are going to predict. $X$ is referred to as the independent variable, which the prediction of $Y$ is based on. In other words, least squares linear regression is to forecast the score of a dependent variable $Y$, from scores of one or more independent variables $X$. For a single independent variable, the predicted process is called simple linear regression. In the case of more than one variable, it is named multiple linear regression.

In simple linear regression, we use the "best-fitting" straight line to describe the practical relationship between $X$ and $Y$. The basic criterion for fitting the best-fitting straight line is to minimize the sum of the squared errors of this model via the set of n points. The equation for defining the straight line is expressed as $Y = A + BX$. As mentioned above, $Y$ and $X$ are the dependent variable (ordinate) and independent variable (abscissa), respectively; $A$ is the intercept, and $B$ is the slope of the straight line. When a random collection of observations is given, the population regression line is assessed by the equation $\hat{y} = b_0 + b_1 x$, where $x$ is the independent

variable, $b_0$ is the constant, $b_1$ is the coefficient between $X$ variable and $Y$ variable.

Simple linear regression is recommended when the following conditions are fulfilled:

1) The dependent variable $Y$ and the independent variable $X$ have a linear relationship. When the condition is satisfied, the scatterplot between $Y$ and $X$ should be linear, and residuals are randomly spread in residual plot.

2) For any given value of $X$, distributions of $Y$ are assumed to share the same standard deviation σ with $X$. Under this circumstance, the residuals will have a stable variability across each value of $X$, which can be clearly claimed in a residual plot.

3) For each value of $X$, $Y$ values have an independent and roughly normal distribution. A histogram or a dotplot is widely imposed to check the normality. When the sample size is large, a little skewness is permitted.

Multiple linear regression is a highly flexible technique that estimates the functional relationship between a batch of independent variables $X$ and a single dependent variable $Y$. For p independent variables $x_1, x_2, \ldots, x_p$, the regression line is expressed as $\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$, which indicates how the mean response $\mu_y$ changes in response to independent variables. Each observed value $y$ shifts from their means $\mu_y$, but has the same standard deviation σ. Due to their variations, the model is defined as DATA = FIT + RESIDUAL, where the presentation of "FIT" term is $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$. And the term "RESIDUAL" is the deviations of the each observed value $y$ from their means $\mu_y$ having a normal distribution with mean 0 and variance σ. Consequently, the basic model for linear regression is expressed as $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i$, where has $p$ independent variables and $n$ observations per variables. $X_{ij}$ presents $i^{th}$ observation of the $j^{th}$ independent variable, $Y_i$ presents the response of the $i^{th}$ observation of the dependent variable ($i$ = 1, 2, ... n, $j$ = 1, 2, ..., p). $\beta_j$ is the estimated parameter, $\varepsilon_i$ is the errors with the assumption of independent and identically distributed normal (Leona S. Aiken et al, 2003).

Regression analysis in biological, behavioral and social sciences are broadly applied to predict potential relationships between variables, as its important role in these fields. For instance,

in Beer's law plots, optical density is imposed against drug concentration; also we could characterize the trend or rate via the slope, even the response is unknown (Sanford Bolton et al, 2009).

### 1.1.2 Robust linear regression

Ordinary least squares are considered as one of the most critical tools in the estimation field because of its attractive properties if their underling assumptions are met. Unfortunately, those assumptions are not usually satisfied in reality. At this point, people began to seek approaches as the remedies of this sort of problem. The concept of robust methods was grown at the early nineteenth century, when the electronic technology was developed rapidly. After the 1960s, robust statistics started to impress people. Up to date, it is a fairly popular topic in statistics.

Formally, the outliers, which do not derive from the same data-generation process, are extremely sensitive to the least square estimation (LSE) in the sample. Existence of outliers results in the inefficiency and bias of the least square estimation, particularly when these outliers are arranged and connected with high leverage points. Alternatively, the term Robust Regression came out as a fitting criterion to characterize the information in the majority of data (John Fox, 2002). These techniques are mostly used to handle the following three categories of problems (Xue Bai, 2012):

1) Outliers in response *y*-direction.

2) The data are contaminated with high leverage outliers that both in *x*-space and *y*-direction.

3) Distributions are presented with a long and heavier tail rather than normal ones.

Numerous robust regression methods have been advanced to remedy these problems, including M-estimates, which is designed to use function $\rho(\varepsilon)$ to replace residual square $\varepsilon_i^2$ so that representing the size of the residual in a fewer extreme way. Least median of squares, as well as least trimmed squares, where a more robust measure of location (like the median or a trimmed mean), is a substitute for the mean parameters (George Seber et al, 2003).

There are two ways to measure robustness. The most general measurement parameter is breakdown point (BP), which is the largest fraction of the data that can be moved arbitrarily before the estimator fails against extreme outliers. Hence, the higher the breakdown point, the more robustness the estimator has. Obviously, a breakdown point of sample mean is 1/n, and ½ is for sample median. Apparently, ½ is the maximum value of BP, since if the majority of data are outliers, it is impossible to differentiate the "good" and "bad" observations. However, despite the sample median attains the highest BP value, its efficiency is still very small.

The other popular method is influence function. Let $\hat{\beta}$ be the estimator of $\beta$ of the original data and $\hat{\beta}_0$ be the estimator from the data that excludes all outliers. Then $\hat{\beta}-\hat{\beta}_0$ is defined as the sensitivity curve of $\beta$. Its asymptotic form is the influence function (IF). When the fraction $\varepsilon$ is small, IF is defined by:

$$IF_{\hat{\beta}}(x_0,F) = \lim_{\varepsilon\to 0^+} \frac{\hat{\beta}_\infty((1-\varepsilon)F+\varepsilon\delta_{x_0})-\hat{\beta}_\infty(F)}{\varepsilon},$$

where $x_0$ is the outlier, $\delta_{x_0}$ is the probability measurement which gives mass 1 to {x}, and $\hat{\beta}_\infty(F)$ is the asymptotic value at $F$. This function indicates how much a single outlier influences the estimation. A bounded IF is required as a robust estimator, so that it would not go to infinity when $x$ becomes arbitrarily large (Xue Bai, 2012).

## 1.2 Nonparametric analysis

Statistically, traditional tests are based on a specific probability distribution (such as the normal distribution) rather than a batch of free parameters. The specific probability distribution is involved in parametric tests, which includes estimation of typical parameters, such as the mean, variance, etc. However, unlike parametric tests, nonparametric tests are characterized as distribution-free and do not request any strict distributional assumptions. Even when the data have a normal distribution, nonparametric analysis may be almost as powerful as the parametric method. But if the assumption is not satisfied, nonparametric tests are commonly more powerful than the parametric method (Marco Marozzi, 2013).

The widespread use of nonparametric analysis is mainly for nominal or ordinal data, or the data with a questionable distribution. Because of the property of distribution-free, non-parametric methods are more robust and broadly applied than the corresponding parametric methods. On the other hand, non-parametric methods have the property of simplicity. Under certain situations, non-parametric methods may be simpler to use, even parametric methods are appropriate. Hence, as the nature of simplicity and broader applicability, non-parametric analysis is an appealing option for this work.

### 1.2.1 Tests by scale

Up to date, different non-parametric approaches have been projected to address either two-sample location or scale problem, or for both. The following content will illustrate for scale first, location second, and scale-location last.

#### 1.2.1.1 Ansari-Bradley test

The Ansari-Bradley two-sample test is a distribution-free rank test, which hypothesizes if each scale parameter is equal to each other or not when the two underlying populations are assumed to have the same median (A. R. Ansari, R. A. Bradley, 1960). This test is aimed at testing whether two populations have the common variability or not.

Suppose $F$ and $G$ are two distribution functions corresponding to population $X$ and $Y$, whose sample size are $m$ and $n$ respectively. The null hypothesis of interest is that $X$ and $Y$ populations hold the same probability distribution even this distribution is unknown. Formally, the null hypothesis is expressed as

$$H_0: [F(t) = G(t), \text{ for every } t].$$

Apparently, the alternative hypothesis of the two-sample dispersion problem is that population $X$ has a larger or less variability than population $Y$. Thus, let the two-sample scale model be

$$F(t) = H\left(\frac{t-\theta_1}{\eta_1}\right) \text{ and } G(t) = H\left(\frac{t-\theta_2}{\eta_2}\right), -\infty < t < \infty,$$

Where $H$ ($u$) is a continuous distribution function with median 0, and $F$ ($\theta_1$) = $G$ ($\theta_2$) = $\frac{1}{2}$. Therefore, population medians of X and Y are $\theta_1$ and $\theta_2$, respectively. However, in such a case, we further restrict that $\theta_1 = \theta_2$, which assumes that $X$ population and $Y$ population have the identical median $\theta_1$ and $\theta_2$. Under this assumption, this equal distribution can be simplified as

$$\frac{X-\theta}{\eta_2} \underline{d} \frac{Y-\theta}{\eta_2},$$

where $\theta$ is the shared median. Even if median $\theta_1 \neq \theta_2$, but both are known, each sample can be shifted as $X_1 - \theta_1, \ldots, X_m - \theta_1$ and $Y_1 - \theta_2, \ldots, Y_n - \theta_2$ to satisfy the assumption of the common median $\theta$.

With the assumption above, the ratio of scales: $\gamma = (\eta_1/\eta_2)$ is the parameter of interest. If the variance exists for population $X$ and $Y$, then

$$\gamma^2 = [\frac{Var(X)}{Var(Y)}].$$

In addition, in order to compute the Ansari-Bradley test statistic $C$, values of merged observations $N = m + n$ is ordered from the smallest to the largest. Assign the score 1 is the combination of the least and greatest observation, the score 2 is the combination of the second least and second greatest, and so on. Let $R_j$ be score assigned to $Y_j$ ($j$ = *1,...,n*), and set the test statistic to

$$C = \sum_{j=1}^{n} R_j.$$

For *One-Sided Upper-Tail Test*, the hypotheses are

$$H_0: \gamma^2 = 1 \text{ versus } H_a: \gamma^2 > 1,$$

at the significance level of $\alpha$,

Reject $H_0$ if $C \geq c_\alpha$; otherwise do not reject.

For *One-Sided Lower-Tail Test*, the hypotheses are

$$H_0: \gamma^2 = 1 \text{ versus } H_a: \gamma^2 < 1,$$

at the significance level of $\alpha$,

Reject $H_0$ if $C \leq [c_{1-\alpha} - 1]$; otherwise do not reject.

While for *Two-sided Test*, we test

$$H_0: \gamma^2 = 1 \text{ versus } H_a: \gamma^2 \neq 1,$$

at the significance level of $\alpha$,

Reject $H_0$ if $C \geq c_{\alpha/2}$ or $C \leq [c_{1-\alpha/2} - 1]$; otherwise do not reject.

For approximation of large sample sizes, expected values and variances of $C$ are required to be known when the null hypothesis is true (Myles Hollander et al, 1999). As the group of scores assigned to the jointly ordered population $X$ and $Y$ relies on whether the integer $N = m + n$ is an even or odd number, the discussion of the expected value and variance is split into two parts. When the null hypothesis $H_0$ is true and integer $N$ is even, the expected value and variance of $C$ are

$$E_0(C) = \frac{n(N+2)}{4}$$

and

$$Var_0(C) = \frac{mn(N+2)(N-2)}{48(N-1)},$$

respectively.

When the null hypothesis $H_0$ is true and integer $N$ is odd, the expected value and variance of $C$ are

$$E_0(C) = \frac{n(N+1)}{4N}$$

and

$$Var_0(C) = \frac{mn(N+1)(3+N^2)}{48N^2},$$

respectively.

For overall $N$ (even or odd), the standardized $C$ can be defined as

$$C^* = \frac{C - E_0(C)}{\{Var_0(C)\}^{1/2}},$$

where $C^*$ has a standard normal $N(0,1)$ distribution. Accordingly, the theoretical hypothesis for *One-Sided Upper-Tail Test* is

Reject $H_0$ if $C^* \geq z_\alpha$; otherwise do not reject.

For *One-Sided Lower-Tail Test*, the hypothesis is

Reject $H_0$ if $C^* \leq -z_\alpha$; otherwise do not reject.

Plus, for *Two-Sided Tail Test*, the hypothesis is

Reject $H_0$ if $C^* \geq z_{\alpha/2}$ or $C^* \leq -z_{\alpha/2}$; otherwise do not reject.

1.2.2 Tests by location

Next, several nonparametric methods that are used to test two-sample relationship by location will be introduced.

1.2.2.1 Tukey's test

Tukey's test is a one-step procedure for multiple comparisons in statistics. It is also called Tukey's honestly significant difference test or Tukey's HSD. This method is used after the performance of ANOVA, to determine which groups have a significant difference than others. Some other alternative multiple comparison tests, including Sheffé's test and Dunnett's test, which are merely applied to two groups of observations. However, when the number of groups is greater than two, a one-way analysis of variance (ANOVA) will become the accurate analysis to estimate the possible existence of significant difference among means of each population. If the ANOVA procedure drives to a conclusion that the group means in the sample significantly differ, Tukey's test will come in handy, as it is able to exactly identify which groups of the means are different from the other.

Tukey's test primarily compares the differences between each pair of means expressed as

$$\mu_i - \mu_j,$$

and then investigates whether each pairwise difference of means is greater than the expected standard error. However, in this set, the Tukey approach is conserved when all compared groups have unequal sample sizes. Since only when all sample sizes are equal, the confidence coefficient are exactly $1 - \alpha$. Otherwise, the confidence coefficient is greater than $1 - \alpha$.

The Tukey's method utilizes the *studentized range distribution.* Suppose that we pick a sample size $n$ from $k$ variables with the identically normal distribution $N(\mu, \sigma)$. First, we calculate the absolution value of $\overline{y_i} - \overline{y_j}$, which is the difference between the means of each population. Moreover, the critical difference, also known as half width of confidence interval is computed, defined as

$$\text{half width} = \frac{q_{K;N-K;\alpha}}{\sqrt{2}} S_p \sqrt{\frac{2}{n}}.$$

Afterwards, difference of the absolution value between two means $|\overline{y_i} - \overline{y_j}|$ is compared with the critical half width. If

$$|\overline{y_i} - \overline{y_j}| > \text{half width}, i, j = 1, \ldots, k \text{ and } i \neq j,$$

we say that $\mu_i$ and $\mu_j$ are significantly different, otherwise they are not. Meanwhile, Tukey's method can be employed to construct confidence intervals for all pairwise differences between sample means so that the familywise confidence level is controlled. The typical formula of confidence intervals for differences between pairwise means is

$$(\overline{y_i} - \overline{y_j}) \pm \frac{q_{\kappa;N-\kappa;\alpha}}{\sqrt{2}} S_p \sqrt{\frac{2}{n}}.$$

As mentioned previously, sample sizes have to be equal when the *studentized range distribution* is used (Tukey, John, 1949).

Nevertheless, there is another way to deal with unequal sample size. In 1965, Clyde Kramer stated that the remedy is to compute the estimated standard deviation for each pairwise comparison separately, and this method for unequal sample sizes is also referred as the Tukey–Kramer method. Unlike equal sample sizes mentioned above, the critical difference in this case is

$$\frac{q_{\kappa;N-\kappa;\alpha}}{\sqrt{2}} S_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}.$$

The confidence interval is as follows:

$$(\overline{y_i} - \overline{y_j}) \pm \frac{q_{\kappa;N-\kappa;\alpha}}{\sqrt{2}} S_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}},$$

where $n_i$ and $n_j$ are the sizes from groups *i* and *j* respectively (Clyde Young Kramer, 1956).

Like all post-hoc tests that are conducted after an ANOVA test, the Tukey's test is also faint, because the test has to be performed after the data collection. Compared with the Tukey's test, a preplanned test yielding significant results after data collection, will be more robust and powerful.

1.2.2.3 Nemenyi test

The Nemenyi test (P. B. Nemenyi, 1963), named after Peter Nemenyi, is another case of post-hoc test, like Tukey's test mentioned before (Peter Nemenyi, 1963). It is intended to identify all classifiers with each other after the statistical multiple comparisons. Multiple comparisons are followed to determine whether the median effect of a single, baseline group is different than that of the remaining $k-1$ treatments individually. The treatment versus control associated multiple comparison is upon the average of the joint rank of the whole sample observations *N*. It results in the decisions of the different effects among the control and each $k-1$ treatments. Hence, the test is obviously a one-sided test in practice.

Primarily, we regard treatment 1 as the single baseline control. Furthermore, suppose $N^*$ is the least common multiple of the sample size $n_1, \ldots, n_k$, and then all observations *N* are jointly ranked. First, set $R_1, \ldots, R_k$ as the average of each joint rank coupled with treatments $1, \ldots, k$ separately; then, calculate the differences of each *k*-1 treatments $R_{\cdot u} - R_{\cdot 1}$ ($u = 2, \ldots, k$). In this manner,

$$\text{Decide } \tau_u > \tau_1 \text{ if } N^*(R_{\cdot u} - R_{\cdot 1}) > y_\alpha^*; \text{ otherwise decide } \tau_u = \tau_1,$$

Where the experiment-wise error rate is $\alpha$, and the experiment-wise error equal to $\alpha$ is made by the constant $y_\alpha^*$. To meet this condition,

$$P_0\{N^*(R_{\cdot u} - R_{\cdot 1}) < y_\alpha^*, u = 2, \ldots, k\} = 1 - \alpha,$$

where the probability $P_0(\cdot)$ is processed under $H_0$.

When sample size is large as well as $H_0$ is true, the distribution of the *k*-1 treatment vectors $(R_{\cdot u} - R_{\cdot 1}, R_{\cdot u} - R_{\cdot 1}, \ldots, R_{\cdot u} - R_{\cdot 1})$ is asymptotically $(k-1)$-variate normal, which is like *min* $(n_1, \ldots, n_k)$ tends to infinity. In particular, if $n_1 = b$ and $n_2 = \cdots = n_k = n$, both *n* and *b* values are large, the critical value $y_\alpha^*$ is replaced by $[N(N+1)/12]^{1/2}[(1/b) + (1/n)]^{1/2} N^* m_\alpha^*$. $m_\alpha^*$ is the upper $\alpha$th percentile point, from distribution of $(k-1)$ $N(0,1)$ variables at most, where the common correlation is $\rho = n/(b+n)$. Hence, for equal treatment sample size $n_2 = \cdots = n_k = n$, the large-sample approximation of The Nemenyi test is

$$\text{Decide } \tau_u > \tau_1 \text{ if } N^*(R_{\cdot u} - R_{\cdot 1}) > m_\alpha^* \left[\frac{N(N+1)}{12}\right]^{1/2} \left(\frac{1}{b} + \frac{1}{n}\right)^{1/2};$$

$$\text{otherwise decide } \tau_u = \tau_1, u = 2, \ldots, k.$$

Unlike equal treatments sample sizes, for arbitrary case of unequal treatments sample sizes, Bonferroni's Inequality is used by Dunn (Dunn, 1964) to process the large-sample approximation that

$$\text{Decide } \tau_u > \tau_1 \text{ if } N^*(R_{\cdot u} - R_{\cdot 1}) > z_\alpha^* \left[\frac{N(N+1)}{12}\right]^{1/2} \left(\frac{1}{n_1} + \frac{1}{n_u}\right)^{1/2};$$

otherwise decide $\tau_u = \tau_1$, $u = 2,\ldots, k$,

where $\alpha^* = \alpha/(k-1)$.

In conclusion, the Nemenyi test is analogous to Tukey's test, and is imposed to compare all classifiers with each other. Classifiers will have a significant difference if the corresponding average ranks differ by the critical values.

### 1.2.3 Tests by scale and location

In the last part of this section, a test with a combination of location and scale tests will be described.

#### 1.2.3.1 Lepage test

As mentioned above, -Mann-Whitney test, raised by Wilcoxon (Mann, Henry B., Whitney, Donald R. 1947), is one of the most popular two-sample rank methods for testing equality of location parameters of two underlying populations. The Ansari-Bradley test, suggested by (A. R. Ansari, R. A. Bradley, 1960), is characterized as the test to identify the equality of the scale parameters with the equal medians by means of a rank test. However, when the location remains constant, the Wilcoxon-Mann-Whitney test cannot reflect the changes in the scale parameters. Alternatively, when the scale remains constant, the Ansari-Bradley test cannot reflect changes in the location parameters. The Lepage test, in terms of the combination of Wilcoxon rank-sum and the Ansari–Bradley statistics, will be the solution for both problems. The Lepage test is designed to test the hypothesis of equality of both location and scale parameters of two populations versus the alternative that at least for one of the parameters the equality does not hold (Yves Legape, 1973).

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be the random samples from population 1 and 2 separately with the assumptions of independence within and between each population. The assessment of interest is whether either the location parameters $\theta_1$ and $\theta_2$ or the scale parameters $\eta_1$ and $\eta_2$ of each population differ or not. Therefore, according to the statement that both the location and scale parameters are equal, the null hypothesis of interest is

$$H_0: \theta_1 = \theta_2 \; and \; \eta_1 = \eta_2,$$

versus the alternative hypothesis

$$H_0: \theta_1 \neq \theta_2 \; and/or \; \eta_1 \neq \eta_2.$$

In order to calculate the Lepage two-sample location-scale statistic *D*, firstly, rank all the united observations $N = m + n$ *X*-values *Y*-values from the smallest to greatest. Let $W = \sum_{j=1}^{n} S_j$ be the Wilcoxon rank sum statistic, where $S_j$ is the joint rank of $Y_j$ $(j = 1, \dots, n)$. Furthermore, let $C = \sum_{j=1}^{n} R_j$ be the Ansari-Bradley scale test statistic, where $R_j$ is the score assigned to $Y_j$ $(j = 1, \dots, n)$. Thus, the Lepage rank statistic *D* is defined as follows,

$$D = \frac{[W - E_0(W)]^2}{var_0(W)} + \frac{[C - E_0(C)]^2}{var_0(C)},$$

where $E_0(W)$, $E_0(C)$, $var_0(W)$ and $var_0(C)$ are expected values and variances of *W* and *C*, respectively, under $H_0$. When this formula is substituted by the standardized forms of the Wilcoxon rank-sum statistic and Ansari-Bradley scale statistic, respectively, the Lepage statistic *D* can be expressed as

$$D = (W^*)^2 + (C^*)^2.$$

At the significant level of $\alpha$,

Reject $H_0$ if $D \geq d_\alpha$; otherwise do not reject,

where $d_\alpha$ makes the type I error probability equal to $\alpha$.

When the populations have large sample sizes, the Lepage statistic *D* has a chi-square distribution with two degrees of freedom under $H_0$. Hence, the large-sample approximation for the accurate level $\alpha$ is written by

Reject $H_0$ if $D \geq \chi^2_{2,\alpha}$; otherwise do not reject;

Where $\chi^2_{2,\alpha}$ is the upper percentile of the chi-square distribution whose degrees of freedom is two.

