

NAIST-IS-MT0551010

修士論文

時系列マイクロアレイにおける解析手法の提案

~RSI 法の開発~

伊藤 遼佑

2007 年 2 月 1 日

奈良先端科学技術大学院大学
情報科学研究科 情報生命科学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士(工学) 授与の要件として提出した修士論文である。

伊藤遼佑

審査委員：

金谷重彦教授 (主指導教員)

小笠原直毅教授 (副指導教員)

黒川顕教授 (副指導教員)

時系列マイクロアレイにおける解析手法の提案

～RSI法の開発～

伊藤遼佑

内容梗概

二色蛍光マイクロアレイは、遺伝子全体の発現変化をダイナミックに捉えるツールとして精力的に研究が行われている。この技術は、転写産物の絶対量を定量化するには不向きだが、適切なコントロールサンプルとの対比により、大量の遺伝子に対してその相対変化量を包括的に求めることができる。通常、二色蛍光マイクロアレイで多変量のデータを作成する場合、変量間の比較を容易にするためにコントロールを固定して対数比により評価する。しかし、この実験デザインは、遺伝子発現の変化を連続的に追っていく時系列データを得たい時、精度を落とし望ましくない。本研究では、コントロールを固定するデザインの欠点を大腸菌野生株の時系列マイクロアレイデータを用いて実験的に示した。また、遺伝子発現の連続的な変化をより精度よく測定するために、コントロールを直前の時系列の点にとるマイクロアレイデータを作成し、前述のデザインよりも精度の高いデータが得られることも示した。また、両デザインに基づくデータを統合する方法(RSI法:Reference-Sequential Integration Method)も考案した。

キーワード

マイクロアレイ, タイムシリーズ, 遺伝子発現変化, デザイン, ばらつき

*奈良先端科学技術大学院大学情報科学研究科情報生命科学専攻修士論文,
NAIST-ISMT0551010, 2007年2月1日.

An approach for time-series microarray analysis based on integrated experimental designs

Ryosuke Ito

Abstract

The time-series microarray analyses is a highly effective approach for interpreting the gene functions. Here we focus on the cDNA microarray for time-series analysis. cDNA microarray can make a direct comparison between two mRNA samples. But there are some difficulties when we deal with the microarray data. Firstly, the variance of data is large and its amount is different for each array and even for each gene. So we have to get the reliable data. Secondly, normalization is performed under the assumption that expression ratio of all the genes are almost normally distributed around zero value. If we compare the largely differing expressions, the distribution will spread widely from zero. Therefore, the normalization cannot be carried out successfully. Consequently, designing the appropriate pairing in microarray experiments are crucial for controlling the quality of data. Previously, Yang and Speed proposed the experimental design for time-series microarray analysis and they evaluated synthetically the practicability at the view point of variability, mRNA resources and so on. However the evaluation of experimental design based on real data has not been performed. Therefore, we evaluated two designs, reference design and sequential design, by using experimental data. Further more, we integrated the expression data from each design so that the changes of gene expression are determined correctly. Then we discuss about how our integrated gene expression patterns are suit for time-series microarray analysis.

Keywords:

Microarray, experimental design, time-series, gene expression, data variance

.*Master's Thesis, Department of Bioinformatics and Genomics, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT0551010, February 1, 2007.

目次

1 序論

- 1. 1 背景
- 1. 2 先行研究
- 1. 3 時系列マイクロアレイ実験デザインの問題点
- 1. 4 目的

2 提案手法

- 2. 1 時系列マイクロアレイの問題点と提案手法
- 2. 2 時系列マイクロアレイデータの定義
- 2. 3 区間ごとの最適なデザイン
- 2. 4 区間ごとの変化の判定
- 2. 5 発現系列の構築

3 実験結果と解析結果および考察

- 3. 1 Growth Curve
- 3. 2 マイクロアレイデータ
- 3. 3 データ再現性のデザインによる違い
- 3. 4 リファレンス・デザインとシーケンシャル・デザインの違
い
 - 3. 4. 1 隣り合う2点間の変化
 - 3. 4. 2 両デザインで異なる発現変化を示す領域
- 3. 5 リファレンス・デザインとシーケンシャル・デザインの利
点と欠点
- 3. 6 リファレンス・デザインとシーケンシャル・デザインの統
合
 - 3. 6. 1 両デザインで同じ発現パターンを示す遺伝子
 - 3. 6. 2 デザインの選択

- 3. 6. 3 全ての二点間の変化
- 3. 6. 4 統合時系列パターン
- 3. 7 統合データに基づく rpoS レギュロンの解析
- 3. 8 まとめ
- 3. 9 今後の課題

4 本研究で用いた手法

4. 1 実験手法

- 4. 1. 1 菌株、培地、培養条件および RNA 抽出
- 4. 1. 2 蛍光標識色素のラベリング
- 4. 1. 3 ハイブリダイゼーションとスポット検出
- 4. 1. 4 ウェスタンブロット法

4. 2 実験デザインとデータの前処理

- 2. 2. 1 マイクロアレイ法におけるデータの規格化
- 2. 2. 2 不正確なデータの除去
- 2. 2. 3 Log-ratio の再現性の評価方法
- 2. 2. 4 リファレンス・デザインとシーケンシャル・デザインで似たパターンを検出するのに用いた方法

5 謝辞

6 参考文献

7 学会発表

補足図

付表

1 序論

1. 1 背景

バクテリアは、環境からの刺激に応じて遺伝子の発現状態を制御する機構を持っている。この遺伝子の機能を解明するためのアプローチとして、cDNA マイクロアレイを用いて遺伝子の時系列変化を発現プロファイルとして解析することが有効であると考えられている[Dharmadi and Gonzalez 2004]。

cDNA マイクロアレイは、二つのサンプル間の遺伝子発現状態をほぼ全ての遺伝子について、同時に比較可能な技術である。cDNA マイクロアレイのような二色蛍光型のマイクロアレイでは、コントロールサンプルの mRNA とターゲットサンプルの mRNA からそれぞれ逆転写反応により得られた cDNA に、蛍光標識を付加し、それらを同一スライド上で競合ハイブリダイゼーションさせることにより、mRNA の量を蛍光強度 (intensity) として推定する。競合ハイブリダイゼーションを行うのは、cDNA マイクロアレイでは mRNA の絶対量を数値化できるほど技術的に精度が高くないため、相対的な mRNA の量の比としてしか定量化できないためである。しかも、得られた蛍光強度はその値が大きくなるに依り、ばらつきが大きくなるという特徴がある[Chen *et al.* 1997]。遺伝子の発現量が細胞の状態によって異なるため、厳密には mRNA 量の相対変化を遺伝子間で同じ尺度では扱えない [Durbin and Rocke 2004]。つまり、基準の異なる相対値同士は比べられないという問題である。通常、マイクロアレイデータの解析では、遺伝子間で相対変化を比較しやすくするために、Log 変換が用いられ、コントロールサンプルの遺伝子発現量に対するターゲットサンプルの遺伝子発現量の相対比として表される。また Log 変換には、蛍光強度のばらつきの違いを軽減するという意味もある。[Huber *et al.* 2002]。コントロールを C、ターゲットを T として表すと、この相対比は $\log(T/C)$ として表される。この遺伝子発現量の相対比を、Log-ratio と呼ぶ。しかし、一般に対数オーダーで発現変化を扱っても Log-ratio の値によってそのばらつき方は一定にすることは難しく、特に蛍光強度の低い領域でばらつきの増加が見られる。

二色蛍光型マイクロアレイ解析ではもう一つ重要なことがある。それは、二つの蛍光標識間に蛍光強度の偏向が認められ、これを規格化する必要があることである。この規格化は、比較したサンプル間でほとんどの遺伝子について発現量に違いがないという前提のもとに行う [Dudoit *et al.* 2002]。つまり、Log-ratio は平均が 0 となる正規分布に従うと仮定し、変化量の大きい遺伝子を当該分布から評価することができる。遺伝子発現変化を時系列として捉えるためには、これらの二色蛍光型マイクロアレイ特有の問題点に加え、さらなる問題点を考慮しなければならぬ。

1. 2 先行研究

上記のような特徴を持つ二色蛍光型マイクロアレイであるが、二色蛍光型マイクロアレイを複数用いて遺伝子発現の時系列変化を捉えるためには、実験デザインを設計することが重要であると考えられる。なぜなら、コントロールの取り方が **Log-ratio** 値の質を左右する要因となるためである。ここで実験デザインとは、発現プロファイルを作製する際の変量としてのマイクロアレイ実験を、コントロールサンプルとターゲットサンプルでどのように組み合わせるかを設計することである。さらには繰り返し実験をどれだけ行うかということも含まれる。二色蛍光型マイクロアレイを用いた様々な実験デザインが報告されている[Churchill 2002: Kerr and Churchill 2001: Conway and Schoolnik 2003]が、時系列マイクロアレイに特化した実験デザインについては、Yang と Speed による研究がある[Yang and Speed 2002]。彼らは、変化を見たい点同士が直接マイクロアレイで比較されるような実験デザインを選択するのが良いとしている。例えば、**図 1.1** で **Design I** は T1 からの変化を段階的に見る場合に適しているし、**Design II** は隣同士の点で微妙な変化を見たい場合に適していると述べている。その他にも、さらにデータの正確さをあげるため、複雑なデザインが提案されている。しかし、Yang and Speed 2002 ではばらつきを理論的に比較しているのみで、実際の実験データに基づく評価が行われていない。

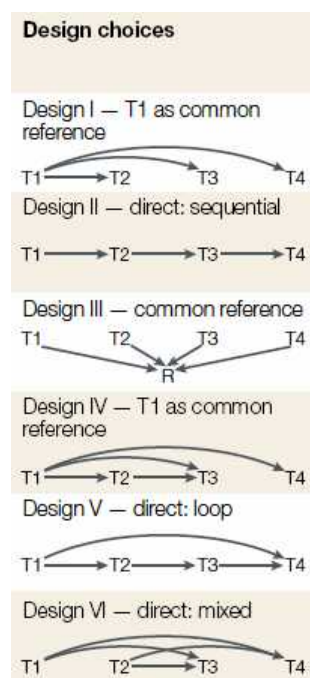


図 1.1 : Yang and Speed 2002 の提案する時系列マイクロアレイ実験デザイン

1.3 時系列マイクロアレイ実験デザインの問題点

我々は、時系列解析で重要なのは、隣り合う二点間の変化であると考え。なぜなら、遺伝子発現の変化は連続的であり[Hakamada *et al.* 2006; Liu *et al.* 2005]、その連続的な変化を追うためには、時々刻々の変化を見る必要があるからである。遺伝子の発現変化は、環境からのストレスに対して、急激に変化するものや徐々に変化するものなど様々な存在すると考えられる。時系列変化のパターンで大切なことは全ての点の関係が決めることだが、遺伝子発現変化のような多様な動きに対応するためには、少なくとも隣り合う二点間の変化を正確にとらえる事が重要である。そこで、本研究では、Yang and Speed 2002 の提案したデザイン(図 1.1)のうち Design II と Design III に着目し、これらのデザインが時系列マイクロアレイ解析を行う際どのような問題点を持ち、どのようなデータの特徴が得られるかを評価する。Design II は隣り合う二点の変化を直接見るデザインとして採用している。Design III は現在最もよく使用されているデザイン[Yang and Speed 2002; Dobbin and Simon 2002]として採用している。ここでは、Yang and Speed 2002 の提案に即して、Design II をシーケンシャル・デザイン(Sequential Design)、Design III をリファレンス・デザイン(Reference Design)と呼ぶことにする。図 1.1 のとおりリファレンス・デザインでは、コントロールとして共通のタイムポイントを使う。シーケンシャル・デザインでは隣り合う 2 点間でハイブリダイゼーションを行い、直前のタイムポイントに対する発現変化を見る。まず、時系列解析にこれらのデザインを用いるにあたり考慮しなければならない問題点は、発現変化を直接的に見るか間接的に見るかという問題に加え、遺伝子発現全体が大きく変遷していくという問題である。本研究では大腸菌(*E. coli* K-12)解析対象とするが、大腸菌は LB 培地で培養していくと、培地内の栄養物量や自らの排泄物などの影響により対数増殖期から変遷期を経て定常期へと遺伝子発現状態を変化させる。ここで、リファレンス・デザインを用いた場合に、コントロール点として対数増殖期を取ると、定常期の発現変化を見るために、大きく異なる発現状態のサンプルを比べなければならないことになる。この場合、得られる各遺伝子の蛍光強度は二つのサンプルで大きく異なることになり、従って前述のとおり、Logratio のばらつきを増大させる原因になると考えられる。そのために、リファレンス・デザインを用いて隣り合う二点間の変化を比べる場合、データの精度が低いために変化を検出できない可能性が考えられる。データのばらつきの増加は、変化を間接的に見るときにも生じるが、時系列解析の場合コントロールサンプルとターゲットサンプルの組み合わせもばらつきを増大させる重要な因子である。それに対して、シーケンシャル・デザインでは近接する二点の間で発現状態を直接比べるので、この問題は軽減されると考えられる。また、発現状態の大きく異なるサンプル間でハイブリダイゼーションした場合、前述のデータの規格化における問題がより顕著になると考えられる。なぜなら、規格化をするにあたり全体的に発現変化がなく Logratio 値がゼロを中心に分布するということが仮定しているため、発現状態の大きく異なるサンプルを比べた場合は、この仮定に反するために規格化がうまく行えないという危険性があるからである[Engelen *et al.* 2006]。さらに、蛍光強度の値が小さい場合、つまり mRNA が少ししか存在しない場合は、Logratio の値は大きくばらつく [Durbin *et al.* 2002]。

そのため、蛍光強度の微弱なものやマイナスになるものは、解析に使用できない値として排除される。リファレンス・デザインの共通コントロールサンプルにおいて、蛍光強度が小さいか検知されなかった遺伝子については、培養の後半で発現が上昇してきてもデータとして検出すら出来ないという問題も考えられる。

一方で、時系列プロファイルの扱いにも様々な手法が提案されている[Bar-Joseph *et al.* 2004]。時系列データを ANOVA モデルに適用するもの[Park *et al.* 2002]、時系列変化を曲線で近似するアプローチ[Hakamada *et al.* 2006; Liu *et al.* 2005; Storey *et al.* 2005]、時系列プロファイルから UP と DOWN の関係を抽出して扱うなどのアプローチ[Willbrand *et al.* 2005; Peddada *et al.* 2003] など様々な方法がある。しかし、細胞周期のように長期間の時系列データとは異なり、比較的短い期間での少ないサンプリングポイントしか少ない遺伝子発現プロファイルの解析手法は確立されているとは言えない[Liu 2005]。しかも、マイクロアレイの価格が年々下がっているといっても、大規模な時系列発現解析を行うのにマイクロアレイ実験数はまだ重要な因子である[Ernst *et al.* 2005]。従って、より少ないタイムポイント数で正確な遺伝子発現の時系列変化を捉えられるデザインが求められる。

1. 4 目的

本研究の目的は二色蛍光マイクロアレイを用いて時系列マイクロアレイ解析を行ううえでの最適なアプローチを提案することである。そのため、異なる二つの実験デザインに基づいて得られたマイクロアレイデータがどのような性質を持つのかを統計的に評価する。本研究で用いる時系列マイクロアレイデータは、大腸菌野生株 K-12 を Luria-Bertani(LB)培地で培養した時の遺伝子発現変化についてである。そして、二つの実験デザインの特徴を生かし、双方の欠点を補完しあった時系列遺伝子発現変化パターン構築法及び、解析法を提案する。それにより、遺伝子機能解析に貢献できると考える。

2 提案手法

2.1 時系列マイクロアレイの問題点と提案手法

二色蛍光マイクロアレイを用いて時系列な発現変化を捉えるためには、様々な技術的な課題を克服しなければならない。まず、二色蛍光法では遺伝子発現を二つの状態間で相対的に比較しているという点に着目する。同一マイクロアレイ上で遺伝子同士の変化を比べようとした場合、遺伝子ごとに発現量は異なるため、遺伝子ごとの相対変化量(Log-ratio)は比べることができない。また、変化量の小さい領域、つまり Log-ratio がゼロ付近の値は信頼性が低く、微妙な変化なのか測定時の変動なのか見分けが付きにくい。しかも、発現量が小さい場合に Log-ratio が大きくばらつき、数値としての信頼性が低くなるという問題もある。従って、二色蛍光法では通常 Log-ratio の数値には意味を求めず、信頼ある Log-ratio に対してある閾値以上の変化を持つものをふるいわけして、顕著に発現変化した遺伝子を推定する。このような技術を用いて連続的な変化を捉えるためには適切な実験デザインを設計し、データの特性に合った解析法を用いることが不可欠である。

時系列の実験デザインは、Yang and Speed 2002 によって提案されているが、ここではその中からリファレンス・デザインとシーケンシャル・デザインに着目した。リファレンス・デザインは時系列の全ての変量間で競合ハイブリダイゼーションのコントロールサンプルを一点に固定している。それによって、遺伝子ごとに変量間が比べることができ、mRNA の資源的な制約も少ないなどの利点から広く用いられるデザインとなっている。しかし、連続的な変化、つまり隣り合う二点間の変化が間接的にしか比べられないという問題点もある。一方、シーケンシャル・デザインは隣り合う二点間の変化を直接競合ハイブリダイゼーションにより見積もる。そのために、微妙な変化を捉えるには適していると考えられている。しかし、このデザインでは全ての実験対でコントロールが異なるという問題点がある。

そこで、本研究ではリファレンス・デザインとシーケンシャル・デザインを併用することにより時系列解析の信頼性向上を目指すことを考えた。提案した手法は、発現系列（時系列データ）における全ての二点間の変化を区間ごとに、より適したデザインで評価することで信頼性を向上させている。また、二点間の変化を UP、DOWN、NOT-CHANGE の三つの状態に集約することで、確実な変化といえるデータのみを使用し、変化の起きた区間を明確に発現系列に反映することができている。これによって、急激な変化とゆるやかな変化を異なるものとして発現系列に反映することができると考え

る。さらに、発現系列の全ての二点対の関係をより適したデザインを用いて評価しているため、遺伝子発現の変化の様子が発現系列に反映され、近い二点間の関係だけでなく遠い二点間の関係も保証されている。通常データの信頼性を上げるためにデータフィルタリングを厳しく行くと、発現プロファイルの欠損値が多くなり多変量解析を行うことが困難になる。逆に、データフィルタリングを行わないと発現プロファイルの欠損値は減るが、信頼性が低くなってしまふ。提案手法は、信頼性の高いデータのみを使用して発現系列を構築できるためこれらの問題点も解消することができるものであると考える。

以下に提案手法の概要を示す。

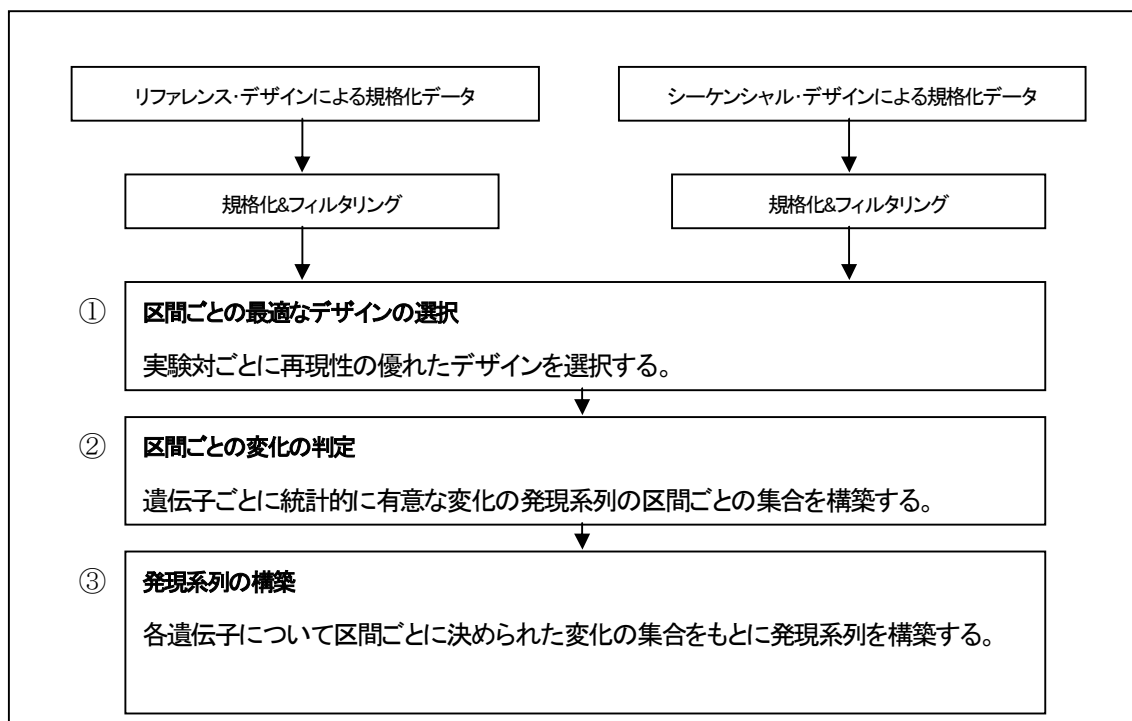


図 2.1 : 提案手法の概要

2.2 時系列マイクロアレイデータの定義

タイムポイント数が T の時系列マイクロアレイ解析について考える。タイムポイントは i または j によって表す。また、この時系列解析では、 $X_1 \cdots X_T$ の時系列サンプル及びコントロールサンプル C が用いられるとする。ここで2つの時系列マイクロアレイ実験デザインを定義する。

(1) リファレンス・デザイン (Reference Design)

コントロールサンプルを C に固定して競合ハイブリダイゼーションを行う。リファレンス・デザインで得たマイクロアレイデータは $MR = (mr_1, \dots, mr_i, \dots, mr_T)$ と表す。従って、このデザインで得られる遺伝子 g のマイクロアレイ mr_i における n 番目の測定に対する相対対数比(Log-ratio)は $r_{ngi} = \log_{10}(X_{ngi} / C_{ng})$ と表される。さらにノーマライズ後のLog-ratioを r'_{ngi} として表す。 \bar{r}_{gi} は繰り返し測定内のLog-ratioの平均値であり、 $\bar{r}_{gi} = \sum_n r'_{ngi} / N$ となる。 N は繰り返し測定の数である。 \bar{r}_{gi} を用い、リファレンス・デザインを用いた遺伝子 g の発現プロファイル(ベクトル)は $R_g = (\bar{r}_{g1}, \bar{r}_{g2}, \dots, \bar{r}_{gT})$ となる。

(2) シーケンシャル・デザイン (Sequential Design)

隣り合う近接した2点間で競合ハイブリダイゼーションを行う。シーケンシャル・デザインで得たマイクロアレイデータは $MS = (ms_2, \dots, ms_i, \dots, ms_T)$ と表す。従って、このデザインで得られる遺伝子 g のマイクロアレイ ms_i における n 番目の測定に対する相対対数比(Log-ratio)は $s_{ngi} = \log_{10}(X_{ngi} / X_{ngi-1})$ と表される。さらにノーマライズ後のLog-ratioを s'_{ngi} として表す。 \bar{s}_{gi} は繰り返し測定内のLog-ratioの平均値であり、 $\bar{s}_{gi} = \sum_n s'_{ngi} / N$ となる。 N は繰り返し測定の数である。 \bar{s}_{gi} を用い、シーケンシャル・デザインを用いた遺伝子 g の発現プロファイル(ベクトル)は $S_g = (\bar{s}_{g2}, \bar{s}_{g3}, \dots, \bar{s}_{gT})$ となる。

また、それぞれのデザインにおいて、遺伝子の総数を G とする。解析時に値を持つもののみの遺伝子の総数が必要なときは G' を用いる。図 2.2 は、リファレンス・デザインとシーケンシャル・デザインを模式的に表したものである。

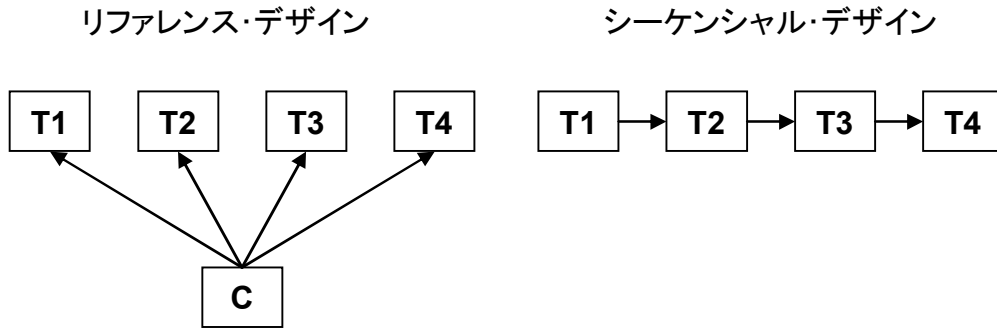


図 2.2 : 時系列マイクロアレイ実験デザイン

2.3 区間ごとの最適なデザイン

次に、タイムポイントの区間ごとの最適なデザインを定義する。区間 $i \sim j$ の変化を評価する際の最適なデザインを ED_{ij} ($ED_{ij} = ED_{ji}$) とし、全ての区間で最適なデザインが選択できたとする。 ED_{ij} の取りうる値は $\{R, S\}$ であり、それぞれリファレンス・デザイン、シーケンシャル・デザインを表している。この時、区間の距離 p ごとに最適なデザインを表すベクトル $ED(p)$ が定義できる。

$$ED(p) = \{ED_{1,1+p}, \dots, ED_{i,i+p}, \dots, ED_{T-p,T}\} \quad (2-1)$$

例えば、距離 $p = 2$ の最適なデザインのベクトル $ED(2)$ は、

$$ED(p) = \{S, \dots, R, \dots, S\}$$

というように値をとる。

最適なデザインの選択は以下の方法を取った。

ある区間 $i \sim j$ を考えたとき、リファレンス・デザインとシーケンシャル・デザインの相対変化量はそれぞれ次のように見積もられる。

$$r_{ngij} = r'_{ngi} - r'_{ngj} \quad (2-2)$$

$$s_{ngij} = \sum_{t=i+1}^j s'_{ngt} \quad (2-3)$$

さらに、それぞれデザインで遺伝子 g ごとの相対変化量の標準偏差が次のように計算される。

$$SD(r)_{gij} = \sqrt{\frac{\sum_n (r'_{ngij} - \bar{r}'_{gij})^2}{N-1}} \quad (2-4)$$

$$SD(s)_{gij} = \sqrt{\frac{\sum_n (s'_{ngij} - \bar{s}'_{gij})^2}{N-1}} \quad (2-5)$$

そして、欠損値を含まない全ての遺伝子について平均値を求める。

$$ave(r)_{ij} = \left(\sum_g SD(r)_{gij} \right) / G' \quad (2-6)$$

$$ave(s)_{ij} = \left(\sum_g SD(s)_{gij} \right) / G' \quad (2-7)$$

$ave(r)_{ij}$ および $ave(s)_{ij}$ は、リファレンス・デザインおよびシーケンシャル・デザインの相対変化量データの再現性を示しており、区間 $i \sim j$ のデザインの選択をこれらの値に基づいて判定する。つまり、 $ave(r)_{ij} \leq ave(s)_{ij}$ の場合はリファレンス・デザイン ($ED_{ij} = R$)、 $ave(r)_{ij} > ave(s)_{ij}$ の場合はシーケンシャル・デザイン ($ED_{ij} = S$)、を選択する。この処理を、すべての区間 $i \sim j$ ($i < j$) について行い、最適なデザインを決定する。

2.4 区間ごとの変化の判定

次に遺伝子 g の区間 $i \sim j$ ($i < j$) での変化を D_{gij} と定義する。 D_{gij} は3つの値{UP, DOWN, NOT}を取りうる。これらの値はそれぞれ、タイムポイント i から j へ発現上昇、発現減少、変化なしという変化の様子を表している。変化の判定は区間 $i \sim j$ ごとに判定されたデザイン ED_{ij} の相対変化量を用いて t 検定を用いて行う。以下に、それぞれのデザインごとに判定法の詳細を説明する。

(1) リファレンス・デザインにおける変化の判定

リファレンス・デザインにおいて区間 $i \sim j$ ($i < j$) の変化を判定するには、Welch の異なる分散を考慮した場合の t 検定を用いる。リファレンス・デザインにおける Welch の修正 t 値は

$$t_{gij} = \frac{\bar{r}_{gj} - \bar{r}_{gi}}{\sqrt{\frac{v_{gj}}{N_{gj}} + \frac{v_{gi}}{N_{gi}}}} \quad (2-8)$$

となる。ここで、 v_{gi} はタイムポイント i のノーマライズ済み Log-ratio r'_{ngi} の分散である。 N_{gi} は欠損値を除いた繰り返し測定数であり、どちらかのタイムポイントで N_g が3に満たないとき

は検定を行わない。この t 値を用いて自由度 df の t 分布から片側 p 値を算出しその有意性を判定し変化の状態を決定する。自由度 df は

$$df = \frac{\left(\frac{v_{gj}}{n_{gj}} + \frac{v_{gi}}{n_{gi}} \right)^2}{\frac{\left(\frac{v_{gj}}{n_{gj}} \right)^2}{n_{gj}-1} + \frac{\left(\frac{v_{gi}}{n_{gi}} \right)^2}{n_{gi}-1}} \quad (2-9)$$

でもとまる。 p 値の有意水準を th とし有意性の判定を行う。さらに相対変化量 $r_{gij} = \bar{r}_{gj} - \bar{r}_{gi}$ の信頼区間を $r_{gij} \geq upth$ または $r_{gij} \leq downth$ とする。従って、区間 $i \sim j$ ($i < j$) でのリファレンス・デザインに基づく変化の値は

$$D_{ij} = \begin{cases} UP & : p \leq th \cap r_{gij} \geq upth \\ DOWN & : p \leq th \cap r_{gij} \leq downth \\ NOT & : others \end{cases} \quad (2-10)$$

として割り当てられる。

(2) シーケンシャル・デザインにおける変化の判定

シーケンシャル・デザインにおいて区間 $i \sim j$ ($i < j$) の変化を判定するには、比べるタイムポイント間の距離によって評価の方法が異なる。いずれも、相対変化量のデータ数が 3 に満たない時は、検定を行うことが出来ない。

i) 距離 1

まず、距離が 1 の場合、つまり $(i-1) \sim i$ の変化を見る場合、遺伝子 g の相対変化量は \bar{s}_{gi} である。変化の有意性を判定するための t 値は

$$t_{g(i-1),i} = \frac{\bar{s}_{gi}}{\sqrt{\frac{v_{gi}}{n_{gi}}}} \quad (2-11)$$

で表される。ここで、 v_{gi} はタイムポイント i のノーマライズ済み Log-ratio s'_{ngi} の分散である。そして、この t 値を用いて一標本 t 検定を行い有意な発現変化かどうかを p 値に基づき判定する。判定基準は以下である。

$$D_{(i-1),i} \begin{cases} UP & : p \leq th \cap \bar{s}_{gi} \geq upth \\ DOWN & : p \leq th \cap \bar{s}_{gi} \leq downth \\ NOT & : others \end{cases} \quad (2-12)$$

ii) 距離 2

距離が 2 の場合、つまり $(i-1) \sim (i+1)$ の変化を見る場合、リファレンス・デザインと同様の Welch の t 検定を用いる。 t 値は

$$t_{g(i-1),(i+1)} = \frac{\bar{s}_{g_i} + \bar{s}_{g_{i+1}}}{\sqrt{\frac{v_{g_i}}{n_{g_i}} + \frac{v_{g_{i+1}}}{n_{g_{i+1}}}}} \quad (2-13)$$

である。相対変化量は $s_{g(i-1),(i+1)} = \bar{s}_{g_i} + \bar{s}_{g_{i+1}}$ で表され、変化の判定基準は以下である。

$$D_{(i-1),(i+1)} = \begin{cases} UP & : p \leq th \cap s_{g(i-1),(i+1)} \geq upth \\ DOWN & : p \leq th \cap s_{g(i-1),(i+1)} \leq downth \\ NOT & : others \end{cases} \quad (2-14)$$

iii) 距離 3 以上

距離が 3 以上で p の場合、つまり $(i-1) \sim (i+p-1)$ の場合遺伝子 g の相対変化量は

$$s_{ng(i-1),(i+p-1)} = \sum_{t=i}^{i+p-1} s'_{ngt} \quad (2-15)$$

である。

この場合の t 値は

$$t_{g(i-1),(i+p-1)} = \frac{\bar{s}_{g(i-1),(i+p-1)}}{\sqrt{\frac{v_{g(i-1),(i+p-1)}}{n_{g(i-1),(i+p-1)}}}}} \quad (2-16)$$

である。ここで、 $\bar{s}_{g(i-1),(i+p-1)}$ は $s_{ng(i-1),(i+p-1)}$ の平均値である。 $n_{g(i-1),(i+p-1)}$ は欠損値を含まない相対変化量の測定数であるが、相対変化量 $s_{ng(i-1),(i+p-1)}$ の計算時に s'_{ngt} が欠損値を含むと

$s_{ng(i-1),(i+p-1)}$ は欠損値となる。このときの相対変化量は $s_{g(i-1),(i+p-1)} = \left(\sum_n s_{ng(i-1),(i+p-1)} \right) / N$ で

あるので、変化の基準は

$$D_{(i-1),(i+p-1)} = \begin{cases} UP & : p \leq th \cap s_{g(i-1),(i+p-1)} \geq upth \\ DOWN & : p \leq th \cap s_{g(i-1),(i+p-1)} \leq downth \\ NOT & : others \end{cases} \quad (2-17)$$

として判定される。

これにより、遺伝子 g の全ての区間 $i \sim j$ ($i < j$) での変化の集合 D_g が求まる。

$$D_g = \{D_{g12}, D_{g13}, \dots, D_{g1T}, D_{g23}, D_{g24}, \dots, D_{g2T}, \dots, D_{gT-1T}\} \quad (2-18)$$

また、この変化の集合 D_g は区間の距離 p ごとの部分集合から成り、 $D_g = \{D_g(1), D_g(2), \dots, D_g(T-1)\}$ で表される。各部分集合は

$$D_g(p) = \{D_{g1,1+p}, D_{g2,2+p}, \dots, D_{gt,t+p}, \dots, D_{gT-p,T}\} \quad (2-19)$$

であり、 $D_g(p)$ のサイズは $T-p$ である。これらの変化の集合は、区間ごとに最適なデザインのもとに構築されるため、より信頼の高いものとなっていると考えられる。

2.5 発現系列の構築

遺伝子 g の全ての区間 $i \sim j$ ($i < j$) での変化の集合 D_g をもとに、 D_g を満たす発現系列 E_g を構築する。発現系列を構築するにあたり留意する点を次にまとめる。

