論文題目

Elucidation of the plasmid effect on the host at single-cell level
(プラスミドが宿主に及ぼす影響の一細胞レベルでの解明)

References

Acknowledgements

Supplementary materials

1. Growth curve and sampling points of each strain

2. Cell phenotype of each sampling points

論文の内容の要旨

応用生命工学専攻

平成 29 年度博士課程入学

氏名　　張 荟 婷

指導教員　野尻秀昭

論文題目

Elucidation of the plasmid effect on the host at single-cell level
(プラスミドが宿主に及ぼす影響の一細胞レベルでの解明)

Chapter 1. Introduction

Plasmids are circular or linear extrachromosomal replicons which are found in many microorganisms. Plasmids can provide functional benefits to the host such as antibiotic-resistance and novel metabolic capacity for xenobiotics. Plasmids are able to spread such functional genes horizontally among bacteria through conjugative transfer, and promote the rapid evolution and adaption of bacteria (Trevors, 1986). Conjugation is the process by which DNA is transferred from one bacterium to another via cell-cell contact or a bridge-like connection between two cells. Many studies have clarified the details of the plasmid effects on host cell physiology, suggesting that the host factors possibly affect the recipient range and conjugative transfer frequency of the plasmid (Shintani *et al.*, 2014) and that the factors expressed from transferred plasmids may affect the expression network of their host chromosome (Takahashi *et al.*, 2015). However, all of these studies were done at population level, and this may hide some variation among the individual cells. That is the averaged data cannot show the diversity of the responses in individual cells, and therefore, single cell level analysis will be necessary. Meanwhile, previous studies on crosstalk between plasmids and their host were performed with plasmid-harboring strains after long-term cultivation in which the host were presumably fully adapted to the plasmid, the interaction among the adaptation process to the conjugated plasmid is still unknown. The objective of my research is to uncover the plasmid effect on the host during plasmid adaptation process (Fig.1) at single-cell level. For this purpose, I employed two techniques, one is a combination of fluorescence-labelling system and flow cytometry, another is based on the label-free and non-invasive Raman spectroscopy.

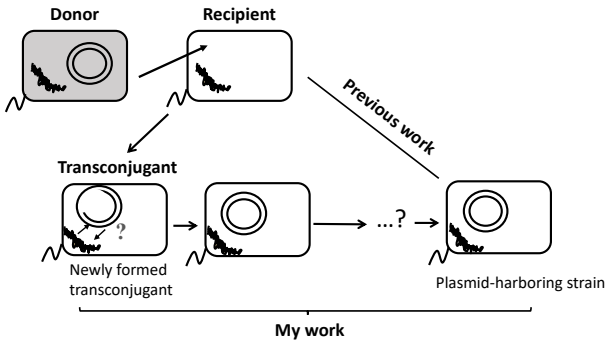


Fig. 1. Hypothetical diagram of plasmid adaptation process after conjugative transfer. Conjugation requires contact (mating process) between donor (plasmid-harboring) and recipient (plasmid-free) cells, leading to

formation of transconjugants (recipient cell acquired plasmid). Newly formed transconjugants were cultured, and then they will be “plasmid-harboring strain” after plasmid-host adaptation.

Chapter 2. Elucidation of the host adaptation process after plasmid conjugation using transcriptome analysis

In this chapter, 2-hour mating assay was performed with *P. putida* SM1443*(*pBP136*::gfp*) (donor) and *P. putida* KT2440RG (recipient). In this system, *gfp* gene conceding reporter protein GFP was encoded on incompatibility group (Inc) P-1β plasmid pBP136*::gfp* under the regulation of *lac* promoter. GFP expression was repressed by LacI in donor and was derepressed in transconjugant. Flow cytometer was introduced to detect and separate 10^7 of newly-formed transconjugant cells. Part of the collected cells (named “2-h”) were continually cultured for several days (transfer culture to fresh medium every day). Log phase of “1-day” and “3-day” cultures were collected as plasmid-adaptive-process samples. Transcriptome analysis was performed on the duplicated samples of recipient cells (KT), newly formed transconjugant cells (2-h), adaptive-process cells (1-day and 3-day) and 3-day transconjugant with undergoing cell sorting (3-day-sort) respectively. Compared to transcriptome of 3-day cells, there is a significant change in 3-day-sorted cells which indicated that flow cytometric cell sorting has effect on transcriptome analysis (Fig. 2). Initially, I want to explore host adaptation process start from 2-h transconjugant, however, due to effect of cell sorter, it was only used to compare with 3-day-sort transconjugant. A total number of 242 differentially expressed genes (DEGs) were discovered between the two samples, the mainly up-regulated genes are related to signal transduction mechanisms protein and mainly down-regulated genes are related to inorganic ion transport and metabolism (Table 1). The transferred plasmid has drastic effect on host indeed, compared to recipient, a total number of 1475 and 1404 DEGs were discovered from 1-day and 3-day transconjugants respectively. Function of differential expressed genes in the 2 combination are the similar. Meanwhile, the gene expression between 1-day to 3-day transconjugant are not changed too much, it may indicates plasmid adaptation process nearly complete, and the sampling points should be set more earlier to catch more intense period during plasmid host adaptation in the future work.

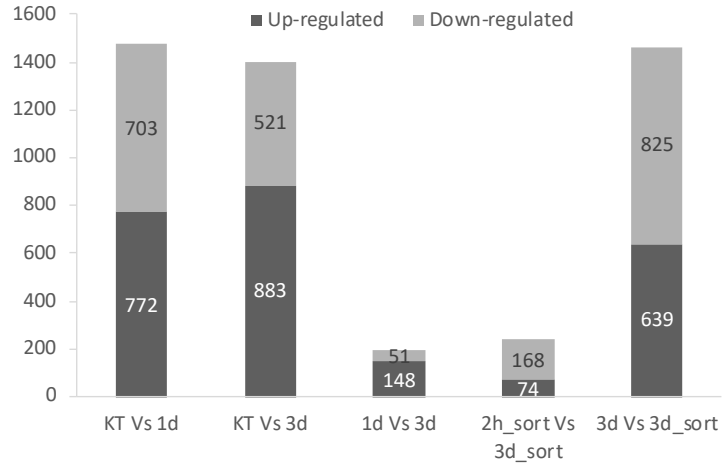


Fig. 2. Up-, down- regulated genes detected during plasmid-adaptation process samples. KT Vs 2h: gene expression of 1-day transconjugant was compared with it of KT, so do others.

Table 1. Distribution list of main differentially expressed genes based on COG classifier.

	COG	All	UP-regulated	All %	COG	Down-regulated	All %
KT Vs 1d	-	1104	91	8.2	-	172	15.6
	E	480	90	18.8	E	71	14.8
	C	274	75	27.4	S	69	16.4
KT Vs 3d	-	1104	129	11.7	-	113	10.2
	E	480	75	15.6	E	61	12.7
	C	274	67	24.5	S	55	13.1
1d Vs 3d	-	1104	63	5.7	E	14	2.9
	L	200	14	7.0	R	7	1.5
	S	420	14	3.3	-	6	0.5
2h_sort Vs 3d_sort	-	1104	22	2.0	-	33	3.0
	T	232	8	3.4	P	20	7.8
	E	480	7	1.5	C	14	5.1

[C] Energy production and conversion, [E] Amino acid transport and metabolism, [L] Replication, recombination and repair, [P] Inorganic ion transport and metabolism, [R] General function prediction only, [S] Function unknown. [T] Signal transduction mechanisms.

Chapter 3. Discrimination of plasmid-harboring and -free strains using single-cell Raman spectroscopy

Raman spectroscopy, label-free and non-invasive detection technique, has already been used in several studies on single cells(Li *et al.*, 2012). In this part, on the basis of Raman spectroscopy techniques, I developed a method to distinguish between plasmid-harboring and -free bacterial strains at the single-cell level without harming or modifying cells. *P. putida* KT2440 and *E. coli* W3110 were used as the hosts for the plasmid pB10 and plasmid RP4. Each strain with or without plasmid pB10 or RP4 was cultured in minimum medium supplied by glucose or succinate as the sole carbon source. Raman spectra of single cells from log and stationary phase were recorded, denoised using Singular Value Decomposition (SVD), and followed by classification of plasmid-harboring and -free strains using Random Forest (RF) algorithm (Lee *et al.*, 2013). For discriminating, half of spectra (30-50) of each strain were selected to constructing training dataset to building the classification model and the remaining half spectra were used to forming test dataset to validating the classification accuracy. Applicaiton of the combination of Raman spectroscopy technique and RF algorithm to *P. putida* KT2440 strains at both log and stationary phases successfully discriminated plasmid-free strain from pB10- or RP4-harboring strains (Fig. 3). On the other hand, I was also able to distinguish plasmid-free *E. coli* W3110 strain from pB10- or RP4-harboring strains with Raman spectra at log phase, but failed to distinquishing these strinas at stationary phase. I am still looking for the accurate reason.

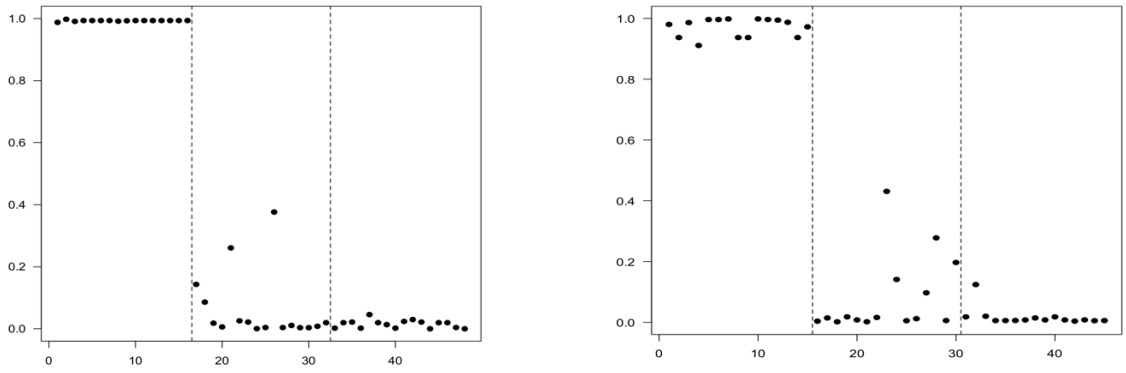


Fig. 3. Score plot of validation result from spectra of *P. putida* KT2440 strains of log-phase (left) and stationary phase (right) samples. X axis: 3 parts from left to right represent to *P. putida* KT2440, *P. putida* KT2440(RP4), *P. putida* KT2440(pB10) respectively. Y axis: reliability score of the plasmid-free, (1 means plasmid-free and 0 means plasmid-harboring). Each dot represents validation score of one Raman spectrum from a single cell.

Chapter 4. Summary and future prospect

Compared to previous study, my study is the first time to explore plasmid effect on host specially for plasmid adaptation process. I explored two techniques for biological analysis at single cell level. With the combination of flow cytometry and RNA-seq analysis, I found that flow cytometric cell sorting change transcriptome of cells with large extent. This result indicates that it can be solved by applying same treatment of each sample with cell sorter in the future. Furtherly, the earlier sampling point (lag phase before 1-day sampling) should be set to exploring host adaptation process. To this end, single (or a smaller number of) cell transcriptome technique should be combined, because current cell sorting cannot be adopted to gather enough cell numbers for RNA extraction. RNA extraction method should be modified to obtain RNA from less cells like 10^5 even 10^3 cells in the future work. On the other hand, I successfully distinguished the plasmid-harboring cells from plasmid-free cells with Raman spectra. It provides the possibility to detecting the plasmid transfer using Raman spectroscopy in future. If single cell-picking technologies, such as laser microdissection or optical tweezer, are combined, this new method can be applied to clarify the interaction between plasmid and host cell at the single cell level without any sampling preparation or modification of cells such as GFP-labeling or mutation construction.

Chapter 1. Introduction

1.1 What is plasmid

Plasmids are small circular DNA molecules that can be replicated independently outside the chromosome (Trevors, 1986). They are usually much smaller than bacterial chromosome ranging from one to several hundred kilobases and in copy number from one to hundred per cell. In general, the presence or absence of plasmids does not play decisive roles in host cell survival, but they provide one or more functional benefits to the host which promote their own and host survival, such as antibiotic-resistance, production of toxins, some carry genes that give cells extra physiological and metabolic capacity like degradation of organic compounds and even increase their pathogenicity in some bacteria. Plasmids are the most common vectors of genetic engineering.

1.2 Study on the interaction between plasmid and host

Plasmids play a central role in the adaptation and evolution of prokaryotes (Gogarten & Townsend 2005), they are able to spread function genes horizontally among bacteria through conjugative transfer. Conjugation requires contact between donor (plasmid –harboring) and recipient (plasmid-free) cells (mating process), leading to formation of transconjugants (recipient cells acquired transferred plasmid). It is the process by which DNA is transferred from one bacterium to another via cell-cell contact. The process composed two steps: DNA preparation, and mating bridge formation. DNA preparation includes the process of relaxosome formation. A relaxosome is the DNA/protein complex created by the binding of certain transfer proteins to regions within the origin of transfer (oriT).

Shintani and collogues compared the chromosomal RNA maps of three pCAR1-containing *Pseudomonas* strains with each pCAR1-free strain, found that the possession of pCAR1 altered gene expression related to the iron acquisition systems in each host, expression of the major siderophore pyoverdine was greater in two of plasmid-containing strains than in their plasmid-free host strains [1]. Takahashi and collogues compared the phenotypes and transcriptomes of the same plasmid harboring and plasmid-free strains in time-series, it was indicated that the transcriptional alterations were commonly large at the transition and early stationary phase, the extent of the alterations differed between hosts [2].