### 1.2.4 Rank Estimation of Linear Models

The rank comparison of the two-sample location (Wilcoxon-Mann-Whitney) and multiple sample ANOVA are special cases of a general linear hypotheses tests in the linear model framework. A general linear model relates a set of p predictors (**X**) collected on n subjects to their response (Y) using a plane. For subject k, this is given by

$$Y_k = \alpha + \beta_1 X_{k1} + \cdots + \beta_p X_{kp} + \varepsilon_k$$

for $k = 1, \ldots, n$. This is usually written in matrix form as

$$\boldsymbol{Y} = \alpha \mathbf{1}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{Y}$ is an $n\times1$ vector of responses, $\boldsymbol{X}$ is an $n\times p$ matrix of predictors, $\boldsymbol{\varepsilon}$ is an $n\times1$ vector of random errors, and $\mathbf{1}_n$ is an $n\times1$ vector of ones. Once again, our main interest is to estimate and test hypotheses about the $p\times1$ vector of slope parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$**.**$\boldsymbol{\beta}$ One way to achieve this is by $\boldsymbol{\beta}$determining the asymptotic distribution of the estimator of $\boldsymbol{\beta}$. Classically $\boldsymbol{\beta}$ is estimated by the method of least squares (minimizing the Euclidean norm of the residuals). The resulting estimator is

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}.$$

If the errors $\varepsilon_1, \ldots, \varepsilon_n$ are iid from a distribution $F$ that has variance $\sigma_F^2$, then $\widehat{\boldsymbol{\beta}}$ follows an approximate $p$-dimensional Gaussian distribution with mean $\boldsymbol{\beta}$ and covariance matrix $\sigma_F^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$. One may use this asymptotic distribution to construct a Wald-type test of the significance. For instance the two-sample t-test is equivalent to a test of a single slope parameter$X_j$ $H_0: \beta_j = 0$ versus $H_A: \beta_j \neq 0$ using the statistic

$$T_j = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}_{jj}}}$$

where $(\boldsymbol{X}^T\boldsymbol{X})^{-1}_{jj}$ is the *(j,j)*th entry of the matrix $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ and $\hat{\sigma}^2 = \frac{(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})^T(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})}{n-p-1}$ is the estimate of the model variance. T$T_j n-p-1$he level $\gamma$ test of significance is performed by comparing $|T_j|$ with upper $\gamma/2$ percentile of the t distribution with $n-p-1$ degrees of freedom. The ANOVA F test may also be constructed as a special case of linear model testing.

Heiler and Willers (1988) have shown that the $\widetilde{\boldsymbol{\beta}}$ follows an asymptotic $p$ dimensional Gaussian distribution given by $N\left(\boldsymbol{\beta}, \tau_\varphi^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\right)$, where $\tau_\varphi^2$ is defined as

$$\tau_\varphi^2 = \int_0^1 \varphi(u)\left\{-\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}\right\}du$$

and represents a scale parameter that is analogous to the standard deviation in least squares estimation. The quantity $\tau_\varphi^2$ reduces to $\tau_F^2$ defined above for the linear score case given by $\varphi(u) = \sqrt{12}(u-1/2)$. A consistent estimator of $\tau_\varphi^2$ is given in Koul, Sievers and McKean (1987). We can use this estimator $\tilde{\tau}^2$ of $\tau_\varphi^2$ along with the asymptotic distribution to construct test statistics for testing various types of hypotheses. For instance, a Wald type *t* test for the significance of the *j*th individual slope, $1 \le j \le p$, can be as

$$W_j = \frac{\tilde{\beta}_j}{\sqrt{\tilde{\tau}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}_{jj}}}$$

and the null hypothesis $H_0: \beta_j = 0$ is rejected in favor of $H_A: \beta_j \neq 0$ if $|T_j| > t_{n-p-1}(\gamma/2)$, $t_{n-p-1}(\gamma/2)$ is the upper $\gamma/2$ percentile of the $t$ distribution with $n-p-1$ degrees of freedom. This is the asymptotic version of the Wilcoxon-Mann-Whitney test when the problem is a two sample comparison problem. Notice that the asymptotic distributions of the least squares estimator $\widehat{\boldsymbol{\beta}}$ and the rank estimator $\widetilde{\boldsymbol{\beta}}$ differ only in their scale parameters $\sigma_F^2$ and $\tau_\varphi^2$,

respectively. It is, thus, obvious that the ARE of the rank estimator with respect to the least squares estimator is

$$ARE(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}) = \frac{\sigma_F^2}{\tau_\varphi^2}.$$

The fitting of this is implemented in the R package Rfit (Kloke and McKean, 2012). This setup allows us to perform several rank-based post-hoc comparisons, like the Tukey-Kramer based on ranks, directly (Kloke and McKean, 2012).

# Chapter 2: Background, Design, and Methods

## 2.1 Literature review

This section principally introduces the opportunistic human pathogen termed *Pseudomonas aeruginosa* and the regulation of the *vfr* gene controlling expressions of virulence factors. In addition to the *vfr* gene regulation in *Pseudomonas aeruginosa,* potential effects of temperatures and carbon sources will be demonstrated as well.

### 2.1.1 *vfr* gene regulation

The gram-negative bacterium *Pseudomonas aeruginosa* is a human pathogen that can trigger infection in immunosuppressed people or colonize in the lungs of patients who suffer from a genetic disorder: cystic fibrosis, caused by mutations in the CF-transmembrane conductance regulator (Hoiby, 1974, Reynolds, 1975, Hoiby, 1995). The versatility of this ubiquitous bacterium depends on a large genome size as well as numerous transcriptional regulators (Stover et al., 2000), which enable this organism to live almost everywhere and withstand severe environmental challenges.

Vfr (virulence factor regulator), encoded by the *vfr* gene, is a global transcriptional regulator of gene expression that affects expression of over 100 genes in the pathogen *Pseudomonas aeruginosa*, including virulence genes encoding virulence factors (wolfgang et al., 2003, west et al., 1994).

Vfr, the *vfr* gene product, which belongs to the winged-helix family, is a cAMP binding protein that is 67% identical and 91% similar to the cAMP receptor protein (CRP) of Escherichia coli (West et al., 1994). In *E. coli*, the level of cAMP mediates CRP functions that characterize as a variety of regulators (Aiba et al., 1982, 11). Primarily, CRP controls some genes involved in catabolic repression by glucose (Botsford & Harman, 1992, Kolb et al., 1993, Ullmann & Danchin, 1983). Likewise, CRP serves as an activator for lactose and arabinose operon, and a repressor in genes for adenylate cyclase (Majerfeld et al., 1981). However, unlike CRP, Vfr failed to present catabolic repression control (Suh, 2002).

To date, Vfr has been stated to regulate many important factors, including Exotoxin A and protease IV (West et al., 1994), two quorum sensing regulators LasR (Albus et al., 1997) and RhlR (Medina et al., 2003), twitching motility (Beatson et al., 2002), type III secretion (Wolfgang et al., 2003), flagellar (Dasgupta et al., 2002), and RpoS (Bertani et al., 2003). Although much study has been conducted to understand the role of Vfr in *P. aeruginosa* pathogenesis, not much is known about the genetic regulation of *vfr* expression. In our laboratory, we have already discovered that tmRNA negatively regulates *vfr* expression (Wu and Suh, unpublished data). We are also interested in exploring any other putative regulators of expression of the *vfr* gene in *P. aeruginosa*.

## 2.2 Temperature and carbon source

During the exponential phase of the growth, once microorganisms start to consume the preferred carbon source, repression of some specific genes, encoding the catabolic enzymes required for assimilation of the subsequent carbon source, will be presented, even though the substrate of the subsequent carbon source is always available in the culture. This process is known as catabolite repression control (CRC), analogous to the exhaustion of the preferred carbon source (Magasanik, 1961). This phenomenon definitely controls the synthesis of

catabolic enzymes, which could be induced by its substrate and repressed by its product or products of catabolic pathways.

Some catabolic enzymes are subject to repression by glucose. For instance, the most outstanding Gram-negative enteric bacteria *Escherichia coil*, prefer glucose as the carbon source (D.N. Collier et al, 1996). Unlike *E. coli*, my target organism *P. aeruginosa* actually does not utilize glucose as the preferred carbon source (Wolff, J. A. et. al, 1991). Alternatively, succinate is preferentially utilized over other carbon sources in this bacterium (SJ Suh et. al, 1999). Therefore, effects of various carbon sources on the regulation of my target gene *vfr* are valuable and worthy of study.

In addition to different carbon sources, bacteria growth is affected by temperature as well. Theoretically, bacteria can be cultured at all temperatures from the freezing point of water to the temperature higher than 100 °C at large depths of the ocean. Most bacteria are mesophilic, which grow best at 30-37°C. Optimum temperature for growth of common pathogenic bacteria including *P. aeruginosa* is 37°C. The optimum temperature in which the bacterium thrives lies somewhere between these maximum and minimum points. When temperatures are below the minimum extremity, bacterial growth will be ceased. However, if the temperature is approaching above the maximum, bacteria will be killed rapidly. In the view of the critical effects of the temperature in bacterium, alerted temperatures are studied in my research.

## 2.3 Motivation and goal

According to the literature view, I hypothesize that there are some putative regulators controlling *vfr* expression. In order to achieve my goals, firstly, I will isolate and characterize putative regulators of *vfr* gene expression. After several rounds of selection and screen, *gacS* was found that served as the potential *vfr* gene regulator. This gene encodes for the sensor kinase in the GacS/GacA two-component regulatory system, while the other gene *gacA* in this system encodes for a response regulator. Theoretically, *GacS* could not function as the regulator of gene expression, since its role is just a sensor kinase. For this reason, we hypothesize that instead of *gacS*, regulation of *vfr* gene expression is actually depending on *gacA*. Subsequently,

the result that both *gacS* and *gacA* mutants that have alerted *vfr* gene expression, verified that our hypothesis is accepted. Thus, either *gacS* mutant or *gacA* mutant will be involved in the following work, as well as the *ssrA* mutant. In brief, the experimental subjects of interest are 4 strains in total. One is the wild type strain, which is the control of the normal *vfr* gene expression, and the remaining three strains are mutants with three mutated *vfr* regulator genes (*gacS*, *gacA* and *ssrA*). In this case, we have four particular target questions we need to address, which are i) Does temperature affect the *vfr* gene regulation? ii) If so, which temperature has the most powerful effect? iii) Does carbon source affect the *vfr* gene regulation? iv) If so, which sort of the carbon source and its concentration affects the *vfr* gene regulation most?

In order to remedy the first two questions, *vfr* gene expression among four strains at three different growth levels, were tested at three temperature gradients including 30 ℃, 37 ℃ and 42 ℃, and 37℃ is the optimal temperature of *P. aeruginosa.* To address the last two questions, *vfr* gene expression among four strains at three different growth levels were tested when bacterium were fed in the L Broth culture supplemented with three different concentration-gradients of three different types of carbon sources. They are glucose, succinate and glycerol. And each observation of three mutants above is compared with the wild type strain at the same growth level in the identical temperature or media. Later, several statistical approaches were utilized to fit the models of data and clarify the variability.

Since the dataset has an unknown distribution probability, a series of nonparametric tests will be applied to analyze the prospective effects of temperature and carbon sources on *vfr* gene regulation. When the condition of the applied test is just restricted to two groups rather than more groups, the dataset will be split into several subsets. According to the statistical output, biological significances will be given to answer these core questions above.

## 2.4 Exploration of *vfr* regulators in *Pseudomonas aeruginosa*

In order to discover and isolate putative regulators of *vfr* gene expression, we took a non-predictable approach to discover *P. aeruginosa* mutants with altered *vfr*::*lacZ* expression. A mini-transposon, Tn5-B30 (Lorenzo et al., 1990), was introduced into the *PAO1* derivative strain

carrying a *vfr*::*lacZ* fusion via triparental conjugation (Leong et al., 1982). Transposon insertion mutants were selected as tetracycline resistant ($Tc^r$) colonies growing at 37°C on Pseudomonas Isolation Agar (PIA) plus tetracycline. These $Tc^r$ colonies were subsequently replicated onto PIA plates supplemented with the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and the insertion-mutants with altered *vfr*::*lacZ* expression were identified as potential regulatory mutants. These replicated colonies displaying altered expressions of *vfr*::*lacZ* with different degrees of blue color will be picked and patched onto fresh PIA with carbenicillin plates as potential candidates to have transposon insertions in genes encoding for regulators of *vfr* expression. After two rounds of screen on plates, putative mutants were rescreened in 96-well plate β-galactosidase assays (Kevin et al., 2002) In brief, we cultivated cells in 96-well micro titer plates in LB broth supplemented with carbenicillin per well (for maintenance of *vfr*::*lacZ*) overnight at 37°C. The next day, 96-well plates β-galactosidase assay will be executed as described by Miller (Miller et al., 1972) to measure *vfr*::*lacZ* expression in the putative mutants. For further accuracy, those potential mutants that consistently presented altered *vfrA*::*lacZ* expression on micro titer plate assays will be selected and analyzed in test tube assays.

In order to test *vfr*::*lacZ* expression during late stationary phase in test tube β-galactosidase assays, putative *vfr* regulator mutants will be cultivated in LB broth containing carbenicillin overnight. In the following days, the β-galactosidase activity of stored cell culture will be assayed to determine the expression of *vfr*::*lacZ*. Each putative regulatory mutant will be growing at 37 °C and assayed at least three independent times to validate the expression of the *vfr*::*lacZ* fusion phenotype.

Arbitrary PCR was executed to classify transposon insertion sites in each mutant. In the first round of arbitrary PCR, Tn5Ext and the other arbitrary primer at a random site downstream were involved to improve random binding. At the second round, the 100-fold-diluted product of the first round will be served as the template, working with Tn5Int and the other nested downstream primer to amplify the specific Tn5 insertion-fragment. Later, all of the amplified arbitrary PCR products were sequenced for DNA mapping. From the sequence results, we found the transposon was inserted in the *gacS*, so combined with the previous discussion, we would study the regulation of three *vfr* regulators encoded by the three genes *gacS*, *gacA* and *ssrA*.

2.5 Growth curve assays to assess *vfr* expression and regulation

For performance of growth curve assays, bacterial cells grew in the applicable medium in 200ml flasks for 24 hours within a shaker at the assigned temperatures. For instance, for analysis of the effects of temperature on *vfr* regulation, bacterial cells were cultured at 30°C, 37°C and 42°C, respectively. For the analysis of effects of carbon source on *vfr* regulation, the growing temperature was 37°C consistently. To make the culture thrive, the overnight culture of cells cultivated in L broth was diluted 1:100 into the proper medium and the cells were incubated for additional 14 hours via shaking at the original temperature. Cell growth was examined as absorbance 600.

2.6 Effects of temperature and carbon source on *vfr* gene expression and regulation

The effect of temperature on *vfr* gene expression and regulation of a *gacA*, *gacS* and *ssrA* mutants were tested by raising *P. aeruginosa* in L broth with carbenicillin at 30°C, 37°C, or 42°C, respectively. The effect of the carbon sources were tested by subculturing the overnight cultures in L broth as well as the NCE minimal medium supplemented with 1mM glucose, 2mM glucose, 10 mM glucose, 2mM succinate, 4mM succinate, 20mM succinate, 1% glycerol, 2% glycerol, or 10% glycerol, respectively, as the unique carbon source. Each culture growing within various media or temperatures was collected at three gradually higher growth levels: middle log phase, late log phase and stationary phase. All of collected cultures were tested by tube β-galactosidase assays to gain the amount of *vfr* expression.

2.7 Statistical methods

The two datasets, temperature-data and carbon-source-data, were analyzed by several methods. The relationships between the dependent variable (the amount of *vfr* expression) and all independent variables were first roughly estimated by the linear regression model after naming all dummy variables. In addition, robust regression approach was used to fit the enhanced model

when outliers appeared in the data.

As the core methods in this study, a set of nonparametric tests was utilized. Tukey's test for multiple comparisons came after performing ANOVA. The Ansari-Bradley test and the Lepage test were conducted by R for two-sample scale and scale-location analysis, respectively.

## Chapter 3: Results of effects of temperature on *vfr* gene expression and regulation

This chapter provided selective results of several analyses related to effects of temperature on *vfr* gene expression and regulation.

The total sample size for temperature set was 144 observations assigned to 4 strains in 3 temperatures in 3 growth levels, and each group had 4 observations collected from two independent times. The total sample size for carbon source set was 288 observations assigned to 4 strains in 10 medias in 3 growth levels. Since the total amount of compared carbon source supplemented NCE medias are too much to collect all data in one time, data collection were completed from three times, each time just contained one sort of three concentrated carbon source added NCE medias. And L broth, which was the controlled media, was repeated in each time's collection. Hence, for strains grown in NCE media supplemented Glucose, Succinate and Glycerol as the sole carbon sources, each group has 6 observations collected from three growth levels. For strains grown in L broth, each group has 18 observations collected in three repeated times.

3.1 Overall description

The descriptive statistics were shown in Table 1 as followed, including names of strains, temperatures and growth levels, median of the amount of *vfr* gene expression, mean +/- standard error, minimum/maximum by 36 groups. Also, the data plot classified by time*temperature*strain treatment was presented in Figure 1 Appendix 1. Apparently, the largest median of *vfr* expression was produced by *ssrA* mutant in stationary phase at 30℃. The smallest median of *vfr* expression was generated via the wild type strain in middle-log phase at 42℃. In addition, the boxplots in Figure 1, 2 and 3 gave more visualized details. When strains were inoculated under three different temperatures, *vfr* expression under 42℃ had the smallest median and mean while 30℃ had the largest ones and 37℃ was in the middle. Among three different growth levels, the stationary phase had the greatest mean and median and the middle-log phase had the least. Among four strains, the order of the means and medians of *vfr* expression in four strains was *ssrA* mutant > *gacA* mutant > *gacS* mutant > wild type, indicating that mutations of *ssrA*, *gacS* and *gacA* genes will induce *vfr* expression in *Pseudomonas aeruginosa*. Corresponding to the previous results, the amplified *vfr* gene expressions in three mutants in all temperatures showed that *ssrA*, *gacS* and *gacA* are three putative repressors of *vfr* that down-regulated *vfr* expression.

Table 1: Descriptive Table for temperature-effect of *vfr* regulation

| Time | Temp | Strain | N | Median | Mean | Std Error | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| Late-log | High | WT | 4 | 708.35 | 721.69 | 15.36 | 702.48 | 767.59 |
| | | *gacA* | 4 | 1438.78 | 1452.32 | 134.79 | 1216.06 | 1715.69 |
| | | *gacS* | 4 | 1217.37 | 1207.73 | 19.77 | 1154.00 | 1242.16 |
| | | *ssra* | 4 | 2514.30 | 2521.62 | 149.46 | 2176.57 | 2881.31 |
| | Low | WT | 4 | 1111.31 | 1113.35 | 5.10 | 1104.63 | 1126.14 |
| | | *gacA* | 4 | 1956.68 | 1950.20 | 19.77 | 1896.41 | 1991.03 |
| | | *gacS* | 4 | 1615.45 | 1612.72 | 38.39 | 1539.20 | 1680.81 |
| | | *ssra* | 4 | 2348.40 | 2362.20 | 103.48 | 2165.68 | 2586.31 |
| | Optimal | WT | 4 | 1151.97 | 1165.99 | 59.39 | 1037.95 | 1322.05 |
| | | *gacA* | 4 | 2094.17 | 2078.57 | 43.17 | 1973.91 | 2152.01 |
| | | *gacS* | 4 | 1580.87 | 1556.50 | 47.25 | 1421.84 | 1642.39 |
| | | *ssra* | 4 | 2028.96 | 2053.46 | 42.66 | 1984.01 | 2171.93 |
| Mid-log | High | WT | 4 | 311.35 | 311.25 | 1.74 | 307.05 | 315.28 |
| | | *gacA* | 4 | 721.75 | 725.11 | 13.04 | 697.90 | 759.04 |
| | | *gacS* | 4 | 446.62 | 445.91 | 3.05 | 439.18 | 451.20 |
| | | *ssra* | 4 | 1306.35 | 1290.11 | 89.34 | 1071.98 | 1475.78 |
| | Low | WT | 4 | 906.44 | 891.10 | 25.25 | 819.91 | 931.62 |
| | | *gacA* | 4 | 1469.45 | 1466.97 | 6.79 | 1449.65 | 1479.34 |
| | | *gacS* | 4 | 1151.23 | 1153.07 | 19.91 | 1112.27 | 1197.56 |
| | | *ssra* | 4 | 1928.89 | 1932.94 | 53.47 | 1808.29 | 2065.68 |
| | Optimal | WT | 4 | 957.07 | 949.07 | 19.72 | 900.67 | 981.47 |
| | | *gacA* | 4 | 1610.43 | 1624.56 | 26.09 | 1578.04 | 1699.36 |
| | | *gacS* | 4 | 1437.49 | 1413.62 | 30.72 | 1322.50 | 1456.97 |

| Time | Temp | Strain | N | Median | Mean | Std Error | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| | | *ssra* | 4 | 1529.11 | 1534.20 | 39.04 | 1451.78 | 1626.81 |
| Stationary | High | WT | 4 | 1048.17 | 1049.86 | 22.69 | 996.84 | 1106.27 |
| | | *gacA* | 4 | 1993.90 | 2002.84 | 100.26 | 1815.69 | 2207.88 |
| | | *gacS* | 4 | 1367.40 | 1365.40 | 16.46 | 1332.09 | 1394.71 |
| | | *ssra* | 4 | 2938.83 | 2932.61 | 130.08 | 2613.13 | 3239.66 |
| | Low | WT | 4 | 1398.31 | 1399.90 | 6.89 | 1386.01 | 1416.98 |
| | | *gacA* | 4 | 3044.31 | 2886.51 | 228.86 | 2214.94 | 3242.48 |
| | | *gacS* | 4 | 2118.48 | 2120.89 | 56.38 | 1986.01 | 2260.60 |
| | | *ssra* | 4 | 3321.81 | 3295.72 | 56.16 | 3148.68 | 3390.57 |
| | Optimal | WT | 4 | 1475.67 | 1479.41 | 26.48 | 1426.61 | 1539.69 |
| | | *gacA* | 4 | 2570.35 | 2583.57 | 25.67 | 2541.56 | 2652.01 |
| | | *gacS* | 4 | 1968.79 | 1981.53 | 20.81 | 1947.58 | 2040.97 |
| | | *ssra* | 4 | 2615.85 | 2629.39 | 76.15 | 2477.06 | 2808.80 |