1. 全ての区間の変化の関係をくまなく取り入れた発現系列を構築する。
2. より近い区間の変化を優先する。
3. 発現系列が整数格子上で形成されるようにする。

全ての点の関係を考慮することで、発現系列が mRNA の時間変化を正確に表せるようになるを考える。簡単のためこれ以降、添え字 g は省略する。よって、発現系列は $E = \{e_1, e_2, \dots, e_T\}$ で表される。ここでは変化の遷移点が整数格子状を動くように発現系列を更新する。そのために、更新の基点となる点を $B = \{b_1, b_2, \dots, b_T\}$ で定義し、 b_i が 1 のとき基点、0 のとき基点ではないとする。そして、基点が整数格子状を動くという制約を本アルゴリズムでは課す。発現系列および基点の初期値は $e_1 = e_2 = \dots = e_T = 0$ 、 $b_1 = b_2 = \dots = b_T = 0$ とする。さらに、更新時の制約条件を定義する。上側制限を各タイムポイントで定義し、そのベクトルを $Lu = \{u_1, u_2, \dots, u_T\}$ とする。初期値は $u_1 = u_2 = \dots = u_T = -1$ とし、制約がないことを表す。例えば、 $u_3 = 1$ である時、発現系列のタイ

ムポイント3はタイムポイント1以上になれないという制約があることを意味している ($e_3 < e_1$ を満たす)。下側制限を同様に $Ld = \{d_1, d_2, \dots, d_T\}$ で定義する。これらの制約条件は、点と点の関係を崩す更新を行わないようにするために設定している。

A. 初期発現系列の形成

初期発現系列の形成アルゴリズムは、発現系列の基本となる骨組みを、より細かい区間の変化の情報を優先して構築する。アルゴリズムを図 2.3 に示す。

```

flag = true


= 1:
while flag = true :
    flag = false
    for i = 1 to  $T/p$ :
        if  $e_{i+p} - e_{i+p-1} = \dots = e_{i+1} - e_i = 0$ :
            flag = true
            if  $D_{i,i+p} = UP$ : (A)
                for  $t = i+1$  to  $T$  do:
                    if  $t < i+p$ : (A-1)
                         $e_t = e_t + (t-i)/p$ : (A-2)
                    else (A-2)
                         $e_t = e_t + 1$ 
                 $b_i = b_{i+p} = 1$  (A-3)
                 $d_{i+p} = i$  (A-4)
            if  $D_{i,i+p} = DOWN$ : (B)
                for  $t = i+1$  to  $T$  do:
                    if  $t < i+p$ :
                         $e_t = e_t - (t-i)/p$ :
                    else
                         $e_t = e_t - 1$ 
                 $b_i = b_{i+p} = 1$ 
                 $u_{i+p} = i$ 
         $p = p + 1$ 
        if  $p \geq T$ : (C)
            flag = false :
return  $E$  (D)


```

図 2.3 : 初期パターン構築アルゴリズム

図 2.3 に示したアルゴリズムでは、初めに発現系列が距離 $p = 1$ の変化情報を示す部分集合 $D(1)$ を満たすように E を組み上げ、さらに基点 B を設定するという操作を行っている。そして、順次区間距離 p を増やして行き発現系列を構築する。

タイムポイント i と $i+1$ ($p = 1$) を考えたとき、発現変化の情報は $D_{i,i+1}$ であるので、 $e_{i+1} - e_i$ が $D_{i,i+1}$ を満たすような発現系列の形成が可能である。初期発現系列の形成では、 $e_{i+1} - e_i = 0$ (発現系列が平坦の状態) である場合に、 $D_{i,i+1} = UP$ の時、 $e_{i+1} \sim e_T$ を上方向に平行移動することによりその関係を発現系列に反映する(A)。アルゴリズムでは、離れた基点間の更新量を考慮しなければな

らず、変更する点 i と点 $i+p$ の間の更新量を求めている(A-1)。それ以外の場合の更新は (A-2) で行う。さらに (A-3) では、基点を設定している。このように基点を設定するのは、区間の距離によらず有意な変化を一律に扱うためである。また、上向きに変更をした場合、(A-4) のように下側制限を設定する。つまり、発現系列の点 $i+p$ は点 i 以下になることが出来ない ($e_{i+p} > e_i$ を満たす) ようにし、更新により点と点の関係が崩れないようにしている。 $D_{i,i+1} = DOWN$ の時 $e_{i+1} \sim e_T$ を下方向に平行移動する事により同様にタイムポイント i と $i+1$ の関係を反映する(B)。

さらに、今度は p を増やし(C)、部分集合 $D(2)$ 、つまり、区間の距離が2であるものについて考える。タイムポイント i と $i+2$ の区間を考えた時、 $e_{i+1} - e_i = e_{i+2} - e_{i+1} = 0$ (発現系列 E の区間 $i \sim i+2$ が平坦) に対し、 $D_{i,i+2}$ の情報を反映させなければならない。 $D_{i,i+2} = UP$ の時、 $e_{i+2} \sim e_T$ を上方向に平行移動し、 $D_{i,i+2} = DOWN$ の時 $e_{i+2} \sim e_T$ を下方向に平行移動する。

このプロセスを $e_{i+p} - e_{i+p-1} = \dots = e_{i+1} - e_i = 0$ を満たす区間が存在しなくなる区間距離 p まで、同様に続ける。そして、初期発現系列 E を得る(D)。

初期発現系列の形成例

ここでは、 $T = 11$ である場合の発現系列 $E = \{e_1, e_2, \dots, e_{11}\}$ を例にとり、初期発現系列の形成アルゴリズムについて説明を行う。初期化時の発現系列は図 2.4 のようになる。発現系列の各点は初期状態で全てゼロで、基点 b_i も設定されていない状態である。

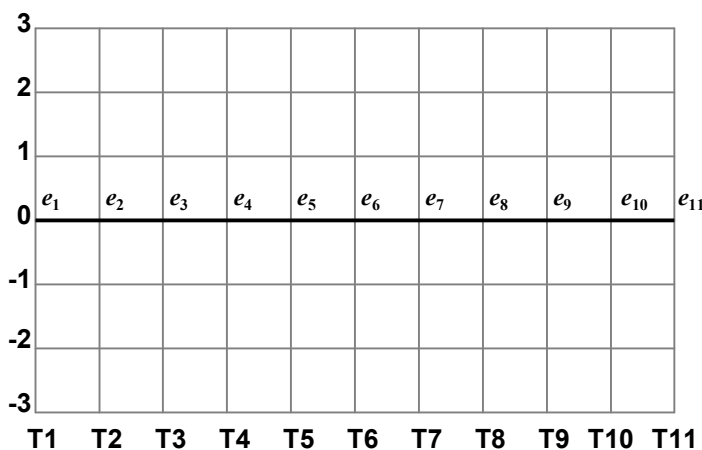


図 2.4 : 初期発現系列の構築 1

初期発現系列形成アルゴリズムでは、まず区間距離が 1 の変化の集合 $D(1)$ について考える。

$D(1) = \{N, N, U, N, N, N, N, D, N, N\}$ であった時、発現系列は図 2.5 の様に更新される。更新は $D_{3,4} = UP$ 、 $D_{8,9} = DOWN$ を基に区間 3-4 で行われており b_3 と b_4 が基点として設定されている (図 2.5 中丸で示す)。この時、下側制約も新たに設定される (図 2.5T 字で示す)。区間 8-9 についても同様の変更が行われている。

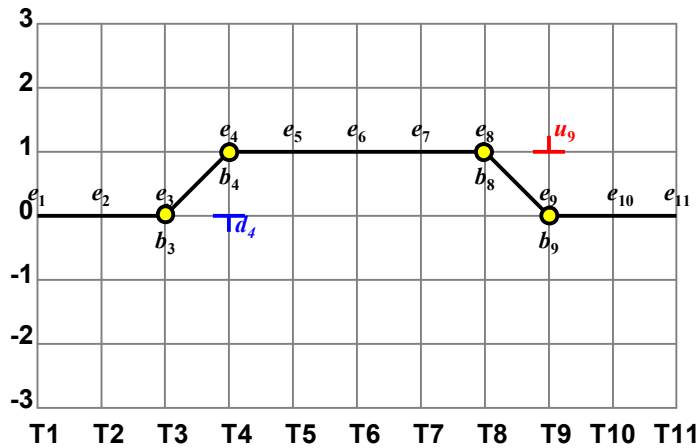


図 2.5 : 初期発現系列の構築 2

区間距離が 2 以上の平坦な区間 ($e_{i+1} - e_i = e_{i+2} - e_{i+1} = 0$) があるので、 $D(2)$ についても更新を行う。 $D(2) = \{N, U, U, U, N, N, D, N, U\}$ であった時、発現系列は図 2.6 の様に更新される。図 2.6 では $D_{4,6} = UP$ 、 $D_{9,11} = UP$ を基に区間 4-6 及び 9-11 について更新されており、制約条件も付加されている。

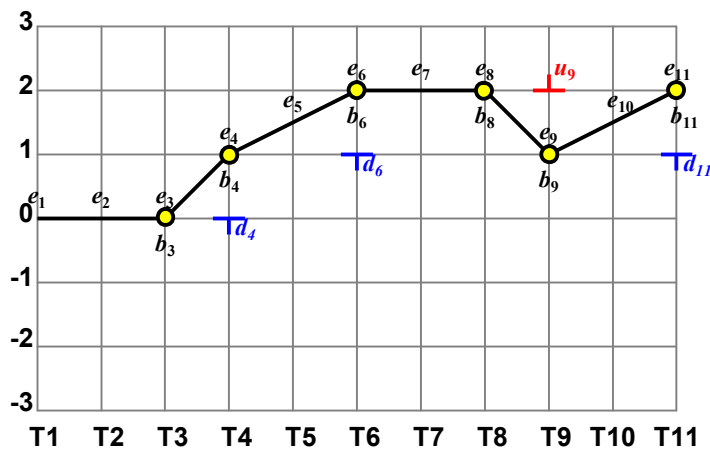


図 2.6 : 初期発現系列の構築 3

次に距離が3以上の平坦な区間($e_{i+1} - e_i = e_{i+2} - e_{i+1} = e_{i+3} - e_{i+2} = 0$)はもうないので、ここで初期発現系列構築アルゴリズムは終了し、初期発現系列を完成とする。ここで設定された、基点(図 2.6 中に○で示した点)が整数格子上を動くように次の更新アルゴリズムは行われる。

B. 初期発現系列からの更新

初期発現系列の構築では、いわば発現系列の骨組みを作った。次に、この骨組みを全ての変化の情報を満たすように組み上げていく処理を行う。この処理は、(1)更新の開始点の選択、(2)更新量の算出、(3)発現系列の更新、(4)制約条件ベクトルの更新、の四つのプロセスで構成される。区間 $i \sim i+p$ ごとに発現系列が $D_{i,i+p}$ を満たしているかどうかを検証し、満たしていない場合は、それぞれ $D_{i,i+p} = UP$ 、 $D_{i,i+p} = DOWN$ を満たすように発現系列の更新を行う。 $D_{i,i+p} = UP$ を満たすように発現系列を上向きに更新するアルゴリズムを図 2.7 に示す。


```

if  $D_{i+i_p} = \text{UP}$  and  $e_{i+p} \cdot e_i \leq 0$ : (a)
  for  $t = i+p-1$  to  $i+1$ :
    if  $b_t = 1$  and  $e_{t+1} \cdot e_t > 0$ : (b)
       $start = t$ 
       $end = \text{next base point}(start)$ 
      if  $end = \text{right edge}$ :
         $end = T$ 
       $w = 1$  (c)
      if  $(start < i+p < end)$  and  $(b_{i+p} = 0)$ : (c-1)
        do  $w++$ ,  $dif = \frac{e_{i+p} - e_{start}}{e_{end} - e_{start}} \times w$ :
          until  $(e_{i+p} + dif) > e_i$ :
        else: (c-2)
          do  $w++$  until  $(e_{i+p} + w) > e_i$ :
           $TE = E$  (d)
           $next = \text{next base point}(start)$ 
          for  $f = start + 1$  to  $T$ :
            if  $u_f = -1$ : (d-1)
              if  $f < next$ : (d-1.1)
                 $te_f = te_f + \frac{f - start}{next - start} \times w$ 
              else: (d-1.2)
                 $te_f = te_f + w$ 
            else: (d-2)
              if  $te_f + w > te_{u_f}$ : (d-2.1)
                if  $b_f = 1$ : (d-2.1.1)
                   $n = \text{previous base point}(\emptyset)$ 
                  for  $s = n$  to  $f-1$ :
                     $te_s = te_s + \frac{f - s}{f - n} \times w$ 
                  break
                else: (d-2.1.2)
                   $n = \text{previous base point}(\emptyset)$ 
                   $m = \text{next base point}(\emptyset)$ 
                  if  $m = \text{right edge}$  and  $b_m = 0$ : (d-2.1.2.1)
                     $m = n$ 
                     $n = \text{previous base point}(\emptyset)$ 
                    for  $s = n$  to  $m-1$ :
                       $te_s = te_s + \frac{m - s}{m - n} \times w$ 
                    break

```

図 2.7 : 上昇方向への発現系列の更新アルゴリズム

<pre> else: flag= true for s= n to m-1: $te_s = te_s + \frac{m-s}{m-n} \times w$ if $u_s \neq -1$ and $te_s > te_{u_s}$: flag= false break if flag= true: break else: m= n n= previous base point (n) for s= n to m-1: $te_s = te_s + \frac{m-s}{m-n} \times w$ for s= m to T: $te_s = e_s$ break else: $te_f = te_f + w$ if next base point (start) \neq next base point (f) and $f \geq i+p$: E= TE break if $D_{i,i+p} = UP$: if $d_{i+p} = -1$: $d_{i+p} = i$ else if $e_{d_{i+p}} < e_i$: $d_{i+p} = i$ </pre>	<p>(d-2.1.2.2)</p> <p>(d-2.1.2.2.1)</p> <p>(d-2.1.2.2.2)</p> <p>(d-2.2)</p> <p>(e)</p> <p>(f-1)</p> <p>(f-2)</p>
--	--

図 2.7 : 上昇方向への発現系列の更新アルゴリズム(続き)

タイムポイント i から $i+p$ へ発現上昇を示す ($D_{i,i+p} = UP$) にも関わらず、発現系列で e_i が e_{i+p} に等しいかそれより大きい時 ($e_i \geq e_{i+p}$) (a)、更新を行う。それ以外の場合は、すでに発現系列が $D_{i,i+p} = UP$ を満たしているなのでこの区間での更新は行わない。

ここで、 e_{i+p} を e_i よりも大きくしたいので、このアルゴリズムではすでに上向きで基点が設定された区間 ($start \sim end$) に着目する(b)。 $i+p$ に最も近いそのような点を探索し、 $start$ とする。更新は $start$ を開始点として制約条件 Lu 、 Ld を考慮しながら更新を行う。更新量は w として e_{i+p} が e_i より大きくなる最小の整数値となるように求める(c)。アルゴリズムでは、比較対象の点 $i+p$ が基

点でない場合(c-1)と、基点の場合(c-2)とで、それぞれ更新量を求めている。

次に、開始点 $start$ と更新量 w の基で、発現系列のコピー $TE = \{te_1, te_2, \dots, te_T\}$ を用いた更新を考える(d)。上向きに更新する場合、上側制限 u_f を超えない限り右側の点にかけて更新ができる。まず、点 f に上側制限が設定されていない場合($u_f = -1$)(d-1)、点 f について更新を行う。この時、 $start$ 点の次の基点から更新量 w だけ上側に平行移動するので、 $start$ 点から次の基点までは更新量を内挿補完してもとめ(d-1.1)、それ以降は更新量 w 更新する(d-1.2)。上側制限が設定されている場合(d-2)は、上側制限を越えない限り更新を続け(d-2.2)、上側制限を超える時更新を終了する(d-2.1)。上側制限を越えた点 f が基点である時(d-2.1.1)は、前の基点から点 f まで更新して終了する。点 f が基点でなかった場合(d-2.1.2)は、点 f の両側の基点 n, m を考える。 m が発現系列の右端の点でかつ基点でない時(d-2.1.2.1)、更新はさらに n より一つ手前の基点までとする。そうでない場合は、 n までの更新を考え、 $n \sim m$ の更新時に上側制限を越えるか超えないかを判定し、超えない時は(d-2.1.2.2.1)で、超える時は(d-2.1.2.2.2)で更新を行う。

$start \sim f$ の更新を発現系列のコピー TE を用いて考えたが、ここで f の範囲を検証する。 f が $i + p$ より発現系列の右側の点で、かつ $start$ と f の間に一つ以上基点が存在する時、元の発現系列 E を更新する(e)。この条件に合わない場合、次の $start$ の候補について同様の処理を行い(b)発現系列 E が更新できるから、 $start$ の候補がなくなるまで続ける。

最後に、下側の制限は $D_{i,i+p} = UP$ の時に必ず検証し、更新の必要があれば新たな制限を設定する(f)。

下向き ($D_{i,i+p} = DOWN$) に更新を行う場合は、上向きの場合と逆向きに同様の処理を施す(図 2.8)。変化なし ($D_{i,i+p} = NOT - CHANGE$) の情報が与えられた時は、更新は行わない。

更新アルゴリズムを区間距離の短いほうから順に ($p = 1 \rightarrow T - 1$)、適用していくことにより全ての区間での変化の情報を反映した発現系列を構築する。

```

if  $D_{i+i_p} = \text{DOWN}$  and  $e_{i+i_p} - e_i \geq 0$ :
  for  $t = i + p - 1$  to  $i + 1$ :
    if  $b_t = 1$  and  $e_{t+1} - e_t < 0$ :
       $start = t$ 
       $end = \text{next base point}(start)$ 
      if  $end = \text{right edge}$ :
         $end = T$ 
       $w = -1$ 
      if  $(start < i + p < end)$  and  $(b_{i+p} = 0)$ :
        do  $w^-$ ,  $dif = \frac{e_{i+p} - e_{start}}{e_{end} - e_{start}} \times w$ :
          until  $(e_{i+p} + dif) < e_i$ 
        else:
          do  $w^-$  until  $(e_{i+p} + w) < e_i$ 
         $TE = E$ 
         $next = \text{next base point}(start)$ 
        for  $f = start + 1$  to  $T$ :
          if  $d_f = -1$ :
            if  $f < next$ :
               $te_f = te_f + \frac{f - start}{next - start} \times w$ 
            else:
               $te_f = te_f + w$ 
          else:
            if  $te_f + w < te_{u_f}$ :
              if  $b_f = 1$ :
                 $n = \text{previous base point}(\emptyset)$ 
                for  $s = n$  to  $f - 1$ :
                   $te_s = te_s + \frac{f - s}{f - n} \times w$ 
                break
              else:
                 $n = \text{previous base point}(\emptyset)$ 
                 $m = \text{next base point}(\emptyset)$ 
                if  $m = \text{right edge}$  and  $b_m = 0$ :
                   $m = n$ 
                   $n = \text{previous base point}(\emptyset)$ 
                  for  $s = n$  to  $m - 1$ :
                     $te_s = te_s + \frac{m - s}{m - n} \times w$ 
                  break

```

図 2.8 : 減少方向への発現系列の更新アルゴリズム

```

else:
    flag= true
    for s= n to m-1:
         $te_s = te_s + \frac{m-s}{m-n} \times w$ 
        if  $d_s \neq -1$  and  $te_s < te_{d_s}$ :
            flag= false
            break
    if flag= true:
        break
    else:
        m= n
        n= previous base point (n)
        for s= n to m-1:
             $te_s = te_s + \frac{m-s}{m-n} \times w$ 
        for s= m to T:
             $te_s = e_s$ 
        break
else:
     $te_f = te_f + w$ 
    if next base point (start)  $\neq$  next base point (f) and  $f \geq i+p$ :
         $E = TE$ 
        break
if  $D_{i,i+p} = \text{DOWN}$ :
    if  $u_{i+p} = -1$ :
         $u_{i+p} = i$ 
    else if  $e_{u_{i+p}} > e_i$ :
         $u_{i+p} = i$ 

```

図 2.8 : 減少方向への発現系列の更新アルゴリズム(続き)

例

前例の初期発現系列をさらに更新を行う例をここでは示す。変化の集合が次のように構成されていた場合を考える。

$$\begin{aligned}
 D(1) &= \{D_{12}, D_{23}, D_{34}, D_{45}, D_{56}, D_{67}, D_{78}, D_{89}, D_{910}, D_{1011}\} = \{N, N, U, N, N, N, N, D, N, N\} \\
 D(2) &= \{D_{13}, D_{24}, D_{35}, D_{46}, D_{57}, D_{68}, D_{79}, D_{810}, D_{911}\} = \{N, U, U, U, N, N, D, N, U\} \\
 D(3) &= \{D_{14}, D_{25}, D_{36}, D_{47}, D_{58}, D_{69}, D_{710}, D_{811}\} = \{U, U, U, N, N, D, D, N\} \\
 D(4) &= \{D_{15}, D_{26}, D_{37}, D_{48}, D_{59}, D_{610}, D_{711}\} = \{U, U, U, D, D, D, D\} \\
 D(5) &= \{D_{16}, D_{27}, D_{38}, D_{49}, D_{510}, D_{611}\} = \{U, N, U, D, D, D\} \\
 D(6) &= \{D_{17}, D_{28}, D_{39}, D_{410}, D_{511}\} = \{N, N, N, D, D\} \\
 D(7) &= \{D_{18}, D_{29}, D_{310}, D_{411}\} = \{N, D, N, D\} \\
 D(8) &= \{D_{19}, D_{210}, D_{311}\} = \{D, D, N\} \\
 D(9) &= \{D_{110}, D_{111}\} = \{D, D\} \\
 D(10) &= \{D_{111}\} = \{D\}
 \end{aligned} \tag{2-20}$$

初めに変化の情報 $D_{i,i+p}$ と発現系列の矛盾が生じるのは区間 4-8 である。 $D_{4,8} = DOWN$ であるので、下向きの更新を考える。しかし区間 4-8 に下向きの区間がないのでここでの更新は行わない(図 2.9)。図 2.9 では、四角枠で囲った点が比較の対象となっている点を表している。

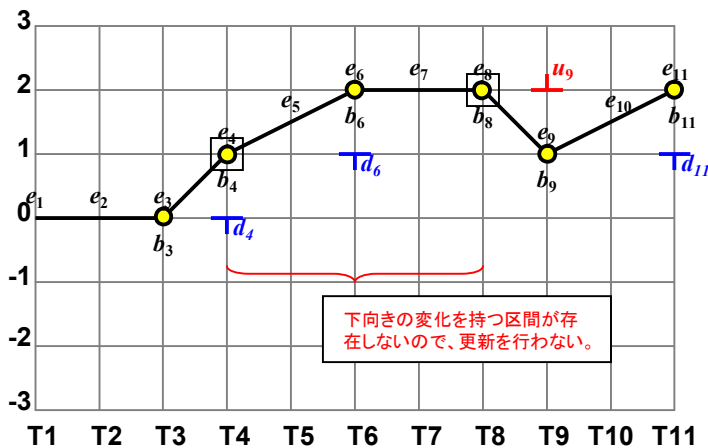


図 2.9 : 更新の例 1

順次変化の情報と発現系列を照らし合わせて行き、次に矛盾が生じるのは区間 7-11 である。
 $D_{7,11} = DOWN$ であるので、下向きを更新を考える。区間 7-11 の間に下向きの区間 8-9 があるの
 でここから更新を行う。制約条件を加味すると、更新区間は 8-11 となる。更新量は $w = 1$ である。
 更新の様子を図 2.10 に示す。

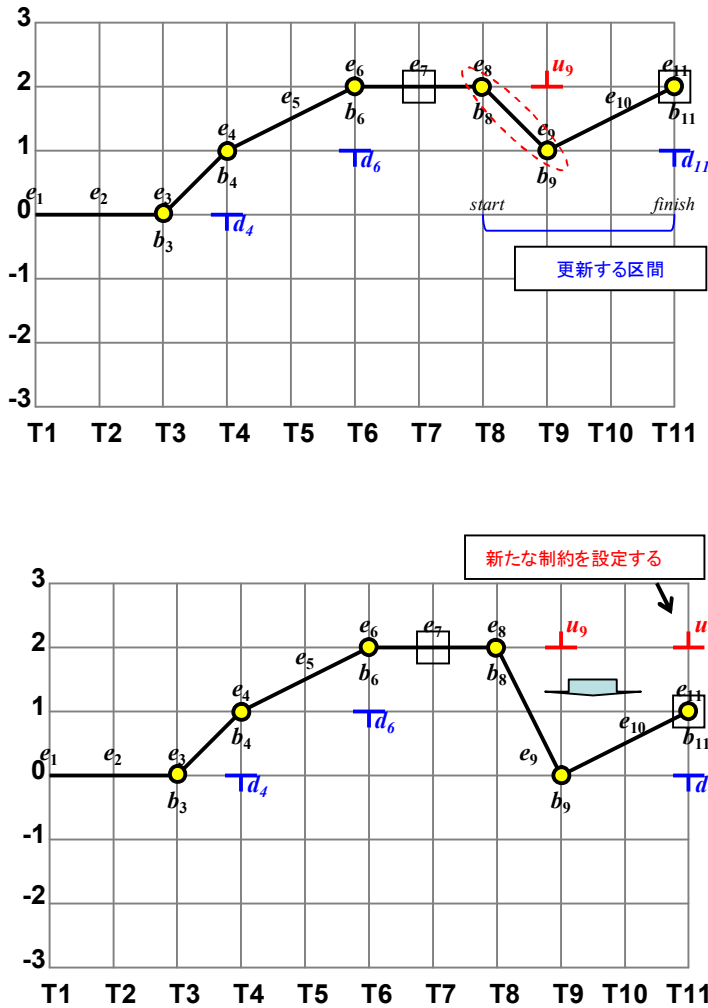


図 2.10 : 更新の例 2

次に更新の対象となる区間は 2-9 であり同様に更新を行うと図 2.11 のようになる。この時、下側
 制約 $d_{11} = 9$ は更新と共に自動で動く。また、上側制約 u_9 の更新が行われている。

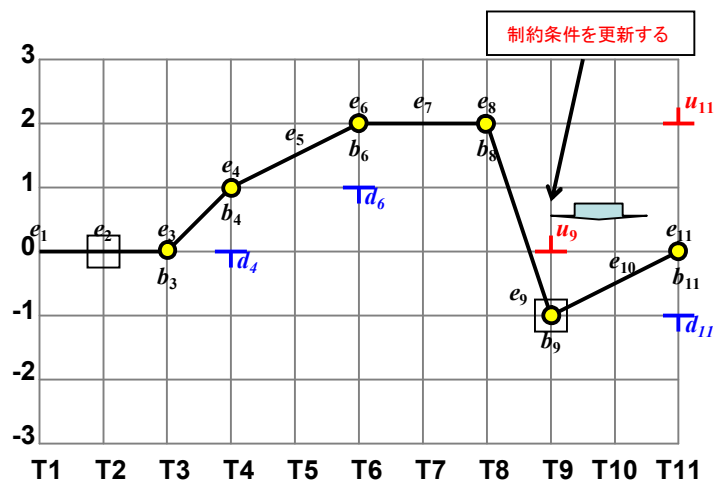
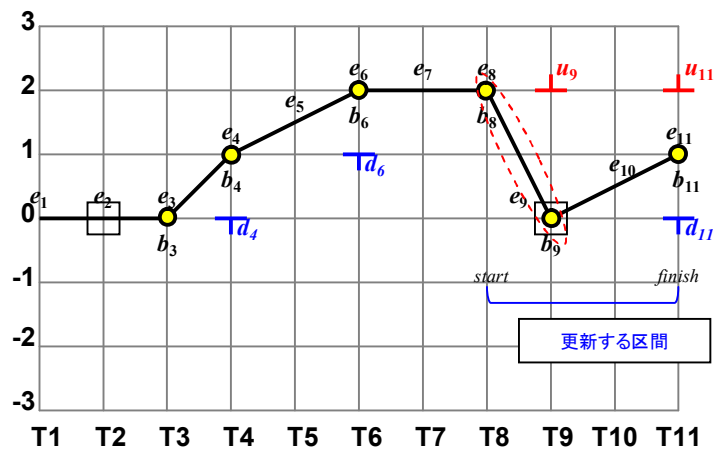


図 2.11 : 更新の例 3

さらに、更新を続けると区間 2-11 で更新ができる(図 2.12)。

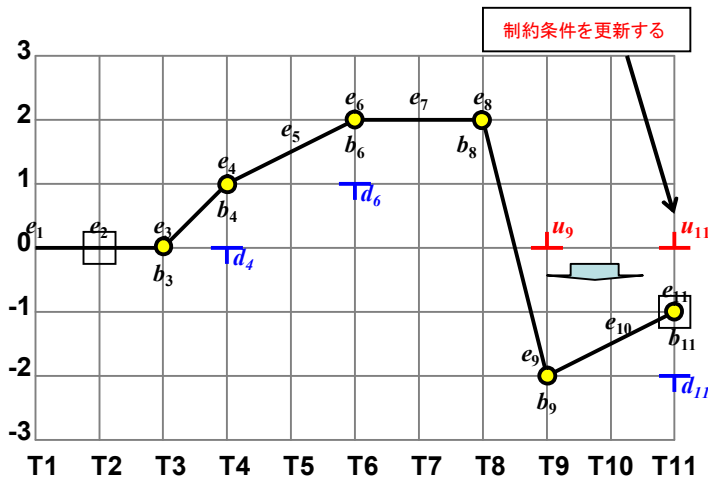
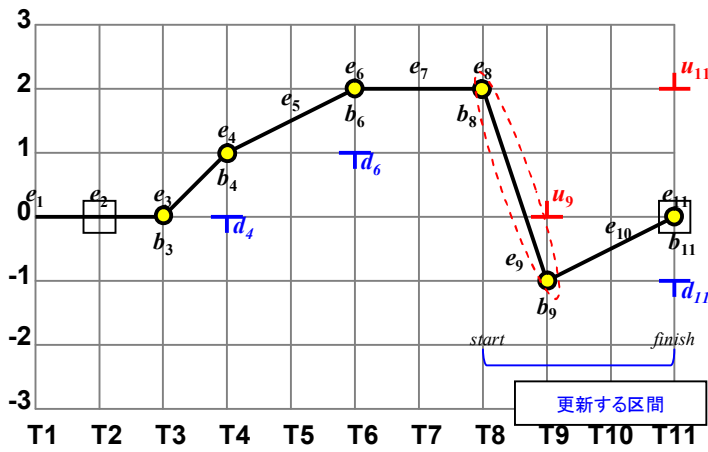


図 2.12 : 更新の例 4

全ての区間において、変化の情報の検証及び更新を終えると発現系列の完成である。

3 実験結果と解析結果および考察

本章では、解析を行うためのデータ取得についてまず説明し、得られたデータに基づいてリファレンス・デザインとシーケンシャル・デザインの特性を評価する。そして、本研究で提案した手法を適用し考察を行う。

3. 1 Growth Curve

解析する対象は、*E.coli*の野生株 K-12 を 20 の LB 培地で培養した系である。培養条件は、37°C一定、攪拌速度 300rpm、で、30 ジャーファーマンターを用いて培養した。表 3.1 のようにサンプリングを行った結果図 3.1 の Growth Curve を得た。表 3.1 で ID はマイクロアレイ実験に用いたサンプルのタイムポイントを表す。タイムポイントは図 3.1 と対応している。表 3.1 のサンプリングポイントは対数増殖期後期から定常期へかけて遺伝子発現が大きく劇的に変わると考え、恣意的に決めた。図 3.1 には培地の pH の変化も合わせて示してある。pH の変化を見ると、T6 と T7 の間で切り替えが起こっており、菌体内でなんらかの変化が起こっていることが予想される。

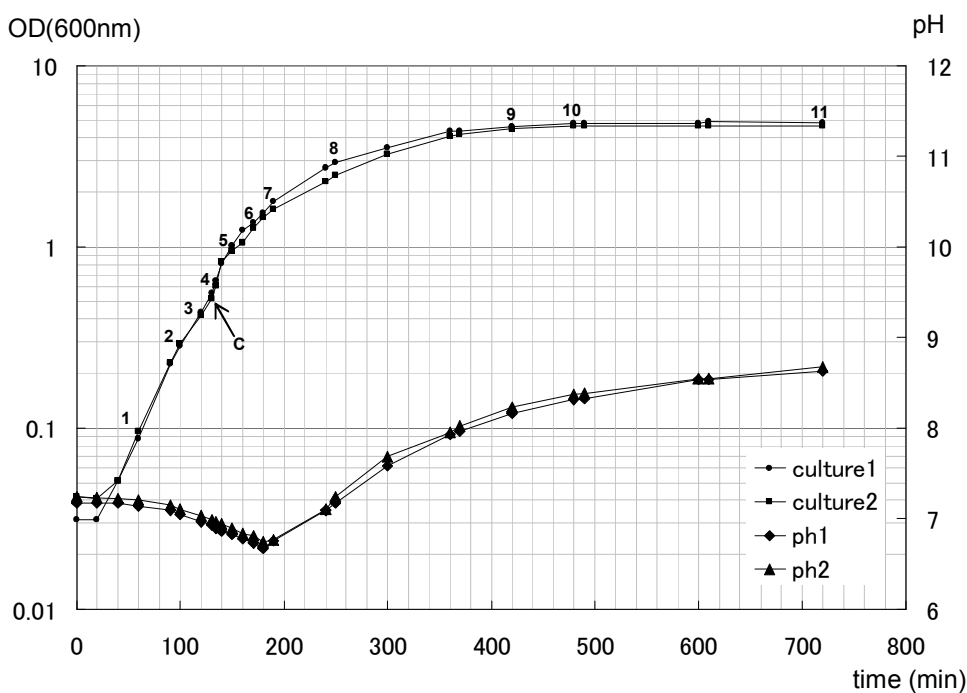


図 3.1 : Growth Curve

表 3.1 : サンプルングポイント

ID	culture 1			culture 2	
	min	od	ph	od	ph
	0	0.031	7.18	0.042	7.24
	20	0.031	7.18	0.041	7.23
	40	0.051	7.17	0.051	7.22
T1	60	0.087	7.14	0.096	7.21
	90	0.227	7.09	0.228	7.15
T2	100	0.285	7.05	0.29	7.1
T3	120	0.435	6.97	0.42	7.04
C	130	0.554	6.93	0.517	6.99
T4	135	0.649	6.89	0.606	6.96
	140	0.81	6.86	0.83	6.94
T5	150	1.02	6.83	0.95	6.9
	160	1.24	6.78	1.05	6.84
T6	170	1.36	6.73	1.27	6.81
	180	1.54	6.68	1.45	6.75
T7	190	1.79	6.76	1.61	6.77
	240	2.73	7.09	2.28	7.1
T8	250	2.9	7.18	2.48	7.25
	300	3.51	7.58	3.25	7.69
	360	4.34	7.93	4.07	7.95
	370	4.36	7.96	4.19	8.02
T9	420	4.62	8.16	4.46	8.23
T10	480	4.82	8.31	4.64	8.37
	490	4.8	8.33	4.68	8.38
	600	4.8	8.53	4.67	8.55
	610	4.89	8.54	4.65	8.55
T11	720	4.85	8.63	4.66	8.67