In previous research, the mechanism of plasmid transfer was explored. Plasmid contain genes conferring resistance to antibiotics and metals, the utilization of carbon compounds, virulence or symbiosis determinants(Heuer & Smalla 2012, Top et al 1995). Plasmid conjugative transfer requires the expression of transfer (tra) genes involved in DNA transfer and replication (Dtr) and in Mating pair formation (Mpf). Dtr genes are required for processing the DNA with a relaxase, which cuts the DNA molecule at the nic site of the origin of transfer (oriT), and remains attached to the single strand of DNA, which will be transferred responding to an undefined signal (Koraimann and Wagner, 2014). This complex, together with other proteins, makes up the relaxosome, which directs the strand to the type 4 secretion system (T4SS), a multiprotein complex located in the membrane of the cell. The Mpf encodes the genes that are responsible for synthesis of the T4SS. The relaxase, with the DNA strand attached, transfers to the recipient cell through the pore formed by the T4SS, finally the plasmid needs to be established in the recipient, with the help of ssDNA binding, anti-restriction and

SOS inhibition proteins to generate stable transconjugants. (Alvarez-Martinez & Christie 2009, Ding & Hynes 2009, Koraimann & Wagner 2014).

The first step in plasmid acquisition by a new host is the physical arrival of the plasmid in the cell. Conjugation is considered the most important mechanism of plasmid transmission and the conjugative process has been extensively studied.

1.3 Technique of single-cell analysis

1.3.1 Flow cytometry

Flow cytometry is a widely used for isolating specific cell populations, high-throughput method to enrich for rare cells such as hematopoietic stem cells (Radbruch and Recktenwald, 1995;Will and Steidl, 2010). However, these methods are not without drawbacks, since they can introduce artificial stress on cells and change their expression profile (Van Den Brink et al., 2017). Flow cytometry-isolated cells are exposed to high pressure during sorting and these osmotic and pressure changes introduced to cells during cell sorting and handling can induce change to the cell expression profile of multiple cell types (Xiong et al., 2002; Romerosantacreu et al., 2009; Van Den Brink et al.,
2017).

1.3.2 Raman spectroscopy

1.3. 1 Raman effect

As Fig.1 shows, Raman scattering is a light scattering effect. When a monochromatic light scattering onto a molecule, photons interact with sample and energy can be gained (anti-Stokes) or lost (Stokes), since Stokes process is normally more intense than anti-Stokes process, the conventional Raman spectroscopy record Stokes spectrum (Nakamoto, 2006). Raman spectroscopy is based on molecular vibrations, energy from light interact with molecular vibrations and release as scattered light. The shift of energy (wavelength) between incident and scattered photo corresponding to the energy excited by a particular molecular vibration, therefore, each molecule has its specific Raman spectrum (Smith et al 2016). For the cell sample, it contains a large number of different molecules, the spectra information of all components in cell requiring complex data interpretation, potential of Raman spectroscopy in cell analysis cannot be ignored.

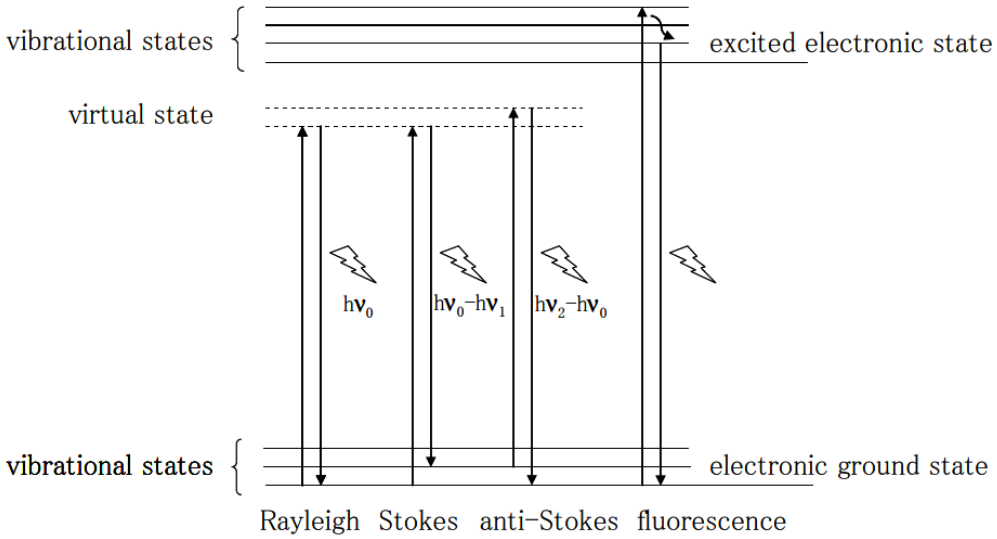


Fig.1 Diagram of different type of scattering light, when light interact with molecule, it can be excited to a virtual state and release with several ways: a. return to original state, photon energy (wavelength) is the same with incident light which is called Rayleigh scattering; b. return to a higher vibrational state, lower energy than the incident light which is called Stokes scattering; c. return to a higher energy than incident light which is called anti-Stokes scattering.

1.3.2 Raman spectra

A Raman spectrum is a diagram of Raman intensity and Raman shift (wavenumber). The relationship between energy and wavenumber is as below:

$$E = h\nu = hc/\lambda = hc\varpi$$

E = energy of light

h = Plank constant

ν = frequency of light

c = speed of light

ϖ = wavenumber of light

$$\text{Raman Shift} = 1E7 * (1/\lambda_{inc} - 1/\lambda_{Sto})$$

λ_{inc}: Wavelength of incident light

λ_{Sto}: Wavelength of Stokes scattering

Raman spectroscopy is a fast and robust technique. One of the key advantages of Raman spectroscopy is the ability to determine the underlying chemical structure of a cell: DNA, proteins and lipids which can be visualized according to their vibrational spectra without any label or stain prior to measurement (Palonpon et al 2013). The combination of Raman spectroscopy with microscopy allows non-destructive detection of single cell. Since it provides valuable biochemical information from living cells in short time, allowing for the real-time observation and monitoring of cellular processes in vitro, it makes Raman spectroscopy apart from other technique.

The confocal Raman microscope was developed in 1990, it recorded Raman spectra of single human cells (Puppels *et al.*, 1990), from then on, many similar phenomena were obtained from human cells. Some groups reported the classification of bacteria using Raman spectroscopy. Some investigation on cell components of spores of single bacterial cell. Recently, various studies on identification of bacteria were carried out with combination of Raman spectra and data classification methods; with combination of Raman spectra and linear discriminant analysis (LDA) and on artificial neural networks (ANNs). according to refere Maquelin identified unknown microbes from blood. (Maquelin *et al.*, 2003). Petra's group identified bacteria by means of a micro-Raman analysis together with support vector machine (SVM) technique.
(Rösch *et al.*, 2005; Kumar *et al.*, 2020).

Raman application

Molecular polarizability, the molecular property which control intensity.
Raman scattering is a shift from an exciting frequency and should be labelled >cm-1 but it is usually use cm-1 with the delta implied.

1.4 Thesis summary

Two kinds of visualization systems were used in my study: Raman spectroscopy, *lacI*^{q}-$P_{A1/03/04}$-*gfp* fluorescence system. Raman spectroscopy is a non-invasive, label-free technology. It can detect the vibrations of chemical bonds in molecules such as nucleic acids, protein, lipids and carbohydrates which generate strong Raman spectra signals, enabling differentiation of cells in different conditions. Here, I assessed whether Raman spectroscopy can be used to distinguish plasmid-harboring and plasmid-free strains. This study is a collaboration with Prof. Shigeto (Kwansei Gakuin University) and Dr. Matsui. At the same time, the *lacI*^{q}-$P_{A1/03/04}$-*gfp* fluorescence system is used visualization of plasmid transfer with flow cytometry. Different time-point samples of transconjugant will be collected for RNA-seq to clarify the expression pattern of the genes during the adaptation process of transferred plasmid in host.

Objective of my research

Previous research just study the interaction between plasmid-harboring and plasmid-free cells, but we don't know what happened after plasmid transfer in initial time of tansfered plasmid in host.

Chapter 2. Clarification of adaptation process between plasmid and host using *lacI*^{q}- $P_{A1/03/04}$-*gfp* fluorescence system

IncP-1β plasmid pBP136 from the human pathogen *Bordetella pertussis* is a highly transferable plasmid in various host cells. In previous report, the *Pseudomonas putida* SMDBS(pBP136::*gfp*) as a donor (A *dapB*-deleted strain of *P. putida* SM1443 carries the *lacI*^{q} transcriptional repressor of the $P_{A1/03/04}$ promoter on its chromosome) of plasmid pBP136::*gfp* (in which a $P_{A1/04/03}$-*gfp* cassette inserted into plasmid pBP136) was mated with soil bacteria, GFP-expressing transconjugants were detected and separated at the single-cell level by

flow cytometry successfully [4]. Based on this, I selected *P. putida* SM1443(pBP136::*gfp*) as donor and *P. putida* KT2440RG as recipient to detect plasmid transfer in the *lacI*^{q}-$P_{A1/03/04}$-*gfp* fluorescence system. Fluorescence expression is repressed by LacI in donor and is derepressed when pBP136::*gfp* transferred into recipient, so that the transferred plasmid will be detected with GFP fluorescence.

To detect the plasmid conjugative transfer in transconjugants immediately, I will perform time-lapse observation of GFP fluorescence after mating *P. putida* SM1443 (pBP136::*gfp*) and *P. putida* KT2440RG with flow cytometry soon, different time-point samples of transconjugant will be used for RNA-seq. Besides, microdevice and droplet technology with time-lapse imaging will also be used for real-time monitoring of plasmid conjugative transfer.

LB	Per liter
Yeast extract	5 g
Tryptone	10 g
NaCl	10 g

NMM4	Per liter
Na2HPO4	2.2 g
KH2PO4	0.8 g
NH4NO3	3.0 g
MgSO4 · 7H2O	0.2 g
FeCl3 · 6H2O	0.01 g
CaCl2 · 2H2O	0.01 g

M9	Per liter
Na2HPO4	2.2 g
KH2PO4	0.8 g
NH4NO3	3.0 g
MgSO4 · 7H2O	0.2 g
FeCl3 · 6H2O	0.01 g
CaCl2 · 2H2O	0.01 g

PBS	Per liter
Na2HPO4.7H2O	2.56 g
NaCl	8 g
KCl	0.2 g

KH2PO4	0.2 g

2.2 Flow cytometry parameters settings.

Time-course fluorescence detection of mating cultures with flow cytometry

To check if the transconjugant can be detected with *lacI*^{q} -$P_{A1/03/04}$-*gfp* fluorescence system and the plasmid transfer process can be reflected with GFP fluorescence variety, mating assay were performed, and fluorescence analysis were carried out using flow cytometry.

The filter and liquid mating assay were carried out with *P. putida* SM1443 (pBP136::*gfp*) as donor and *P. putida* KT2440RG, *E. coli* DH5α as recipients as follows:

1. All strains were pre-cultured overnight in Luria-Bertani (LB) broth with appropriate antibiotics, rifampin (Rif) and kanamycin (Km) for *P. putida* SM1443 (pBP136::*gfp*), rifampin (Rif) and gentamicin (Gm) for *P. putida* KT2440RG. Antibiotics were used at final concentrations of 25 μg/mL for Rif, 50 μg/mL for Km, 30 μg/mL for Gm.
2. 4 mL cultures were collected in 2 mL centrifuge tubes and washed with 2 mL carbon free (CF) buffer for twice;
3. OD_{600} of donor was adjusted to 0.2 and recipient to 2, equal volumes donor and recipient were mixed in 2-mL centrifuge tubes, 400 μL mixture were transferred on 0.22-μm-pore-size membrane filter (Millipore, Billerica, MA, USA) on glass microanalysis filter holders and filtering flasks (Millipore), the filters were placed on LB agar plates and incubated at 30°C for serial times.
4. Filter with cells after serially incubation time was collected and washed with 1 mL CF buffer and diluted to 10^{-1}, the resultant cultures were transferred through the gap into 5 mL round bottom polystyrene test tube (Corning, Falcon®) for fluorescence measurement using with flow cytometry MoFlo™ XDP (Beckman Coulter).
5. The resultant data was analyzed by FlowJo software.

For sample preparation: Liquid mating mixture were centrifuged and resuspended with 1 mL CF buffer, membrane filters were washed with 1 mL CF buffer, both of the resultant cultures were diluted to 10^{-1} and transferred into 5 mL round bottom polystyrene test tube (Corning, Falcon®). The resultant data was analyzed by the software of FlowJo.