Figure 1: Boxplot for *vfr* gene expression under 3 temperatures

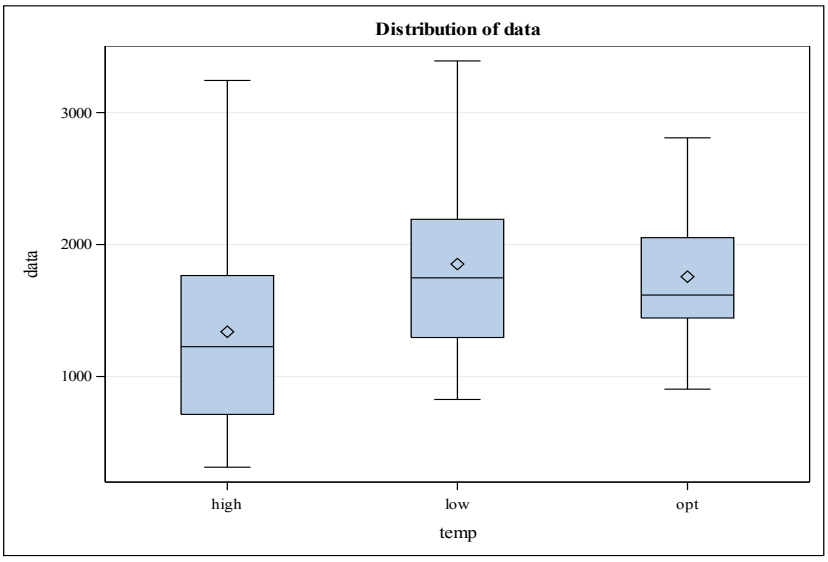


Figure 2: Boxplot for *vfr* gene expression under 4 strains under all temperatures

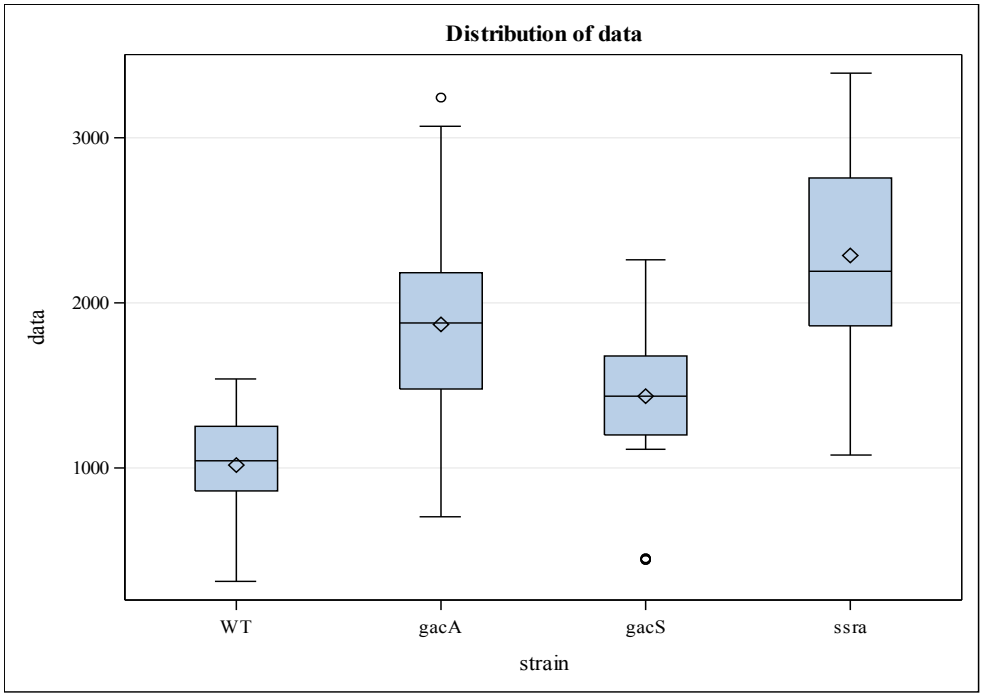
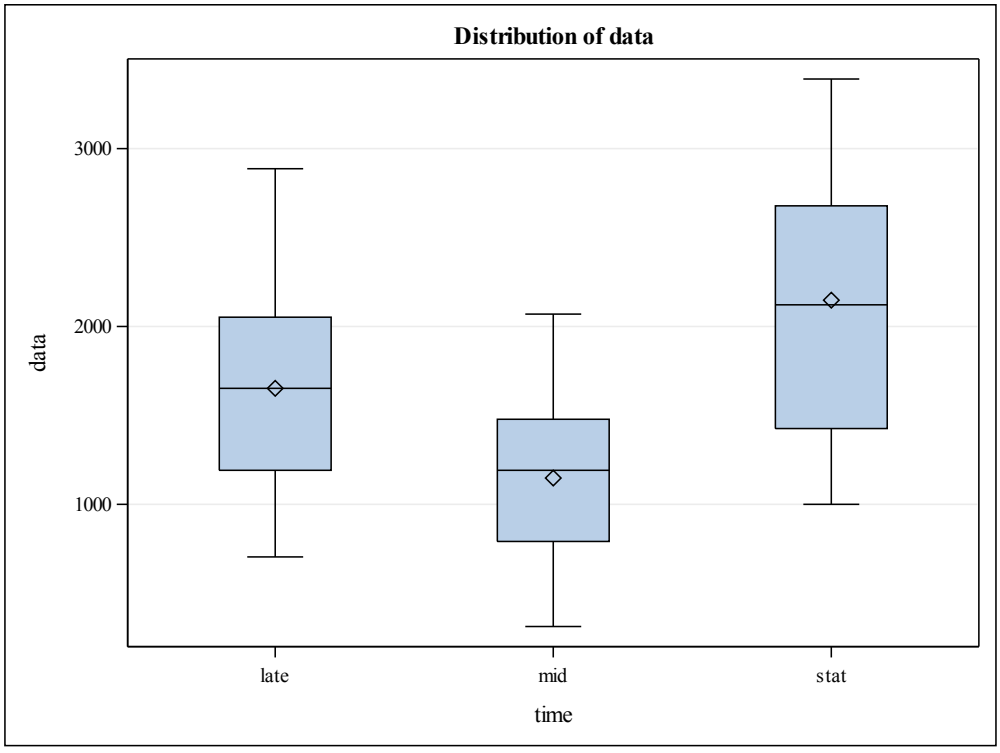



Figure 3: Boxplot for *vfr* gene expression under 3 growth levels under all temperatures



3.2 Linear regression

At the first step, the full linear regression model was fitted by strains, temperature-treatment and growth levels, and all of them were coded as indicator variables since they were categorical variables. As displayed in Table 2, the ANOVA F test had a significant result, indicating that the model was fitted. From Table 3, we noticed that R-Square was 0.872169. From the linear regression model estimate in Table 4, we noticed that 1) wild type strain and *gacS* mutant had the least and second least *vfr* expression, respectively, while *ssrA* mutant and *gacA* mutant had the most and second most *vfr* expression, 2) the high-temperature-treatment had decreased *vfr* expression and low-temperature-treatment had increased *vfr* expression, 3) the lowest growth level had the least *vfr* expression and highest growth level had the most *vfr* expression, as we expected. The P value for low-temperature-treatment was near 0.05 but still not significant.

Table 2: ANOVA table for the temperature effect

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Model** | 7 | 63763898.02 | 9109128.29 | 132.56 | <.0001 |
| **Error** | 136 | 9345702.06 | 68718.40 | | |
| **Corrected Total** | 143 | 73109600.08 | | | |

Table 3: R-square of the liner regression for the temperature effect

| R-Square | Coeff Var | Root MSE | data Mean |
|---|---|---|---|
| 0.872169 | 15.92441 | 262.1419 | 1646.164 |

Table 4: The liner regression for the temperature effect

| Parameter | | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| **Intercept** | | 2889.380729 | 61.78744815 | 46.76 | <.0001 |
| **Strain** | **WT** | -1274.513320 | 61.78744815 | -20.63 | <.0001 |
| **Strain** | ***gacA*** | -420.176871 | 61.78744815 | -6.80 | <.0001 |
| **Strain** | ***gacS*** | -854.987341 | 61.78744815 | -13.84 | <.0001 |
| **Strain** | ***ssra*** | 0.000000 | . | . | . |
| **Temp** | **High** | -418.616711 | 53.50949974 | -7.82 | <.0001 |
| **Temp** | **Low** | 94.643160 | 53.50949974 | 1.77 | 0.0792 |
| **Temp** | **Opt** | 0.000000 | . | . | . |
| **Time** | **Late** | -494.274546 | 53.50949974 | -9.24 | <.0001 |
| **Time** | **Mid** | -999.142770 | 53.50949974 | -18.67 | <.0001 |
| **Time** | **Stat** | 0.000000 | . | . | . |

In addition, robust regression model was applied to raise the robustness and power for the general regression model as the appearance of outliers. Table 5 verified the superiority of the robust regression model because of the advanced R-square value. Table 6 presented the modified estimates via robust regression and other related descriptive statistics. Analogous to P value of low-temperature-treatment in Table 4, in Table 6 it became larger and kept the insignificant difference. Furthermore, Figure 4 and 5 exposed that observation 48 was an apparent outlier. From Figure 6, we concluded that the normal assumption was not satisfied because of its heavy tail and distinguished curve from that of the normal distribution. Also, the QQ-plot (Figure 7) further verified the unsatisfied assumption of normality.

Table 5: R-square of the robust regression for the temperature effect

| Statistic | Value |
|---|---|
| **R-Square** | 0.7247 |
| **AICR** | 165.3385 |
| **BICR** | 192.5779 |
| **Deviance** | 7140040 |

Table 6: Parameter Estimates of the robust regression for the temperature effect

| Parameter | | Estimate | Standard Error | 95% Confidence Limits | | Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|---|
| **Intercept** | | 2767.973 | 55.5702 | 2659.05 | 2876.889 | 2481.07 | <.0001 |
| **Strain** | WT | -1147.73 | 55.5702 | -1256.6 | -1038.8 | 426.57 | <.0001 |
| **Strain** | *gacA* | -318.958 | 55.5702 | -427.87 | -210.04 | 32.94 | <.0001 |
| **Strain** | *gacS* | -733.814 | 55.5702 | -842.72 | -624.89 | 174.38 | <.0001 |
| **Strain** | *ssra* | 0.0000 | . | . | . | . | . |
| **Temp** | High | -503.609 | 48.1252 | -597.93 | -409.28 | 109.51 | <.0001 |
| **Temp** | Low | 57.2665 | 48.1252 | -37.05 | 151.59 | 1.42 | 0.2341 |
| **Temp** | Opt | 0.0000 | . | . | . | . | . |
| **Time** | Mid | -921.295 | 48.1252 | -1015.6 | -826.97 | 366.48 | <.0001 |
| **Time** | Late | -463.182 | 48.1252 | -557.50 | -368.85 | 92.63 | <.0001 |
| **Time** | Stat | 0.0000 | . | . | . | . | . |

Figure 4: Outlier and leverage diagnostics for response variable of the temperature effect

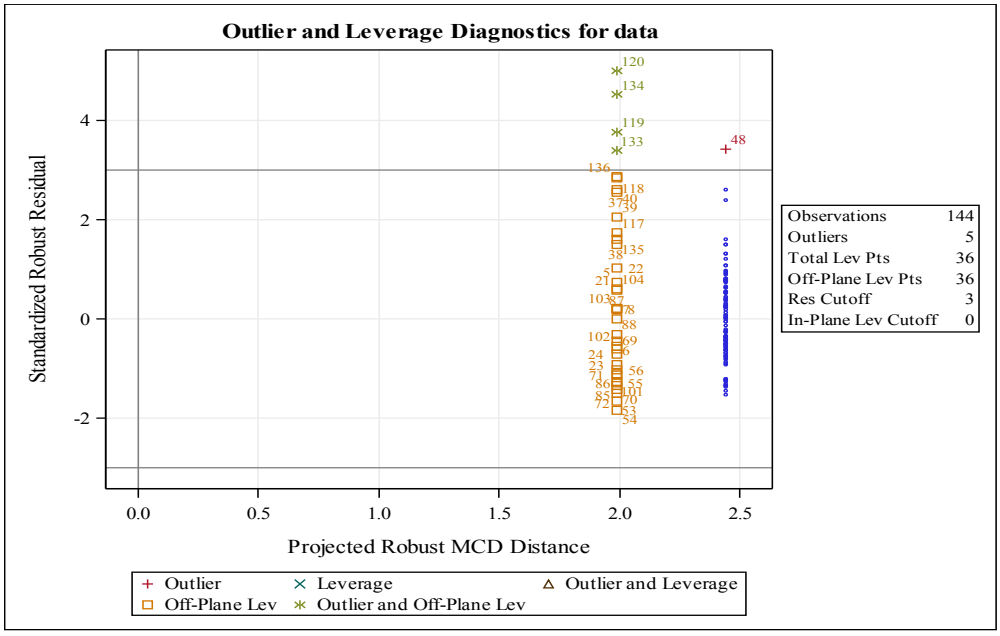



Figure 5: Leverage diagnostics for response variable of the temperature effect

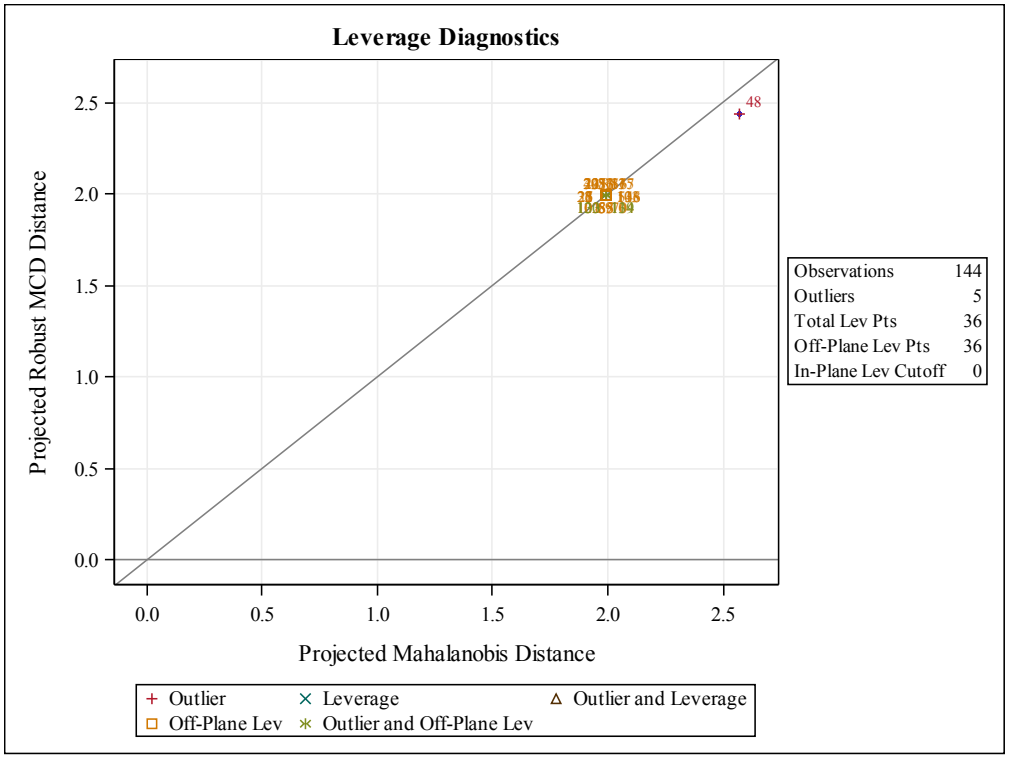


Figure 6: Distribution of residuals for response variable of the temperature effect

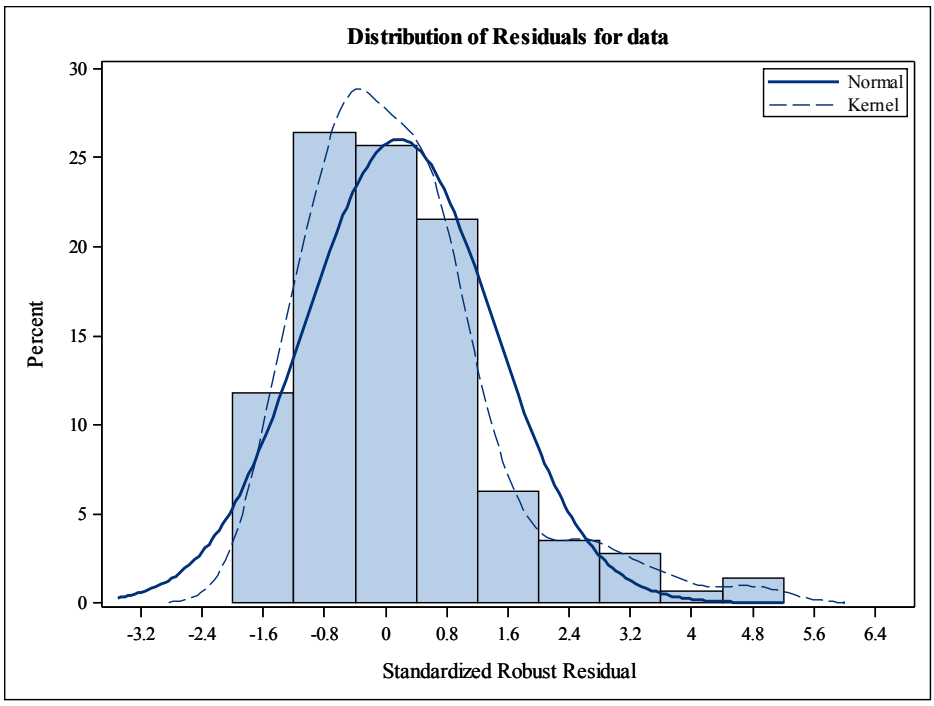
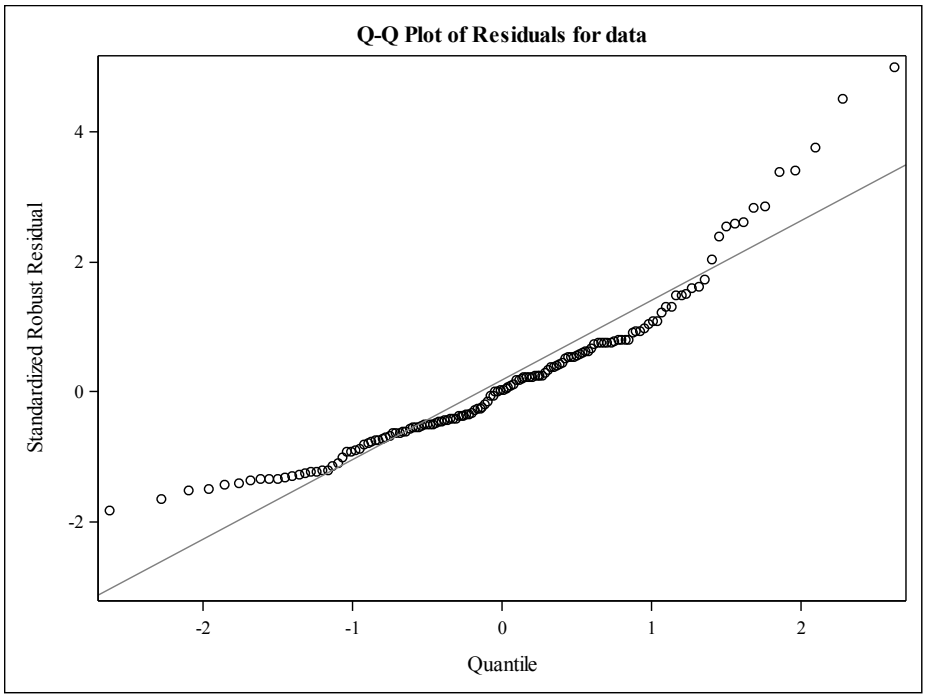



Figure 7: QQ-plot of residuals for response variable of the temperature effect



3.3 Nonparametric analysis

In this section, results are grouped by several nonparametric approaches. Primarily, ANOVA were executed to identify the significances of all variables as well as their interactions. Primarily, a model without interactions was fitted by ANOVA. As shown in Table 7, the P value of the "colonies" variable revealed that there is no difference between different colonies at each time measurement. Hence, we eliminated this variable, and started to explore the interaction between three variables, which showed that all the interactions were significant and needed to be included, as presented in Table 8. In addition, because of the characteristics of high efficiency and robustness to outliers in the response space (John Kloke and Joseph McKean, 2012), rank-based regression models were applied to fit the same model via ranking the residuals. Consequently, all of the P values were also presented to be significant in this rank-based model in Table 9. Furthermore, we performed the drop in dispersion tests by dropping one factor each time, to reconfirm the significance of each variable and their interactions. According to the results presented in Table 10 and 11 that all of P values were smaller than 0.05, we concluded that temperatures, strains and growth levels mutually affected *vfr* expression in *P. aeruginosa.*

Table 7: The model of temperatures without interaction by GLM procedure

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Colonies** | 1 | 45.04 | 45.04 | 0.00 | 0.9797 |
| **Strain** | 3 | 32642002.05 | 10880667.35 | 157.17 | <.0001 |
| **Temp** | 2 | 7162127.57 | 3581063.78 | 51.73 | <.0001 |
| **Time** | 2 | 23959768.40 | 11979884.20 | 173.05 | <.0001 |

Table 8: The model of temperatures with interaction by GLM procedure

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Strain** | 3 | 32642002.05 | 10880667.35 | 566.30 | <.0001 |
| **Temp** | 2 | 7162127.57 | 3581063.78 | 186.38 | <.0001 |
| **Time** | 2 | 23959768.40 | 11979884.20 | 623.51 | <.0001 |
| **Temp*Time** | 4 | 1261726.32 | 315431.58 | 16.42 | <.0001 |
| **Strain*Time** | 6 | 2344125.70 | 390687.62 | 20.33 | <.0001 |
| **Strain*Temp** | 6 | 3114815.95 | 519135.99 | 27.02 | <.0001 |
| **strain*temp*time** | 12 | 549947.86 | 45828.99 | 2.39 | 0.0090 |
| **Residuals** | 108 | 2075086 | 19214 | | |

Table 9: The model of temperature with interaction by rank-based GLM procedure

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Strain** | 3 | 121931 | 40644 | 708.332 | <.0001 |
| **Temp** | 2 | 27798 | 13899 | 242.227 | <.0001 |
| **Time** | 2 | 71673 | 35837 | 624.552 | <.0001 |
| **Temp * Time** | 4 | 2243 | 561 | 9. 773 | <.0001 |
| **Strain* Time** | 6 | 4858 | 810 | 14.112 | <.0001 |
| **Strain* Temp** | 6 | 8775 | 1462 | 25.487 | <.0001 |
| **Strain* Temp* Time** | 12 | 5343 | 445 | 7. 760 | <.0001 |
| **Residuals** | 108 | 6197 | 57 | | |

Table 10: The full model of temperature in the Drop test

| Source | DF | Type I SS | RSS | AIC | F Value | Pr > F |
|---|---|---|---|---|---|---|
| **Strain** | 3 | 6939298 | 9014384 | 1656.4 | 120.3876 | <.0001 |
| **Temp** | 2 | 875388 | 2950474 | 1497.6 | 22.7802 | <.0001 |
| **Time** | 2 | 3286000 | 5361086 | 1583.6 | 85.5116 | <.0001 |
| **Temp*Time** | 4 | 358110 | 2433196 | 1465.8 | 4.6595 | .0002 |
| **Strain*Time** | 6 | 1207304 | 3282390 | 1504.9 | 10.4726 | <.0001 |
| **Strain*Temp** | 6 | 1442762 | 3517849 | 1514.9 | 12.5150 | <.0001 |
| **Temp*Strain*time** | 54 | 549948 | 2625034 | 1460.8 | 2.3852 | 0.009 |

Table 11: Drop in Dispersion Test of temperature data

| **Model :** data = temp + time + strain + temp*time + temp*strain + strain*time | |
|---|---|
| **F-Statistic** | **p-value** |
| 580.94 | <.0001 |
| **Model :** data = temp + time + strain + temp*time + temp*strain | |
| **F-Statistic** | **p-value** |
| 22.597 | 0.000 |
| **Model :** data = temp + time + strain + temp*time + strain*time | |
| **F-Statistic** | **p-value** |
| 24.901 | 0.000 |
| **Model :** data = temp + time + strain + temp*strain + strain*time | |
| **F-Statistic** | **p-value** |
| 17.084 | <.0001 |
| **Model :** data = strain*time | |
| **F-Statistic** | **p-value** |
| 37.12 | 0.000 |

| **Model :** data = temp*strain | |
|---|---|
| **F-Statistic** | **p-value** |
| 66.519 | 0.000 |
| **Model :** data = temp*time | |
| **F-Statistic** | **p-value** |
| 64.846 | 0.000 |

3.3.1 Effects of temperatures on *vfr* gene expression

Robust Tukey's method and rank-based Tukey's test, both characterized as the analyses with the higher power and robustness, were executed for the study of temperatures on *vfr* gene expression. Compared with the output from the general Tukey's test displayed in Table 12, the output from the robust Tukey in Table 13 had the corresponding outcome as the general one, revealing that there is no significant difference of *vfr* expression in *P. aeruginosa* between 37℃ versus 30℃ or 42℃.

