博士論文

Establishing the Reference-Genes System for Accurate Quantification of Gene Expression

遺伝子発現を正確に定量するためのレファレンスジーンシステムの確立

国立大学法人　横浜国立大学大学院

環境情報学府

常　青

Qing Chang

March 2009

Table of contents

Synopsis

The bioremediation has become well developed as a means of easily cleaning up degraded petroleum products. Bacteria have access to a variety of compounds to help them produce energy and nutrients to build more cells. In a few cases the natural conditions at the contaminated site provide all the necessary materials in large enough quantities that bioremediation can occur. Bioremediation requires the construction of engineered system to supply microbe-stimulating materials. A critical factor in deciding whether bioremediation can happen is the activity of the bacteria and the fate of target genes.

In order to understand biological phenomena in molecular levels, the molecular biological methods have been developed. The reverse transcription polymerase chain reaction (RT-PCR) has become the most sensitive methods for characterizing or confirming gene expression patterns and comparing mRNA levels in different samples. The conventional PCR based quantification has the disadvantage of relying on end-point measurement of the amount of DNA produced, which makes it difficult to determine the initial concentration of template DNA. The recent development of real-time RT-PCR combines the amplification and analysis steps of PCR and its sensitivity and specificity and wide dynamic range make it the method of choice for quantitative mRNA levels. However, the purity of RNA samples and the co-extraction of DNA with RNA can interfere with the quantification of mRNA. For the purpose of this study, the optical density ratio (OD_{260}/OD_{280}) and the difference between Cp

$(RT\text{-}negative\ control\ samples)$ and $Cp_{(RT\ sample)}$ have been preformed.

Experimental design and data analysis from real-time, quantitative PCR experiments may be achieved using either relative or absolute quantification. When designing quantitative gene expression studies using real-time PCR, the first question that an investigator should ask is how the data should be presented. If absolute copy number is requirement, then the absolute method should be used. Otherwise, presentation of the relative gene expression should suffice. Relative quantification may be easier to perform then the absolute method because copy number is not required. The selection of valid reference genes is a prerequisite for relative quantification. However, the expression levels of the reference genes did not remain constant under different metabolic conditions or among treatments in the same sample or different growth stages. Although relative quantification in eukaryote have already published, the validation of reference genes for relative quantitative gene expression during bacterial growth phase still suffer lack of precision analysis as well as the systematic comparison. In the present study, systematic evaluation of reference gene and the effect using different strategies on target gene(s) in *Pseudomonas putida* mt-2 during degrading *p*-xylene were reported.

This thesis consists of five chapters. **Chapter 1** describes the general introduction with background, role of bioremediation, basic principle of qRT-PCR, and normalization knowledge. **Chapter 2** investigates RNA sample extracted from collection bacteria during degrading *p*-xylene, the analysis the co-extraction of DNA with RNA, and the confirmation of primers specificity. In **Chapter 3,** a panel of eight candidate reference genes in *Pseudomonas putida* mt-2 estimated by dedicated

validation program (geNorm) is presented. Based on the results of the present study, the expression levels of reference genes still vary under the growth of bacteria. There is not apparent correlation between gene expression stability and gene expression, and the normalization factors (NF_4) can be calculated by geNorm method. **Chapter 4** describes the difference between normalization factor (NF_4) and unstable gene as well as a single stable reference gene, and the influence of different approaches on interesting genes. **Chapter 5** is the general conclusion obtained from each chapter.

Chapter 1

General Introduction

1.1. Introduction

There is a major pollution problem due to organic pollutants all around the world. It has seriously influenced the lives of millions of people and caused many deaths and various healthy deranges (1). Pollution is contamination of a chemical or other agent, which makes part of the environment unstable for intended or desire use. Since toxic chemicals have been released into the environment throughout human life, environment protection and environment remediation play an important role in our life.

Pseudomonas putida mt-2 contains the most extensively characterized catabolic plasmid, which encodes enzymes for the minerlization of toluene, *m*-and *p*-xylene, *m*-ethyltoluene, and 1,3,4-trimethlbenzene (2, 3). Because the catabolic plasmid in the bacteria is considered as the best-characterized example of plasmids, its genetic information has become to a fundamental understanding of molecular biology of catabolic plasmids.

Testing and monitoring of bacterial gene expression have become integral aspects of microbial ecology, bioremediation, and diversity monitoring in environment. The development of molecular biological methods such as PCR (polymerase chain reaction), RT-PCR (reverse transcription-polymerase chain reaction) has become new techniques for identification of microorganisms. The quantification of bacterial activity in the environment is an important topic in microbial ecology and particularly in

bioremediation. Although the first report of real-time PCR was found in 1993, this technology has reached the mainstream, and been receiving importance and employed for the quantification of microorganisms (4). Because real-time PCR provides the most sensitive and flexible method for the detecting expression of single or multiple genes, it has increasingly been dealt with various other strategies for absolute and relative quantification.

In gene expression quantification studies, there are two methods, absolute quantification and relative quantification. Absolute quantification determines the exact copy concentration of target gene by relating the Cp value to a standard curve. On the other hand, there are differences in quality and quantity caused by variations in initial sample amount, possible RNA degradation of sample material, variations in cDNA synthesis efficiency. Therefore, relative quantification is developed and frequently used to estimate the quantitative real-time RT-PCR data in order to avoid these differences. Instead of working with absolute concentration based on a standard curve, target gene expression is related to a stable expressed reference gene simultaneously determined in the same sample. Although quantitative RT-PCR shows that performing is accurate and reproducible, and that convenient gene expression profiling experiment are optimal, there are a number of problems and limitations included in this method. One of the most prominent of these is the choice of one or more adequate control gene(s) to normalize expression results. Consequently, the necessity to identify the valid reference gene and

choose an optimal set of reference genes for each experimental system has been realized.

1.2. Description of Bioremediation

1.2.1. What is Bioremediation?

One of conventional methods for groundwater purge depends on pumping water to the surface and treating it there (5). As they require the withdrawal of large volumes of water to flush contaminants from aquifer solids, these pump–and–treat processes are very slow, and they may leave behind reservoirs of contaminants, which are lighter or denser than water and/or have low solubility. Additional, pump-and-treat methods do not destroy pollutants, but simply bring them to the surface for treatment or disposal elsewhere.

Another conventional method for soil cleanup requires that the contaminated soil is dug up and either incinerated or buried at a secure landfill (6). Soil excavation and incineration may increase the exposure to contaminants for both the workers at the site and nearby residents. Furthermore, excavation and final disposal are extremely costly. Comparing with the treatment of the soil in place, bioremediation reduces both the exposure risk and the cleanup cost.

The most important principle of bioremediation is that microorganisms (mainly bacteria) can be used to destroy hazardous pollutants or transform them to less their harmful forms. The microorganisms act against the pollutants only when they have access to a variety of materials – compounds to help them produce energy and nutrients to generate more cells. Because in a few cases the natural conditions at the contaminated site provide all the essential materials in large enough quantities, bioremediation can occur without human intervention – a process called engineered bioremediation. More often, bioremediation needs the construction of engineered systems in order to supply microbial–stimulating materials, a process also called engineered bioremediation.

The aim in bioremediation is to stimulate microorganisms with nutrients and other chemicals that will enable them to destroy the contaminants native to the contaminated sites, encouraging them to work by supplying them with the optimum levels of nutrients and other chemicals essential for their metabolism. As a result, bioremediation systems are limited by the activities of the native microorganisms.

1.2.2. Bacteria –*Pseudomonas putida* mt-2

Strains of *Pseudomonas* species capable of growth on the aromatic hydrocarbons and *m*- and *p*-xylene can be easily isolated from soil by selective enrichment (7).

Pseudomonas arvilla mt-2, a strain of *Pseudomonas* species, was later renamed *Pseudomonas putida* (arvilla) mt-2 and is usually referred to as *Pseudomonas putida* mt-2. *Pseudomonas putida* mt-2 contains TOL plasmid, which encodes the enzymes that degrade toluene. At present, the best-understand catabolic plasmid is TOL plasmid. The archetype TOL plasmid was first described in 1974 by Williams and Murray (8). Because it is the best-characterized example of a TOL plasmid, a detail review of its catabolic enzymes and genetic structure is provided to develop a fundamental understanding of molecular biology of catabolic plasmids. The plasmid is very large, about 117 kilobase pairs in size, approximately 40 kilobase pairs of which is needed for the catabolic pathway region and regulatory genes (9). The TOL plasmid has been shown to confer on the capacity to degrade not only toluene but also *m*- and *p*-xylene and other benzene derivative. The *xyl* genes of TOL plasmid are organized into two operons referred to as the upper operon and lower operon. Two operons are separated by several thousands base pairs. These genes encoding catabolic enzymes have been named the *xyl* genes. The upper operon, *xylCAB*, encodes the degradation of toluene and xylenes to benzoate and toluenes. The lower operon, *xylDLEFJK*, encodes the degradation of benzoate and toluenes to pyruvate and propanal. These genes, xyl*CAB*, xyl*DZEFJK* and their enzymes are shown in Table 1-1.

Table 1-1: *xyl* proteins encoded by two operons of the TOL plasmid

Gene	Protein
xylA	Xylene oxygenase
xylB	Benzyl alcohol dehydrogenase
xylC	Benzaldehyde dehydrogenase
xylD	Benzoate 1,2-dioxygenase
xylL	4-Methylcyclohexa-3,5-diene-1,2-cis-diol-1-carboxylic acid dehydrogenase
xylE	Catechol 2,3-dioxygenase
xylF	2-Hydroxymuconic semialdehyde hydrolase
xylJ	2-Oxopent-4-enoate hydratase
xylK	4-Hydroxy-2-oxoyalerate aldolase

1.2.3. Biodegradation of Aromatic Hydrocarbons

There are many aromatic hydrocarbons that are important components of petroleum and its refined products. Aromatic compounds are broadly defined as benzene and other compounds that exhibit similar chemical behavior (10). Benzene and substituted benzenes compose the naturally occurring aromatic hydrocarbons. However,

among the most important aromatic petroleum hydrocarbons are benzene, ethylbenzene, toluene (methylbenzene), xylenes (dimethylbenzenes), and the polycyclic aromatic hydrocarbons, of which naphthalene is the simplest representative.

The differences between aromatic hydrocarbons and aliphatic ones are composing and provide a useful method for classifying these compounds. However, the similarities also occur with respect to their biodegradability. For example, although anaerobic biodegradation of aromatic hydrocarbons has been reported (11, 12), it is rare and slow comparing to aerobic biodegradation. As is the case with aliphatic hydrocarbons, aerobic biodegradation of aromatic hydrocarbons involves molecular oxygen as a direct reactant and as the terminal electron acceptor. Finally, many aromatic hydrocarbons can support the growth of bacteria as the sole source of carbon and energy. Although aromatic hydrocarbons are not as biodegradable as normal alkanes and branched alkanes, they are somewhat more easily degradable than the alicyclic hydrocarbons (13, 14).

1.2.4. Biodegradation of *p*-xylene

Xylene has three isomers: *m*-xylene, *o*-xylene and *p*-xylene and contains benzene ring. In mineral oil-contaminated soils and aquifers, *p*-xylene is often of major concern because of its high water solubility and toxicity. Because all of the important aromatic

hydrocarbon that occur in petroleum are derivatives of benzene, a reaction is cleavage of the benzene ring and is common to all pathway that lead to mineralization of aromatic substrates. Under aerobic conditions, monoaromatic hydrocarbons are rapidly mineralized. Oxidation of *p*-xylene provides an example of degrading pathway by which alkyl-benzenes are degraded. Many bacteria that can grow on *p*-xylene as the sole source of carbon and energy first oxidize a methyl group, producing the corresponding toluic acid, and then oxidize the aromatic ring. Molecular oxygen serves as a reactant to insert into the carbon atom bearing the carboxyl group, and decarboxylation accompanies dehydrogenation of the diol to methylcatechol for *p*-xylene catabolism(15). Enzymes can catalyze such reaction in the degrading *p*-xylene. The biodegradation pathway for *p*-xylene using *Pseudomonas putida* mt-2 is shown in Fig.1-1. The relationship between all enzymes and *xyl* genes in TOL plasmid are shown in Table 1-1.

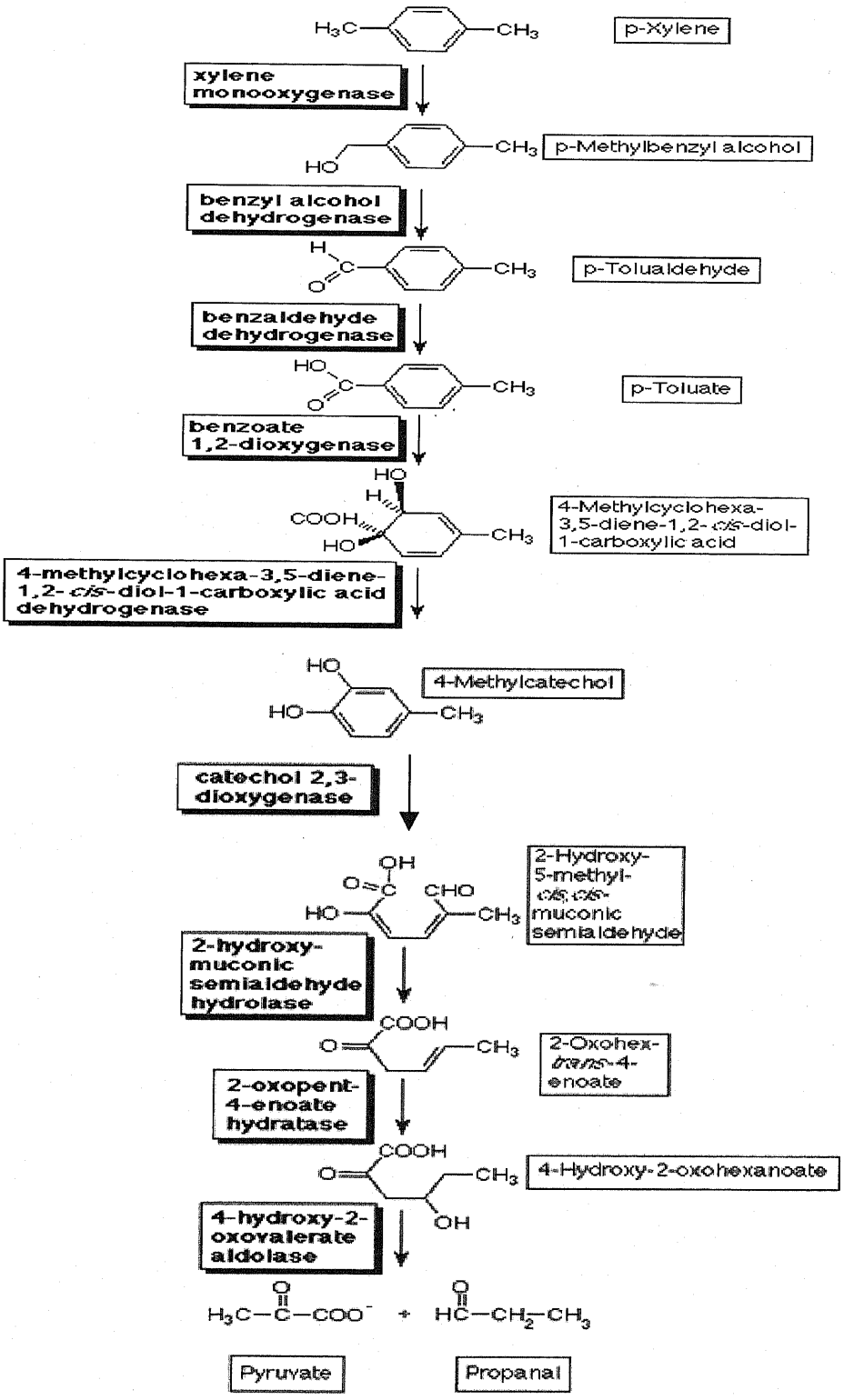


Fig.1-1. *p*-xylene graphical pathway map (http://umbbd.msi.umn.edu/pxy/pxy_image_map.html)

1.2.5. Gene Expression

Gene is a unit of genetic information and gene expression is the process by which gene products are made (16). Not only bacteria existence, but also gene expression of bacteria should be researched for understanding of the behavior and function of bacteria in different environment. Individual genes can be 'switched on' or 'switched off' depending on the requirements and circumstances of the bacteria at a particular time. A number of mechanisms are considered to be responsible for the control of gene expression. Bacteria gene expression in the environment is studied on the level of proteins by detecting specific enzyme (17, 18) or by analyzing the level of mRNA by direct analysis of gene transcript (19, 20).

For mRNA analysis, the transcription of specific bacteria gene can be studied. The quantitative evaluation of catabolic genes in the bacteria that are responsible for the degradation of pollutants is one of the monitoring environments. Quantitative gene expression is useful for investigation of microbial process, such as, bioremediation and the role of microbial function in contaminated environment.

1.3. Quantification of Nucleic Acids by PCR

1.3.1. Quantitative PCR Characteristics

Quantitative PCR means the estimation of amount of DNA or RNA sample using PCR amplification. The amplification reaction is described as at first stochastic ("lag phase"), then exponential ("exponential phase") and finally stagnant ("plateau phase") (21). For most purposes, the lag phase is related primarily to the threshold sensitivity and variability of the particular method used to measure the amplified product. As long as enzyme reaction kinetics are driven by the concentration of primer and dNTP substrate, both of which are present in vast excess, the influence of the concentration of target template on the kinetics of amplification should be insignificant. In the exponential phase the accumulation of product is calculated by the formula $y=N(1+E)^{n}$, where y is the amplification factor, N is the number of input target molecules, E is the amplification efficiency, n is the number of amplification cycles. In practice, the truly exponential amplification is limited to only very few cycles among a typical 40 cycles amplification run and this formula may be accurate only for those few cycles. Subsequently the amplification slows down to constant amplification rates and eventually it reaches the plateau phase. Here, its effects are difficult to foretell or standardize and amplification is affected by limitations of substrate and inhibition of

enzyme. Overall amplification efficiency will also vary with the presence of RT and / or DNA polymerase inhibitions and may be affected by position of the sample in the thermal cycle.

1.3.2. Conventional Quantitative PCR

The words "quantitative" and "PCR" were deemed a contravention, with quantitative PCR an aspiration rather than reality for no long time (22). Conventional PCR is a qualitative estimation answering yes/no questions and variations in reaction components. Thermal cycling conditions and mispriming events during the early stage of PCR will vastly influence the yield of the amplified product. Therefore, conventional quantitative PCR demands extensive validation and tedious controls. At least, it requires improved data recording methods, which are less subjective than band densitometry.

The basic attribute of the PCR, cyclic priming and enzymatic replication of target sequences, not only results in its exquisite sensitivity, but also presents substantial problems when used for quantitative purposes. There are many experimental variables that become exaggerated by the exponential amplification. Quantitative results are affected by mispriming, differences in the kinetics of amplification, variation in reagent purity and reaction composition, and, perhaps most significantly, protocol and operator variability (23). Therefore, it is difficult to attain a consistent relationship between the

amount of starting template and absolute amount of amplified product.

Conventional PCR technique is mainly based on the end point analysis, which is different between samples. As a result, quantification based on endpoint analysis shows poor precision, low sensitivity and time consuming due to the post PCR analysis.

1.3.3. Real-Time Quantitative PCR

Of course, there are the real-time fluorescence-based PCR and RT-PCR assays that investigate the impression of generating actual quantitative data. Both methods are relied on the contention that there is a quantitative relationship between the amount of target gene present at the start of a PCR assay and the amount of product amplified during its exponential phase.

Real-time PCR is based on continuous monitoring of the DNA amplification and since the DNA amplification can be monitored in each and every cycle, so it is a real-time measurement. Figure 1-2 shows a representative amplification plot and defines the term used in the quantification analysis. An amplification plot is the plot of fluorescence signal versus cycle number. In the initial cycles of PCR, there is little change in fluorescence signal. This defines the baseline for the amplification plot.