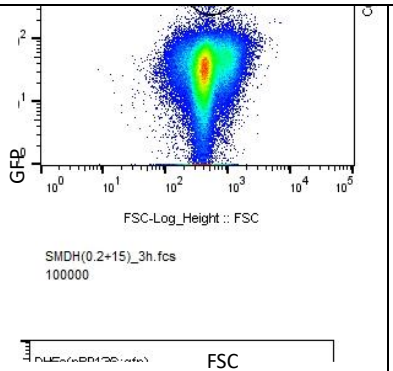
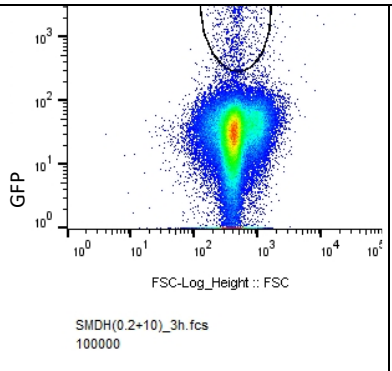
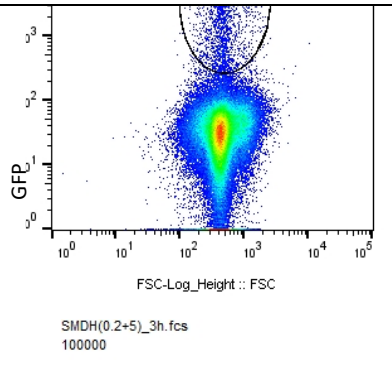
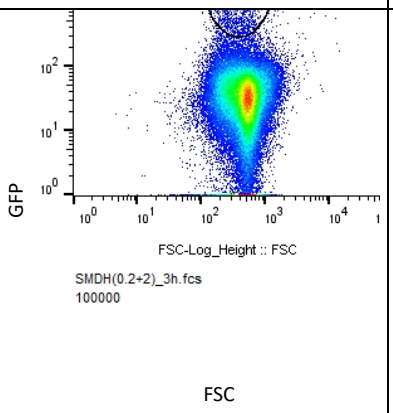
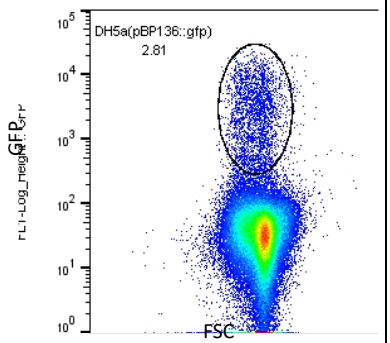
A B C

Fig. 2. Time-course fluorescence detection of filter mating cultures with flow cytometry. SM: *P. putida* SM1443 (pBP136::*gfp*), KT: *P. putida* KT2440RG, DH5α: E. coli DH5α, F: filter mating. X axis represent the intensity of GFP fluorescence, Y axis means the number of cells.

As is shown above, for filter mating assay with *P. putida* SM1443 (pBP136::*gfp*) as donor *P. putida* KT2440RG as recipient, the transconjugant can be detected from 0.4-hour mating represented with the enhanced GFP fluorescence, the plasmid conjugative transfer occurred before this time-point, the GFP intensity became stronger and stronger during the first hour mating (Fig.2A) which may represent the ·fluorescence intensity were increased with time lapse after 1-hour mating (Fig.2B**).** While for the mating with *P. putida* SM1443 (pBP136::*gfp*) and *E.coli* DH5α, it could not detect any fluorescence change maybe because of the low plasmid transfer frequency in this situation. There was not any fluorescence variety in liquid mating in both the recipient of *P. putida* KT2440RG and *E. coli* DH5α (data not shown).

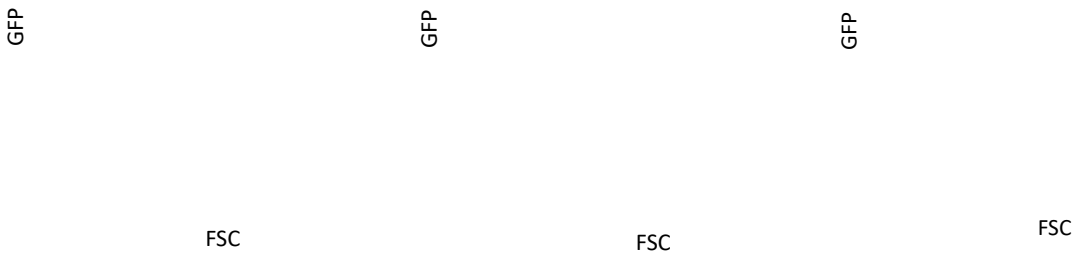
Next, I will try to detect if transconjugant can be detected earlier than 0.4 h, then repeat this experiment using *P. putida* SM1443 (pBP136::*gfp*) as donor *P. putida* KT2440RG as recipient with several replicates, add the only donor and only recipient as control, after stabilize the phenomenon, the transconjugant in different time-point will be collected for RNA sequence.

2.1.1 Optimization of mating condition

		
0.2:2	0.2:5	0.2:10
		
0.2:15	0.2:20	

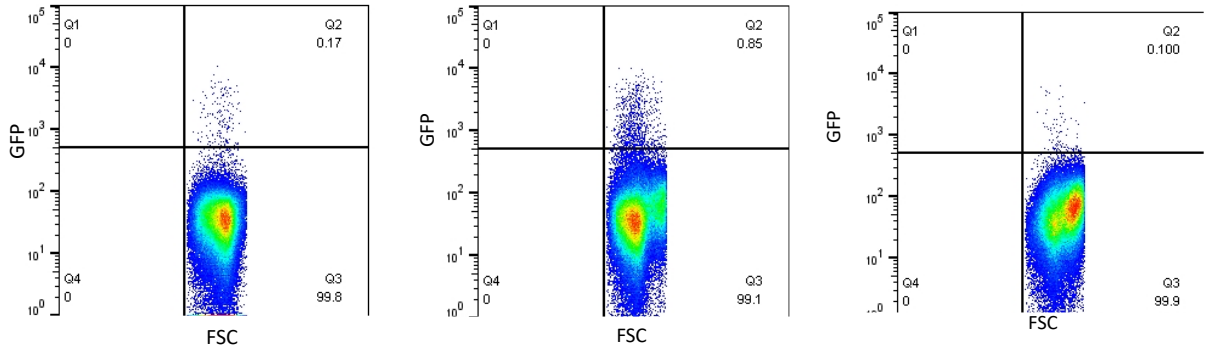
Mating assay for 3 hours.

When increase the number of recipient have no positive effect on transconjugant ratio. So I increase donor



0.2:5 2:5 2:10

At the same time, I want to see if the ratio was changed when using SM1443(pBP136::gfp) as donor and KT2440RG as recipient, higher ratio of transconjugants can be obtained or not,



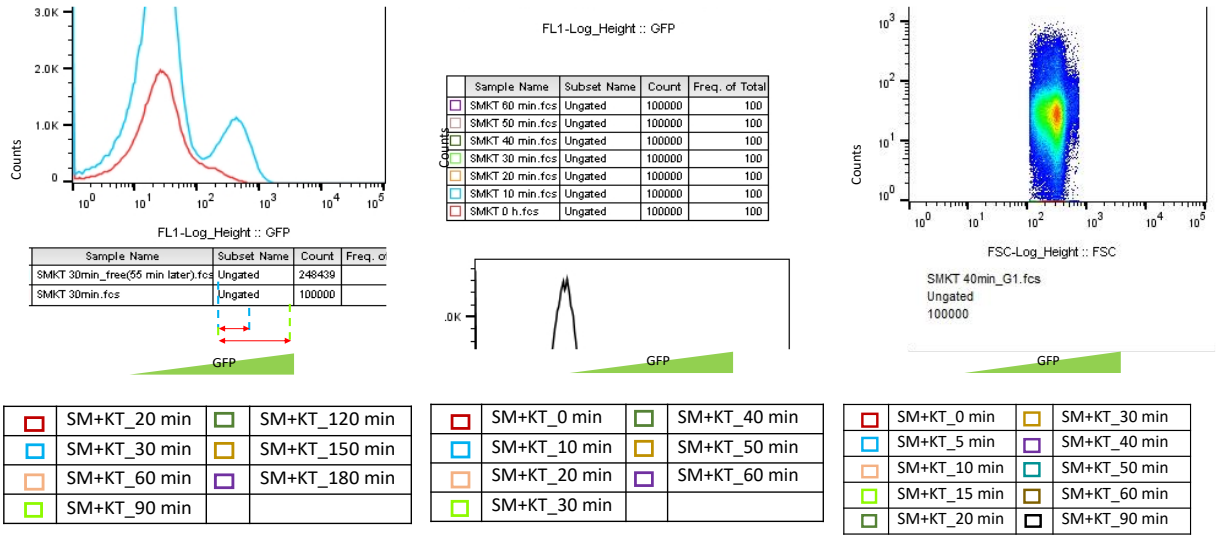
0.2:0.2 0.2:2 2:2

Fig. Filter mating of SM+KT with incubation time of 40 min with different OD_{600} ratio,0.2 : 0.2 means OD_{donor} : $OD_{recipient}$ = 0.2 : 0.2, others are the same.

2.1.2 Stabilization of mating system

To check the reproducibility of this phenomenon and find the initial time of fluorescence variety, I repeat previous experiment (Fig.2 A) and check fluorescence intensity more frequently in first-hour mating (Fig.2 B,C).

Since there was no obvious change of fluorescence intensity with liquid mating in both the recipient of *P. putida* KT2440RG and *E. coli* DH5α, this time I performed filter mating assay only, the accurate process as follows:



A B C

Fig. 2. Time-course fluorescence detection of filter mating cultures with flow cytometry during 3-hours incubation (A),1-hour incubation (B) and 1.5-hour incubation (C). (The data were from different mating assay). SM: *P. putida* SM1443 (pBP136::*gfp*), KT: *P. putida* KT2440RG, X axis represent the intensity of GFP fluorescence, Y axis means the number of cells.

It (Fig.2 A) shows the same tendency with last result that fluorescence intensity shift with time-lapse. To detect initial time of fluorescence intensity change, I took sample for fluorescence observation every 10 minutes until 60-min incubation, there was no obvious shift of fluorescence intensity during first 10-min incubation, it

was become stronger and stronger from 20-min incubation sample until 1 hour (Fig.2 B), it indicated that fluorescence intensity shift which can be detected is initiated during 10-min to 20-min. To catch initial time point, I performed time-course observation experiment with interval of 5 minutes in first 20-min incubation (Fig.2 C), overall fluorescence intensity of this time is weaker than before and transconjugant number is also fewer than before, it may be caused by different condition of cells used for mating. Compare to previously, too small single colony of *P. putida* SM1443 (pBP136::*gfp*) was selected for pre-culture which may led to slow growth speed and different cell activity after 14-hour shaking culture. Next time, I will pay attention to colony size for pre-culture to ensure the same growth and activity condition of donor and recipient for plasmid transfer observation.

2.1.2 Optimization of mating condition

Mating assay for 3 hours.
When increase the number of recipient have no positive effect on transconjugant ratio. So I increase donor

When using *E. coli* DH5α as recipient,
To check the reproducibility of this phenomenon and find the initial time of fluorescence variety, I repeat previous experiment (Fig.2 A) and check fluorescence intensity more frequently in first-hour mating (Fig.2 B,C).

Since there was no obvious change of fluorescence intensity with liquid mating in both the recipient of *P. putida* KT2440RG and *E. coli* DH5α, this time I performed filter mating assay only, the accurate process as follows:

It (Fig.2 A) shows the same tendency with last result that fluorescence intensity shift with time-lapse. To detect initial time of fluorescence intensity change, I took sample for fluorescence observation every 10 minutes until 60-min incubation, there was no obvious shift of fluorescence intensity during first 10-min incubation, it was become stronger and stronger from 20-min incubation sample until 1 hour (Fig.2 B), it indicated that fluorescence intensity shift which can be detected is initiated during 10-min to 20-min. To catch initial time point, I performed time-course observation experiment with interval of 5 minutes in first 20-min incubation (Fig.2 C), overall fluorescence intensity of this time is weaker than before and transconjugant number is also fewer than before, it may be caused by different condition of cells used for mating. Compare to previously, too small single colony of *P. putida* SM1443 (pBP136::*gfp*) was selected for pre-culture which may led to slow growth speed and different cell activity after 14-hour shaking culture. Next time, I will pay attention to colony size for pre-culture to ensure the same growth and activity condition of donor and recipient for plasmid transfer observation.

Sorting of high fluorescence intensity cells

From last result, fluorescence intensity become stronger and stronger with time-lapse during mating, I want to confirm cells with high fluorescence intensity are the transconjugant, so the high-intensity cells from 30-min and 90-min mating samples (Fig.2A short and long arrow) were sorted with flow cytometry. 384-well model

was set for sorting on LB plate with Cm (both the donor and recipient are resistant to Cm) to prevent contamination in sorting chamber, plates incubated at 30°C for 2 days.

A B

Fig.3. Sorting plate of 90-min mating samples (A), transconjugant selection plate (B): All colony were selected from sorting plate A.

Finally, 81 colony from 90-min incubation sample were obtained after 2-day incubation (Fig.3 A), colony were picked with tooth tick to new transconjugant selection plate supplied with Rif, Gm and Km (Fig.3 B), all colony can be grown on selection plate which represent they may the transconjugant, if it is necessary I will do colony PCR to confirm transconjugant in the future experiment. In this condition, the surviving ability of sorted cell is 21%, maybe because of the sorted single cell is weak so that cannot easily survive on antibiotic plate, I will use LB plate for sorting and then transfer the colony on transconjugant selection plate. Some particles with high fluorescence intensity are sorted as cells also lead low efficiency sorting. For 30-min incubation sample I did not obtain any colony of sorting plate maybe because of the low density of transconjugant. Or the fluorescence intensity is not strong enough to separate transconjugant from other cells. I will increase cycle amount from 10000 to 20000 to accumulate more transconjugant to make clear sorting gate, check and modify sorting condition to ensure successful sorting.