Moreover, in order to explore the more specific effects of temperature on *vfr* gene expression at different growth levels, comparisons of *vfr* expression in the wild type strain at three temperatures was categorized by time. Before Tukey's test, rank transformation was performed to take more robustness. Subsequently, multiple comparisons of 37℃ versus 30℃ / 42℃ were tested. The rank-based Tukey's results of each test were unified in Table 14. In mid-log and late-log phases, significant differences appeared between 42℃ and 37℃, but not between 30℃ nd 37℃. However, in the highest growth level stationary phase, there were significant differences of *vfr* gene expression between the 42℃/30℃ and 37℃. In the biological angle, the results indicated that in *P. aeruginosa,* culturing at low temperature (30℃) did not affect the *vfr*

expressions until the stationary phase. However, growing at the high temperature (42℃), had effects on the *vfr* expressions through the whole growth.

Table 12: Tukey's test for 37℃ versus 30℃ / 42℃ comparisons of *vfr* expression

| Strain | Temperature Comparison | Estimate | St Err | t value | Pr(>\|t\|) |
|---|---|---|---|---|---|
| WT | Opt - Low | 503.885 | 195.242 | 2.581 | 0.3016 |
| | Opt - High | 63.369 | 195.242 | 0.325 | 1.0000 |

Table 13: Robust Tukey's test for 37℃ versus 30℃ / 42℃ comparisons of *vfr* expression

| Strain | Temperature Comparison | Estimate | St Err | Lower Bound CI | Upper Bound CI |
|---|---|---|---|---|---|
| WT | Opt - Low | 272.19888 | 180.5028 | -328.4387 | 872.83646 |
| | Opt - High | -417.7899 | 180.5028 | -1018.428 | 182.84767 |

Table 14: Rank-based Tukey's test for 37℃ versus 30℃ / 42℃ comparisons of *vfr* expression

| Strain | Time | Temperature Comparison | Difference Between Rank Means | Simultaneous 95% Confidence Limits | | |
|---|---|---|---|---|---|---|
| WT | Mid-log | Low - Opt | -3.000 | -6.318 | 0.318 | |
| | | High - Opt | -7.500 | -10.818 | -4.182 | *** |
| | Late-log | Low - Opt | -2.000 | -5.823 | 1.823 | |
| | | High - Opt | -7.000 | -10.823 | -3.177 | *** |
| | Stationary | Low - Opt | -4.0000 | -6.3862 | -1.6138 | *** |
| | | High - Opt | -8.0000 | -10.3862 | -5.6138 | *** |

Depending on the principle of the Nemenyi test claimed in the first chapter, rank transformation was conducted first before the test. Differences between the ranks of each group were provided as the test statistics. The critical value was computed by $z_{\alpha}^{*}[\frac{N(N+1)}{12}]^{1/2}(\frac{1}{n_1}+\frac{1}{n_u})^{1/2}$, where $\alpha^{*} = \alpha/(k-1)$ =0.5/(3 − 1), $N$ =12, $n_1 = n_2 = n_3 = 4$.Since all the test statistics were greater than the critical value in Table 16, we concluded that the *vfr* expressions at the low and high temperatures were significantly different from those at the optimal temperatures through the whole growth.

Table 15: Nemenyi test for 37℃ versus 30℃ / 42℃ comparisons of *vfr* expression

| **Strain** | **Time** | **Temperature Comparison** | **Difference Between Rank** | **Critical Value** | |
|---|---|---|---|---|---|
| WT | Mid-log | Opt - Low | 12 | 6.12 | *** |
| | | Opt - High | 30 | | *** |
| | Late-log | Opt - Low | 8 | | *** |
| | | Opt - High | 28 | | *** |
| | Stationary | Opt - Low | 16 | | *** |
| | | Opt - High | 32 | | *** |

The outputs of all multiple comparisons by the Ansari-Bradley test were combined into Table 17. As mentioned before, this analysis was applied for testing the scale of each compared population. Thus, the results indicated that the variance of *vfr* expression in *P. aeruginosa* at 37℃ was different from the variance at 30℃ since late-log phase. However, at the first growth level mid-log phase, *vfr* expressions had an almost equivalent variance between the two temperatures. Meanwhile, *vfr* expression in *P. aeruginosa* had different variances between 37℃

and 42℃, from mid-log phase to late-log phase, but did not persist at the end, which was the stationary phase.

Table 16: Ansari-Bradley test for 37℃ versus 30℃ / 42℃ comparisons of *vfr* expression

| Strain | Time | Temperature Comparison | P Value | |
|---|---|---|---|---|
| WT | Mid-log | Opt - Low | 1 | |
| | | Opt - High | 0.01796048 | *** |
| | Late-log | Opt - Low | 0.01796048 | *** |
| | | Opt - High | 0.08188793 | *** |
| | Stationary | Opt - Low | 0.01796048 | *** |
| | | Opt - High | 0.5541131 | |

We tested comparisons of *vfr* expression of different temperatures by location, and we also tested the comparisons by scale. Now the next analysis will inspect them by both location and scale. In terms of combination, Lepage method is functioned as identifying if two populations both have the same rank mean and the same variance or not. Depending on the outputs of Lepage analysis, which is computed by R in Table 19, none of the comparisons were significantly different, suggesting that temperature did not simultaneously have an effect on *vfr* expression and its range in *P. aeruginosa*.

Table 17: Lepage test for 37℃ versus 30℃ / 42℃ comparisons of *vfr* expression

| Strain | Time | Temperature Comparison | P Value |
|---|---|---|---|

| | | | |
|---|---|---|---|
| WT | Mid-log | Opt - Low | 0.7643343 |
| | | Opt - High | 0.8464817 |
| | Late-log | Opt - Low | 0.1985595 |
| | | Opt - High | 0.8464817 |
| | Stationary | Opt - Low | 0.8464817 |
| | | Opt - High | 0.8464817 |

3.3.2 Effects of temperatures on *vfr* gene regulation

According to the high efficiency and robustness, the general Tukey's method was replaced by the Robust and rank-based Tukey's test, to study the effects of temperatures on *vfr* gene regulation. The output of the general Tukey's test was revealed in Table 20, and the output of the robust Tukey's approach was reflected in Table 21. Compared with the traditional Tukey's method, the Robust Tukey had greatly different results on most tests. Depending on the Robust Tukey's analysis, three mutants had significantly different *vfr* expression from those in wild type at three temperatures, except for *gacA* mutant at 30℃ and *ssra* mutant at 30℃.

Subsequently, in order to investigate the more specific effects of temperatures on *vfr* gene regulation in special growth levels, *vfr* expressions' comparisons at three temperatures were grouped by three growth levels. Consequences of the rank-based Tukey's test were all integrated into Table 22. After rank transformation, each comparison was specified by time and temperature first, and then tested by the rank-based Tukey's methods. The results exposed that all comparisons had significant differences since none of their 95% confidence intervals covered zero. In the biological viewpoint, the consistent results revealed that *gacA*, *gacS* and *ssrA* genes are *vfr* regulators inhibiting *vfr* expressions in *P. aeruginosa* through the whole growth, whether

the temperature is 30℃, 37℃ or 42℃. Therefore, temperature did not affect tmRNA, GacA and GacS dependent *vfr* regulation.

Table 18: Tukey's test for WT versus mutants' comparisons under three temperatures

| **Temperature** | **Strain Comparison** | **Estimate** | **St Err** | **t value** | **Pr(>\|t\|)** | |
|---|---|---|---|---|---|---|
| Low | WT – *ssra* | -1395.498 | 195.242 | -7.148 | <0.01 | *** |
| | WT - *gacA* | -966.442 | 195.242 | -4.950 | <0.01 | *** |
| | WT - *gacS* | -494.110 | 195.242 | -2.531 | 0.3311 | |
| Opt | WT - *ssra* | -874.198 | 195.242 | -4.478 | <0.01 | *** |
| | WT - *gacA* | -897.411 | 195.242 | -4.596 | <0.01 | *** |
| | WT - *gacS* | -452.394 | 195.242 | -2.317 | 0.4697 | |
| High | WT - *ssra* | -1553.844 | 195.242 | -7.959 | <0.01 | *** |
| | WT - *gacA* | -699.157 | 195.242 | -3.581 | 0.0235 | * |
| | WT - *gacS* | -312.074 | 195.242 | -1.598 | 0.9068 | |

Table 19: Robust Tukey's test for WT versus mutants' comparisons under three temperatures

| **Temperature** | **Strain Comparison** | **Estimate** | **St Err** | **Lower Bound CI** | **Upper Bound CI** | |
|---|---|---|---|---|---|---|
| Low | WT - *ssra* | 620.80519 | 180.5028 | 20.16760 | 1221.44277 | *** |
| | WT - *gacA* | -368.80085 | 180.5028 | -969.43843 | 231.83674 | |
| | WT - *gacS* | 748.82471 | 180.5028 | 148.18713 | 1349.46229 | *** |

| | | | | | | |
|---|---|---|---|---|---|---|
| Opt | WT - *ssra* | 9.32777 | 180.5028 | -591.30981 | 609.96536 | |
| | WT - *gacA* | -824.95251 | 180.5028 | -1425.5901 | -224.31493 | *** |
| | WT - *gacS* | -702.75752 | 180.5028 | -1303.3951 | -102.1199 | *** |
| High | WT - *ssra* | 1348.14157 | 180.5028 | 747.50399 | 1948.77916 | *** |
| | WT - *gacA* | 883.81219 | 180.5028 | 283.17461 | 1484.44978 | *** |
| | WT - *gacS* | 943.36707 | 180.5028 | 342.72948 | 1544.00465 | *** |

Table 20: Rank-based Tukey's test for WT versus mutants' comparisons under three temperatures

| **Temperature** | **Time** | **Strain Comparison** | **Difference Between Rank Means** | **Simultaneous 95% Confidence Limits** | | |
|---|---|---|---|---|---|---|
| Low | Mid-Log | *ssra* - WT | 12.0000 | 9.5509 | 14.4491 | *** |
| | | *gacA* - WT | 8.0000 | 5.5509 | 10.4491 | *** |
| | | *gacS* - WT | 4.0000 | 1.5509 | 6.4491 | *** |
| | Late-Log | *ssra* - WT | 12.0000 | 9.5509 | 14.4491 | *** |
| | | *gacA* - WT | 8.0000 | 5.5509 | 10.4491 | *** |
| | | *gacS* - WT | 4.0000 | 1.5509 | 6.4491 | *** |
| | Stationary | *ssra* - WT | 12.0000 | 9.5509 | 14.4491 | *** |
| | | *gacA* - WT | 8.0000 | 5.5509 | 10.4491 | *** |
| | | *gacS* - WT | 4.0000 | 1.5509 | 6.4491 | *** |
| Optimal | Mid-Log | *ssra* - WT | 11.250 | 7.436 | 15.064 | *** |
| | | *gacA* - WT | 8.500 | 4.686 | 12.314 | *** |
| | | *gacS* - WT | 4.250 | 0.436 | 8.064 | *** |
| | Late-Log | *ssra* - WT | 10.000 | 6.051 | 13.949 | *** |
| | | *gacA* - WT | 10.000 | 6.051 | 13.949 | *** |
| | | *gacS* - WT | 4.000 | 0.051 | 7.949 | *** |
| | Stationary | *ssra* - WT | 10.000 | 6.051 | 13.949 | *** |
| | | *gacA* - WT | 10.000 | 6.051 | 13.949 | *** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | *gacS* - WT | 4.000 | 0.051 | 7.949 | *** |
| | | *ssra* - WT | 12.0000 | 9.5509 | 14.4491 | *** |
| | Mid-Log | *gacA* - WT | 8.0000 | 5.5509 | 10.4491 | *** |
| | | *gacS* - WT | 4.0000 | 1.5509 | 6.4491 | *** |
| | | *ssra* - WT | 12.000 | 8.388 | 15.612 | *** |
| High | Late-Log | *gacA* - WT | 7.000 | 3.388 | 10.612 | *** |
| | | *gacS* - WT | 5.000 | 1.388 | 8.612 | *** |
| | | *ssra* - WT | 12.0000 | 9.5509 | 14.4491 | *** |
| | Stationary | *gacA* - WT | 8.0000 | 5.5509 | 10.4491 | *** |
| | | *gacS* - WT | 4.0000 | 1.5509 | 6.4491 | *** |

As mentioned above, the Nemenyi test is also a rank-based analysis that required a rank transformation primarily. And then differences between entire ranks of each group were computed as the test statistics. The critical value of comparisons was computed by $z_{\alpha}^{*}[\frac{N(N+1)}{12}]^{1/2}(\frac{1}{n_1}+\frac{1}{n_u})^{1/2}$, where $\alpha^{*} = \alpha/(k-1)$ =0.05 /(4 − 1), $N = 16$, $n_1 = n_2 = n_3 = n_4 = 4$.Outputs of each comparison were present in Table 24. The results revealed that all the differences between total ranks of each population were greater than the critical value, so neither growing at the low temperature nor at the high temperature would affect tmRNA, GacS and GacA dependent *vfr* regulations in *P. aeruginosa*.

Table 21: Nemenyi test for WT versus 3 mutants comparisons of *vfr* expression

| **Temperature** | **Time** | **Strain Comparison** | **Difference Between Rank** | **Critical Value** | |
|---|---|---|---|---|---|
| | | *ssra* - WT | 48 | | *** |
| | Mid-Log | *gacA* - WT | 32 | | *** |

| | | | | | |
|---|---|---|---|---|---|
| | | *gacS* - WT | 16 | | *** |
| | | *ssra* - WT | 48 | | *** |
| | Late-Log | *gacA* - WT | 32 | | *** |
| Low | | *gacS* - WT | 16 | | *** |
| | | *ssra* - WT | 48 | | *** |
| | Stationary | *gacA* - WT | 32 | 7.17 | *** |
| | | *gacS* - WT | 16 | | *** |
| | | *ssra* - WT | 45 | | *** |
| | Mid-Log | *gacA* - WT | 34 | | *** |
| | | *gacS* - WT | 17 | | *** |
| | | *ssra* - WT | 40 | | *** |
| | Late-Log | *gacA* - WT | 40 | | *** |
| Optimal | | *gacS* - WT | 16 | | *** |
| | | *ssra* - WT | 40 | | *** |
| | Stationary | *gacA* - WT | 40 | | *** |
| | | *gacS* - WT | 16 | | *** |
| | | *ssra* - WT | 48 | | *** |
| | Mid-Log | *gacA* - WT | 32 | | *** |
| | | *gacS* - WT | 16 | | *** |
| | | ssra - WT | 48 | | *** |
| High | Late-Log | *gacA* - WT | 28 | | *** |
| | | *gacS* - WT | 20 | | *** |
| | | *ssra* - WT | 48 | | *** |
| | Stationary | *gacA* - WT | 32 | | *** |
| | | *gacS* - WT | 16 | | *** |

All the results of Ansari-Bradley analyses were computed by R and contained within Table 25. Some comparisons of variances of *vfr* expression had shifted results at low or high temperature, when compared with those at the optimal temperature. Cultivating at 42℃ affected

the variances of tmRNA, GacS and GacA controlled *vfr* regulation of all growth levels, except that the variance of the GacA regulated expression at the stationary phase in *P. aeruginosa*. Meanwhile, culturing at 30℃ affacted the variances of the GacS regulated *vfr* expression at the mid-log phase, and both tmRNA, and GacS regulated *vfr* expressions at the late-log phase in *P. aeruginosa*.

Table 22: Ansari-Bradley test for WT versus mutants' comparisons of *vfr* expression

| **Temperature** | **Time** | **Strain Comparison** | **P Vaule** | |
|---|---|---|---|---|
| Low | Mid-Log | *ssra* - WT | 1 | |
| | | *gacA* - WT | 0.07592696 | *** |
| | | *gacS* - WT | 0.5541131 | |
| | Late-Log | *ssra* - WT | 0.01796048 | *** |
| | | *gacA* - WT | 1 | |
| | | *gacS* - WT | 0.01796048 | *** |
| | Stationary | *ssra* - WT | 0.01796048 | *** |
| | | *gacA* - WT | 0.01796048 | *** |
| | | *gacS* - WT | 0.07592696 | |
| Optimal | Mid-Log | *ssra* - WT | 0.07592696 | |
| | | *gacA* - WT | 0.5541131 | |
| | | *gacS* - WT | 0.2367236 | |
| | Late-Log | *ssra* - WT | 1 | |
| | | *gacA* - WT | 0.5541131 | |
| | | *gacS* - WT | 0.5541131 | |
| | Stationary | *ssra* - WT | 0.01796048 | *** |
| | | *gacA* - WT | 0.03038282 | *** |
| | | *gacS* - WT | 0.2367236 | |
| | | *ssra* - WT | 0.01796048 | *** |

| | | | | |
|---|---|---|---|---|
| High | Mid-Log | *gacA* - WT | 0.01796048 | *** |
| | | *gacS* - WT | 0.01796048 | *** |
| | Late-Log | ssra - WT | 0.08188793 | |
| | | *gacA* - WT | 0.08188793 | |
| | | *gacS* - WT | 0.3843454 | *** |
| | Stationary | *ssra* - WT | 0.07592696 | |
| | | *gacA* - WT | 0.01796048 | *** |
| | | *gacS* - WT | 1 | |

Lepage was finally applied to test two populations by both location and scale, whose results were shown in Table 27. Since none of the tests had a significantly different output, we concluded that temperature did not synchronously affect tmRNA, GacA and GacS controlled *vfr* regulations as well as their variance in *P. aeruginosa*.

Table 23: Lepage test for WT versus 3 mutants comparisons of *vfr* expression

| **Temperature** | **Time** | **Strain Comparison** | **P Vaule** |
|---|---|---|---|
| Low | Mid-Log | *ssra* - WT | 0.8464817 |
| | | *gacA* - WT | 0.8464817 |
| | | *gacS* - WT | 0.8464817 |
| | Late-Log | *ssra* - WT | 0.8464817 |
| | | *gacA* - WT | 0.8464817 |
| | | *gacS* - WT | 0.8464817 |
| | Stationary | *ssra* - WT | 0.8464817 |
| | | *gacA* - WT | 0.8464817 |
| | | *gacS* - WT | 0.8464817 |
| | Mid-Log | *ssra* - WT | 0.8464817 |
| | | *gacA* - WT | 0.8464817 |

| | | | |
|---|---|---|---|
| Optimal | | *gacS* - WT | 0.8464817 |
| | Late-Log | *ssra* - WT | 0.8464817 |
| | | *gacA* - WT | 0.8464817 |
| | | *gacS* - WT | 0.8464817 |
| | Stationary | *ssra* - WT | 0.8464817 |
| | | *gacA* - WT | 0.8464817 |
| | | *gacS* - WT | 0.8464817 |
| High | Mid-Log | *ssra* - WT | 0.8464817 |
| | | *gacA* - WT | 0.8464817 |
| | | *gacS* - WT | 0.8464817 |
| | Late-Log | ssra - WT | 0.8464817 |
| | | *gacA* - WT | 0.8464817 |
| | | *gacS* - WT | 0.8464817 |
| | Stationary | *ssra* - WT | 0.8464817 |
| | | *gacA* - WT | 0.8464817 |
| | | *gacS* - WT | 0.8464817 |

Chapter 4: Results of effects of carbon source on *vfr* gene expression and regulation

In this section, discussed results were related to effects of carbon sources on *vfr* gene expression and regulation.

4.1 Overall description

The descriptive statistics were shown in Table 28 as follows, including names of strains, types of carbon sources with concentration gradients and growth levels, median of the amount of

*vfr* gene expression, mean +/- standard error, minimum/maximum in 120 groups. These plots classified by each time*strain*carbon treatment were also presented in Figure 2 Appendix 1. Obviously, the largest median of *vfr* expression was produced by *ssrA* mutant at the stationary phase when grown in NCE media supplemented 20mM succinate as the sole carbon source. The smallest median of *vfr* expression was produced by the wild type strain at the middle-log phase cultured in NCE media supplemented 1mM glucose. In addition, the boxplot in Figure 8, 9 and 10 gave more visualized details. When *P. aeruginosa* were cultivated in ten medias, the largest median of *vfr* expression appeared in the growth of NCE media with 10% glycerol, while cultivation in L broth had the smallest median of *vfr* expression. Among three different growth levels, the stationary phase had the greatest mean and median of *vfr* expression and the middle-log phase had the least one. Among four strains, the means and medians of *vfr* expression was also arranged as *ssrA* mutant > *gacA* mutant > *gacS* mutant > wild type, which further indicated that mutations of *ssrA*, *gacS* and *gacA* genes induced *vfr* expression in *P. aeruginosa*. Again, the enlarged *vfr* gene expression in three mutants reconfirmed that *ssrA*, *gacS* and *gacA* are repressors that repress *vfr* gene expression.