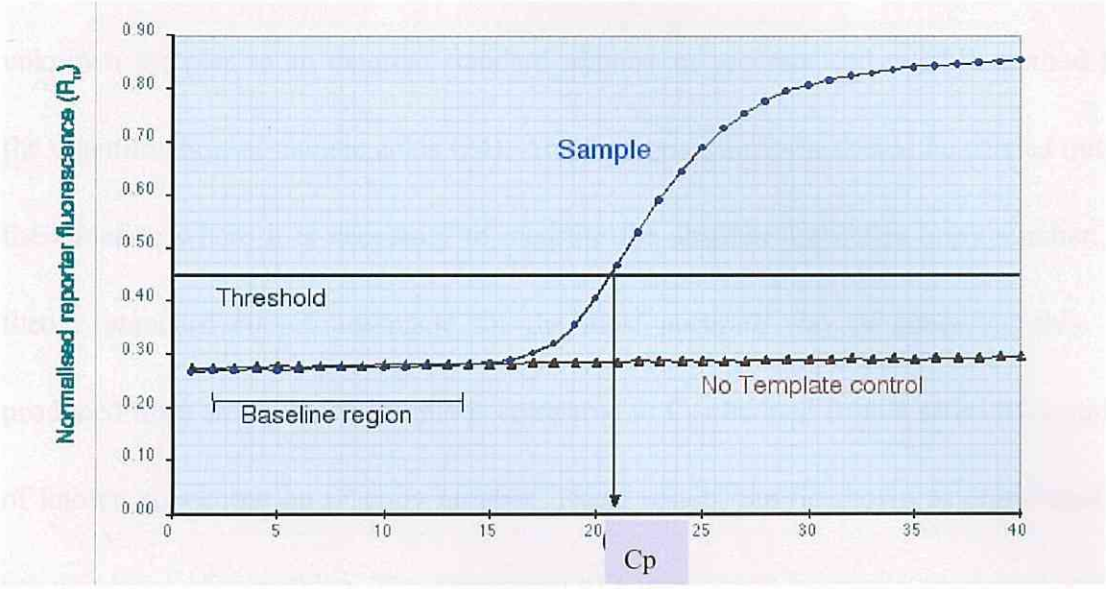


Fig. 1-2. Model of a signal amplification plot for real-time PCR

(http://cgr.otago.ac.nz/slides/taqman/sld012.htm)

An increase in fluorescence above the baseline indicates the detection of accumulated PCR product. A fixed fluorescence threshold can be set above the baseline. The parameter C_p (crossing point) is defined as the fractional signal point at which the fluorescence passes the fixed baseline. To determine starting copy number, the amount of target in unknown samples is accomplished by measuring C_p and using the standard curve. There are two methods of quantification: absolute and relative.

1.4. Quantitative Methods for Gene Expression

1.4.1. Absolute Quantification

Absolute quantification provides an analysis method where the comparison of unknown samples to an external standard obtains an accurate and reliable method for the quantification of nucleic acids (24). Absolute quantification should be carried out in the situation where it is necessary to measure the absolute transcript copy number. In theory, standard curves are made for the most accurate way of achieving this. C_p produced from an unknown sample is compared to C_p obtained from a series of samples of known concentration or copy number. These results can be shown as copy number per unit mass of something. The expression of a target gene is usually compared across many samples, often from different individuals, and sometimes from different tissues. Theoretically, the amount of product from each reaction is proportional to the initial amount of target in each sample. However, since small differences in sample input or reaction conditions can lead to large differences in PCR product yield, the amount of starting material must be quantified with rigorous accuracy to normalize sample data and correct for tube-to-tube differences. Therefore, if absolute quantification is to be accurate, it needs to take that variability into account.

1.4.2. Relative Quantification

The relative quantification changes in gene expression using real-time PCR requires certain equation, assumptions, and the testing of these assumption to properly

analyze the data. Relative quantification is used to compare the changes in steady-state mRNA levels of two or more genes with each other, with one of them as an endogenous reference. So, the number of target gene copies is normalized to the reference gene, for example, a suitable reference gene and then all samples are expressed as an n-fold difference relative to that mRNA. Theoretically, relative quantification should be superior to and far more convenient than absolute quantification. This is because the result is a ratio, hence, RNA concentration is irrelevant. The underlying justification for the relative quantification method is as follows:

1. Based on the amplification equation of PCR for target gene and normalizer, the amount of amplification molecules at the threshold cycle is given by:

$$X_T = T_0 (1+E_T)^{Cp,t}$$

Where X_T is the number of target copies at the threshold cycle of the reaction, T_0 is the initial number of target copies, E_T is the efficiency of target amplification, $C_{p,t}$ indicates the threshold cycle for target amplification.

2. A similar equation for the endogenous reference reaction is:

$$X_R = R_0 (1+E_R)^{Cp,r}$$

Where X_R is the number of reference gene copies at the threshold cycle, R_0 is the initial number of reference gene copies, E_R is the efficiency of reference gene amplification, $C_{p,r}$ is the threshold cycle for reference gene amplification.

3. Dividing X_T by X_R gives the ratio of target gene copies (T) to reference gene copies (R) at the threshold cycle and normalizes the expression of target gene:

$$T_0/R_0\ (1+E)^{(Cp,t-Cp,r)} = K,$$

This equation is based on the precondition that the efficiencies of target gene and reference gene amplification are approximately equal. Where $C_{p,t}$-$C_{p,r}$ is the difference in threshold cycles for target gene and reference gene, and K is a constant. T_0/R_0 is the normalized amount of target and the equation can be rearranged to

$$T_0/R_0 = K/(1+E)^{(Cp,t-Cp,r)} \quad \text{or} \quad T_0/R_0 = K/(1+E)^{\Delta Cp},$$

Comparing serial dilutions of target gene and reference gene simultaneously, the plot of log input amount versus ΔC_p (C_p target gene – C_p reference gene) has a slope of approximately 0. In practice, anything less than 0.1 is deemed to be acceptable.

K is dependent on (1) the reporter dye used with the probe or primer, (2) sequence context effects on the fluorescence properties of the probe or primer, (3) purity of probe or primer, and (4) efficiency of probe or primer cleavage; the exact value of K need not

be equal to 1. However, influencing parameters are assumed to vary only negligibly among single samples so that K is assumed to be equal and thus does not affect the comparison of calculated relative ratios. The efficiency of PCR provides the information about the amplification rate and varies from 0 to 1. The rate equal 1 (=100%) means that in each cycle the number of copies is doubled. The efficiency can be calculated from the slope of a standard curve:

$$E = 10^{-1/s} - 1 \quad ,$$

Where E is the run efficiency and s is the slope of generated from standard curve.

There are two main problems with the relative quantification to an internal control, which is rRNA or a reference gene approach:

This approach tends to introduce a significant statistical bias that results in misleading biological explanation (25). This is particularly true when there are vast differences in the expression levels of target and normalizer or when the target gene is expressed at very low levels. In this case the relationship between target gene and reference gene levels may not regress to a zero intercept. This is important, because the relationship between the two factors may not be linear at very low target gene copy numbers. Interestingly, this problem with the ratio method has been described before, albeit in another context (26), but seems to have been forgotten.

It is difficult to find a reference gene whose expression is constant and against which the target gene copy numbers can be normalized during the experimental conditions. However, under certain circumstances, if it can be shown that experimental treatment does not affect the reference gene chosen as the normalizer, relative quantification can be useful. There are several mathematical models that calculate relative expression ratios, some of which correcting for differences in amplification efficiency and some not.

1.4.2.1. Reference Gene (Housekeeping Gene)

Bacterial genomes usually contain several thousand different genes. Some of the gene products are required by the cell under all growth conditions and these genes are called housekeeking genes. They are always turn on and are believed to be constitutively expressed and minimally regulated. These genes include the genes that encode such proteins as DNA polymerase, RNA polymerase, and DNA gyrase. Many other gene products are required under specific growth conditions. They include enzymes that synthesize amino acids, break down specific sugars, or respond to a specific environmental condition such as DNA damage. Housekeeping genes have been used widely as reference genes for relative quantitative RT-PCR analysis. Reference genes must be expressed at some level all of the time. Frequently, as the cell grows

faster, more of the reference gene products are essential. Even if the cell grows very slow growth, some of each reference gene product is made. When choosing a housekeeping gene as a reference for relative quantification, one must identify a gene whose expression level remains relatively constant for a certain experimental set-up. In fact, it is usually necessary to test a panel of housekeeping genes experimentally to find one that is not regulated in the investigated system. Since choosing an appropriate reference gene is critical for accurate quantitative RNA analysis, the behavior of candidate genes in different cell types and cell metabolic stages should be carefully examined.

1.4.2.2. Normalization

Data normalization in real-time RT-PCR is a further significant marker in gene quantitative analysis. Data normalization, while a vital aspect of experimental design (27), remains a real problem for absolute quantification (28) and it is impossible to provide general recommendation about the most appropriate procedure. The reason is different experimental setups, targets, and samples sources are so divergent, that no single set of rules or even recommendations can be correct for every one. RT-PCR specific errors in the quantification of mRNA transcripts are easily mixed by any variation in the amount of starting material between samples. The ideal internal standard

should be expressed at a constant level amount different cell, at all stages of development, and should be unaffected by the experimental treatment. In addition, an endogenous control should also be expressed at roughly the same level as the RNA under study. Because any one single RNA with a constant expression level dose not exist in all of these situations (29), various reference genes, rRNA, and total RNA are most commonly used to normalize gene expression patterns.

A recently systematic analysis and comparison of their usefulness has concluded that a single reference gene should not be used for normalization. (30). The recent demonstration of the effectiveness of normalization against the geometric mean of multiple carefully selected reference gene is interesting (31).

Figure 1-7 describes the data analysis before and after normalization is different (35).The amount of assayed mRNA may fluctuate before normalization is shown in Fig. 1-7 A. The reason is due to differences in variations in initial sample amount; RNA degradation; variations in sample loading/pipetting errors; variations in cDNA synthesis efficiency. Figure1-7 B shows that the amount of mRNA decreased fluctuation after normalization to one reference gene. Figure1-7 C shows that the amount of mRNA didn't fluctuate after normalization to more than one reference gene. The purpose of normalization is to remove the sampling differences. However, the main problem in relative quantification of mRNA expression analysis is selection of an appropriate control gene.

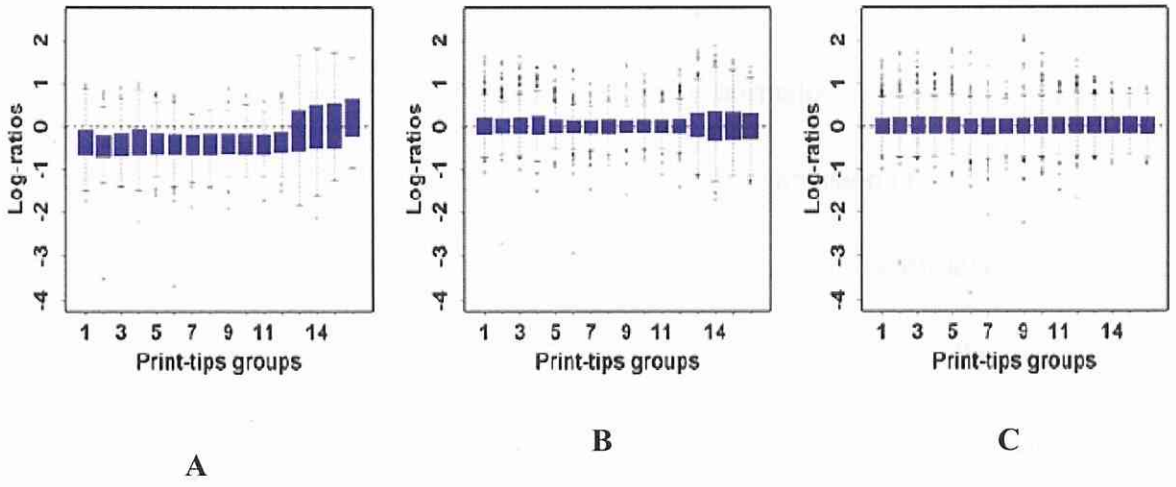


Fig.1-7: Data analysis before (A) and after normalization to one reference gene (B) and after normalization to more than one reference genes (C) (35).

1.4.2.3. GeNorm Program

Several variables need to be controlled for in gene-expression analysis, such as the amount of starting material, enzymatic efficiencies, and differences between cells in overall transcriptional activity. Various strategies have been applied to normalize these variations.

To date, internal genes are most frequently used to normalize the mRNA fraction. This internal control – often referred to as a reference gene – should not vary in the cells under investigation, or in response to experimental treatment. With the increased

sensitivity, reproducibility and large dynamic range of real-time RT-PCR methods, the requirements for a proper internal control gene have become increasingly stringent. Thus, expression levels of target gene are usually normalized to expression levels of internal control genes that are supposed to show stable expression in cell of interest.

GeNorm method (34) relies on the principle that the expression level ratio of two ideal internal control genes is identical in all samples, regardless of the experimental condition or cell type. In this way, variation of the expression ratios of two reference genes reflects the fact one (or both) of the genes is (are) not constantly expression, with increasing variation in ratio corresponding to decreasing expression stability. For every control gene, the method determined the pairwise variation with all other control genes as the standard deviation of logarithmically transformed expression ratios, and defined the internal control gene-stability measure, M, as the average pairwise variation of a particular gene with all other control genes. Calculation of the control gene stability M was performed according to Vandesompele (34). In brief, real-time RT-PCR expression levels $\alpha_{i,j}$ of n internal control genes are determined in m samples. An array $A_{j,k}$ of m elements is calculated for every combination of two internal control genes j and k, consisting of log2-transformed expression ratios $\alpha_{i,j}/\alpha_{i,k}$ (Equation 1). The pairwise variation $V_{j,k}$ for the control genes j and k represents the SD of $A_{j,k}$ elements (Equation 2). The gene stability measure M_j for control gene j is the arithmetic mean of all pairwise variations $V_{j,k}$ (Equation 3).

$(j \neq k)$

$$A_{j,k} = \left\{\log_2\left(\frac{a_{1,j}}{a_{1,k}}\right), \log_2\left(\frac{a_{2,j}}{a_{2,k}}\right), \log_2\left(\frac{a_{3,j}}{a_{3,k}}\right).....\log_2\left(\frac{\alpha_{i,j}}{\alpha_{i,k}}\right)\right\} \quad (1)$$

(here, i=1,2,3.....m)

$$V_{j,k} = st.dev.\left(A_{j,k}\right) \quad (2)$$

$$M_j = \frac{\sum_{k=1}^{n} V_{j,k}}{n-1} \quad (3)$$

Genes with the lowest M values have the most stable expression. Assuming that the control genes are not co-regulated, stepwise exclusion of gene with the highest M value results in a combination of two constitutively expression reference genes that have the most stable expression in tested samples.

In order to measure expression levels accurately, normalization by multiple reference genes instead of one is required. Normalization factor (NFn) were determined for each sample by calculating the geometrical mean of expression levels of the performing control genes (n).

For i sample, normalization factor (NFn,i) was calculated as following equation(4),

$$NF_{n,i} = \sqrt[n]{a_{i,1} \times a_{i,2} \times a_{i,3} \ldots\ldots a_{i,n}} \quad (4)$$

$$P = \left\{ \log_2 \left(\frac{NF_n}{NF_{n+1}} \right)_1, \log_2 \left(\frac{NF_n}{NF_{n+1}} \right)_2, \ldots . \log_2 \left(\frac{NF_n}{NF_{N+1}} \right)_i \right\} \quad (5)$$

(here, i=1,2,3.....m,)

$$V_{n/n+1} = st.dev.(P) \quad (6)$$

It is obvious that an accurate normalization factor should not include the rather unstable genes that were observed in samples. Additionally, the geNorm software made a pairwise variation $V_{n/n+1}$ that was calculated between the two sequential normalization factors (NF_n and NF_{n+1}) for all samples within the same cell to determine the optimal number of genes necessary (equation 6). A large value means that the added gene has a significant effect and should be included for calculation of the normalization factor.

1.5. Objectives

1.5.1. Evaluation of Reference Genes Stability during Biodegradation

In order to quickly purify polluted groundwater and toxic waste sites, the

bioremediation needs that bacteria remain higher active to degrade pollutants. It is very important to quantify gene expression in bioremediation. The two most commonly used methods to study data from real-time quantitative PCR experiments are absolute quantification and relative quantification. Relative quantification relates the PCR signal of target gene transcript and reference gene transcript. Although the growth of bacteria is a physiological process, target gene transcript and reference gene transcript would vary. Therefore, the reference genes expression need to be analyze under investigation condition.

Use of the real-time polymerase chain reaction (PCR) to amplify cDNA products reverse transcribed from mRNA is on the way to become a routine tool in molecular biology to analyze low amount gene expression. Real-time PCR is easy to perform, provides the necessary accuracy and produces reliable as well as rapid quantification results. Reference genes are often used for the relative quantification of target genes in gene expression studies in eukaryote. However, the stability of reference genes for relative quantitative gene expression in the bioremediation has not been described enough detail.

In the present study, pure culture was performed in *Pseudomonas putida* mt-2 in the presence of *p*-xylene. We used eight reference genes (*rpoN*, *rpoD*, *dbhA*, *phaF*, *16S rRNA, gst, lexA, and atkA*) during degradation of *p*-xylene for the systematic comparison of the stability among reference genes explained detail in chapter 3. We

also analyzed the relationship between gene expression and stability of reference gene.

1.5.2. Effect of Different Strategies on Target Genes Expression Level

Although real-time RT-PCR is widely used to quantitative biologically relevant changes in mRNA levels, remaining a number of problems associated with its use. These include the inherent variability of RNA, variability of protocols that may copurify inhibitors, and different reverse transcription and PCR efficiencies. Consequently, it is important that an accurate method of normalization is chosen to control for this error. Unfortunately, universal and idea reference gene dose not appear in all experiments. Therefore, a suitable and selective reference genes require to be studied in bioremediation. The aim of the reference control gene is to normalize the PCR data for the amount of RNA added to the reverse transcript reaction. Normalization remains one of real-time PCR most difficult problem. The validation of reference gene is very important for relative quantification of target gene(s). In this study, RT and real-time PCR were performed at optimal condition. We identified and selected a set of suitable reference genes. Different normalization strategies have been investigated for normalizing real-time PCR data. We illustrated the valid reference genes using two target genes, *xylA* and *xylE*.

REFERENCES

1. Reardon, K. F., Mosteller, D. C., and Bull Rogers J. D.: Biodegradation kinetics of benzene, toluene, and phenol as single and mixed substrates for *Pseudomonas putida* F1. Biotechnol. Bioeng. 2000, 69, 4: 385-400.

2. Kunz, D. A., and P. J. Chapman. Catabolism of pseudocumene and 3-ethyltoluene by Pseudomonas putida (arvilla) mt-2: evidence for new functions of the TOL (pWW0) plasmid. J. Bacteriol., 1981, 146: 179-191.

3. Worsey, M. J. and Williams, P. A.: Metabolism of toluene and xylenes by *pseudomonas putida* (*arvilla*) mt-2: evidence for a new function of TOL plasmid. J. Bacteriol., 1975, 124: 7-13

4. Ginzinger, D. G.: Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. Exp. Hematol. 2002, 30: 503-512.

5. Travis, C.C., and Doty, C.B.: Can contaminated aquifers at Superfund sites be remediated?: Environ. Sci. and Technol., 1990, 24:1464-1466.

6. Qifei H, Qi W, Lu D, Beidou X. and Binyan Z.: The current situation of solid waste management in China. J Mater Cycles Waste Manag. 2006, 8: 63–69.

7. Worsey, J. M. Franklin, F. C. H. and Williams, P. A.: Regulation of the degradative pathway enzymes coded for by the TOL plasmid (pWW0) from *Pseudomonas putida* mt-2. J. Bacteriol. 1978, 134: 757-764.

8. Williams, P. A., and K. Murray. Metabolism of benzoate and the methylbenzoates

by Pseudomonas putida (arvilla) mt-2: evidence for the existence of a TOL plasmid. J. Bacteriol., 1974, 120: 416-423

9. Burlage, R. S., Hooper, S. W. and Sayler, G. S.: MINIREVIEW: The TOL (pWW0) catabolic plasmid. Appl. Environ Microbiol., 1989, 55, 1323-1328.

10. Rittmann, B. E., Seagren, E, and Wrenn, B. A.: in situ bioremediation. 2nd. Ed New Jersey, USA.1992, p31.

11. Grbic-Galic, D., and Vogel, T.M.: Transformation of toluene and benzene by mixed methanogenic cultures. Appl. Environ. Microbiol. 1987, 53: 254-260.

12. Zeyer, J., Kuhn, E. P. and Schwarzenbach, R. P.: Rapid microbial mineralization of toluene and 1,3-dimethylbenzene in the absence of molecular oxygen. Appl. Environ. Microbiol. 1986, 52: 944-947.

13. Perry, J. J.: Microbial metabolism of cyclic alkanes, In R.M. Atlas (ed.), Petroleum Microbiology. Macmillan Publishing Co., New York. 1984, p. 61-97.

14. Leahy, J. G. and Colwell, R. R.: Microbial degradation of hydrocarbons in the environment. Microbiol. Rev. 1990, 54: 305-315.

15. Ribbons, D. W. and Eaton, R. W.: Chemical transformations of aromatic hydrocarbons that support the growth of microorganisms, In A. M. Chakrabarty(ed.), Biodegradation and Detoxification of Environmental Pollutants. CRC Press, Inc., Boca Raton, FL. 1982, p. 59-84.

16. Robert F. W.: molecular biology, the 2nd Ed. McGraw-Hill Companies, Inc. 2002,

p826.