RNA-seq of transconjugants from different mating time

One colony from a freshly streaked freezer
stock was inoculated into 5 ml TSB, which was incubated for
_16 hr, and subsequently a 4.88-ml culture was transferred into
5ml TSB (_10 generations per day). The same culture volume
was transferred daily for 10 days (representing 100 generations
of growth)

Chapter 3. Discrimination of plasmid-harboring and -free strains with single-cell Raman spectroscopy

3.1 Introduction

The important and challenging point of my research is to detect the plasmid transfer and capture the newly formed transconjugants immediately. Raman spectroscopy as a fast, label-free and noinsive technique have been used for single-cell analysis. Because it has ability to measure compounds inside the cell, I prospect that plasmid carriage strains have different metabolics with plasmid-free, which may lead to different Raman spectra. If this it can be reflected in Raman spectra, then Raman spectroscopy can be used for detecting plasmid transfer. So the objective of this work is to explore the possibilities to obtain information about cell compounds from bacterial Raman spectra. Whether the spectra reflect difference between plasmid-harboring and -free strains, whether the difference reflect some useful information about cell metabolism, whether plasmid-harboring and -free strains can be distinguished directly from Raman spectra. This part of work starts with cultivation of microbiology, measurement of Raman spectra on microbes, data processing and possible contribution of cell compounds to bacterial Raman spectra.

Pseudomonas putida KT2440 is the best characterized saprophytic *Pseudomonas* that has retained its ability to survive and function in the environment. *Escherichia coli* is a bacterium in the human and animal gut. Both strains are the preferred hosts for cloning and gene expression experiments of Gram-negative bacteria. The incompatibility group (IncP-1β) plasmid pB10 and RP4 (IncP-1α) are broad host range. It is known that both plasmids can be transferred between *Pseudomonas* hosts or *E. coli* hosts with high frequencies.

If this technique can be applied to detecting plasmid transfer, combined with single cell-picking technologies, such as laser microdissection or optical tweezer, this new method can be applied to clarify the interaction between plasmid and host cell at the level of single cells in real time. For this hypothesis,

If single cell-picking technologies, such as laser microdissection or optical tweezer, are combined, this new method can be applied to clarify the interaction between plasmid and host cell at the level of single cells in real time. It may be caused by different growth conditions of strains or some other effects like sampling preparation process, etc.

Combination of the type of energy source, carbon source, lead to specific type of metabolism,

3.2 Materials and methods

3.2.1 Cultivation of microorganisms

All strains were pre-cultured into LB medium at 30°C with incubator shaker (BR-23FP MR, Taitec) overnight, add chloramphenicol (Cm 30 μg/mL) and tetracycline (Tc 12.5 μg/mL), Cm 30 μg/mL and kanamycin (Km 50 μg/mL), Km 50 μg/mL for *P. putida* KT2440(pB10), *P. putida* KT2440(RP4) and *P. putida* KT2440 respectively; same concentration of Tc and Km to *E. coli* W3110(pB10), *E. coli* W3110(RP4) respectively.

The overnight cultured cells were centrifuged with 12000rpm (Beckman Coulter Microfuge 16) and washed with PBS buffer for twice, transferred into 300 mL flask containing 100-mL-minimal mediums with primary OD_{600} = 0.025. NMM-4 [15 mM Na_2HPO_4, 5 mM KH_2PO_4, 35 mM NH_4NO_3, 0.06 mM $FeCl_3 \cdot 6H_2O$, 0.8 mM $MgSO_4 \cdot 7H_2O$, 0.06 mM $CaCl_2 \cdot 2H_2O$, 0.1% (w/v) Succinate] was used for *P. putida* KT2440 cultivation and M9 medium [22 mM KH_2PO_4, 8.6 mM NaCl, 48 mM $NaHPO_4$, 19 mM NH_4Cl, 2 mM $MgSO_4 \cdot 7H_2O$, 0.1 mM $CaCl_2$ and 0.2% (w/v) D-Glucose] was used for *E. coli* W3110 cultivation.

3.2.2 Raman sample preparation

To determine sampling points of log- and stationary phase, optical density (OD) at 600 nm was measured every 2 hours using V-750 UV-Visible Spectrophotometer (Jasco, Japan). 50μL cell culture from 2 different growth conditions were taken for Raman spectroscopic measurement, 40-50 cells were randomly selected from each sample for observation; for stationary phase sample, the culture was 10 times dilution before Raman spectroscopy observation.

3.2.3 Raman spectroscopy

Raman spectra of bacteria on slide were recorded with a laboratory-built confocal Raman micro spectrometer (Huang et al 2012). It uses a 632.8 ‐ nm output of a He–Ne laser (3 mW at sample point) (Thorlabs, HNL210L) as the Raman excitation light, TE-cooled CCD detector (AndoriDus DU401-BV) as detector. The Raman spectrum was recorded with a 40

s exposure time. The spectral intensities were calibrated using a standard light source (Princeton Instruments, IntelliCal). All measurements were done at room temperature.

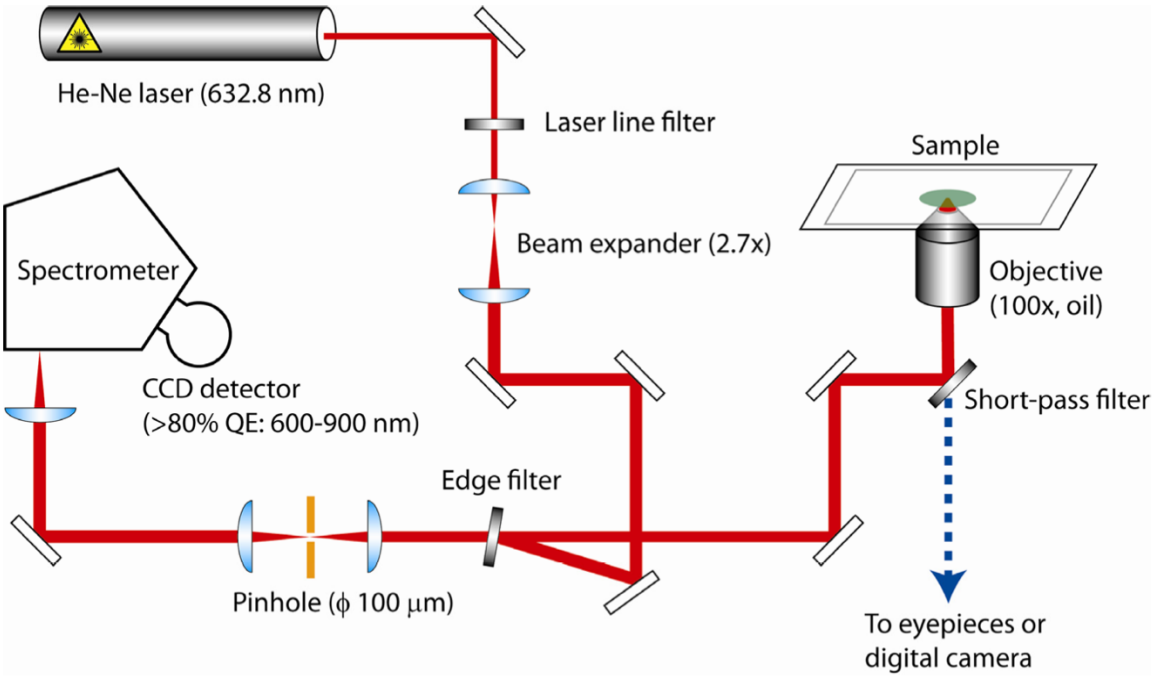


Figure S3. Schematic diagram of the laboratory-built confocal Raman microspectrometer used in this work. See text for details. laser beam was introduced into an inverted microscope (Nikon, TE2000‐U, customized) by a pair of an edge filter and a dichroic mirror. The beam was focused onto the sample with a 100×, NA = 1.3, oil‐immersion objective, and the backward Raman scattering was collected with the same objective.

detected with an electron‐multiplying charge‐coupled device detector (Andor Technology, Newton).

3.2.4 Data analysis

3.2.4.1 Spectra pre-processing

Raw spectra data is CCD pixel index as the horizontal axis, the horizontal axis from pixel index to Raman shift by calibration with 2 steps: 1. Convert pixel induces into wavelength and 2. Conversion from wavelength to Raman shift

For wavelength calibration, wavelengths of Ne have been determined accurately in literatures, it is used as standards.

Wavelength = K0+K1*x+K2*x^2

(Excitation wavelength = 632.991 nm)

Raman spectra data is loaded as Figure.1. Raman shift /cm $^{-1}$ for horizontal axis, Intensity/ counts for vertical.

The spectrum data were pre-treated with Igor Pro (WaveMetrics) to denoise by using singular value decomposition (SVD) which is considerably improved the signal-to-noise ratio without losing any characteristic features in the raw spectrum [3].

3.2.4.2 Setup and validation of predictive model

Raman spectra contains large amount of information which lead to problems for analysis: like each spectrum conclude overlapping elements from different compounds, spectra from different species show various subtle changes, it is difficult to see the different with eye directly, so the machine learning method Random forest (RF) is introduced for classification analysis. Radom forest is machining learning method for classification, regression and other tasks. It is operated by constructing a multitude of decision trees at training time, then outputting the mode of the class of the individual trees.

In my case, 50% numbers of spectra obtained from plasmid-harboring and –free strains were used to building training dataset, left 50% data were introduced for validation, the validation results can evaluate how good our classifier is. Take *P. putida* KT2440 strains as an example, when do classification analysis with Raman spectra obtained from their log phase samples, a total of 90 spectra (30 spectra for each) collected from *P. putida* KT2440(pB10), *P. putida* KT2440(RP4) and *P. putida* KT2440, 45 spectra (15 spectra for each) of them were used as training set to constructing a number of decision trees (this time is 500), the left were used as test set for vibration. After decision trees were built, each spectrum from test set was go through all nodes of the decision trees, each tree gives "votes" for a class, the forest chooses the classification having the most votes (over all the trees in the forest).

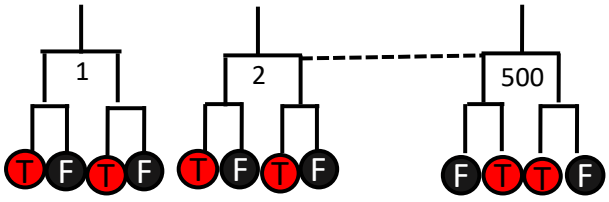


Figure 1. A decision tree generated by RF. Node values correspond to Raman shift and the terminal nodes show outcome class. In my case, I defined spectrum of plasmid-free strain as

T (=1), spectrum of plasmid-harboring as F (=0). After one spectrum go through all trees, it got the final vote of T classification with m times, the final score of T classification is m/500.

Raman spectroscopy is based on inelastic scattering of photons following

Table 1. Sampling point of each strain in log phase (Log) and stationary phase (Sta)

Strain name	Incubation time (Log)	OD_{600}	Incubation time (Sta)	OD_{600}
E. coli DH5α	5 h	0.355	22 h10 min	0.948
E. coli DH5α(pB10)	6 h 41 min	0.535	23 h 5 min	0.891
E. coli DH5α(RP4)	7 h 10 min	0.528	24 h 40 min	0.918
P.putida KT2440	7 h 40 min	0.57	25 h 20 min	1.164
P.putida KT2440(pB10)	8 h 40 min	0.418	26 h 20 min	1.262
P.putida KT2440(RP4)	21 h 30 min	0.239	27 h 20 min	0.84

This time, six kinds of strains *E. coli* W3110(pB10), *E. coli* W3110(RP4) and *E. coli* W3110, *P. resinovorans* CA10dm4RG, *P. resinovorans* CA10dm4RG (pWW0) and *P. resinovorans* CA10dm4RG (NAH7K2) were taken to Prof. Shigeto's lab for Raman spectroscopy observation. Log and stationary phase samples of each strain were taken based on the growth curve that I made in our lab previously. The process as follows:

1. Strains were pre-cultured into LB liquid medium overnight supplied with antibiotic when necessary;
2. Resulted cells were incubated in 100 mL minimal medium supplied with proper carbon source (NMM-4 medium with 0.1% (w/v) succinate for *P. resinovorans* CA10dm4RG strains, M9 medium with 0.2% d-glucose for E. *coli* stains) at 30°C, 120 rpm.
3.

The spectrum data were pre-treated with Igor Pro (WaveMetrics) to denoised by using singular value decomposition (SVD) since it is considerably improved the signal-to-noise ratio without losing any characteristic features in the raw spectrum [3]. After pre-treatment of all spectrums, single-cell classification of the resultant spectrum was implemented with an effective machine learning system named Random forest, it is a classification model that works by constructing a multitude of decision trees at training time and predict the unknown class based on the known example [4]. The difference of Raman spectrum between plasmid-

harboring and each plasmid-free strain in log phase and stationary phase were analyzed respectively. For example, when analyzed the log phase samples of *P. resinovorans* CA10dm4RG, *P. resinovorans* CA10dm4RG (pWW0) and *P. resinovorans* CA10dm4RG (NAH7K2) (each strain 30 repeat *3 strains) were input into database, 45 were used to construct decision trees (each strain 15 spectrum*3 strains), the left 45 were predicted for classification. The classification result shows the plasmid free strain *P. resinovorans* CA10dm4RG can be distinguished from plasmid-harboring strains *P. resinovorans* CA10dm4RG (pWW0) and *P. resinovorans* CA10dm4RG (NAH7K2) from Raman spectrums (Fig.3A). The *E. coli* W3110 also can be distinguished from *E. coli* W3110(pB10) and *E. coli* W3110(RP4) from their Raman spectrums (Fig.3B).