3. 2 マイクロアレイデータ

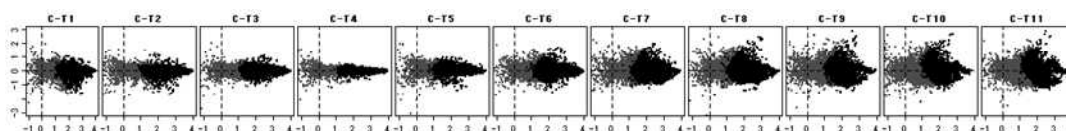
図 3.1 のように、Growth Curve 上の 11 のタイムポイントについて適時、適量サンプリングし、リファレンス・デザインとシーケンシャル・デザインそれぞれのデザインに基づいてマイクロアレイ実験を行った。繰り返し実験回数は 2 回である。その結果、リファレンス・デザインについては 11 枚のマイクロアレイデータが 2 セット。シーケンシャル・デザインにおいては 10 枚のマイクロアレイデータが 2 セット。計 42 枚のマイクロアレイデータを得た。今回の実験で得たデータの概観図を補足図 1 に示す。得られた、全てのマイクロアレイに対して、MA-plot に基づくマイクロアレイデータの規格化(4.2.1 節)を施した。

図 3.2 に補正後の MA-plot を時間軸方向に並べて示す。各 MA-plot で灰色の点は微小な蛍光強度の除去(4.2.2 節(1))により省かれた点を示している。各プロットにおいて縦軸は Log-ratio、横軸は蛍

光強度の大きさを表している。図 3.2 を見て分かるように、リファレンス・デザインでは、タイムポイントが後になるほど、MA-plot のちらばりが大きくなっていくという特徴が見て取れる。リファレンス・デザインでは、各点におけるマイクロアレイで、固定のコントロールサンプルとの発現変化を見ているので、一度変化が起こると、次の点ではその変化を保持した上で新たな変化を捉えなければならない。つまり、遺伝子発現状態の大きく異なる対数増殖期と定常期を比較したような場合には、Log-ratio のふれ幅が全体的に大きく y 軸方向に広がっている。従って、データの規格化時にベースラインが正しく定まらないため、Log-ratio が 0 付近の値を実験間で比べる時に間違った結果を生む危険性がある。一方で、シーケンシャル・デザインでは、隣り合う二点間の変化を直接マイクロアレイで比較するので、図 3.2 下段のような縦軸方向にひろがりの少ない MA-plot が得られる。そのため、マイクロアレイデータの規格化の問題を軽減できると考えられる。

また、繰り返し実験間で整合性の取れないデータの除去についての結果を補足図 2、3 に示す。これより、蛍光強度の小さい領域に Log-ratio の大きくばらつくデータが多く存在することが分かる。また、ばらつきの大きなデータが排除され、残ったデータ全体にわたりばらつき方が均一化されていることが分かる。従って、今後の解析ではデータの除去後に残ったデータについて評価、及び解析を行う。

リファレンス・デザイン



シーケンシャル・デザイン

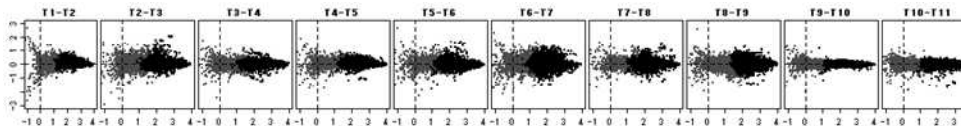
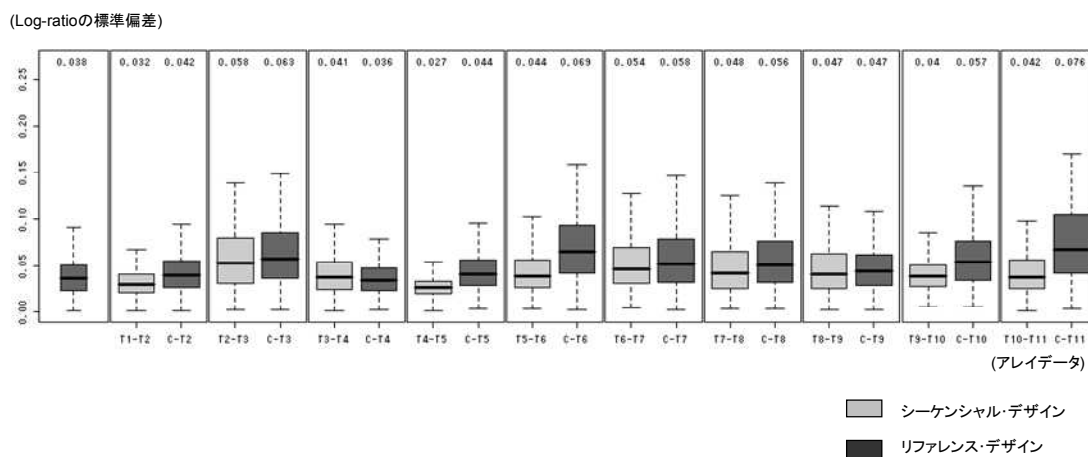


図3.2: マイクロアレイデータのMA-plotによる図示

3. 3 データ再現性のデザインによる違い

本節では、Log-ratio の再現性がリファレンス・デザインとシーケンシャル・デザインでどのように異なるかを比較した結果について示す。図 3.3 は、各遺伝子について求めた Log-ratio 値の標準偏差がマイクロアレイデータセットごとにどのように分布するかを、箱ヒゲ図を用いて表したものである。箱ヒゲ図は、データの第一四分位点と第三四分位点を示す箱の中に、第二四分位点(中央値)を示し、第一四分位点と第三四分位点の差の 1.5 倍の領域を箱の両側にとりその中に入るデータの最大値または最小値をヒゲとして上下に付加したものである。ヒゲの外側に来るデータは外れ値としてプロットするが、ここでは省略している。箱ヒゲ図の各枠内で左側がシーケンシャル・デザインに対応し、右側がリファレンス・デザインに対応している。図 3.3 より、リファレンス・デザインでの ratio 値のばらつきは、シーケンシャル・デザインの Log-ratio 値のばらつきに比べ、大きくなる傾向にあることが伺える。また、リファレンス・デザインでは、コントロールポイントから離れると Log-ratio 値のばらつきが徐々に大きくなる傾向も見て取れる。つまり、Log-ratio 計算時の蛍光強度の大きさが比較状態間で大きく異なる場合に、ばらつきがより大きくなることがわかる。また、シーケンシャル・デザインでも変化が大きく起こった[T2-T3]や[T6-T7]のデータでばらつき方が大きくなっている。従って、実験デザインの違いが、得られるデータの質にも影響を与えていることが示された。このようなばらつき方の増加は統計的な解析に影響を与えると考えられる。



Log-ratio分散の平均とその標準偏差の箱ヒゲ図。中心の太線は分布の中央値を示す。Outlierは図示していない。それぞれのタイムポイントで、左側がシーケンシャル・デザイン、右側がリファレンス・デザインに対応する。各枠の上には、標準偏差の平均値が記してある。

図3.3: Log-ratioの箱ヒゲ図(標準偏差の分布)

3. 4 リファレンス・デザインとシーケンシャル・デザインの違い

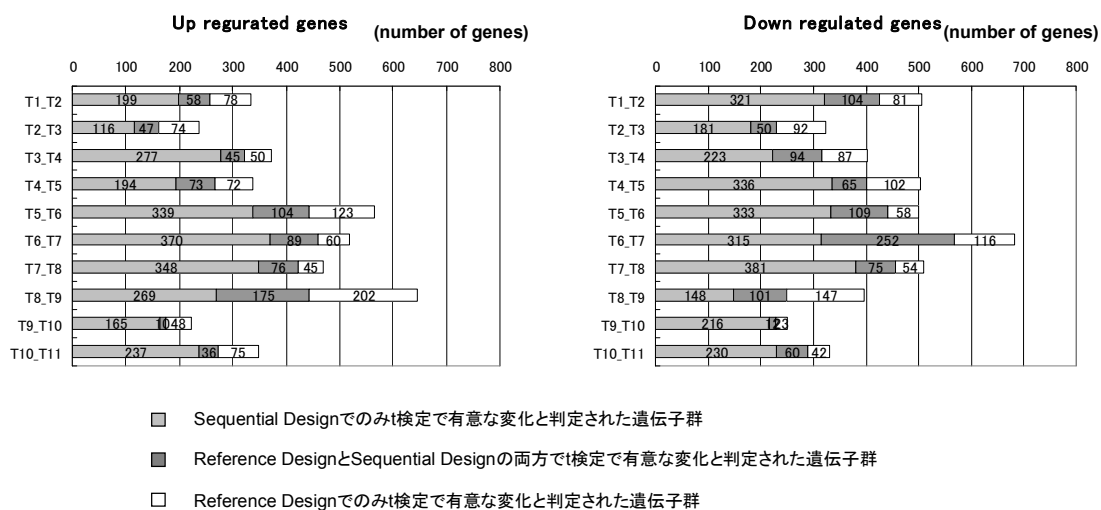
3.3 節でそれぞれのデザインから得られるデータの再現性の違いについて述べた。ここでは、データの精度が異なることによって解析に与える影響を調べる。また、データの精度以外にもそれぞれのデザインを用いた時の利点及び欠点を、データを解析することにより検証する。

3. 4. 1 隣り合う 2 点間の変化

リファレンス・デザインで顕著な Log-ratio 値のばらつきの増加は、遺伝子発現変化を精度よく捉えるための障害になると考えられる。なぜなら、発現の変化を統計的に有意であるかどうかを判定する場合、 t 検定がしばしば用いられるが、このような検定では値のばらつき、すなわち標準偏差が考慮されるため、ばらつきの大きなデータは検定で有意性が得られなくなる。各デザインに基づくデータの特徴が、検定結果にどのように影響を与えるかを検証するために、 t 検定を用いて統計的に有意な変化をもつ遺伝子の個数を調べた。時系列の解析では、連続的な変化を捉えることを優先課題と考え、ここでは隣り合う二点間の変化に着目している。シーケンシャル・デザインにおいては、一標本 t 検定を適用し p-value が 0.01 以下となる遺伝子を統計的に有意に発現変化した遺伝子として選別した。リファレンス・デザインにおいては、2 点のアレイデータを比べることでその 2 点間の発現変化を見ることができるので、2 枚のアレイデータについて二標本の Welch の t 検定を適用した。そして、同様に p-value が 0.01 以下となる遺伝子を得た。結果を図 3.4 に示す。図 3.4 で X 軸は統計的に有意に変化と判定された遺伝子の個数である。縦軸は上から下へ向かって隣り合う 2 点間の時系列を示す。灰色のバーはシーケンシャル・デザインのみで統計的に有意な発現変化をしたと判定された遺伝子、黒色のバーは両デザインで統計的に有意な発現変化をしたと判定された遺伝子、白色のバーはリファレンス・デザインでのみ統計的に有意な発現変化と判定された遺伝子の個数を表す。図 3.4 から、シーケンシャル・デザインの方が変化を統計的に有意と判定できる遺伝子の個数が多いことがわかる。前述の通り、シーケンシャル・デザインに基づいて得られた Log-ratio 値のほうが、全体的にばらつきが小さい精度のよいデータであるために、このように検定結果にも影響を与えていると考えられる。また、両デザインで有意に変化と判定された群については、Log-ratio 値の絶対値は比較的大きく、シーケンシャル・デザインのみで有意に変化と判定された群については、Log-ratio 値の絶対値は 0 付近に分布する。つまり、隣り合う 2 点間の微小な変化をシーケンシャル・デザインはリファレンス・

デザインよりも感度よく捉えることができていると考えられる。ここで、シーケンシャル・デザインでもリファレンス・デザインでも有意と判定された遺伝子の発現変化というのは、非常に信頼が高いと考えられる。

[Yang and Speed 2002; Peixoto *et al.* 2006]は、着目する2点間を直接同じマイクロアレイで競合ハイブリダイゼーションさせて変化を見た場合(シーケンシャル・デザイン)の方が、固定のコントロール点を介して間接的に変化を見た場合(リファレンス・デザイン)よりも、Log-ratio 値の潜在的なばらつきは小さくなり、変化をより精度よく捉えられ、アレイ間で比べるよりもアレイ内で比べた方が、Log-ratio 値の精度がよくなる事を主張している。さらに、今回のような発現状態が大きく異なるサンプル間を比べるときの問題がある場合はよりその差が顕著になっていると考えられる。



$p \leq 0.01$

図3.4: 隣り合う2点間で有意に発現変化した遺伝子の個数

3. 4. 2 両デザインで異なる発現変化を示す領域

3.4.1 節で隣り合う 2 点間の発現変化をそれぞれのデザインについて t 検定を用いて統計的に評価した。しかし、統計的に有意と判定された郡の中にも、両デザインで発現変化の向きが矛盾しているものがいくつか存在する。つまり、一方のデザインで Log-ratio 値は発現上昇を示しているにも関わらず、もう一方のデザインでは発現減少を示しているというような場合である。この矛盾は Log-ratio 値でゼロ付近に分布し、その範囲はそれぞれ異なる(図 3.5)。

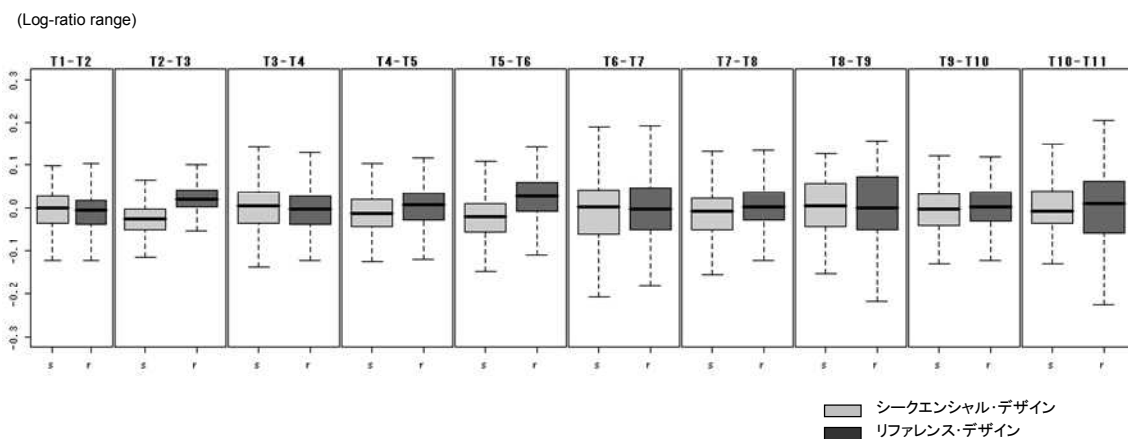
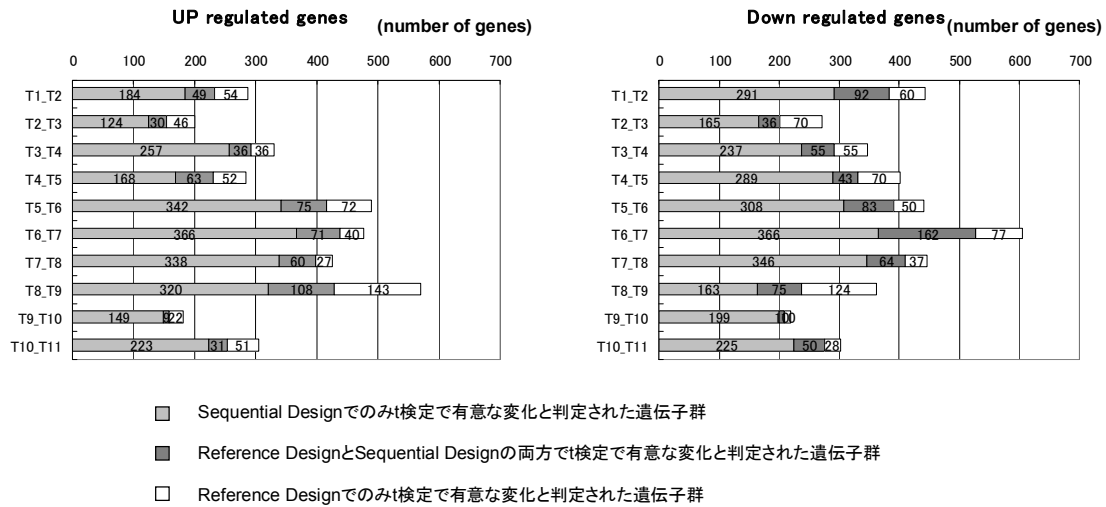


図3.5: 両デザインで変化の向きが矛盾するデータにおけるLog-ratioの存在する範囲

そこで、これらの矛盾の Log-ratio をもつ遺伝子は信頼がおけない値として、特異的に発現変化した遺伝子群から除外することができる。リファレンス・デザインでは 2 枚のマイクロアレイから変化を判定するので、両隣の 2 点間で矛盾が生じた遺伝子についても考慮しなければならない。例えば [T4-T5] の変化を考えた場合、[T3-T4]、[T5-T6] で生じた矛盾が T4 と T5 のマイクロアレイのデータが原因である可能性がある。したがって、リファレンス・デザインでは前後の 2 点間の解析で生じた矛盾についても合わせて除いた。矛盾を示す遺伝子を図 3.4 の結果から除外したものを、図 3.6 に示す。図 3.6 でもシーケンシャル・デザインの方が、全体としてリファレンス・デザインよりも多くの遺伝子を有意と判定していることがわかる。隣り合う 2 点間の変化についてはシーケンシャル・デザインの方がリファレンス・デザインよりも Log-ratio の再現性がよいという結果を 3.3 節で得たが、図 3.6 でもその傾向が見て取れる。



$p \leq 0.01$
 二つのデザインでLog-ratioの矛盾している遺伝子を除いた。

図3.6: 隣り合う2点間で有意に発現変化した遺伝子の個数(矛盾カット)

片方のデザインでのみ有意な変化と判定されるのには二つの原因があると考えられる。一つ目は、蛍光強度が低く解析が行えない場合であり、二つ目はデータのばらつきが大きくなり統計的な有意性が得られない場合である。二つ目の要因は統計的な有意性の判定基準に左右されるので、ここでは一つ目の蛍光強度が低く解析が行えない場合について考察する。表 3.2 は、片方のデザインで統計的に有意と判定されたがもう一方のデザインでは蛍光強度が微弱で解析不能であった遺伝子の数を示したものである。

表 3.2 : 他方のデザインでは解析不能である遺伝子数

	T1_T2	T2_T3	T3_T4	T4_T5	T5_T6	T6_T7	T7_T8	T8_T9	T9_T10	T10_T11
シーケンシャルでのみ有意(UP)だがリファレンスでは解析不能	110	68	125	99	188	230	180	129	106	149
シーケンシャルでのみ有意(DOWN)だがリファレンスでは解析不能	151	65	78	102	135	139	129	90	141	149
リファレンスでのみ有意(UP)だがシーケンシャルでは解析不能	27	30	25	31	48	23	20	98	0	0
リファレンスでのみ有意(DOWN)だがシーケンシャルでは解析不能	46	50	43	47	33	56	21	102	0	0

表 3.2 より、どちらのデザインでも他方で解析不能であった遺伝子が存在する。しかし、この結果の持つ意味は、リファレンス・デザインとシーケンシャル・デザインで異なるを考える。なぜなら、シ

シーケンシャル・デザインでのみ有意と判定された場合、リファレンス・デザインで解析不能となる可能性は比較状態間の蛍光強度が低かった場合のみならず、固定のコントロールサンプルで蛍光強度が低く解析が行えない場合があるからである。一方、リファレンス・デザインでのみ有意と判定された場合というのは、シーケンシャル・デザインで比較したどちらかの状態で蛍光強度が低い場合である。これをリファレンス・デザインで統計的に有意と判定したからといって蛍光強度が低いもの同士を比べているので結果も疑わしい。従って、リファレンス・デザインでのみ有意と判定された場合はあまり信頼が置けない。いずれにしろ、以上の結果で、両方のデザインで共に有意と判定される変化というのは非常に信頼があると考えられる(附表 1、2)。

3. 5 リファレンス・デザインとシーケンシャル・デザインの利点と欠点

表 3.3 はリファレンス・デザインとシーケンシャル・デザインの利点と欠点をまとめたものである。3.4.2 節で解析したように、隣り合う 2 点間の変化は直接見る場合はシーケンシャル・デザインのデータが適していると言えるが、シーケンシャル・デザインのデータだけでは、時系列変化をより詳しく見ることはできない。なぜなら、マイクロアレイ実験で得られるデータはあくまで 2 点間の相対値であるので、コントロールの異なる相対値同士を正確に比べることはできない。例えば $10 \rightarrow 100 \rightarrow 1000 \rightarrow 10000$ と蛍光強度が変化していく場合を考えたときシーケンシャル・デザインでは $\log_{10}(100/10)$ と $\log_{10}(1000/100)$ と $\log_{10}(10000/1000)$ の三つの時系列データができ、どれも同じ値を示すが、通常マイクロアレイでは蛍光強度の大きさに依存してデータのばらつき方が変わるため、 $\log_{10}(100/10)$ と $\log_{10}(10000/1000)$ の変化を同列に扱うには難がある。mRNA の量の変化の割合をこの相対値(Log-ratio)が反映しているかも疑わしい。シーケンシャル・デザインで遠くの 2 点間を比べたいときには、このような同列に比べていかかわからないような相対値同士を足し合わせて 2 点間で発現が上昇したか減少したかを解析しなければならず、それは全く信頼の置ける結果にはならない。一方で、リファレンス・デザインによるデータは、それ自体ではばらつきが大きくなるかもしれないが遠い 2 点間を比べる場合、シーケンシャル・デザインよりも都合がよいと考えられる。なぜならリファレンス・デザインではコントロール点を介してどの 2 点間でも 2 枚のマイクロ

アレイデータを比べるだけで済むからである。表 3.3 にまとめたように、遺伝子発現の時系列パターンをより正確なパターンとして捉えたい場合、どちらか一方のデザインだけでは不完全であることがわかる。そこで、重要なのは、これら二つのデザインの利点を生かしながら両方のデザインから得られる情報を統合することである。そこで次節では、デザインの統合について述べる。

表 3.3 : シーケンシャル・デザインとリファレンス・デザインの利点と欠点

	Sequential Design	Reference Design
利点	直近の2点間の変化をより正確に精度よく捉えることができる。	コントロール点を介してどの2点間でも2枚のマイクロアレイを比べるだけで済む。
	データの規格化の条件にあっている。	共通するコントロールに対する変化なので変量間が比べやすい。
	値のばらつきがReference Designよりも小さなデータが得られる	
欠点	遠くの点同士の変化を見るのには向かない。	発現状態の大きく異なる点同士で実験を行わなければならない、データの規格化が不正確になる。
		コントロール依存で変化を見ることが出来ない遺伝子が存在する。
		値のばらつきがSequential Designよりも大きいデータが得られる。

3. 6 リファレンス・デザインとシーケンシャル・デザインの統合

これまでの解析により、隣り合う 2 点間の変化を直接比べるにはシーケンシャル・デザインが適していることが示された。連続した変化を離散的ないくつかのタイムポイントで表現しようとする場合、すべての任意の 2 点間の関係が決められることが望ましい。しかし、マイクロアレイで測定できるのは遺伝子発現変化の相対値であり、しかもその値はコントロールが変わると比較することが難しい。任意の 2 点間の遺伝子発現変化を決めたい場合、すべての 2 点間でマイクロアレイ実験を行う必要があるが、それは現実的ではない。そこで、本研究では、リファレンス・デザインとシーケンシャル・デザインの 2 つのそれぞれ異なる特徴を持った実験デザインを用いて全ての 2 点間の組み合わせで発現が上昇したか減少したかをより信頼ある情報として捉えることで、正確な時系列パターンを構築する手法を提案している。ここでは、この手法を今回の実験データに適用した結果について述べる。

3. 6. 1 両デザインで同じ発現パターンを示す遺伝子

もし、マイクロアレイが mRNA の量の変化を完全に正確に測れるならば、どちらのデザインを用いても相互の Log-ratio 値を補完できるはずである。つまり、隣り合う 2 点間のシーケンシャル・デザインに基づく変化とリファレンス・デザインで表される変化を対応させる場合は、 $\log(S_{t+1}/S_t) = \log(R_{t+1}/C) - \log(R_t/C)$ が成り立つ。ここで、タイムポイントを t で、 S_t, R_t はタイムポイント t における蛍光強度を表す。しかし実際には、リファレンス・デザインでコントロールサンプル C が介在することにより実験誤差が混入し、この等式は正確には成り立たない。しかも、蛍光強度の大きさの組み合わせによってばらつき方の異なる値を示すと考えられる。しかしながら、シーケンシャル・デザインとリファレンス・デザインで同じ変化を表現しているデータというのは非常に正確であるはずである。そこで、シーケンシャル・デザインとリファレンス・デザインで同じ変化を示すものがどれだけ存在するかを相関係数で評価することにより調べた。

リファレンス・デザインとシーケンシャル・デザインで非常によく似た発現変化のパターンを持つものが大腸菌約 4,300 遺伝子のうち 106 遺伝子得られた。さらに、若干の違いはあるがほぼ同じ発現変化のパターンを持つものが 121 遺伝子得られた。それ以外については、あまりパターンが似ていないか、データの欠損が多く相関係数を計算できなかったものである。これらの発現プロファイルはリ

ファレンス・デザイン、シーケンシャル・デザイン共に支持しており、非常に信頼の高いものであると考えられる。しかし、これらの発現プロファイルには欠損値が含まれており、従来の多変量解析手法を用いるには不適切であり、より多くの遺伝子について正確な時系列遺伝子変化のパターンの作製が望まれる。本研究では、マイクロアレイデータの信頼性を上げるために不正確なデータの除去を行っているが、データ除去の基準を厳しく設定すればするほど、欠損値が多くなり発現プロファイルとして扱うのが困難となる。逆に、データの除去を行わないと多くの遺伝子について欠損値のないパターンを得られるが、信頼の低い変化を含む発現変化パターン(ランダムな変動を含む、本当の発現変化を反映していないパターン)となってしまうというトレードオフの関係にあるのである。本論文では、より信頼のおける正確な変化の情報に基づいて時系列変化パターンを後から構築することで、このトレードオフの関係を克服できていると考える。

3. 6. 2 デザインの選択

デザインの選択は $\text{Log-ratio } M_{gs}$ の標準偏差のマイクロアレイ全体における平均値が小さい方を再現性がよいとして評価することによって行った(2.3節)。その結果を図3.7に示す。図3.7では from ~to の区間で優れたデザインをリファレンス・デザインなら R、シーケンシャル・デザインなら S の文字で記している。その結果、二点間の距離が短いほどシーケンシャル・デザインの方で再現性がよく、遠い二点間になるほどリファレンス・デザインの方で相対的に再現性がよいことがわかった。二点間の距離が2、つまり $T_i \sim T_{i+2}$ のときもシーケンシャル・デザインの再現性がよい。それは、シーケンシャル・デザインの方が元のデータの再現性がよいため、使用するマイクロアレイの枚数が同じでもシーケンシャル・デザインの方がよくなると考えられる。二点間の距離が増えると徐々にリファレンス・デザインを用いた方がいいことがデータからも示された。

		to									
		T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
from	T1	S	S	R	R	R	R	R	R	R	R
	T2		S	R	R	S	R	R	R	R	R
	T3			S	S	S	R	R	R	R	R
	T4				S	S	R	R	R	R	R
	T5					S	S	R	R	R	R
	T6						S	S	R	R	R
	T7							S	S	R	R
	T8								S	R	R
	T9									S	S
	T10										S

図 3.7 : 区間ごとのデザインの選択結果

3. 6. 3 全ての二点間の変化

区間ごとに決められた各デザインに基づいて t 検定を用いて二点間の変化を求めた。詳細は、2.4 節で述べている。ここでは、 t 検定の有意性の判定における p -value の閾値は $th = 0.001$ と厳しめに設定した。また、Log-ratio の変化の基準とする閾値パラメータは $upth = 0.2$ 、 $downth = -0.2$ に設定した。これは、3.4.2 節での矛盾の存在する領域をもとに決定した。図 3.8 に 4 つの遺伝子について、変化がどこでおこっているかを示したマップの例を示す。

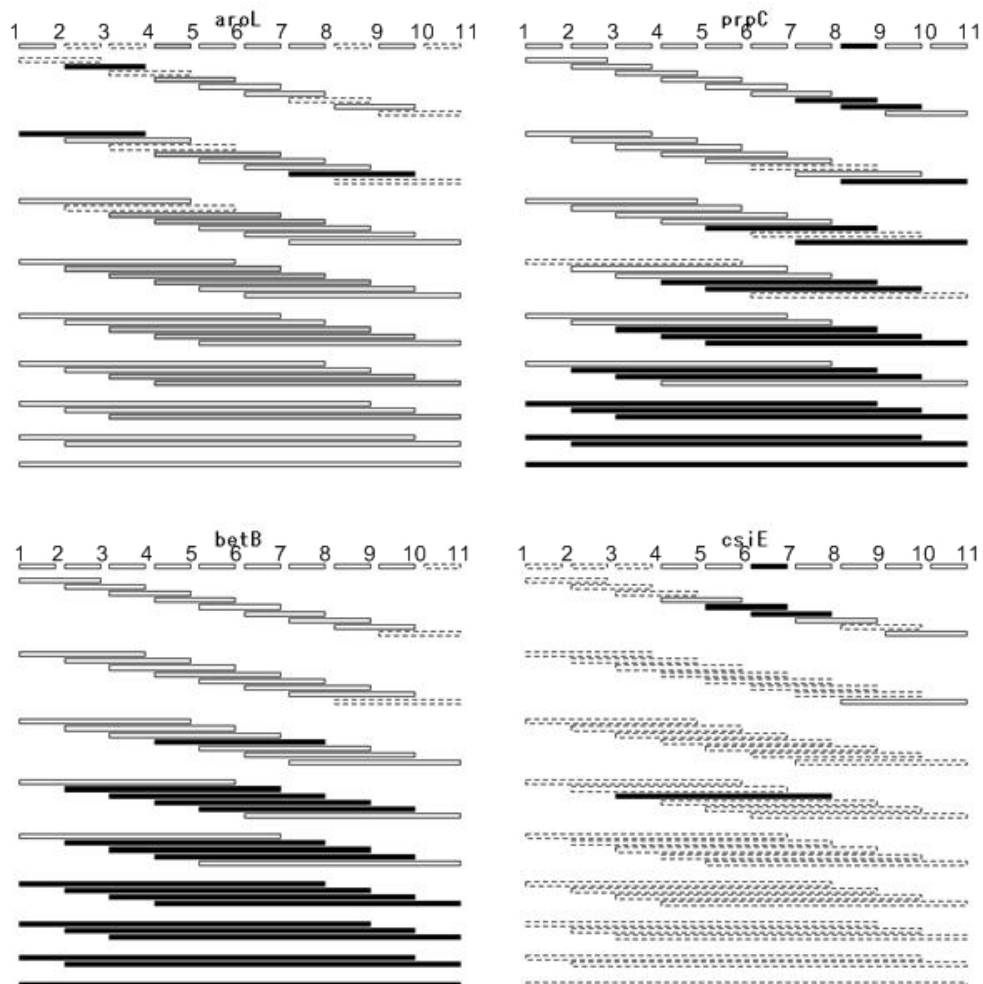


図 3.8 : 統合データから求めた変化のマップ

図 3.8 で発現上昇は黒塗りのバーで、発現減少は灰色のバーで、変化なしを点線と実践の白抜きのパ
ーで示している。点線は、データが欠損値となり検定が行えなかった部分を示している。また、バー
の長さでどのタイムポイントとタイムポイントの間の変化についてであるかを示している。例えば
遺伝子 *aroL*(図 3.8 左上) について区間 3-11 で発現減少と判定されているが、この変化はより細かい
区間 4-5 の変化に集約されていることがわかる。シーケンシャル・デザインを用いることで、より信
頼のある細かい区間の変化を反映していると考えられる。また、区間 2-3 及び 7-10 で発現上昇を示
しており、この遺伝子の発現変化の概略は上がって下がって再び上がるというようなものであること
が推察される。図 3.8 右上の *prpC* 遺伝子についても、区間 8-9 の発現上昇に全体の変化が集約され
ている例である。図 3.8 左下はシーケンシャル・デザインでは捉えられない変化をリファレンス・デザ
インによって捉えられた場合の例である。図 3.8 右下はリファレンス・デザインの共通コントロール
において発現量が少なく解析の行えない遺伝子に関して、シーケンシャル・デザインを用いること
により変化を捉えられた例である。これらのように、より信頼のおけるデザインのデータを用いて求め
た時系列変化のマップは少なくともどちらか一方を用いた時系列変化のパターンよりも信頼ある変
化を表すことができると考えられる。そこで、この変化のマップから全ての二点間の変化を矛盾なく
示す発現系列を計算機により算出した結果を、次節では考察する。

3. 6. 4 統合発現系列

図 3.9 は図 3.8 の変化のマップについて時系列パターンを構築した結果の例である。実線はシーケ
ンシャル・デザインの実測値(Logratio)を、荒い点線はリファレンス・デザインの実測値(Logratio)
を、細かい点線は統合したデータをパターン各点の値の二乗和が 1 になるように規格化してそれぞ
れ表す。シーケンシャル・デザインでは、便宜上、Logratio を順次足し合わせてプロットしている。
途中値の抜け落ちている箇所はデータフィルタリングにより排除した値である。図 3.9 を見てわかる
ように統合パターンは両デザインのデータを簡略化しつつ、変化の重要な部分を反映したものとなっ
ていることがわかる。例えば遺伝子 *prpC* の例(図 3.9 右上)において、リファレンス・デザイン、シー
ケンシャル・デザイン共に時系列変化は後半で徐々に上昇している。しかし、シーケンシャル・デザ
インを用いることにより確実に変化のある箇所だけが反映された発現系列が構築されていることがわ
かる。遺伝子 *betB*(図 3.9 左下)は、シーケンシャル・デザインでは捉えられない微妙な変化をリファ
レンス・デザインで緩やかな変化として捉えられていることを示す例である。遺伝子 *csiE*(図 3.9 右下)
はシーケンシャル・デザインを元に途中から発現上昇してきた遺伝子について発現系列として反映で

きている。これらのように、統合パターンの良い所は変化の形状がそのまま信頼おけるということである。しかも、どの点と点の間で変化が生じたかが明確であるため、遺伝子発現の制御機構の解明に役立つと考えられる。