17. Heitzer, A., Webb, O. F., Thonnard, J. E. and Sayler, G. S. Specific and Quantitative Assessment of Naphthalene and Salicylate Bioavailability by Using a Bioluminescent Catabolic Reporter Bacterium Appl. Environ. Microbiol. 1992, 58: 1839—1846.

18. King, R. J., Short, K. and Seidler, R. J.: Assay for detection and enumeration of genetically engineered microorganisms which is based on the activity of a deregulated 2,4-dichlorophenoxyacetate monooxygenase. Appl. Environ. Microbiol., 1991, 57,6: 1790-1792.

19. Matin, A.: Molecular analysis of the starvation stress in *Escherichia coli.* FEMS Microbiol. Ecol., 1990, 74: 185-196.

20. Jeffrey, W. H., Nazaret, S. and Von Haven, R. Improved method for the recovery of mRNA from aquatic samples: application to detecting mer gene expression. Appl. Environ. Microbiol. 1994, 60: 1814-1821.

21. Wittwer, C. T., Herrmann, M. G., Moss, A. A. and Rasmussen, R. P.: Continuous fluorescence monitoring of rapid cycle DNA amplification. Biotechniques, 1997, 22: 130-138.

22. Ferre, F.: Quantitative or semi-quantitative PCR: reality versus myth. PCR methods. Genome Res. 1992, 2: 1-9.

23. Bustin, S. A.: Quantification of mRNA using real-time reverse transcription PCR

(RT-PCR): trends and problems. J Mol. Endocrinol., 2002, 29: 23-39.

24. Ke, L. D., Chen, Z. and Yung, W. K.: A reliability test of standard-based quantitative PCR: exogenous vs endogenous standards. Mol cell probes. 2000, 14: 127-135.

25. Hocquette, J. F. and Brandstetter, A. M., Common practice in molecular biology may introduce statistical bias and misleading biological interpretation. J. Nutr Biochem. 2002,13: 370-377.

26. Poehman, E. T. and Toth, M. J.: Mathematical ratios lead to spurious conclusions regarding age- and sex- related differences in resting metabolic rate. Ann J Clin Nutr 1995, 61: 482-485.

27. Karge, W. H., Schaefer, E. J. and Ordovas, J. M.: Quantification of mRNA by polymerase chain reaction (PCR) using an internal standard and a nonradioactive detection method. Methods Mol boil. 1998, 110: 43-61.

28. Thellin, O., Zorzi, W., Lakaye, B., Borman, D. B., Coumans, B., Hennen, G., Grisar, T., Igout, A. and Heinen, E., Housepeeking genes as internal standards: Use and limits. J Biotechnol. 1999, 75: 291-295.

29. Haberhausen, G., Pinsl, J., Kuhn, CC. and Markert-Hahn, C.: Comparative study of different standardization concepts in quantitative competitive reverse transcription PCR assays. J Clin Microbiol. 1998, 36: 628-633.

30. Tricarco, C., Pinzani, P., Bianchi, S., Paglierani, M., Distante, V., Pazzagli, M,

Bustin, SA. and Orlando, C.: Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single reference genes is inappropriate for human tissue biopsies. Anal Biochem. 2002, 309: 293-300.

31. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy N., De paepe, A. and Speleman, F.: Accurate normalization of real-time geometric averaging of multiple internal control genes. Genome boil. 2002, 3: research0034.1-0034.11.

32. Yang, Y. H., Dudoit, S. and Luu, P.: Normalization for cDNA microarry data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res., 2002, 30, 4:e15.

Chapter 2

Investigation of RNA Quality and Confirmation of PCR Primers Specificity

Abstract

Real-time quantitative reverse-transcription polymerase-chain-reaction (qRT-PCR) reaction is the sensitive and reliable quantitative method that can be used for the measurement of low abundant mRNA gene expression. However, many factors may influence the sensitivity of procedure, such as, RNA quality and designed primers. The assessment of RNA quality is a important first step in obtaining meaningful gene expression data. Using high quality RNA is a key element for the successful application of modern molecular biological method. The complete synthesis of the primers is also precondition to get ideal quantitative gene expression result. They were checked by electrophoresis agarose gel and melting curve. Quantification of DNA contaminant was detected for a precaution against false positive results. In this study, our result showed the RNA samples extracted from culture using RNAspin Mini RNA Isolation kit (GE Healthcare UK Limited, Buckinghamshire, UK) were good and the designed primers were completely produced, and also the residual DNA level was regarded as negligible because the different between $C_{(p,\ RT\text{-}negative\ control\ samples)}$ and $C_{(p,\ RT\ sample)}$ was greater than 4.

2.1. Introduction

For quantifying mRNA levels, SYBR Green I-based real-time reverse transcription PCR (real-time RT-PCR) method is widely used (1). This technology has the advantage of sensitivity, speed, high throughput and high degree of potential automation compared to the compared to the conventional quantification methods, such as, Northern blot analysis or RNase protection assay. The accuracy of gene expression estimation is considered as influenced by the quality of RNA and primers. Purities of RNA and primers are critical elements for the overall success of RNA-based analyses. It is preferable to use high-quality RNA and primers as starting point in molecular biology. Therefore, any kind of array applications of the used total RNA should be checked. It is well known that RNA is sensitive to degradation by postmortem processes and inadequate sample handling or storage (2). The quality and quantity of purified RNA is variable after the extraction during long storage rather unstable (3). As a result, several steps during bacteria handling must be carefully controlled to preserve the quality of RNA samples.

Using a set of primers have proven to be a very accurate and reproducible tool for gene quantification (4) and thus this system has been selected for the quantification of catabolic genes. Oligonucleotide primers are generally synthesized in the range 18-30 bases, though it is possible to amplify low complexity DNA with shorter primers.

Primer sequences should have similar G+C content, minimal secondary structure and low complement to each other. DNA contamination in RNA samples due to co-extraction should be removed by DNase treatment for accurate quantitative mRNA. However, there is a possibility of presence of contaminating DNA in RNA samples even after DNase treatment, DNA contamination may interfere with the subsequent quantification of target genes.

In this study, our aim is to investigate RNA samples and designed primers quality to preserve meaning results. In addition, the differences between Cp $_{\text{(RT-negative control samples)}}$ and $Cp_{\text{(RT samples)}}$ were carried out to get an accurate estimation of gene expression.

2.2. Experimental Section

2.2.1. Materials for Culture

P. putida mt-2 (DSM 3931), was obtained from Japan Collection of Microorganisms RIKEN BioResource Center. A growth medium or culture medium is difined as a liquid or gelatinous substance containing nutrients in which microorganisms or tissues are cultivated for scientific purposes (5).

Two growth mediums, Nutrient Agar NO. 2 (6) and M9 minimal medium (7), were

used in the study as follow (Table 2-1 and Table 2-2, respectively):

Table 2-1: The component of nutrient agar No. 2

Nutrient Agar No. 2	
Peptone	10.0 g
Meat extract	10.0 g
NaCl	5.0 g
Distilled water	1.0 L
Adjust pH to 7.0-7.2	

Table 2-2: The component of M9 minimal medium

M9 MINIMAL MEDIUM	
5 × M9 salts	200.00 ml
A9 solution	20.00 ml
$MgSO_4$ (1 M)	1.00 ml
Ammonium iron citrate 6%	1.00
Sterile deionised H_2O to 1 liter,	Adjust pH to 7.4.

5 × M9 salts:	
$Na_2HPO_4 \times 7\ H_20$	64.0 g
KH_2PO_4	15.0 g
NaCl	2.5 g
NH_4Cl	5.0 g
Sterile deionised H_2O to 1 liter	
A9 Solution:	
HBO_3	0.30 g
$ZnCl_2$	0.05 g
$MnCl_2 \times 4\ H_20$	0.30 g
$CoCl_2$	0.20 g
$CuCl_2 \times \mathbf{2}\ H_20$	0.01 g
$NiCl_2 \times \mathbf{6}\ H_20$	0.02 g
$NaMO_4 \times \mathbf{2}\ H_20$	0.03 g
Desionsed H_20 to a final volume of 1 liter	

2.2.2. Growth of the Bacterial Strain

P. putida mt-2 was inoculated in nutrient agar NO. 2 (without agar) at 30°C on a rotary shaker at 110 rpm for 20 h and the cell pellets were collected and washed twice in

PBS (phosphate-buffered saline) and stored at –80°C until use.

P. putida mt-2 was put into the flask (500 ml) containing the M9 mineral medium (300 ml) with succinate (10%, 9ml) and was incubated at 30°C on a rotary shaker at 110 rpm for pre-culture. After sufficient growth (1.2 OD), the pellet was centrifuged and was washed by PBS. Washed pellet was transferred in M9 mineral medium in the presence of *p*-xylene vapor as the sole carbon source. Vials (50 ml) containing the mineral medium (12 ml) and the bacterial were used for degrading culture. A small amount of cotton with *p*-xylene (9.5 μl) was added to a tube, which was then placed in the vial (Fig. 2-1).

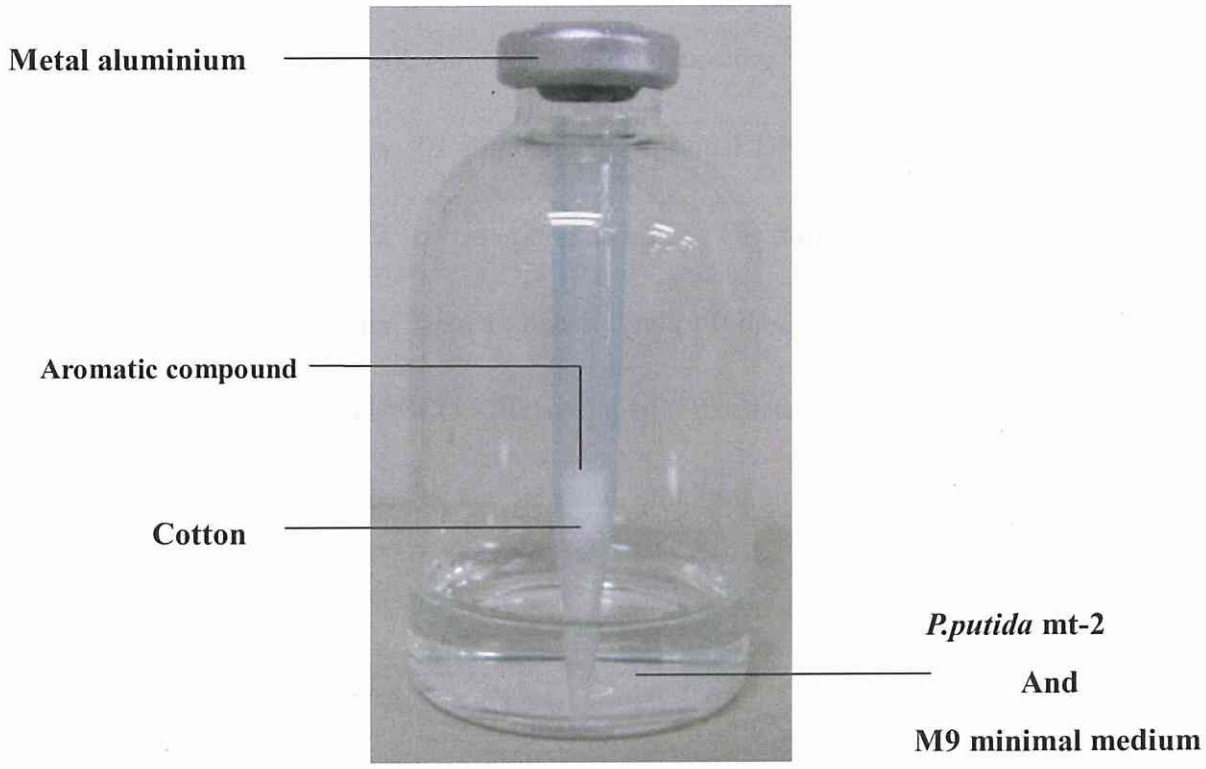


Fig.2-1. The vial used for the growth of *Pseudomonas putida mt-2* in the presence of aromatic.

Three replicate vials were incubated at 30°C on a rotary shaker at 110 rpm. Bacterial growth was monitored by determining the optical density at 600 nm (OD_{600}) and cells were obtained from 1.5-ml cell cultures by centrifugation at 4°C for 10 min at 6000 rpm for different time intervals. The cell pellets were washed twice in phosphate-buffered saline and stored at –80°C until use.

2.2.3. RNA Extraction

Total RNA from frozen cell pellets was extracted using an RNAspin Mini RNA Isolation kit (GE Healthcare UK Limited, Buckinghamshire, UK) in accordance with the manufacturer's instructions. To eliminate the residual DNA, RNase-free DNase I treatment was performed during the isolation procedure. The purity of RNA sample was routinely determined from the optical density ratio (OD_{260}/OD_{280}) using a V-550 UV/VIS spectrophotometer (JASCO, Tokyo, Japan). Extracted RNA was checked by electrophoresis on a 1% agarose gel, staining with ethidium bromide, and viewed with a UV transilluminator.

2.2.4. cDNA Synthesis

Reverse transcription reactions were performed using an Exscript® RT regent kit

in accordance with the manufacturer's instructions (Takara Bio Inc., Otsu, Japan) with random 6-mers to obtain cDNAs from the mRNAs (4 μl). Reaction conditions were as follows: heating at 42°C for 15 min for reverse transcription reaction, heating at 95°C for 2 min for enzyme denaturation, and rapid cooling at 4°C in a total volume of 20 μl. To check for the presence of residual DNA, reverse transcriptase control samples (RT-negative control samples) were prepared for each RNA sample using the identical cDNA synthesis procedure except for omission of reverse transcriptase. All cDNA samples and RT-control samples were diluted 1:5 and stored at –20°C for use as templates in real-time PCR analysis.

2.2.5. Quantitative Real-time PCR

Primers were designed for the reference genes and target genes using Primer3 (8) and were obtained from Invitrogen Japan (Tokyo, Japan) (Table 2-3). PCR amplification and analysis were performed using a LightCycler instrument (Roche, Mannheim, Germany), software version 3.5 (Roche Diagnostics) and SYBR® Premix Ex Taq™ (Takara Bio Inc.) according to the manufacturer's recommendations. Briefly, the final PCR mix included 0.2 μM each primer, 2 μl diluted cDNA, and 10 μl of SYBR® Premix Ex Taq™ (2×) in a final volume of 20 μl. Cycling conditions consisted of heating to 95°C for 10 s, followed by 45 cycles of 5 s at 95°C, and 20 s at 60°C. .

Finally, melting curve analysis was performed.

Table 2-3: Characteristics of reference genes-specific real-time PCR assays

Name	5'-3' Primer sequence	Amplicon (bp)	PCR efficiency
rpoN	Forward: gttaaggctttgcaccag	108	0.95
	Reverse: gatttcatcgacctgctc		
rpoD	Forward: cgatggaaatcaccagac	143	0.92
	Reverse: gctgatcgaccttgagac		
dbhA	Forward: gctggccttgatcttga	138	0.97
	Reverse: attctcgacagcatcacc		
phaF	Forward: tggggtagttgaagatgc	135	0.93
	Reverse: ccaagattcaggacgaag		
16S rRNA	Forward: ccgtgtctcagttccagt	104	0.94
	Reverse: tgagcctaggtcggatta		
gst	Forward: ggacattctcaagggtga	109	0.91
	Reverse: gttggactcccacaggta		
lexA	Forward: tcaccagctcctgttctt	112	0.94
	Reverse: gacgaagtcaccgtcaag		
atkA	Forward: gtagtcggcaaaggtctg	114	0.93
	Reverse: caacttctgggtcgacat		

For relative quantification of gene expression, a standard curve was generated using a four-fold serial dilution of cDNA stock solution. Crossing point (Cp) and PCR efficiency were derived from the standards curve. The Cp value is considered as the signal point when the fluorescence rises appreciably above the baseline and is inversely correlated to cDNA concentration. Stock solution of cDNA, negative control (distilled water), and RT-negative control samples were included in each run. Duplicate measurements for each cDNA sample were performed and their mean values were calculated. Amplified products were analyzed by agarose gel electrophoresis using 2% agarose gels in Tris-acetate-EDTA buffer. The gels were then stained with ethidium bromide.

2.3. Results and Discussion

2.3.1. RNA Quality

Real-time RT-PCR requires careful assay design and reaction optimization to maximize the sensitivity and to get reliable quantitative results. RNA samples were examined for purity and gel electrophoresis on a 1% agarose gel. The range of the optical density ratio (OD_{260}/OD_{280}) was from 1.86 to 2.00 (Fig. 2-2). It indicated that all RNA samples were high purity. From the appearance of the two bands in the Fig.

2-3, the presence of 16s rRNA and 23s rRNA bands were confirmed. Since DNA can be co-extracted with RNA, another band was also observed and it has been considered as the contaminating DNA band (Fig. 2-3). In order to eliminate the contaminating DNA, DNase treatment was carried out. After DNase treatment, No other band was found in the region (Fig. 2-4). So, the contaminating DNA is not in extracted RNA samples. All RNA samples were pure and protein-free isolated. Thus, it was found that the DNase treatment is necessary and effective in obtaining sufficiently pure RNA.

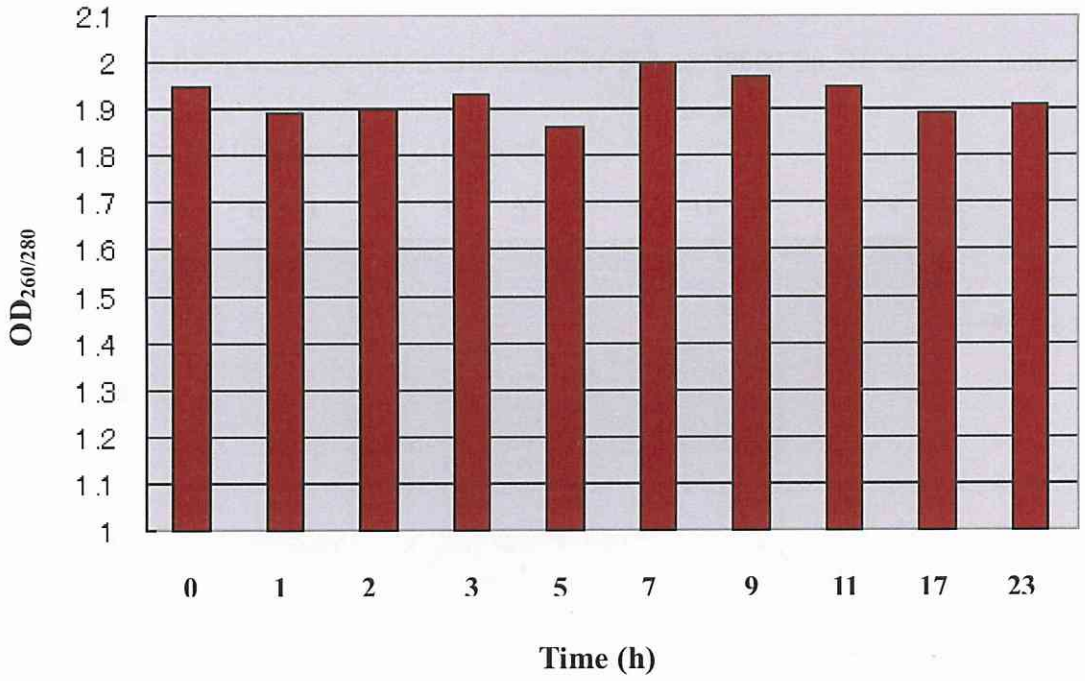


Fig. 2-2. The purity of RNA extracted from *P. putida* mt-2 culture growth with *p*-xylene as the carbon source. 0-23 indicates the time sample to collection.

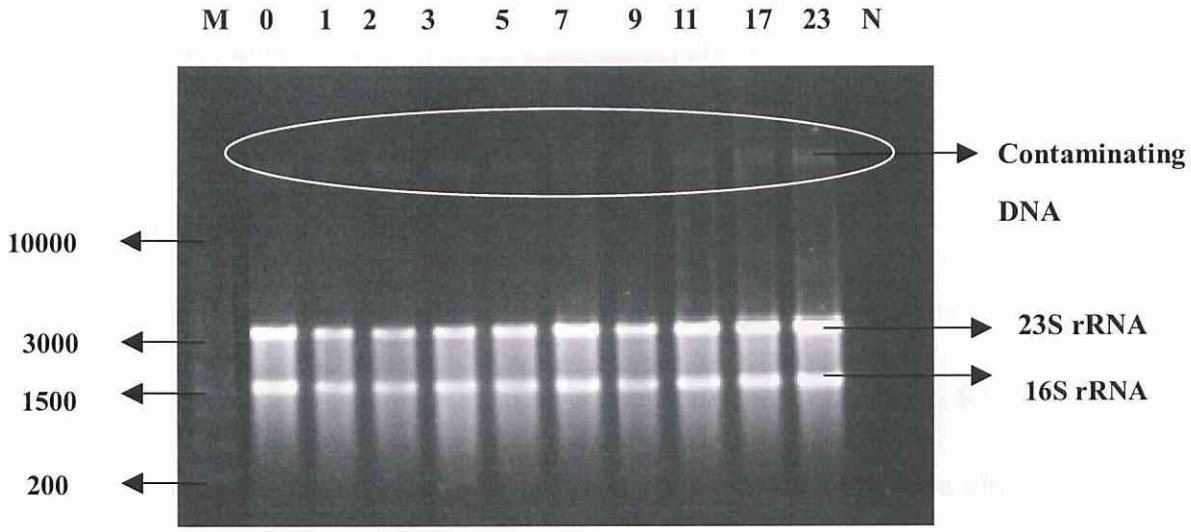


Fig.2-3. Agarose (1%) gel electrophoresis of total RNA extracted from *P. putida* mt-2 culture growth with *p*-xylene as the carbon source before DNase treatment. Lanes: M Perfect RNA markers with a size range of 200 to 10000 bp. N: negative control.0-23 indicates the time sample to collection.