Table 24: Descriptive Table for carbon-source-effect on *vfr* regulation

| **Time** | **Strain** | **Carbon** | **N** | **Median** | **Mean** | **Std Error** | **Minim** | **Maximum** |
|---|---|---|---|---|---|---|---|---|
| Mid-Log | WT | 1%-Gly | 2 | 785.01 | 785.01 | 1.31 | 783.70 | 786.31 |
| | | 10%-Gly | 2 | 1561.24 | 1561.24 | 13.84 | 1547.40 | 1575.08 |
| | | 10mM-Glu | 2 | 975.77 | 975.77 | 14.88 | 960.88 | 990.65 |
| | | 1mM-Glu | 2 | 659.15 | 659.15 | 10.03 | 649.12 | 669.17 |
| | | 2%-Gly | 2 | 901.04 | 901.04 | 0.00 | 901.04 | 901.04 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 20mM-Suc | 2 | 1107.59 | 1107.59 | 16.05 | 1091.53 | 1123.64 |
| | 2mM-Glu | 2 | 899.47 | 899.47 | 31.75 | 867.72 | 931.22 |
| | 2mM-Suc | 2 | 1034.40 | 1034.40 | 23.64 | 1010.76 | 1058.04 |
| | 4mM-Suc | 2 | 1058.38 | 1058.38 | 27.39 | 1030.99 | 1085.77 |
| | Lb | 6 | 907.59 | 913.54 | 51.39 | 768.36 | 1077.76 |
| *gacA* | 1%-Gly | 2 | 1251.43 | 1251.43 | 3.05 | 1248.37 | 1254.48 |
| Mutant | 10%-Gly | 2 | 2235.14 | 2235.14 | 12.92 | 2222.22 | 2248.06 |
| | 10mM-Glu | 2 | 1290.05 | 1290.05 | 47.16 | 1242.89 | 1337.22 |
| | 1mM-Glu | 2 | 878.59 | 878.59 | 5.60 | 873.00 | 884.19 |
| | 2%-Gly | 2 | 1418.62 | 1418.62 | 0.00 | 1418.62 | 1418.62 |
| | 20mM-Suc | 2 | 1451.63 | 1451.63 | 6.78 | 1444.85 | 1458.42 |
| | 2mM-Glu | 2 | 1170.70 | 1170.70 | 42.06 | 1128.64 | 1212.76 |
| | 2mM-Suc | 2 | 1355.81 | 1355.81 | 0.00 | 1355.81 | 1355.81 |
| | 4mM-Suc | 2 | 1445.33 | 1445.33 | 13.96 | 1431.36 | 1459.29 |
| | Lb | 6 | 1176.11 | 1238.52 | 55.39 | 1113.16 | 1419.84 |
| *gacS* | 1%-Gly | 2 | 1118.13 | 1118.13 | 20.26 | 1097.88 | 1138.39 |
| Mutant | 10%-Gly | 2 | 2117.84 | 2117.84 | 21.57 | 2096.27 | 2139.41 |
| | 10mM-Glu | 2 | 1180.36 | 1180.36 | 3.15 | 1177.21 | 1183.51 |
| | 1mM-Glu | 2 | 820.88 | 820.88 | 10.01 | 810.87 | 830.89 |
| | 2%-Gly | 2 | 1298.26 | 1298.26 | 14.94 | 1283.33 | 1313.20 |
| | 20mM-Suc | 2 | 1317.34 | 1317.34 | 19.60 | 1297.73 | 1336.94 |
| | 2mM-Glu | 2 | 1097.63 | 1097.63 | 3.85 | 1093.78 | 1101.48 |
| | 2mM-Suc | 2 | 1211.79 | 1211.79 | 17.98 | 1193.82 | 1229.77 |
| | 4mM-Suc | 2 | 1286.60 | 1286.60 | 36.97 | 1249.63 | 1323.57 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Lb | 6 | 1104.25 | 1144.23 | 42.87 | 1036.49 | 1285.29 |
| | *ssra* | 1%-Gly | 2 | 2516.54 | 2516.54 | 36.57 | 2479.97 | 2553.12 |
| | Mutant | 10%-Gly | 2 | 4741.80 | 4741.80 | 250.86 | 4490.94 | 4992.66 |
| | | 10mM-Glu | 2 | 1473.28 | 1473.28 | 10.62 | 1462.67 | 1483.90 |
| | | 1mM-Glu | 2 | 1041.39 | 1041.39 | 26.63 | 1014.76 | 1068.02 |
| | | 2%-Gly | 2 | 2804.32 | 2804.32 | 7.77 | 2796.55 | 2812.09 |
| | | 20mM-Suc | 2 | 2321.72 | 2321.72 | 0.00 | 2321.72 | 2321.72 |
| | | 2mM-Glu | 2 | 1276.79 | 1276.79 | 30.61 | 1246.17 | 1307.40 |
| | | 2mM-Suc | 2 | 2152.91 | 2152.91 | 12.23 | 2140.67 | 2165.14 |
| | | 4mM-Suc | 2 | 2305.58 | 2305.58 | 79.24 | 2226.35 | 2384.82 |
| | | Lb | 6 | 1882.72 | 1949.43 | 158.00 | 1498.62 | 2477.99 |
| Late-Log | WT | 1%-Gly | 2 | 1314.60 | 1314.60 | 1.19 | 1313.40 | 1315.79 |
| | | 10%-Gly | 2 | 1936.97 | 1936.97 | 40.62 | 1896.36 | 1977.59 |
| | | 10mM-Glu | 2 | 1740.93 | 1740.93 | 58.37 | 1682.57 | 1799.30 |
| | | 1mM-Glu | 2 | 1538.29 | 1538.29 | 30.46 | 1507.83 | 1568.76 |
| | | 2%-Gly | 2 | 1685.31 | 1685.31 | 4.86 | 1680.45 | 1690.18 |
| | | 20mM-Suc | 2 | 1784.47 | 1784.47 | 28.52 | 1755.95 | 1813.00 |
| | | 2mM-Glu | 2 | 1704.08 | 1704.08 | 13.61 | 1690.48 | 1717.69 |
| | | 2mM-Suc | 2 | 1406.59 | 1406.59 | 26.09 | 1380.51 | 1432.68 |
| | | 4mM-Suc | 2 | 1422.16 | 1422.16 | 36.26 | 1385.89 | 1458.42 |
| | | Lb | 6 | 1335.03 | 1324.56 | 77.51 | 1107.23 | 1534.10 |
| | *gacA* | 1%-Gly | 2 | 1993.55 | 1993.55 | 19.26 | 1974.29 | 2012.81 |
| | Mutant | 10%-Gly | 2 | 2864.12 | 2864.12 | 14.94 | 2849.17 | 2879.06 |
| | | 10mM-Glu | 2 | 2306.05 | 2306.05 | 19.06 | 2287.00 | 2325.11 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1mM-Glu | 2 | 2021.62 | 2021.62 | 36.38 | 1985.24 | 2058.00 |
| | 2%-Gly | 2 | 2480.78 | 2480.78 | 4.99 | 2475.79 | 2485.77 |
| | 20mM-Suc | 2 | 2452.99 | 2452.99 | 30.53 | 2422.47 | 2483.52 |
| | 2mM-Glu | 2 | 2232.64 | 2232.64 | 10.19 | 2222.45 | 2242.84 |
| | 2mM-Suc | 2 | 1860.46 | 1860.46 | 23.35 | 1837.11 | 1883.80 |
| | 4mM-Suc | 2 | 1901.27 | 1901.27 | 35.35 | 1865.93 | 1936.62 |
| | Lb | 6 | 1768.35 | 1806.36 | 136.21 | 1436.74 | 2201.63 |
| *gacS* | 1%-Gly | 2 | 1767.70 | 1767.70 | 9.99 | 1757.71 | 1777.69 |
| Mutant | 10%-Gly | 2 | 2485.77 | 2485.77 | 0.00 | 2485.77 | 2485.77 |
| | 10mM-Glu | 2 | 2123.53 | 2123.53 | 4.51 | 2119.03 | 2128.04 |
| | 1mM-Glu | 2 | 1848.28 | 1848.28 | 9.88 | 1838.40 | 1858.17 |
| | 2%-Gly | 2 | 2256.16 | 2256.16 | 19.97 | 2236.20 | 2276.13 |
| | 20mM-Suc | 2 | 2150.69 | 2150.69 | 41.99 | 2108.70 | 2192.68 |
| | 2mM-Glu | 2 | 2049.97 | 2049.97 | 73.82 | 1976.15 | 2123.79 |
| | 2mM-Suc | 2 | 1714.94 | 1714.94 | 32.63 | 1682.31 | 1747.57 |
| | 4mM-Suc | 2 | 1736.70 | 1736.70 | 2.05 | 1734.64 | 1738.75 |
| | Lb | 6 | 1625.77 | 1657.62 | 118.02 | 1336.28 | 2011.18 |
| *ssra* | 1%-Gly | 2 | 5714.29 | 5714.29 | 28.57 | 5685.71 | 5742.86 |
| Mutant | 10%-Gly | 2 | 6168.18 | 6168.18 | 0.00 | 6168.18 | 6168.18 |
| | 10mM-Glu | 2 | 3926.33 | 3926.33 | 43.12 | 3883.21 | 3969.45 |
| | 1mM-Glu | 2 | 3706.74 | 3706.74 | 95.88 | 3610.85 | 3802.62 |
| | 2%-Gly | 2 | 5907.28 | 5907.28 | 79.47 | 5827.81 | 5986.75 |
| | 20mM-Suc | 2 | 4392.31 | 4392.31 | 64.47 | 4327.84 | 4456.78 |
| | 2mM-Glu | 2 | 3828.14 | 3828.14 | 89.85 | 3738.29 | 3917.98 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 2mM-Suc | 2 | 3226.62 | 3226.62 | 360.65 | 2865.96 | 3587.27 |
| | | 4mM-Suc | 2 | 3712.36 | 3712.36 | 290.72 | 3421.64 | 4003.09 |
| | | Lb | 6 | 3150.36 | 3116.29 | 163.69 | 2632.83 | 3570.81 |
| Stationary | WT | 1%-Gly | 2 | 2708.33 | 2708.33 | 40.06 | 2668.27 | 2748.40 |
| | | 10%-Gly | 2 | 2976.05 | 2976.05 | 18.04 | 2958.01 | 2994.08 |
| | | 10mM-Glu | 2 | 2667.05 | 2667.05 | 9.09 | 2657.96 | 2676.15 |
| | | 1mM-Glu | 2 | 2209.97 | 2209.97 | 0.00 | 2209.97 | 2209.97 |
| | | 2%-Gly | 2 | 2838.01 | 2838.01 | 87.34 | 2750.67 | 2925.35 |
| | | 20mM-Suc | 2 | 2502.89 | 2502.89 | 64.07 | 2438.82 | 2566.96 |
| | | 2mM-Glu | 2 | 2242.81 | 2242.81 | 5.13 | 2237.68 | 2247.94 |
| | | 2mM-Suc | 2 | 1825.62 | 1825.62 | 13.89 | 1811.73 | 1839.51 |
| | | 4mM-Suc | 2 | 1837.62 | 1837.62 | 5.21 | 1832.42 | 1842.83 |
| | | Lb | 6 | 1985.14 | 1966.52 | 45.07 | 1822.92 | 2075.94 |
| | *gacA* | 1%-Gly | 2 | 4030.03 | 4030.03 | 18.80 | 4011.23 | 4048.84 |
| | Mutant | 10%-Gly | 2 | 4479.68 | 4479.68 | 6.25 | 4473.43 | 4485.93 |
| | | 10mM-Glu | 2 | 3497.62 | 3497.62 | 5.51 | 3492.11 | 3503.13 |
| | | 1mM-Glu | 2 | 2880.49 | 2880.49 | 24.72 | 2855.77 | 2905.21 |
| | | 2%-Gly | 2 | 4110.04 | 4110.04 | 35.03 | 4075.01 | 4145.07 |
| | | 20mM-Suc | 2 | 3217.50 | 3217.50 | 34.78 | 3182.72 | 3252.29 |
| | | 2mM-Glu | 2 | 2918.90 | 2918.90 | 24.12 | 2894.77 | 2943.02 |
| | | 2mM-Suc | 2 | 2194.74 | 2194.74 | 5.34 | 2189.40 | 2200.08 |
| | | 4mM-Suc | 2 | 2388.10 | 2388.10 | 14.30 | 2373.80 | 2402.40 |
| | | Lb | 6 | 2638.89 | 2649.93 | 26.44 | 2575.21 | 2766.82 |
| | *gacS* | 1%-Gly | 2 | 3673.45 | 3673.45 | 12.84 | 3660.60 | 3686.29 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Mutant | 10%-Gly | 2 | 4094.89 | 4094.89 | 80.06 | 4014.83 | 4174.95 |
| | 10mM-Glu | 2 | 3397.38 | 3397.38 | 1.64 | 3395.74 | 3399.02 |
| | 1mM-Glu | 2 | 2834.50 | 2834.50 | 1.81 | 2832.69 | 2836.30 |
| | 2%-Gly | 2 | 3965.53 | 3965.53 | 67.49 | 3898.04 | 4033.02 |
| | 20mM-Suc | 2 | 2962.96 | 2962.96 | 17.64 | 2945.33 | 2980.60 |
| | 2mM-Glu | 2 | 2868.17 | 2868.17 | 73.54 | 2794.63 | 2941.72 |
| | 2mM-Suc | 2 | 2132.23 | 2132.23 | 80.16 | 2052.07 | 2212.39 |
| | 4mM-Suc | 2 | 2309.57 | 2309.57 | 55.99 | 2253.58 | 2365.56 |
| | Lb | 6 | 2528.00 | 2504.21 | 27.83 | 2421.23 | 2576.74 |
| *ssra* | 1%-Gly | 2 | 9048.62 | 9048.62 | 109.61 | 8939.01 | 9158.22 |
| Mutant | 10%-Gly | 2 | 11526.25 | 11526.25 | 190.52 | 11335.73 | 11716.77 |
| | 10mM-Glu | 2 | 10671.65 | 10671.65 | 123.77 | 10547.88 | 10795.42 |
| | 1mM-Glu | 2 | 9757.23 | 9757.23 | 68.93 | 9688.30 | 9826.15 |
| | 2%-Gly | 2 | 9790.77 | 9790.77 | 357.65 | 9433.12 | 10148.43 |
| | 20mM-Suc | 2 | 14853.23 | 14853.23 | 388.09 | 14465.14 | 15241.32 |
| | 2mM-Glu | 2 | 10134.20 | 10134.20 | 222.86 | 9911.34 | 10357.05 |
| | 2mM-Suc | 2 | 8735.51 | 8735.51 | 60.18 | 8675.33 | 8795.68 |
| | 4mM-Suc | 2 | 10374.90 | 10374.90 | 100.24 | 10274.66 | 10475.14 |
| | Lb | 6 | 8185.62 | 8112.63 | 486.50 | 6651.07 | 9549.23 |

Figure 8: Boxplot for *vfr* gene expression when fed with 10 carbon sources

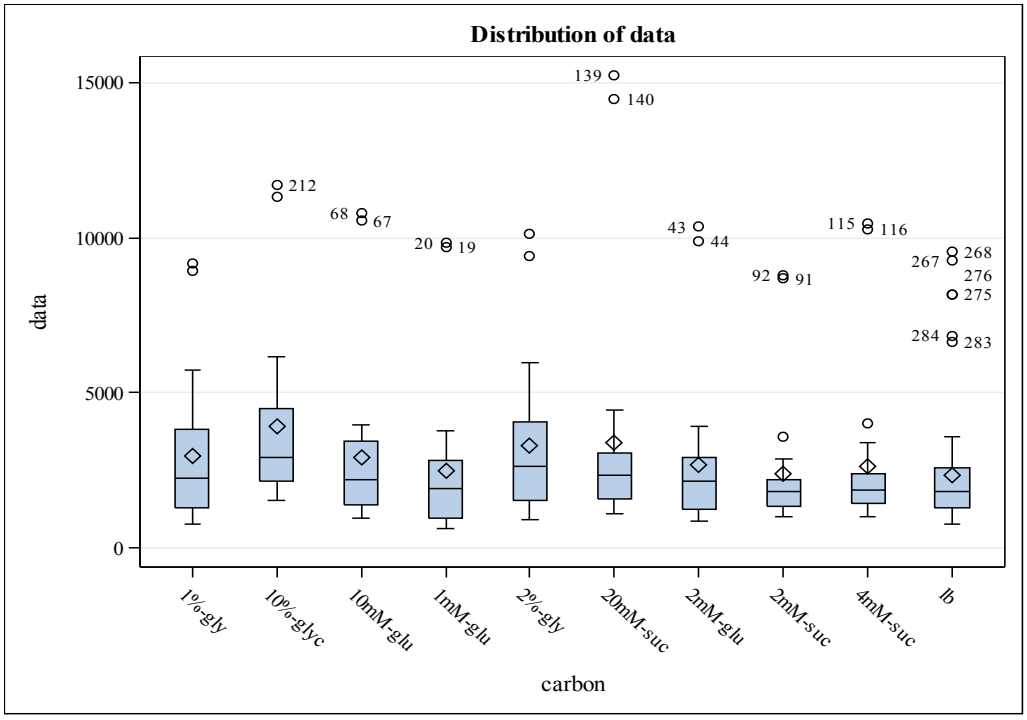



Figure 9: Boxplot for *vfr* gene expression of 4 strains in all medias

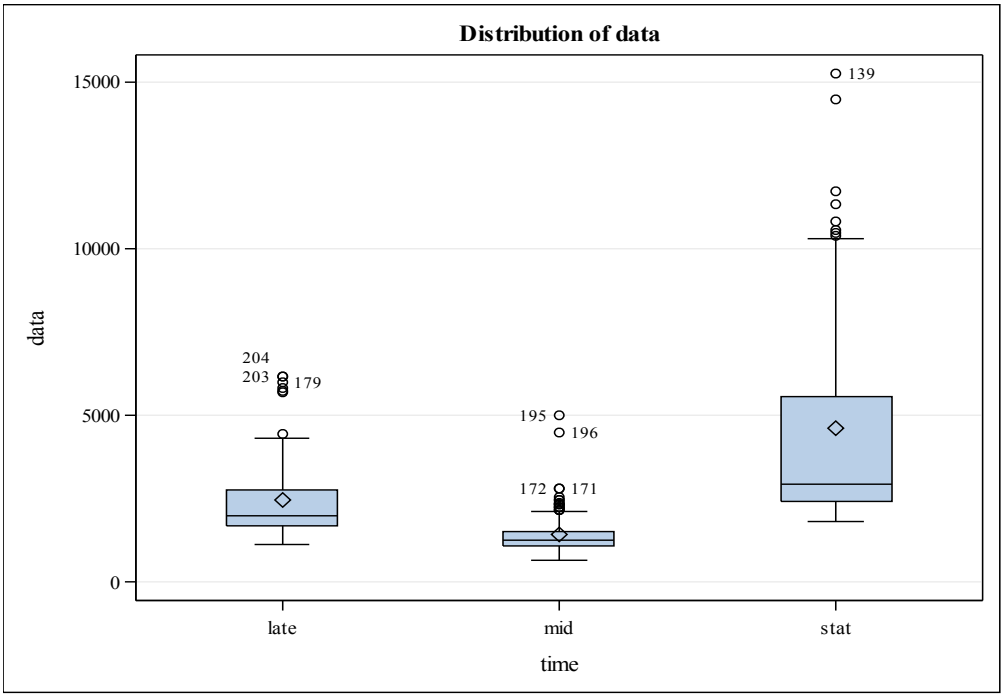
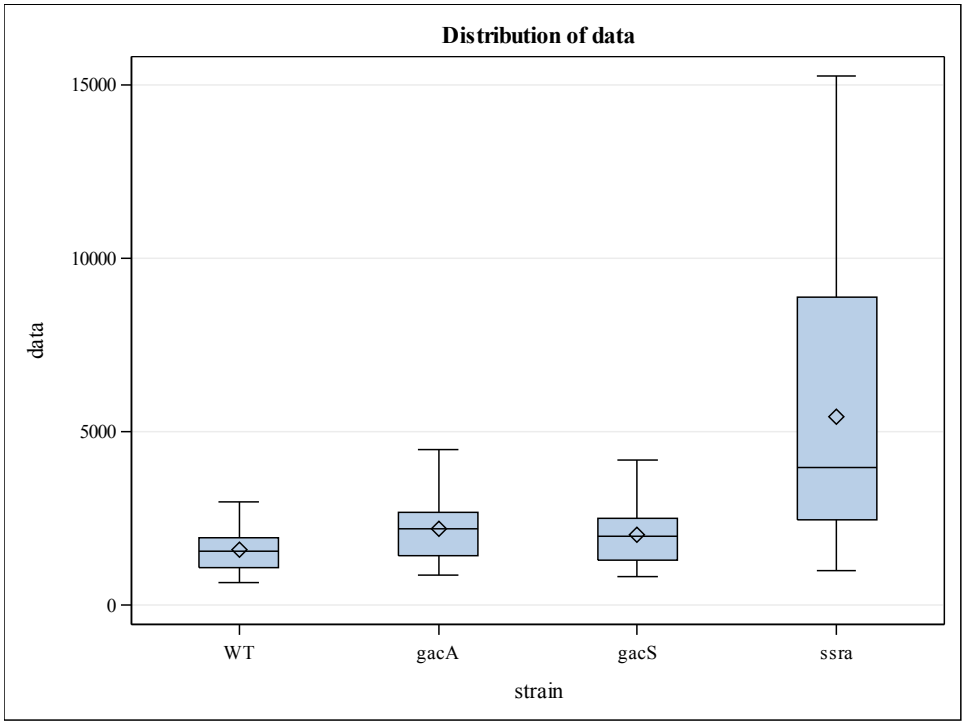


Figure 10: Boxplot for *vfr* gene expression under 4 strains in all medias




4.2 Linear regression

Similar to the last section, the linear regression model was primarily fitted by variables of strains, temperature-treatment and growth levels, and each variable was coded as dummy variables respectively. As displayed in Table 29, the F test of ANOVA was significant, demonstrating that the model was fitted. From Table 30, R-Square was 0.872169. The linear regression detected that 1) wild type strain and *gacS* mutant had the least and second least *vfr* expression, respectively, while *ssrA* mutant and *gacA* mutant had the most and second most *vfr* expression among four strains of *P. aeruginosa*, 2) when the 2mM-succinate supplied NCE media as the sole carbon source, *P. aeruginosa* had the least increased *vfr* expression and when the 20mM-succinate as the sole carbon source, *P. aeruginosa* had the most increased *vfr* expression, compared with the least *vfr* expression *P. aeruginosa* in L broth, 3) the lowest

growth level had the least *vfr* expression and the highest growth level had the most *vfr* expression as expected. Since the P value in all three concentrated glucose supplemented medias as well as low and median concentrated succinate supplemented medias were not significantly different when compared with the control media L broth, shown in Table 31, it is concluded that the inductions of *vfr* expression were not considerable.