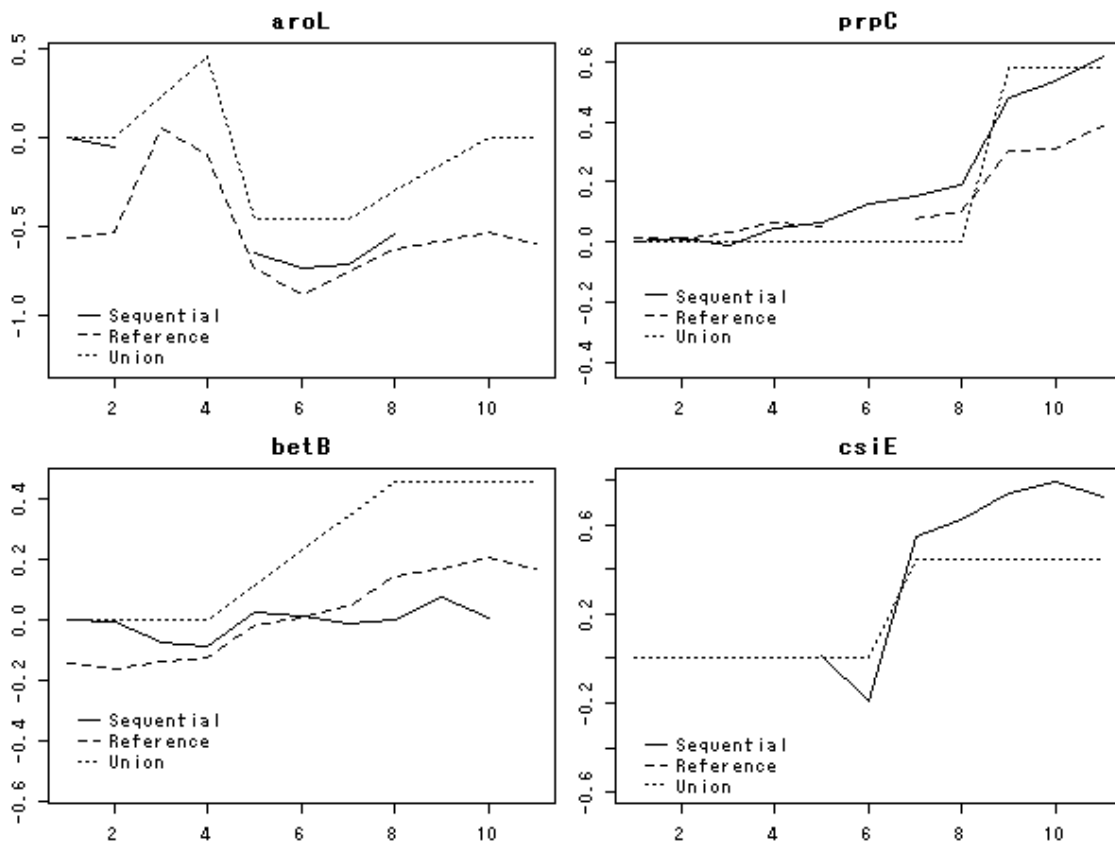


図 3.9 : 統合したパターンと実測値との比較

3.7 統合データに基づく rpoS レギュロンの解析

マイクロアレイ実験と同様のタイムポイントで、定常期シグマ σ^s についてタンパク量の変動をウエスタンブロッティングにより見た。結果を図 3.10 に示す。

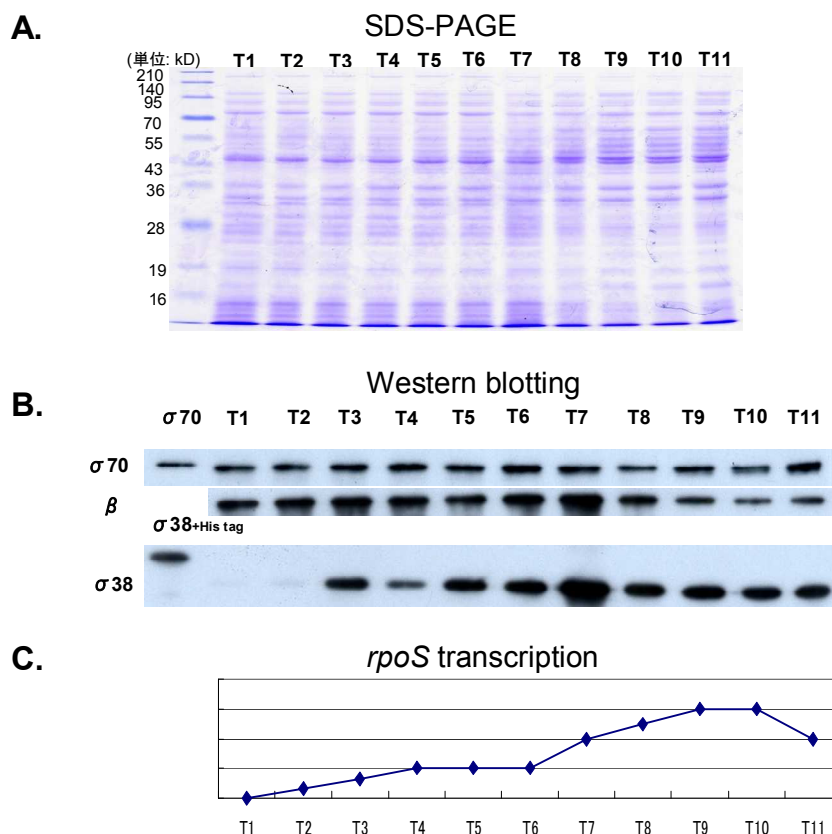
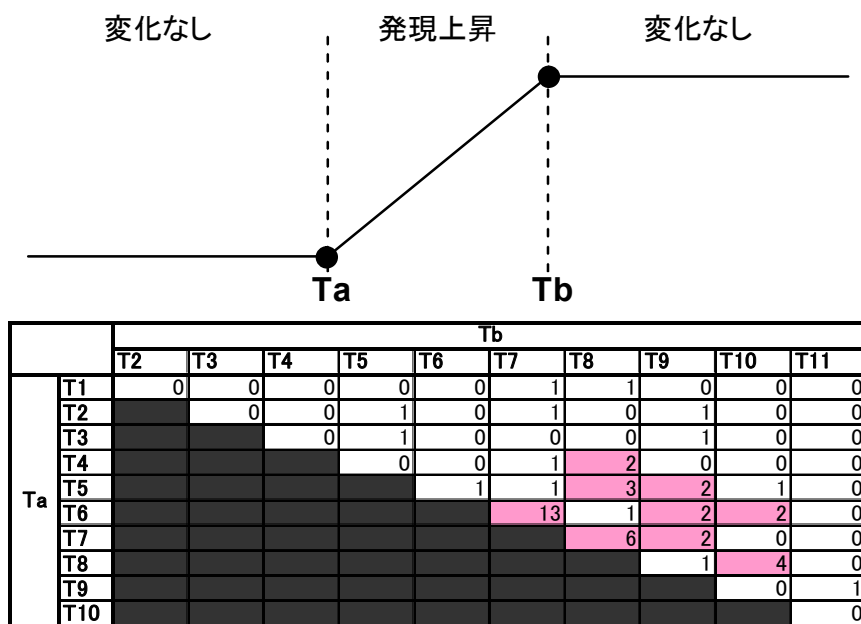


図3.10: sigma 38の時系列変化

図 3.10 より T3(OD=0.4)で σ^s ($\sigma 38$)は現れ始め T7(OD=1.7)で急激に増えていることがわかる。この結果はBordes *et al.* 2003 の結果とほぼ一致する。また、Bordes *et al.* 2003 では遺伝子 *osmE* がOD=1.3~2 の間で急激に、 σ^s に誘導されると述べている。また図 3.10C には本研究で構築した *rpoS* 遺伝子の遺伝子発現変化の時系列パターンを示している。図 3.10B の *sigmaS* の変動と一致していることが見て取れる。

図 3.11 は、*rpoS* レギュロンの統合パターンを変化の開始と終わりの位置で分類した結果である。*rpoS* レギュロンは、 σ^s が認識するプロモーターを持ち σ^s に制御されると考えられる遺伝子群であ

る。図 3.11 では Ta-Tb で発現上昇を表す統合パターンを分類し、分類されたグループの中に既知の *rpoS* レギュロンがいくつ存在するかをカウントしている(図 3.11 内表)。これを見ると、T6-T7 で発現上昇する統合パターンに *rpoS* レギュロンが多く含まれていることがわかる。また、T6-T7 を挟む区間でも *rpoS* レギュロンがいくつか分類されている(T4-T8、T5-T8、T5-T9、T6-T9、T6-T10)。隣り合う 2 点間では有意な変化として捉えられない変化も、統合パターンでは離れた点同士で有意な変化として反映されている。従って、T6-T7 を挟んだ *rpoS* レギュロンのゆるやかな発現変化を捉えることが出来ていると考えられる。また、Bordes *et al.* 2003 では遺伝子 *osmE* が OD=1.3~2 の間で急激に、 σ^s に誘導されると述べているが、これは、本研究における T6-T7 の区間(OD=1.3~OD=1.7)と一致している。従って、T6-T7 で誘導される時系列変化パターンを示す遺伝子は、 σ^s によってなんらかの形で制御されていることを示唆していると考えられる。また、図 3.11 表では、T7 以降で初めて発現上昇を示すパターンにも *rpoS* レギュロンがいくつか分類された。これらの遺伝子が T6-T7 に遅れて σ^s に誘導されるかを検証することは興味深い。



Ta~Tbの区間での発現上昇に着目し統合パターンを分類した。表はその中にある、*rpoS*レギュロンの個数を示す。マークされた区間は、*rpoS*レギュロンが2個以上分類された箇所である。

図3.11: 変化の位置による*rpoS*レギュロンの分類

3.8 まとめ

本研究では、本来、二状態間の相対変化を見るための技術である二色蛍光マイクロアレイを用いて遺伝子発現の時系列変化を捉える試みを実験デザインの段階に踏み込んで行った。実験デザインには、近接した二点間を直接比べる利点を持つシーケンシャル・デザインと、より離れた点同士でも比較的相対変化量の比べられるリファレンス・デザインとを用い、これらを実際のマイクロアレイ実験を行い、検証を行った。そして、それぞれの実験デザインの特徴を活かして、正確でより信頼のある発現変化のパターンの構築法を提案した。さらに、提案手法を用いることにより、変化の区間に基づいた分類ができることを、*rpoS* レギュロンに関する解析を行うことで示した。

3.9 今後の課題

構築した発現系列が本当の遺伝子発現量の時間変化を反映したものであるかの検証をさらに行う必要がある。これは、文献による調査が行えると考える。また、オペロンは同じ発現系列を示すべきだという視点から、発現系列構築時の各種パラメータの設定などを検討することができないかということも考えている。本提案手法の問題点としては、変化の判定時の間違いがそのまま発現系列に反映されるため、変化を変化として正確に判定することが求められることである。現段階では、全ての二点間の変化のマップを作る際に、*t* 検定の *p-value* の閾値や *Log-ratio* の閾値を恣意的に決めている。今後は、より正確に変化を捉えることが出来る最適な閾値を得ることが必要となる。それは、統合パターンの結果とすでにわかっている個々の遺伝子の発現変化を比べることにより可能となると思われる。また、Huber *et al.* 2002 に提案されているような、*Log-ratio* 値のばらつきを安定化させる拡張型 *Log* 変換を試して見ることも興味深い。

4 本研究で用いた手法

4. 1 実験手法

本研究で用いた、諸実験手法についてここでは説明する。

4. 1. 1 菌株、培地、培養条件およびRNA抽出

大腸菌野生株 K-12 (W3110)を 2,000ml の Luria-Bertani(LB)液体培地内で 3,000ml ジャーファーマンターを用いて培養した。培養開始時の LB 培地の pH は 7.4 に調整、オーバーナイトで前培養した培地を 250 倍希釈して添加した。培養条件は温度を 37°Cの一定、攪拌速度を 300rpm、空気送り量 20/h である。細胞の増殖の様子は、吸光度計により 600nm の波長の透過度で測定した。回収した培養液は直ちに RNA protect Bacteria Reagent (Qiagen)試薬を添加し、5,000rpm の遠心操作により菌体を分離し-80°Cで冷凍保管した。RNA の精製は、RNA protect(Qiagen), RNeasy Mini Kit(Qiagen), RNase-free DNase を使用して行った。精製は、添付のマニュアルに従い行い、-80°Cで保管した。また、精製物へのゲノム DNA のコンタミネーションはゲル電気泳動で確認した。

4. 1. 2 蛍光標識色素のラベリング

蛍光標識は total RNA を逆転写酵素により相補 DNA に逆転写させ、その際に取り込んだ Amino-allyl 修飾された塩基とのカップリング反応により行われる。Total RNA 15 μ g と 1.2 μ g の random primer(Invitrogen)を総量 31 μ l になるように Nuclease free water (Ambion)に入れ、70°C で 10 分加熱し、25°Cでインキュベートしプライマーを RNA 鎖に結合させた。逆転写反応は、逆転写酵素に SuperScript III (Invitrogen)、バッファーに 1 \times 1stStrand buffer および 10mM DTT、ヌクレオチドに 5mM dATP、5mM dGTP、5mM dCTP、0.25mM dTTP、0.25mM AA-dUTP(TAKARA)を使用し行った。この反応は、25°Cで 10 分、37°Cで 60 分、42°Cでオーバーナイト、70°Cで 10 分連続して行った。残留した RNA 鋳型は、20 μ l の 1N NaOH の添加で 65°C、

30分反応されることにより加水分解し、その後20 μ lの1N HClの添加で反応を中和させた。cDNAの精製は、CyScribe GFX Purification Kit (GE Healthcare) を用いてマニュアルどおりに行い、蛍光色素標識のためCy3またはCy5のNHS エステルを加え4時間インキュベートした。15 μ lの4M hydroxylamineを入れ暗所で15分インキュベートすることにより反応を停止させ、ラベリングしたcDNAの精製は再びCyScribe GFX Purification Kit を用いて行った。

4. 1. 3 ハイブリダイゼーションとスポット検出

マイクロアレイスライドのプリハイブリダイゼーションを滅菌ろ過したプリハイブリダイゼーション溶液の元、3時間42 $^{\circ}$ Cで行った。プリハイブリダイゼーション溶液の組成は25%formamide、5 \times SSC、10mg/l BSA、0.1% SDSである。プリハイブリダイゼーション後スライドをmiliQ水と80%エタノールで洗い、1000gで5分遠心して乾燥させた。ハイブリダイゼーション溶液の組成は25%formamide、5 \times SSC、0.1% SDS、0.1 μ g poly(A)/ml、1 \times Denhardt's solution、100pmol Cy3 and Cy5 combined probeである。蛍光標識後のcDNAを入れたハイブリダイゼーション溶液を95 $^{\circ}$ Cで3分加熱し、42 $^{\circ}$ C16時間ハイブリダイゼーション反応を行った。ハイブリダイゼーション反応にはadvalytixのハイブリダイゼーション装置(ArrayBooster)を使用した。ハイブリダイゼーション後、スライドを洗い、1000gで5分の遠心により乾燥させた。そして、Fuji FLA-8000 スキャナーでスキャンし、スポット抽出はFuji filmのArray gauge ver.2.0ソフトウェアを使用して行った。

4. 1. 4 ウェスタンブロット法

培養細胞をTCA沈殿し、タンパク質を得た。得られたタンパク質を10%のSDSゲルに電気泳動後、ニトロセルロース膜に転写した。各レーンのタンパク質総量はSDS-PAGEを行くマシーブリリアントブルー(CBB)染色をした際の発色強度で合わせた。転写したニトロセルロース膜を5%スキムミルクでブロッキングし、1次抗体に反応させた。1次抗体にはMAb (sigma70, sigmaS, beta) subunit of E.coli RNA polymerase (neoclone)を使用した。2次抗体としてGoat Anti-Mouse IgG(H+L)-HRP Conjugate (BIO-RAD)を用いた。シグナルの検出はECLキット(GE-Healthcare)を使用した。

4. 2 実験デザインとデータの前処理

マイクロアレイデータを解析する上での前処理法を説明する。また、本研究で用いたデータの評価法についてもこの節で説明している。

4. 2. 1 マイクロアレイ法におけるデータの規格化

DNA アレイの基本原理は DNA のハイブリダイゼーション(hybridization)であり、サザンブロットやノーザンブロットを大規模かつ並列で行うことで数千から数万個の遺伝子の発現量を測定する方法である。DNA アレイ実験には、GeneChip 法、スポット型アレイ法(cDNA マイクロアレイ法)、および Serial Analysis of Gene Expression (SAGE)法がある。

GeneChip 法は Affymetrix 社が開発した方法であり、各遺伝子の検出に対して最大 40 塩基 (通常は 11-20 塩基程度) のオリゴヌクレオチドが使われる。各遺伝子について、他の遺伝子と比べて最も相同性の低い領域をいくつか選択し、それぞれの領域に対して 11-20 塩基からなる完全相補オリゴヌクレオチドとこの配列と 1 塩基置換が中央付近にある不完全オリゴヌクレオチドに対する蛍光強度差から発現量を測定する。

二色蛍光マイクロアレイ(cDNA マイクロアレイ)は、基準となる系(コントロール実験)と目的とする系 (ターゲット実験) それぞれから mRNA を抽出し、それぞれの系において二つの蛍光色素 (例えば、Cy3 と Cy5) によりラベルされた cDNA を、マイクロアレイ上にスポットした cDNA と競合ハイブリダイゼーションすることにより、それぞれの実験における 1000-10,000 種類の遺伝子の発現量の相違を検出する方法である(図 4.1)。

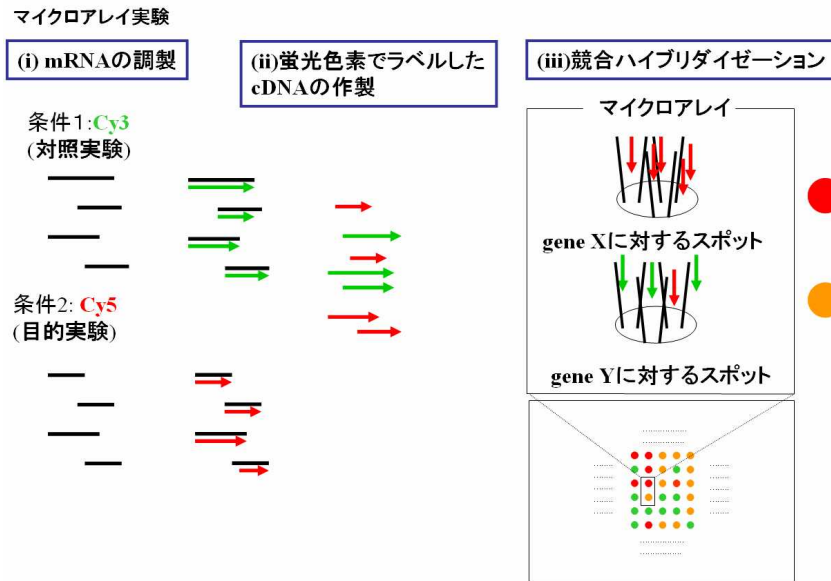


図 4.1 : マイクロアレイ実験

二つのDNAマイクロアレイ測定データを比較する場合、それぞれの測定データの大多数の遺伝子の発現強度に差がないと仮定し、データを規格化する。たとえば、スポット型アレイ法(cDNA マイクロアレイ法)において、Cy3 と Cy5 の二種類の蛍光色素で cDNA を標識した場合、蛍光色素の蛍光強度が不安定であること、また、標識する過程における cDNA への取り込み効率が異なることから、画像処理の過程で得られたこれらの二色の蛍光強度にはランダムノイズとバイアスの二つの誤差が含まれる。シグナル強度が低いほどこれらの誤差の影響を受けやすい。これらの誤差を軽減化する方法として MA プロット法(MA-plot)によりデータを規格化する[Dudoit *et al.* 2002: Quackenbush 2002]。それぞれのスポットに対する蛍光強度の対数比を $M_i = \log(T_i / C_i)$ 、平均対数強度(averaged logarithmic intensity)を $A_i = \{\log(T_i) + \log(C_i)\} / 2$ とする。 A_i を横軸、 M_i を縦軸にプロットすることにより平均対数強度に対する偏り誤差を評価することができる。図 4.2 上段は MA プロットの一例である。この図から平均対数強度が低い領域において対照実験におけるシグナル強度が低い傾向にあることがわかる。

このような偏り誤差を補正するために $M_i = 0$ の近傍において A_i と M_i の関数

$$M_i = f(A_i) + \varepsilon_i \tag{4-1}$$

を求める。ここで ε_i は i 番目のスポットに対する残差である。式(2-1)により得た関数 f をもとにそれぞれの A_i に対するベースラインを求め、これを

$$\hat{M}_i = f(A_i) \tag{4-2}$$

とする。規格化対数比(normalization Log-ratio)は M_i と \hat{M}_i との差 (式(4-3)) により求められる。

$$\tilde{M}_i = M_i - \hat{M}_i \quad (4-3)$$

補正前(図 4.2 上段)および補正後 (図 4.2 下段) の MA プロットを図 4.2 に示す。この図から明らかのように下段において $M_i = 0$ に対して対称的な分布となっていることから、蛍光色素特異的なシグナルによる偏り誤差は軽減されたことになる。この方法では、大多数のスポットに対しては発現強度に大きな差がないことが前提となっている。また、 A_i と M_i の関数 f を求める過程で $M_i = 0$ の近傍の統計的規定法は[Tseng *et al.* 2001]により提案されている。

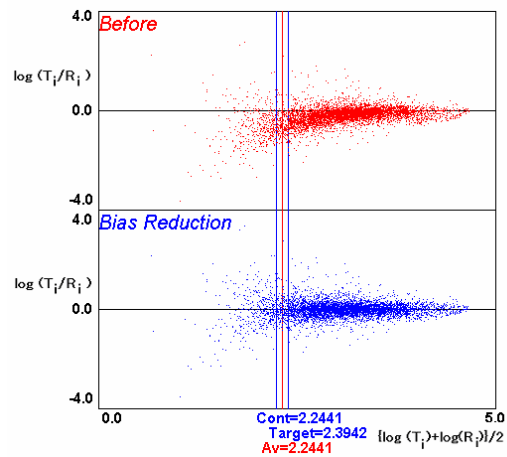


図 4.2 : MA プロットに基づくマイクロアレイデータの規格化

4. 2. 2 不正確なデータの除去

本研究で用いた不正確なデータの除去法は、(1)微小蛍光強度の除去、(2)繰り返し実験間で不整合なデータの除去、の二つのプロセスにより構成される。(1)微小な蛍光強度を切り捨てる手法は、Gao *et al.* 2003、Asyali *et al.* 2004、Quackenbush 2002 などで提案されている。ここでは Quackenbush 2002 のバックグラウンド強度の分布に基づいた手法を用いた。(2)繰り返し実験間での不整合なデータの除去法も、Quackenbush 2002 に基づいている。

(1)微小蛍光強度の除去

蛍光強度がバックグラウンド強度の標準偏差の2倍の範囲に入ってしまうときは、蛍光強度が解析に不十分な値を示すものとして除去する。各スポットのターゲット蛍光強度 T_g とコントロール蛍光強度 C_g は、スポットの平均強度とローカルなバックグラウンド強度によりもとまる。

$$T_g = T_g^{spot} - T_g^{background} \quad (4-4)$$

$$C_g = C_g^{spot} - C_g^{background} \quad (4-5)$$

バックグラウンド強度の全てのスポットでの標準偏差をもとに、ターゲットとコントロール共にバックグラウンド強度の 2σ 以上の強度を持つスポットのみを解析に用いる(4-6)。

$$T_g^{spot} > 2 \times \sigma(T^{background}) \cap C_g^{spot} > 2 \times \sigma(C^{background}) \quad (4-6)$$

(2)繰り返し実験間で不整合なデータの除去

蛍光標識間のバイアスを規格化後の Log-ratio の値を、二回の繰り返し実験間で x 軸と y 軸にプロットすると図 4.3 のようになる。x-y 座標において傾き 45 度の直線(図 4.3 u 軸)が再現良く得られたデータを表す。つまり、この直線から大きく外れたデータを不正確な値として除去しなければならない。ここでは、v 軸方向のデータの標準偏差 σ を算出し u 軸から $2 * \sigma$ の範囲(図 4.3 点線)から外れるデータを除去する。

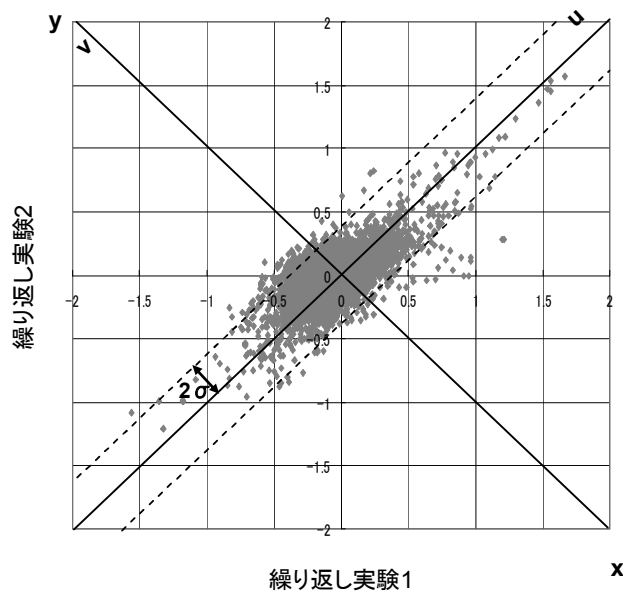


図 4.3 : 繰り返し実験間の散布図

以降のデータ解析では、微小蛍光強度の除去、及び繰り返し実験間で不整合なデータの除去を施したデータを用いる。従って、比較する実験対間のどちらか一方でデータが除去されて欠損値となっている場合、比較を行うことはできない。

4. 2. 3 Log-ratio の再現性の評価方法

ある実験対 m のある遺伝子 g について繰り返し数 n の遺伝子の相対発現変化量(Log-ratio: M_{ngm}) が得られたとするとその再現性は、 N 個のスポットの標準偏差 SD_{gm} で評価することができる。そして、この各時点の個々の遺伝子に対する標準偏差 SD_{gm} を全遺伝子について分布を見ることにより、マイクロアレイ内のデータ全体として、データの質を評価することが出来る。個々の遺伝子 g の標準偏差 SD_{gm} は次式で表される。

$$SD_{gm} = \sqrt{\frac{\sum_s (M_{ngm} - \bar{M}_{gm})^2}{N-1}} \quad (4-7)$$

\overline{M}_{gm} は Log-ratio M_{ngm} の平均値である。この個々の遺伝子ごとの標準偏差を各マイクロアレイ内の遺伝子全体で分布を見ることで、実験対ごとのマイクロアレイデータの再現性の評価を行う。

4. 2. 4 リファレンス・デザインとシーケンシャル・デザインで似たパターンを検出するのに用いた方法

本研究では、mRNA の量的変化を 2 つの異なる見方で測定することを試み、それぞれの見え方の違いや特徴を明らかにすることも目的の一つである。理想的には、どちらのデザインでも等価な時系列変化が見られることが望ましい。それぞれの、デザインで検出できた発現パターンが、対応する遺伝子でどれだけ似ているかを評価するために相関係数を用いた。

シーケンシャル・デザインで得られた遺伝子 g の発現プロファイル $\bar{s}_g = s_{g2} \ s_{g3} \ \cdots \ s_{gi} \ \cdots \ s_{gT}$ 、リファレンス・デザインにより得られた遺伝子 g の発現プロファイル $r_{g1} \ r_{g2} \ \cdots \ r_{gi} \ \cdots \ r_{gT}$ とする。これらの系列は、同一の変化を違う角度から見たデータといえるので、相互に比較できるように値を変換する必要がある。そこで、リファレンス・デザインの発現プロファイルの隣同士の差を取ることで、シーケンシャル・デザインのデータと比較できるようにした。 $r'_{gi} = r_{gi} - r_{gi-1}$ のように差を取り、新たなリファレンス・デザインの系列 $\bar{r}'_g = r'_{g2} \ r'_{g3} \ \cdots \ r'_{gi} \ \cdots \ r'_{gT}$ を作った。 \bar{s}_g と \bar{r}'_g の相関を取ることで、両方で同じような変化を表しているパターンを検出できる。相関係数は次式で求められる。

$$r_{ag} = \frac{\sum_{t=2}^T (s_{gt} - \bar{s}_g)(r'_{gt} - \bar{r}'_g)}{\sqrt{\sum_{t=2}^T (s_{gt} - \bar{s}_g)^2 \sum_{t=2}^T (r'_{gt} - \bar{r}'_g)^2}} \quad (4-8)$$

この相関の取り方は、隣り合う 2 点間の差しか見ておらず、パターンの類似度があいまいな時にはあまり機能しない。そこで、さらに厳密にパターンの類似度を見るために、2 点間の ratio 値の差を全ての点の間の組み合わせで算出し、それに基づく相関係数も求めた。シーケンシャル・デザインで全ての点の組み合わせで差を算出すると次式となる。

$$s_{gx}'' = \begin{cases} \sum_{t=i}^j s_{gt} & (i=j) \\ s_{gt} & (i < j) \end{cases} \begin{cases} 2 < i < T \\ 2 < j < T \\ i \leq j \\ x = j(j-3)/2 + i \end{cases} \quad (4-9)$$

一方、リファレンス・デザインで全ての点の組み合わせで差を算出すると次式となる。

$$r_{gx}'' = r_{gk} - r_{gl} \begin{cases} 2 < k < T \\ 1 < l < T-1 \\ l < k \\ x = k(k-3)/2 + l + 1 \end{cases} \quad (4-10)$$

これらを、用いた相関係数は

$$r_{bg} = \frac{\sum_{x=1}^{T(T-1)/2} (s_{gx}'' - \bar{s}_g'') (r_{gx}'' - \bar{r}_g'')}{\sqrt{\sum_{x=2}^{T(T-1)/2} (s_{gx}'' - \bar{s}_g'')^2 \sum_{x=2}^{T(T-1)/2} (r_{gx}'' - \bar{r}_g'')^2}} \quad (4-11)$$

としてもとまり、より厳密にパターンの類似度を測る指標となる。これら二つの相関係数を用いて、ここでは表 4.1 の基準に基づいて似たパターンを選定した。

表 4.1 : 相関係数の閾値

	r_{ai}	r_{bi}
波形がほぼ同じとみなされる	$r_{ai} > 0.9$	$r_{bi} > 0.5$
波形がだいたい同じとみなされる	$0.9 \geq r_{ai} > 0.8$ $0.8 \geq r_{ai} > 0.7$ $0.7 \geq r_{ai} > 0.6$ $0.6 \geq r_{ai} > 0.5$	and $r_{bi} > 0.6$ and $r_{bi} > 0.7$ and $r_{bi} > 0.8$ and $r_{bi} > 0.85$
波形に相違が見られる	others	

5 謝辞

本研究を遂行するにあたり、終始適切な御助言を頂き、また丁寧に御指導してくださった金谷重彦教授に心より感謝申し上げます。また、国内外の多数の学会や研究会に参加させていただき誠にありがとうございます。

システム細胞学講座の大島拓助手には、実験を行うにあたり多くの有益なご指導、ご鞭撻を頂き、また、常に本研究に対するスタンスや生物学的なものの考え方をアドバイスしていただき、心より感謝いたします。

システム細胞学講座の小笠原直毅教授には実験を行うにあたり、最適な環境を提供していただき、また多くの適確な御助言を頂き、心より感謝を申し上げます。

また本研究を遂行するにあたり、様々な方々に貴重なご助言、ご指導を賜りました。システム細胞学講座の石川周助手、小林和夫助手をはじめ諸先輩方、修士学生の方々には実験を遂行するにあたり親切な指導をして頂き心より感謝いたします。最後に比較ゲノム講座のスタッフ、先輩、修士学生の皆様には心温まる雰囲気の中楽しく一緒に研究をさせていただき心より感謝いたします。本当にありがとうございます。

6 参考文献

[Asyali *et al.* 2004]

Asyali, M. H., Shoukri, M. M., Demirkaya, O. D. and Khabar, K. S. A. : Assessment of reliability of microarray data and estimation of signal thresholds using mixture modeling. *Nucleic Acids Research* 2004, **32(8)**: 2323-2335.

[Bar-Joseph *et al.* 2004]

Bar-Joseph, Z.: Analyzing time series gene expression data. *Bioinformatics* 2004, **20(16)**: 2493-2503.

[Bordes *et al.* 2003]

Bordes, P., Conter, A., Morales, V., Bouvier, J., Kolb, A. And Gutierrez, Claude.: DNA supercoiling contributes to disconnect σ^s accumulation from σ^s -dependent transcription in *Escherichia coli*. *Molecular Microbiology* 2003, **48(2)**: 561-571.

[Chang *et al.* 2002]

Chang, D. E., Smalley, D. J. and Conway, T.: Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. *Molecular Microbiology* 2002, **45(2)**: 289-306.

[Chen *et al.* 1997]

Chen, Y., Dougherty, E. R., and Bittner, M. L.: Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Journal of Biomedical Optics* 1997, **2(4)**: 364-374.

[Churchill 2002]

Churchill, G. A.: Fundamentals of experimental design for cDNA microarrays. *Nature Genetics Supplement* 2002, **32**: 490-495.

[Conway and Schoolnik 2003]

Conway, T. and Schoolnik, G. K.: Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. *Molecular Microbiology* 2003, **47(4)**: 879-889

[Dobbin and Simon 2002]

Dobbin, K. and Simon, R.: Comparison of microarray designs for class comparison and class discovery. *Bioinformatics* 2002, **18(11)**: 1438-1445.

[Durbin and Rocke 2004]

Durbin, B. P. and Rocke, D.M.: Variance-stabilizing transformations for two-color microarrays. *Bioinformatics* 2004, **20(5)**: 660-667.

[Durbin *et al.* 2002]

Durbin, B. P., Hardin, J. S., Hawkins, D. M. and Rocke, D. M.: A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics* 2002, **18**: S105-S110.

[Dudoit *et al.* 2002]

Dudoit, S., Yang, Y. H., Callow, M. J. and Speed, T. P.: Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* 2002, **12**: 111-139.

[Dharmadi and Gonzalez 2004]

Dharmadi, Y. and Gonzalez, R.: DNA microarrays: Experimental issues, data analysis, and application to bacterial systems. *Biotechnology Progress* 2004, **20**: 1309-1324.

[Engelen *et al.* 2006]

Engelen, K., Naudts, B., Moor, B. D. and Marchal, K.: A calibration method for estimating absolute expression levels from microarray data. *Bioinformatics* 2006, **22(10)**: 1251-1258.

[Ernst *et al.* 2005]

Ernst, J. Nau, G. J. and Bar-Joseph, Z.: Clustering short time series gene expression data. *Bioinformatics* 2005, **21**: i159-i168.

[Gao *et al.* 2003]

Gao, X., Fu, X., Li, Tao., Zi, J., Luo, Y., Wei, Q., Zeng, E., Xie, Y., Li, Y. and Mao, Y.: Determining a detectable threshold of signal intensity in cDNA microarray based on accumulated distribution. *Journal of Biochemistry and Molecular Biology* 2003, **36(6)**: 558-564.

[Hakamada *et al.* 2006]

Hakamada, K., Okamoto, M. and Hanai, T.: Novel technique for preprocessing high dimensional time-course data from DNA microarray: mathematical model-based clustering. *Bioinformatics* 2006, **22(7)**: 843-848.

[Huber *et al.* 2002]

Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A., and Vingron, M.: Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 2002, **18**: S96-S104

[Jishage *et al.* 1996]

Jishage, M., Iwata, A., Ueda, S. and Ishihama, A.: Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: Intracellular levels of four species of sigma subunit under various growth conditions. *Journal of Bacteriology* 1996, **178(18)**: 5447-5451.

[Kerr and Churchill 2001]

Kerr, M. K. and Churchill, G. A.: Experimental design for gene expression microarrays. *Biostatistics* 2001, **2(2)**: 183-201.

[Liu *et al.* 2005]

Liu, H., Tarima, S., Borders, A. S., Getchell, T. V., Getchell, M. L. and Stromberg, A. J.: Quadratic regression analysis for gene discovery and pattern recognition for non-cyclic short time-course microarray experiments. *BMC Bioinformatics* 2005, **6(106)**.