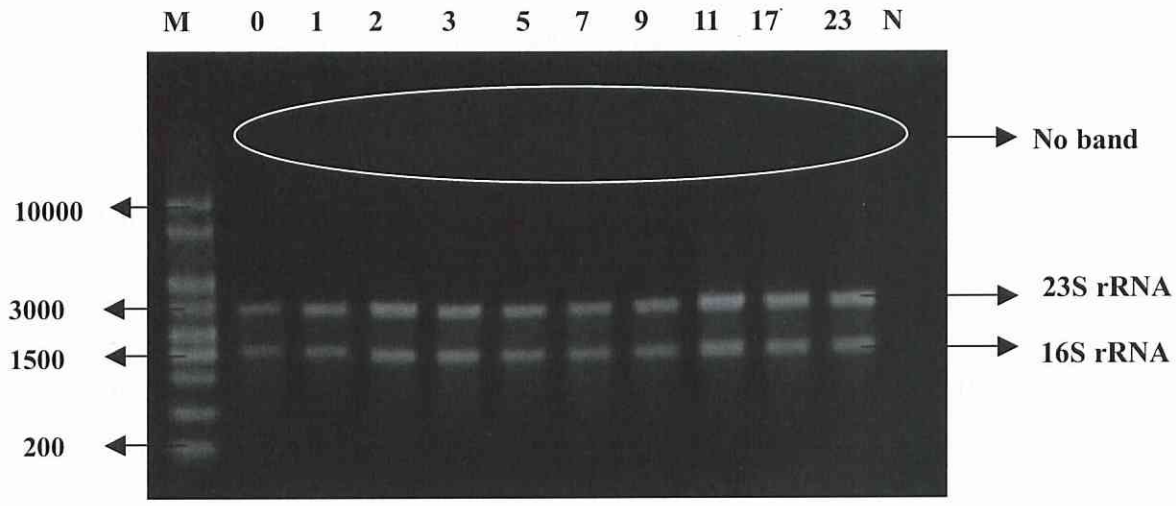


Fig.2-4. Agarose (1%) gel electrophoresis of total RNA extracted from *P. putida* mt-2 culture growth with p-xylene as the carbon source after DNase treatment. Lanes: M Perfect RNA markers with a size range of 200 to 10000 bp. N: negative control. 0-23 indicates the time sample to collection.

2.3.2. Difference between Cp $_{(RT\text{-}negative\ control\ samples)}$ and Cp $_{(RT\ samples)}$

Gene level studies have recently become popular because of the developments in molecular techniques that have allowed extensive investigation of gene expression. One of the problems with mRNA quantification is the DNA contamination in RNA samples due to co-extraction. DNase treatment is usually performed to selective eliminate the DNA contaminant from RNA samples using RNase-free DNase step and then the samples are used for the reverse transcription (RT-PCR). However, DNase treatment was performed before quantitative mRNA and there is a possibility that the RNA samples may still contain the DNA contaminant. Therefore, DNA contamination may interfere in the quantification of mRNA. So, quantification of DNA contaminant is also detected for a precaution against false positive results.

ΔCp values (Cp $_{(RT\text{-}negative\ control\ samples)}$ and Cp $_{(RT\ samples)}$) were calculated to compare the amounts of the RT-control sample and cDNA sample. As ΔCp value was found to be greater 4 (Fig. 2-5), the residual DNA level was regarded as negligible (9).

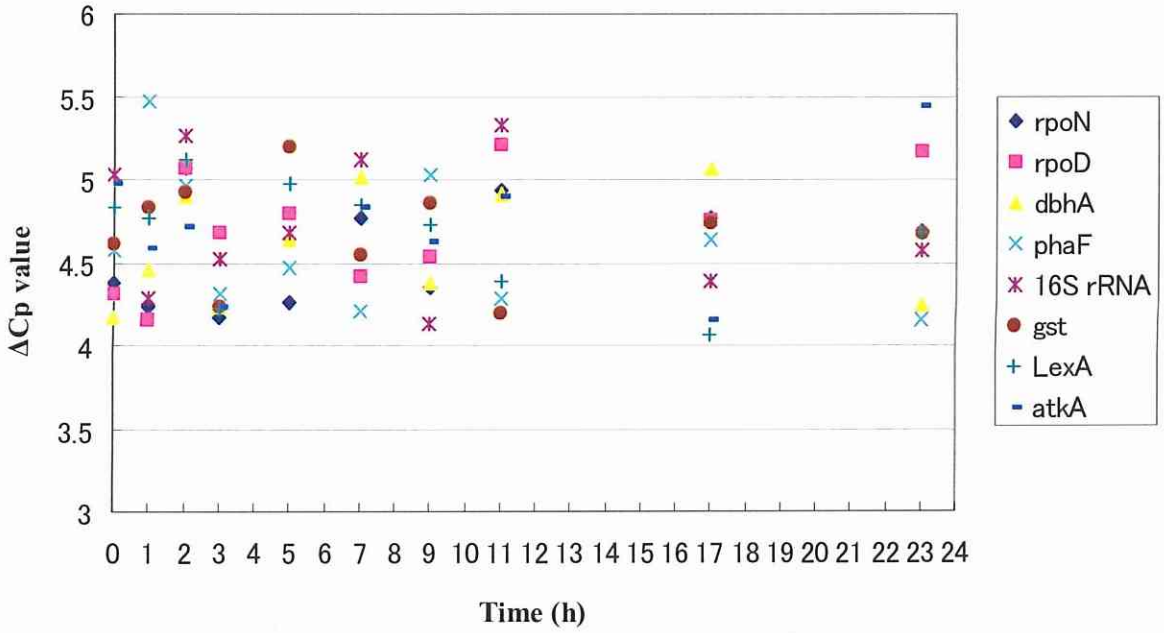


Fig. 2-5. Difference between Cp $_{(RT\text{-}negative\ control\ samples)}$ and Cp $_{(RT\ sample)}$ in reference genes of all samples.

2.3.3. Real-time PCR Specificity

Following real-time PCR, the resulting PCR products were checked to confirm the amplification specificity. Melting curve analysis is an inversion of PCR detection method, in which increases in fluorescence are detected during cycling (10). In order to visualize nonspecific PCR, a melting curve analysis can be performed. Different fragments will usually appear separating melting peaks. The PCR products of reference gene (*rpoN*) and other reference genes were analyzed by melting curve analysis as

shown in Fig. 2-6 A and 2-6 B, respectively. Only a single melting peak at the same melting temperature was produced for PCR product. Melting curve analysis did not detect any primer dimmers or other side-product. The result indicated the PCR reaction is accurate and high specificity.

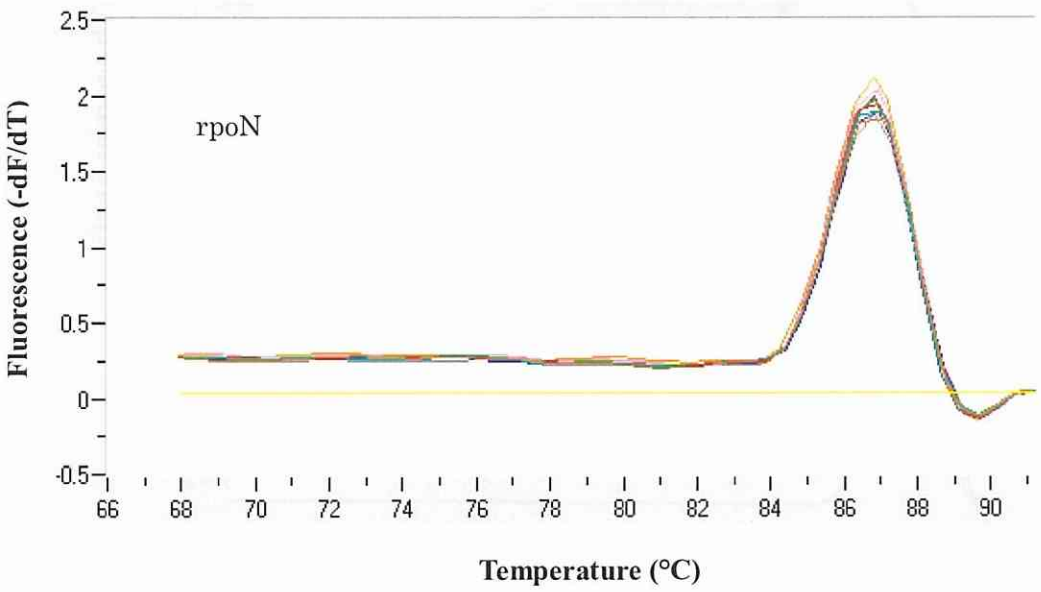


Fig. 2-6 A. The real-time PCR was specific. In real-time PCR, a melting curve analysis was performed to demonstrate the specificity of the reactions.

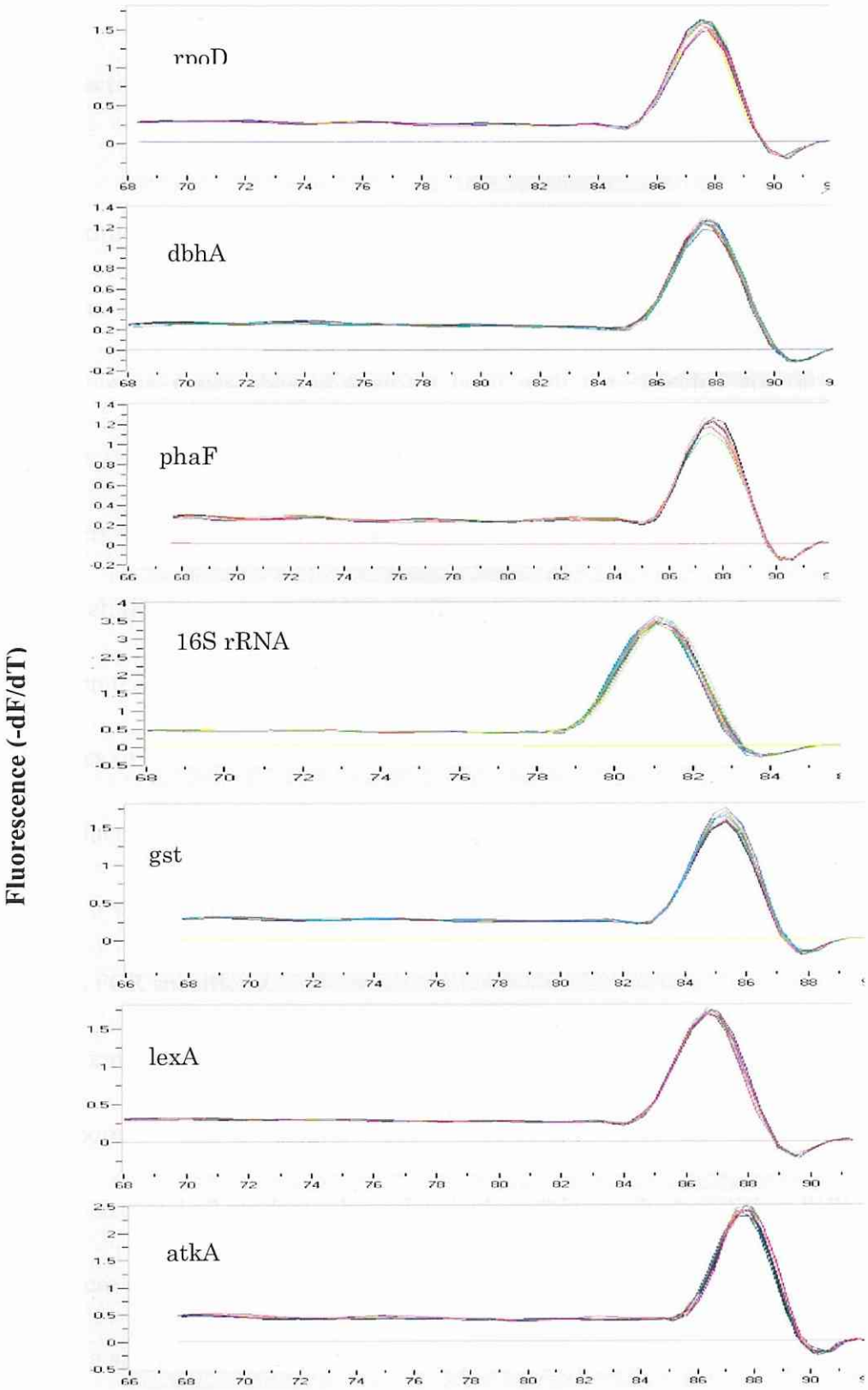

rpoD
dbhA
phaF
16S rRNA
gst
lexA
atkA
Fluorescence (-dF/dT)
Temperature (°C)

Fig. 2-6 B

2.3.4. Electrophoresis of RT-PCR

In order to confirm PCR specificity, gel electrophoresis was preformed. Electrophoresis analysis of all the amplified products for reference gene (*rpoN*) and other reference genes showed a single band with the expected size in all samples (Fig.2-7A and Fig.2-7B). The result indicated that non-specific PCR products with the primer sets were not detected in the analyzed temperature range. Because the secret of the gel's ability to separate DNAs of different sizes lies in friction, every PCR product also generated prominent bands with expected sizes in the gel electrophoresis analysis. If the rates of desired band and other band move different and primer dimmers or other side-products fall without the range of the standards, it reveals PCR non-specificity. In contrast, a single band with the desired size was found in the range of standards. It describes PCR specificity.

The amplification specificity was checked by both melting curve analysis and gel electrophoresis. Agarose gel electrophoresis demonstrated a single band with the expected size, and all products showed a single melting peak on real-time PCR. Two results were in agreement. Therefore, only primers sets that produce a single melting peak and a single band of expected size on gel were used for further analysis.

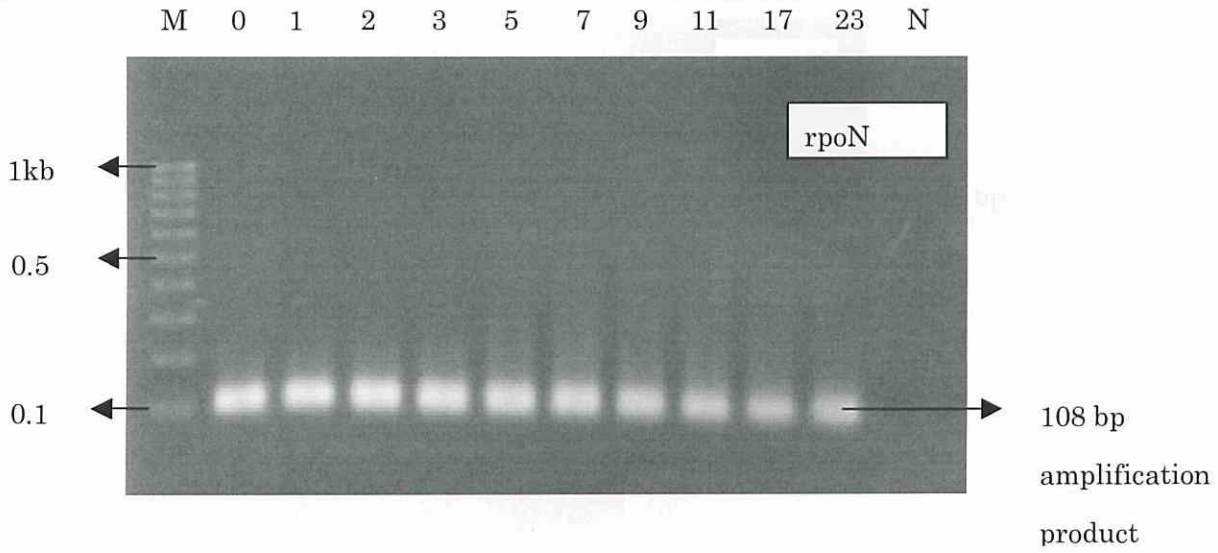


Fig. 2-7 A. Agarose gel electrophoresis (2%) of the RT-PCR products. Lanes: M, 100-bp size markers; 0-23 numbers indicate the incubation time (h). N, negative control.

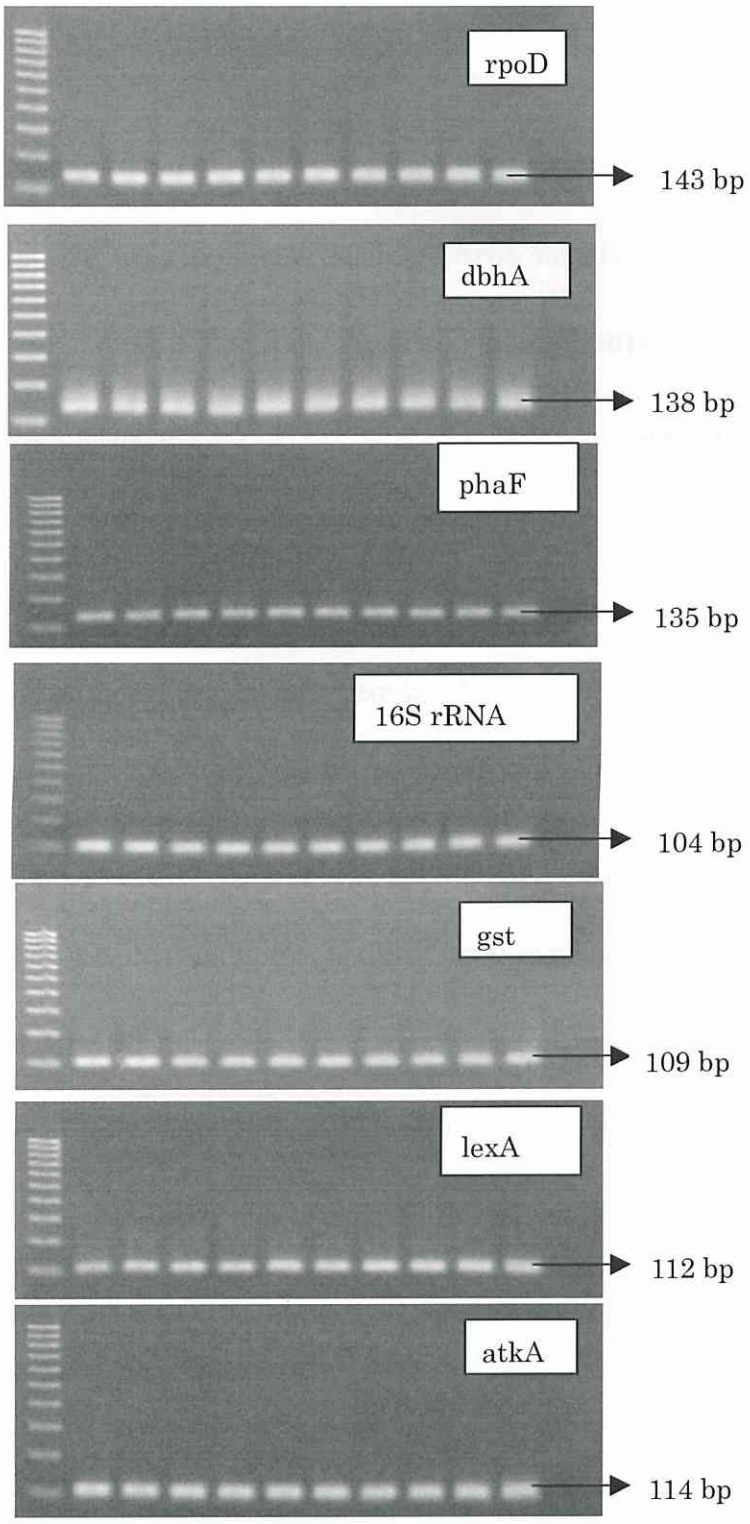

rpoD
143 bp
dbhA
138 bp
phaF
135 bp
16S rRNA
104 bp
gst
109 bp
lexA
112 bp
atkA
114 bp
M 0 1 2 3 5 7 9 11 17 23 N

Fig. 2-7 B

2.4. Conclusions

In conclusion, all works should be made to obtain high-quality RNA sample. The quantitative mRNA has the major problems, such as, RNA samples purity, designed primer sets, and DNA co-extraction. The study illustrated the application real-time PCR and the primer sets were designed. SYBR® Premix Ex Taq™ was used in real-time quantitative PCR because it can analyze the accumulation of primer dimmers and the amplification of non-specific PCR products. The OD_{260}/OD_{280} values showed all RNA samples were purity and designed primer sets were good. Because one melting peak at the same melting temperature was produced for PCR product and each PCR product also generated prominent bands with expected sizes in the gel electrophoresis analysis, it demonstrated quantitative PCR condition is optimal. Only primers sets were used for further analysis. In according to ΔCp values that were greater 4 and the residual DNA level as negligible, the all RNA samples can be used for the future quantitative gene expression. Our results suggest that the investigation of RNA samples purity and the confirmation of primers specificity are important step in further gene expression analysis.