3.2 Results and discussion

The large amount of information present in the Raman spectrum creates problems for analysis – each spectrum (see Figure 2 above) comprises overlapping elements from a number of different components. Spectra obtained from individual cells of the same species, and of different species, all show numerous subtle changes, too many to be distinguished by eye .However, multivariate techniques including principal components analysis (PCA), discriminant functional analysis (DFA) and hierarchical cluster analysis (HCA) allow these many subtle spectral differences to be clarified and used to classify cell types.Discrimination of SpeciesThree bacterial species (Acinetobacter sp., ADP1 E. coli DH5α and Pseudomonas fluorescens SBW25) were studied, with cells taken from three distinct growth stages for each (corresponding to incubations of 4h, 8h, and 22h). Initial multivariate analysis (Figure 3) showed excellent clustering for the three bacterial species, clearing indicating the power of Raman to distinguish the physiological differences between cells of different species.It is interesting to note that this experiment showed the clustering to be robust to species phenotypic differences, despite temporal differences in cellular physiology during the phases of growth.This demonstrates that whatever growth stage a particular cell is in, Raman provides a facile method for species identificatio

3.2.1 Bulk spectra

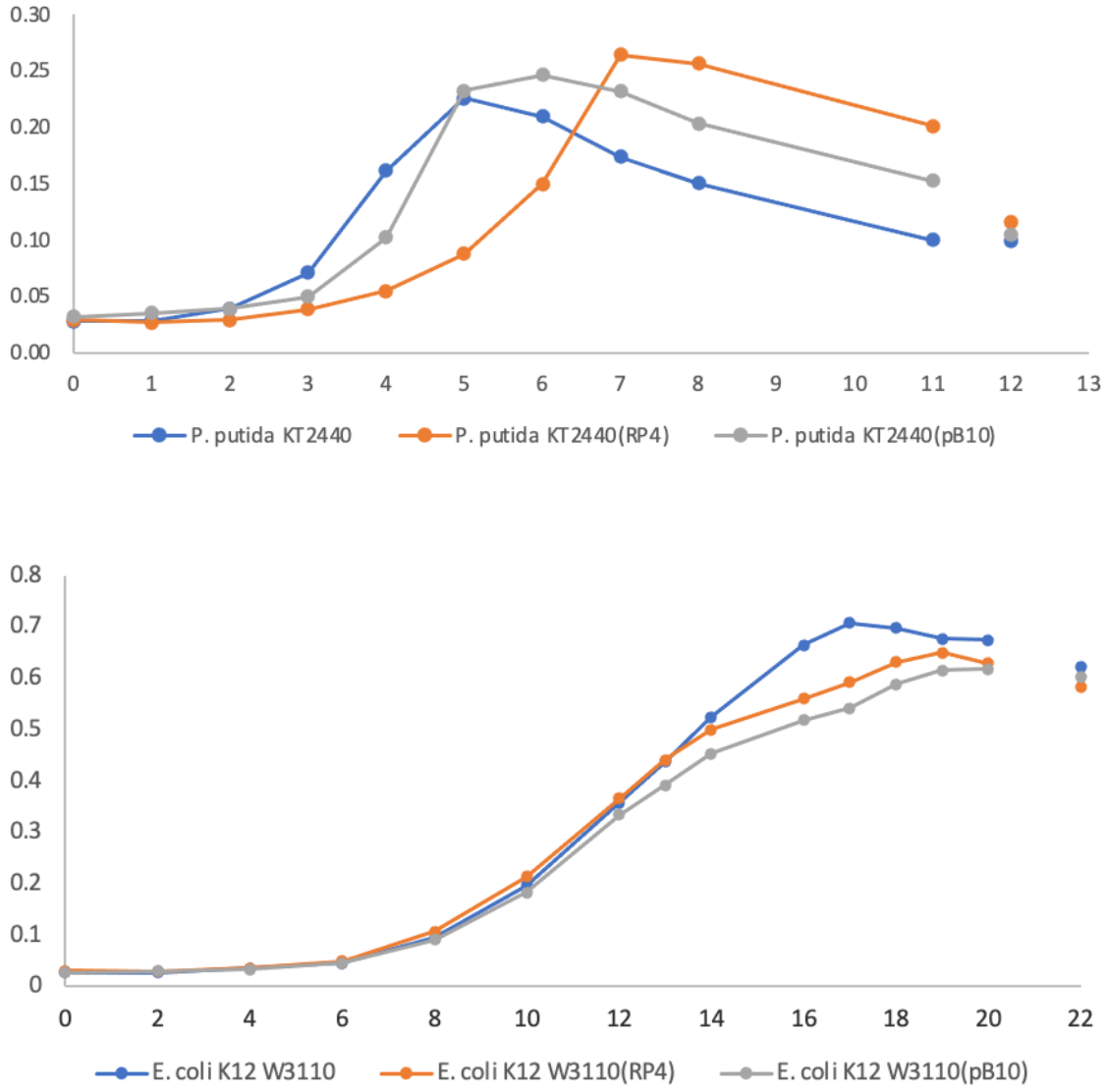
The spectra were recorded with integration time of 40s directly from minimal medium (40 to 50 repetitions [Table 1]). The spectrum of E. coli W3110 strain is dominated by bands at, the spectrum of P. putida KT2440 strains.

For a distinct identification of different strains, a reliable data analysis method is required.

Band assignments for Raman spectra of E. *coli* W3110 strains.

3.1 Materials and methods

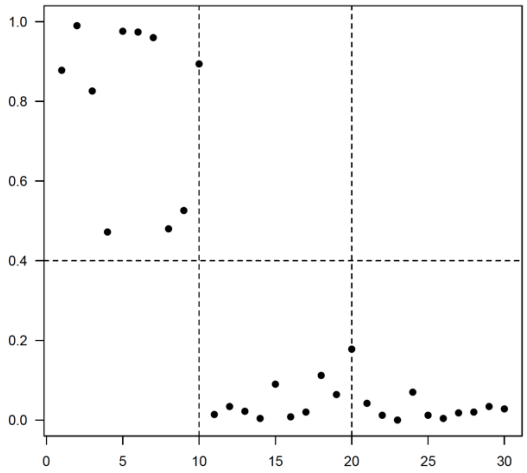
Cells were pre-cultured overnight in LB with antibiotic, and culture to minimal medium with primary OD_{600} = 0.025, shaking for 200 rpm 30C.



Reference growth curve

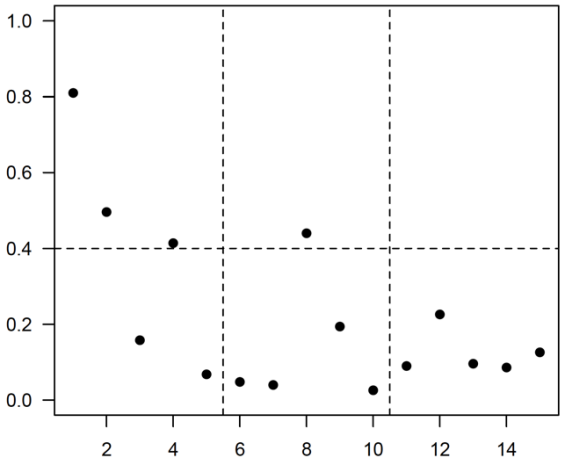
Strains	First time	Second time
P. putid KT2440	3.5 h	9.5h
P. putid KT2440(pB10)	5.5h	10h
P. putid KT2440(RP4)	6h	22h
E. coli W3110	8h	24h
E. coli W3110(pB10)	9h	25h
E. coli W3110(RP4)	10h	26.5h

First time

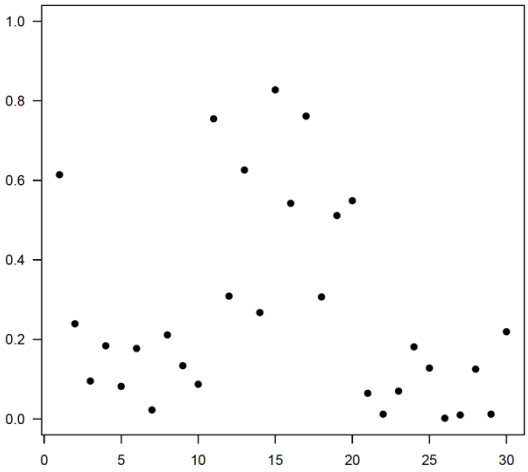


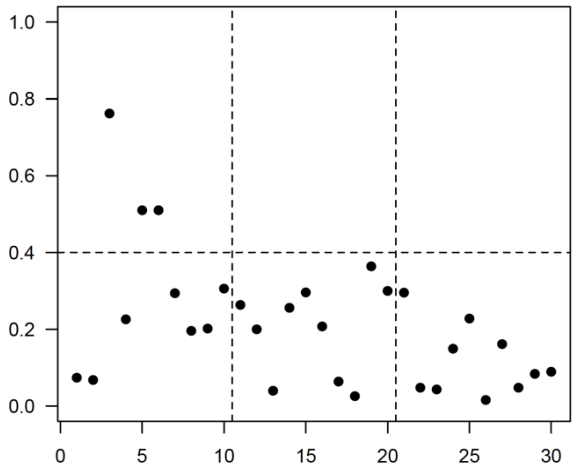
KT KT(pB10) KT(RP4)

Second time



KT KT(pB10) KT(RP4)

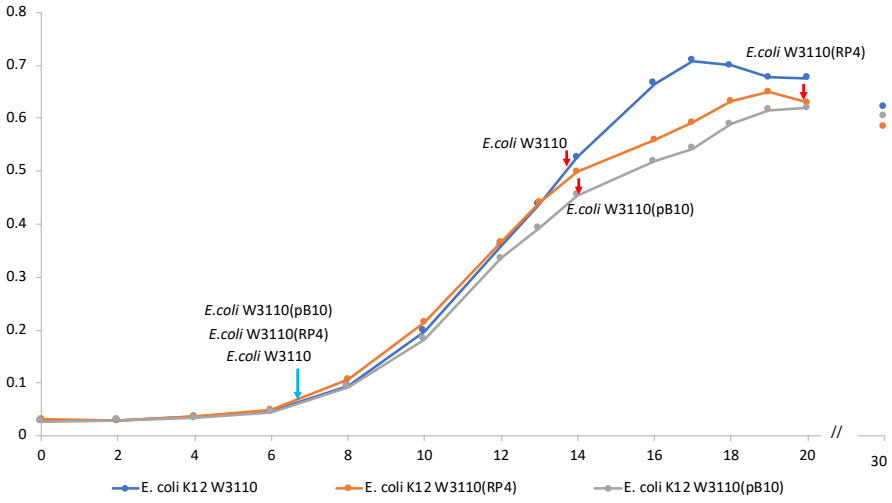




W3110 W3110(pB10) W3110(RP4) W3110 W3110(pB10) W3110(RP4)

Conclusion:

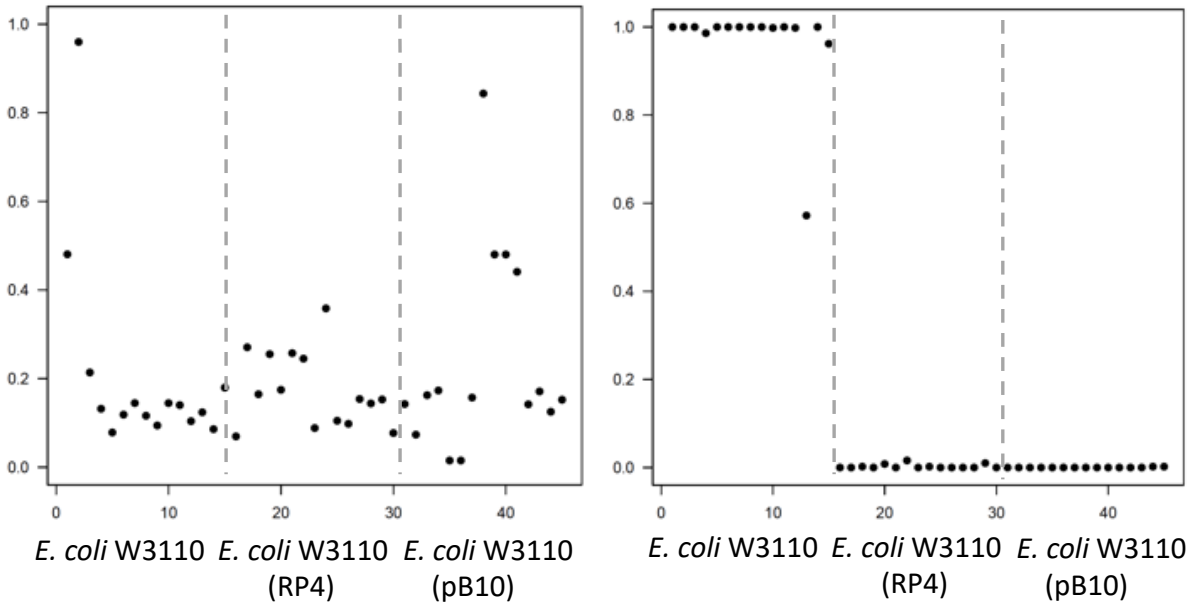
Difference seems can be detected, need more cells for confirmation.