[Park *et al.* 2002]

Park, T., Yi, S. G., Lee, S., Lee, S. Y., Yoo, D. H., Ahn, J. I. and Lee Y. S.: Statistical test for identifying differentially expressed genes in time-course microarray experiments. *Bioinformatics* 2003, **19(6)**: 694-703.

[Peddada *et al.* 2003]

Peddada, S. D., Lobenhofer, E. K., Li, L., Afshari, C. A., Weinberg, C. R. And Umbach, D. M.: Gene selection and clustering for time-course and dose-response microarray experiments using order-restricted inference. *Bioinformatics* 2003, **19(7)**: 834-841.

[Peixoto *et al.* 2006]

Peixoto, B. R., Vencio, R. ZN., Egidio, C. M., Mota-Vieira, L., Verjovski-Almeida, S. and Reis, E. M.: Evaluation of reference-based two-color methods for measurement of gene expression ratios using spotted cDNA microarrays. *BMC Genomics* 2006, **7**:35.

[Quackenbush 2002]

Quackenbush, J.: Microarray data normalization and transformation. *Nature Genetics* 2002, **32**: 496-501.

[Storey *et al.* 2005]

Storey, J. D., Xiao, W., Leek, J. T., Tompkins, R. G. and Davis, R. W.: Significance analysis of time course microarray experiments. *PNAS* 2005, **102(36)**: 12837-12842.

[Tseng *et al.* 2001]

Tseng, G. C., Oh M-K., Rohlin, L., Liao J.C. and Wong, W. H. : Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Research* 2001, **29(12)**: 2549-2557.

[Troyanskaya *et al.* 2002]

Troyanskaya, O. G., Garber, M. E., Brown, P. O., Botstein, D. and Altman, R. B.: Nonparametric methods for identifying differentially expressed genes in microarray data. *Bioinformatics* 2002, **18**: 1454-1461.

[Willbrand *et al.* 2005]

Willbrand, K., Radvanyi, F., Nadal, J. P., Thiery, J. P. and Fink, T. M. A.: Identifying genes from up-down properties of microarray expression series. *Bioinformatics* 2005, **21(20)**: 3859-3864.

[Yang and Speed 2002]

Yang, Y. H. and Speed, T.: Design issues for cDNA microarray experiments. *Nature Reviews Genetics* 2002, **3**: 579-588.

7 学会発表

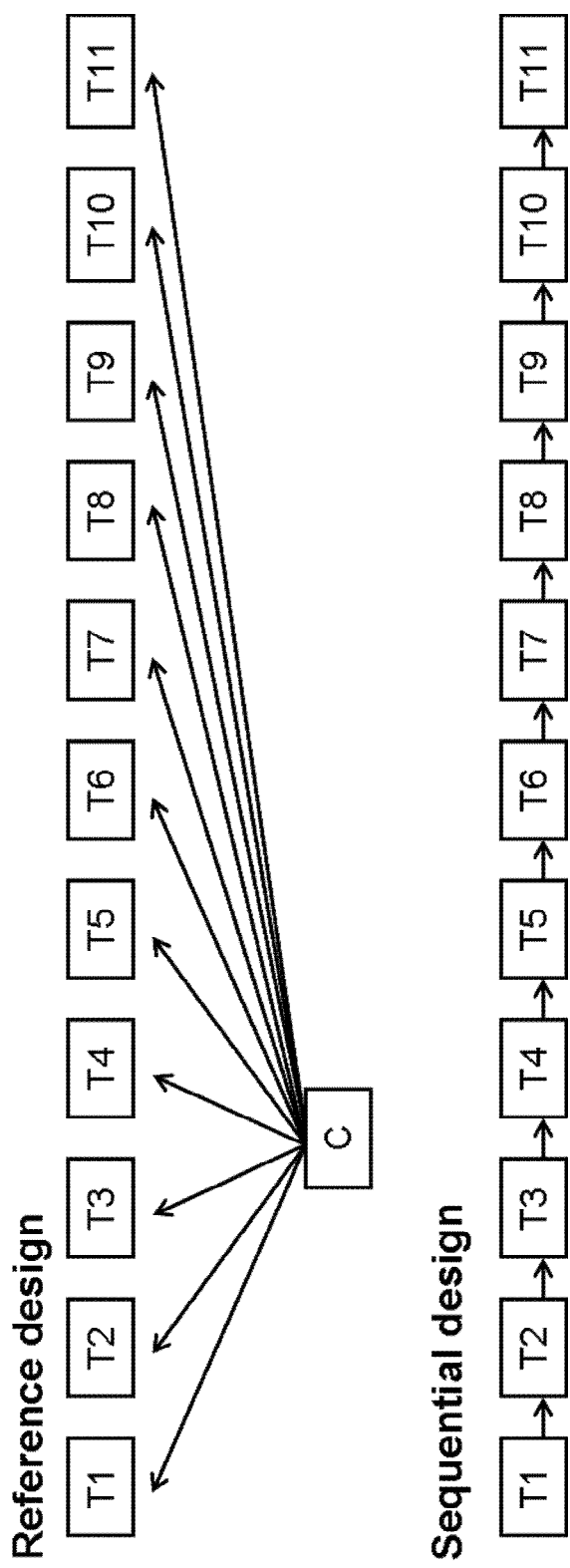
本論文の内容について一部、以下の学会等で発表した。

【口頭発表】

1. 第5回情報処理学会バイオ情報学研究会
時系列マイクロアレイ実験デザインの発現解析に及ぼす影響
伊藤遼佑・高橋弘喜・大島拓・小笠原 直毅・Md. Altaf-Ul-Amin・金谷 重彦・黒川 顕
沖縄先端科学技術大学院大学ゲストハウス
2. Experimental Design for Time-Series Microarray Analysis
Ryosuke Ito, Hiroki Takahashi, Taku Oshima, Naotake Ogasawara, Md. Altaf-Ul-Amin,
Ken Kurokawa, Shigehiko Kanaya
Bioinformatics Department of Max Planck Institute for Molecular Plant Physiology

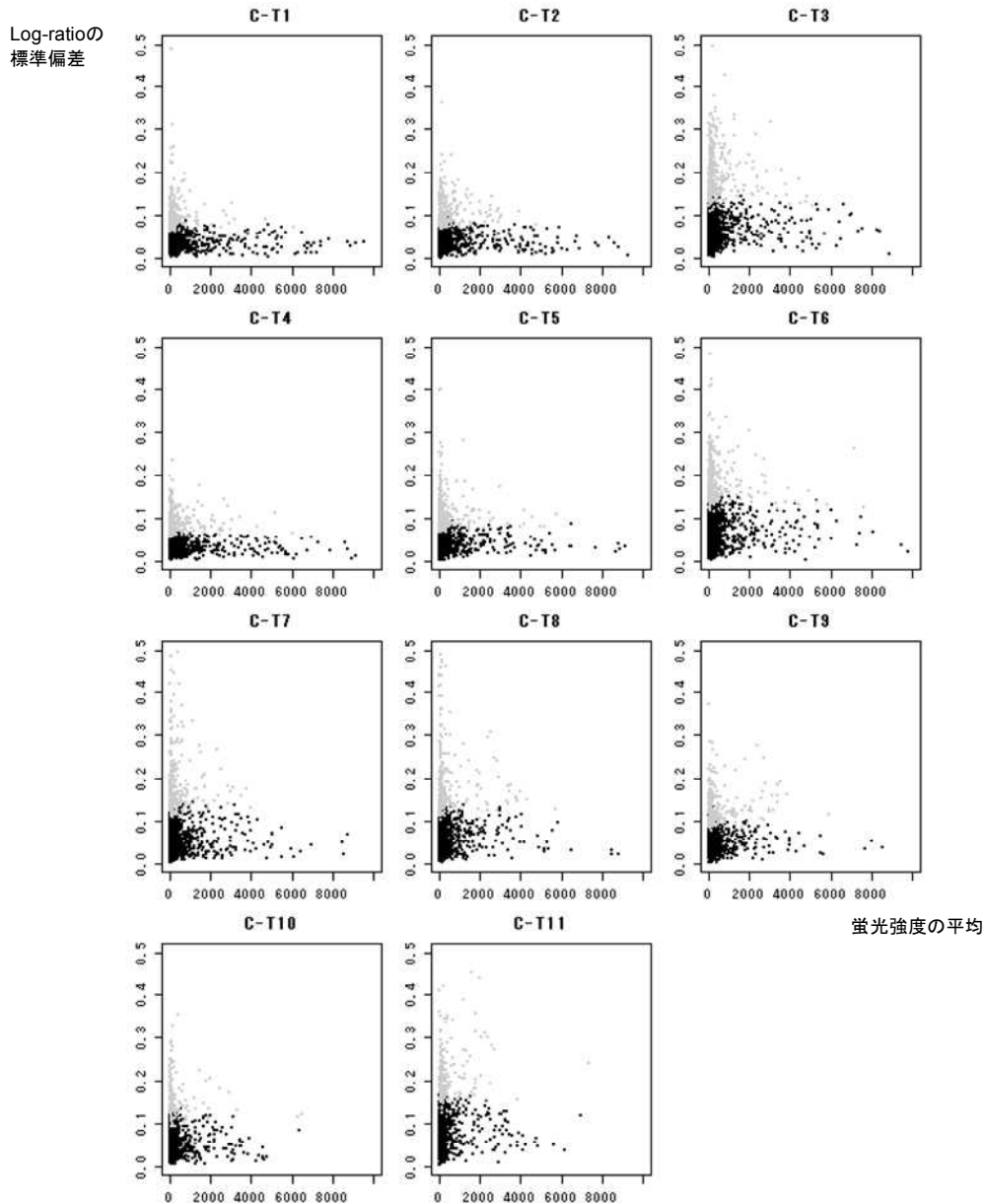
【ポスター発表】

3. The 17th International Conference on Genome Informatics. (GIW2006)
Experimental Design for Time-Series Microarray Analysis
Ryosuke Ito, Hiroki Takahashi, Taku Oshima, Naotake Ogasawara, Md. Altaf-Ul-Amin, Ken Kurokawa,
Shigehiko Kanaya
PACIFICO YOKOHAMA
4. The 5th Asia Pacific Bioinformatics Conference
Experimental Design for Time-Series Microarray Analysis
Ryosuke Ito, Hiroki Takahashi, Taku Oshima, Naotake Ogasawara, Md. Altaf-Ul-Amin, Ken Kurokawa,
Shigehiko Kanaya
The university of Hong Kong



補足図1:タイムポイントとマイクロアレイデザイン

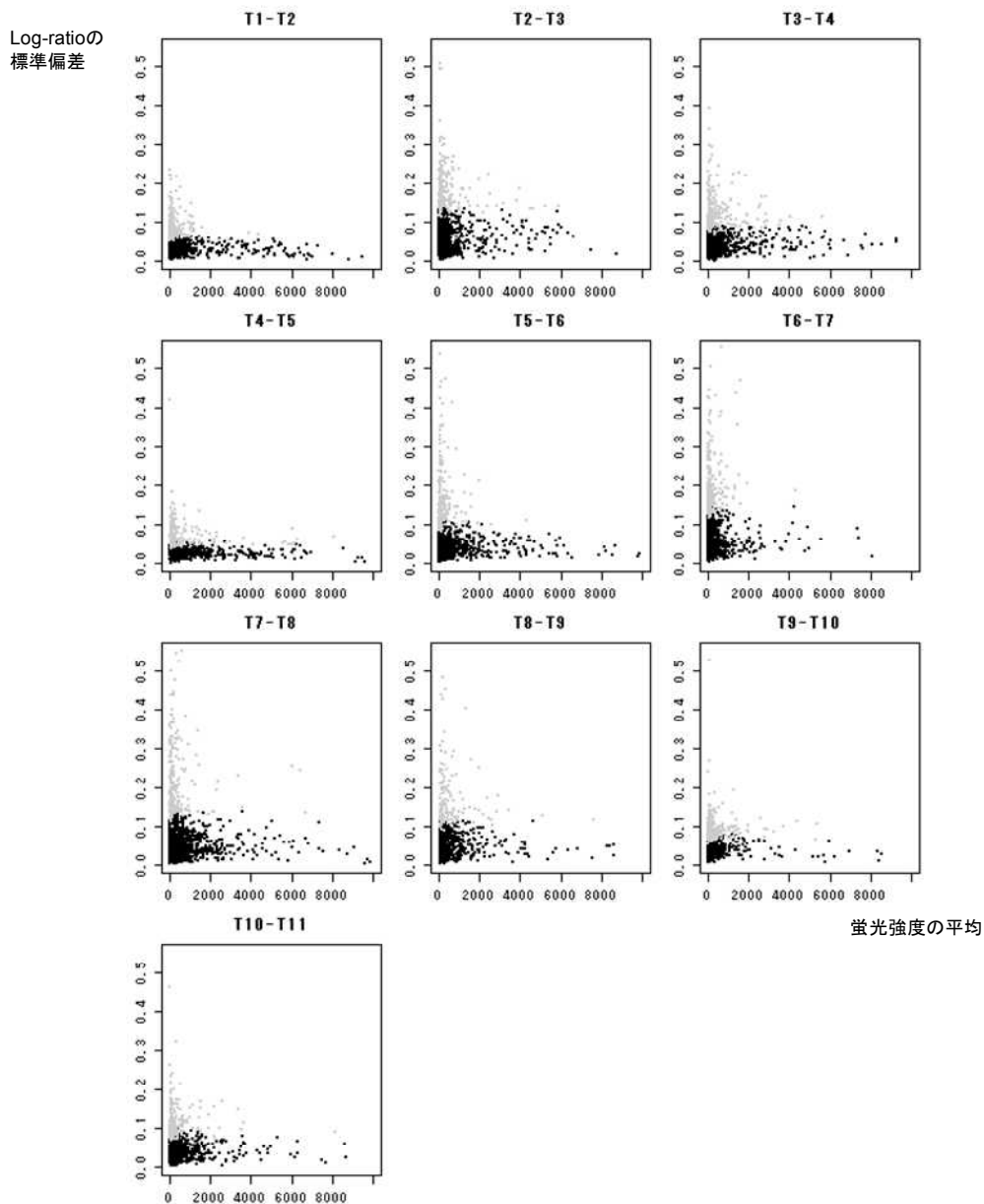
リファレンス・デザインでのLog-ratioの標準偏差の分布



各図は、マイクロアレイデータごとの蛍光強度の大きさに対するLog-ratioのばらつき方をプロットしたものである。横軸は蛍光強度の平均を表し、 $A=(Cy3+Cy5)/2$ である。縦軸は遺伝子ごとのLog-ratioの標準偏差を表す。この時点で、微小な蛍光強度は除いているにもかかわらず、蛍光強度の小さな領域に不良な値が多く存在することがわかる。各図で灰色でプロットしてあるのは、繰り返し実験間で不整合なデータとして除去されたものである。除去後のデータのばらつき方は安定していると考えられる。

補足図2 : 蛍光強度の大きさとLog-ratioのばらつきの関係

シーケンシャル・デザインでのLog-ratioの標準偏差の分布



各図は、マイクロアレイデータごとの蛍光強度の大きさに対するLog-ratioのばらつき方をプロットしたものである。横軸は蛍光強度の平均を表し、 $A=(Cy3+Cy5)/2$ である。縦軸は遺伝子ごとのLog-ratioの標準偏差を表す。この時点で、微小な蛍光強度は除いているにもかかわらず、蛍光強度の小さな領域に不良な値が多く存在することがわかる。各図で灰色でプロットしてあるのは、繰り返し実験間で不整合なデータとして除去されたものである。除去後のデータのばらつき方は安定していると考えられる。

補足図2 : 蛍光強度の大きさとLog-ratioのばらつきの関係

付表1 :各タイムポイント間で発現上昇と判定された遺伝子

b no.	gene	cogID	function.
T1-T2			
b0001	<i>thrL</i>	-	thr operon leader peptide
b0104	<i>guaC</i>	COG0516F	guanosine 5'-monophosphate oxidoreductase
b0131	<i>panD</i>	-	aspartate 1-decarboxylase precursor
b0145	<i>dkxA</i>	COG1734T	DNA-binding transcriptional regulator of rRNA transcription, DnaK suppressor protein
b0222	<i>lpcA</i>	COG0279G	phosphoheptose isomerase
b0259,b0552,b0656,b1331,b2030,b2192,b2982,b3218,b3505	<i>insH-1,insH-2,insH-3,insH-4,insH-7,insH-8,insH-9,insH-10,insH-11</i>	COG3039L,C OG3039L,CO G3039L,COG3 039L,COG303 9L,COG3039L ,COG3039L,C OG3039L,CO	CP4-6 prophage; IS5 transposase and trans-activator ,DLP12 prophage; IS5 transposase and trans-activator ,IS5 transposase and trans-activator ,IS5 transposase and trans-activator ,IS5 transposase and trans-activator ,IS5 transposase and trans-activator ,IS5 transposase and trans-activator ,IS5 transposase and trans-activator
b0334	<i>prpD</i>	-	2-methylcitrate dehydratase
b0432	<i>cyoA</i>	COG1622C	cytochrome o ubiquinol oxidase subunit II
b0814	<i>ompX</i>	COG3637M	outer membrane protein X
b0963	<i>mgsA</i>	-	methylglyoxal synthase
b1101	<i>ptsG</i>	COG1263G	fused glucose-specific PTS enzymes: IIB component/IIC component
b1174	<i>minE</i>	COG0851D	cell division topological specificity factor MinE
b1241	<i>adhE</i>	COG1012C, COG1454C	fused acetaldehyde-CoA dehydrogenase/iron-dependent alcohol dehydrogenase/pyruvate-formate lyase deactivase
b1370	<i>insH-5</i>	COG3039L	Rac prophage; IS5 transposase and trans-activator
b1479	<i>sfcA</i>	COG0281C	malate dehydrogenase, (decarboxylating, NAD-requiring) (malic enzyme)
b1614	<i>ydgA</i>	COG5339S	hypothetical protein b1614
b1641	<i>slyB</i>	COG3133M	outer membrane lipoprotein
b1808	<i>yoaA</i>	COG1199KL	conserved protein with nucleoside triphosphate hydrolase domain
b1913	<i>uvrC</i>	COG0322L	excinuclease ABC subunit C
b1914	<i>uvrY</i>	COG2197TK	response regulator
b1920	<i>fliY</i>	COG0834ET	cystine transporter subunit
b2175	<i>spr</i>	COG0791M	predicted peptidase, outer membrane lipoprotein
b2277	<i>nuoM</i>	COG1008C	NADH dehydrogenase subunit M
b2278	<i>nuoL</i>	COG1009CP	NADH dehydrogenase subunit L
b2279	<i>nuoK</i>	COG0713C	NADH dehydrogenase subunit K
b2282	<i>nuoH</i>	COG1005C	NADH dehydrogenase subunit H
b2415	<i>ptsH</i>	COG1925G	phosphohistidinoprotein-hexose phosphotransferase component of PTS system (Hpr)
b2431	<i>yfeX</i>	COG2837P	hypothetical protein b2431
b2453	<i>eutG</i>	COG1454C	predicted alcohol dehydrogenase in ethanolamine utilization
b2500	<i>purN</i>	-	phosphoribosylglycinamide formyltransferase
b2553	<i>glnB</i>	COG0347E	regulatory protein P-II for glutamine synthetase
b2579	<i>yfiD</i>	COG3445R	pyruvate formate lyase subunit
b2618	<i>yjfF</i>	COG2914S	hypothetical protein b2618
b2619	<i>yjfG</i>	COG2867I	hypothetical protein b2619
b2684	<i>mprA</i>	COG1846K	DNA-binding transcriptional repressor of microcin B17 synthesis and multidrug efflux
b2750	<i>cysC</i>	COG0529P	adenylylsulfate kinase
b2797	<i>sdaB</i>	COG1760E	L-serine deaminase II
b2926	<i>pgk</i>	-	phosphoglycerate kinase
b3049	<i>glgS</i>	-	glycogen synthesis protein GlgS
b3357	<i>crp</i>	COG0664T	DNA-binding transcriptional dual regulator
b3748	<i>rbsD</i>	COG1869G	predicted cytoplasmic sugar-binding protein
b3891	<i>fdhE</i>	-	formate dehydrogenase accessory protein FdhE
b3893	<i>fdoH</i>	COG0437C	formate dehydrogenase-O, Fe-S subunit
b4142	<i>groS</i>	COG0234O	co-chaperonin GroES
b4144	<i>yjeI</i>	-	hypothetical protein b4144
b4232	<i>fbp</i>	COG0158G	fructose-1,6-bisphosphatase
b4466	<i>yghJ</i>	-	predicted inner membrane lipoprotein
b4471	<i>tdcG</i>	COG1760E	L-serine dehydratase 3
T2-T3			
b0028	<i>fkpB</i>	COG1047O	FKBP-type peptidyl-prolyl cis-trans isomerase (rotamase)
b0583	<i>entD</i>	COG2977Q	phosphopantetheinyltransferase component of enterobactin synthase multienzyme complex
b0586	<i>entF</i>	COG1020Q, COG3319Q	enterobactin synthase multienzyme complex component, ATP-dependent
b0595	<i>entB</i>	COG1535Q, COG3433Q	isochorismatase
b0756	<i>galM</i>	COG2017G	galactose-1-epimerase (mutarotase)
b0757	<i>galK</i>	COG0153G	galactokinase
b0928	<i>aspC</i>	COG1448E	aspartate aminotransferase, PLP-dependent
b0970	<i>yccA</i>	COG0670R	inner membrane protein
b1101	<i>ptsG</i>	COG1263G	fused glucose-specific PTS enzymes: IIB component/IIC component
b1604	<i>ydgH</i>	-	hypothetical protein b1604
b1905	<i>fn</i>	COG1528P	ferritin iron storage protein (cytoplasmic)
b2155	<i>cirA</i>	COG4771P	ferric iron-catecholate outer membrane transporter

b2209	<i>eco</i>	-	ecotin precursor
b2277	<i>nuoM</i>	COG1008C	NADH dehydrogenase subunit M
b2278	<i>nuoL</i>	COG1009CP	NADH dehydrogenase subunit L
b2279	<i>nuoK</i>	COG0713C	NADH dehydrogenase subunit K
b2281	<i>nuoI</i>	COG1143C	NADH dehydrogenase subunit I
b2283	<i>nuoG</i>	COG1034C	NADH dehydrogenase subunit G
b3097	<i>yqjC</i>	-	hypothetical protein b3097
b3257	<i>yhdT</i>	-	conserved inner membrane protein
b3347	<i>fkpA</i>	COG0545O	FKBP-type peptidyl-prolyl cis-trans isomerase (rotamase)
b3506	<i>slp</i>	-	outer membrane lipoprotein
b3754	<i>hsrA</i>	-	predicted multidrug or homocysteine efflux system
b3925	<i>glpX</i>	-	fructose 1,6-bisphosphatase II
b4232	<i>fbp</i>	COG0158G	fructose-1,6-bisphosphatase
b4238	<i>nrdD</i>	COG1328F	anaerobic ribonucleoside triphosphate reductase
b4239	<i>treC</i>	-	trehalose-6-P hydrolase
b4322	<i>uxuA</i>	COG1312G	mannonate dehydratase
b4381	<i>deoC</i>	COG0274F	deoxyribose-phosphate aldolase
b4392	<i>slt</i>	COG0741M	lytic murein transglycosylase, soluble

T3-T4

b0121	<i>speE</i>	COG0421E	spermidine synthase
b0159	<i>null</i>	COG0775F	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase
b0169	<i>rpsB</i>	-	30S ribosomal protein S2
b0307	<i>ykgF</i>	COG1139C	predicted amino acid dehydrogenase with NAD(P)-binding domain and ferridoxin-like domain
b0355	<i>frmB</i>	COG0627R	predicted esterase
b0380	<i>yaiZ</i>	-	predicted inner membrane protein
b0481	<i>ybaK</i>	COG2606S	hypothetical protein b0481
b0488	<i>ybbJ</i>	COG1585OU	conserved inner membrane protein
b0614	<i>citX</i>	COG3697HI	2'-(5'-triphosphoribosyl)-3'-dephospho-CoA:apo-citrate lyase
b0750	<i>nadA</i>	-	quinolinate synthetase
b0809	<i>glnQ</i>	COG1126E	glutamine ABC transporter ATP-binding protein
b0826	<i>moeB</i>	COG0476H	molybdopterin biosynthesis protein MoeB
b0852	<i>rimK</i>	COG0189HJ	ribosomal protein S6 modification protein
b0928	<i>aspC</i>	COG1448E	aspartate aminotransferase, PLP-dependent
b0970	<i>yccA</i>	COG0670R	inner membrane protein
b0992	<i>yccM</i>	COG0348C	predicted 4Fe-4S membrane protein
b1101	<i>ptsG</i>	COG1263G	fused glucose-specific PTS enzymes: IIB component/IIC component
b1241	<i>adhE</i>	COG1012C, COG1454C	fused acetaldehyde-CoA dehydrogenase/iron-dependent alcohol dehydrogenase/pyruvate-formate lyase deactivase
b1281	<i>pyrF</i>	COG0284F	orotidine 5'-phosphate decarboxylase
b1282	<i>yciH</i>	-	translation initiation factor Sui1
b1286	<i>rnb</i>	-	exoribonuclease II
b1288	<i>fabI</i>	COG0623I	enoyl-(acyl carrier protein) reductase
b1326	<i>mpaA</i>	COG2866E, COG3608R	murein peptide amidase A
b1563	<i>relE</i>	COG2026JD	Qin prophage; toxin of the RelE-RelB toxin-antitoxin system
b1564	<i>relB</i>	COG3077L	Qin prophage; bifunctional antitoxin of the RelE-RelB toxin-antitoxin system/ transcriptional repressor
b1618	<i>uidR</i>	COG1309K	DNA-binding transcriptional repressor
b1656	<i>sodB</i>	COG0605P	superoxide dismutase, Fe
b1711	<i>btuC</i>	COG4139H	vitamin B12-transporter permease
b1866	<i>aspS</i>	COG0173J	aspartyl-tRNA synthetase
b2146	<i>yeiT</i>	COG0493ER	predicted oxidoreductase
b2172	<i>yeiQ</i>	-	predicted dehydrogenase, NAD-dependent
b2215	<i>ompC</i>	COG3203M	outer membrane porin protein C
b2276	<i>nuoN</i>	COG1007C	NADH dehydrogenase subunit N
b2309	<i>hisJ</i>	COG0834ET	histidine/lysine/arginine/ornithine transporter subunit
b3294	<i>rplQ</i>	COG0203J	50S ribosomal protein L17
b3340	<i>fusA</i>	COG0480J	elongation factor EF-2

T4-T5

b0001	<i>thrL</i>	-	thr operon leader peptide
b0004	<i>thrC</i>	COG0498E	threonine synthase
b0118	<i>acnB</i>	COG1049C	aconitate hydratase
b0146	<i>sfsA</i>	-	sugar fermentation stimulation protein A
b0240	<i>crI</i>	-	DNA-binding transcriptional regulator
b0361,b1402,b 1578,b1996,b2 860,b3045,b42 73	<i>insD-1,insD- 2,insD-7,insD- 3,insD-4,insD- 5,insD-6</i>	COG2801L,C OG2801L,CO G2801L,COG2 801L,COG280 1L,COG2801L COG2801L	IS2 insertion element transposase InsAB', IS2 insertion element transposase InsAB', QIn prophage; IS2 insertion element transposase InsAB', C-ter fragment (pseudogene), CP4-44 prophage; IS2 insertion element transposase InsAB', IS2 insertion element transposase InsAB', IS2 insertion element transposase InsAB', KpLE2 phage-like element; IS2 insertion element transposase InsAB'
b0517	<i>allD</i>	-	ureidoglycolate dehydrogenase
b0658	<i>ybeX</i>	COG4535P	predicted ion transport

b0720	<i>gltA</i>	COG0372C	citrate synthase
b0723	<i>sdhA</i>	COG1053C	succinate dehydrogenase flavoprotein subunit
b0724	<i>sdhB</i>	COG0479C	succinate dehydrogenase, FeS subunit
b0755	<i>gpmA</i>	COG0588G	phosphoglyceromutase
b0811	<i>glnH</i>	COG0834ET	glutamine ABC transporter periplasmic protein
b0889	<i>lrp</i>	COG1522K	DNA-binding transcriptional dual regulator, leucine-binding
b0907	<i>serC</i>	COG1932HE	phosphoserine aminotransferase
b0937	<i>ssuE</i>	COG0431R	NAD(P)H-dependent FMN reductase
b1056	<i>yceI</i>	-	hypothetical protein b1056
b1136	<i>icd</i>	COG0538C	isocitrate dehydrogenase
b1190	<i>dadX</i>	COG0787M	alanine racemase
b1243	<i>oppA</i>	COG4166E	oligopeptide transporter subunit
b1247	<i>oppF</i>	COG4608E	oligopeptide transporter subunit
b1324	<i>tpx</i>	COG2077O	thiol peroxidase
b1415	<i>aldA</i>	COG1012C	aldehyde dehydrogenase A, NAD-linked
b1431	<i>ydcL</i>	-	predicted lipoprotein
b1452	<i>yncE</i>	COG3391S	hypothetical protein b1452
b1619	<i>hdhA</i>	COG10281QR	7-alpha-hydroxysteroid dehydrogenase
b1777	<i>yeaC</i>	-	hypothetical protein b1777
b1832	<i>yebR</i>	COG1956T	hypothetical protein b1832
b2094	<i>gatA</i>	COG1762GT	galactitol-specific enzyme IIA component of PTS
b2096	<i>gatY</i>	COG0191G	D-tagatose 1,6-bisphosphate aldolase 2, catalytic subunit
b2414	<i>cysK</i>	COG0031E	cysteine synthase A, O-acetylserine sulphydrylase A subunit
b2463	<i>maeB</i>	COG0280C, COG0281C	malic enzyme
b2600	<i>tyrA</i>	COG0287E, COG1605E	fused chorismate mutase T/prephenate dehydrogenase
b2703	<i>srIE</i>	-	glucitol/sorbitol-specific enzyme IIB component of PTS
b2704	<i>srIB</i>	-	glucitol/sorbitol-specific enzyme IIA component of PTS
b2705	<i>srID</i>	COG10281QR	3-ketoacyl-(acyl-carrier-protein) reductase
b2749	<i>ygbE</i>	-	conserved inner membrane protein
b2904	<i>gcvH</i>	COG0509E	glycine cleavage system protein H
b2905	<i>gcvT</i>	COG0404E	glycine cleavage system aminomethyltransferase T
b2913	<i>serA</i>	COG0111HE	D-3-phosphoglycerate dehydrogenase
b2979	<i>gldD</i>	COG0277C	glycolate oxidase subunit, FAD-linked
b3005	<i>exbD</i>	COG0848U	membrane spanning protein in TonB-ExbB-ExbD complex
b3358	<i>yhfK</i>	COG1289S	conserved inner membrane protein
b3403	<i>pck</i>	COG1866C	phosphoenolpyruvate carboxykinase
b3417	<i>malP</i>	COG0058G	maltodextrin phosphorylase
b3429	<i>glgA</i>	COG0297G	glycogen synthase
b3433	<i>asd</i>	COG0136E	aspartate-semialdehyde dehydrogenase
b3607	<i>cysE</i>	COG1045E	serine acetyltransferase
b3748	<i>rbsD</i>	COG1869G	predicted cytoplasmic sugar-binding protein
b3754	<i>hsrA</i>	-	predicted multidrug or homocysteine efflux system
b3847	<i>pepQ</i>	COG0006E	proline dipeptidase
b3888	<i>yjiD</i>	-	predicted acetyltransferase
b3908	<i>sodA</i>	COG0605P	superoxide dismutase, Mn
b3939	<i>metB</i>	COG0626E	cystathionine gamma-synthase
b4015	<i>aceA</i>	COG2224C	isocitrate lyase
b4142	<i>groS</i>	COG0234O	co-chaperonin GroES
b4222	<i>ytfP</i>	COG2105S	hypothetical protein b4222
b4232	<i>fbp</i>	COG0158G	fructose-1,6-bisphosphatase
b4239	<i>treC</i>	-	trehalose-6-P hydrolase
b4291	<i>fecA</i>	COG4772P	KpLE2 phage-like element; ferric citrate outer membrane transporter
b4332	<i>yjiJ</i>	-	predicted inner membrane protein
b4367	<i>fhuF</i>	COG4114R	ferric iron reductase involved in ferric hydroximate transport
b4401	<i>arcA</i>	COG0745TK	DNA-binding response regulator in two-component regulatory system with ArcB or CpxA