REFERENCES

1. Bustin, S.A.: Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J. Mol.Endocrinol., 2002, 29: 23-39.

2. Perez-Novo, C.A., Claeys, C., Speleman, F., Gauwenberge, P.V., Bachert, C. and Vandesompele, J.: Impact of RNA quality on reference gene expression stability. Biotechniques 2005, 39: 52-56.

3. Bustin, S.A., Benes, V., Nolan, T. and Pfaffl, M. W.: Quantitative real-time RT-PCR-a perspective. J. Mol.Endocrinol., 2005, 34: 597-601.

4. Gibson, U. E., Heid, C. A. and William, P.M.: A novel method for real time quantitative RT-PCR. Genome Res., 1996, 6: 995-1001.

5. http://www.thefreedictionary.com/culture+medium.

6. Akira S., Catalogue of strains 6th Ed., Japan Collection of Microorganisms. 1995, p488.

7. Harayama, S., Leppik, R. A., Rekik, M., Mermod, N., Lehrbach, P. R., Reineke, W. and Timmis, K.N.: Gene order of the TOL catabolic plasmid upper pathway operon and oxidation of both toluene and benzyl alcohol by xylA product. J. Bacteriol., 1986, 167: 455-461.

8. http://frodo.wi.mit.edu/cgi-bin/primer3 /primer3.cgi.

9. Rudi, K., Nogva, H. K., Naterstad, K., Dromtorp, S. M., Bredholt, S. and Holck, A.: Subtyping *Listeria monocytogenes* through the combined analyses of genotype and

expression of the *hlyA* virulence determinant. J. Appl. Microbiol. 2003, 94: 3862-3871.

10. Changsoo L., Jaai K., Seung. Gu S. and Seokhwan H.: Abslute and relative QPCR quantification of plasimd copy number in *Escherichia coli*. J. Biotechnol. 2006, 123: 273-280.

Chapter 3

Evaluation of Reference Genes in *Pseudomonas putida* mt-2 in Presence of *p*-xylene

Abstract

The quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR) is an important and very useful tool for the quantification of gene expression. qRT-PCR relies on accurate normalization of gene expression data. Reference genes are used as normalizer for relative quantification of target genes in gene expression studies. However, the expression level of these genes may vary among cells, and may change under certain circumstance. In this study, we performed real-time PCR to investigate the expression of eight reference genes (*rpoN, rpoD, dbhA, phaF, 16S rRNA, gst, lexA, and atkA*) in *Pseudomonas putida* mt-2 during degradation of *p*-xylene. The dedicated validation program (geNorm) was used to rank the eight reference genes from best to worst and ascertain the most suitable reference from these candidates. *rpoN, rpoD*, genes were the most stable reference genes, while the *phaF* and *dbhA* genes showed unstable expressed genes by the method. The normalization factors (NF_4) were obtained from the four most suitable reference genes (*rpoN,-rpoD,-16S rRNA*, and *atkA*) by geNorm method.

3.1. Introduction

Detection of bacterial gene expression in environment has become an active field of research in microbial ecology, bioremediation, and diversity monitoring. The recently developed quantitative real-time PCR method is extremely sensitive, and rapid, has a broad dynamic quantification range, and is relatively easy to perform (1, 2). The method allows the detection of amplicaon accumulation since it is performed using intercalating dyes such as SYBR Green I, rather than by conventional end-point analysis.

Relative quantitative real-time PCR can be applied to normalization of target gene expression to some form of control, to assess variations in sample input, extraction, reaction efficiencies, and RNA quality among samples. The expression levels of reference genes should remain relatively constant among different samples and experimental conditions (3). If the requirements are not fulfilled, then normalization to varying internal references can lead to increased "noise" or erroneous results (4). Thus, the use of valid reference genes is a prerequisite for accurate gene quantification.

Bacterial reference genes have been used for normalization of gene expression (5-9). However, several studies indicated that the expression levels of the reference genes did not remain constant under different metabolic conditions or among treatments in the same sample or different growth stages (10-13). Identification of candidate genes that are at least minimally regulated under the conditions investigated and preferably the

inclusion of more than one reference gene in the analysis are important for the accuracy of quantitative real-time PCR test (14,15). For prokaryotic mRNA quantification, the expression stability of different reference genes was validated for *Pseudomonas aeruginosa* and *Actinobacillus pleuropneumoniae* (15).

The suitability of reference genes for relative quantitative gene expression in the environment has not been investigated in sufficient detail to date. To find suitable reference genes, we evaluated the expressional stability in a panel of eight candidate reference genes (*rpoN, rpoD, dbhA, phaF, 16S rRNA, gst, lexA, and atkA*) in *Pseudomonas putida* mt-2 during degradation of *p*-xylene using the geNorm program (14). The selected genes take part in critical functions. The *rpoD* gene encodes the housekeeping sigma factor σ^{70}, which is a critical housekeeping gene (16). The *phaF* gene is involved in a ring-hydroxylation system (17). The *dbhA* gene encodes the DNA-binding protein HU-alpha (18). The *rpoN* gene encodes the alternative sigma factor σ^{54}, which was found to be involved in growth phase-dependent activation of promoters of various genes (19). The *gst* gene encodes glutathione S-transferase family protein (20). The *lexA* gene encodes a LexA repressor that is able to bind the *recA* gene promoter region (21). The *atkA* gene encodes cation-transporting ATPase (22). 16S rRNA is an important part of the ribosomal complex (13). 16S rRNA was the most abundant among the genes tested all the time as determined from the Cp value, which was consistent with the results reported previously (13).

The goal of our study is the systematic comparison of the stability among reference genes in *Pseudomonas putida* mt-2 during biodegradation of *p*-xylene using the geNorm program and the relationship between gene expression and stability. In addition, the aim is to determine which genes are most appropriate for the normalization of gene expression and normalization factors can be calculated from the number of most suitable reference genes by the geNorm program.

3.2. Experimental Section

3.2.1. Materials for Culture and Growth of the Bacterial Strain

Pseudomonas putida mt-2 is the original isolate of *P. putida* bearing the archetypical TOL plasmid (23). Japan Collection of Microorganisms RIKEN BioResource Center provided *P. putida* mt-2 (DSM 3931). The bacteria were cultured at 30°C on a rotary shaker at 110 rpm in M9 mineral medium (24) of 12ml to expose to vapors of *p*-xylene as the sole carbon source. To this end, the release of inducer was slowed down by generating vapors of *p*-xylene (9.5 μl) from small amount of cotton. Three replicate vials were incubated.

The liquid cultures were centrifuged at 4°C for 10 min at 6000 rpm, cells pellets washed twice in phosphate-buffered saline, and then stored at –80°C until use. The

detail culture procedure has been introduced in the experimental section of chapter 2.

3.2.2. Isolation of Total RNA

Total RNA was extracted from frozen cell pellets by an RNAspin Mini RNA Isolation kit (GE Healthcare UK Limited, Buckinghamshire, UK) in accordance with the manufacturer's instructions. To eliminate the residual DNA, RNase-free DNase I treatment was performed during the isolation procedure.

3.2.3. RT Step

Reverse transcription reactions were performed using an Exscript® RT regent kit in accordance with the manufacturer's instructions (Takara Bio Inc., Otsu, Japan) with random 6-mers to obtain cDNAs from the mRNAs (4 μl). Reaction conditions were as follows: heating at 42°C for 15 min for reverse transcription reaction, heating at 95°C for 2 min for enzyme denaturation, and rapid cooling at 4°C in a total volume of 20 μl. To check for the presence of residual DNA, reverse transcriptase control samples (RT-negative control samples) were prepared for each RNA sample using the identical cDNA synthesis procedure except for omission of reverse transcriptase. All cDNA samples and RT-control samples were diluted 1:5 and stored at –20°C for use as

templates in real-time PCR analysis.

3.2.4. Quantitative Real-time PCR

Primer3 (25) was used to design the all primers. The primers were obtained from Invitrogen Japan (Tokyo, Japan). PCR amplification condition and analysis were introduced in the experimental section of chapter 2.

3.2.5. Data Analysis

The mathematical equations used for the analysis of gene expression data are as follow (26):

$$X_n = X_0 \times (E+1)^n \quad \text{..... (1)}$$

Here, X_n is the amount of PCR product at cycle n, X_0 is the beginning amount PCR template (which we want to know) and E is amplification efficiency which can have a value between 0 (no amplification) and 1 (doubling of the PCR product in each amplification cycle).

Since fluorescence signal is proportional to the accumulation of PCR amplification

product, the equation 1 can be written as:

$$F_n = F_0 \times (E+1)^n \quad \text{..... (2)}$$

where, F is fluorescence signal after background subtraction.

Then, the beginning of fluorescence F_0 can be calculated as:

$$F_0 = \frac{F_n}{(E+1)^n} \quad \text{..... (3)}$$

where F_n is the signal of dye fluorescence at cycle n and F_0 the theoretical beginning fluorescence that is proportional to the amount of beginning PCR template. In real-time PCR technique, n in the equation can be replaced by Cp:

$$F_0 = \frac{F_{Cp}}{(E+1)^{Cp}} \quad \text{..... (4)}$$

Cp that is the fluorescence signal needed to rise above baseline is detected. F_{Cp} represents the threshold fluorescence, which can be set for each of the compared

amplification individually, or a threshold value (F_{Cp}) common for all compared amplification can be used. Amplification efficiency (E) for each gene was calculated using the equation of the standard curve:

$$Cp = -\frac{1}{\log(E+1)} \times \log F_0 + \frac{\log F}{\log(E+1)} \quad \text{..... (5)}$$

$$E = 10^{\frac{1}{slope}} - 1 \quad \text{..... (6)}$$

$$slope = -\frac{1}{\log(E+1)} \quad \text{..... (7)}$$

To compare the amounts of the cDNA sample and RT-control sample, ΔCp was calculated. As ΔCp of >4 was found in this study, the residual DNA level was regarded as negligible (27).

To compare the stability of candidate reference genes, we applied geNorm, version 3.5 (14). The geNorm software depends on the assumption that the expression ratio of two ideal internal control genes is identical in all samples, regardless of the experimental conditions. Cp values were transformed into relative quantities by equation 8 for analysis with geNorm software manual (28). The calculation procedures

of the stability M values and the normalization factors (NF_n) have been introduced in chapter 1.

$$a_{i,j} = (E+1)^{\Delta C_{pi,j}} \quad \ldots\ldots\ldots\ldots\ldots 8$$

(here: $\Delta Cp_{i,j} = Cp_{j(Lowest)} - Cp_{i,j}$

$C_{p\,j(Lowest)}$: Lowest crossing point for j gene among all samples.

$C_{p\,i,j}$: Crossing point for j gene in i sample)

3.3. Results and Discussion

3.3.1. Standard Curve for Reference Genes

Standard curves of the analyzed reference genes are shown in Fig. 3- 1 and were generated by using logarithmic values of standard RNA amounts vs the crossing point (Cp) that is the cycle number when the threshold fluorescence is reached. The coefficients of the standard curves indicate that the efficiencies of reverse transcription and the PCR amplification efficiencies are similar in all dilutions. Amplification efficiency (E) representative for each reference gene was calculated from equation 6 in Experimental Section. The coefficient values (R^2) were shown in Table 3-1. Detection and quantification were linear over the range of the crossing point examined. The result showed that these standard curves can be used for relative quantification.

Table 3-1. The coefficient values of reference genes

Gene	*rpoN*	*rpoD*	*lexA*	*dbhA*	*atkA*	*gst*	*16S rRNA*	*phaF*
R^2	0.9956	0.9941	0.9952	0.9959	0.9987	0.9972	0.9964	0.9983

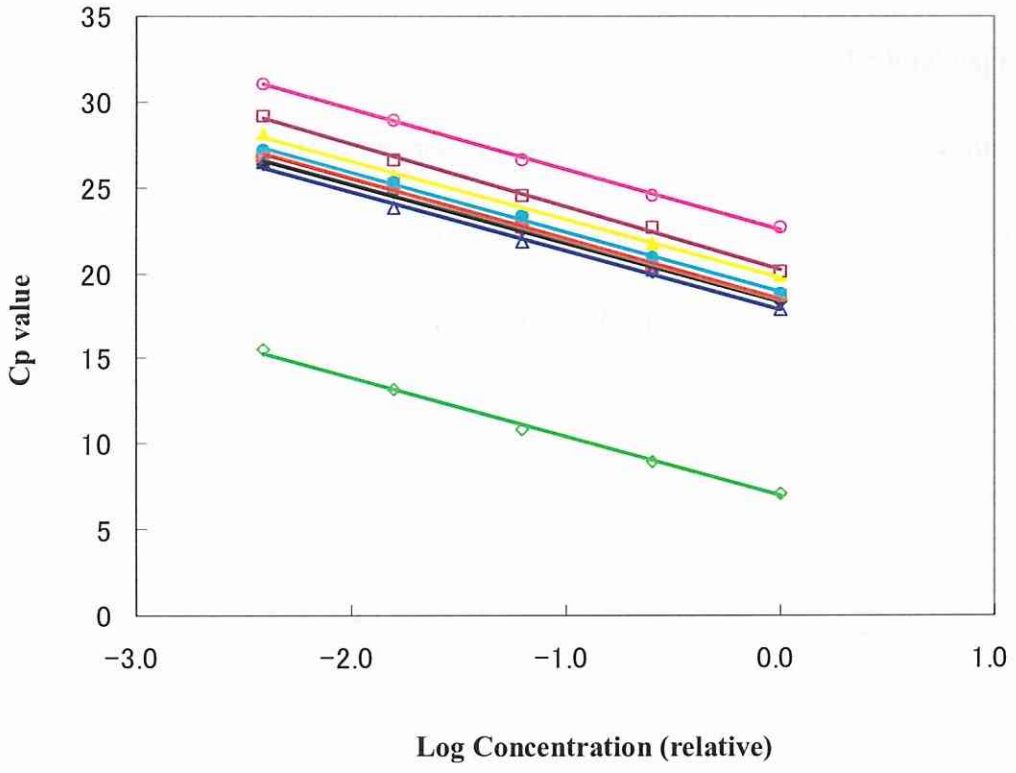


Fig. 3-1. Standard curves for reference genes. *rpoN* (black curve and closed diamonds); *rpoD* (red curve and closed squares); *lexA* (blue curve and open triangles); *dbhA* (yellow curve and closed triangles); *atkA* (purple curve and open circles); *gst* (amaranth curve and open squares); *16S rRNA* (green curve and open diamonds); *phaF* (azure curve and closed circles). Quantities of standard RNA were expressed as dilution factors of RNA preparation (1, 0.25, 0.0625, 0.015625, 0.0039062).

3.3.2. Expression Profiling of the Candidate Reference Genes

The expression levels of the eight candidate reference genes were investigated, and are shown depending on Cp-values as a box-plot in Fig. 3-2. In this study, ten collection samples for each reference gene were included. The median expression level and interquartile values are shown in boxes and the whiskers indicated the total expression ranges. The eight candidate reference genes studied showed a wide expression range, with Cp values between 9 and 28. Figure 3-2 indicated that two arbitrary lines separated these reference genes into three groups (29). Genes with lower expression levels showed higher Cp values and genes with higher expression levels showed lower Cp values. Among the genes tested, *16S rRNA* showed the highest level of expression with a mean (±SD) Cp value of 11.03 ± 0.86, while *atkA* showed the lowest level of expression with a mean (±SD) Cp value of 27.20 ± 0.61. Genes spanning a maximal expression range were *rpoN* (1.36), *rpoD* (1.44), *dbhA* (2.88), *phaF* (2.37), and *16S rRNA* (2.32), *gst* (2.35), *lexA* (2.47), *atkA* (1.71), indicating changes in the expression level of reference genes in the bacteria during biodegradation of *p*-xylene.

In order to avoid co-regulated genes, the eight employed reference genes (*rpoN*, *rpoD*, *dbhA*, *phaF*, *16S rRNA*, *gst*, *lexA*, *atkA*) (8, 9) belonging to different functional classes were selected in this study. Based on SYBR Green detection, a real-time qPCR assay was preformed for expression profiling of these reference genes. Although the

growth of bacteria is a physiological process, we found variable expression levels of not only the target gene but also the reference genes. These observations were consistent with the results of previous studies in both eukaryotic (30, 31) and prokaryotic (13) cells. Ideal and universal reference genes do not exist. Therefore, the stability of reference genes in prokaryotic should be tested under the investigated conditions.

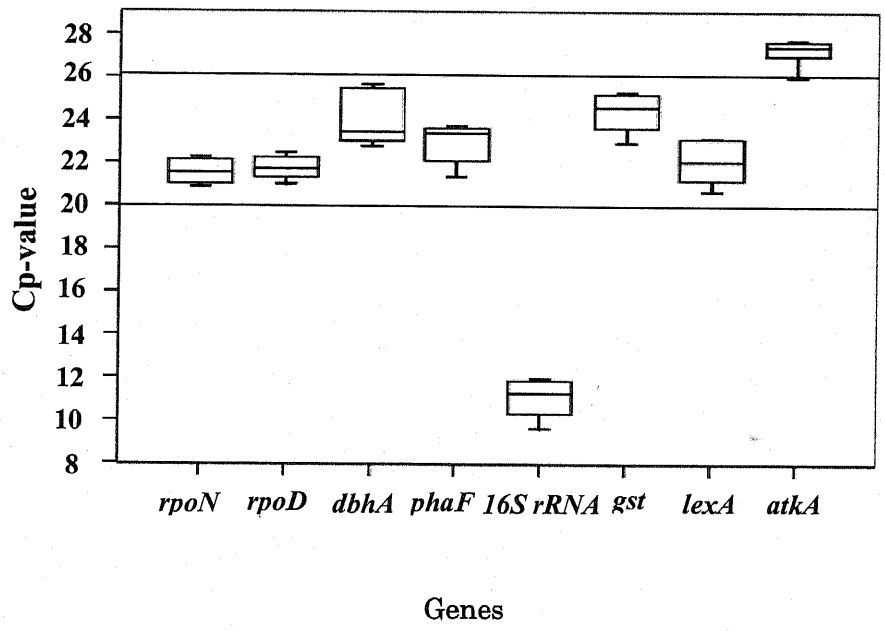


Fig. 3-2. Real-time PCR Cp-values in collected samples. The distributions of expression levels of candidate reference genes are shown as median (lines), lower and upper quartiles (boxes), and ranges (whiskers) (n=10). Different expressed reference genes were divided into 3 groups by the arbitrary lines at Cp 20 and 26.

3.3.3. Candidate Reference Genes Expression Stability Measured by Software Program

The essential requirement of a candidate gene for normalization purposes is its invariable expression in each search situation. Therefore, particular validation of potential reference genes for the respective conditions is needed. It is a very time consuming and labor-intensive process to search for an appropriate reference gene, and various software and methods have been suggested to simplify the search. Several approaches have been recommended to identify suitable reference genes from candidate genes. (32, 33-35). In this study, we applied geNorm to objectify our results. The program has often been used in other studies to find suitable reference genes from a set of candidates (14, 30, 36-39).

To identify the reference gene with the most stable expression, the data were analyzed using the geNorm software package (14), which chooses appropriate reference genes by calculating the gene expression stability measure, the *M* value. The value is the mean pairwise variation for a gene compared with all other tested control genes. Increasing *M* value corresponds to decreasing expression stability. The gene with the highest *M* value is excluded and the gene with the lower *M* value is regarded as the most stable gene. A new *M* value is calculated and this calculation procedure is preformed again until only two genes remain. These two genes have the lowest *M* value and are the

most stably expressed. According to their increasing expression stability (decreasing *M* value), the stability ranking of the eight reference genes is shown in Fig. 3-3. All of the genes studied appeared to show high expression stability with low *M* values less than 0.6, which were below the default limit of 1.5 in the geNorm program.

