

# Microbial growth models – which is better and why?

PokMan HO

Department of Life Sciences, Faculty of Natural Sciences,

Imperial College London



**Imperial College**  
**London**

Approximate Word Count: 1180

# Microbial growth models – which is better and why?

PokMan HO (CID: 01786076)

## Abstract

## Introduction

Phenological models are expected to fit data trends within its biological field. Yet due to different reasons, models developed and published from one sample may not fit the others. These reasons can be data variabilities, confounding factors, inaccurate assumptions or models being too-specific. This data-mining project is aimed at comparing and contrasting published phenological models on microbial population size data, highlighting which is a better model under what conditions. The hypotheses are:

- published phenological models are significantly better fits than polynomials for describing microbial growths;
- appropriate phenological model(s) is/are identifiable through distinguishable microbial population size parameter values; and
- parameter values of each phenological model from successful fits are clusters with well-defined boundaries between models.

## Methods

Microbial population sizes data were given, sourced from ten different publications (Table1). The collection contained different microbial clades growing at various conditions for varying times. Also, these data were recorded in multiple time and population units. Some of the population data were direct counts while some were not.

Experimental microbial population growth data library were divided into individual data subsets through six filters (“Temperature (in °C)”, “Microbial clade”, “growth substrate mate-

rials”, “experimental replicate number”, “population data recording unit” and “data source”). Records with data unit “OD\_595” were scaled into optical density percentages (i.e. data\*100) to facilitate general analyses workflow. Independent (or explanatory) variable was “Time (hr)” and dependent (or response) variable was “population size”.

Some raw data were recorded in minutes (instead of hour). This record artifact was not corrected because of two reasons: 1. shape of curves were the main concern instead of independent variable’s scale; and 2. the unit was consistent within each data subset.

## Model assessment

Six candidate models were assessed, four phenological and two polynomial equations. They were “Verhulst (classical)”<sup>1</sup>, “modified Gompertz”<sup>2</sup>, “Baranyi”<sup>3</sup>, “Buchanan”<sup>4</sup>, “quadratic” and “cubic”. Non-linear least square (NLLS) method was used only on the four phenological models and linear approach was taken on the two polynomials. Starting values selection (for phenological models only) was described below:

Initial (N0) and final (K) population sizes were selected to be the minimum and maximum values of each data subset respectively. Maximum growth rate (r.max) and relative time lag (t.lag) were obtained through recursive linear modelling with shrinking independent range (from maximum and minimum). In this project 5% was chosen as the shrinking threshold assuming this resolution was sufficient for initiating NLLS fits. The final slope value (i.e. r.max) would be a positive finite number with the highest R<sup>2</sup> value (i.e. best-fit slope on considered data). The x-intercept for that slope was the t.lag. Time which this r.max linear model intersected with K was regarded as the time achieving carrying capacity (t.K). Population data was then classified into three groups (gx) according to the time:  $g1 \leq t.lag < g2 < t.K \leq g3$ . Inputs for phenological models were listed below (popn & time were the dependent and independent variables respectively):

Verhulst (classical):  $popn = f(N0, K, r.max, time)$

modified Gompertz:  $popn = f(N0, K, r.max, time, t.lag)$

Baranyi:  $popn = f(N0, K, r.max, time, t.lag)$

Buchanan:  $popn = f(N0, K, r.max, time, t.lag, gx)$

All test starting values were than sampled from normal distribution with mean as the estimated value and standard deviation (sd) of 1. The sd value was chosen because of different reasons for each parameters. N0 and K were directly extracted from the raw experimental

data. Hence a small sd was precise assuming this extraction was an accurate estimate. r.max was a guesstimated value. So a large sd would be sufficient for getting the “true” value with fair accuracy. 100 trials were done as a optimal value under a trade-off between efficiency and accuracy.

AIC<sup>5-7</sup> was used to select optimal parameter values (i.e. meeting the “minimal AIC +2”<sup>8</sup> criteria) within each phenological model and best candidate model(s) for a data subset. All available parameter sets were included in principal component analysis (PCA). AIC tolerance threshold was expanded to  $\min(\text{AIC})+2^8$  to incorporate more accepted models for analyses.

## Statistical analysis

Kruskal test was used for identify the best-fit model among all included model because the count was categorical and not assumed being normally-distributed. Pairwise Nemenyi comparisons would be carried out to identify the best test if p-value of the test was significant.