T5-T6

b0006	<i>yaaA</i>	COG3022S	hypothetical protein b0006
b0020	<i>nhaR</i>	COG0583K	DNA-binding transcriptional activator
b0104	<i>guaC</i>	COG0516F	guanosine 5'-monophosphate oxidoreductase
b0118	<i>acnB</i>	COG1049C	aconitate hydratase
b0154	<i>hemL</i>	COG0001H	glutamate-1-semialdehyde aminotransferase
b0171	<i>pyrH</i>	COG0528F	uridylate kinase
b0337	<i>codA</i>	COG0402FR	cytosine deaminase
b0415	<i>ribE</i>	-	riboflavin synthase subunit beta
b0431	<i>cyoB</i>	COG0843C	cytochrome o ubiquinol oxidase subunit I
b0452	<i>tesB</i>	COG1946I	acyl-CoA thioesterase II
b0684	<i>fldA</i>	COG0716C	flavodoxin I
b0779	<i>uvrB</i>	COG0556L	excinuclease ABC subunit B
b0888	<i>trxB</i>	COG0492O	thioredoxin reductase, FAD/NAD(P)-binding
b1004	<i>wrbA</i>	COG0655R	TrpR binding protein WrbA

b1049	<i>mdoH</i>	COG2943M	glucosyltransferase MdoH
b1063	<i>yceB</i>	-	predicted lipoprotein
b1108	<i>ycjP</i>	COG3150R	hypothetical protein b1108
b1131	<i>purB</i>	COG0015F	adenylosuccinate lyase
b1136	<i>icd</i>	COG0538C	isocitrate dehydrogenase
b1232	<i>purU</i>	COG0788F	formyltetrahydrofolate deformylase
b1243	<i>oppA</i>	COG4166E	oligopeptide transporter subunit
b1272	<i>sohB</i>	COG06160U	predicted inner membrane peptidase
b1273	<i>yciN</i>	-	hypothetical protein b1273
b1282	<i>yciH</i>	-	translation initiation factor Sui1
b1295	<i>ymjA</i>	-	hypothetical protein b1295
b1378	<i>ydbK</i>	COG0674C, COG1013C, COG1014C	fused predicted pyruvate-flavodoxin oxidoreductase: conserved protein/conserved protein/FeS binding protein
b1479	<i>sfcA</i>	COG0281C	malate dehydrogenase, (decarboxylating, NAD-requiring) (malic enzyme)
b1584	<i>speG</i>	COG1670J	spermidine N1-acetyltransferase
b1654	<i>ydhD</i>	COG0278O	hypothetical protein b1654
b1661	<i>cfa</i>	COG2230M	cyclopropane fatty acyl phospholipid synthase (unsaturated-phospholipid methyltransferase)
b1702	<i>pps</i>	COG0574G	phosphoenolpyruvate synthase
b1794	<i>yeaP</i>	-	predicted diguanylate cyclase
b1814	<i>sdaA</i>	COG1760E	L-serine deaminase I
b1905	<i>fin</i>	COG1528P	ferritin iron storage protein (cytoplasmic)
b1913	<i>uvrC</i>	COG0322L	excinuclease ABC subunit C
b1919	<i>yedO</i>	COG2515E	D-cysteine desulfhydrase
b1952	<i>dsrB</i>	-	hypothetical protein b1952
b2016	<i>yeeZ</i>	COG0451MG	predicted epimerase, with NAD(P)-binding Rossmann-fold domain
b2029	<i>gnd</i>	COG0362G	6-phosphogluconate dehydrogenase
b2049	<i>cpsB</i>	COG0662G, COG0836M	mannose-1-phosphate guanyltransferase
b2066	<i>udk</i>	COG0572F	uridine kinase
b2113	<i>mrp</i>	COG0489D	antiporter inner membrane protein
b2153	<i>folE</i>	COG0302H	GTP cyclohydrolase I
b2154	<i>yeiG</i>	COG0627R	predicted esterase
b2231	<i>gyrA</i>	COG0188L	DNA gyrase subunit A
b2300	<i>yfcE</i>	COG0622R	phosphodiesterase
b2324	<i>trmC</i>	COG0665E, COG4121S	hypothetical protein b2324
b2326	<i>yfcM</i>	-	hypothetical protein b2326
b2346	<i>vacJ</i>	-	predicted lipoprotein
b2496	<i>hda</i>	COG0593L	DNA replication initiation factor
b2498	<i>upp</i>	COG0035F	uracil phosphoribosyltransferase
b2499	<i>purM</i>	COG0150F	phosphoribosylaminoimidazole synthetase
b2509	<i>xseA</i>	COG1570L	exodeoxyribonuclease VII large subunit
b2523	<i>pepB</i>	-	aminopeptidase B
b2560	<i>yfhB</i>	COG0560E	hypothetical protein b2560
b2610	<i>ffh</i>	COG0541U	Signal Recognition Particle (SRP) component with 4.5S RNA (ffs)
b2780	<i>pyrG</i>	COG0504F	CTP synthetase
b2908	<i>pepP</i>	COG0006E	proline aminopeptidase P II
b3018	<i>plsC</i>	COG0204I	1-acyl-sn-glycerol-3-phosphate acyltransferase
b3052	<i>rfaE</i>	COG2870M	fused heptose 7-phosphate kinase/heptose 1-phosphate adenylyltransferase
b3065	<i>rpsU</i>	COG0828J	30S ribosomal protein S21
b3203	<i>yhbH</i>	COG1544J	predicted ribosome-associated, sigma 54 modulation protein
b3204	<i>ptsN</i>	COG1762GT	sugar-specific enzyme IIA component of PTS
b3425	<i>glpE</i>	COG0607P	thiosulfate sulfurtransferase
b3648	<i>gmk</i>	COG0194F	guanylate kinase
b3729	<i>glmS</i>	COG0449M	D-fructose-6-phosphate amidotransferase
b3764	<i>yifE</i>	-	hypothetical protein b3764
b3858	<i>yihD</i>	-	hypothetical protein b3858
b3908	<i>sodA</i>	COG0605P	superoxide dismutase, Mn
b3961	<i>oxyR</i>	COG0583K	DNA-binding transcriptional dual regulator
b4145	<i>yjeJ</i>	-	hypothetical protein b4145
b4328	<i>iadA</i>	COG0044F, COG1820G	isoaspartyl dipeptidase
b4373	<i>rimI</i>	COG0456R	ribosomal-protein-alanine N-acetyltransferase
b4483	<i>tatD</i>	COG0084L	DNase, magnesium-dependent
b4493,b1416	<i>gapC</i>	-	-

T6-T7

b0187	<i>yaeR</i>	COG0346E	predicted lyase
b0189	<i>rof</i>	-	modulator of Rho-dependent transcription termination
b0280	<i>yagN</i>	-	CP4-6 prophage; predicted protein
b0376	<i>ampH</i>	COG1680V	beta-lactamase/D-alanine carboxypeptidase
b0380	<i>yaiZ</i>	-	predicted inner membrane protein

b0460	<i>hha</i>	-	modulator of gene expression, with H-NS
b0476	<i>aes</i>	COG0657I	acetyl esterase
b0652	<i>gltL</i>	COG1126E	glutamate and aspartate transporter subunit
b0712	<i>ybgK</i>	-	predicted enzyme subunit
b0729	<i>sucD</i>	COG0074C	succinyl-CoA synthetase subunit alpha
b0750	<i>nadA</i>	-	quinolinate synthetase
b0791	<i>ybhQ</i>	-	predicted inner membrane protein
b0865	<i>ybjP</i>	-	predicted lipoprotein
b0932	<i>pepN</i>	COG0308E	aminopeptidase N
b0964	<i>yccT</i>	-	hypothetical protein b0964
b1150	<i>ymfR</i>	-	e14 prophage; predicted protein
b1175	<i>minD</i>	COG2894D	membrane ATPase of the MinC-MinD-MinE system
b1325	<i>ycjG</i>	COG4948MR	L-Ala-D/L-Glu epimerase
b1329	<i>mppA</i>	COG4166E	murein tripeptide (L-ala-gamma-D-glutamyl-meso-DAP) transporter subunit
b1333	<i>uspE</i>	COG0589T	stress-induced protein
b1399	<i>paaX</i>	COG3327K	DNA-binding transcriptional repressor of phenylacetic acid degradation, aryl-CoA responsive
b1511	<i>lsrK</i>	COG1070G	autoinducer-2 (AI-2) kinase
b1594	<i>dgsA</i>	COG1940KG	DNA-binding transcriptional repressor
b1614	<i>ydgA</i>	COG5339S	hypothetical protein b1614
b1682	<i>sufC</i>	COG0396O	cysteine desulfurase ATPase component
b1819	<i>manZ</i>	-	mannose-specific enzyme IID component of PTS
b1853	<i>yebK</i>	COG1737K	predicted DNA-binding transcriptional regulator
b2076	<i>mdtC</i>	COG0841V	multidrug efflux system, subunit C
b2095	<i>gatZ</i>	COG4573G	D-tagatose 1,6-bisphosphate aldolase 2, subunit
b2170	<i>setB</i>	-	lactose/glucose efflux system
b2244	<i>yfaD</i>	COG5464S	hypothetical protein b2244
b2308	<i>hisQ</i>	COG4215E	histidine/lysine/arginine/ornithine transporter subunit
b2309	<i>hisJ</i>	COG0834ET	histidine/lysine/arginine/ornithine transporter subunit
b2365	<i>dsdX</i>	COG2610GE	predicted transporter
b2388	<i>glk</i>	COG0837G	glucokinase
b2463	<i>maeB</i>	COG0280C, COG0281C	malic enzyme
b2521	<i>sseA</i>	COG2897P	3-mercaptopyruvate sulfurtransferase
b2552	<i>hmp</i>	COG1017C, COG1018C	fused nitric oxide dioxygenase/dihydropteridine reductase 2
b2572	<i>rseA</i>	COG3073T	anti-sigma factor
b2584	<i>yfiQ</i>	COG1042C	fused predicted acyl-CoA synthetase: NAD(P)-binding subunit/ATP-binding subunit
b2600	<i>tyrA</i>	COG0287E, COG1605E	fused chorismate mutase T/prephenate dehydrogenase
b2661	<i>gabD</i>	COG1012C	succinate-semialdehyde dehydrogenase I, NADP-dependent
b2705	<i>srhD</i>	COG10281QR	3-ketoacyl-(acyl-carrier-protein) reductase
b2741	<i>rpoS</i>	COG0568K	RNA polymerase sigma factor
b2833	<i>ygdR</i>	-	hypothetical protein b2833
b3023	<i>ygiV</i>	COG3449L	predicted transcriptional regulator
b3084	<i>ygiO</i>	COG2813J	predicted methyltransferase small domain
b3089	<i>sstT</i>	COG3633E	sodium:serine/threonine symporter
b3269	<i>yhdX</i>	COG4597E	predicted amino-acid transporter subunit
b3417	<i>malP</i>	COG0058G	maltodextrin phosphorylase
b3428	<i>glgP</i>	COG0058G	glycogen phosphorylase
b3448	<i>yhhA</i>	-	hypothetical protein b3448
b3495	<i>uspA</i>	COG0589T	universal stress global response regulator
b3601	<i>mtlR</i>	-	DNA-binding repressor
b3672	<i>ivbL</i>	-	ilvB operon leader peptide
b3678	<i>yidJ</i>	COG3119P	predicted sulfatase/phosphatase
b3708	<i>maA</i>	COG3033E	tryptophanase/L-cysteine desulfhydrase, PLP-dependent
b3748	<i>rbsD</i>	COG1869G	predicted cytoplasmic sugar-binding protein
b3752	<i>rbsK</i>	COG0524G	ribokinase
b3862	<i>yihG</i>	COG0204I	predicted endonuclease
b3874	<i>yihN</i>	-	predicted transporter
b3888	<i>yiiD</i>	-	predicted acetyltransferase
b3922	<i>yiiS</i>	-	hypothetical protein b3922
b4035	<i>malK</i>	COG3839G	fused maltose transport subunit, ATP-binding component of ABC superfamily/regulatory protein
b4059	<i>ssb</i>	COG0629L	single-strand DNA-binding protein
b4117	<i>adiA</i>	COG1982E	biodegradative arginine decarboxylase
b4178	<i>yjeB</i>	COG1959K	predicted DNA-binding transcriptional regulator
b4208	<i>cycA</i>	COG1113E	D-alanine/D-serine/glycine transporter
b4214	<i>cysQ</i>	COG1218P	PAPS (adenosine 3'-phosphate 5'-phosphosulfate) 3'(2'),5'-bisphosphate nucleotidase
b4401	<i>arcA</i>	COG0745TK	DNA-binding response regulator in two-component regulatory system with ArcB or CpxA
b4470	<i>yhaM</i>	-	hypothetical protein b4470

T7-T8

b0023	<i>rpsT</i>	COG0268J	30S ribosomal protein S20
b0025	<i>ribF</i>	COG0196H	hypothetical protein b0025

b0121	<i>speE</i>	COG0421E	spermidine synthase
b0171	<i>pyrH</i>	COG0528F	uridylate kinase
b0179	<i>lpxD</i>	COG1044M	UDP-3-O-
b0211	<i>mltD</i>	COG0741M	predicted membrane-bound lytic murein transglycosylase D
b0336	<i>codB</i>	COG1457F	cytosine transporter
b0388	<i>aroL</i>	COG0703E	shikimate kinase II
b0410	<i>yajD</i>	-	hypothetical protein b0410
b0421	<i>ispA</i>	COG0142H	geranyltransferase
b0484	<i>copA</i>	COG2217P	copper transporter
b0595	<i>entB</i>	COG1535Q, COG3433Q	isochorismatase
b0658	<i>ybeX</i>	COG4535P	predicted ion transport
b0710	<i>ybgI</i>	COG0327S	conserved metal-binding protein
b0740	<i>tolB</i>	COG0823U	translocation protein TolB precursor
b0741	<i>pal</i>	COG2885M	peptidoglycan-associated outer membrane lipoprotein
b0742	<i>ybgF</i>	COG1729S	hypothetical protein b0742
b0864	<i>artP</i>	COG4161E	arginine transporter subunit
b0911	<i>rpsA</i>	COG0539J	30S ribosomal protein S1
b0927	<i>ycbL</i>	COG0491R	predicted metal-binding enzyme
b0964	<i>yccT</i>	-	hypothetical protein b0964
b0967	<i>yccW</i>	COG1092R	predicted methyltransferase
b0970	<i>yccA</i>	COG0670R	inner membrane protein
b1045	<i>ymdB</i>	-	hypothetical protein b1045
b1091	<i>fabH</i>	COG0332I	3-oxoacyl-(acyl carrier protein) synthase
b1207	<i>prsA</i>	COG0462FE	ribose-phosphate pyrophosphokinase
b1215	<i>kdsA</i>	COG2877M	2-dehydro-3-deoxyphosphooctonate aldolase
b1263	<i>trpD</i>	COG0512EH, COG0547E	bifunctional indole-3-glycerol-phosphate synthase/anthranilate phosphoribosyltransferase
b1466	<i>narW</i>	COG2180C	nitrate reductase 2 (NRZ), delta subunit (assembly subunit)
b1515	<i>lsrD</i>	COG1172G	AI2 transporter
b1563	<i>relE</i>	COG2026JD	Qin prophage; toxin of the RelE-RelB toxin-antitoxin system
b1704	<i>aroH</i>	COG0722E	3-deoxy-D-arabino-heptulosonate-7-phosphate synthase, tryptophan repressible
b1814	<i>sdaA</i>	COG1760E	L-serine deaminase I
b1824	<i>yobF</i>	-	hypothetical protein b1824
b1831	<i>proQ</i>	COG3109T	putative solute/DNA competence effector
b1913	<i>uvrC</i>	COG0322L	excinuclease ABC subunit C
b2144	<i>sanA</i>	COG2949S	hypothetical protein b2144
b2168	<i>fruK</i>	COG1105G	1-phosphofructokinase
b2175	<i>spr</i>	COG0791M	predicted peptidase, outer membrane lipoprotein
b2297	<i>pta</i>	COG0280C, COG0857R	phosphate acetyltransferase
b2318	<i>truA</i>	COG0101J	tRNA pseudouridine synthase A
b2393	<i>nupC</i>	-	nucleoside (except guanosine) transporter
b2432	<i>yfeY</i>	-	hypothetical protein b2432
b2571	<i>rseB</i>	COG3026T	periplasmic negative regulator of sigmaE
b2593	<i>yfiH</i>	COG1496S	hypothetical protein b2593
b2697	<i>alaS</i>	COG0013J	alanyl-tRNA synthetase
b2780	<i>pyrG</i>	COG0504F	CTP synthetase
b2813	<i>mltA</i>	COG2821M	membrane-bound lytic murein transglycosylase A
b3037	<i>ygiB</i>	COG5463S	conserved outer membrane protein
b3065	<i>rpsU</i>	COG0828J	30S ribosomal protein S21
b3190	<i>yrbA</i>	COG5007K	predicted DNA-binding transcriptional regulator
b3250	<i>mreC</i>	COG1792M	cell wall structural complex MreBCD transmembrane component MreC
b3283	<i>yrdD</i>	COG0551L	predicted DNA topoisomerase
b3347	<i>fkpA</i>	COG0545O	FKBP-type peptidyl-prolyl cis-trans isomerase (rotamase)
b3511	<i>hdeD</i>	COG3247S	acid-resistance membrane protein
b3866	<i>yihI</i>	-	hypothetical protein b3866
b4041	<i>plsB</i>	COG2937I	glycerol-3-phosphate acyltransferase
b4129	<i>lysU</i>	COG1190J	lysine tRNA synthetase, inducible
b4170	<i>mutL</i>	COG0323L	DNA mismatch repair protein
b4428,b1420	<i>hokB,mokB</i>	-,-	toxic polypeptide, small ,regulatory peptide

T8-T9

b0009	<i>mog</i>	-	molybdenum cofactor biosynthesis protein
b0146	<i>sfsA</i>	-	sugar fermentation stimulation protein A
b0198	<i>metI</i>	COG2011P	DL-methionine transporter subunit
b0233	<i>yafO</i>	-	predicted toxin of the YafO-YafN toxin-antitoxin system
b0260	<i>mmuP</i>	COG0833E	CP4-6 prophage; predicted S-methylmethionine transporter
b0261	<i>mmuM</i>	COG2040E	homocysteine methyltransferase
b0319	<i>yahE</i>	-	hypothetical protein b0319
b0324	<i>yahJ</i>	COG0402FR	deaminase
b0333	<i>prpC</i>	COG0372C	2-methylcitrate synthase
b0357	<i>frmR</i>	COG1937S	regulator protein that represses frmRAB operon

b2878	<i>ygfK</i>	COG0493ER	predicted oxidoreductase, Fe-S subunit
b2937	<i>speB</i>	COG0010E	agmatinase
b2979	<i>glcD</i>	COG0277C	glycolate oxidase subunit, FAD-linked
b3075	<i>ebgR</i>	COG1609K	DNA-binding transcriptional repressor
b3136	<i>agaS</i>	COG2222M	tagatose-6-phosphate ketose/aldose isomerase
b3183	<i>obgE</i>	COG0536R	GTPase involved in cell partitioning and DNA repair
b3332	<i>gspK</i>	COG3156U	general secretory pathway component, cryptic
b3361	<i>fic</i>	COG2184D	stationary-phase protein, cell division
b3362	<i>yhfG</i>	-	hypothetical protein b3362
b3471	<i>yhhQ</i>	COG1738S	conserved inner membrane protein
b3521	<i>yhjC</i>	COG0583K	predicted DNA-binding transcriptional regulator
b3581	<i>sgbH</i>	COG0269G	3-keto-L-gulonate 6-phosphate decarboxylase
b3672	<i>ivbL</i>	-	ilvB operon leader peptide
b3882	<i>yihU</i>	COG2084I	predicted oxidoreductase with NAD(P)-binding Rossmann-fold domain
b3905	<i>rhaS</i>	-	DNA-binding transcriptional activator, L-rhamnose-binding
b4024	<i>lysC</i>	COG0527E	aspartate kinase III
b4058	<i>uvrA</i>	COG0178L	excinuclease ABC subunit A
b4075	<i>nrjF</i>	COG3088O	heme lyase (NrjEFG) for insertion of heme into c552, subunit NrjF
b4080	<i>yjcP</i>	COG1538MU	predicted outer membrane factor of efflux pump
b4106	<i>phnC</i>	COG3638P	phosphonate/organophosphate ester transporter subunit
b4205	<i>ytfA</i>	-	predicted transcriptional regulator
b4301	<i>sgcE</i>	COG0036G	ribose-phosphate 3-epimerase
b4328	<i>iadA</i>	COG0044F, COG1820G	isoaspartyl dipeptidase
b4345	<i>mcrC</i>	COG4268V	5-methylcytosine-specific restriction enzyme McrBC, subunit McrC
b4484	<i>cpxP</i>	COG3678UNT P	periplasmic protein combats stress
b4494,b1720	<i>arpB</i>	-	-
b4500,b2227	<i>yfaS</i>	-	-
b4581,b0625	<i>ybeM</i>	-	-

T9-T10

b0040	<i>caiT</i>	COG1292M	L-carnitine/gamma-butyrobetaine antiporter
b0464	<i>acrR</i>	COG1309K	DNA-binding transcriptional repressor
b0565	<i>ompT</i>	COG4571M	DLP12 prophage; outer membrane protease VII (outer membrane protein 3b)
b0590	<i>fepD</i>	COG0609P	iron-enterobactin transporter subunit
b0862	<i>artQ</i>	COG4215E	arginine transporter subunit
b1290	<i>sapF</i>	COG4167V	predicted antimicrobial peptide transporter subunit
b2419	<i>yfeK</i>	-	hypothetical protein b2419
b3329	<i>gspH</i>	-	predicted general secretory pathway component, cryptic
b3646	<i>yicG</i>	COG2860S	conserved inner membrane protein

T10-T11

b0077	<i>ilvI</i>	COG0028EH	acetolactate synthase III large subunit
b0096	<i>lpxC</i>	-	UDP-3-O-
b0261	<i>mmuM</i>	COG2040E	homocysteine methyltransferase
b0317	<i>yahC</i>	-	predicted inner membrane protein
b0512	<i>allB</i>	COG0044F	allantoinase
b0551	<i>ybcQ</i>	-	DLP12 prophage; predicted antitermination protein
b0590	<i>fepD</i>	COG0609P	iron-enterobactin transporter subunit
b0690	<i>ybfG</i>	-	hypothetical protein b0690
b0981	<i>yecC</i>	COG0489D, COG3206M	cryptic autophosphorylating protein tyrosine kinase Etk
b1114	<i>mfd</i>	COG1197LK	transcription-repair coupling factor
b1477	<i>yddM</i>	COG3093R	predicted DNA-binding transcriptional regulator
b1563	<i>relE</i>	COG2026JD	Qin prophage; toxin of the RelE-RelB toxin-antitoxin system
b1564	<i>relB</i>	COG3077L	Qin prophage; bifunctional antitoxin of the RelE-RelB toxin-antitoxin system/ transcriptional repressor
b1648	<i>ydhL</i>	COG3313R	hypothetical protein b1648
b1790	<i>yeaM</i>	-	predicted DNA-binding transcriptional regulator
b1814	<i>sdaA</i>	COG1760E	L-serine deaminase I
b1873	<i>torY</i>	COG3005C	TMAO reductase III (TorYZ), cytochrome c-type subunit
b1874	<i>cutC</i>	COG3142P	copper homeostasis protein
b2049	<i>cpsB</i>	COG0662G, COG0836M	mannose-1-phosphate guanylttransferase
b2194	<i>ccmH</i>	COG3088O, COG4235O	heme lyase, CcmH subunit
b2247	<i>yfaW</i>	COG4948MR	predicted enolase
b2529	<i>iscU</i>	COG0822C	scaffold protein
b2620	<i>smpB</i>	COG0691O	SsrA-binding protein
b2797	<i>sdaB</i>	COG1760E	L-serine deaminase II
b3049	<i>glgS</i>	-	glycogen synthesis protein GlgS
b3329	<i>gspH</i>	-	predicted general secretory pathway component, cryptic

b3350	<i>kefB</i>	COG0475P, COG1226P	glutathione-regulated potassium-efflux system protein
b3905	<i>rhaS</i>	-	DNA-binding transcriptional activator, L-rhamnose-binding
b4084	<i>alsK</i>	COG1940KG	D-allose kinase
b4205	<i>ytfA</i>	-	predicted transcriptional regulator
b4484	<i>cpxP</i>	COG3678UNT P	periplasmic protein combats stress

付表2 :各タイムポイント間で発現減少と判定された遺伝子

b no.	gene	cogID	function.
T1-T2			
b0080	<i>fruR</i>	COG1609K	DNA-binding transcriptional dual regulator
b0173	<i>dxr</i>	-	1-deoxy-D-xylulose 5-phosphate reductoisomerase
b0176	<i>yaeL</i>	COG0750M	zinc metalloproteinase
b0184	<i>dnaE</i>	COG0587L	DNA polymerase III subunit alpha
b0207	<i>dkgB</i>	COG0656R	2,5-diketo-D-gluconate reductase B
b0211	<i>mltD</i>	COG0741M	predicted membrane-bound lytic murein transglycosylase D
b0239	<i>frsA</i>	-	fermentation/respiration switch protein
b0410	<i>yqjD</i>	-	hypothetical protein b0410
b0423	<i>thiI</i>	COG0301H, COG0607P	thiamine biosynthesis protein ThiI
b0460	<i>hha</i>	-	modulator of gene expression, with H-NS
b0461	<i>ybaJ</i>	-	hypothetical protein b0461
b0469	<i>apt</i>	COG0503F	adenine phosphoribosyltransferase
b0524	<i>lpxH</i>	COG2908S	UDP-2,3-diacetylglucosamine hydrolase
b0591	<i>ybdA</i>	-	predicted transporter
b0594	<i>entE</i>	COG1021Q	2,3-dihydroxybenzoate-AMP ligase component of enterobactin synthase multienzyme complex
b0596	<i>entA</i>	COG1028IQR	2,3-dihydroxybenzoate-2,3-dehydrogenase
b0635	<i>mrdA</i>	COG0768M	transpeptidase involved in peptidoglycan synthesis (penicillin-binding protein 2)
b0638	<i>cobC</i>	COG0406G	predicted alpha-ribazole-5'-P phosphatase
b0639	<i>nadD</i>	COG1057H	nicotinic acid mononucleotide adenyltransferase
b0640	<i>holA</i>	COG1466L	DNA polymerase III subunit delta
b0674	<i>asnB</i>	COG0367E	asparagine synthetase B
b0733	<i>cydA</i>	COG1271C	cytochrome d terminal oxidase, subunit I
b0804	<i>ybiX</i>	COG3128S	putative hydroxylase
b0815	<i>ybiP</i>	COG2194R	predicted hydrolase, inner membrane
b0904	<i>focA</i>	COG2116P	formate transporter
b0910	<i>cmk</i>	COG0283F	cytidylate kinase
b0951	<i>pqiB</i>	COG3008R	paraquat-inducible protein B
b0990	<i>cspG</i>	-	DNA-binding transcriptional regulator
b1049	<i>mdoH</i>	COG2943M	glucosyltransferase MdoH
b1134	<i>ymfB</i>	COG0494LR	bifunctional thiamin pyrimidine pyrophosphate hydrolase/ thiamin pyrophosphate hydrolase
b1211	<i>prfA</i>	COG0216J	peptide chain release factor 1
b1214	<i>ycaA</i>	COG2912S	predicted transcriptional regulator
b1216	<i>chaA</i>	COG0387P	calcium/sodium:proton antiporter
b1243	<i>oppA</i>	COG4166E	oligopeptide transporter subunit
b1245	<i>oppC</i>	COG1173EP	oligopeptide transporter subunit
b1246	<i>oppD</i>	COG0444EP	oligopeptide transporter ATP-binding component
b1266	<i>yciV</i>	COG0613R	hypothetical protein b1266
b1269	<i>rluB</i>	COG1187J	23S rRNA pseudouridylylase synthase
b1413	<i>hrpA</i>	COG1643L	ATP-dependent helicase
b1451	<i>yncD</i>	COG1629P	predicted iron outer membrane transporter
b1557	<i>cspB</i>	-	Qin prophage; cold shock protein
b1558	<i>cspF</i>	-	Qin prophage; cold shock protein
b1600	<i>mdtJ</i>	COG2076P	multidrug efflux system transporter
b1627	<i>rsxA</i>	COG4657C	Na(+)-translocating NADH-quinone reductase subunit E
b1630	<i>rsxD</i>	COG4658C	electron transport complex protein RnfD
b1631	<i>rsxG</i>	COG4659C	electron transport complex protein RnfG
b1632	<i>rsxE</i>	COG4660C	NADH-ubiquinone oxidoreductase
b1663	<i>mdtK</i>	COG0534V	multidrug efflux protein NorM
b1713	<i>pheT</i>	COG0072J, COG0073R	phenylalanyl-tRNA synthetase beta subunit
b1714	<i>pheS</i>	COG0016J	phenylalanyl-tRNA synthetase alpha subunit
b1804	<i>rnd</i>	COG0349J	ribonuclease D
b1805	<i>fadD</i>	COG03181Q	acyl-CoA synthase
b1830	<i>prc</i>	COG0793M	carboxy-terminal protease for penicillin-binding protein 3
b2155	<i>cirA</i>	COG4771P	ferric iron-catecholide outer membrane transporter
b2183	<i>rsuA</i>	COG1187J	16S rRNA pseudouridylylase 516 synthase
b2211	<i>yojI</i>	COG4615QP	fused predicted multidrug transport subunits of ABC superfamily: membrane component/ATP-binding
b2214	<i>yojL</i>	-	predicted thiamine biosynthesis lipoprotein
b2314	<i>dedD</i>	COG3147S	hypothetical protein b2314
b2511	<i>der</i>	COG1160R	GTP-binding protein EngA
b2526	<i>hscA</i>	COG0443O	chaperone protein HscA
b2527	<i>hscB</i>	COG1076O	co-chaperone HscB
b2910	<i>zapA</i>	COG3027S	protein that localizes to the cytokinetic ring
b2946	<i>yggJ</i>	COG1385S	hypothetical protein b2946
b3006	<i>exbB</i>	COG0811U	membrane spanning protein in TonB-ExbB-ExbD complex
b3034	<i>nudF</i>	COG0494LR	ADP-ribose pyrophosphatase
b3163	<i>nlpI</i>	COG4785R	hypothetical protein b3163
b3166	<i>truB</i>	COG0130J	tRNA pseudouridine synthase B
b3170	<i>yhbC</i>	COG0779S	hypothetical protein b3170
b3199	<i>yrbK</i>	COG3117S	hypothetical protein b3199
b3203	<i>yhbH</i>	COG1544J	predicted ribosome-associated, sigma 54 modulation protein

b3248	<i>yhdE</i>	COG0424D	Maf-like protein
b3399	<i>yrfG</i>	COG1011R	predicted hydrolase
b3493	<i>pitA</i>	COG0306P	phosphate transporter, low-affinity
b3556	<i>cspA</i>	-	major cold shock protein
b3628	<i>rfaB</i>	COG0438M	UDP-D-galactose:(glucosyl)lipopolysaccharide-1,6-D-galactosyltransferase
b3631	<i>rfaG</i>	COG0438M	glucosyltransferase I
b3633	<i>kdtA</i>	COG1519M	3-deoxy-D-manno-octulosonic-acid transferase
b3702	<i>dnaA</i>	COG0593L	chromosomal replication initiation protein
b3704	<i>rnpA</i>	COG0594J	ribonuclease P
b3790	<i>rffC</i>	-	TDP-fucosamine acetyltransferase
b3792	<i>wzxE</i>	-	O-antigen translocase
b3859	<i>yihE</i>	COG2334R	predicted kinase
b3871	<i>bipA</i>	COG1217T	GTP-binding protein
b4018	<i>iclR</i>	COG1414K	DNA-binding transcriptional repressor
b4049	<i>yjbN</i>	COG0042J	tRNA-dihydrouridine synthase A
b4077	<i>glpP</i>	COG1301C	glutamate/aspartate:proton symporter
b4125	<i>dcuS</i>	COG3290T	sensory histidine kinase in two-component regulatory system with DcuR, regulator of anaerobic fumarate respiration
b4170	<i>mutL</i>	COG0323L	DNA mismatch repair protein
b4208	<i>cycA</i>	COG1113E	D-alanine/D-serine/glycine transporter
b4293	<i>fecI</i>	COG1595K	KpLE2 phage-like element; RNA polymerase, sigma 19 factor
b4469	<i>ygiQ</i>	COG1032C	hypothetical protein b4469
b4484	<i>cpxP</i>	COG3678UNT	periplasmic protein combats stress
		P	

T2-T3

b0115	<i>aceF</i>	COG0508C	dihydroliipoamide acetyltransferase
b0116	<i>lpd</i>	COG1249C	dihydroliipoamide dehydrogenase
b0121	<i>speE</i>	COG0421E	spermidine synthase
b0125	<i>hpt</i>	COG0634F	hypoxanthine-guanine phosphoribosyltransferase
b0282	<i>yagP</i>	COG0583K	predicted transcriptional regulator
b0614	<i>citX</i>	COG3697HI	2'-(5'-triphosphoribosyl)-3'-dephospho-CoA:apo-citrate lyase
b0628	<i>lipA</i>	COG0320H	lipoyl synthase
b0804	<i>ybiX</i>	COG3128S	putative hydroxylase
b0849	<i>grxA</i>	COG0695O	glutaredoxin 1, redox coenzyme for ribonucleotide reductase (RNR1a)
b0888	<i>trxB</i>	COG0492O	thioredoxin reductase, FAD/NAD(P)-binding
b0911	<i>rpsA</i>	COG0539J	30S ribosomal protein S1
b0990	<i>cspG</i>	-	DNA-binding transcriptional regulator
b0992	<i>yccM</i>	COG0348C	predicted 4Fe-4S membrane protein
b1088	<i>yceD</i>	COG1399R	hypothetical protein b1088
b1286	<i>rmb</i>	-	exoribonuclease II
b1378	<i>ydbK</i>	COG0674C, COG1013C, COG1014C	fused predicted pyruvate-flavodoxin oxidoreductase: conserved protein/conserved protein/FeS binding protein
b1557	<i>cspB</i>	-	Qin prophage; cold shock protein
b2297	<i>pta</i>	COG0280C, COG0857R	phosphate acetyltransferase
b2518	<i>ndk</i>	-	nucleoside diphosphate kinase
b2526	<i>hscA</i>	COG0443O	chaperone protein HscA
b2810	<i>csdA</i>	COG0520E	cysteine sulfinate desulfinate
b2947	<i>gshB</i>	COG0189HJ	glutathione synthetase
b3247	<i>rng</i>	COG1530J	ribonuclease G
b3250	<i>mreC</i>	COG1792M	cell wall structural complex MreBCD transmembrane component MreC
b3294	<i>rplQ</i>	COG0203J	50S ribosomal protein L17
b3527	<i>yhjJ</i>	COG0612R	predicted zinc-dependent peptidase
b3606	<i>yibK</i>	COG0219J	predicted rRNA methylase
b3699	<i>gyrB</i>	COG0187L	DNA gyrase subunit B
b3748	<i>rbsD</i>	COG1869G	predicted cytoplasmic sugar-binding protein
b3782	<i>rhoL</i>	-	rho operon leader peptide
b3843	<i>ubiD</i>	-	3-octaprenyl-4-hydroxybenzoate decarboxylase
b3965	<i>trmA</i>	COG2265J	tRNA (uracil-5-)-methyltransferase
b3988	<i>rpoC</i>	COG0086K	DNA-directed RNA polymerase subunit beta'
b4332	<i>yjiJ</i>	-	predicted inner membrane protein
b4371	<i>rsmC</i>	COG2813J	16S ribosomal RNA m2G1207 methyltransferase
b4490,b1016	<i>ycdN</i>	-	-