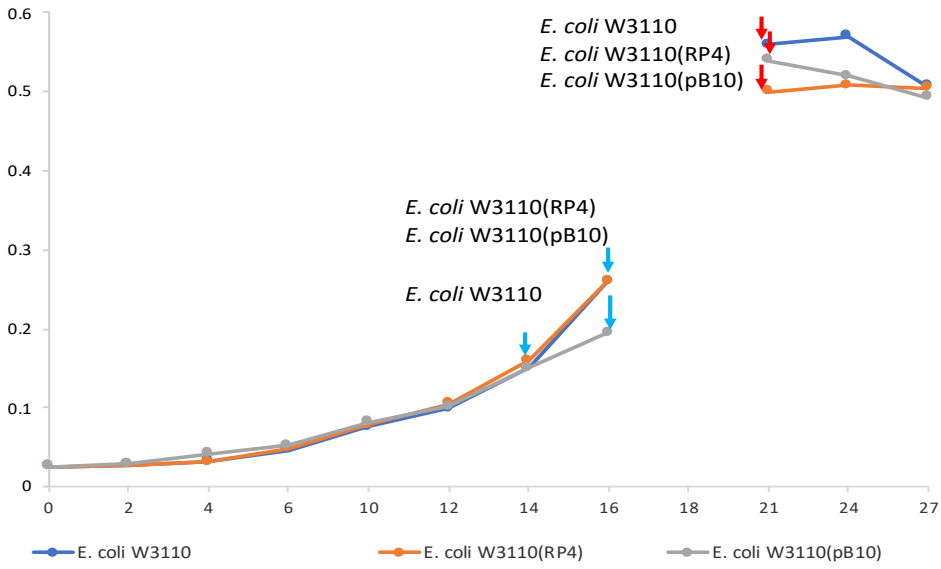
Reference growth curve with real-time sampling points

Strains	First time	OD_{600}	Second time	OD_{600}
E. coli W3110	7h		24h	0.45
E. coli W3110(pB10)	9h	0.065	26h	0.458

E. coli W3110(RP4)	8h	0.0473	25h	0.437

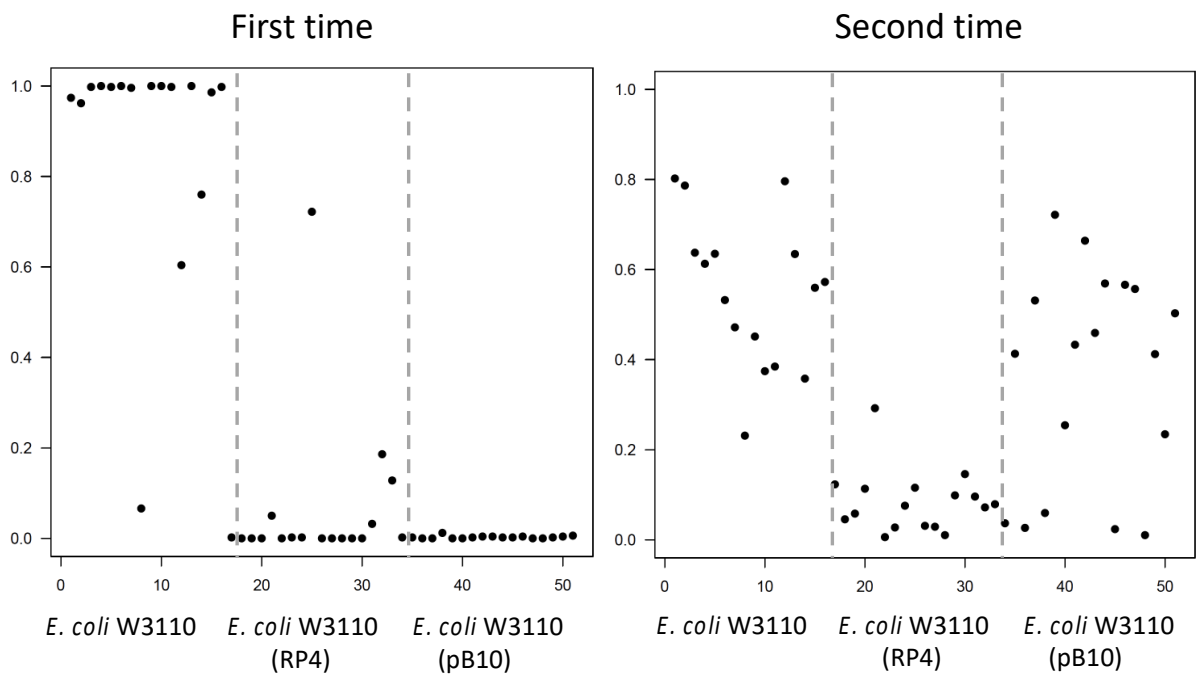


Point: Growth curve should be taken from real time.



Growth curve and sampling points in real-time

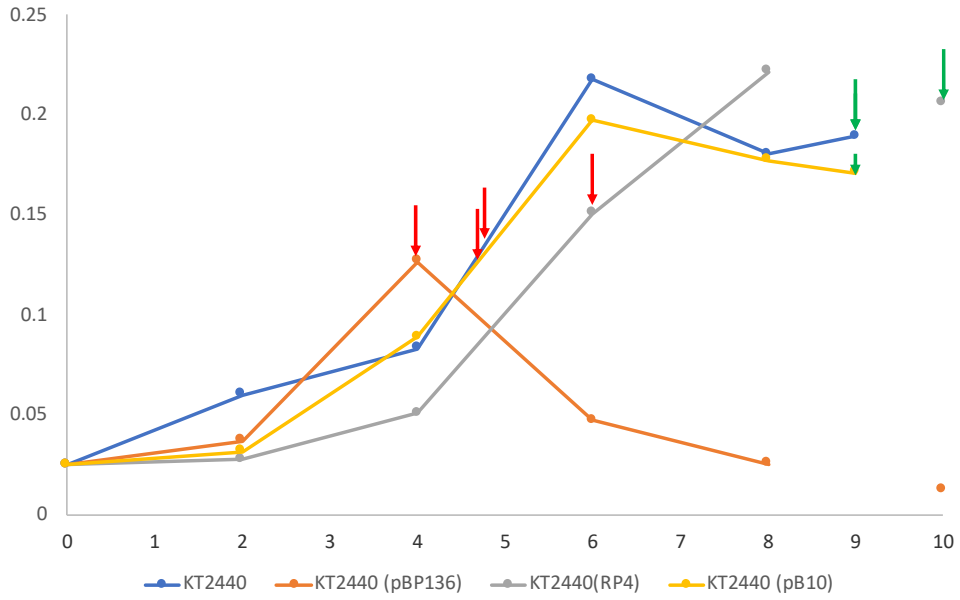
Strains	First time	OD_{600}	Second time	OD_{600}
E. coli W3110	7h		24h	0.45
E. coli W3110(pB10)	9h	0.065	26h	0.458
E. coli W3110(RP4)	8h	0.0473	25h	0.437



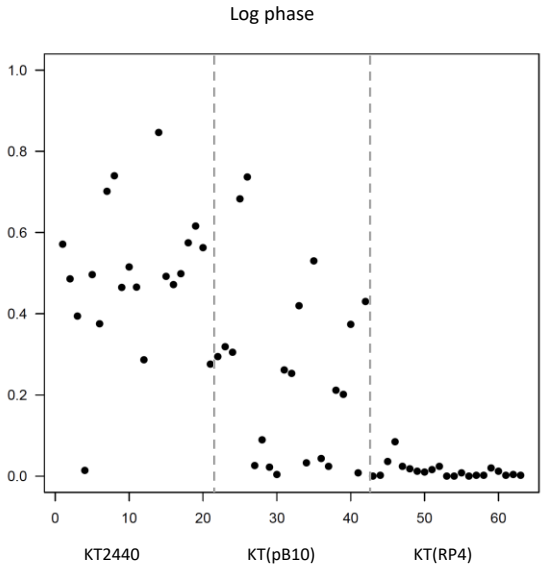
Conclusion:

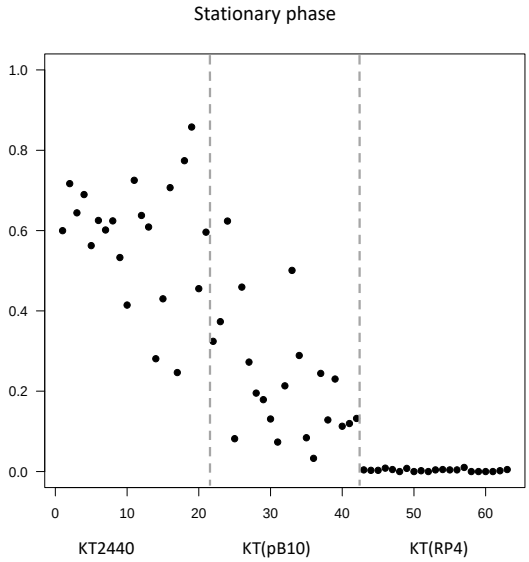
For *E. coli* W3110 strains, log phase sample can be distinguished

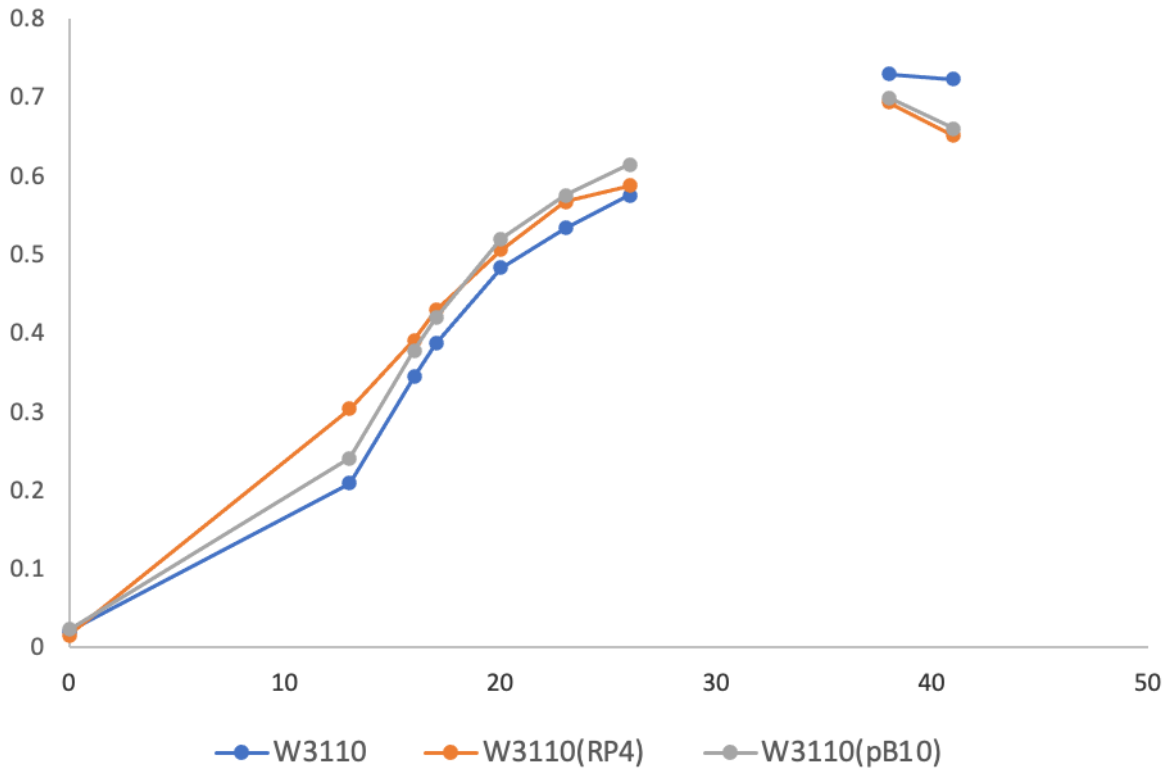
For CA10dm4RG strains, not so clearly, maybe because of late time points of 2 strains, need batch culture and take log phase sample in from same time-points.



Strains	Log phase	Real OD	Stationary phase	Real OD
P. putid KT2440	5 h	0.148	9h	0.1895
P. putid KT2440(pB10)	5h	0.1353	9h	0.17
P. putid KT2440(RP4)	6h	0.151	10h	0.2058







Strains	First time	OD_{600}	Second time	OD_{600}
E. coli W3110	17 h		24h	0.45
E. coli W3110(pB10)	13 h	0.065	26h	0.458
E. coli W3110(RP4)	16 h	0.0473	25h	0.437

3.2 Results and discussion

3.3 Discussion

To checking generality of this method, more plasmids from different Incompatibility (Inc) groups will be introduced for Raman spectroscopy observation (Table 1). The research in our lab shows that various plasmids, including IncP-1, P-7, P-9, W group plasmids did not load fitness cost on *P. resinovorans* CA10dm4 host, it is insensitive to various plasmid. As a control, the *P. resinovorans* CA10dm4 will be introduced as another host to check if this property can be reflected by Raman spectrum.

Table 1. The host strains and plasmids used in Raman spectroscopy.

Host	Plasmid

	IncP-1	IncP-7	IncP-9	IncW
E. coli W3110	RP4	pCAR1	pWW0	R388
P. putida KT2440	pB10	Rms148	NAH7K2	
P. resinovorans CA10dm4	pB136	pDK1K		

IncP: Incompatibility groups are classified based on a *Pseudomonas* strain as a recipient, and IncP-1 corresponds to IncP in the *Escherichia coli* plasmid classification.

As table 1 shows, the combination between the two hosts *P. putida* KT2440, *E. coli* W3110 and plasmids pB10, RP4 were used for Raman spectroscopy and classification analysis. The combination of these two hosts and *P. resinovorans* CA10dm4 with the left plasmids will be used for Raman spectroscopy observation in my future work to predict an integral regular pattern of plasmid harboring with Raman spectrums.