Parameter weights were assessed across phenological models by PCA R-way analysis method on natural-logged parameter data. Datasets would be expected clustering together if parameter(s) were representing the observed data. Then with Kruskal-Wallis test, each parameter was tested for statistical differences across phenological models. Post-hoc Tukey pairwise comparisons would be carried out upon significance.

## Main Assumptions

- there was no negative population growth throughout experiments from source publications. Data not fitted this assumption were set to zeros;
- all parameter estimates converged to global optimal using NLLS method.

## Computing tools

R (ver 3.6.0)<sup>9</sup> was used with packages “minpack.lm”<sup>10</sup> (NLLS), “stats”<sup>9</sup> (Kruskal test and PCA) and “PMCMR”<sup>11</sup> (post-hoc Nemenyi pairwise comparisons). Python (ver 3.7.3)<sup>12</sup> was used with package “subprocess”<sup>12</sup> (streamline project workflow).

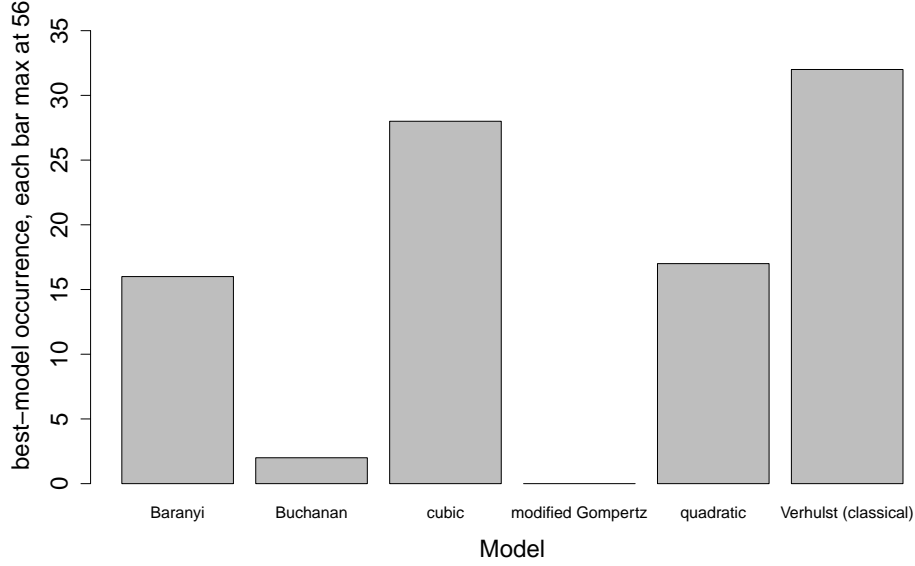


Figure 1: Barplot showing the number of “best model” identification under AIC model-selection methods with “ Kruskal-Wallis rank sum test ” statistic  $X^2 = 5$  ,  $df = 5$  ,  $p = 0.42$

84 From Fig.1, large fluctuations between each model to be described as “best-fit” were observed.  
 85 However the occurrence difference was not statistical significant. Among the counts, there were  
 86 39 datasets with more than one “best-fit” models. Verhulst (classical) and cubic were the top  
 87 two models selected as “best-fit” for the 56 datasets ( 32 for Verhulst (classical) and 28 for  
 88 cubic). There are 9 datasets calling both “best-fit” at the same trial. Between Baranyi and  
 89 quadratic, the counts were 16 and 17 respectively with 7 datasets calling both models “best-fit”.  
 90 The only outstanding performance was from modified Gompertz, which 0 datasets were called  
 91 it as “best-fit”.