Table 25: ANOVA table for the carbon sources effect

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Model** | 14 | 1230212100 | 87872293 | 49.47 | <.0001 |
| **Error** | 273 | 484929811 | 1776300 | | |
| **Corrected Total** | 287 | 1715141910 | | | |

Table 26: R-square of the liner regression for the carbon sources effect

| R-Square | Coeff Var | Root MSE | Data Mean |
|---|---|---|---|
| 0.717265 | 47.18574 | 1332.779 | 2824.538 |

Table 27: The liner regression for the carbon sources effect

| Parameter | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|
| **Intercept** | 6741.557770 | 235.6042546 | 28.61 | <.0001 |
| **Strain WT** | -3825.174646 | 222.1298214 | -17.22 | <.0001 |
| **Strain *gacA* mutant** | -3228.658049 | 222.1298214 | -14.54 | <.0001 |

| Parameter | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|
| **Strain *gacS* mutant** | -3386.238007 | 222.1298214 | -15.24 | <.0001 |
| **Strain *ssra* mutant** | 0.000000 | . | . | . |
| **Carbon 1%-glycerol** | 628.152928 | 314.1390061 | 2.00 | 0.0465 |
| **Carbon 2%-glycerol** | 922.691764 | 314.1390061 | 2.94 | 0.0036 |
| **Carbon 10%-glycerol** | 1567.008239 | 314.1390061 | 4.99 | <.0001 |
| **Carbon 1mM-glucose** | 151.107948 | 314.1390061 | 0.48 | 0.6309 |
| **Carbon 2mM-glucose** | 336.638357 | 314.1390061 | 1.07 | 0.2848 |
| **Carbon 10mM-glucose** | 572.182216 | 314.1390061 | 1.82 | 0.0696 |
| **Carbon 2mM-succinate** | 38.981007 | 314.1390061 | 0.12 | 0.9013 |
| **Carbon 4mM-succinate** | 282.894868 | 314.1390061 | 0.90 | 0.3686 |
| **Carbon 20mM-succinate** | 1010.958310 | 314.1390061 | 3.22 | 0.0014 |
| **Lb** | 0.000000 | . | . | . |
| **Time Late** | -2151.646112 | 192.3700683 | -11.18 | <.0001 |
| **Time Mid** | -3147.015020 | 192.3700683 | -16.36 | <.0001 |
| **Time Stat** | 0.000000 | . | . | . |

In addition, robust regression was applied to provide the robustness and power for the regression as the appearance of some outliers. In Table 32, the advanced R-square value given by robust regression was smaller, since the correctly treated outlier made the regression model less fitted. Table 33 presented the modified estimates of each variable and their related descriptive statistics, compared with the estimates in linear regression in Table 31. Unlike P value in Table 31, some estimates were significant. Only 2mM-succinate, 4mM-succinate and 1mM-glucose

added medias had insignificant differences of *vfr* expression, compared with that of L broth. In addition, it was apparent that Figure 11 and 12 exposed that a lot of observations were outliers. From Figure 13, it was concluded that the normal assumption was not satisfied because of the heavy tail and sharp peak. Also, the QQ-plot (Figure 14) further verified that the distribution was far away from the assumption of normality.

Table 28: R-square of the robust regression for the carbon sources effect

| **Statistic** | **Value** |
|---|---|
| **R-Square** | 0.5804 |
| **AICR** | 480.3470 |
| **BICR** | 546.6177 |
| **Deviance** | 50531225 |

Table 29: Parameter Estimates of the robust regression for the carbon sources effect

| **Parameter** | | **Estimate** | **Standard Error** | **95% Confidence Limits** | | **Chi-Square** | **Pr > ChiSq** |
|---|---|---|---|---|---|---|---|
| **Intercept** | | 3985.200 | 58.8995 | 3869.75 | 4100.64 | 4578.02 | <.0001 |
| **Strain** | WT | -1842.30 | 55.5309 | -1951.14 | -1733.46 | 1100.65 | <.0001 |
| **Strain** | *gacA* | -1272.94 | 55.5309 | -1381.78 | -1164.10 | 525.47 | <.0001 |
| **Strain** | *gacS* | -1419.58 | 55.5309 | -1528.42 | -1310.74 | 653.51 | <.0001 |
| **Strain** | *ssra* | 0.0000 | . | . | . | . | . |
| **Carbon** | 1%-glyerol | 281.1467 | 78.5326 | 127.225 | 435.067 | 12.82 | 0.0003 |

| Parameter | | Estimate | Standard Error | 95% Confidence Limits | | Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|---|
| **Carbon** | 2%-glycerol | 573.9854 | 78.5326 | 420.0643 | 727.9065 | 53.42 | <.0001 |
| **Carbon** | 10%-glycero | 1095.566 | 78.5326 | 941.6449 | 1249.487 | 194.62 | <.0001 |
| **Carbon** | 1mM-glucose | 131.8655 | 78.5326 | -22.0556 | 285.7866 | 2.82 | 0.0931 |
| **Carbon** | 2mM-glucoe | 287.5792 | 78.5326 | 133.6581 | 441.5003 | 13.41 | 0.0003 |
| **Carbon** | 10mM-glucose | 499.2865 | 78.5326 | 345.3654 | 653.2076 | 40.42 | <.0001 |
| **Carbon** | 2mM-succinate | -27.1858 | 78.5326 | -181.107 | 126.7353 | 0.12 | 0.7292 |
| **Carbon** | 4mM-succinate | 78.5399 | 78.5326 | -75.3812 | 232.4610 | 1.00 | 0.3173 |
| **Carbon** | 20mM-succinate | 455.2493 | 78.5326 | 301.3282 | 609.1704 | 33.60 | <.0001 |
| **Carbon** | Lb | 0.0000 | . | . | . | . | . |
| **Time** | Mid | -1624.52 | 48.0912 | -1718.77 | -1530.2 | 1141.08 | <.0001 |
| **Time** | Late | -836.384 | 48.0912 | -930.641 | -742.12 | 302.47 | <.0001 |
| **Time** | Stat | 0.0000 | . | . | . | . | . |

Figure 11: Outlier and leverage diagnostics for response variable of the carbon sources effect

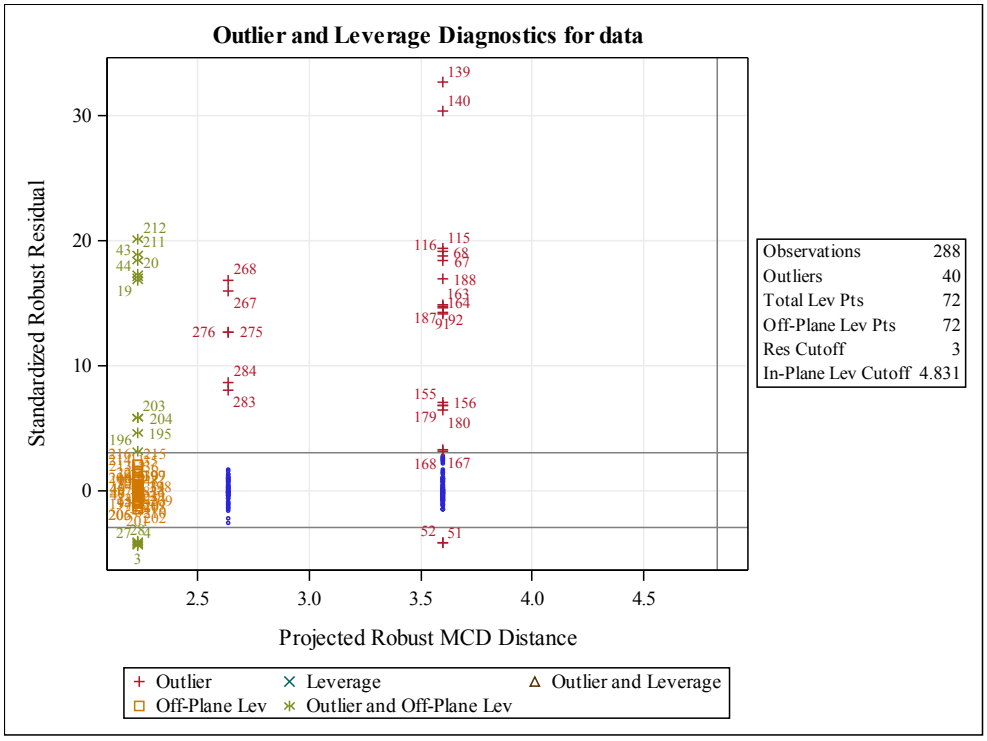



Figure 12: Leverage diagnostics for response variable of the carbon sources effect

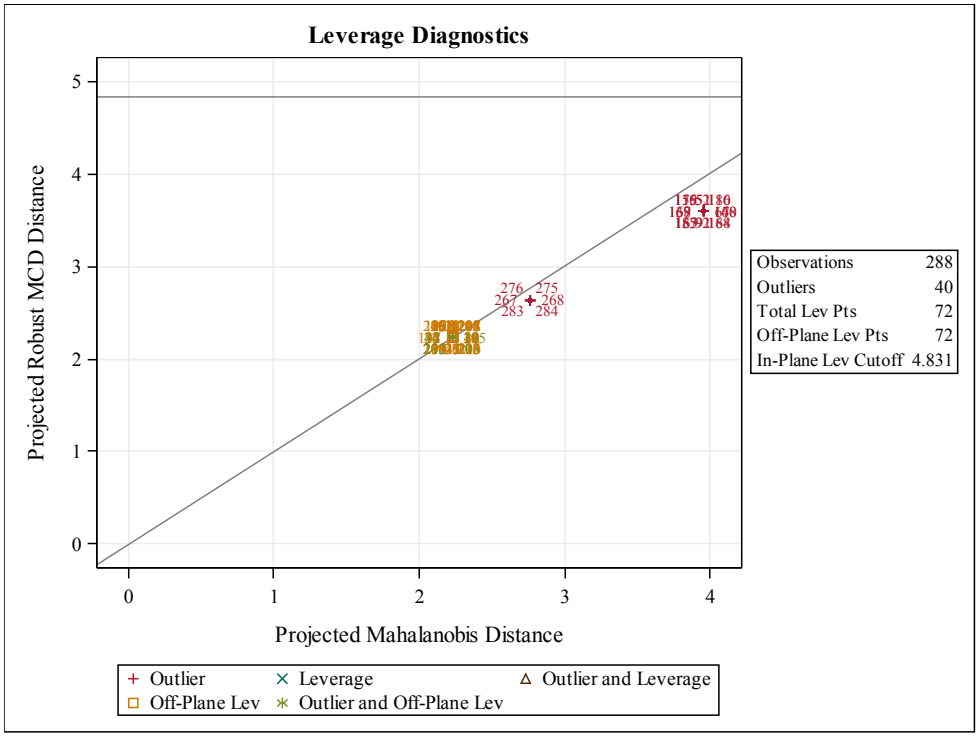


Figure 13: Distribution of residuals for response variable of the carbon sources effect

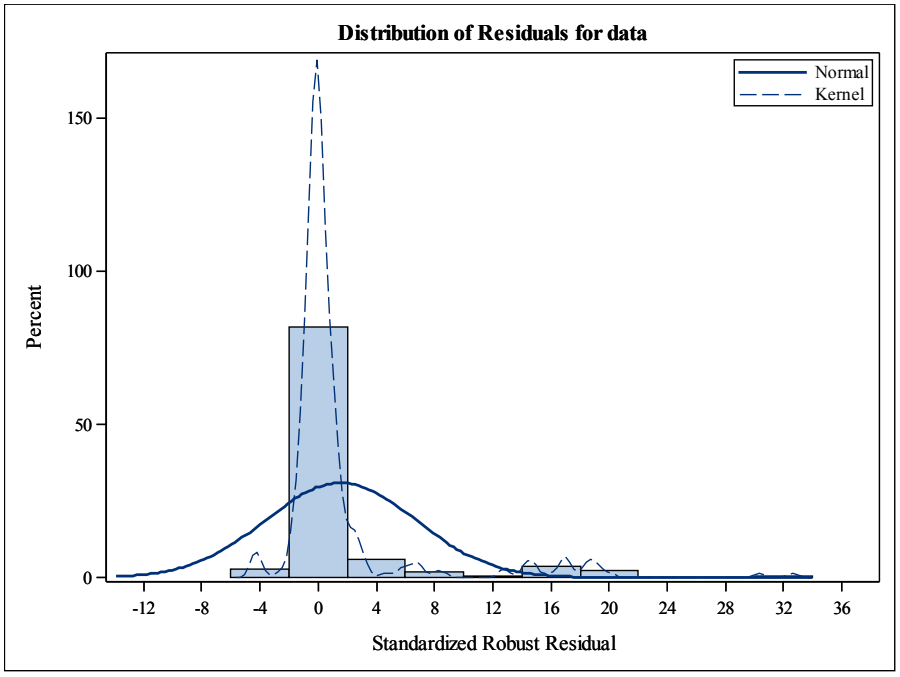
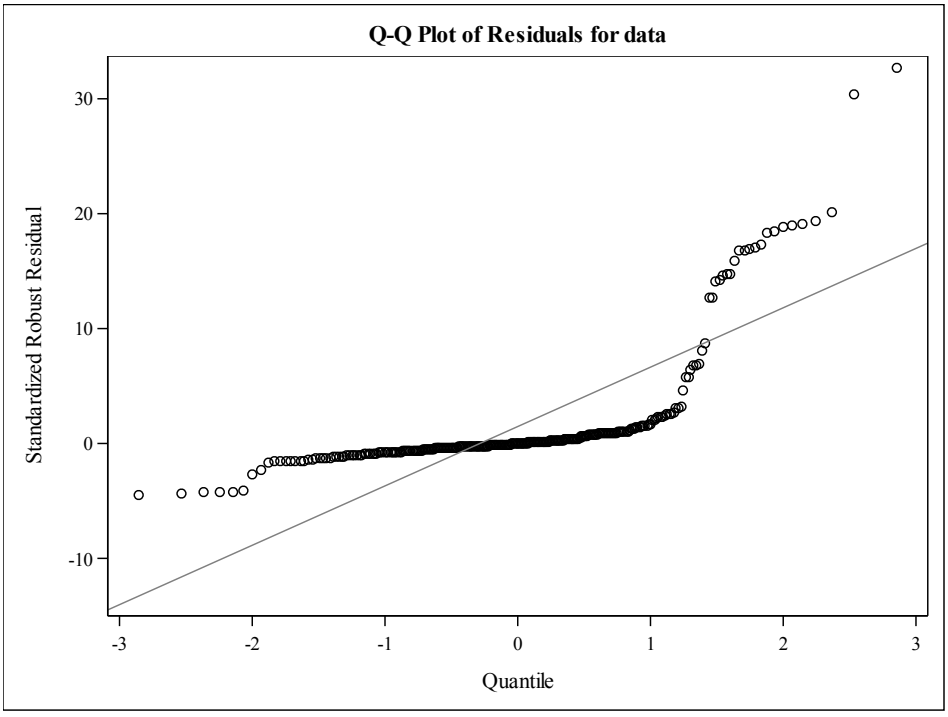



Figure 14: QQ-plot of residuals for response variable of the carbon sources effect



4.3 Nonparametric analysis

Corresponding to the last section, outcomes of this section are presented depending on the types of nonparametric analyses as well. Primarily, ANOVA were implemented to identify the significances of all variables. As shown in Table 34, the P value of the "colonies" variable was insignificant, which revealed that this variable should be disregarded. And then we discovered the interactions between all variables, and the results revealed that all of them were significant as shown in Table 35. In addition, analogous to the last section, the rank-based regression model, which was exceedingly efficient and robust when outliers appeared, was also performed to fit the same model. As shown in Table 36, all of the P values also were shown to be insignificant. Furthermore, we applied the drop in dispersion tests that verified the significance of each variable again, by dropping one each time, respectively. As the results presented in Table 37 and 38 indicate, all P values were also less than 0.05, we concluded that carbon sources, strains and growth levels mutually affected the *vfr* expression in *P. aeruginosa.*

Table 30: The model of carbon sources without interaction by GLM procedure

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
|---|---|---|---|---|---|
| **Carbon** | 9 | 65712348.1 | 7301372.0 | 4.11 | <.0001 |
| **Strain** | 3 | 667730350.4 | 222576783.5 | 125.30 | <.0001 |
| **Time** | 2 | 496769401.2 | 248384700.6 | 139.83 | <.0001 |

Table 31: The model of carbon sources with interaction by GLM procedure

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Carbon** | 9 | 65712348.1 | 7301372.0 | 104.37 | <.0001 |
| **Strain** | 3 | 667730350.4 | 222576783.5 | 3181.78 | <.0001 |
| **Time** | 2 | 496769401.2 | 248384700.6 | 3550.71 | <.0001 |
| **Carbon* Strain** | 27 | 31479034.2 | 1165890.2 | 16.67 | <.0001 |
| **Strain * Time** | 6 | 375341842.5 | 62556973.8 | 894.27 | <.0001 |
| **Carbon * Time** | 18 | 20969517.4 | 1164973.2 | 16.65 | <.0001 |
| **Carbon * Strain * Time** | 54 | 45387230.0 | 840504.3 | 12.02 | <.0001 |
| **Residuals** | 168 | 11752187 | 69953 | | |

Table 32: The model of carbon sources with interaction by rank-based GLM procedure

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Carbon** | 9 | 162389 | 18043 | 110.024 | <.0001 |
| **Strain** | 3 | 630756 | 210252 | 1282.080 | <.0001 |
| **Time** | 2 | 1018388 | 509194 | 3104.974 | <.0001 |
| **Carbon* Strain** | 27 | 20185 | 748 | 4.559 | <.0001 |
| **Strain * Time** | 6 | 16692 | 2782 | 16.964 | <.0001 |
| **Carbon * Time** | 18 | 71684 | 3982 | 24.284 | <.0001 |
| **Carbon * Strain * Time** | 54 | 42980 | 796 | 4.853 | <.0001 |
| **Residuals** | 168 | 27551 | 164 | | |

Table 33: The full model of carbon sources in the Drop test

| Source | DF | Type I SS | RSS | AIC | F Value | Pr > F |
|---|---|---|---|---|---|---|
| **Carbon** | 9 | 2554671 | 14306858 | 3336.2 | 4.0577 | <.0001 |
| **Strain** | 3 | 24746989 | 36499176 | 3618.0 | 117.9212 | <.0001 |
| **Time** | 2 | 8279125 | 20031312 | 3447.2 | 59.1759 | <.0001 |
| **Carbon* Strain** | 27 | 15550800 | 27302986 | 3486.3 | 8.2334 | <.0001 |
| **Strain * Time** | 6 | 13369717 | 25121903 | 3504.4 | 31.8538 | <.0001 |
| **Carbon * Time** | 18 | 5229362 | 16981549 | 3367.6 | 4.1530 | <.0001 |
| **Carbon * Strain * Time** | 54 | 45387230 | 57139416 | 3645.0 | 12.0152 | <.0001 |

Table 34: Drop in Dispersion Test of carbon source data

| **Model :** data = carbon + time + strain + carbon*time + carbon*strain + strain*time | |
|---|---|
| **F-Statistic** | **p-value** |
| 24.542 | 0.000 |
| **Model :** data = carbon + time + strain + carbon*time + carbon*strain | |
| **F-Statistic** | **p-value** |
| 408.93 | 0.000 |
| **Model :** data = carbon + time + strain + carbon*time + strain*time | |
| **F-Statistic** | **p-value** |
| 11.907 | 0.000 |
| **Model :** data = carbon + time + strain + carbon*strain + strain*time | |
| **F-Statistic** | **p-value** |
| 22.955 | 0.000 |
| **Model :** data = strain*time | |
| **F-Statistic** | **p-value** |
| 24.176 | 0.000 |
| **Model :** data = carbon*strain | |

| F-Statistic | p-value |
|---|---|
| 190.09 | 0.000 |
| **Model :** data = carbon*time | |
| **F-Statistic** | **p-value** |
| 131.67 | 0.000 |

4.3.1 Effects of carbon sources on *vfr* expression

For the purpose of studying the effects of carbon sources on *vfr* expression in wild types, the dataset was just filtered with the wild type strain. One reason was that the sample size was very limited, and all segments in this section dropped the time group. The other reason was because time effects would have the synchronized effect for each strain, like the lowest growth level would always have the least *vfr* expression and the highest growth level would always have the most *vfr* expression. Thus, this carbon source set was not taking the time effects into account.

Before Tukey's test, rank transformation was accomplished first to reconstruct the data. And then we tested each comparison of *vfr* expressions cultivated in carbon sources added NCE medias versus L broth. The rank-based Tukey's tests' results were fused in Table 39. As given in the table, all of the tests' consequences exposed that *vfr* expressions of *P. aeryginosa* did not have the significant differences in different medias.

To sum up, *vfr* expression in *P. aeruginosa* would not be affected when treated with glucose, glycerol and succinate as the sole carbon sources.

Table 35: Rank-based Tukey's test for LB versus carbon sources added medias' comparisons

| Strain | Carbon sources Comparison | Difference Between Rank Means | Simultaneous 95% Confidence Limits | |
|---|---|---|---|---|
| WT | 1%-glycerol - LB | 23.222 | -5.075 | 51.519 |
| | 2%-glycerol - LB | 12.889 | -15.408 | 41.186 |
| | 10%-glycerol – LB | 9.722 | -18.575 | 38.019 |
| | 2mM-succinate - LB | 8.722 | -19.575 | 37.019 |
| | 4mM-succinate - LB | 6.722 | -21.575 | 35.019 |
| | 20mM-succinate - LB | 2.056 | -26.242 | 30.353 |
| | 1mM-glucose - LB | 1.722 | -26.575 | 30.019 |
| | 2mM-glucose - LB | 1.056 | -27.242 | 29.353 |
| | 10mM-glucose - LB | 0.556 | -27.742 | 28.853 |

However, depending on the principle of Nemenyi analysis, this method was not applicable in this case, since the unequal sample sizes appeared between the L broth and the remaining medias. For instance, the L broth group had 18 observations collected from the three repeated growths, while the remaining tested medias' group just had 6 observations. In this circumstance, the Nemenyi test was dropped in testing the effects of carbon sources on *vfr* expression.

According to Ansari-Bradley analyses, the growth in all treated medias has similar variances of *vfr* expression when compared with that in L broth, reflected in Table 41. The outcomes indicated that the variance of *vfr* expression in *P. aeruginosa* was also not affected when the glucose, glycerol or succinate were served as the carbon sources during the whole growth. In brief, carbon sources like glycerol, glucose or succinate also did not affect the variance of *vfr* expression in *P. aeruginosa*.