Denoise with PCA

Huang C-K, Ando M, Hamaguchi H-o, Shigeto S. 2012. Disentangling Dynamic Changes of Multiple Cellular Components during the Yeast Cell Cycle by in Vivo Multivariate Raman Imaging. *Analytical Chemistry* 84: 5661-68

Raman spectroscopy also presents disadvantages. For instance, the Raman signals of certain compounds can be quite weak, making them difficult to detect or undetectable. The Raman signal of certain compounds can be composed of several peaks, or be unknown. Also, the

background of samples can interfere with the Raman signal of bacteria. The equipment can be quite costly, depending on the type of Raman spectroscope.

Chapter 3. Discrimination of plasmid-harboring and -free strains with single-cell Raman spectroscopy

3.1 Introduction

The important and challenging point of my research is to detect the plasmid transfer and capture the newly formed transconjugants immediately. Raman spectroscopy as a fast, label-free and noinsive technique have been used for single-cell analysis. Because it has ability to measure compounds inside the cell, I prospect that plasmid carriage strains have different metabolics with plasmid-free, which may lead to different Raman spectra. If this it can be reflected in Raman spectra, then Raman spectroscopy can be used for detecting plasmid transfer. So the objective of this work is to explore the possibilities to obtain information about cell compounds from bacterial Raman spectra. Whether the spectra reflect difference between plasmid-harboring and -free strains, whether the difference reflect some useful information about cell metabolism, whether plasmid-harboring and -free strains can be distinguished directly from Raman spectra. This part of work starts with cultivation of microbiology, measurement of Raman spectra on microbes, data processing and possible contribution of cell compounds to bacterial Raman spectra.

Pseudomonas putida KT2440 is the best characterized saprophytic *Pseudomonas* that has retained its ability to survive and function in the environment. *Escherichia coli* is a bacterium in the human and animal gut. Both strains are the preferred hosts for cloning and gene expression experiments of Gram-negative bacteria. The incompatibility group (IncP-1β) plasmid pB10 and RP4 (IncP-1α) are broad host range. It is known that both plasmids can be transferred between *Pseudomonas* hosts or *E. coli* hosts with high frequencies.

If this technique can be applied to detecting plasmid transfer, combined with single cell-picking technologies, such as laser microdissection or optical tweezer, this new method can be applied to clarify the interaction between plasmid and host cell at the level of single cells in real time. For this hypothesis,

If single cell-picking technologies, such as laser microdissection or optical tweezer, are combined, this new method can be applied to clarify the interaction between plasmid and host

cell at the level of single cells in real time. It may be caused by different growth conditions of strains or some other effects like sampling preparation process, etc.

Combination of the type of energy source, carbon source, lead to specific type of metabolism,

3.2 Materials and methods

3.2.1 Cultivation of microorganisms

All strains were pre-cultured into LB medium at 30°C with incubator shaker (BR-23FP MR, Taitec) overnight, add chloramphenicol (Cm 30 μg/mL) and tetracycline (Tc 12.5 μg/mL), Cm 30 μg/mL and kanamycin (Km 50 μg/mL), Km 50 μg/mL for *P. putida* KT2440(pB10), *P. putida* KT2440(RP4) and *P. putida* KT2440 respectively; same concentration of Tc and Km to *E. coli* W3110(pB10), *E. coli* W3110(RP4) respectively.

The overnight cultured cells were centrifuged with 12000rpm (Beckman Coulter Microfuge 16) and washed with PBS buffer for twice, transferred into 300 mL flask containing 100-mL-minimal mediums with primary OD_{600} = 0.025. NMM-4 [15 mM Na_2HPO_4, 5 mM KH_2PO_4, 35 mM NH_4NO_3, 0.06 mM $FeCl_3 \cdot 6H_2O$, 0.8 mM $MgSO_4 \cdot 7H_2O$, 0.06 mM $CaCl_2 \cdot 2H_2O$, 0.1% (w/v) Succinate] was used for *P. putida* KT2440 cultivation and M9 medium [22 mM KH_2PO_4, 8.6 mM NaCl, 48 mM $NaHPO_4$, 19 mM NH_4Cl, 2 mM $MgSO_4 \cdot 7H_2O$, 0.1 mM $CaCl_2$ and 0.2% (w/v) D-Glucose] was used for *E. coli* W3110 cultivation.

3.2.2 Raman sample preparation

To determine sampling points of log- and stationary phase, optical density (OD) at 600 nm was measured every 2 hours using V-750 UV-Visible Spectrophotometer (Jasco, Japan). 50μL cell culture from 2 different growth conditions were taken for Raman spectroscopic measurement, 40-50 cells were randomly selected from each sample for observation; for stationary phase sample, the culture was 10 times dilution before Raman spectroscopy observation.

3.2.3 Raman spectroscopy

Raman spectra of bacteria on slide were recorded with a laboratory-built confocal Raman micro spectrometer (Huang et al 2012). It uses a 632.8‐nm output of a He–Ne laser (3 mW at sample point) (Thorlabs, HNL210L) as the Raman excitation light, TE-cooled CCD detector (AndoriDus DU401-BV) as detector. The Raman spectrum was recorded with a 40 s exposure time. The spectral intensities were calibrated using a standard light source (Princeton Instruments, IntelliCal). All measurements were done at room temperature.

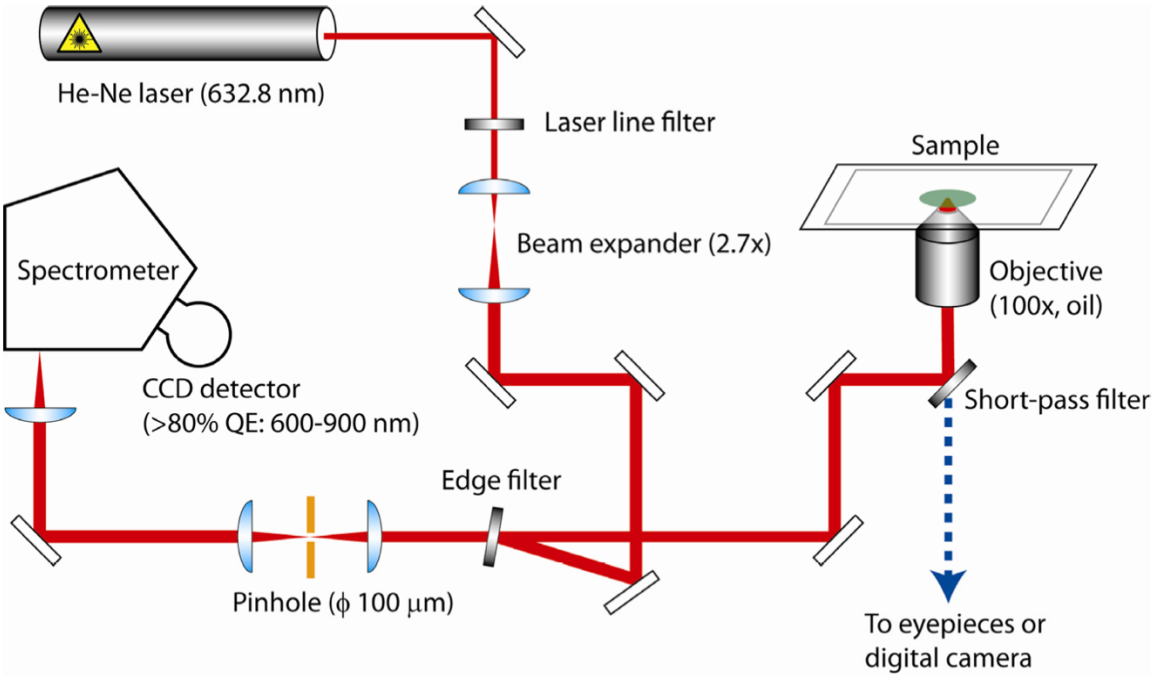


Figure S3. Schematic diagram of the laboratory-built confocal Raman microspectrometer used in this work. See text for details. laser beam was introduced into an inverted microscope (Nikon, TE2000‐U, customized) by a pair of an edge filter and a dichroic mirror. The beam was focused onto the sample with a 100×, NA = 1.3, oil‐immersion objective, and the backward Raman scattering was collected with the same objective.

detected with an electron‐multiplying charge‐coupled device detector (Andor Technology, Newton).

3.2.4 Data analysis

3.2.4.1 Spectra pre-processing

Raw spectra data is CCD pixel index as the horizontal axis, the horizontal axis from pixel index to Raman shift by calibration with 2 steps: 1. Convert pixel induces into wavelength and 2. Conversion from wavelength to Raman shift

For wavelength calibration, wavelengths of Ne have been determined accurately in literatures, it is used as standards.

Wavelength = K0+K1*x+K2*x$^{\wedge 2}$

(Excitation wavelength = 632.991 nm)

Raman spectra data is loaded as Figure.1. Raman shift /cm $^{-1}$ for horizontal axis, Intensity/ counts for vertical.

The spectrum data were pre-treated with Igor Pro (WaveMetrics) to denoise by using singular value decomposition (SVD) which is considerably improved the signal-to-noise ratio without losing any characteristic features in the raw spectrum [3].

3.2.4.2 Setup and validation of predictive model

Raman spectra contains large amount of information which lead to problems for analysis: like each spectrum conclude overlapping elements from different compounds, spectra from different species show various subtle changes, it is difficult to see the different with eye directly, so the machine learning method Random forest (RF) is introduced for classification analysis. Radom forest is machining learning method for classification, regression and other tasks. It is operated by constructing a multitude of decision trees at training time, then outputting the mode of the class of the individual trees.

In my case, 50% numbers of spectra obtained from plasmid-harboring and –free strains were used to building training dataset, left 50% data were introduced for validation, the validation results can evaluate how good our classifier is. Take *P. putida* KT2440 strains as an example, when do classification analysis with Raman spectra obtained from their log phase samples, a total of 90 spectra (30 spectra for each) collected from *P. putida* KT2440(pB10), *P. putida* KT2440(RP4) and *P. putida* KT2440, 45 spectra (15 spectra for each) of them were used as training set to constructing a number of decision trees (this time is 500), the left were used as test set for vibration. After decision trees were built, each spectrum from test set was go through all nodes of the decision trees, each tree gives "votes" for a class, the forest chooses the classification having the most votes (over all the trees in the forest).

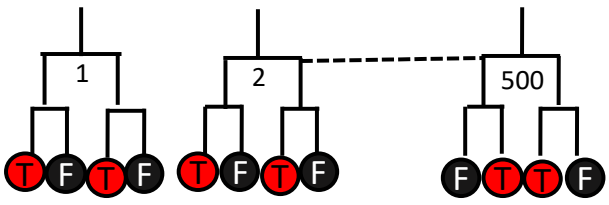


Figure 1. A decision tree generated by RF. Node values correspond to Raman shift and the terminal nodes show outcome class. In my case, I defined spectrum of plasmid-free strain as T (=1), spectrum of plasmid-harboring as F (=0). After one spectrum go through all trees, it got the final vote of T classification with m times, the final score of T classification is m/500.

Raman spectroscopy is based on inelastic scattering of photons following

Table 1. Sampling point of each strain in log phase (Log) and stationary phase (Sta)

Strain name	Incubation time (Log)	OD_{600}	Incubation time (Sta)	OD_{600}
E. coli DH5α	5 h	0.355	22 h10 min	0.948
E. coli DH5α(pB10)	6 h 41 min	0.535	23 h 5 min	0.891
E. coli DH5α(RP4)	7 h 10 min	0.528	24 h 40 min	0.918
P.putida KT2440	7 h 40 min	0.57	25 h 20 min	1.164
P.putida KT2440(pB10)	8 h 40 min	0.418	26 h 20 min	1.262
P.putida KT2440(RP4)	21 h 30 min	0.239	27 h 20 min	0.84

This time, six kinds of strains *E. coli* W3110(pB10), *E. coli* W3110(RP4) and *E. coli* W3110, *P. resinovorans* CA10dm4RG, *P. resinovorans* CA10dm4RG (pWW0) and *P. resinovorans* CA10dm4RG (NAH7K2) were taken to Prof. Shigeto's lab for Raman spectroscopy observation. Log and stationary phase samples of each strain were taken based on the growth curve that I made in our lab previously. The process as follows:

1. Strains were pre-cultured into LB liquid medium overnight supplied with antibiotic when necessary;
2. Resulted cells were incubated in 100 mL minimal medium supplied with proper carbon source (NMM-4 medium with 0.1% (w/v) succinate for *P. resinovorans* CA10dm4RG strains, M9 medium with 0.2% d-glucose for E. *coli* stains) at 30°C, 120 rpm.
3.

The spectrum data were pre-treated with Igor Pro (WaveMetrics) to denoised by using singular value decomposition (SVD) since it is considerably improved the signal-to-noise ratio without losing any characteristic features in the raw spectrum [3]. After pre-treatment of all spectrums, single-cell classification of the resultant spectrum was implemented with an effective machine learning system named Random forest, it is a classification model that works by constructing a multitude of decision trees at training time and predict the unknown class based on the known example [4]. The difference of Raman spectrum between plasmid-harboring and each plasmid-free strain in log phase and stationary phase were analyzed respectively. For example, when analyzed the log phase samples of *P. resinovorans*

CA10dm4RG, *P. resinovorans* CA10dm4RG (pWW0) and *P. resinovorans* CA10dm4RG (NAH7K2) (each strain 30 repeat *3 strains) were input into database, 45 were used to construct decision trees (each strain 15 spectrum*3 strains), the left 45 were predicted for classification. The classification result shows the plasmid free strain *P. resinovorans* CA10dm4RG can be distinguished from plasmid-harboring strains *P. resinovorans* CA10dm4RG (pWW0) and *P. resinovorans* CA10dm4RG (NAH7K2) from Raman spectrums (Fig.3A). The *E. coli* W3110 also can be distinguished from *E. coli* W3110(pB10) and *E. coli* W3110(RP4) from their Raman spectrums (Fig.3B).