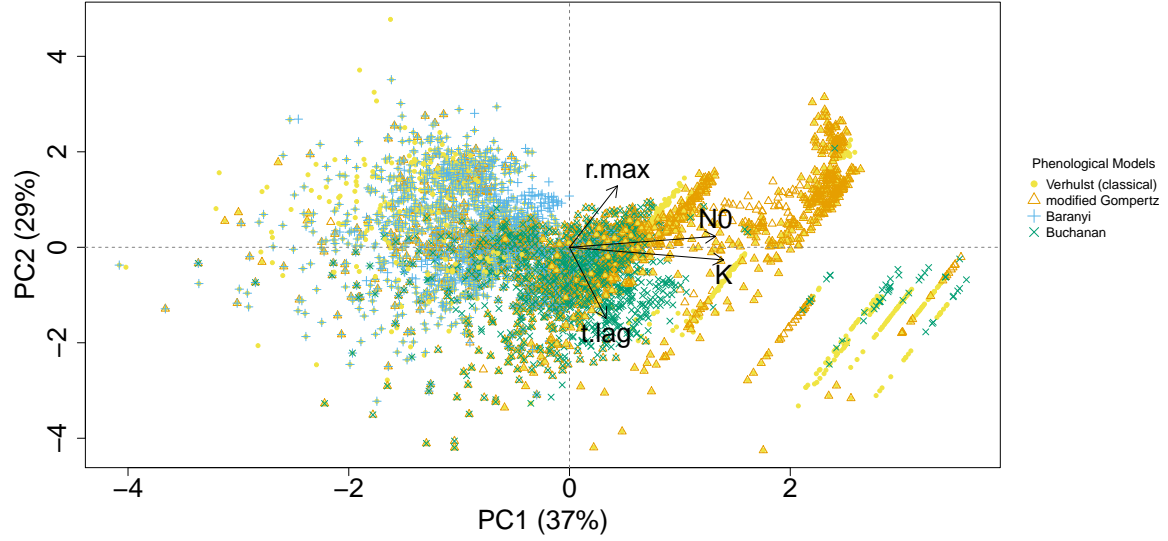


Figure 2: Biplot of Principal Component Analysis (PCA) comparing phenological models using estimated parameter values with “minimal AIC +2”<sup>8</sup> evaluations.

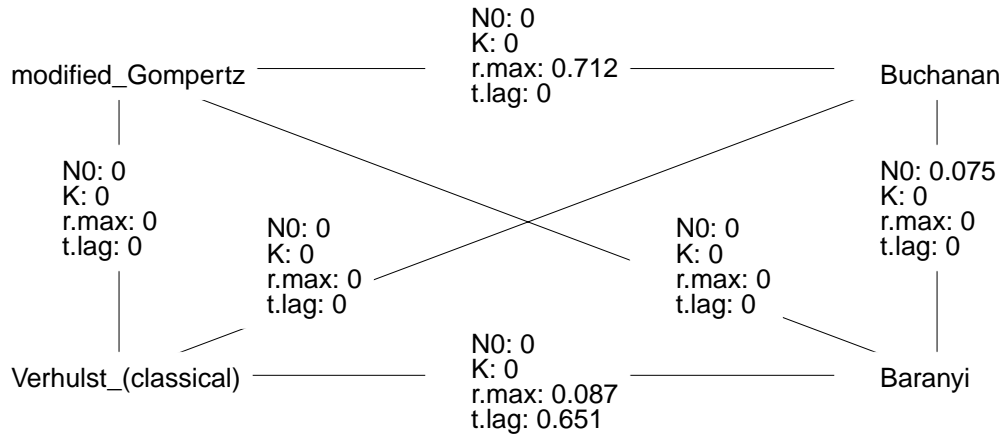


Figure 3: P-value summary between models on the four parameters under post-hoc Tukey-Dist pairwise comparison from Kruskal-Wallis Test. Kruskal tests for all four factors were significant (  $N0 : X^2 = 388.75$  ,  $df = 3$  ,  $p\text{-value} = 0$  ;  $K : X^2 = 1562.4$  ,  $df = 3$  ,  $p\text{-value} = 0$  ;  $r.\text{max} : X^2 = 127.53$  ,  $df = 3$  ,  $p\text{-value} = 0$  ;  $t.\text{lag} : X^2 = 1383.59$  ,  $df = 3$  ,  $p\text{-value} = 0$  ).