Table 36: Ansari-Bradley test for *vfr* expression in LB versus carbon sources added medias' comparisons

| Strain | Carbon sources Comparison | P Value |
|---|---|---|
| WT | 1%-glycerol - LB | 0.5038702 |
| | 2%-glycerol - LB | 0.2285139 |
| | 10%-glycerol – LB | 0.6883851 |
| | 2mM-succinate - LB | 0.1414243 |
| | 4mM-succinate - LB | 0.1414243 |
| | 20mM-succinate - LB | 0.4224979 |
| | 1mM-glucose - LB | 0.6881267 |
| | 2mM-glucose - LB | 0.8936536 |
| | 10mM-glucose - LB | 0.1414243 |

Finally, as the combination of all tests above, Lepage analysis was a joint test to investigate whether carbon sources have effects on both *vfr* expression and its variance in *Pseudomonas aeruginosa.* According to the output shown in Table 43, none of groups had significant difference on both aspects, reconfirming the results before. In summary, treatment of glycerol, glucose and succinate as the carbon sources would not simultaneously affect *vfr* expression as well as its variance in *Pseudomonas*.

Table 37: Lepage test for *vfr* expression in LB versus carbon sources added medias' comparisons

| Strain | Carbon sources Comparison | P Value |
|---|---|---|
| | 1%-glycerol - LB | 0.645376 |
| | 2%-glycerol - LB | 0.5627974 |

| | | |
|---|---|---|
| WT | 10%-glycerol – LB | 0.966435 |
| | 2mM-succinate - LB | 0.4092001 |
| | 4mM-succinate - LB | 0.4091665 |
| | 20mM-succinate - LB | 0.9918033 |
| | 1mM-glucose - LB | 0.2760861 |
| | 2mM-glucose - LB | 0.4837118 |
| | 10mM-glucose - LB | 0.7220136 |

4.3.2 Effects of carbon sources on *vfr* gene regulation

In order to inspect the effects of carbon sources on *vfr* gene regulation, rank-based Tukey's tests were first performed to analyze the amount of *vfr* expressions in the wild type versus three mutants, and results were revealed in Table 44. After the rank transformation, datasets were classified into subgroups, which divided by ten medias for two strains' comparisons. The *vfr* expressions' comparisons in the *ssrA* mutant versus wild type all showed significant differences in all succinate and glycerol supplemented medias as well as L broth, except for the glucose supplied media. However, in the comparisons of *gacA* mutant versus wild type, the significant differences were not conserved at all in all succinate, glucose and glycerol supplemented medias compared with that in L broth. For the *vfr* expressions between in *gacS* mutant versus wild type, all tests had insignificant results in all medias including L broth.

Hence, the results of rank-based Tukey's methods directed that carbon sources did not alter GacS controlled *vfr* regulation, since the results presented that GacS regulations in *P. aeruginosa* were not greatly induced *vfr* expression in all medias including the L broth. Moreover, the GacA dependent *vfr* regulations became less effective in all three carbon sources supplemented NCE medias compared with that in L broth. Nevertheless, tmRNA (as the product of *ssra* gene)

induced *vfr* expressions were not affected when grown in succinate and glycerol supplemented NCE medias, but would be affected in the glucose supplied one.

Table 38: Rank-based Tukey's test for WT versus mutant comparisons in ten medias

| Carbon sources | Strain Comparison | Difference Between Rank Means | Simultaneous 95% Confidence Limits | | |
|---|---|---|---|---|---|
| 2mM succinate | *ssra* - WT | 13.500 | 5.801 | 21.199 | *** |
| | *gacA* - WT | 5.333 | -2.366 | 13.033 | |
| | *gacS* - WT | 3.167 | -4.533 | 10.866 | |
| 4mM succinate | *ssra* - WT | 14.333 | 7.176 | 21.491 | *** |
| | *gacA* - WT | 6.833 | -0.324 | 13.991 | |
| | *gacS* - WT | 3.500 | -3.658 | 10.658 | |
| 20mM succinate | *ssra* - WT | 10.833 | 1.833 | 19.834 | *** |
| | *gacA* - WT | 5.000 | -4.000 | 14.000 | |
| | *gacS* - WT | 2.167 | -6.834 | 11.167 | |
| 1mM glucose | *ssra* - WT | 8.667 | -1.180 | 18.514 | |
| | *gacA* - WT | 4.000 | -5.847 | 13.847 | |
| | *gacS* - WT | 2.000 | -7.847 | 11.847 | |
| 2mM glucose | *ssra* - WT | 8.833 | -0.993 | 18.659 | |
| | *gacA* - WT | 4.167 | -5.659 | 13.993 | |
| | *gacS* - WT | 2.333 | -7.493 | 12.159 | |
| 10mM glucose | *ssra* - WT | 8.667 | -1.183 | 18.516 | |
| | *gacA* - WT | 4.000 | -5.850 | 13.850 | |
| | *gacS* - WT | 2.000 | -7.850 | 11.850 | |
| 1% glycerol | *ssra* - WT | 11.333 | 2.595 | 20.072 | *** |
| | *gacA* - WT | 4.000 | -4.739 | 12.739 | |
| | *gacS* - WT | 2.000 | -6.739 | 10.739 | |

| | | | | | |
|---|---|---|---|---|---|
| 2% glycerol | *ssra* - WT | 12.000 | 3.560 | 20.440 | *** |
| | *gacA* - WT | 4.333 | -4.107 | 12.774 | |
| | *gacS* - WT | 2.333 | -6.107 | 10.774 | |
| 10% glycerol | *ssra* - WT | 15.333 | 8.865 | 21.802 | *** |
| | *gacA* - WT | 6.000 | -0.468 | 12.468 | |
| | *gacS* - WT | 4.000 | -2.468 | 10.468 | |
| L broth | *ssra* - WT | 35.556 | 22.179 | 48.932 | *** |
| | *gacA* - WT | 15.444 | 2.068 | 28.821 | *** |
| | *gacS* - WT | 9.889 | -3.488 | 23.266 | |

The Nemenyi test results to some tests above, shown in Table 46. Differences between the total ranks of one group were applied as the test statistics in this test. The critical value of each comparison was computed by $z_{\alpha}^{*}[\frac{N(N+1)}{12}]^{1/2}(\frac{1}{n_1}+\frac{1}{n_u})^{1/2}$, where $\alpha^{*} = \alpha/(k-1)$ =$0.5/(4-1)$, $N$ =24, $n_1 = n_2 = n_3 = n_4 = 6$ in 9 treated medias and $N$ =72, $n_1 = n_2 = n_3 = n_4 = 18$ in L broth. Since all of test statistics were larger than the critical value, it was suggested that carbon sources had no effects on tmRNA-, GacA- and GacS- dependent *vfr* regulations, as all the three regulators (tmRNA, GacA and GacS) consistently regulated *vfr* expression perfectly during the growths.

Table 39: Nemenyi test for WT versus mutant comparisons in ten medias

| **Carbon sources** | **Strain Comparison** | **Difference Between Rank** | **Critical Value** | |
|---|---|---|---|---|
| 2mM succinate | *ssra* - WT | 81 | 8.70 | *** |
| | *gacA* - WT | 32 | 8.70 | *** |
| | *gacS* - WT | 19 | 8.70 | *** |
| | *ssra* - WT | 86 | 8.70 | *** |

| | | | | |
|---|---|---|---|---|
| 4mM succinate | *gacA* - WT | 41 | 8.70 | *** |
| | *gacS* - WT | 21 | 8.70 | *** |
| 20mM succinate | *ssra* - WT | 54 | 8.70 | *** |
| | *gacA* - WT | 30 | 8.70 | *** |
| | *gacS* - WT | 13 | 8.70 | *** |
| 1mM glucose | *ssra* - WT | 52 | 8.70 | *** |
| | *gacA* - WT | 24 | 8.70 | *** |
| | *gacS* - WT | 12 | 8.70 | *** |
| 2mM glucose | *ssra* - WT | 53 | 8.70 | *** |
| | *gacA* - WT | 25 | 8.70 | *** |
| | *gacS* - WT | 14 | 8.70 | *** |
| 10mM glucose | *ssra* - WT | 52 | 8.70 | *** |
| | *gacA* - WT | 24 | 8.70 | *** |
| | *gacS* - WT | 12 | 8.70 | *** |
| 1% glycerol | *ssra* - WT | 68 | 8.70 | *** |
| | *gacA* - WT | 24 | 8.70 | *** |
| | *gacS* - WT | 12 | 8.70 | *** |
| 2% glycerol | *ssra* - WT | 72 | 8.70 | *** |
| | *gacA* - WT | 26 | 8.70 | *** |
| | *gacS* - WT | 14 | 8.70 | *** |
| 10% glycerol | *ssra* - WT | 92 | 8.70 | *** |
| | *gacA* - WT | 36 | 8.70 | *** |
| | *gacS* - WT | 24 | 8.70 | *** |
| L broth | *ssra* - WT | 640 | 14.86 | *** |
| | *gacA* - WT | 278 | 14.86 | *** |
| | *gacS* - WT | 178 | 14.86 | *** |

To test the differences of *vfr* expressions' variance between wild types versus other mutants in ten medias, Ansari-Bradley tests were accomplished by R. However, all the tests consistently presented insignificant differences except for the variance of *vfr* expression between *ssra* mutant

and wild type in 2mM succinate supplied NCE media (P value =0.052). In summary, all the variance of GacA and GacS dependent *vfr* regulations were not affected by carbon sources through the whole growth in *P. aeruginosa*. For the variance of tmRNA dependent *vfr* regulations, they were not affected by carbon sources except for 2mM succinate.

Table 40: Ansari-Bradley test for WT versus mutant comparisons of *vfr* expression in ten medias

| **Carbon sources** | **Strain Comparison** | **p Value** |
|---|---|---|
| 2mM succinate | *ssra* - WT | 0.05213509 |
| | *gacA* - WT | 0.7444195 |
| | *gacS* - WT | 0.3315441 |
| 4mM succinate | *ssra* - WT | 0.1055878 |
| | *gacA* - WT | 0.3315441 |
| | *gacS* - WT | 0.3315441 |
| 20mM succinate | *ssra* - WT | 0.1030947 |
| | *gacA* - WT | 0.1954314 |
| | *gacS* - WT | 0.1055878 |
| 1mM glucose | *ssra* - WT | 0.1030947 |
| | *gacA* - WT | 0.1030947 |
| | *gacS* - WT | 0.3280648 |
| 2mM glucose | *ssra* - WT | 0.1055878 |
| | *gacA* - WT | 0.3315441 |
| | *gacS* - WT | 0.1055878 |
| 10mM glucose | *ssra* - WT | 0.3315441 |
| | *gacA* - WT | 0.3315441 |
| | *gacS* - WT | 0.3315441 |
| 1% glycerol | *ssra* - WT | 0.1055878 |
| | *gacA* - WT | 0.1055878 |

| | | |
|---|---|---|
| | *gacS* - WT | 0.1055878 |
| 2% glycerol | *ssra* - WT | 0.1030947 |
| | *gacA* - WT | 0.1005954 |
| | *gacS* - WT | 0.1030947 |
| 10% glycerol | *ssra* - WT | 0.5784348 |
| | *gacA* - WT | 0.3315441 |
| | *gacS* - WT | 0.7811381 |
| LB | *ssra* - WT | 0.6501416 |
| | *gacA* - WT | 0.27806925 |
| | *gacS* - WT | 0.65962772 |

Again, Lepage test was applied to test significant differences between both *vfr* regulation and their regulated range in *P. aeruginosa* in different medias. As shown in Table 49 as follows, the test revealed that none of groups had significant differences from the other, indicating that carbon sources did not simultaneously affect *vfr* regulation and their variances.

Table 41: Lepage test for WT versus mutant comparisons of *vfr* expression in ten medias

| **Carbon sources** | **Strain Comparison** | **p Value** |
|---|---|---|
| 2mM succinate | *ssra* - WT | 0 .8910234 |
| | *gacA* - WT | 0.6063641 |
| | *gacS* - WT | 0.9872613 |
| 4mM succinate | *ssra* - WT | 0.8910234 |
| | *gacA* - WT | 0.8935382 |
| | *gacS* - WT | 0.9872613 |
| 20mM succinate | *ssra* - WT | 0.9325798 |
| | *gacA* - WT | 0.9325513 |
| | *gacS* - WT | 0.9872613 |

| | | |
|---|---|---|
| 1mM glucose | *ssra* - WT | 0.417401 |
| | *gacA* - WT | 0.9872613 |
| | *gacS* - WT | 0.9872613 |
| 2mM glucose | *ssra* - WT | 0.417401 |
| | *gacA* - WT | 0.9325513 |
| | *gacS* - WT | 0.9872613 |
| 10mM glucose | *ssra* - WT | 0.417401 |
| | *gacA* - WT | 0.9872613 |
| | *gacS* - WT | 0.9872613 |
| 1% glycerol | *ssra* - WT | 0.9325798 |
| | *gacA* - WT | 0.9872613 |
| | *gacS* - WT | 0.9872613 |
| 2% glycerol | *ssra* - WT | 0.8662783 |
| | *gacA* - WT | 0.9872613 |
| | *gacS* - WT | 0.9872613 |
| 10% glycerol | *ssra* - WT | 0.8910234 |
| | *gacA* - WT | 0.417401 |
| | *gacS* - WT | 0.417401 |
| LB | *ssra* - WT | 0.9083371 |
| | *gacA* - WT | 0.9872613 |
| | *gacS* - WT | 0.9872613 |

# Chapter 5: Discussion and Conclusion

## 5.1 Interpretation

This is a comprehensive experimental research combining temperature study and carbon source study together. The conclusion will be made by discussing results both from the temperature and carbon source study.

### 5.1.1 Effects of temperature on *vfr* expression

All the outputs of nonparametric analyses for the effects of temperatures on *vfr* expression were integrated in Table 50. Statistically, according to the results of Rank-based Tukey's test, we detected that there were some significant differences in *vfr* expression in *P. aeruginosa* between the high (42℃) and the optimal (37℃) temperature, which occurred through the whole growth. However, there were not significant differences in *vfr* expression in *P. aeruginosa* between the low (30℃) and the optimal (37℃) temperature until the stationary phase. Unlike the rank-based Tukey analysis, the Nemenyi test presented various results, where *vfr* expression in *P. aeruginosa* under the low or high temperature both differed from those under the optimal temperature through the whole growth.

Meanwhile, besides comparing the mean values in each group, we also compared variances of each population by using the Ansari-Bradley method. According to the results of Ansari-Bradley analysis, there were significant differences in the variances of *vfr* expression in *P. aeruginosa* between the high temperature and optimal temperature until late-log phase, but not held until the end, which is stationary phase. However, the significant differences in the

variances of *vfr* expression in *P. aeruginosa* between the low temperature and optimal temperature did not present until the late-log phase.

Finally, according the Lepage analysis, which is applied to test two populations both by location and by scale, neither *vfr* expression nor their variances in *P. aeruginosa* under the low temperature or high temperature were significantly different from those under the optimal temperature through the whole growth.

Table 42: Summary of all nonparametric analyses for 37℃ versus 30℃ / 42℃ comparisons of *vfr* expression

| Strain | Time | Temperature Comparison | Rank-based Tukey's | Nemenyi | Ansari-Bradley | Lepage |
|---|---|---|---|---|---|---|
| | Mid-log | low - opt | | *** | | |
| | | high - opt | *** | *** | *** | |
| WT | Late-log | low - opt | | *** | *** | |
| | | high - opt | *** | *** | *** | |
| | Stationary | low - opt | *** | *** | *** | |
| | | high - opt | *** | *** | | |

Biologically, depending on the outcomes of Rank-based Tukey's, the high temperature (42℃) affected the *vfr* expression in *P. aeruginosa* through the whole growth, and the low temperature (30℃) would not have an effect on *vfr* expression until the stationary phase, when compared with that under the optimal temperature (37℃). However, according to the Nemenyi analysis, both the high temperature and the low temperature had an effect on *vfr* expression in *P. aeruginosa* through the whole growth.

In addition, Ansari-Bradley test suggested that the high temperature affected the variances of *vfr* expression in *P. aeruginosa* from mid-log to late-log phase, but not to the end. For the effect of the low temperature on the variances of *vfr* expression in *P. aeruginosa*, the Ansari-Bradley test discovered that the low temperature would not affect the variances until the late-log phase. Lastly, the Lepage test exposed that temperature did not simultaneously both affect the mean value and the variance of *vfr* expression in *P. aeruginosa* at any growth levels.

### 5.1.2 Effects of temperature on *vfr* regulation

All the outputs of nonparametric analyses for the effects of temperatures on *vfr* regulation were combined into Table 51. Statistically, the two analyses (Rank-based Tukey and Nemenyi), which were all applied to test each group by location, consistently recommended all of tmRNA, GacA and GacS controlled *vfr* expressions in *P. aeruginosa* were different from that in the wild type through the whole growth.

Moreover, the Ansari-Bradley test discovered that, during the mid-log phase, unlike the optimal temperature, high temperature presented the noticeable differences between the variances *vfr* expression in all three mutants and the wild type. In the late-log, the variance of GacS controlled *vfr* expression significantly differed from that in the wild type, which did not appear under the optimal temperature. In the stationary phase, the significant difference between the variance of *vfr* expression in the *ssra* mutant and the wild type disappeared under the high temperature, when compared with that under the optimal temperature. For the comparison between low and optimal temperature, the variances of GacA regulated *vfr* expression was shifted in the mid-log phase, the variances of tmRNA, and GacS controlled ones were reversed

in the late-log phase, while none of these three *vfr* regulations were switched in the stationary phase.

According to the Lepage method, neither high nor low temperature had an effect on both *vfr* regulation and their range in *P. aeruginosa,* since none of the tests presented the significant different between two populations.

Table 43: Summary of all nonparametric analyses for WT versus mutants' comparisons under 3 temperatures

| **Tempe -rature** | **Time** | **Strain Comparison** | **Rank-based Tukey's** | **Nemen -yi** | **Ansari-Bradley** | **Lepage** |
|---|---|---|---|---|---|---|
| low | mid-log | WT - *ssra* | *** | *** | | |
| | | WT - *gacA* | *** | *** | *** | |
| | | WT - *gacS* | *** | *** | | |
| | late-log | WT - *ssra* | *** | *** | *** | |
| | | WT - *gacA* | *** | *** | | |
| | | WT - *gacS* | *** | *** | *** | |
| | Station -ary | WT - *ssra* | *** | *** | *** | |
| | | WT - *gacA* | *** | *** | *** | |
| | | WT - *gacS* | *** | *** | | |
| optimal | mid-log | WT - *ssra* | *** | *** | | |
| | | WT - *gacA* | *** | *** | | |
| | | WT - *gacS* | *** | *** | | |
| | late-log | WT - *ssra* | *** | *** | | |
| | | WT - *gacA* | *** | *** | | |
| | | WT - *gacS* | *** | *** | | |

| | | | | | |
|---|---|---|---|---|---|
| | Station-ary | WT - *ssra* | *** | *** | *** |
| | | WT - *gacA* | *** | *** | *** |
| | | WT - *gacS* | *** | *** | |
| high | mid-log | WT - *ssra* | *** | *** | *** |
| | | WT - *gacA* | *** | *** | *** |
| | | WT - *gacS* | *** | *** | *** |
| | late-log | WT - *ssra* | *** | *** | |
| | | WT - *gacA* | *** | *** | |
| | | WT - *gacS* | *** | *** | *** |
| | Station-ary | WT - *ssra* | *** | *** | |
| | | WT - *gacA* | *** | *** | *** |
| | | WT - *gacS* | *** | *** | |

Biologically, Rank-based Tukey and Nemenyi analyses suggested that temperature did not affect tmRNA, GacA and GacS dependent *vfr* regulations in *P. aeruginos,* since all of their regulations were impactful.

Furthermore, the Ansari-Bradley method showed that high temperature affected all of the variances of tmRNA, GacA and GacS controlled *vfr* expression in mid-log, but only GacS regulated *vfr* expression in late-log, and only tmRNA dependent expression at the stationary phase in *P. aeruginosa*. When cultured at 30℃, the low temperature affected the variances of GacA regulated *vfr* expression at the mid-log phase, and affected the variances of tmRNA and GacS regulated expression at the late-log phase in *P. aeruginosa*, while at the stationary phase, none of the variances of these *vfr* regulations were altered by the low temperature.

Finally, according to the Lepage method, neither high nor low temperature had an effect on both *vfr* regulation and their ranges in *P. aeruginosa,* since none of the tests presented the significant different between each group.

5.1.3 Effects of carbon source on *vfr* expression

All nonparametric analyses for comparisons of LB versus carbon sources added medias were presented in Table 52. Statistically, none of Glucose, Glycerol or Succinate added NCE medias had significantly different *vfr* expressions in *P. aeruginosa* from that in LB through the whole growth, nor their variances, due to the consequences of all these five nonparametric analyses. Biologically, during the growth of *P. aeruginosa,* supplements of Glucose, Glycerol or Succinate as the sole carbon sources would not affect *vfr* expressions and their variances, compared with those cultured in LB.

Table 44: Summary of all nonparametric analyses for comparisons of LB versus carbon sources added medias

| Strain | Carbon sources Comparison | Rank-based Tukey's | Neme-nyi | Ansari-Bradley | Lepage |
|---|---|---|---|---|---|
| WT | 1%-glycerol - LB | | | | |
| | 2%-glycerol - LB | | | | |
| | 10%-glycerol – LB | | | | |
| | 2mM-succinate - LB | | | | |
| | 4mM-succinate - LB | | | | |
| | 20mM-succinate - LB | | | | |
| | 1mM-glucose - LB | | | | |

| |
|---|
| 2mM-glucose - LB |
| 10mM-glucose - LB |

5.1.4 Effects of carbon sources on *vfr* regulation

As presented in Table 53, Rank-based Tukey's test showed that *vfr* expression in *gacS* mutant were not different from that in wild type in all medias, including LB. It also revealed that the differences between *vfr* expression in *gacA* mutant and wild type were not significant any more in all other Glucose, Glycerol or Succinate added NCE medias, when compared with that in LB, excepted for that in 4mM Succinate supplemented media. For the comparisons of the *vfr* expression between *ssrA* mutant and wild type, all of them were different from each other in various medias except for glucose supplied one.

Ansari-Bradley analysis specified that none of the variances of *vfr* expressions in three mutants were significantly different from those in wild type in any medias, including LB.

According to the results of Lepage test, none of tmRNA, GacA and GacS dependent *vfr* regulation and their variances were noticeable.

Biologically speaking, the Rank-based Tukey's method recommended that 1) Cultivation of *P. aeruginosa* in Glucose, Glycerol or Succinate supplemented medias would not affect the GacS controlled *vfr* expression, compared with that in LB. As all of *vfr* expressions were not significantly induced by GacS in *P. aeruginosa*. 2) However, GacA dependent *vfr* regulations were not remarkably induced any more in *P. aeruginosa,* when cultured in Glucose, Glycerol or Succinate supplemented medias, except for 4mM Succinate supplemented one, compared with LB. 3) The tmRNA controlled *vfr* expression was not affected by Glycerol or Succinate

supplemented medias, but became less impressed in Glucose supplied medias, compared with LB.