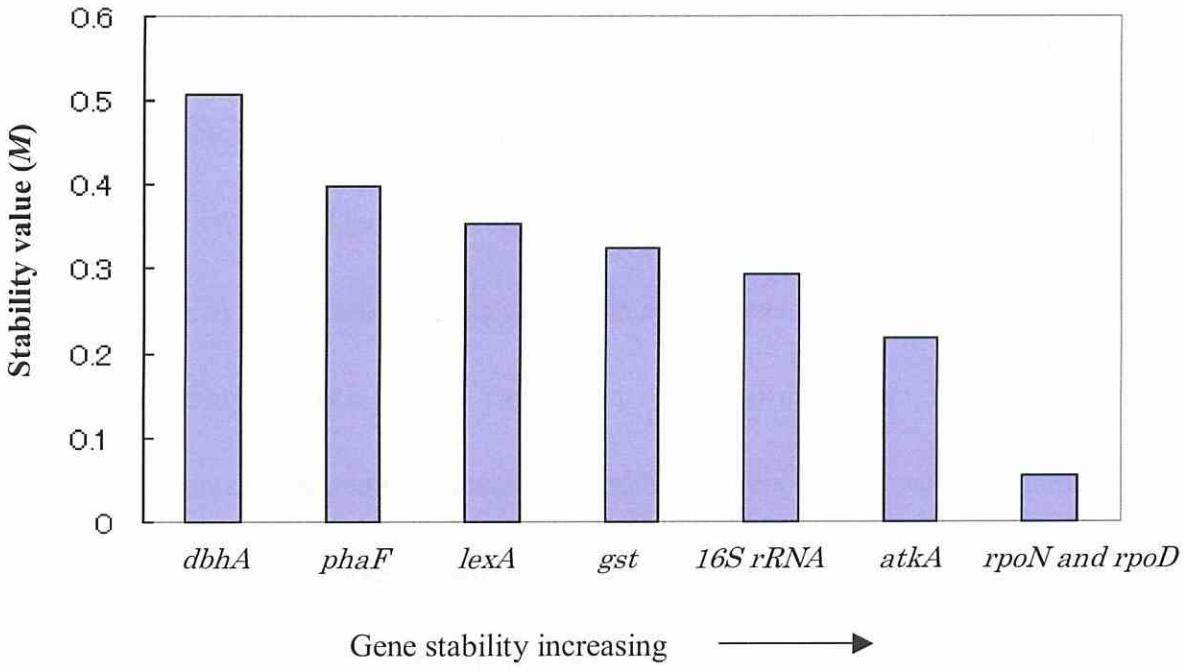


Fig. 3-3. Stability ranking of the reference genes by geNorm. Average expression stability value *M* after excluding unstable gene at every step was calculated. Genes are ranked from left to right to increasing expression stability (decreasing *M* value).

3.3.4. Normalization Factor Calculated by GeNorm Method

It has been suggested that normalization should be based on more than one reference gene alone to achieve more accurate normalization (14, 12). The geNorm software made a pairwise variation ($V_{n/n+1}$) that was calculated between the two sequential normalization factors for determination of the optimal number of genes necessary. The pairwise variation analysis showed that $V_{2/3}$ value was 0.099 (Fig. 3-4), which was below the cutoff value of 0.15. The cutoff value was considered as the limit beneath which it would not be necessary to include additional reference genes for normalization. A large variation means that the added gene has a significant effect and should preferably be included for calculation of the normalization factor. Since the geNorm software suggested that at least three stable reference genes were recommended and lower *V* value did not essentially decrease when no less than five reference genes were included, we used four reference genes (*rpoN*, *rpoD*, *16S rRNA*, and *atkA*) for normalization factor calculation of gene profiling studies in this experiment. By calculating the normalization factor (NF_4) that is the geometric mean of *rpoN-rpoD-16S rRNA-atkA* genes, the most accurate way of normalization can be obtained.

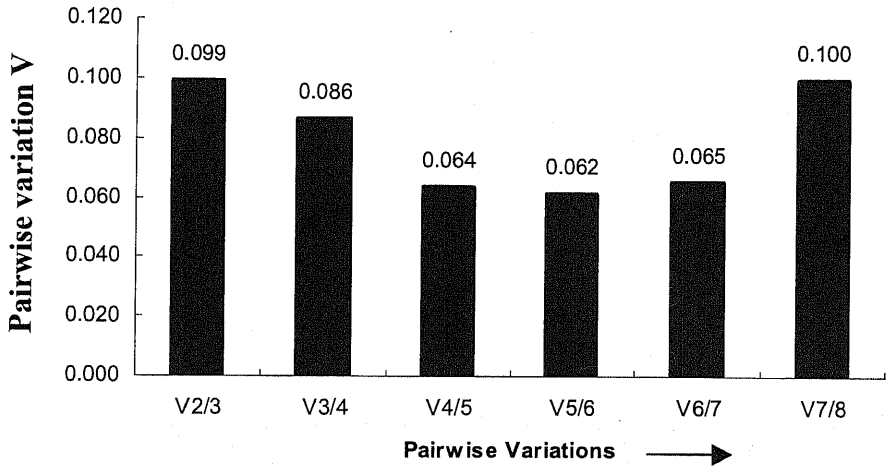


Fig. 3-4. Optimal number of control genes for normalization. Pairwise variability $V_{n/n+1}$ showed the change in normalization accuracy with stepwise addition of more reference genes according to the stability ranking of the reference genes.

3.3.5. The Relationship between Average Gene Expression and Gene Stability

Although the average Cp values of eight reference genes studies cover a wide dynamic range (average Cp = 11-27), there was no correlation (coefficient of determination, $R^2 = 0.002$) between the average Cp values and the *M* values by geNorm (Fig. 3-5). As mentioned above, four genes, *rpoN*, *rpoD*, *16S rRNA*, and *atkA*, were considered as the most suitable reference genes for the calculation of normalization factors. These genes represented the three arbitrarily defined levels of low, intermediate, and high gene expression (Fig. 3-2). These stable reference genes, minimally regulated

in response to experimental treatments, are of paramount importance in the relative quantification of target gene expression under different physiological conditions.

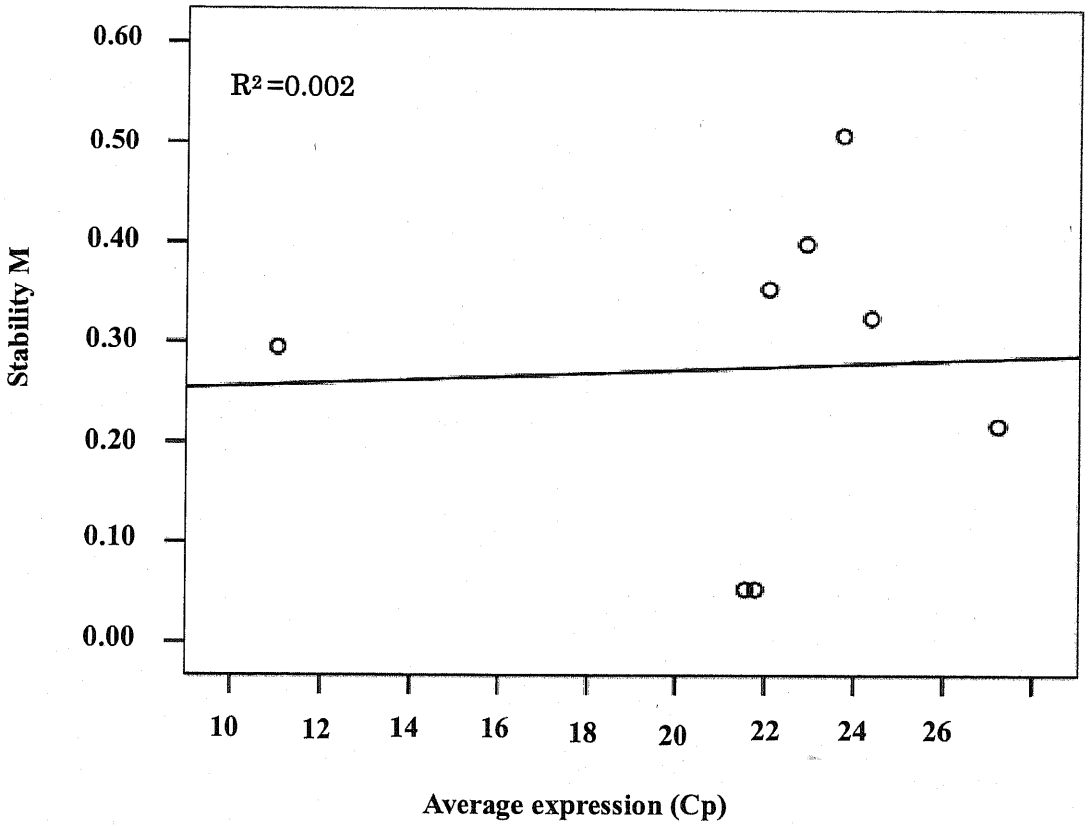


Fig. 3-5. Expression stability (*M*) versus average expression (Cp) by geNorm. According to the diverse scatters, there was no apparent correlation (R^2 = 0.002) between gene expression stability and expression values.

3.4. Conclusions

In order to normalize the variations in the amount of starting material, enzymatic efficiencies, and between cells in overall transcriptional activity, various strategies have been applied to normalization these variations. The selection of stable reference genes is the major problem for normalization purpose. Our approach has let to investigate eight reference genes fairly stably expressed *in P. putida* mt-2 throughout degradation of *p*-xylene. Our results reveal there are the significant variations in reference genes during bacterial growth by geNorm programs, and shows there is not apparent correlation between gene expression stability and gene expression. *rpoN*, *rpoD*, genes were the most stable reference genes, while the *phaF* and *dbhA* genes showed unstable expressed genes. Additional, lower *V* value did not significantly decrease when no less than five reference genes were included, we determine to use four reference genes (*rpoN*, *rpoD*, *16S rRNA*, and *atkA*) for normalization factor calculation of gene profiling studies in this experiment. The normalization factors (NF_4) were obtained from the geometric mean of four most stable reference genes (*rpoN*,-*rpoD*,-*16S rRNA*, and *atkA*) by geNorm method and can be used to normalize target genes expressions for accurate quantification.

REFERENCES

1. Ginzinger, D.: Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. Exp. Hematol., 2002, 30: 503-512.

2. Klein, D.: Quantification using real-time PCR technology: applications and limitations. Trends Mol. Med., 2002, 8: 257-260.

3. Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A. and Heinen, E.: Housekeeping genes as internal standards: use and limits. J. Biotechnol., 1999, 75: 291-295.

4. Bustin, S. A.: Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J. Mol. Endocrinol., 2000, 25: 169-193.

5. Botteldoorn, N., Van Coillie, E., Grijspeerdt, K., Werbrouck, H., Haesebrouck, F., Donne, E., D'Haese, E., Heyndrickx, M., Pasmans, Frank. and Herman, L.: Real-time reverse transcription PCR for the quantification of the *mntH* expression of *Salmonella enterica* as a function of growth phase and phagosome-like conditions. J. Microbiol. Methods., 2006, 66: 125-135.

6. Werbrouck, H., Grijspeerdt, K., Bottleldoorn, N., Van Pamel, E., Rijpens, N., Van Damme, J., Uyttendaele, M., Herman, L. and Van Coillie, E.: Differential *inlA* and *inlB* expression and interaction with human intestinal and liver cells by *listeria monocytogenes* strains of different origins. Appl. Environ. Microbiol., 2006, 72:

3862-3871.

7. Sue, D., Fink, D., Wiedmann, M. and Boor, K. J.: σ^B-dependent gene induction and expression in *Listeria monocytogenes* during osmotic and acid stress conditions simulating the intestinal environment. Microbiology., 2004, 150: 3843-3855.

8. Velazquez, F., Parro, V. and De Lorenzo, V.: Inferring the genetic network of *m*-xylene metabolism through expression profiling of the *xyl* genes of *Pseudomonas putida* mt-2. Mol. Microbiol., 2005, 57: 1557-1569.

9. Conte, E., Catara, V., Greco, S., Russo, M., alicata, R., strano, L., Lombardo, A., Di silvestro, S. and Catara, A.: Regulation of polyhydroxyalkanoate synthases (*phaC1* and *phaC2*) gene expression in *Pseudomonas corrugata*. Appl. Microbiol. Biotechnol., 2006, 72: 1054-1062.

10. Schmittgen, T. D. and Zakrajsek, B. A.: Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. J. Biochem. Biophys. Methods., 2000, 46, 69-81.

11. Goidin, D., Mamessier, A., Staquet, M. J., Schmitt, D. and Berthier-Vergnes, O.: Ribosomal 18S RNA prevails over glyceraldhyde-3-phosphate dehydrogenase and beta-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulation. Anal. Biochem., 2001, 295: 17-21.

12. Huggestt, J., Dheda, K., Bustin, S. and Zumla, A.: Real-time RT-PCR normalization;

strategies and considerations. Genes Immune., 2005, 6: 279-284.

13. Vandecasteele, S. J., Peetermans, W. E., Merckx, R., and Van, E.: Quantification of expression of *Staphylococcus epidemidis* housekeeping genes with Taqman quantitative PCR during in vitro growth and under different conditions. J. Bacteriol., 2001, 183: 7094-7101.

14. Vandesompele, J., DePreter, K., Pattyn, F., Poppe, B., VanRoy, N., Depaepe, A. and Speleman, F.: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol., 2002, 3, research0034.1-research0034.

15. Nielsen, K. K. and Boye, M.: Real-Time quantitative reverse transcription-PCR analysis of expression stability of *Actinobacillus pleuropneumoniae* housekeeping genes during in vitro growth under iron-depleted conditions. Appl. Environ. Microbiol., 2005, 71: 2949–2954.

16. 23. Schnider, U., Keel, C., Blumer, C., Troxler, J., Defago, G. and Haas, D.: Amplification of the housekeeping sigma factor in *Pseudomonas fluorescens* CHA0 enhances antibiotic production and improves biocontrol abilities. J. Bacteriol., 1995, 177: 5387-5392.

17. Olivera, E. R., Minambres, B., Garcia, B., Muniz, C., Moreno, M. A., Ferrandez, A., Diaz, E., Garcia, J. L., and Luengo J. M.: Molecular characterization of the phenylacetic acid catabolic pathway in *pseudomonas putida* U: The phenylacetyl-CoA catabolon.

Proc. Natl. Acad. Sci., 1998, 95: 6419-6424.

18. Kano, Y., Ogawa, T., Ogura, T., Hiraga, S., Okazaki, T. and Imamoto, F.: Participation of the histone-like protein HU and of IHF in minichromosomal maintenance in *Escherichia coli*. Gene., 1991, 103: 25-30.

19. Kim, Y., Watrud, L. S. and Matin, A.: A carbon starvation survival gene of *Pseudomonas putida* is regulated by sigma 54. J. Bacteriol., 1995, 177: 1850-1859

20. Hincheung, L., Yannpyng, S. T., Yensheng, L. and Chen-pei, D.T.: A molecular genetic approach for the identification of essential residues in human glutathione S-transferase function in *Escherichia coli*. J. Biol. Chem., 1995, 270: 99-109.

21. Jordi, C., Paul, C., Gerard, M., Susana, C., Ben, A. and Jordi B.: The *Leptospira interrogans lexA* gene is not autoregulated. J. Bioteriol., 2005, 186: 5841-5845.

22. Joann, H., Patti, M., Deborah, L., Mullen., Joseph T., Genshi Z., Timothy, I. M., Thalia, I. N., and Richard, J.: Gene disruption studies of penicillin-binding proteins 1a, 1b, 2a in *Streptococcus pneumoniae*. J. Bioteriol., 1999, 181: 6552-6555.

23. Greated, A., lambertsen, L., Williams, P. A. and Thomas, C. M.: Complete sequence of the IncP-9 TOL plasmid pWW0 from Pseudomonas putida. Environ. Microbiol., 2002, 4: 856-871.

24. Harayama, S., Leppik, R. A., Rekik, M., Mermod, N., Lehrbach, P. R., Reineke, W. and Timmis, K. N.: Gene order of the TOL catabolic plasmid upper pathway operon and oxidation of both toluene and benzyl alcohol by the *xylA* product. J. Bacteriol., 1986,

167: 455-461.

25. http://frodo.wi.mit.edu/cgi-bin/primer3 /primer3.cgi.

26. Stefan, C., Alexandra, B. and Juraj, K.: relative quantification of mRNA: comparison of methods currently used for real-time PCR data analysis. BMC Molecular Biology, 2007, 8: 113. http://www. Biomedcentral.com/1471-2199/8/113.

27. Rudi, K., Nogva, H. K., Naterstad, K., Dromtorp, S. M., Bredholt, S. and Holck, A.: Subtyping *Listeria monocytogenes* through the combined analyses of genotype and expression of the *hlyA* virulence determinant. J. Appl. Microbiol., 2003, 94: 3862-3871.

28. http://medgen.ugent.be/~jvdesomp/genorm/.

29. Ohl, F., Jung, M., Xu, C., Stephan, C., Rabien, A., Burkhardt, M., Nitsche, A., Kristiansen, G., Radonic, A. and Jung, K.: Gene expression studies in prostate cancer tissue: which reference gene should be selected? J. Mol. Med., 2005, 83: 1014-1024.

30. chen, J., Rider, D. A. and Ruan, R.: Identification of valid housekeeping genes and antioxidant enzyme gene expression change in the aging rat liver. J. Gerontol., 2006, 61: 20-27.

31. Ohl, F., Jung, M., Radonic, A., Sachs, M., Loening, S. and Jung, K.: Identification and validation of suitable endogenous reference genes for gene expression studies of human bladder cancer. J. Urol., 2006, 175: 1915-1920.

32. Andersen, C. L, Jensen, J.L. and Orntoft, T. F.: Normalization of real-time quantitative reverse transcription-PCR data: a modelbased variance estimation approach

to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res., 2004, 64: 5245-5250.

33. Haller, F., Kulle, B., Schwager, S., Gunawan, B., Sultmann, H. et al: Equivalence test in quantitative reverse transcription polymerase chain reaction: confirmation of reference genes suitable for normalization. Anal. Biochem., 2004, 335: 1.

34. Pfaffl, M. W., Tichopad, A., Prgomet, C. and Neuvians, T. P.: Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper—Excel-based tool using pair-wise correlations. Biotechnol Lett., 2004, 26: 509.

35. Szabo, A., Perou, C. M., Karaca, M., Perreard, L., Quackenbush, J. F. and Bernard, P. S.: Statistical modeling for selecting housekeeper genes. Genome Biol., 2004, 5: R59.

36. McGann, P., Ivanek, R., Wiedmann, M., and Boor, K. J.: Temperature-dependent expression of *Listeria monocytogenes* internalin and internalin-like genes suggests functional diversity of these proteins among the listeriae. Appl. Environ. Microbiol., 2007, 73: 2806-2814.

37. Biederman, J., Yee, J. and Cortes, P.: Validation of internal control genes for gene expression analysis in diabetic glomerulosclerosis. Kidney Int., 2004, 66: 2308-2314.

38. Kim, S. and Kim, T.: Selection of optimal internal controls for gene expression profiling of liver disease. Biotechniques., 2003, 35: 456-458.

39. Tricarico, C., Pinzani, P., Bianchi, S., Paglierani, M., Distante, V., Pazzagli, M.,

Bustin, S. A., and Orlando, C.: Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. Anal. Biochem., 2002, 309: 293-300.

Chapter 4

Influence of Normalization on Target Genes in *Pseudomonas putida* mt-2

Proving the effectiveness of our results for accurate relative quantification

Abstract

In determining relative gene expression by quantitative measurements of mRNA levels using real-time quantitative PCR, internal standards such as reference genes are essential. In this study, we used geNorm method to analyze the stable expression of eight reference genes (*rpoN, rpoD, dbhA, phaF, 16S rRNA, gst, lexA, and atkA*) in *Pseudomonas putida* mt-2 during degradation of *p*-xylene. According to their expression stability. The method revealed that the *rpoN, rpoD, 16S rRNA,* and *atkA* genes were suitable reference genes, while the *phaF* and *dbhA* genes showed unstable expressed genes. The levels of expression of target genes, *xylA* and *xylE*, normalized with unstable reference gene alone, *phaF* or *dbhA,* showed significant different behavior compared with those normalized with the normalization factor (NF_4) obtained from the four suitable reference genes (*rpoN-rpoD-16S rRNA-atkA*) using geNorm software. In addition to over- or underestimation of target gene expression, (i) the delay in maximal gene expression, (ii) the increasing gene expression without inducer, (iii) the highest relative expression of *xylA* and *xylE* at the same time were observed using unstable reference genes, though it is experimentally shown that the expression of the two genes reach a maximum differently. Our study indicates that a valid set of reference genes covering a broad expression range is recommended as a normalizer to have an accurate relative quantification of the target gene(s) transcript in many microbial processes.