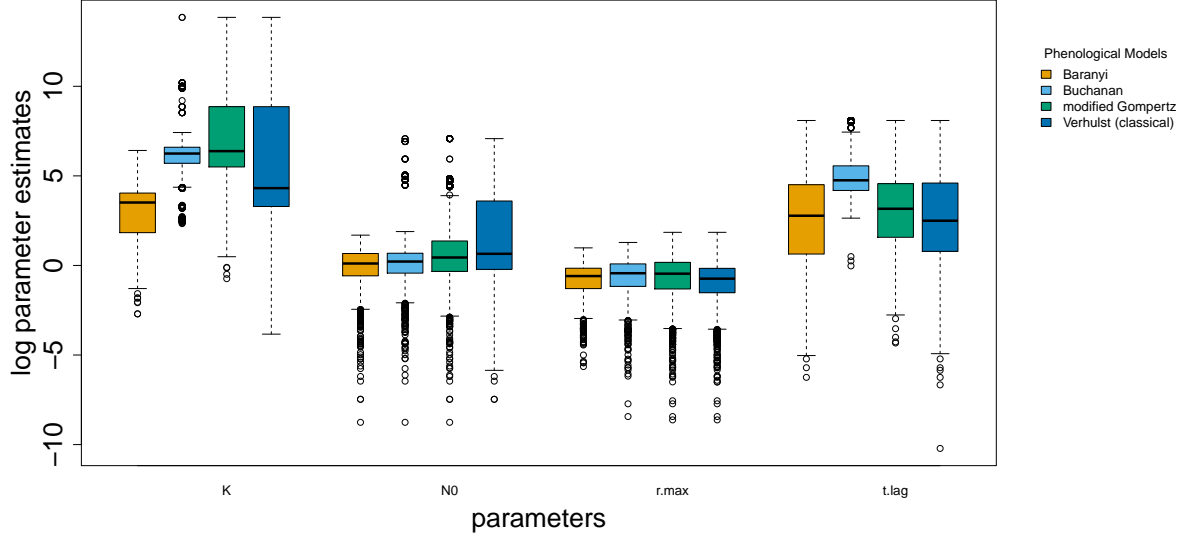


Figure 4: Boxplot of log parameter values grouped by phenological models. Statistical results were summarized in Fig.3

In Fig.2, principal component 1 (PC1) was capturing 37 % variability. It was composed approximately by 0.66 N0, 0.7 K, 0.22 r.max and 0.17 t.lag. PC2 was capturing 29 % variability. It was composed approximately by 0.11 N0, -0.13 K, 0.64 r.max and -0.74 t.lag. There were 51 datasets with phenological models fitting, although they may not be the “best-fit” ones. Datasets 23, 27, 36, 52, 53 were strictly limited to polynomial-fitting (Fig.5). Verhulst (classical) was having the widest neutral coverage across parameter space (Fig.2,4). All other three models (modified Gompertz, Baranyi and Buchanan) were generally modelling within the Verhulst (classical) coverage (Fig.4). modified Gompertz was the second widest coverage model but Verhulst (classical) was evaluated better if both equations fitted the same dataset (Fig.1). More successful trials were towards positive responses for N0, K and r.max. Baranyi was a specific model more specified in describing datasets with negative responses towards most parameter factors (all except r.max). Baranyi had the strictest r.max acceptance for successful NLLS modelling (Fig.4). Buchanan had the narrowest parameter ranges in most parameters (all except r.max, Fig.4). Datasets describable by this model were generally neutral responses towards all four parameters (Fig.2). In the analysis for individual parameters, the parameter value ranges overlapping between phenological models (Fig.4). Hence the differences were not observable although most “differences” were statistically significant (Fig.3).