T3-T4

b0031	<i>dapB</i>	COG0289E	dihydrodipicolinate reductase
b0053	<i>surA</i>	COG0760O	peptidyl-prolyl cis-trans isomerase (PPIase)
b0090	<i>murG</i>	COG0707M	N-acetylglucosaminyl transferase
b0092	<i>ddlB</i>	COG1181M	D-alanylalanine synthetase
b0094	<i>ftsA</i>	COG0849D	cell division protein

b0361,b1402,b1578,b1996,b2860,b3045,b4273	<i>insD-1,insD-2,insD-7,insD-3,insD-4,insD-5,insD-6</i>	COG2801L,COG2801L,COG2801L,COG2801L,COG2801L,COG2801L	IS2 insertion element transposase InsAB', IS2 insertion element transposase InsAB', QIn prophage; IS2 insertion element transposase InsAB', C-ter fragment (pseudogene), CP4-44 prophage; IS2 insertion element transposase InsAB', IS2 insertion element transposase InsAB', IS2 insertion element transposase InsAB', KpLE2 phage-like element; IS2 insertion element transposase InsAB'
b0658	<i>ybeX</i>	COG4535P	predicted ion transport
b0756	<i>galM</i>	COG2017G	galactose-1-epimerase (mutarotase)
b1667	<i>ydhR</i>	-	hypothetical protein b1667
b1780	<i>yeaD</i>	COG0676G	hypothetical protein b1780
b1913	<i>uvrC</i>	COG0322L	excinuclease ABC subunit C
b2094	<i>gata</i>	COG1762GT	galactitol-specific enzyme IIA component of PTS
b2096	<i>gatY</i>	COG0191G	D-tagatose 1,6-bisphosphate aldolase 2, catalytic subunit
b2323	<i>fabB</i>	COG0304IQ	3-oxoacyl-(acyl carrier protein) synthase
b2326	<i>yfcM</i>	-	hypothetical protein b2326
b2327	<i>yfcA</i>	COG0730R	conserved inner membrane protein
b2366	<i>dsdA</i>	COG3048E	D-serine dehydratase
b2504	<i>yfgG</i>	-	hypothetical protein b2504
b2600	<i>tyrA</i>	COG0287E, COG1605E, COG0208F	fused chorismate mutase T/prephenate dehydrogenase
b2676	<i>nrdF</i>	COG0208F	ribonucleotide-diphosphate reductase beta subunit
b2683	<i>ygaH</i>	-	predicted inner membrane protein
b2783	<i>chpR</i>	COG2336T	antitoxin of the ChpA-ChpR toxin-antitoxin system
b2830	<i>nudH</i>	COG0494LR	dinucleoside polyphosphate hydrolase
b2904	<i>gcvH</i>	COG0509E	glycine cleavage system protein H
b2991	<i>hybF</i>	-	protein involved with the maturation of hydrogenases 1 and 2
b3005	<i>exbD</i>	COG0848U	membrane spanning protein in TonB-ExbB-ExbD complex
b3084	<i>ygiO</i>	COG2813J	predicted methyltransferase small domain
b3097	<i>yqjC</i>	-	hypothetical protein b3097
b3114	<i>tdcE</i>	COG1882C	pyruvate formate-lyase 4/2-ketobutyrate formate-lyase
b3203	<i>yhbH</i>	COG1544J	predicted ribosome-associated, sigma 54 modulation protein
b3204	<i>ptsN</i>	COG1762GT	sugar-specific enzyme IIA component of PTS
b3250	<i>mreC</i>	COG1792M	cell wall structural complex MreBCD transmembrane component MreC
b3388	<i>damX</i>	COG3266S	hypothetical protein b3388
b3426	<i>glpD</i>	COG0578C	sn-glycerol-3-phosphate dehydrogenase, aerobic, FAD/NAD(P)-binding
b3428	<i>glgP</i>	COG0058G	glycogen phosphorylase
b3430	<i>glgC</i>	COG0448G	glucose-1-phosphate adenylyltransferase
b3506	<i>slp</i>	-	outer membrane lipoprotein
b3607	<i>cysE</i>	COG1045E	serine acetyltransferase
b3608	<i>gpsA</i>	COG0240C	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase
b3702	<i>dnaA</i>	COG0593L	chromosomal replication initiation protein
b3741	<i>gidA</i>	COG0445D	glucose-inhibited division protein A
b3813	<i>uvrD</i>	COG0210L	DNA-dependent ATPase I and helicase II
b3996	<i>nudC</i>	COG2816L	NADH pyrophosphatase
b4025	<i>pgi</i>	COG0166G	glucose-6-phosphate isomerase
b4049	<i>yjbN</i>	COG0042J	tRNA-dihydrouridine synthase A
b4174	<i>hflK</i>	COG0330O	modulator for HflB protease specific for phage lambda cII repressor
b4175	<i>hflC</i>	COG0330O	modulator for HflB protease specific for phage lambda cII repressor
b4178	<i>yjeB</i>	COG1959K	predicted DNA-binding transcriptional regulator
b4262	<i>yjgQ</i>	-	conserved inner membrane protein
b4290	<i>fecB</i>	COG4594P	KpLE2 phage-like element; iron-dicitrate transporter subunit
b4291	<i>fecA</i>	COG4772P	KpLE2 phage-like element; ferric citrate outer membrane transporter
b4352	<i>yjiA</i>	COG0523R	predicted GTPase
b4367	<i>fhuF</i>	COG4114R	ferric iron reductase involved in ferric hydroximate transport
b4470_1	<i>yhaM</i>	-	-
b4493,b1417	<i>gapC</i>	-	-

T4-T5

b0104	<i>guaC</i>	COG0516F	guanosine 5'-monophosphate oxidoreductase
b0170	<i>tsf</i>	-	elongation factor Ts
b0171	<i>pyrH</i>	COG0528F	uridylylate kinase
b0355	<i>frmB</i>	COG0627R	predicted esterase
b0388	<i>aroL</i>	COG0703E	shikimate kinase II
b0586	<i>entF</i>	COG1020Q, COG3319Q	enterobactin synthase multienzyme complex component, ATP-dependent
b0595	<i>entB</i>	COG1535Q, COG3433Q	isochorismatase
b0791	<i>ybhQ</i>	-	predicted inner membrane protein
b0887	<i>cydD</i>	COG4988CO	fused cysteine transporter subunits of ABC superfamily: membrane component/ATP-binding component
b0910	<i>cmk</i>	COG0283F	cytidylate kinase
b1090	<i>plsX</i>	COG0416I	fatty acid/phospholipid synthesis protein
b1269	<i>rluB</i>	COG1187J	23S rRNA pseudouridylylate synthase
b1299	<i>puuR</i>	COG1396K	DNA-binding transcriptional repressor
b1406	<i>ydbC</i>	COG0667C	predicted oxidoreductase, NAD(P)-binding
b1537	<i>ydeJ</i>	COG1546R	competence damage-inducible protein A

b1618	<i>uidR</i>	COG1309K	DNA-binding transcriptional repressor
b1676	<i>pykF</i>	-	pyruvate kinase
b1714	<i>pheS</i>	COG0016J	phenylalanyl-tRNA synthetase alpha subunit
b1719	<i>thrS</i>	COG0441J	threonyl-tRNA synthetase
b1800	<i>yeaU</i>	COG0473CE	predicted dehydrogenase
b1871	<i>yecP</i>	-	predicted S-adenosyl-L-methionine-dependent methyltransferase
b1876	<i>argS</i>	COG0018J	arginyl-tRNA synthetase
b1902	<i>yecI</i>	COG1528P	predicted ferritin-like protein
b2156	<i>lysP</i>	COG0833E	lysine transporter
b2415	<i>ptsH</i>	COG1925G	phosphohistidinoprotein-hexose phosphotransferase component of PTS system (Hpr)
b2417	<i>crr</i>	-	glucose-specific PTS system enzyme IIA component
b2499	<i>purM</i>	COG0150F	phosphoribosylaminoimidazole synthetase
b2511	<i>der</i>	COG1160R	GTP-binding protein EngA
b2606	<i>rplS</i>	-	50S ribosomal protein L19
b2684	<i>mprA</i>	COG1846K	DNA-binding transcriptional repressor of microcin B17 synthesis and multidrug efflux
b2780	<i>pyrG</i>	COG0504F	CTP synthetase
b3261	<i>fis</i>	COG2901KL	DNA-binding protein Fis
b3390	<i>aroK</i>	COG0703E	shikimate kinase I
b3493	<i>pitA</i>	COG0306P	phosphate transporter, low-affinity
b3604	<i>lldR</i>	COG2186K	DNA-binding transcriptional repressor
b3606	<i>yibK</i>	COG0219J	predicted rRNA methylase
b3703	<i>rpmH</i>	COG0230J	50S ribosomal protein L34
b3740	<i>gidB</i>	COG0357M	glucose-inhibited division protein B
b3894	<i>fdoG</i>	COG0243C	formate dehydrogenase-O, large subunit
b3936	<i>rpmE</i>	COG0254J	50S ribosomal subunit protein L31
b4170	<i>mutL</i>	COG0323L	DNA mismatch repair protein
b4177	<i>purA</i>	-	adenylosuccinate synthetase
b4466	<i>yghJ</i>	-	predicted inner membrane lipoprotein

T5-T6

b0064	<i>araC</i>	-	DNA-binding transcriptional dual regulator
b0076	<i>leuO</i>	COG0583K	leucine transcriptional activator
b0084	<i>fisI</i>	COG0768M	transpeptidase involved in septal peptidoglycan synthesis (penicillin-binding protein 3)
b0145	<i>dksA</i>	COG1734T	DNA-binding transcriptional regulator of rRNA transcription, DnaK suppressor protein
b0180	<i>fabZ</i>	COG0764I	(3R)-hydroxymyristoyl ACP dehydratase
b0307	<i>ykgF</i>	COG1139C	predicted amino acid dehydrogenase with NAD(P)-binding domain and ferridoxin-like domain
b0344	<i>lacZ</i>	COG3250G	beta-D-galactosidase
b0380	<i>yaiZ</i>	-	predicted inner membrane protein
b0409	<i>secF</i>	COG0341U	protein export protein SecF
b0469	<i>apt</i>	COG0503F	adenine phosphoribosyltransferase
b0679	<i>nagE</i>	COG1263G	fused N-acetyl glucosamine specific PTS enzyme: IIC, IIB, and IIA components
b0820	<i>ybiT</i>	COG0488R	fused predicted transporter subunits of ABC superfamily: ATP-binding components
b0846	<i>ybjK</i>	COG3226S	predicted DNA-binding transcriptional regulator
b0880	<i>cspD</i>	-	cold shock protein homolog
b0908	<i>aroA</i>	COG0128E	3-phosphoshikimate 1-carboxyvinyltransferase
b0927	<i>ycbL</i>	COG0491R	predicted metal-binding enzyme
b0954	<i>fabA</i>	COG0764I	3-hydroxydecanoyl-ACP dehydratase
b1036	<i>ycdZ</i>	-	predicted inner membrane protein
b1090	<i>plsX</i>	COG0416I	fatty acid/phospholipid synthesis protein
b1269	<i>rluB</i>	COG1187J	23S rRNA pseudouridylylase synthase
b1329	<i>mppA</i>	COG4166E	murein tripeptide (L-ala-gamma-D-glutamyl-meso-DAP) transporter subunit
b1406	<i>ydbC</i>	COG0667C	predicted oxidoreductase, NAD(P)-binding
b1424	<i>mdoD</i>	-	glucan biosynthesis protein, periplasmic
b1511	<i>lsrK</i>	COG1070G	autoinducer-2 (AI-2) kinase
b1609	<i>rstB</i>	COG0642T	sensory histidine kinase in two-component regulatory system with RstA
b1612	<i>fumA</i>	COG1838C	fumarate hydratase (fumarase A), aerobic Class I
b1618	<i>uidR</i>	COG1309K	DNA-binding transcriptional repressor
b1711	<i>btuC</i>	COG4139H	vitamin B12-transporter permease
b1713	<i>pheT</i>	COG0072J,	phenylalanyl-tRNA synthetase beta subunit
		COG0073R	
b1717	<i>rpmI</i>	COG0291J	50S ribosomal protein L35
b1768	<i>pncA</i>	COG1335Q	nicotinamidase/pyrazinamidase
b1819	<i>manZ</i>	-	mannose-specific enzyme IID component of PTS
b1864	<i>yebC</i>	-	hypothetical protein b1864
b1865	<i>nudB</i>	COG0494LR	dATP pyrophosphohydrolase
b1867	<i>yecD</i>	COG1335Q	predicted hydrolase
b2021	<i>hisC</i>	COG0079E	histidinol-phosphate aminotransferase
b2033	<i>wbbJ</i>	COG0110R	predicted acyl transferase
b2037	<i>rfbX</i>	COG2244R	predicted polysoprenol-linked O-antigen transporter
b2041	<i>rfbB</i>	COG1088M	dTDP-glucose 4,6 dehydratase, NAD(P)-binding
b2144	<i>sanA</i>	COG2949S	hypothetical protein b2144
b2159	<i>nfo</i>	COG0648L	endonuclease IV
b2288	<i>nuoA</i>	COG0838C	NADH dehydrogenase subunit A
b2412	<i>zipA</i>	COG3115D	cell division protein ZipA

b2463	<i>maeB</i>	COG0280C, COG0281C	malic enzyme
b2515	<i>ispG</i>	-	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase
b2600	<i>tyrA</i>	COG0287E, COG1605E	fused chorismate mutase T/prephenate dehydrogenase
b2606	<i>rplS</i>	-	50S ribosomal protein L19
b2607	<i>trmD</i>	-	tRNA (guanine-N(1)-)-methyltransferase
b2704	<i>srlB</i>	-	glucitol/sorbitol-specific enzyme IIA component of PTS
b2705	<i>srlD</i>	COG1028IQR	3-ketoacyl-(acyl-carrier-protein) reductase
b2749	<i>ygbE</i>	-	conserved inner membrane protein
b3005	<i>exbD</i>	COG0848U	membrane spanning protein in TonB-ExbB-ExbD complex
b3066	<i>dnaG</i>	COG0358L	DNA primase
b3084	<i>ygiO</i>	COG2813J	predicted methyltransferase small domain
b3170	<i>yhbC</i>	COG0779S	hypothetical protein b3170
b3178	<i>ftsH</i>	COG0465O	protease, ATP-dependent zinc-metallo
b3179	<i>rrmJ</i>	COG0293J	23S rRNA methyltransferase
b3283	<i>yrdD</i>	COG0551L	predicted DNA topoisomerase
b3294	<i>rplQ</i>	COG0203J	50S ribosomal protein L17
b3297	<i>rpsK</i>	COG0100J	30S ribosomal protein S11
b3301	<i>rplO</i>	COG0200J	50S ribosomal protein L15
b3303	<i>rpsE</i>	COG0098J	30S ribosomal protein S5
b3305	<i>rplF</i>	COG0097J	50S ribosomal protein L6
b3307	<i>rpsN</i>	COG0199J	30S ribosomal protein S14
b3312	<i>rpmC</i>	-	50S ribosomal protein L29
b3315	<i>rplV</i>	-	50S ribosomal protein L22
b3316	<i>rpsS</i>	COG0185J	30S ribosomal protein S19
b3319	<i>rplD</i>	-	50S ribosomal protein L4
b3390	<i>aroK</i>	COG0703E	shikimate kinase I
b3415	<i>gntT</i>	COG2610GE	gluconate transporter, high-affinity GNT I system
b3417	<i>malP</i>	COG0058G	maltodextrin phosphorylase
b3748	<i>rhsD</i>	COG1869G	predicted cytoplasmic sugar-binding protein
b3754	<i>hsrA</i>	-	predicted multidrug or homocysteine efflux system
b3776	<i>yifO</i>	COG3692S	hypothetical protein b3776
b3871	<i>bipA</i>	COG1217T	GTP-binding protein
b3981	<i>secE</i>	COG0690U	translocase
b4059	<i>ssb</i>	COG0629L	single-strand DNA-binding protein
b4117	<i>adiA</i>	COG1982E	biodegradative arginine decarboxylase
b4170	<i>mutL</i>	COG0323L	DNA mismatch repair protein
b4234	<i>yjgA</i>	-	hypothetical protein b4234
b4322	<i>uxuA</i>	COG1312G	mannonate dehydratase
b4466	<i>yghJ</i>	-	predicted inner membrane lipoprotein
b4499,b2115	<i>molR</i>	-	-

T6-T7

b0023	<i>rpsT</i>	COG0268J	30S ribosomal protein S20
b0026	<i>ileS</i>	COG0060J	isoleucyl-tRNA synthetase
b0028	<i>fkpB</i>	COG1047O	FKBP-type peptidyl-prolyl cis-trans isomerase (rotamase)
b0076	<i>leuO</i>	COG0583K	leucine transcriptional activator
b0096	<i>lpxC</i>	-	UDP-3-O-
b0103	<i>coaE</i>	COG0237H	dephospho-CoA kinase
b0115	<i>aceF</i>	COG0508C	dihydroipoamide acetyltransferase
b0159	<i>null</i>	COG0775F	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase
b0161	<i>degP</i>	COG0265O	serine endoprotease (protease Do), membrane-associated
b0169	<i>rpsB</i>	-	30S ribosomal protein S2
b0170	<i>tsf</i>	-	elongation factor Ts
b0171	<i>pyrH</i>	COG0528F	uridylate kinase
b0172	<i>fir</i>	-	ribosome releasing factor
b0176	<i>yaeL</i>	COG0750M	zinc metalloproteinase
b0177	<i>yaeT</i>	COG4775M	hypothetical protein b0177
b0179	<i>lpxD</i>	COG1044M	UDP-3-O-
b0180	<i>fabZ</i>	COG0764I	(3R)-hydroxymyristoyl ACP dehydratase
b0211	<i>mltD</i>	COG0741M	predicted membrane-bound lytic murein transglycosylase D
b0336	<i>codB</i>	COG1457F	cytosine transporter
b0354	<i>yaiL</i>	-	nucleoprotein/polynucleotide-associated enzyme
b0369	<i>hemB</i>	-	delta-aminolevulinic acid dehydratase
b0406	<i>tgt</i>	COG0343J	queuine tRNA-ribosyltransferase
b0407	<i>yajC</i>	COG1862U	preprotein translocase subunit YajC
b0415	<i>ribE</i>	-	riboflavin synthase subunit beta
b0416	<i>nusB</i>	COG0781K	transcription antitermination protein NusB
b0421	<i>ispA</i>	COG0142H	geranyltranstransferase
b0432	<i>cyoA</i>	COG1622C	cytochrome o ubiquinol oxidase subunit II
b0436	<i>tig</i>	COG0544O	trigger factor
b0469	<i>apt</i>	COG0503F	adenine phosphoribosyltransferase
b0472	<i>recR</i>	COG0353L	recombination protein RecR

b0525	<i>ppiB</i>	COG0652O	peptidyl-prolyl cis-trans isomerase B (rotamase B)
b0529	<i>folD</i>	COG0190H	bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate
b0613	<i>citG</i>	COG1767H	triphosphoribosyl-dephospho-CoA transferase
b0632	<i>dacA</i>	COG1686M	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 5)
b0642	<i>leuS</i>	COG0495J	leucyl-tRNA synthetase
b0680	<i>glnS</i>	COG0008J	glutamyl-tRNA synthetase
b0683	<i>fur</i>	COG0735P	ferric uptake regulator
b0687	<i>seqA</i>	-	regulatory protein for replication initiation
b0740	<i>tolB</i>	COG0823U	translocation protein TolB precursor
b0741	<i>pal</i>	COG2885M	peptidoglycan-associated outer membrane lipoprotein
b0742	<i>ybgF</i>	COG1729S	hypothetical protein b0742
b0766	<i>ybhA</i>	COG0561R	predicted hydrolase
b0893	<i>serS</i>	COG0172J	seryl-tRNA synthetase
b0921	<i>smtA</i>	COG0500QR	predicted S-adenosyl-L-methionine-dependent methyltransferase
b1028	<i>ymdE</i>	COG0331I	hypothetical protein b1028
b1034	<i>ycdX</i>	COG1387ER	hypothetical protein b1034
b1035	<i>ycdY</i>	COG3381R	hypothetical protein b1035
b1055	<i>yceA</i>	COG1054R	hypothetical protein b1055
b1088	<i>yceD</i>	COG1399R	hypothetical protein b1088
b1089	<i>rpmF</i>	COG0333J	50S ribosomal protein L32
b1090	<i>plsX</i>	COG0416I	fatty acid/phospholipid synthesis protein
b1091	<i>fabH</i>	COG0332I	3-oxoacyl-(acyl carrier protein) synthase
b1093	<i>fabG</i>	COG1028IQR	3-oxoacyl-
b1094	<i>acpP</i>	COG0236IQ	acyl carrier protein
b1132	<i>hflD</i>	COG2915R	hypothetical protein b1132
b1207	<i>prsA</i>	COG0462FE	ribose-phosphate pyrophosphokinase
b1209	<i>lolB</i>	-	outer membrane lipoprotein LolB precursor
b1214	<i>yehA</i>	COG2912S	predicted transcriptional regulator
b1215	<i>kdsA</i>	COG2877M	2-dehydro-3-deoxyphosphooxonate aldolase
b1232	<i>purU</i>	COG0788F	formyltetrahydrofolate deformylase
b1272	<i>sohB</i>	COG0616OU	predicted inner membrane peptidase
b1288	<i>fabI</i>	COG0623I	enoyl-(acyl carrier protein) reductase
b1452	<i>yncE</i>	COG3391S	hypothetical protein b1452
b1479	<i>sfcA</i>	COG0281C	malate dehydrogenase, (decarboxylating, NAD-requiring) (malic enzyme)
b1537	<i>ydeJ</i>	COG1546R	competence damage-inducible protein A
b1550	<i>gnsB</i>	-	Qin prophage; predicted protein
b1583	<i>ymfB</i>	-	hypothetical protein b1583
b1638	<i>pdxH</i>	COG0259H	pyridoxamine 5'-phosphate oxidase
b1641	<i>slyB</i>	COG3133M	outer membrane lipoprotein
b1653	<i>lhr</i>	COG1201R	predicted ATP-dependent helicase
b1654	<i>ydhD</i>	COG0278O	hypothetical protein b1654
b1702	<i>pps</i>	COG0574G	phosphoenolpyruvate synthase
b1713	<i>pheT</i>	COG0072J, COG0073R	phenylalanyl-tRNA synthetase beta subunit
b1714	<i>pheS</i>	COG0016J	phenylalanyl-tRNA synthetase alpha subunit
b1717	<i>rpmI</i>	COG0291J	50S ribosomal protein L35
b1718	<i>infC</i>	COG0290J	translation initiation factor IF-3
b1782	<i>mipA</i>	-	scaffolding protein for murein synthesizing machinery
b1870	<i>yecO</i>	COG0500QR	predicted methyltransferase
b2029	<i>gnd</i>	COG0362G	6-phosphogluconate dehydrogenase
b2041	<i>rfbB</i>	COG1088M	dTDP-glucose 4,6 dehydratase, NAD(P)-binding
b2113	<i>mrp</i>	COG0489D	antiporter inner membrane protein
b2153	<i>folE</i>	COG0302H	GTP cyclohydrolase I
b2183	<i>rsuA</i>	COG1187J	16S rRNA pseudouridylate 516 synthase
b2185	<i>rplY</i>	COG1825J	50S ribosomal protein L25
b2187	<i>yefL</i>	-	hypothetical protein b2187
b2193	<i>narP</i>	COG2197TK	DNA-binding response regulator in two-component regulatory system with NarQ or NarX
b2231	<i>gyrA</i>	COG0188L	DNA gyrase subunit A
b2300	<i>yfcE</i>	COG0622R	phosphodiesterase
b2318	<i>truA</i>	COG0101J	tRNA pseudouridine synthase A
b2320	<i>pdxB</i>	COG0111HE	erythronate-4-phosphate dehydrogenase
b2328	<i>mepA</i>	COG3770M	penicillin-insensitive murein endopeptidase
b2346	<i>vacJ</i>	-	predicted lipoprotein
b2400	<i>gltX</i>	COG0008J	glutamyl-tRNA synthetase
b2420	<i>yfeS</i>	COG3831S, COG4884S	hypothetical protein b2420
b2478	<i>dapA</i>	COG0329EM	dihydrodipicolinate synthase
b2518	<i>ndk</i>	-	nucleoside diphosphate kinase
b2553	<i>glnB</i>	COG0347E	regulatory protein P-II for glutamine synthetase
b2564	<i>pdxJ</i>	-	pyridoxal phosphate biosynthetic protein
b2568	<i>lepB</i>	COG0681U	leader peptidase (signal peptidase I)
b2608	<i>rimM</i>	COG0806J	16S rRNA-processing protein
b2610	<i>ffh</i>	COG0541U	Signal Recognition Particle (SRP) component with 4.5S RNA (ffs)
b2620	<i>smpB</i>	COG0691O	SsrA-binding protein
b2697	<i>alaS</i>	COG0013J	alanyl-tRNA synthetase

b2746	<i>ispF</i>	-	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
b2780	<i>pyrG</i>	COG0504F	CTP synthetase
b2827	<i>thyA</i>	-	thymidylate synthase
b2908	<i>pepP</i>	COG0006E	proline aminopeptidase P II
b2937	<i>speB</i>	COG0010E	agmatinase
b3005	<i>exbD</i>	COG0848U	membrane spanning protein in TonB-ExbB-ExbD complex
b3006	<i>exbB</i>	COG0811U	membrane spanning protein in TonB-ExbB-ExbD complex
b3034	<i>nudF</i>	COG0494LR	ADP-ribose pyrophosphatase
b3042	<i>yqiC</i>	-	hypothetical protein b3042
b3052	<i>rfaE</i>	COG2870M	fused heptose 7-phosphate kinase/heptose 1-phosphate adenyltransferase
b3065	<i>rpsU</i>	COG0828J	30S ribosomal protein S21
b3066	<i>dnaG</i>	COG0358L	DNA primase
b3164	<i>pnp</i>	COG1185J	polynucleotide phosphorylase/polyadenylase
b3175	<i>secG</i>	COG1314U	protein-export membrane protein
b3180	<i>yhbY</i>	-	predicted RNA-binding protein
b3197	<i>kdsD</i>	COG0517R, COG0794M	D-arabinose 5-phosphate isomerase
b3212	<i>gluB</i>	COG0067E, COG0069E, COG0070E	glutamate synthase, large subunit
b3230	<i>rpsI</i>	-	30S ribosomal protein S9
b3231	<i>rplM</i>	-	50S ribosomal protein L13
b3250	<i>mreC</i>	COG1792M	cell wall structural complex MreBCD transmembrane component MreC
b3261	<i>fis</i>	COG2901KL	DNA-binding protein Fis
b3296	<i>rpsD</i>	COG0522J	30S ribosomal protein S4
b3298	<i>rpsM</i>	COG0099J	30S ribosomal protein S13
b3306	<i>rpsH</i>	-	30S ribosomal protein S8
b3307	<i>rpsN</i>	COG0199J	30S ribosomal protein S14
b3339,b3980	<i>tufA,tufB</i>	COG0050J,CO G0050J	protein chain elongation factor EF-Tu (duplicate of tufB) ,protein chain elongation factor EF-Tu (duplicate of tufA)
b3340	<i>fusA</i>	COG0480J	elongation factor EF-2
b3341	<i>rpsG</i>	COG0049J	30S ribosomal protein S7
b3347	<i>fkpA</i>	COG0545O	FKBP-type peptidyl-prolyl cis-trans isomerase (rotamase)
b3425	<i>glpE</i>	COG0607P	thiosulfate sulfurtransferase
b3458	<i>livK</i>	COG0683E	leucine transporter subunit
b3493	<i>pitA</i>	COG0306P	phosphate transporter, low-affinity
b3620	<i>rfaF</i>	COG0859M	ADP-heptose:LPS heptosyltransferase II
b3636	<i>rpmG</i>	COG0267J	50S ribosomal protein L33
b3639	<i>dff</i>	COG0452H	bifunctional phosphopantothencysteine decarboxylase/phosphopantothenate synthase
b3640	<i>dut</i>	-	deoxyuridine 5'-triphosphate nucleotidohydrolase
b3641	<i>ttk</i>	COG1309K	nucleoid occlusion protein
b3648	<i>gmk</i>	COG0194F	guanylate kinase
b3649	<i>rpoZ</i>	COG1758K	DNA-directed RNA polymerase subunit omega
b3729	<i>glmS</i>	COG0449M	D-fructose-6-phosphate amidotransferase
b3733	<i>atpG</i>	COG0224C	F0F1 ATP synthase subunit gamma
b3737	<i>atpE</i>	COG0636C	F0F1 ATP synthase subunit C
b3744	<i>asnA</i>	COG2502E	asparagine synthetase AsnA
b3775	<i>ppiC</i>	COG0760O	peptidyl-prolyl cis-trans isomerase C (rotamase C)
b3776	<i>yifO</i>	COG3692S	hypothetical protein b3776
b3826	<i>yigL</i>	COG0561R	predicted hydrolase
b3860	<i>dsbA</i>	-	periplasmic protein disulfide isomerase I
b3936	<i>rpmE</i>	COG0254J	50S ribosomal subunit protein L31
b3961	<i>oxyR</i>	COG0583K	DNA-binding transcriptional dual regulator
b3972	<i>murB</i>	COG0812M	UDP-N-acetylenolpyruvoylglucosamine reductase
b3985	<i>rplJ</i>	COG0244J	50S ribosomal protein L10
b3986	<i>rplL</i>	COG0222J	50S ribosomal protein L7/L12
b3988	<i>rpoC</i>	COG0086K	DNA-directed RNA polymerase subunit beta'
b4041	<i>plsB</i>	COG2937I	glycerol-3-phosphate acyltransferase
b4222	<i>yifP</i>	COG2105S	hypothetical protein b4222
b4232	<i>fbp</i>	COG0158G	fructose-1,6-bisphosphatase
b4348	<i>hsdS</i>	COG0732V	specificity determinant for hsdM and hsdR
b4391	<i>yijK</i>	COG0488R	fused predicted transporter subunits of ABC superfamily: ATP-binding components
b4472	<i>yhdP</i>	COG3164S	conserved membrane protein, predicted transporter

T7-T8

b0104	<i>guaC</i>	COG0516F	guanosine 5'-monophosphate oxidoreductase
b0118	<i>acnB</i>	COG1049C	aconitate hydratase
b0167	<i>glnD</i>	COG2844O	PII uridylyl-transferase
b0240	<i>crl</i>	-	DNA-binding transcriptional regulator
b0346	<i>mhpR</i>	COG1414K	DNA-binding transcriptional activator, 3HPP-binding
b0654	<i>gluJ</i>	COG0765E	glutamate and aspartate transporter subunit
b0655	<i>gluI</i>	COG0834ET	glutamate and aspartate transporter subunit
b0722	<i>sdhD</i>	COG2142C	succinate dehydrogenase cytochrome b556 small membrane subunit
b0759	<i>galE</i>	COG1087M	UDP-galactose-4-epimerase
b0786	<i>ybhL</i>	COG0670R	predicted inner membrane protein

b0963	<i>mgsA</i>	-	methylglyoxal synthase
b0991	<i>ymcE</i>	-	cold shock gene
b1136	<i>icd</i>	COG0538C	isocitrate dehydrogenase
b1399	<i>paaX</i>	COG3327K	DNA-binding transcriptional repressor of phenylacetic acid degradation, aryl-CoA responsive
b1734	<i>chbF</i>	COG1486G	cryptic phospho-beta-glucosidase, NAD(P)-binding
b1749	<i>xthA</i>	COG0708L	exonuclease III
b2244	<i>yfaD</i>	COG5464S	hypothetical protein b2244
b2290	<i>yfbQ</i>	COG0436E	aspartate aminotransferase
b2294	<i>yfbU</i>	-	hypothetical protein b2294
b2308	<i>hisQ</i>	COG4215E	histidine/lysine/arginine/ornithine transporter subunit
b2309	<i>hisJ</i>	COG0834ET	histidine/lysine/arginine/ornithine transporter subunit
b2310	<i>argT</i>	COG0834ET	lysine/arginine/ornithine transporter subunit
b2346	<i>vacJ</i>	-	predicted lipoprotein
b2369	<i>evgA</i>	COG2197TK	DNA-binding response regulator in two-component regulatory system with EvgS
b2370	<i>evgS</i>	COG0642T, COG0834ET	hybrid sensory histidine kinase in two-component regulatory system with EvgA
b2371	<i>yfaE</i>	COG1804C	predicted CoA-transferase, NAD(P)-binding
b2374	<i>frc</i>	COG1804C	formyl-coenzyme A transferase
b2472	<i>dapE</i>	COG0624E	succinyl-diaminopimelate desuccinylase
b2552	<i>hmp</i>	COG1017C, COG1018C	fused nitric oxide dioxygenase/dihydropteridine reductase 2
b2703	<i>srlE</i>	-	glucitol/sorbitol-specific enzyme IIB component of PTS
b2774	<i>ygeW</i>	COG1028IQR	predicted deoxygluconate dehydrogenase
b2788	<i>gudX</i>	COG4948MR	predicted glucarate dehydratase
b3039	<i>zupT</i>	COG3384S	predicted dioxygenase
b3089	<i>ssiT</i>	COG3633E	sodium:serine/threonine symporter
b3103	<i>yhaH</i>	COG3152S	predicted inner membrane protein
b3209	<i>elbB</i>	COG3155Q	isoprenoid biosynthesis protein with amidotransferase-like domain
b3397	<i>nudE</i>	COG0494LR	ADP-ribose diphosphatase
b3399	<i>yrfG</i>	COG1011R	predicted hydrolase
b3438	<i>gntR</i>	COG1609K	DNA-binding transcriptional repressor
b3469	<i>zntA</i>	COG2217P	zinc, cobalt and lead efflux system
b3549	<i>tag</i>	-	3-methyl-adenine DNA glycosylase I, constitutive
b3575	<i>miaK</i>	-	2,3-diketo-L-gulonate dehydrogenase, NADH-dependent
b3599	<i>mtlA</i>	COG2213G, COG4668G	fused mannitol-specific PTS enzymes: IIA components/IIB components/IIC components
b3601	<i>mtlR</i>	-	DNA-binding repressor
b3604	<i>lldR</i>	COG2186K	DNA-binding transcriptional repressor
b3605	<i>lldD</i>	COG1304C	L-lactate dehydrogenase, FMN-linked
b3616	<i>tdh</i>	COG1063ER	L-threonine 3-dehydrogenase
b3748	<i>rhsD</i>	COG1869G	predicted cytoplasmic sugar-binding protein
b3773	<i>ilvY</i>	COG0583K	DNA-binding transcriptional dual regulator
b3789	<i>rffH</i>	COG1209M	glucose-1-phosphate thymidyltransferase
b3843	<i>ubiD</i>	-	3-octaprenyl-4-hydroxybenzoate decarboxylase
b3847	<i>pepQ</i>	COG0006E	proline dipeptidase
b3862	<i>yihG</i>	COG0204I	predicted endonuclease
b3894	<i>fdoG</i>	COG0243C	formate dehydrogenase-O, large subunit
b3963	<i>fabR</i>	COG1309K	DNA-binding transcriptional repressor
b4024	<i>lysC</i>	COG0527E	aspartate kinase III
b4035	<i>malK</i>	COG3839G	fused maltose transport subunit, ATP-binding component of ABC superfamily/regulatory protein
b4111	<i>proP</i>	-	proline/glycine betaine transporter
b4178	<i>yjeB</i>	COG1959K	predicted DNA-binding transcriptional regulator
b4208	<i>cycA</i>	COG1113E	D-alanine/D-serine/glycine transporter
b4214	<i>cysQ</i>	COG1218P	PAPS (adenosine 3'-phosphate 5'-phosphosulfate) 3'(2'),5'-bisphosphate nucleotidase
b4342	<i>yjiT</i>	-	hypothetical protein b4342
b4394	<i>yjiX</i>	-	NTPase
b4579	<i>yaiX</i>	-	-