4.1. Introduction

Many Gram-negative bacteria can utilize toxic aromatic compounds as the only carbon and energy sources. Typically, when *Pseudomonas putida* mt-2 encounters *p*-xylene in the medium, the aromatic compound is sensed both as a growth substrate to metabolize (1) and as an environmental stressor to endure (2, 3) *Pseudomonas putida* mt-2 has the TOL plasmid that encodes the enzymes for the oxidative catabolism of *p*-xylene. The compound is degraded by the progressive oxidation of a methyl side chain to carboxylic acid, followed by oxygenative cleavage of the aromatic ring and finally changed into the Krebs cycle by means of the products of lower operon (4, 5). Xylene oxygenase encoded by the *xylA* gene is the key enzyme in the upper degradation pathway for *p*-xylene. The first step from *p*-xylene to *p*-methylbenzyl alcohol catalyzed by the *xylA* gene product is the key step in the degradation pathway. Catechol 2,3-oxygenase, the product of *xylE*, is the enzyme in the lower degradation pathway for *p*-toluate that is produced from the upper degradation pathway (6).

A recent development, real-time fluorescence-based RT-PCR, integrates the amplification and analysis steps of the PCR, and its sensitivity, specificity, and wide dynamic range make it the method for quantitative steady-state mRNA levels. (7).

However, variability in protocols used for samples acquisition and RNA template isolation can introduce errors into the analysis process. Gene expression studies in

bioremediation have not been investigated in sufficient detail to date. For that purpose, the relative quantification of reverse transcription-PCR (RT-PCR) data is the method of choice to determine gene expression (7, 8).

This method is based on the normalization of the target gene expression on any stably expressed internal reference gene, a so-called reference gene, measured in the same biological material. A crucial problem involved here is finding such a suitable reference gene, which has to be tested and verified under defined study conditions.

For quantitative mRNA studies, choosing a valid internal control for monitoring intersample variation is mandatory. In literature, several single reference genes or reference gene indexes summarizing two or more reference genes have been used for relative quantification(9-12). The use of multiple control genes results in a much more accurate and reliable normalization of gene expression data.

In this study, our aim is to identify the valid set of reference genes for accurate normalization were illustrated using two target genes, *xylA* and *xylE*, which are well-studied genes involved in *p*-xylene degradation and analyze the effects of using the different normalization approaches on gene of interest.

4.2. Experimental Section

4.2.1. Bacterial Strain and Growth Conditions

Pseudomonas putida mt-2 is the original isolate of *P. putida* bearing the archetypical TOL plasmid. (13). Japan Collection of Microorganisms RIKEN BioResource Center provided the bacteria. Growth conditions and procedures were introduced in the experimental section of chapter 2. The optical density at 600 nm (OD_{600}) was measured for the bacterial growth and at designated time point,

4.2.2. Chemical and Analysis of *p*-xylene

P-xylene (>98% pure) was used as the carbon source and obtained from Wake Pure Chemical Industries, Ltd. Gas chromatography-mass spectrometry (GC-MS) analysis of *p*-xylene concentration was carried out with a Hewlett-Packard 5973 mass spectrometer connected to a 6890 gas chromatography fitted with a fused silica capillary column (PH-5; 0.25 by 30m; film thickness, 0.25μm). The following conditions were used for GC: 2.0 ml of high-purity helium per min, on-column injection mode; oven temperature, 50°C for 1min; thermal gradient 40°C/min to 170°C, where it was held for 5 min, then increased at 10 °C/min to 280°C, and then held at 280°C for 1 min. 1 ml of aqueous

samples were extracted by 1 ml hexane. The hexane layer was removed and 1μl sample was injected into GC-MS to analyze.

4.2.3. RNA Extraction

The extraction process of total RNA and an RNAspin Mini RNA Isolation kit was introduced in the experimental section of the chapter 2.

4.2.4. cDNA Synthesis

The synthesis procedure of cDNA and an Exscript® RT regent kit were introduced in the experimental section of chapter 2.

4.2.5. Real-time PCR Assays

Primer3 (14) was used to design the primers of target genes. The primers were obtained from Invitrogen Japan (Tokyo, Japan) (Table 4-1). PCR amplification and analysis were performed using a LightCycler instrument (Roche, Mannheim, Germany), software version 3.5 (Roche Diagnostics) and SYBR® Premix Ex Taq™ (Takara Bio Inc.) according to the manufacturer's recommendations. Briefly, the final PCR mix

included 0.2 μM each primer, 2 μl diluted cDNA, and 10 μl of SYBR® Premix Ex Taq™ (2×) in a final volume of 20 μl. Cycling conditions consisted of heating to 95°C for 10 s, followed by 45 cycles of 5 s at 95°C, and 20 s at 60°C. Finally, melting curve analysis and gel electrophoresis were performed. Two results were in agreement. Only primers sets that produce a single melting peak and a single band of expected size on gel were used for further analysis.

Table 4-1. Characteristics of target genes-specific real-time PCR assays

Name	5'-3' Primer sequence	Amplicon (bp)	PCR efficiency
xylA	Forward: cagccgtttctgcttact	132	0.97
	Reverse: tatcagtccggctatcgt		
xylE	Forward: agcatcctcatccacaac	113	0.93
	Reverse: gccgtgtctatctgaagg		

4.2.6. Data Analysis

The calculation of the amplification efficiency (E) for each gene was introduced in

the experimental section in chapter 3.

To compare the amounts of the cDNA sample and RT-control sample, ΔCp was calculated. As ΔCp of >4 was found in this study, the residual DNA level was regarded as negligible (15).

To determine the differences between unstable genes and the normalization factor, NF_n, which is the geometric mean of n control genes (14), we calculated the discrepancies using the following formula:

$$difference\ (\%) = \frac{Gene_u - NF_n}{NF_n} \times 100$$

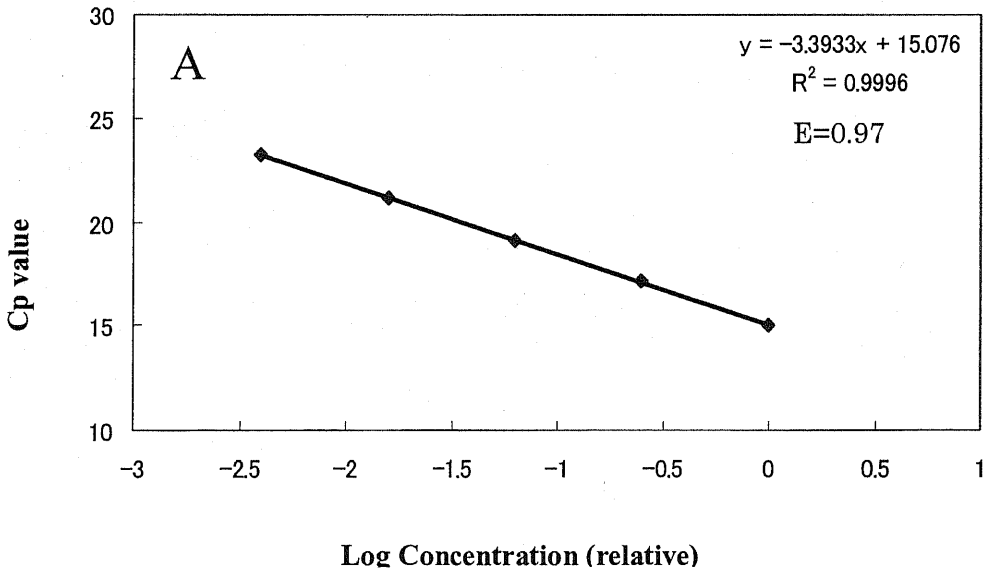
In the formula, difference (%) is the percentage difference between ($Gene_u$ - NF_n) and NF_n, and the terms $Gene_u$ and NF_n are unstable gene and normalization factor, respectively.

The normalized target gene expression levels can be calculated by dividing the target gene by the normalization factor (NF_4) or single stable reference gene expression level or unstable reference gene expression level for each sample according to geNorm software manual (16).

4.3. Results and Discussion

4.3.1. Standard Curve for Target Genes

Relative standard curves of the analyzed target genes are shown in Fig.4- 1 and were generated by using logarithmic values of standard RNA amounts vs the crossing point (Cp) that is the cycle number when the threshold fluorescence is reached. The coefficients of determination (R^2) for *xylA* and *xylE* were 0.9996 and 0.9971, respectively. Amplification efficiency (E) representative for each reference gene was calculated from equation 6 in Experimental Section of chapter 3. Detection and quantification were linear over the range of the target gene crossing point examined. The result showed that these standard curves could be used for relative quantification.



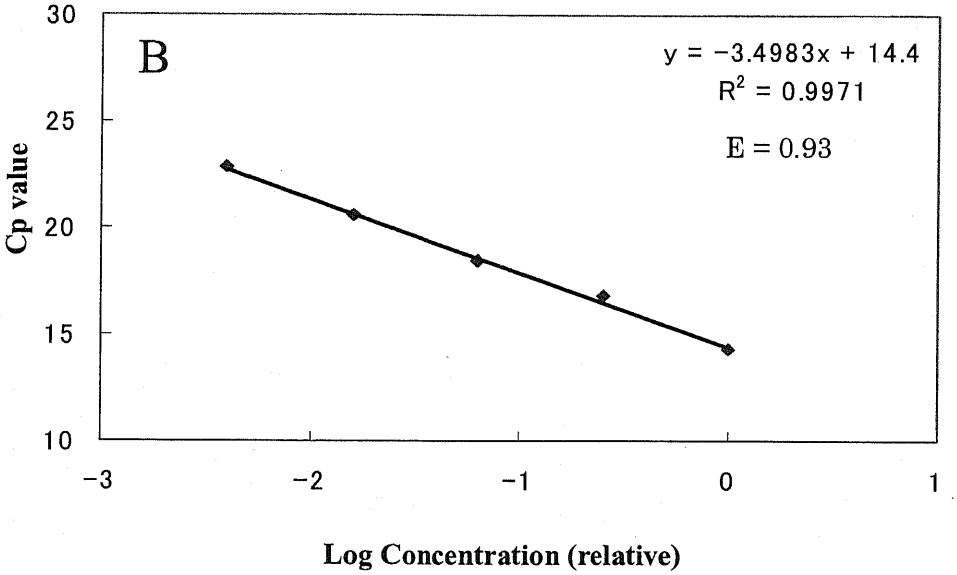


Fig. 4-1. Relative standard curves for (A) *xylA* and (B) *xylE*. Quantities of standard RNA were expressed as dilution factors of RNA preparation (1, 0.25, 0.0625, 0.015625, 0.0039062).

4.3.2. Standard Curve for *p*-xylene

A series solution of *p*-xylene from 2ppm to 500ppm was prepared for making the standard curve of *p*-xylene and these concentrations were monitored by GC-MS. The results showed strong linear relationship (0.9979, Fig. 4-2) and can be used as the standard curve.

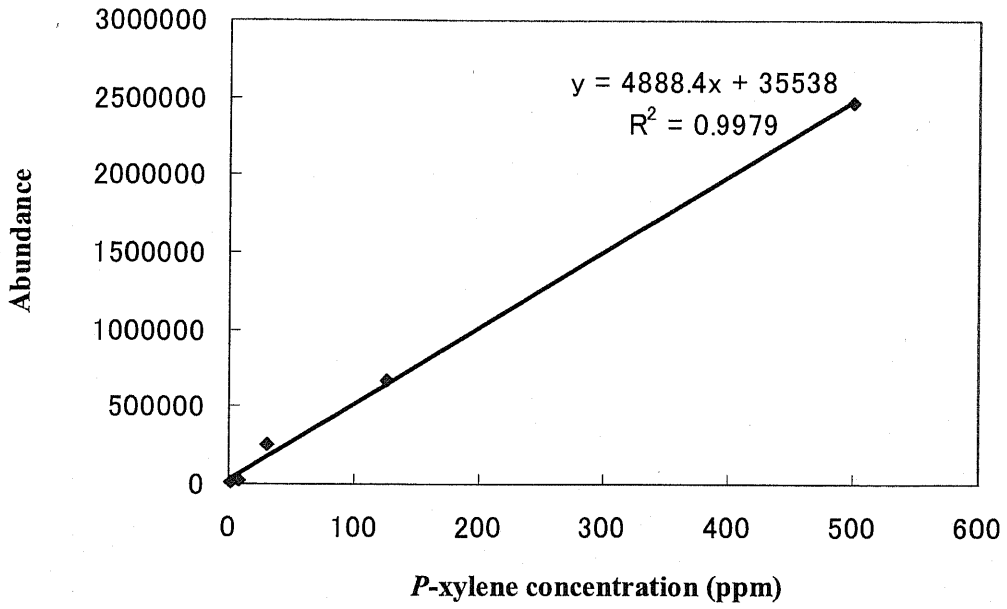


Fig. 4-2. Standard curve of *p*-xylene.

4.3.3. Growth Bacteria and Degradation of *p*-xylene

Figure 4-3 shows the growth curves of *Pseudomonas putida* mt-2 and the change of *p*-xylene concentration and their blanks. It is well known that there are four stages in bacterial growth: lag phase, exponential growth phase, stationary and death phases. When *p*-xylene was added, OD_{600} decreased at the beginning of growth, and then OD_{600} quickly increased and reached into 1.1 at about 10 hours. The reason is that *p*-xylene is toxic for the bacteria. And then the bacteria adapted for p-xylene and degraded *p*-xylene to grow. After 10 hours, OD_{600} didn't significantly change. Because the growth curve

without *p*-xylene did not significantly change, it indicated *p*-xylene is sole carbon source.

During the initial experimental period, the concentration of *p*-xylene in medium showed quickly increasing because *p*-xylene entered the medium from cotton with *p*-xylene, and then decreasing after degradation. The complete degradation time of *p*-xylene was about 7 hours. The concentration of *p*-xylene in medium didn't significantly show any change without bacteria. So, the vial can be considered as sealed.

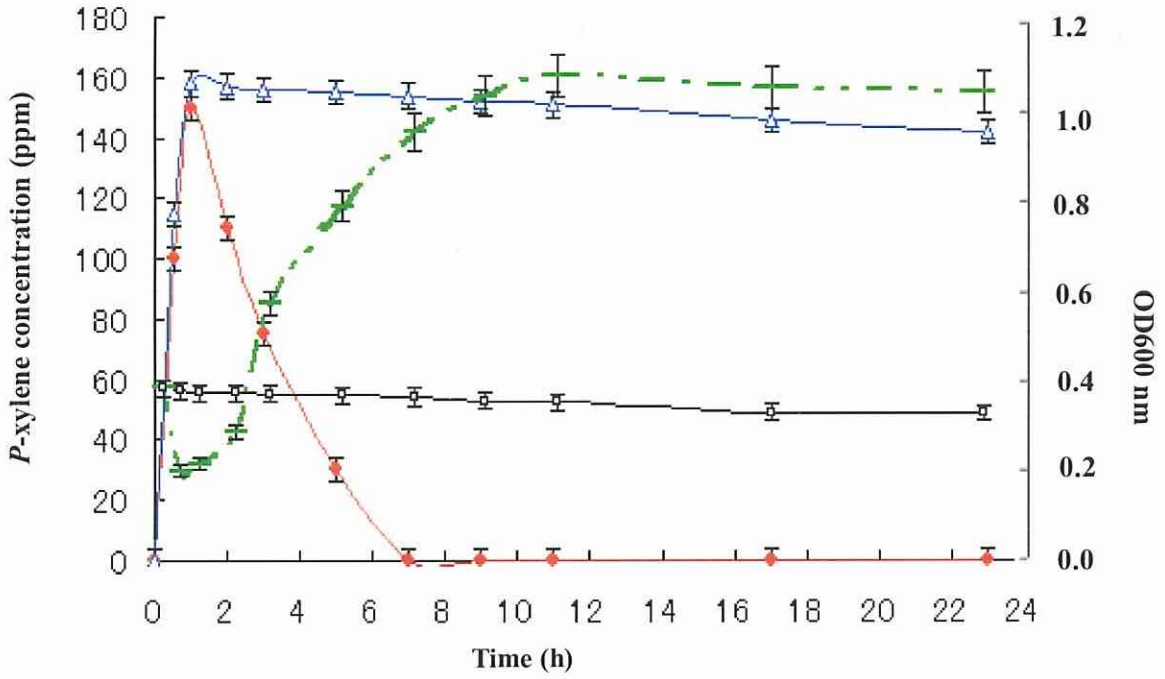


Fig. 4-3. Growth curve and *p*-xylene concentration. The left y-axis represents *p*-xylene concentration, with bacteria (red curve and close circles) and without bacteria (blue curve and open triangles). The right y-axis represents the growth curve, with *p*-xylene (green curve and the dashed line) and without *p*-xylene (black curve and open squares). The error bars show standard deviation.

4.3.4. Difference between NF_4 and Unstable Reference Genes

To compare the difference between normalization factors (NF_4) and unstable reference genes, we calculated the discrepancies depending on the formula in the experimental section. The discrepancies are shown in Table 4-2. The table clearly shows the large differences between NF_4 and unstable genes alone, *dbhA* or *phaF*. The wide range of minimum and maximum differed markedly between –62.87% and +66.86%, and these values indicated that if gene expressions were normalized to an unstable reference gene, the target gene expressions would be over- or underestimated. The error data was caused by unstable reference gene. This supported the extensive variation in expression of various unstable reference genes. Therefore, validation of reference gene is absolutely important for accurate gene expression quantification during bacterial growth.

Table 4-2: Difference (%) between normalization factor (NF_4) and unstable reference genes (*phaF* and *dbhA*)

Time (h)	0	1	2	3	5	7	9	11	17	23
Difference (%)[a]	–35.41	–9.02	+7.35	–16.17	–62.87	–57.18	–46.23	–42.89	–37.33	–37.23
Difference (%)[b]	–56.99	–37.90	–22.75	+19.50	+52.41	+52.67	+66.86	+22.53	–54.93	–56.88

+ and – represent over and under normalization factor (NF_4), respectively.

a: percentage difference between (*phaF* – NF_4) and NF_4

b: percentage difference between (*dbhA* – NF_4) and NF_4

4.3.5. Expression Levels of *xylA* Gene using NF_4 and Single Stable Reference Genes as Normalizer

In order to discuss the difference between normalization factor (NF_4) and single stable reference gene, *rpoN* or *16S rRNA,* as normalizer, the expression levels of *xylA* gene after normalizing to three different strategies is shown in Fig. 4-4. The tendency of three curves is similar. However, compare to normalization factor (NF_4), the expression levels of *xylA* gene using *rpoN* or *16S rRNA* as a control gene give significantly different levels, average decreasing 18% or increasing 34%, respectively. These results clearly provided evidence that a conventional normalization strategy based on a single

reference gene leads to erroneous normalization and reveal the need for a more reliable normalization method. Therefore, we determine to use four stable reference genes for normalization of gene expression levels in *Pseudomonas putida* mt-2 during degradation of *p*-xylene.

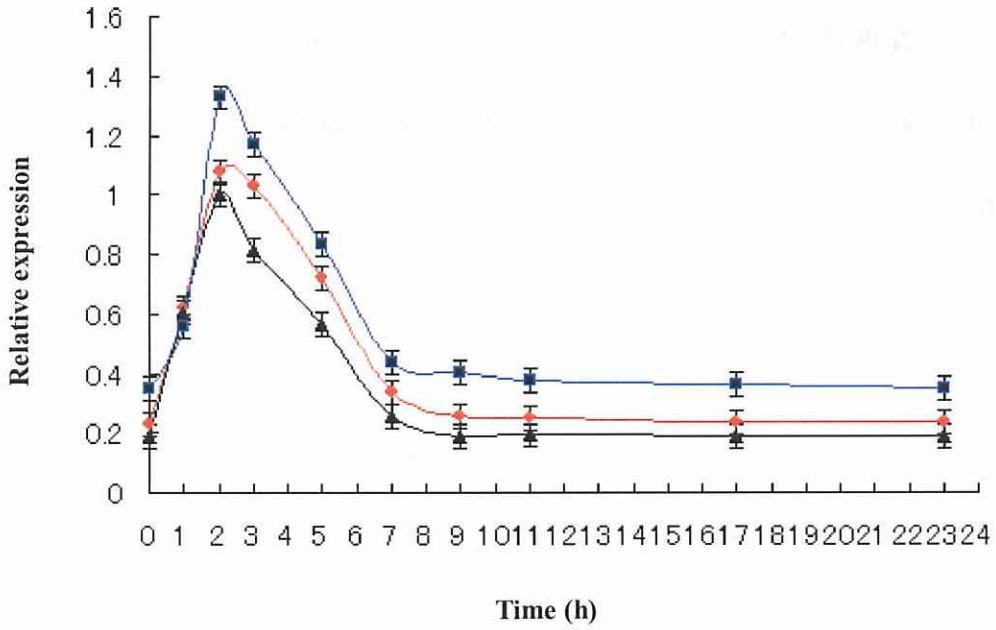


Fig. 4-4. Relative expression of *xylA* gene with normalization factor (NF_4) compared to single stable reference gene. Three lines represent relative expression of target gene depending on three different normalization strategies [red curve and closed circles, *xylA*/NF_4; black curve and closed triangles, *xylA*/*rpoN*; blue curve and closed squares, *xylA*/*16S rRNA*]. The error bars show standard deviation.