Ansari-Bradley analysis indicated that none of the variances of GacA and tmRNA controlled *vfr* expression were affected by Glycerol, Glucose or Succinate supplied medias. The test also suggested that the variances of GacS dependent *vfr* expression were not affected by any of these three carbon sauces.

Finally, the Lepage analysis shown that none of Glycerol, Glucose or Succinate had significantly different tmRNA, GacA and GacS dependent *vfr* regulation from that in LB, as well as their variances.

Table 45: Summary of all nonparametric analyses for WT versus mutant comparisons in 10 medias

| **Carbon sources** | **Strain Comparison** | **Rank-based Tukey's** | **Neme-nyi** | **Ansari-Bradley** | **Lepage** |
|---|---|---|---|---|---|
| 2mM succinate | *ssra* - WT | *** | *** | | |
| | *gacA* - WT | | *** | | |
| | *gacS* - WT | | *** | | |
| 4mM succinate | *ssra* - WT | *** | *** | | |
| | *gacA* - WT | | *** | | |
| | *gacS* - WT | | *** | | |
| 20mM succinate | *ssra* - WT | *** | *** | | |
| | *gacA* - WT | | *** | | |
| | *gacS* - WT | | *** | | |
| 1mM glucose | *ssra* - WT | | *** | | |
| | *gacA* - WT | | *** | | |
| | *gacS* - WT | | *** | | |

| | | | |
|---|---|---|---|
| | *ssra* - WT | | *** |
| 2mM | *gacA* - WT | | *** |
| glucose | *gacS* - WT | | *** |
| | *ssra* - WT | | *** |
| 10mM | *gacA* - WT | | *** |
| glucose | *gacS* - WT | | *** |
| | *ssra* - WT | *** | *** |
| 1% glycerol | *gacA* - WT | | *** |
| | *gacS* - WT | | *** |
| | *ssra* - WT | *** | *** |
| 2% glycerol | *gacA* - WT | | *** |
| | *gacS* - WT | | *** |
| | *ssra* - WT | *** | *** |
| 10% glycerol | *gacA* - WT | | *** |
| | *gacS* - WT | | *** |
| | *ssra* - WT | *** | *** |
| LB | *gacA* - WT | *** | *** |
| | *gacS* - WT | | *** |

Among the two location-based multiple comparison methods (Rank-based Tukey's, and Nemenyi), we concluded that Nemenyi test held higher power and robustness than the Tukey test. This higher power and robustness led to narrow confidence intervals, because zero was not included by the Nemenyi test, while zero was covered by the Tukey method. The Lepage test for location-scale comparisons had very low power and weak robustness. As we see, even some of the comparisons suggested either significantly different mean values, or a significantly different variance by both location-based and scale-based analyses. The Lepage test was not able to detect them as the combination of location-based and scale-based methods.

The main results of the study can be summarized as follows: 1) The high (42℃) temperature affected *vfr* expression in *P. aeruginosa* as well as their variances through the whole growth. The low (30℃) temperature also affected *vfr* expression and their variances in *P. aeruginosa*, except the variances in the stationary phase. 2) The low temperature did not affect the tmRNA, GacA and GacS dependent *vfr* regulations, and their variances, since all of the three regulators regulated *vfr* expressions impressively. Also, the high temperature did not affect the tmRNA, GacA and GacS dependent regulations, and their variances excluding the variances of the GacA and GacS regulated ones in the mid-log phase. 3) When Glycerol, Glucose or Succinate supplied as the sole carbon sources in NCE medias, *vfr* expressions in *P. aeruginosa* were not be affected, nor their variances. 4) The tmRNA, GacA and GacS controlled *vfr* expressions in *P. aeruginosa* was not affected by the supply of Glycerol, Glucose or Succinate as the sole carbon sources, as all of the regulations functioned well. The variances of GacS dependent regulation was not affected by the supply of Glycerol, Glucose or Succinate neither, as none of the *vfr* expression in *gacS* mutant were different from that in wild type. Conversely, the variances of GacA dependent *vfr* regulations was altered by supplement of all these carbon sources, which was not significant any more. The variances of tmRNA dependent *vfr* regulations were not affected by the supply of Glycerol or Succinate, but were affected by Glucose supplied NCE media, which turned out to be insignificant.

## 5.2 limitations of the study

Even though this study provided critical insight into the effect of temperature and carbon source on *vfr* gene expression and regulation, there were still several limitations on the experimental design and statistical inference approaches.

The first issue, which may result in bias of our study is associated with the experimental design. The amount of *vfr* expression was collected from two independent time periods. In this case, there would be some bias generating in the data collection since gene expression level will have more or less variety in each incubation, despite controlling for variety. Besides, the $\beta$-glucosidase assay, which was used to test the *vfr* expressions for each strain, has a relatively large system error because of the rapid reaction time, which explains why a considerable difference occurred between the duplicates.

Another problem came from how we define each growth level. For the strain under the same growth level, they just had similar rather than consistent cell density, since cell density were shifted over the time, and it is hard to ensure each measurement of the grown culture exactly had the same cell density. This problem may cause slight fluctuation of the *vfr* expression level for the strains under the same conditions.

Moreover, there is a limitation for the limited sample size as well. Although the total observation-number was not small, the sample size for each multiple comparisons after segments was small. To redeem this problem, more replications need to be conducted to guarantee enough measurements after segments.

## 5.3 Implications of results

For answering the first several questions at the beginning of this thesis, I saw the temperatures (30℃/42℃) affected *vfr* expression but not their regulations controlled by tmRNA, GacA or GacS in *P. aeruginosa*. Carbon sources including Glycerol, Glucose or Succinate, would not have an effect on neither *vfr* expression nor tmRNA, GacA or GacS controlled *vfr* regulations in *P. aeruginosa.* However, for further exploration, the experimental design can be

focused on discovering *vfr* expression and regulation by some more different temperatures or supply of some other carbon sources with a gradient concentration.

## 5.4 Future research

According to the limitations and the unaddressed problems of this work, the continuous studies need to be investigated in the future. In order to solve the problem of the system errors of the $\beta$- glucosidase assay and the limited sample size, more replication need to be accomplished to reduce the experiment error. To diminish the fluctuation of the cell density of the collected culture at the same growth level, OD values need to be read highly frequently.

In addition, the approach that I test the effect of temperatures and carbon sauces on *vfr* regulation can be applied and developed to test the effects of some other factors in response to *vfr* regulation, like nitrogen sources or oxygen concentration.

Reference

[1] John Fox, Robust Regression, January 2002.
[2] George A. F. Seber, Alan J. Lee, Linear Regression Analysis, 2nd Edition, Biometrical Journal, February 2003, ISBN: 978-0-471-41540-4.
[3] Xue Bai, Robust Linear Regression, 2012.
[4] Marco Marozzi, Nonparametric Simultaneous Tests for Location and Scale Testing: A Comparison of Several Methods, 2013. DOI:10.1080/03610918.2012.665546
[5] Tukey, John. Comparing Individual Means in the Analysis of Variance. Biometrics, Vol. 5, No. 2 (Jun., 1949), pp. 99-114.
[6] Clyde Young Kramer. Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications. Biometrics, Vol. 12, No. 3 (Sep., 1956), pp. 307-310.
[7] Nadim Nachar, the Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. 2008. Tutorials in Quantitative Methods for Psychology, vol. 4(1), p. 13-20.
[8] Nemenyi, P.B. (1963) Distribution-free Multiple Comparisons. PhD thesis, Princeton University.
[9] Dunnett, C. W. (1964). "New tables for multiple comparisons with a control". Biometrics 20 (3): 482–491. JSTOR 2528490
[10] P. B. Nemenyi. “Distribution-free multiple comparisons”. PhD thesis, Princeton University, 1963.
[11] Dunn, Olive Jean (1964). "Multiple comparisons using rank sums". Technometrics 6 (3): 241–252.
[12] Levene, Howard (1960). Ingram Olkin, Harold Hotelling, et alia, ed. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford University Press. pp. 278–292.
[13] Brown, Morton B.; Forsythe, Alan B. (1974), "'Robust tests for equality of variances", Journal of the American Statistical Association 69: 364–367,
[14] Yves Lepage. A Table for a Combined Wilcoxon Ansari-Bradley Statistic. *Biometrika*. Vol. 60, No. 1 (Apr., 1973), pp. 113-116
[15] Mann, Henry B.; Whitney, Donald R. (1947). "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other". Annals of Mathematical Statistics 18 (1): 50–60.
[16] A. R. Ansari and R. A. Bradley: Rank-sum test for dispersions. Ann. Math. Statist. 31 (1960), 1174-1189.
[17] MAGASANIK, B. (1961). Catabolite repression. Cold Spring Harb. Symp. quant. Biol. 26, 249.
[18] Collier DN1, Hager PW, Phibbs PV Jr. Catabolite repression control in the Pseudomonads. Res Microbiol. 1996 Jul-Sep;147(6-7):551-61.

[19] Wolff, J. A., C. H. MacGregor, R. C. Eisenberg, and P. V. Phibbs, Jr. 1991. Isolation and characterization of catabolite repression control mutants of Pseudomonas aeruginosa PAO. J. Bacteriol. 173:4700-4706.
[20] Sang-Jin Suh, Laura Silo-Suh, Donald E. Woods, Daniel J. Hassett, Susan E. H. West, and Dennis E. Ohman. Effect of rpoS Mutation on the Stress Response and Expression of Virulence Factors in Pseudomonas aeruginosa. J. Bacteriol. July 1999 vol. 181 no. 13 3890-3897
[21] W. J. Conover and Ronald L. Iman. Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics. Aug, 1981. Taylor & Francis, Ltd. on behalf of the American Statistical Association.
[22] Rfit: Rank-based Estimation for Linear Models. John D. Kloke and Joseph W. McKean. The R Journal Vol. 4/2, December 2012. ISSN 2073-4859.
[23] Hettmansperger, Thomas P. ; McKean, Joseph W. Robust nonparametric statistical methods. Second edition. Monographs on Statistics and Applied Probability, 119. CRC Press, Boca Raton, FL, 2011. xviii+535 pp. ISBN: 978-1-4398-0908-2
[24] Jaeckel, Louis A. Estimating regression coefficients by minimizing the dispersion of the residuals. Ann. Math. Statist. 43 (1972), 1449--1458.
[25] Heiler, Siegfried ; Willers, Reinhart . Asymptotic normality of R-estimates in the linear model. Statistics 19 (1988), no. 2, 173--184.

Appendix 1: Figures

Figure 15: Plot of the temperature dataset classified by time*temperature*strain treatments

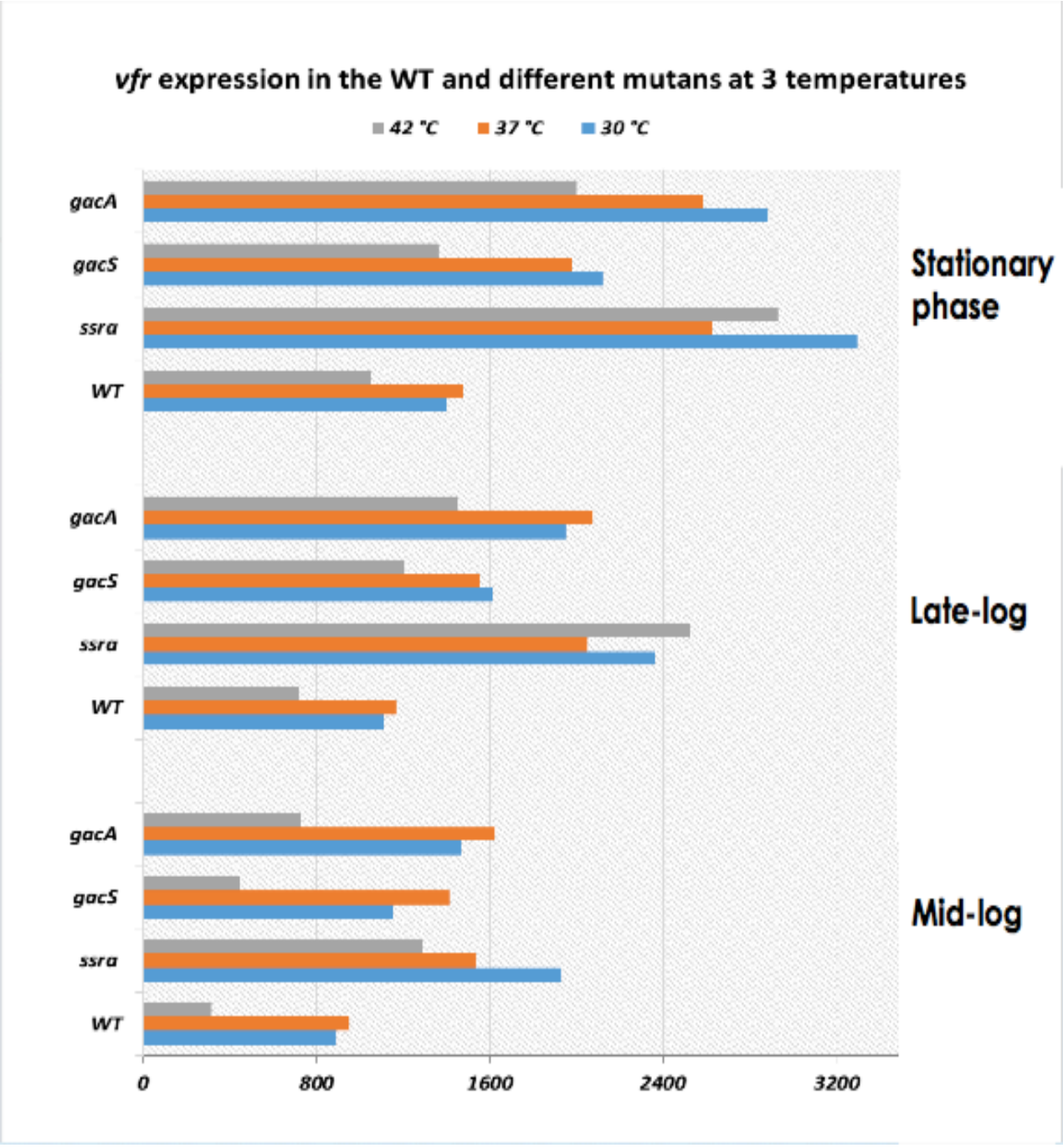
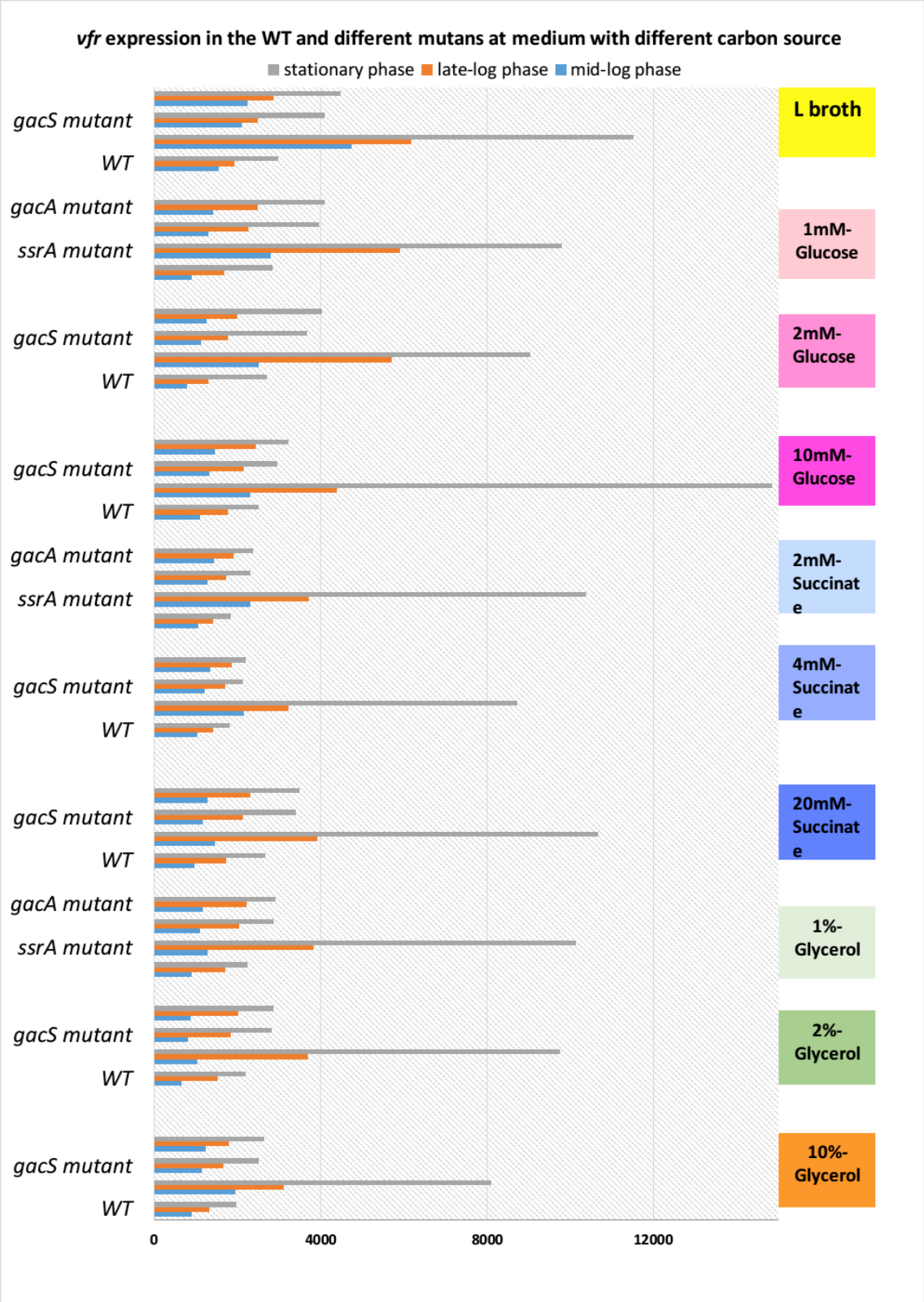


Figure 16: Plot of the carbon dataset classified by time*strain*carbon treatments



Appendix 2: Partial SAS Code

Temperature SAS Code:

```
/***nonparametric analysis***/
proc glm data = temp1;
class strain temp time colonies;
model data = colonies strain temp time;
random colonies;
lsmeans strain temp time;
run;
proc glm data = temp1;
class strain  temp  time;
model data = strain temp time temp*time time*strain temp*strain
temp*strain*time;
lsmeans temp |time| strain /pdiff;
run;
proc glm data = temp1;
class strain  temp  time;
model data = strain temp time temp*time time*strain temp*strain
temp*strain*time;
means temp time strain /tukey;
run;
proc sort data =temp1;
by time;
run;
proc glm data = temp1;
class temp strain;
model data = temp strain strain*temp;
lsmeans temp*strain /pdiff adjust=tukey;
by time;
run;
proc sort data =temp1;
by strain;
run;
proc glm data = temp1;
class temp time;
model data = temp time temp*time;
lsmeans temp*time / pdiff adjust=tukey;
by strain;
run;
proc sort data =temp1;
by temp;
run;
proc glm data = temp1;
class strain time;
model data = strain time time*strain;
lsmeans time*strain / pdiff adjust=tukey;
by temp;
run;
/*******WT at 3 temp*******/
proc sort data =temp1 out=newtemp2;
by strain time;
run;
proc rank data=newtemp2 out=nt2 ties=mean;
ranks rankdata2;
```

```
var data;
run;
proc print;
run;
proc glm data = nt2;
class temp;
model rankdata2 = temp;
means temp /dunnett ('opt') hovtest = BF;
by strain time;
run;
/****split data  +  Tukey's(strain)****/
data highlate;
set temp1;
if (temp eq 'high' && time eq 'late');
run;
proc rank data=highlate out=hmrank ties=mean;
ranks rankdata;
var data;
run;
proc print;
run;
proc glm data = hmrank;
class strain;
model rankdata = strain;
means strain / tukey dunnett('WT');
run;
quit;
/****temp  +  Tukey's****/
data highlate;
set temp1;
if (strain eq 'WT' && time eq 'late');
run;
proc rank data=highlate out=hmrank ties=mean;
ranks rankdata;
var data;
run;
proc print;
run;
proc glm data = hmrank;
class temp;
model rankdata = temp;
means temp / tukey dunnett('opt');
run;
quit;
/***PROC MEAN***/
proc means data = temp1 maxdec=2 N median mean stderr min max ;
class  time temp strain;
var data ;
output out=means1 n=n median=median mean=mean stderr=stderr min=min max=max;
run;
```

Carbon Source SAS code:

```
ata carbon;
input carbon$ strain$ time$ data;
logdata = log(data);
```

```
if strain = 'WT' then WT = 1; else WT = 0;
if strain = 'ssra' then ssra = 1; else ssra = 0;
if strain = 'gacS' then gacS = 1; else gacS = 0;
if time = 'mid' then mid = 1; else mid = 0;
if time = 'late' then late = 1; else late = 0;
if carbon eq "1mM-glu" then x1 = 1; else x1 = 0;
if carbon = '2mM-glu' then x2 = 1; else x2 = 0;
if carbon = '10mM-glu' then x3 = 1; else x3 = 0;
if carbon = '2mM-suc' then x4 = 1; else x4 = 0;
if carbon = '4mM-suc' then x5 = 1; else x5 = 0;
if carbon = '20mM-suc' then x6 = 1; else x6 = 0;
if carbon = '1%-gly' then x7 = 1; else x7 = 0;
if carbon = '2%-gly' then x8 = 1; else x8 = 0;
if carbon = '10%-gly' then x9 = 1; else x9 = 0;
datalines;
……;
proc print data = carbon;
run;
/****regression model***/
proc glm data = carbon;
class strain carbon time;
model data = strain carbon time/solution;
run;
proc robustreg data = carbon
plots = (rdplot ddplot reshistogram resqqplot);
class strain carbon time;
model data = strain carbon time;
run;
/**anova**/
proc glm data = carbon;
class carbon strain time;
model data = carbon strain time;
lsmeans strain carbon time;
run;
proc glm data = carbon;
class carbon strain time;
model data = carbon strain time carbon*strain strain*time carbon*time
carbon*time*strain;
run;
/****Tukey's(1)****/
data highlate;
set carbon;
if (carbon eq '1mM-glu');
run;
proc rank data=highlate out=hmrank ties=mean;
ranks rankdata;
var data;
run;
proc print;
run;
proc glm data = hmrank;
class strain;
model rankdata = strain;
means strain / tukey dunnett('WT');
run;
quit;
/**** Tukey's 2****/
```

```
data highlate;
set carbon;
if (strain = 'WT');
run;
proc rank data=highlate out=hmrank ties=mean;
ranks rankdata;
var data;
run;
proc print;
run;
proc glm data = hmrank;
class carbon;
model rankdata = carbon;
means carbon / tukey dunnett('lb');
run;
quit;
```