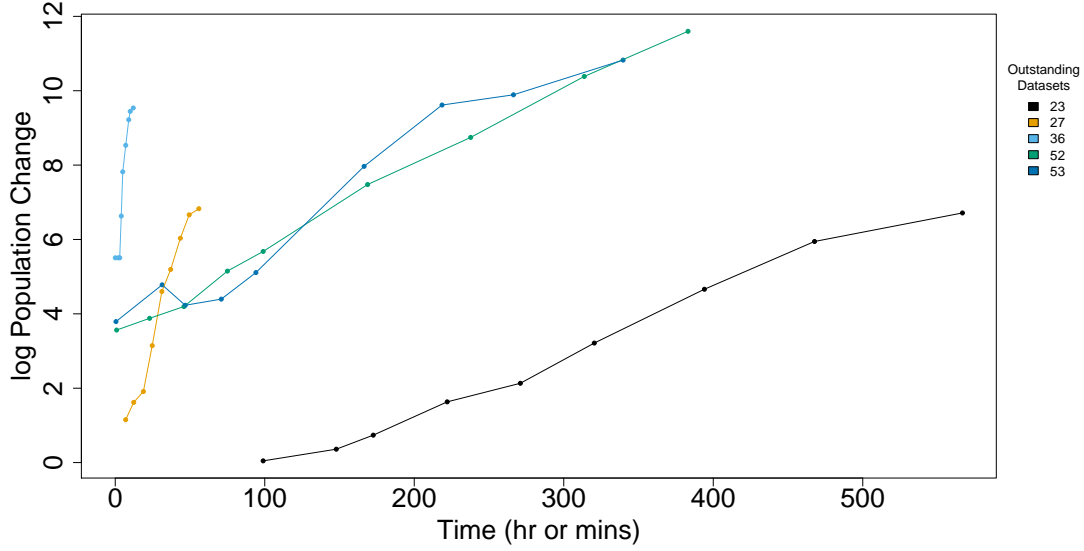


Figure 5: Line plot of datasets restricted to polynomial fits. Dataset details could be found in Table 1

## Discussion

AIC is considered the most suitable model-selection approach within model and between models in this project. Unlike BIC, AIC are more accurate with small sample sizes<sup>13,14</sup> and sparse data<sup>14</sup>. AIC did not assume a “true model” was under examination<sup>15–17</sup>. Since candidate models were not “nested model”, BIC is not a better choice than AIC<sup>18</sup>. Hence the use of only AIC as model-selection criterion should be justified.

Although Baranyi and Buchanan were observed occupying different parameter space (Fig.2), these differences were not statistically significant (Fig.1).

## Conclusion

Published phenological models were data-specific, which none of them were found significantly performing better than the others. Although most of the parameter values are significantly different between models, their ranges are superimposing with one another. Phenological models correlate with parameters differently, but the correlations are unobservable through plotting a log-linear logistic growth curve. There were assumptions embedded within the phenological models which have limited its ability to describe data without a distinct sigmoid shape.



## Code and Data Availability

All scripts and data used for this report were publicly available at GitHub.

## References

1. McKendrick, A. & Pai, M. K. XLV.—the rate of multiplication of micro-organisms: a mathematical study. *Proceedings of the Royal Society of Edinburgh* **31**, 649–653 (1912).
2. Gil, M. M., Brandão, T. R. & Silva, C. L. A modified Gompertz model to predict microbial inactivation under time-varying temperature conditions. *Journal of Food Engineering* **76**. Bugdeath, 89–94. ISSN: 0260-8774. <http://www.sciencedirect.com/science/article/pii/S0260877405003389> (2006).
3. Baranyi, J., McClure, P., Sutherland, J. & Roberts, T. Modeling bacterial growth responses. *Journal of industrial microbiology* **12**, 190–194 (1993).
4. Buchanan, R., Golden, M. & Whiting, R. Differentiation of the effects of pH and lactic or acetic acid concentration on the kinetics of *Listeria monocytogenes* inactivation. *Journal of Food Protection* **56**, 474–478 (1993).
5. Johnson, J. B. & Omland, K. S. Model selection in ecology and evolution. *Trends in ecology & evolution* **19**, 101–108 (2004).
6. Akaike, H. in *Selected papers of hirotugu akaike* 199–213 (Springer, 1998).
7. Burnham, K. & Anderson, D. Model selection and multimodel inference: a practical information-theoretic approach. *Ecological Modelling*.
8. Burnham, K. P. & Anderson, D. R. Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research* **33**, 261–304 (2004).
9. R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria, 2019). <https://www.R-project.org/>.
10. Elzhov, T. V., Mullen, K. M., Spiess, A.-N. & Bolker, B. *minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds* R package version 1.2-1 (2016). <https://CRAN.R-project.org/package=minpack.lm>.