T8-T9

b0023	<i>rpsT</i>	COG0268J	30S ribosomal protein S20
b0094	<i>ftsA</i>	COG0849D	cell division protein
b0116	<i>lpd</i>	COG1249C	dihydroliipoamide dehydrogenase
b0179	<i>lpxD</i>	COG1044M	UDP-3-O-
b0280	<i>yagN</i>	-	CP4-6 prophage; predicted protein
b0407	<i>yajC</i>	COG1862U	preprotein translocase subunit YajC
b0430	<i>cyoC</i>	COG1845C	cytochrome o ubiquinol oxidase subunit III
b0436	<i>tig</i>	COG0544O	trigger factor
b0653	<i>gltK</i>	COG0765E	glutamate and aspartate transporter subunit
b0723	<i>sdhA</i>	COG1053C	succinate dehydrogenase flavoprotein subunit
b0811	<i>glnH</i>	COG0834ET	glutamine ABC transporter periplasmic protein
b0911	<i>rpsA</i>	COG0539J	30S ribosomal protein S1
b0926	<i>ycbK</i>	COG3108S	hypothetical protein b0926
b0937	<i>ssuE</i>	COG0431R	NAD(P)H-dependent FMN reductase
b1101	<i>ptsG</i>	COG1263G	fused glucose-specific PTS enzymes: IIB component/IIC component

b1136	<i>icd</i>	COG0538C	isocitrate dehydrogenase
b1237	<i>hns</i>	COG2916R	global DNA-binding transcriptional dual regulator H-NS
b1243	<i>oppA</i>	COG4166E	oligopeptide transporter subunit
b1251	<i>ycil</i>	-	predicted enzyme
b1324	<i>tpx</i>	COG2077O	thiol peroxidase
b1329	<i>mppA</i>	COG4166E	murein tripeptide (L-ala-gamma-D-glutamyl-meso-DAP) transporter subunit
b1614	<i>ydgA</i>	COG5339S	hypothetical protein b1614
b1654	<i>ydhD</i>	COG0278O	hypothetical protein b1654
b1719	<i>thrS</i>	COG0441J	threonyl-tRNA synthetase
b1823	<i>cspC</i>	-	stress protein, member of the CspA-family
b1824	<i>yobF</i>	-	hypothetical protein b1824
b1832	<i>yebR</i>	COG1956T	hypothetical protein b1832
b2096	<i>gaiY</i>	COG0191G	D-tagatose 1,6-bisphosphate aldolase 2, catalytic subunit
b2144	<i>sanA</i>	COG2949S	hypothetical protein b2144
b2234	<i>nrdA</i>	COG0209F	ribonucleotide-diphosphate reductase alpha subunit
b2287	<i>nuoB</i>	COG0377C	NADH dehydrogenase subunit B
b2415	<i>ptsH</i>	COG1925G	phosphohistidinoprotein-hexose phosphotransferase component of PTS system (Hpr)
b2478	<i>dapA</i>	COG0329EM	dihydrodipicolinate synthase
b2551	<i>glyA</i>	COG0112E	serine hydroxymethyltransferase
b2557	<i>purL</i>	COG0046F, COG0047F	phosphoribosylformylglycinamide synthase
b2585	<i>pssA</i>	COG1502I	phosphatidylserine synthase
b2697	<i>alaS</i>	COG0013J	alanyl-tRNA synthetase
b2780	<i>pyrG</i>	COG0504F	CTP synthetase
b2913	<i>serA</i>	COG0111HE	D-3-phosphoglycerate dehydrogenase
b2926	<i>pgk</i>	-	phosphoglycerate kinase
b2935	<i>tktA</i>	COG0021G	transketolase 1, thiamin-binding
b2959	<i>yggL</i>	-	hypothetical protein b2959
b2960	<i>yggH</i>	COG0220R	tRNA(m7G46)-methyltransferase
b3038	<i>ygiC</i>	-	predicted enzyme
b3204	<i>ptsN</i>	COG1762GT	sugar-specific enzyme IIA component of PTS
b3311	<i>rpsQ</i>	COG0186J	30S ribosomal protein S17
b3339,b3980	<i>tufA,tufB</i>	COG0050J,CO G0050J	protein chain elongation factor EF-Tu (duplicate of tufB) ,protein chain elongation factor EF-Tu (duplicate of tufA)
b3357	<i>crp</i>	COG0664T	DNA-binding transcriptional dual regulator
b3386	<i>rpe</i>	COG0036G	ribulose-phosphate 3-epimerase
b3418	<i>malT</i>	COG2909K	transcriptional regulator MalT
b3432	<i>glgB</i>	COG0296G	glycogen branching enzyme
b3528	<i>dctA</i>	COG1301C	C4-dicarboxylate transport protein
b3556	<i>cspA</i>	-	major cold shock protein
b3560	<i>glyQ</i>	-	glycyl-tRNA synthetase subunit alpha
b3636	<i>rpmG</i>	COG0267J	50S ribosomal protein L33
b3688	<i>yidQ</i>	-	conserved outer membrane protein
b3737	<i>atpE</i>	COG0636C	F0F1 ATP synthase subunit C
b3752	<i>rbsK</i>	COG0524G	ribokinase
b3782	<i>rhoL</i>	-	rho operon leader peptide
b3783	<i>rho</i>	COG1158K	transcription termination factor Rho
b3805	<i>hemC</i>	-	porphobilinogen deaminase
b3919	<i>tpiA</i>	COG0149G	triosephosphate isomerase
b3928	<i>yiiU</i>	COG3074S	hypothetical protein b3928
b3962	<i>sihA</i>	COG1249C	soluble pyridine nucleotide transhydrogenase
b3964	<i>yijD</i>	-	conserved inner membrane protein
b3984	<i>rpLA</i>	-	50S ribosomal protein L1
b3985	<i>rpIJ</i>	COG0244J	50S ribosomal protein L10
b4000	<i>hupA</i>	-	HU, DNA-binding transcriptional regulator, alpha subunit
b4014	<i>aceB</i>	COG2225C	malate synthase
b4035	<i>malK</i>	COG3839G	fused maltose transport subunit, ATP-binding component of ABC superfamily/regulatory protein
b4177	<i>purA</i>	-	adenylosuccinate synthetase
b4232	<i>fbp</i>	COG0158G	fructose-1,6-bisphosphatase
b4381	<i>deoC</i>	COG0274F	deoxyribose-phosphate aldolase
b4382	<i>deoA</i>	COG0213F	thymidine phosphorylase
b4384	<i>deoD</i>	COG0813F	purine nucleoside phosphorylase

T9-T10

b0220	<i>ivy</i>	-	inhibitor of vertebrate C-lysozyme
b0440	<i>hupB</i>	-	HU, DNA-binding transcriptional regulator, beta subunit
b1315	<i>ycjS</i>	COG0673R	predicted oxidoreductase, NADH-binding
b1651	<i>gloA</i>	COG0346E	glyoxalase I, Ni-dependent
b1819	<i>manZ</i>	-	mannose-specific enzyme IID component of PTS
b3128	<i>garD</i>	-	(D)-galactarate dehydrogenase
b3535	<i>yhjR</i>	-	hypothetical protein b3535
b4076	<i>nrfG</i>	COG4235O	heme lyase (NrfEFG) for insertion of heme into c552, subunit NrfG
b4377	<i>yjjU</i>	COG4667R	predicted esterase
b4428,b1420	<i>hokB,mokB</i>	-,-	toxic polypeptide, small ,regulatory peptide

T10-T11

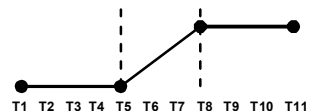
b0023	<i>rpsT</i>	COG0268J	30S ribosomal protein S20
b0169	<i>rpsB</i>	-	30S ribosomal protein S2
b0179	<i>lpxD</i>	COG1044M	UDP-3-O-
b0181	<i>lpxA</i>	COG1043M	UDP-N-acetylglucosamine acyltransferase
b0334	<i>prpD</i>	-	2-methylcitrate dehydratase
b0341	<i>cynX</i>	COG2807P	predicted cyanate transporter
b0407	<i>yajC</i>	COG1862U	preprotein translocase subunit YajC
b0440	<i>hupB</i>	-	HU, DNA-binding transcriptional regulator, beta subunit
b0642	<i>leuS</i>	COG0495J	leucyl-tRNA synthetase
b0710	<i>ybgI</i>	COG0327S	conserved metal-binding protein
b0711	<i>ybgJ</i>	-	predicted enzyme subunit
b0729	<i>sucD</i>	COG0074C	succinyl-CoA synthetase subunit alpha
b0734	<i>cydB</i>	COG1294C	cytochrome d terminal oxidase, subunit II
b0912	<i>ihfB</i>	-	integration host factor subunit beta
b1101	<i>ptsG</i>	COG1263G	fused glucose-specific PTS enzymes: IIB component/IIC component
b1237	<i>hns</i>	COG2916R	global DNA-binding transcriptional dual regulator H-NS
b1247	<i>oppF</i>	COG4608E	oligopeptide transporter subunit
b1676	<i>pykF</i>	-	pyruvate kinase
b2096	<i>gaiY</i>	COG0191G	D-tagatose 1,6-bisphosphate aldolase 2, catalytic subunit
b2265	<i>menF</i>	COG1169HQ	menaquinone-specific isochorismate synthase
b2415	<i>ptsH</i>	COG1925G	phosphohistidinoprotein-hexose phosphotransferase component of PTS system (Hpr)
b2741	<i>rpoS</i>	COG0568K	RNA polymerase sigma factor
b2959	<i>yggL</i>	-	hypothetical protein b2959
b3185	<i>rpmA</i>	-	50S ribosomal protein L27
b3296	<i>rpsD</i>	COG0522J	30S ribosomal protein S4
b3301	<i>rplO</i>	COG0200J	50S ribosomal protein L15
b3303	<i>rpsE</i>	COG0098J	30S ribosomal protein S5
b3309	<i>rplX</i>	COG0198J	50S ribosomal protein L24
b3310	<i>rplN</i>	-	50S ribosomal protein L14
b3313	<i>rplP</i>	COG0197J	50S ribosomal protein L16
b3314	<i>rpsC</i>	COG0092J	30S ribosomal protein S3
b3315	<i>rplV</i>	-	50S ribosomal protein L22
b3317	<i>rplB</i>	COG0090J	50S ribosomal protein L2
b3319	<i>rplD</i>	-	50S ribosomal protein L4
b3320	<i>rplC</i>	-	50S ribosomal protein L3
b3321	<i>rpsJ</i>	COG0051J	30S ribosomal protein S10
b3339,b3980	<i>tufA,tufB</i>	COG0050J,COG0050J	protein chain elongation factor EF-Tu (duplicate of tufB) ,protein chain elongation factor EF-Tu (duplicate of tufA)
b3341	<i>rpsG</i>	COG0049J	30S ribosomal protein S7
b3342	<i>rpsL</i>	-	30S ribosomal protein S12
b3349	<i>slyD</i>	COG1047O	FKBP-type peptidyl prolyl cis-trans isomerase (rotamase)
b3495	<i>uspA</i>	COG0589T	universal stress global response regulator
b3637	<i>rpmB</i>	COG0227J	50S ribosomal protein L28
b3848	<i>yigZ</i>	COG1739S	predicted elongation factor
b3919	<i>tpiA</i>	COG0149G	triosephosphate isomerase
b3986	<i>rplL</i>	COG0222J	50S ribosomal protein L7/L12
b4014	<i>aceB</i>	COG2225C	malate synthase
b4056	<i>yjbQ</i>	-	hypothetical protein b4056
b4172	<i>hfq</i>	COG1923R	RNA-binding protein Hfq
b4201	<i>priB</i>	COG2965L	primosomal replication protein N
b4202	<i>rpsR</i>	-	30S ribosomal protein S18

付表3 : 統合パターンの発現上昇の位置による遺伝子の分類結果

b no.	gene	cogID	function.	rpoS

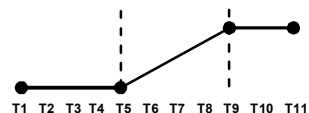
T4-T8

b0312	<i>betB</i>	COG1012C	betaine aldehyde dehydrogenase, NAD-dependent	
b0366	<i>tauB</i>	COG4525P	taurine transporter subunit	
b0522	<i>purK</i>	COG0026F	phosphoribosylaminoimidazole carboxylase	
b0907	<i>serC</i>	COG1932HE	phosphoserine aminotransferase	
b1423	<i>ydcJ</i>	COG5383S	hypothetical protein b1423	
b1661	<i>cfa</i>	COG2230M	cyclopropane fatty acyl phospholipid synthase (unsaturated-phospholipid)	<i>rpoS regulon</i>
b1668	<i>ydhS</i>	COG4529S	conserved protein with FAD/NAD(P)-binding domain	
b1682	<i>sufC</i>	COG0396O	cysteine desulfurase ATPase component	
b1927	<i>amyA</i>	-	cytoplasmic alpha-amylase	
b1995	<i>yoeA_1</i>	-	-	
b3510	<i>hdeA</i>	-	stress response protein acid-resistance protein	<i>rpoS regulon</i>
b3544	<i>dppA</i>	COG0747E	dipeptide transporter	
b3546	<i>eptB</i>	COG2194R	predicted metal dependent hydrolase	
b3831	<i>udp</i>	COG2820F	uridine phosphorylase	
b3888	<i>yjiD</i>	-	predicted acetyltransferase	
b4140	<i>fxsA</i>	COG3030R	inner membrane protein	



T5-T8

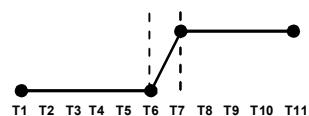
b0997	<i>torA</i>	COG0243C	trimethylamine N-oxide (TMAO) reductase 1, catalytic subunit	
b1466	<i>narW</i>	COG2180C	nitrate reductase 2 (NRZ), delta subunit (assembly subunit)	
b1515	<i>lsrD</i>	COG1172G	AI2 transporter	<i>rpoS regulon</i>
b1516	<i>lsrB</i>	COG1879G	AI2 transporter	<i>rpoS regulon</i>
b1744	<i>astE</i>	-	succinylglutamate desuccinylase	<i>rpoS regulon</i>
b2204	<i>napH</i>	COG0348C	quinol dehydrogenase membrane component	
b2312	<i>purF</i>	COG0034F	amidophosphoribosyltransferase	
b2361	<i>yfdR</i>	COG1896R	CPS-53 (KpLE1) prophage; conserved protein	
b2557	<i>purL</i>	COG0046F, COG0047F	phosphoribosylformylglycinamide synthase	
b3508	<i>yhiD</i>	COG1285S	predicted Mg(2+) transport ATPase inner membrane protein	
b4064	<i>yjcD</i>	COG2252R	predicted permease	
b4171	<i>miaA</i>	COG0324J	tRNA delta(2)-isopentenylpyrophosphate transferase	



T5-T9

b0009	<i>mog</i>	-	molybdenum cofactor biosynthesis protein	
b0146	<i>sfsA</i>	-	sugar fermentation stimulation protein A	
b0162	<i>cdaR</i>	COG3835KT	DNA-binding transcriptional activator	
b0361,b1402,b1996,b2860,b3045,b4273	<i>insD-1,insD-2,insD-3,insD-4,insD-5,insD-6</i>	COG2801L,COG2801L,COG2801L,COG2801L,COG2801L,COG2801L,COG2801L	IS2 insertion element transposase InsAB' ,IS2 insertion element transposase InsAB' ,CP4-44 prophage; IS2 insertion element transposase InsAB' ,IS2 insertion element transposase InsAB' ,IS2 insertion element transposase InsAB' ,KpLE2 phage-like element; IS2 insertion element transposase InsAB'	
b0413	<i>ybaD</i>	-	hypothetical protein b0413	
b0515	<i>yIbA</i>	COG3257R	hypothetical protein b0515	
b0779	<i>uvrB</i>	COG0556L	excinuclease ABC subunit B	
b1004	<i>wrbA</i>	COG0655R	TrpR binding protein WrbA	
b1063	<i>yceB</i>	-	predicted lipoprotein	
b1111	<i>yefQ</i>	COG1309K	predicted DNA-binding transcriptional regulator	
b1115	<i>yefT</i>	COG4763S	predicted inner membrane protein	
b1283	<i>osmB</i>	-	lipoprotein	<i>rpoS regulon</i>
b1423	<i>ydcJ</i>	COG5383S	hypothetical protein b1423	
b1488	<i>ddpX</i>	COG2173M	D-ala-D-ala dipeptidase, Zn-dependent	
b1691	<i>ydiN</i>	-	predicted transporter	
b1774	<i>ydfJ</i>	COG1063ER	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	
b1799	<i>yeaT</i>	COG0583K	predicted DNA-binding transcriptional regulator	
b1814	<i>sdaA</i>	COG1760E	L-serine deaminase I	

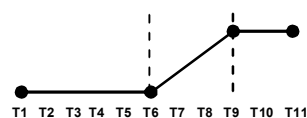
b1843	<i>yobB</i>	COG0388R	hypothetical protein b1843	
b1959	<i>yedA</i>	-	predicted inner membrane protein	
b2016	<i>yeeZ</i>	COG0451MG	predicted epimerase, with NAD(P)-binding Rossmann-fold domain	
b2072	<i>yegK</i>	-	hypothetical protein b2072	
b2439	<i>euL</i>	-	predicted carboxysome structural protein with predicted role in ethanolamine utilization	
b2488	<i>hyfH</i>	COG1143C	hydrogenase 4, Fe-S subunit	
b2802	<i>fucI</i>	COG2407G	L-fucose isomerase	
b2878	<i>ygfK</i>	COG0493ER	predicted oxidoreductase, Fe-S subunit	
b2968	<i>yghD</i>	-	predicted secretion pathway M-type protein, membrane anchored	
b3059	<i>ygiH</i>	COG0344S	hypothetical protein b3059	
b3135	<i>agaA</i>	COG1820G	predicted truncated N-acetylgalactosamine-6-phosphate deacetylase	
b3448	<i>yhhA</i>	-	hypothetical protein b3448	
b3471	<i>yhhQ</i>	COG1738S	conserved inner membrane protein	
b3503	<i>arsC</i>	COG1393P	arsenate reductase	
b3521	<i>yhjC</i>	COG0583K	predicted DNA-binding transcriptional regulator	
b3581	<i>sgbH</i>	COG0269G	3-keto-L-gulonate 6-phosphate decarboxylase	
b3751	<i>rbsB</i>	COG1879G	D-ribose transporter subunit	
b3848	<i>yigZ</i>	COG1739S	predicted elongation factor	
b3882	<i>yihU</i>	COG2084I	predicted oxidoreductase with NAD(P)-binding Rossmann-fold domain	
b4217	<i>ytfK</i>	-	hypothetical protein b4217	<i>rpoS regulon</i>
b4257	<i>yjgN</i>	-	conserved inner membrane protein	
b4328	<i>iadA</i>	COG0044F, COG1820G	isoaspartyl dipeptidase	
b4358	<i>yjjN</i>	COG1063ER	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	
b4460	<i>araH</i>	COG1172G	fused L-arabinose transporter subunits of ABC superfamily: membrane components	



T6-T7

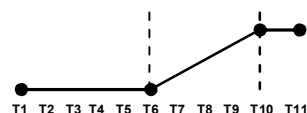
b0155	<i>clcA</i>	COG0038P	chloride channel protein	
b0186	<i>ldcC</i>	COG1982E	lysine decarboxylase 2, constitutive	<i>rpoS regulon</i>
b0187	<i>yaeR</i>	COG0346E	predicted lyase	
b0212	<i>gloB</i>	COG0491R	predicted hydroxyacylglutathione hydrolase	
b0219	<i>yaV</i>	COG0388R	predicted C-N hydrolase family amidase, NAD(P)-binding	
b0391	<i>yaiE</i>	-	hypothetical protein b0391	
b0643	<i>ybeL</i>	-	hypothetical protein b0643	
b0712	<i>ybgK</i>	-	predicted enzyme subunit	
b0750	<i>nadA</i>	-	quinolinate synthetase	
b0865	<i>ybjP</i>	-	predicted lipoprotein	<i>rpoS regulon</i>
b0867	<i>ybjR</i>	COG3023V	predicted amidase and lipoprotein	
b1014	<i>putA</i>	COG4230C	fused DNA-binding transcriptional regulator/proline dehydrogenase/pyrroline-5-carboxylate dehydrogenase	
b1051	<i>msyB</i>	-	hypothetical protein b1051	<i>rpoS regulon</i>
b1065	<i>mdtH</i>	-	predicted drug efflux system	
b1195	<i>ymgE</i>	COG2261S	predicted inner membrane protein	
b1200	<i>dhaK</i>	COG2376G	dihydroxyacetone kinase, N-terminal domain	
b1297	<i>puuA</i>	-	gamma-Glu-putrescine synthase	
b1333	<i>uspE</i>	COG0589T	stress-induced protein	
b1376	<i>uspF</i>	COG0589T	stress-induced protein, ATP-binding protein	
b1423	<i>ydjJ</i>	COG5383S	hypothetical protein b1423	
b1598	<i>ydgD</i>	-	predicted peptidase	
b1606	<i>folM</i>	COG10281QR	short chain dehydrogenase	
b1739	<i>osmE</i>	-	DNA-binding transcriptional activator	<i>rpoS regulon</i>
b1746	<i>astD</i>	COG1012C	succinylglutamic semialdehyde dehydrogenase	<i>rpoS regulon</i>
b1747	<i>astA</i>	-	arginine succinyltransferase	<i>rpoS regulon</i>
b1748	<i>astC</i>	COG4992E	succinylornithine transaminase, PLP-dependent	<i>rpoS regulon</i>
b1896	<i>otsA</i>	COG0380G	trehalose-6-phosphate synthase	<i>rpoS regulon</i>
b2133	<i>dld</i>	COG0277C	D-lactate dehydrogenase, FAD-binding, NADH independent	
b2137	<i>yohF</i>	COG10281QR	predicted oxidoreductase with NAD(P)-binding Rossmann-fold domain	
b2143	<i>cdd</i>	COG0295F	cytidine deaminase	
b2146	<i>yeiT</i>	COG0493ER	predicted oxidoreductase	
b2150	<i>mgIB</i>	COG1879G	methyl-galactoside transporter subunit	<i>rpoS regulon</i>
b2170	<i>setB</i>	-	lactose/glucose efflux system	
b2388	<i>glk</i>	COG0837G	glucokinase	
b2392	<i>mntH</i>	COG1914P	manganese transport protein MntH	
b2453	<i>eutG</i>	COG1454C	predicted alcohol dehydrogenase in ethanolamine utilization	
b2494	<i>yfgC</i>	COG4783R	predicted peptidase	
b2535	<i>csiE</i>	COG3711K	stationary phase inducible protein	<i>rpoS regulon</i>
b2587	<i>kgiP</i>	-	alpha-ketoglutarate transporter	
b2597	<i>yfiA</i>	COG1544J	cold shock protein associated with 30S ribosomal subunit	
b2795	<i>ygdH</i>	COG1611R	hypothetical protein b2795	
b2804	<i>fucU</i>	COG4154G	L-fucose mutarotase	
b2833	<i>ygdR</i>	-	hypothetical protein b2833	
b2835	<i>ygeD</i>	-	predicted inner membrane protein	

b2869	<i>ygeV</i>	COG3829KT	predicted DNA-binding transcriptional regulator	
b3012	<i>dkgA</i>	COG0656R	2,5-diketo-D-gluconate reductase A	
b3023	<i>ygiV</i>	COG3449L	predicted transcriptional regulator	
b3049	<i>glgS</i>	-	glycogen synthesis protein GlgS	<i>rpoS regulon</i>
b3080	<i>ygiK</i>	COG1626G, COG3408G	predicted glycosyl hydrolase	
b3104	<i>yhaI</i>	COG3152S	predicted inner membrane protein	
b3160	<i>yhbW</i>	COG2141C	predicted enzyme	
b3221	<i>yhcH</i>	-	hypothetical protein b3221	
b3239	<i>yhcO</i>	COG2732K	predicted barnase inhibitor	
b3292	<i>zntR</i>	COG0789K	zinc-responsive transcriptional regulator	
b3548	<i>yhjY</i>	-	hypothetical protein b3548	
b3600	<i>mtiD</i>	-	mannitol-1-phosphate 5-dehydrogenase	
b3678	<i>yidJ</i>	COG3119P	predicted sulfatase/phosphatase	
b3830	<i>ysgA</i>	COG0412Q	predicted hydrolase	
b3995	<i>rsd</i>	-	stationary phase protein, binds sigma 70 RNA polymerase subunit	<i>rpoS regulon</i>
b4056	<i>yjbQ</i>	-	hypothetical protein b4056	
b4068	<i>yjch</i>	-	conserved inner membrane protein involved in acetate transport	
b4117	<i>adiA</i>	COG1982E	biodegradative arginine decarboxylase	
b4127	<i>yjAJ</i>	COG2388R	predicted acyltransferase with acyl-CoA N-acyltransferase domain	
b4219	<i>msrA</i>	-	methionine sulfoxide reductase A	
b4235	<i>pmbA</i>	COG0312R	predicted peptidase required for the maturation and secretion of the antibiotic peptide	
b4363	<i>yjjB</i>	COG3610S	conserved inner membrane protein	
b4376	<i>osmY</i>	COG2823R	periplasmic protein	<i>rpoS regulon</i>
b4383	<i>deoB</i>	COG1015G	phosphopentomutase	



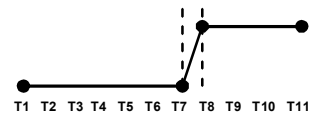
T6-T9

b0300	<i>ykgA</i>	-	predicted DNA-binding transcriptional regulator	
b0419	<i>yajO</i>	COG0667C	predicted oxidoreductase, NAD(P)-binding	
b0812	<i>dps</i>	COG0783P	DNA protection during starvation conditions	<i>rpoS regulon</i>
b0866	<i>ybjQ</i>	-	hypothetical protein b0866	
b0906	<i>ycaP</i>	-	conserved inner membrane protein	
b0955	<i>ycbZ</i>	COG1067O	predicted peptidase	
b0965	<i>yccU</i>	COG1832R	predicted CoA-binding protein with NAD(P)-binding Rossmann-fold domain	
b1150	<i>ymfR</i>	-	e14 prophage; predicted protein	
b1165	<i>ymgA</i>	-	hypothetical protein b1165	
b1188	<i>yecB</i>	COG2719S	hypothetical protein b1188	
b1325	<i>ycjG</i>	COG4948MR	L-Ala-D/L-Glu epimerase	
b1478	<i>adhP</i>	COG1064R	alcohol dehydrogenase	
b1586	<i>ynfD</i>	-	hypothetical protein b1586	
b1724	<i>ydiZ</i>	-	hypothetical protein b1724	
b1725	<i>yniA</i>	COG3001G	predicted phosphotransferase/kinase	
b1783	<i>yeaG</i>	COG2766T	conserved protein with nucleoside triphosphate hydrolase domain	
b1955	<i>yedP</i>	COG3769R	hypothetical protein b1955	
b1967	<i>hchA</i>	COG0693R	chaperone protein HchA	
b1988	<i>nac</i>	COG0583K	DNA-binding transcriptional dual regulator of nitrogen assimilation	
b2021	<i>hisC</i>	COG0079E	histidinol-phosphate aminotransferase	
b2134	<i>pbpG</i>	COG1686M	D-alanyl-D-alanine endopeptidase	
b2151	<i>galS</i>	COG1609K	DNA-binding transcriptional repressor	
b2543	<i>yphA</i>	COG2259S	predicted inner membrane protein	
b2573	<i>rpoE</i>	COG1595K	RNA polymerase, sigma 24 (sigma E) factor	
b2583	<i>yfiP</i>	-	hypothetical protein b2583	
b3050	<i>yqiJ</i>	-	predicted inner membrane protein	
b3494	<i>yhiO</i>	-	universal stress protein UspB	<i>rpoS regulon</i>
b3766	<i>ilvL</i>	-	ilvG operon leader peptide	
b3922	<i>yiiS</i>	-	hypothetical protein b3922	
b4214	<i>cysQ</i>	COG1218P	PAPS (adenosine 3'-phosphate 5'(2'),5'-bisphosphate nucleotidase	
b4568	<i>ytjA</i>	-	hypothetical protein b4568	



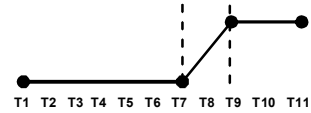
T6-T10

b0435	<i>bolA</i>	COG0271T	regulator of penicillin binding proteins and beta lactamase transcription (morphogene)	<i>rpoS regulon</i>
b0857	<i>potI</i>	COG1177E	putrescine transporter subunit: membrane component of ABC superfamily	
b1327	<i>ycjY</i>	COG1073R	predicted hydrolase	
b1440	<i>ydcS</i>	COG0687E	predicted spermidine/putrescine transporter subunit	
b1712	<i>ihfA</i>	-	integration host factor subunit alpha	<i>rpoS regulon</i>
b2310	<i>argT</i>	COG0834ET	lysine/arginine/ornithine transporter subunit	
b3362	<i>yhfG</i>	-	hypothetical protein b3362	
b3423	<i>glpR</i>	COG1349KG	DNA-binding transcriptional repressor	



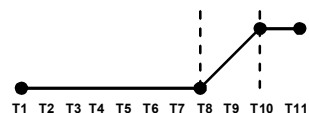
T7-T8

b0273	<i>argF</i>	COG0078E	CP4-6 prophage; ornithine carbamoyltransferase 2, chain F	
b0484	<i>copA</i>	COG2217P	copper transporter	
b0485	<i>ybaS</i>	-	predicted glutaminase	
b0487	<i>cueR</i>	COG0789K	DNA-binding transcriptional activator of copper-responsive regulon genes	
b0713	<i>ybgL</i>	COG1540R	hypothetical protein b0713	
b0927	<i>ycbL</i>	COG0491R	predicted metal-binding enzyme	
b0978	<i>appC</i>	COG1271C	cytochrome bd-II oxidase, subunit I	<i>rpoS regulon</i>
b0979	<i>appB</i>	COG1294C	cytochrome bd-II oxidase, subunit II	<i>rpoS regulon</i>
b1003	<i>yccJ</i>	-	hypothetical protein b1003	
b1343	<i>dbpA</i>	COG0513LKJ	ATP-dependent RNA helicase, specific for 23S rRNA	
b1465	<i>narV</i>	COG2181C	nitrate reductase 2 (NRZ), gamma subunit	
b1512	<i>lsrR</i>	COG2390K	lsr operon transcriptional repressor	
b1518	<i>lsrG</i>	COG1359S	autoinducer-2 (AI-2) modifying protein LsrG	<i>rpoS regulon</i>
b1794	<i>yeaP</i>	-	predicted diguanylate cyclase	
b2789	<i>gudP</i>	-	predicted D-glucarate transporter	
b3511	<i>hdeD</i>	COG3247S	acid-resistance membrane protein	
b3512	<i>gdeE</i>	COG2197TK	DNA-binding transcriptional activator	<i>rpoS regulon</i>
b3513	<i>mdtE</i>	COG0845M	multidrug resistance efflux transporter	
b3588	<i>aldB</i>	COG1012C	aldehyde dehydrogenase B	<i>rpoS regulon</i>
b4187	<i>aidB</i>	COG1960I	isovaleryl CoA dehydrogenase	<i>rpoS regulon</i>



T7-T9

b0077	<i>ilvI</i>	COG0028EH	acetolactate synthase III large subunit	
b0131	<i>panD</i>	-	aspartate 1-decarboxylase precursor	
b0226	<i>dinJ</i>	COG3077L	predicted antitoxin of YafQ-DinJ toxin-antitoxin system	
b0330	<i>prpR</i>	COG1221KT, COG3829KT	DNA-binding transcriptional activator	
b0418	<i>pgpA</i>	COG1267I	phosphatidylglycerophosphatase A	
b0451	<i>amtB</i>	-	ammonium transporter	
b0464	<i>acrR</i>	COG1309K	DNA-binding transcriptional repressor	
b0481	<i>ybaK</i>	COG2606S	hypothetical protein b0481	
b0597	<i>ybdB</i>	COG2050Q	hypothetical protein b0597	
b0604	<i>dsbG</i>	COG1651O	periplasmic disulfide isomerase/thiol-disulphide oxidase	
b0631	<i>ybeD</i>	-	hypothetical protein b0631	
b0904	<i>focA</i>	COG2116P	formate transporter	
b0947	<i>ycbX</i>	COG0633C, COG3217R	predicted 2Fe-2S cluster-containing protein	
b1045	<i>ymdB</i>	-	hypothetical protein b1045	
b1217	<i>chaB</i>	COG4572R	cation transport regulator	
b1477	<i>yddM</i>	COG3093R	predicted DNA-binding transcriptional regulator	
b1482	<i>osmC</i>	COG1764O	osmotically inducible, stress-inducible membrane protein	<i>rpoS regulon</i>
b1534	<i>ydeE</i>	-	predicted transporter	
b1564	<i>relB</i>	COG3077L	Qin prophage; bifunctional antitoxin of the RelE-RelB toxin-antitoxin system/ transcriptional repressor	
b1637	<i>tyrS</i>	COG0162J	tyrosyl-tRNA synthetase	
b1679	<i>sufE</i>	COG2166R	cysteine desulfuration protein SufE	
b1684	<i>sufA</i>	COG0316S	iron-sulfur cluster assembly scaffold protein	
b1836	<i>yebV</i>	-	hypothetical protein b1836	
b1897	<i>otsB</i>	COG1877G	trehalose-6-phosphate phosphatase, biosynthetic	<i>rpoS regulon</i>
b2067	<i>yegE</i>	COG2202T, COG5001T	predicted diguanylate cyclase, GGDEF domain signalling protein	
b2080	<i>yegP</i>	-	hypothetical protein b2080	
b2217	<i>rscB</i>	COG2197TK	DNA-binding response regulator in two-component regulatory system with RcsC and YojN	
b2295	<i>yfbV</i>	-	hypothetical protein b2295	
b2799	<i>fucO</i>	COG1454C	L-1,2-propanediol oxidoreductase	
b3073	<i>ygiG</i>	COG4992E	putrescine:2-oxoglutaric acid aminotransferase, PLP-dependent	
b3392	<i>yrfA</i>	-	hypothetical protein b3392	
b3623	<i>rfaK</i>	COG0859M	lipopolysaccharide core biosynthesis	
b3929	<i>rraA</i>	-	ribonuclease activity regulator protein RraA	
b3958	<i>argC</i>	COG0002E	N-acetyl-gamma-glutamyl-phosphate reductase	
b4058	<i>uvrA</i>	COG0178L	excinuclease ABC subunit A	
b4065	<i>yjcE</i>	COG0025P	predicted cation/proton antiporter	
b4072	<i>nrjC</i>	COG0437C	formate-dependent nitrite reductase, 4Fe4S subunit	
b4137	<i>cutA</i>	COG1324P	copper binding protein, copper sensitivity	



T8-T10

b0143	<i>pcnB</i>	COG0617J	poly(A) polymerase I	
b0912	<i>ihfB</i>	-	integration host factor subunit beta	<i>rpoS regulon</i>
b1013	<i>yedC</i>	COG1309K	predicted DNA-binding transcriptional regulator	
b1106	<i>ycfN</i>	COG0510M	thiamin kinase	
b1166	<i>ymgB</i>	-	hypothetical protein b1166	
b1235	<i>rssB</i>	COG0745TK	response regulator of RpoS	<i>rpoS regulon</i>
b1266	<i>yciV</i>	COG0613R	hypothetical protein b1266	
b1446	<i>ydcY</i>	-	hypothetical protein b1446	
b1528	<i>ydeA</i>	COG2814G	sugar efflux transporter	
b1608	<i>rstA</i>	COG0745TK	DNA-binding response regulator in two-component regulatory system with RstB	
b1647	<i>ydhF</i>	COG4989R	predicted oxidoreductase	
b1648	<i>ydhL</i>	COG3313R	hypothetical protein b1648	
b1806	<i>yeaY</i>	-	predicted lipoprotein	
b2111	<i>yehD</i>	COG3539NU	predicted fimbrial-like adhesin protein	
b2173	<i>yeiR</i>	COG0523R	predicted enzyme	
b2250	<i>yfaZ</i>	-	predicted outer membrane porin protein	
b2418	<i>pdxK</i>	COG2240H	pyridoxine kinase	
b3361	<i>fic</i>	COG2184D	stationary-phase protein, cell division	<i>rpoS regulon</i>
b3862	<i>yihG</i>	COG0204I	predicted endonuclease	<i>rpoS regulon</i>
b4205	<i>ytfA</i>	-	predicted transcriptional regulator	