4.3.6. Relative Quantification of Target Gene *xylA* Normalized to NF_4 and Unstable Reference Gene

Many studies have only concentrated on over- or underestimated target gene expression in eukaryote (17, 18), not on discussing the relative quantification during bacterial growth phase, to describe the validation of reference genes in prokaryote. In the present study, in accordance with the geNorm software manual, we calculated relative expression levels normalized to NF_4 and the unstable genes *dbhA* and *phaF* under given physiological conditions to illustrate the importance of selecting suitable reference genes when using gene expression analysis techniques.

Rapid depletion of *p*-xylene in the culture medium was observed at 7 h (Fig.4-2). Figure 4-5 shows that three relative expression levels of target gene normalized to NF_4 or unstable genes, *dbhA* and *phaF*, increased rapidly because target gene, *xylA*, is the first gene in upper operon of TOL plasmid of *Pseudomonas putida* mt-2, and can be induced immediately after cell encountered *p*-xylene and the gene expression level increased quickly (19). For normalizing to NF_4 and the unstable gene, *dbhA*, both relative expression levels reached their maximums at early-exponential growth, then declining and finally return to their background levels. The relative expression after the depletion of *p*-xylene did not increase with normalization to NF_4, whereas in the case of normalizing to the unstable gene, *dbhA*, the relative expression increased after the

depletion of *p*-xylene at early-stationary phase was surprisingly found in the same experiment. The phenomenon was not consistent with the reality that gene expression does not increase when there is no inducer. The reason was the largest amount of *xylA* transcript appeared at early-exponential phase, but the largest expression level of *dbhA* gene showed later at mid-exponential growth as previously research (20). After entering stationary phase, the expression level of *xylA* gene was lower and changed little, but *dbhA* gene expression level was still high and decreased quickly. Thus, relative expression (*xylA*/*dbhA*) was the lowest at early-stationary phase, and then increased again with decreasing *dbhA* gene expression level.

With normalization to the unstable gene, *phaF*, the relative expression reached a maximum at mid-exponential growth. Subsequently, a rapid decrease was observed. That is, maximal relative gene expression with normalization to the unstable gene *phaF* was about 3 h later than that with normalization to NF_4. The result was contrary to the results of a previous study regarding the maximal gene expression at the initial stages of log phase (19, 21). Because the expression levels of *phaF* gene and *xylA* gene simultaneously reached their highest levels at early-exponential growth, the relative expression did not reach a maximum. At mid-exponential growth, *phaF* gene expression reduced to a very low, but *xylA* gene expression was still high. The result caused the relative expression reached a maximum at this time. Thereafter, the relative expression would decline with decreasing xylA gene expression level.

As mentioned above, the behavior of target gene after normalizing to NF_4 changed regularly as the earlier studies (19, 21), whereas that of target gene after normalizing to *dbhA* gene or *phaF* gene did not, showing delay in maximum relative expression or increasing gene expression without inducer. These results demonstrated that the four reference genes, *rpoN*, *rpoD*, *16S rRNA*, *atkA*, were valid as normalizer, and the two reference genes, *phaF* and *dbhA*, were invalid as normalizer. The results were also consistent with the above findings by the geNorm software.

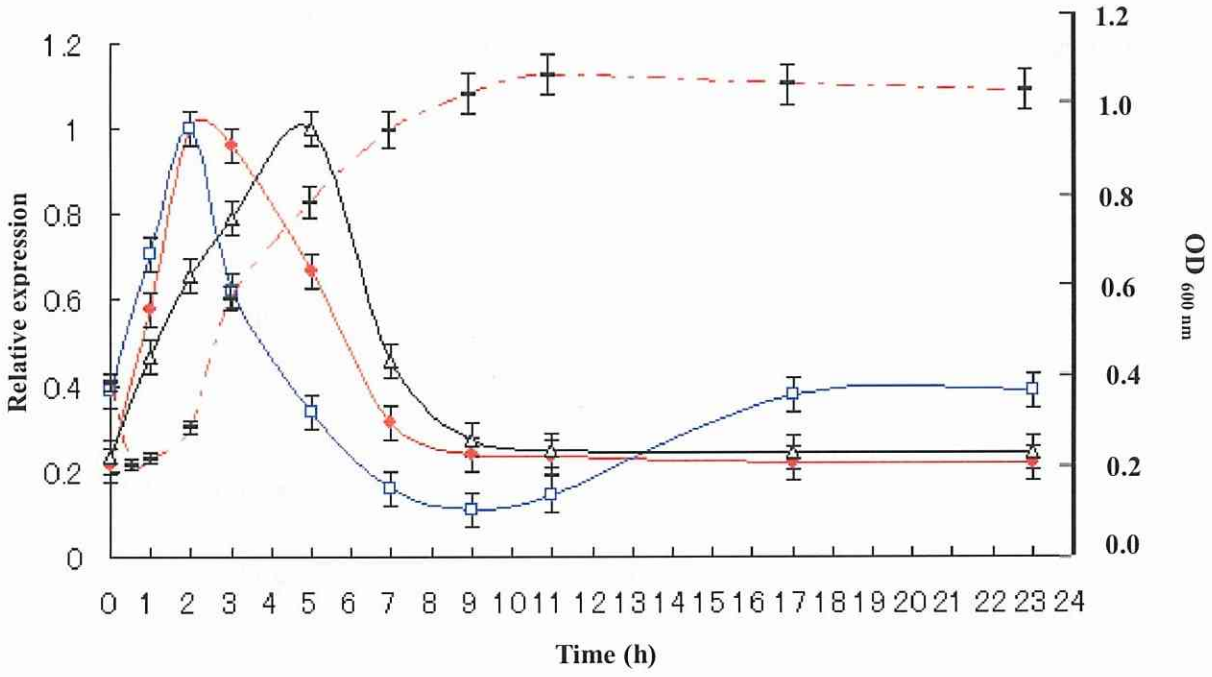


Fig. 4-5. Growth curve and relative expression of *xylA* gene depending on different normalization approaches during the degradation of *p*-xylene. The each highest relative expression is set to 1. The left y-axis, three solid lines represent relative target gene expressions of different normalization to approaches [red curve and closed circles, *xylA*/NF_4; blue curve and open squares, *xylA*/*dbhA*; black curve and open triangles, *xylA*/*phaF*]. The dashed line, referring to the right y-axis, represents the growth curve. The error bars show standard deviation.

4.3.7. The Effect of the Two Target Genes Normalizing to NF_4 and *phaF* Gene

Figure 4-6 shows the two target genes, *xylA* and *xylE*, normalized to NF_4 and unstable gene *phaF*. For normalizing to NF_4, the maximal relative expression of *xylA* gene was earlier than that of *xylE* gene. The result was consistent with the previous described (19). It is caused by two reasons. First, two target genes, *xylA* and *xylE*, were induced by different substrates, *p*-xylene and *p*-toluate, respectively. The original sources of two substrates were different. A mount of *p*-xylene was directly put into the mediate at the initial time. However, the *p*-toluate was the intermediate that was produced from the upper degradation of *p*-xylene. Second, *xylA* gene and *xylE* are in the upper and lower operon, respectively and two operons in TOL plasmid were separated by several thousands base pairs. Therefore, the expression processes of two target genes were different. For normalizing to unstable reference gene *phaF*, both maximal relative expressions of *xylA* gene and *xylE* gene were found at the same time (5h). The reason is *phaF* gene expression was very low at the time, but *xylA* gene and *xylE* gene expression were still high. The result implied the maximal relative expression of other target genes in the TOL plasmid would be found at the same time during the degradation of *p*-xylene. The variation of unstable reference genes could cause confusing, even misleading explanation of gene expression data. Therefore, the relative expression relying on the reference genes chosen describes the significance of validation of reference genes in

degradation.

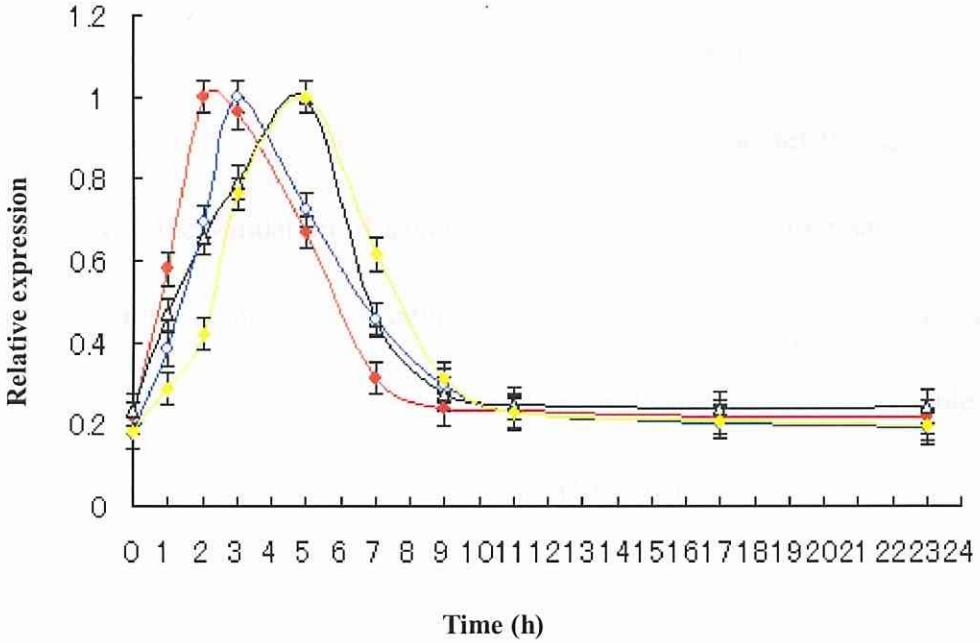


Fig. 4-6. Relative expressions of target genes, *xylA* and *xylE*, normalizing to NF4 and *phaF* gene during the degradation of *p*-xylene. The each highest relative expression is set to 1. Four lines represent relative target gene expressions [red curve and closed circles, *xylA*/NF_4; black curve and open triangles, *xylA*/*phaF*; blue curve and open circles, *xylE*/NF_4; yellow curve and closed diamonds, *xylE*/*phaF*]. The error bars show standard deviation.

4.4 Conclusions

In gene expression researches, relative quantification used different strategies to estimate quantitative real-time reverse transcription-polymerse chain reaction (RT-PCR) data. One of the major prerequisites associated with the relative quantification of target gene is the validation of suitable reference genes. It is necessary to evaluate the effect different strategies on target genes in bacteria growth. In this study, we investigated the difference between normalization factors (NF_4) and unstable reference genes, and comparing normalization factors (NF_4) with one stable reference gene and the relative expressions of *xylA* gene and *xylE* gene to analyze the validation of eight reference genes *in P. putida* mt-2 during degradation of *p*-xylene. Our results reveal using an unstable gene as a normalizer may not only show the wrong target gene expression, such as, delay in the target gene maximum relative expression at mid-exponential phase or still having the increasing relative expression with no inducer, but also show confusing the relationship of gene expressions among all target genes in the bacterium. These erroneous results would be caused by the lack of validation of reference genes, indicating the effects of unstable reference genes on the analysis of target gene expression. In addition, the conventional normalization strategy based on a single reference gene results in the error normalized data. The results also show the use of more than one reference genes has been proposed for normalization because of the

obvious reference gene expression variations in some experiment. A reliable set of normalizing genes covering a wide range of expression appears a potential improved advantage in gene profiling studies of many microbial processes for accurate relative quantification and normalization purposes.

REFERENCES

1. Assinder, S. J. and Willams, P. A.: The TOL plasmids: determinants of the catabolism of toluene and xylenes. Adv microb Physiol., 1990, 31: 1-69.

2. Ramos. J. L., Gallegos, M. T., Marques, S., Ramos-Gonzalez, M. I.,Espinosa-Urgel, M., and Sequra, A.: Responses of gram-negative bateria to certain environmental stressors. Curr Opin Microbiol., 2001, 4: 166-171.

3. Ramos, J. L., Duque, E., Gallegos, M. T., Godoy, Rojas, A., Teran, W., and Segura, A.: Mechanisms of solvent tolerance in gram-negative bacteria. Annu Rew Microbiol., 2002, 56: 743-768.

4. Wong, C. L. and Dunn, N. W.: Transmissible plasmid coding for the degradation of benzoate and *m*-toluate in *Pseudomonas arvilla* mt-2. Genet. Res., 1974, 23: 227-232.

5. Worsey, M. J. and Williams, P. A.: Metabolism of toluene and xylenes by *pseudomonas putida* (*arvilla*) mt-2: evidence for a new function of TOL plasmid. J. Bacteriol., 1975, 124: 7-13.

6. Burlage, R. S., Hooper, S. W. and Sayler, G. S.: MINIREVIEW: The TOL (pWW0) catabolic plasmid. Appl. Environ Microbiol., 1989, 55, 1323-1328.

7. Bustin SA.: Absolute quantification of mRNA using rea-ltime reverse transcription polymerase chain reaction assays. J Mol Endocrinol., 2000, 25: 169-193.

8. Rasmussen R: Quantification on the LightCycler. In Rapid Cycle Real-time PCR,

Methods and Applications Edited by: Meuer S, Wittwer C and Nakagawara K. Heidelberg, Springer Press; 2001, 21-34.

9. Vandesompele, J., De Preter K., Pattyn. F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F.: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002, 3: RESEARCH0034.1- RESEARCH0034.11.

10. Bustin, SA.: Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J. Mol. Endocrinol., 2002, 29: 23-39.

11. Schmid, H., Cohen, CD, Henger, A., Irrgang, S., Schlondorff, D. and Kretzler, M.: Validation of endogenous controls for gene expression analysis in microdissected human renal biopsies. Kidney Int., 2003, 64: 356-360.

12. Ohl, F., Jung, M., Xu, C., Stephan, C., Rabien, A., Burkhardt, M., Nitsche, A., Kristiansen, G., Radonic, A. and Jung, K.: Gene expression studies in prostate cancer tissue: which reference gene should be selected? J. Mol. Med., 2005, 83: 1014-1024.

13. Greated, A., lambertsen, L., Williams, P. A. and Thomas, C. M.: Complete sequence of the IncP-9 TOL plasmid pWW0 from Pseudomonas putida. Environ. Microbiol., 2002, 4: 856-871.

14. http://frodo.wi.mit.edu/cgi-bin/primer3 /primer3.cgi

15. Rudi, K., Nogva, H. K., Naterstad, K., Dromtorp, S. M., Bredholt, S. and Holck, A.: Subtyping *Listeria monocytogenes* through the combined analyses of genotype and

expression of the *hlyA* virulence determinant. J. Appl. Microbiol., 2003, 94: 3862-3871.

16. http://medgen.ugent.be/~jvdesomp/genorm/.

17. Chen, J., Rider, D. A. and Ruan, R.: Identification of valid housekeeping genes and antioxidant enzyme gene expression change in the aging rat liver. J. Gerontol., 2006, 61: 20-27.

18. Biederman, J., Yee, J. and Cortes, P.: Validation of internal control genes for gene expression analysis in diabetic glomerulosclerosis. Kidney Int., 2004, 66: 2308-2314.

19. Velazquez, F., Parro, V. and De Lorenzo, V.: Inferring the genetic network of *m*-xylene metabolism through expression profiling of the *xyl* genes of *Pseudomonas putida* mt-2. Mol. Microbiol., 2005, 57: 1557-1569.

20. Claret, L. and Rouviere-Yaniv, J.: Variation in HU composition during growth of *Escherichia coli*: the heterodimer is required for long term survival. J. Mol. Biol., 1997, 273: 93-104.

21. Meckenstock, R., Steinle, P., van der Meer, J. R. and Snozzi, M.: Quantification of bacterial mRNA involved in degradation of 1,2,4-trichlorobenzene by Pseudomonas sp. Strain P51 from liquid culture and from river sediment by reverse transcriptase PCR (RT/PCR). FEMS Microbiol. Lett., 1998, 167: 123-129.

Chapter 5

General conclusions

5.1. General Conclusions and Outlook

The field of present thesis was Establishing the Reference-Genes System for Accurate Quantification of Gene Expression. In this study, real-time PCR was successfully used to accurately quantify the reference genes and target genes during degradation of *p*-xylene and geNorm method was used to analyze the stability of reference genes. We obtained the results as follows:

In Chapter 1, we illustrated basic knowledge and regarding reported literature in bioremediation to better understand the present thesis. We illustrated the bacteria, *Pseudomonas putida* mt-2, and bioremediation story; comparison the conventional method of ground water cleanup and bioremediation method; Biodegradation pathway of aromatic hydrocarbons and *p*-xylene. Additional, we also described TOL plasmid that is important for bioremediation. Next, conventional quantification PCR and Real-Time quantitative PCR, and reference gene were explained and normalization software was introduced. Finally, we described the objective of the present study.

In Chapter 2, we described all RNA samples were purity and designed primer sets were very well. One melting peak and gel electrophoresis analysis demonstrated that the quantitative PCR condition is optimal. The residual DNA level is negligible because of

ΔCp values > 4. The all RNA samples can be used for the future quantitative gene expression.

Evaluation of Reference Genes in *Pseudomonas putida* mt-2 during Degradation of *p*-xylene was discussed in Chapter 3 and the conclusions were as follows:

1, The expression levels of reference genes were significant changed and no apparent correlation between gene expression stability and gene expression was shown during bacterial growth.

2, According to their expression stability, *rpoN*, *rpoD*, genes were the most stable reference genes, while the *phaF* and *dbhA* genes showed unstable expressed genes.

3. The normalization factors (NF_4) obtained from the geometric mean of four suitable reference genes (*rpoN*,-*rpoD*,-*16S rRNA*, and *atkA*) by geNorm method can be used to normalize target genes expressions for accurate quantification.

4. Four suitable reference genes (*rpoN*,-*rpoD*,-*16S rRNA*, and *atkA*) represent the levels of low, intermediate, and high gene expression.

Influence of Normalization to Different Strategies on Target Genes in *Pseudomonas putida* mt-2 was discussed in Chapter 4 and conclusions were as follows:

1, The expression levels of target genes, *xylA* and *xylE*, normalized with unstable reference gene alone, *phaF* or *dbhA,* showed significant different behavior compared

with those normalized with the normalization factor (NF_4) obtained from the four most suitable reference genes (*rpoN-rpoD-16S rRNA-atkA*) using geNorm software.

2, In addition to over- or underestimation of target gene expression, (i) the delay in maximal gene expression, (ii) the increasing gene expression without inducer, (iii) the highest relative expression of xylA and *xylE* at the same time were observed using unstable reference genes, though it is experimentally shown that the expression of the two genes reach a maximum differently.

3, A reliable set of normalizing genes covering a wide range of expression appears a potential improved advantage in gene profiling studies of many microbial processes for accurate relative quantification and normalization purposes.

Publication

Qing Chang, Takashi Amemiya, Jianbo Liu, Xiaojing Xu, Narasimmalu Rajendran and Kiminori Itoh, "Identification and Validation of Suitable Reference Genes for Quantitative Expression of *xylA* and *xylE* Genes in *Pseudomonas putida* mt-2". *J. Biosci. Bioeng.*, 2009, 107: 210-214.

International Conference Attended

1, Qing Chang, Takashi Amemiya, and Kiminori Itoh, "Identification and Validation of Suitable Reference Genes for Quantitative Gene Expression of *P*-xylene-Induced *xylA* Gene in *Pseudomonas putida* mt-2". Yokohama National University GLOBAL COE PROGRAM International Symposium on Emerging Ecological Risks and Food Security in Asia. Feb., 23, Yokohama National University, Japan. (2008).

2, Qing Chang, Takashi Amemiya, Jianbo Liu and Kiminori Itoh, "Relative Quantitative of Gene Expression During Biodegradation to Select Reliable Reference Genes". 24th Annual International Conference on Soils, Sediments and Water. Oct., 20-23, University of Massachusetts in Amherst, USA (2008).

Acknowledgements

I highly appreciate **Prof. Kiminori Itoh** and **Associate Professor Takashi Amemiya** for supervision and guidance throughout the entire period of my doctoral program. My thanks also go to **Prof. Kiminori Itoh** for giving me the good opportunity and entirely help on my coming and studying, and my living in Japan and I thank **Prof. Narasimmalu Rajendran** for discussion.

I am also grateful to all the former and present members of Prof. Itoh's laboratory for giving me a lot of useful advises and helping study and living. Especially to **Mr. Jianbo Liu, Ms. Xiaojing Xu, Mr. Guohua Liu, Ms Chengshu Lian and Mr. Kenichi Shibata.**

Finally, my heart-felt appreciation goes to my families and ChongQing Technology and Business University in China for their sincere support and constant encouragement to my study all through.