- 151 11. Pohlert, T. *The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR)* R  
152 package (2014). <https://CRAN.R-project.org/package=PMCMR>.
- 153 12. Python Core Team. *Python: A dynamic, open source programming language* Python Soft-  
154 ware Foundation (Delaware, United States, 2019). <https://www.python.org/>.
- 155 13. Acquah, H. D.-G. Comparison of Akaike information criterion (AIC) and Bayesian in-  
156 formation criterion (BIC) in selection of an asymmetric price relationship. *Journal of*  
157 *Development and Agricultural Economics* **2**, 001–006 (2010).
- 158 14. Kuha, J. AIC and BIC: Comparisons of assumptions and performance. *Sociological methods*  
159 *& research* **33**, 188–229 (2004).
- 160 15. Aho, K., Derryberry, D. & Peterson, T. Model selection for ecologists: the worldviews of  
161 AIC and BIC. *Ecology* **95**, 631–636 (2014).
- 162 16. Vrieze, S. I. Model selection and psychological theory: a discussion of the differences  
163 between the Akaike information criterion (AIC) and the Bayesian information criterion  
164 (BIC). *Psychological methods* **17**, 228 (2012).
- 165 17. Yang, Y. Can the strengths of AIC and BIC be shared? A conflict between model inden-  
166 tification and regression estimation. *Biometrika* **92**, 937–950 (2005).
- 167 18. Wang, Y. & Liu, Q. Comparison of Akaike information criterion (AIC) and Bayesian infor-  
168 mation criterion (BIC) in selection of stock–recruitment relationships. *Fisheries Research*  
169 **77**, 220–225 (2006).
- 170 19. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).
- 171 20. Kelley, C. T. *Iterative methods for optimization* (SIAM, 1999).
- 172 21. Turchin, P. *Complex population dynamics: a theoretical/empirical synthesis* (Princeton  
173 university press, 2003).

## 174 Appendix

Table 1: Table showing dataset id details in this project.

id	T°C	clade	substrate	replicate	Source	Pop unit
----	-----	-------	-----------	-----------	--------	----------

1	5	Chryseobacterium balustinum	TSB	1	1	OD_595
2	5	Enterobacter sp	TSB	1	1	OD_595
3	5	Pantoea agglomerans	TSB	1	1	OD_595
4	5	Bacillus pumilus	TSB	1	1	OD_595
5	5	Clavibacter michiganensis	TSB	1	1	OD_595
6	5	Pseudomonas fluorescens	TSB	1	1	OD_595
7	5	Acinetobacter clacoaceticus	TSB	1	1	OD_595
8	5	Stenotrophomonas maltophilia	TSB	1	1	OD_595
9	5	Klebsiella pneumonia	TSB	1	1	OD_595
10	5	Dickeya zeae	TSB	1	1	OD_595
11	5	Pectobacterium carotovorum	TSB	1	1	OD_595
12	15	Chryseobacterium balustinum	TSB	1	1	OD_595
13	25	Chryseobacterium balustinum	TSB	1	1	OD_595
14	35	Chryseobacterium balustinum	TSB	1	1	OD_595
15	5	Tetraselmis tetrahele	ESAW	1	2	N
16	5	Tetraselmis tetrahele	ESAW	2	2	N
17	5	Tetraselmis tetrahele	ESAW	3	2	N
18	5	Tetraselmis tetrahele	ESAW	4	2	N
19	5	Tetraselmis tetrahele	ESAW	5	2	N
20	8	Tetraselmis tetrahele	ESAW	1	2	N
21	16	Tetraselmis tetrahele	ESAW	1	2	N
22	32	Tetraselmis tetrahele	ESAW	1	2	N
23	2	Staphylococcus sp	Raw Chicken Breast	1	3	CFU
24	4	Staphylococcus sp	Raw Chicken Breast	1	3	CFU
25	7	Staphylococcus sp	Raw Chicken Breast	1	3	CFU
26	10	Staphylococcus sp	Raw Chicken Breast	1	3	CFU
27	20	Staphylococcus sp	Raw Chicken Breast	1	3	CFU
28	2	Staphylococcus sp	Salted Chicken Breast	1	3	CFU
29	2	Staphylococcus sp	Cooked Chicken Breast	1	3	CFU
30	2	Pseudomonas sp	Raw Chicken Breast	1	3	CFU
31	2	Aerobic Psychotropic	Raw Chicken Breast	1	3	CFU
32	2	Aerobic Mesophilic	Raw Chicken Breast	1	3	CFU
33	8	Spoilage	Vacuum Beef Striploins	1	4	N
34	8	Escherichia coli	Vacuum Beef Striploins	1	4	N
35	8	Salmonella Typhimurium	Vacuum Beef Striploins	1	4	N

36	10	Spoilage	C02 Beef Striploins	1	4	N
37	12	Spoilage	Vacuum Beef Striploins	1	4	N
38	30	Spoilage	Vacuum Beef Striploins	1	4	N
39	6	<i>Serratia marcescens</i>	Pasteurised Skim Milk	1	5	N
40	6	<i>Serratia marcescens</i>	UHT Skim Milk	1	5	N
41	6	<i>Serratia marcescens</i>	Pasteurised Full-fat Milk	1	5	N
42	6	<i>Serratia marcescens</i>	UHT Full-fat Milk	1	5	N
43	6	<i>Serratia marcescens</i>	Pasteurised Double Cream	1	5	N
44	6	<i>Serratia marcescens</i>	UHT Double Cream	1	5	N
45	0	<i>Arthrobacter</i> sp	TGE agar	1	6	CFU
46	37	<i>Arthrobacter</i> sp	TGE agar	1	6	CFU
47	0	<i>Arthrobacter aureus</i>	TGE agar	1	6	CFU
48	0	<i>Arthrobacter citreus</i>	TGE agar	1	6	CFU
49	0	<i>Arthrobacter globiformis</i>	TGE agar	1	6	CFU
50	0	<i>Arthrobacter simplex</i>	TGE agar	1	6	CFU
51	8	<i>Lactobacillus plantarum</i>	MRS broth	1	7	N
52	4	<i>Weissella viridescens</i>	MRS broth	1	7	N
53	4	<i>Lactobacillus sakei</i>	MRS broth	1	7	N
54	15	<i>Oscillatoria agardhii</i>	Z8	1	8	DryWeight
55	15	<i>Pseudomonas</i> sp	APT Broth	1	9	CFU
56	10	<i>Lactobacillus plantarum</i>	MRS	1	10	N

175 “Source” column publication key:

176

- 1 Bae, Y.M., Zheng, L., Hyun, J.E., Jung, K.S., Heu, S. and Lee, S.Y., 2014. Growth characteristics and biofilm formation of various spoilage bacteria isolated from fresh produce. *Journal of food science*, 79(10), pp.M2072-M2080.
- 2 Bernhardt, J.R., Sunday, J.M. and O'Connor, M.I., 2018. Metabolic theory and the temperature-size rule explain the temperature dependence of population carrying capacity. *The American naturalist*, 192(6), pp.687-697.
- 3 Galarz, L.A., Fonseca, G.G. and Prentice, C., 2016. Predicting bacterial growth in raw, salted, and cooked chicken breast fillets during storage. *Food Science and Technology International*, 22(6), pp.461-474.

- 4 Gill, C.O. and DeLacy, K.M., 1991. Growth of *Escherichia coli* and *Salmonella ty-*  
phimurium on high-pH beef packed under vacuum or carbon dioxide. *International*  
*journal of food microbiology*, 13(1), pp.21-30.
- 5 Phillips, J.D. and Griffiths, M.W., 1987. The relation between temperature and  
growth of bacteria in dairy products. *Food Microbiology*, 4(2), pp.173-185.
- 6 Roth, N.G. and Wheaton, R.B., 1962. Continuity of psychrophilic and mesophilic  
growth characteristics in the genus *Arthrobacter*. *Journal of bacteriology*, 83(3),  
pp.551-555.
- 7 Silva, A.P.R.D., Longhi, D.A., Dalcanton, F. and Aragão, G.M.F.D., 2018. Modelling  
the growth of lactic acid bacteria at different temperatures. *Brazilian Archives of*  
*Biology and Technology*, 61.
- 8 Sivonen, K., 1990. Effects of light, temperature, nitrate, orthophosphate, and bacte-  
ria on growth of and hepatotoxin production by *Oscillatoria agardhii* strains. *Appl.*  
*Environ. Microbiol.*, 56(9), pp.2658-2666.
- 9 Stannard, C.J., Williams, A.P. and Gibbs, P.A., 1985. Temperature/growth relation-  
ships for psychrotrophic food-spoilage bacteria. *Food Microbiology*, 2(2), pp.115-122.
- 10 Zwietering, M.H., De Wit, J.C., Cuppers, H.G.A.M. and Van't Riet, K., 1994. Model-  
ing of bacterial growth with shifts in temperature. *Appl. Environ. Microbiol.*, 60(1),  
pp.204-213.