3.2 Results and discussion

The large amount of information present in the Raman spectrum creates problems for analysis – each spectrum (see Figure 2 above) comprises overlapping elements from a number of different components. Spectra obtained from individual cells of the same species, and of different species, all show numerous subtle changes, too many to be distinguished by eye .However, multivariate techniques including principal components analysis (PCA), discriminant functional analysis (DFA) and hierarchical cluster analysis (HCA) allow these many subtle spectral differences to be clarified and used to classify cell types.Discrimination of SpeciesThree bacterial species (Acinetobacter sp., ADP1 E. coli DH5α and Pseudomonas fluorescens SBW25) were studied, with cells taken from three distinct growth stages for each (corresponding to incubations of 4h, 8h, and 22h). Initial multivariate analysis (Figure 3) showed excellent clustering for the three bacterial species, clearing indicating the power of Raman to distinguish the physiological differences between cells of different species.It is interesting to note that this experiment showed the clustering to be robust to species phenotypic differences, despite temporal differences in cellular physiology during the phases of growth.This demonstrates that whatever growth stage a particular cell is in, Raman provides a facile method for species identificatio

3.2.1 Bulk spectra

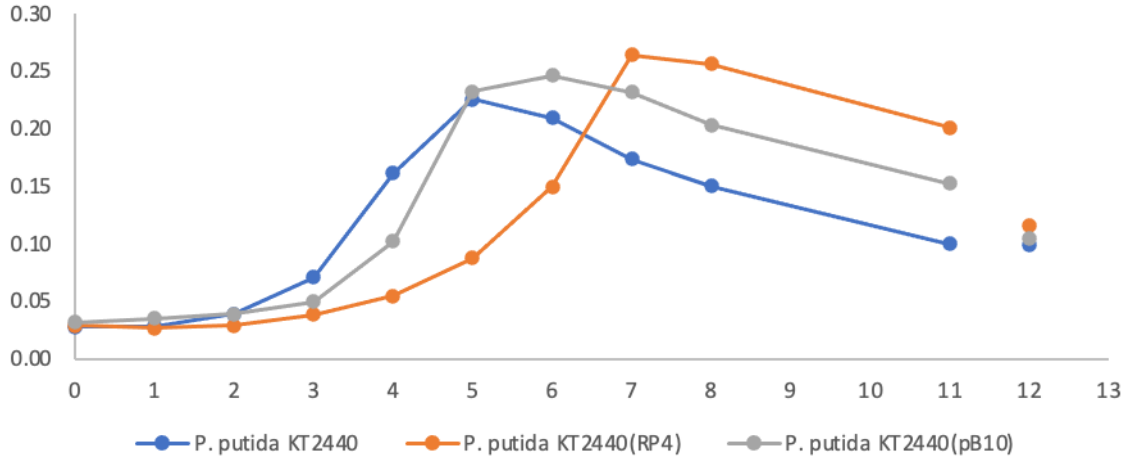
The spectra were recorded with integration time of 40s directly from minimal medium (40 to 50 repetitions [Table 1]). The spectrum of E. coli W3110 strain is dominated by bands at, the spectrum of P. putida KT2440 strains.

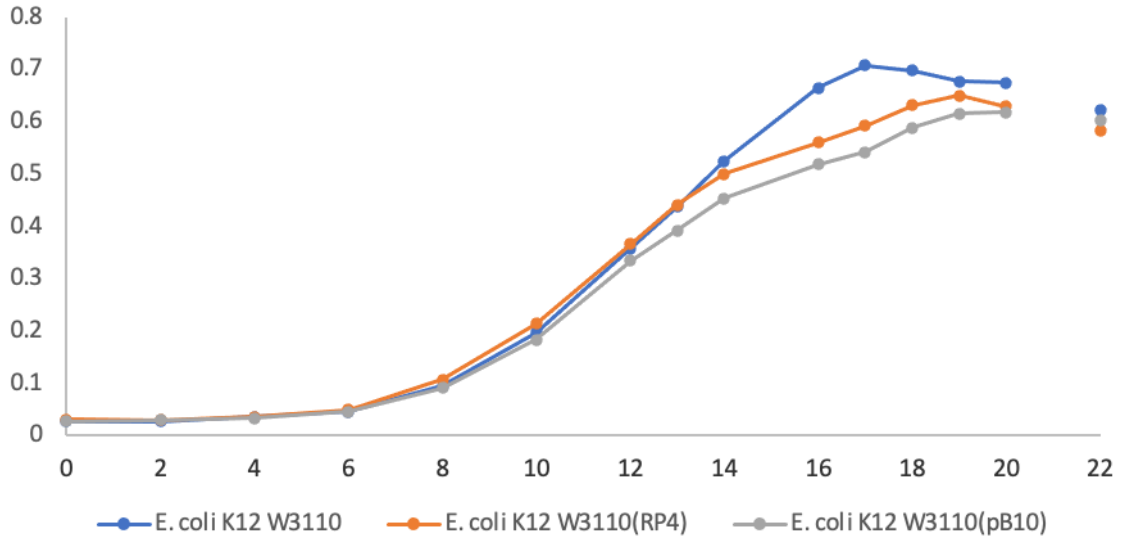
For a distinct identification of different strains, a reliable data analysis method is required.

Band assignments for Raman spectra of E. *coli* W3110 strains.

3.1 Materials and methods

Cells were pre-cultured overnight in LB with antibiotic, and culture to minimal medium with primary OD_{600} = 0.025, shaking for 200 rpm 30C.



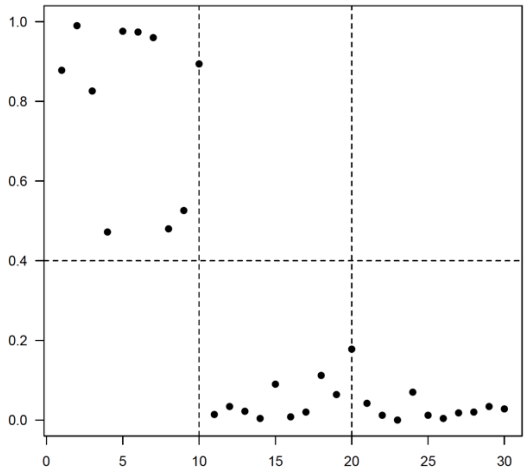


Reference growth curve

Strains	First time	Second time

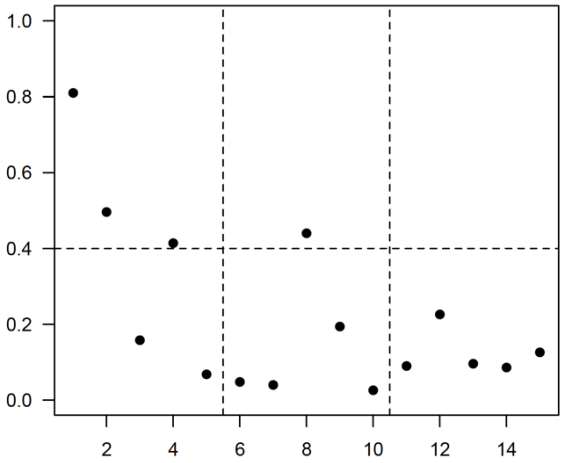
P. putid KT2440	3.5 h	9.5h
P. putid KT2440(pB10)	5.5h	10h
P. putid KT2440(RP4)	6h	22h
E. coli W3110	8h	24h
E. coli W3110(pB10)	9h	25h
E. coli W3110(RP4)	10h	26.5h

First time

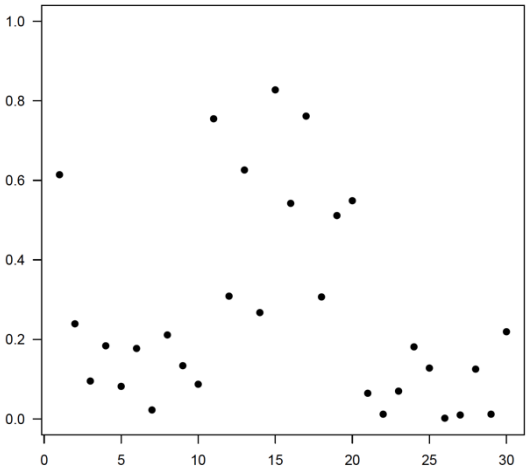


KT KT(pB10) KT(RP4)

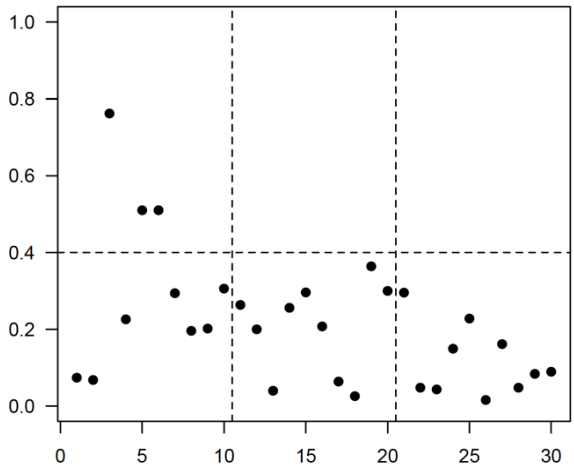
Second time



KT KT(pB10) KT(RP4)



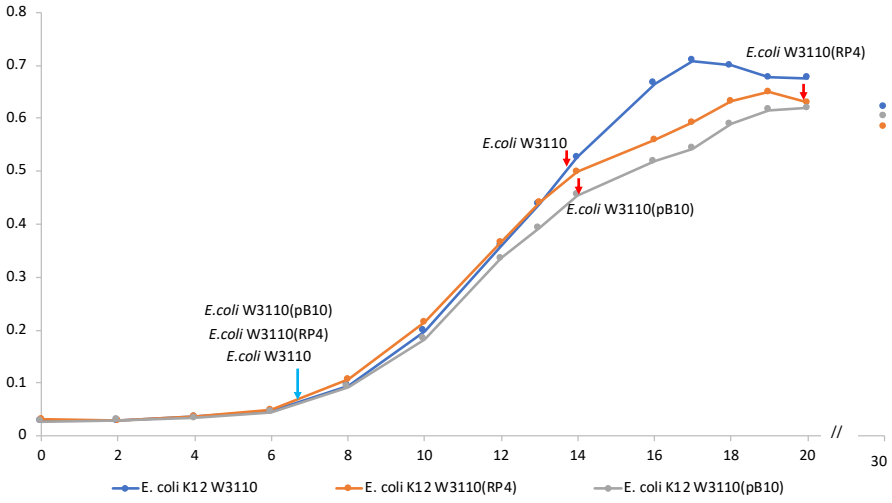
W3110 W3110(pB10) W3110(RP4)



W3110 W3110(pB10) W3110(RP4)

Conclusion:

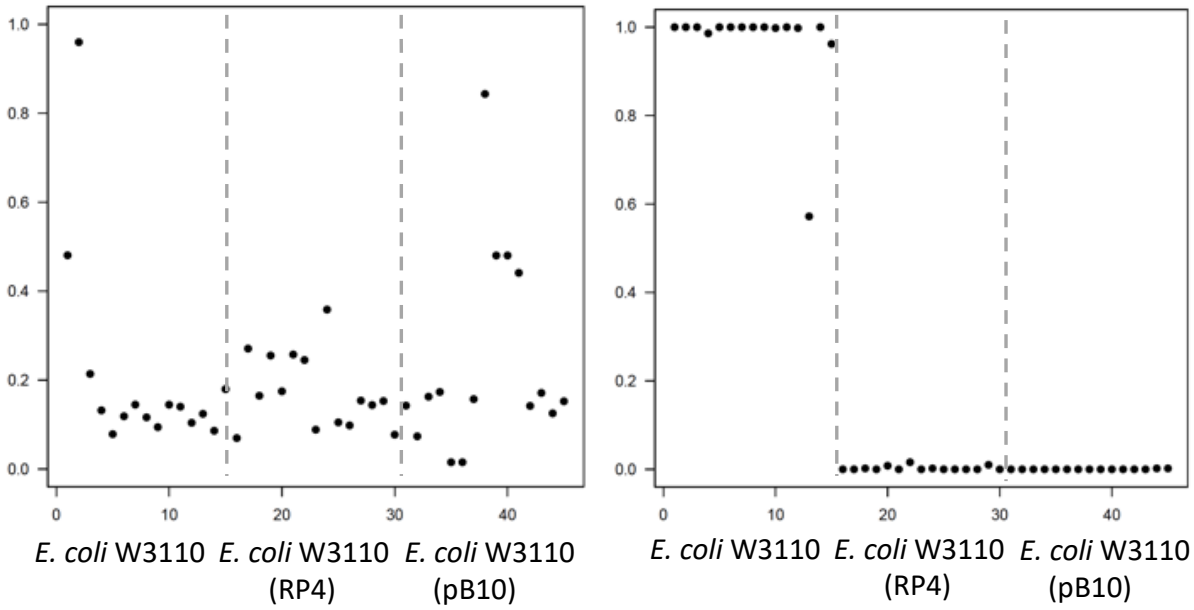
Difference seems can be detected, need more cells for confirmation.



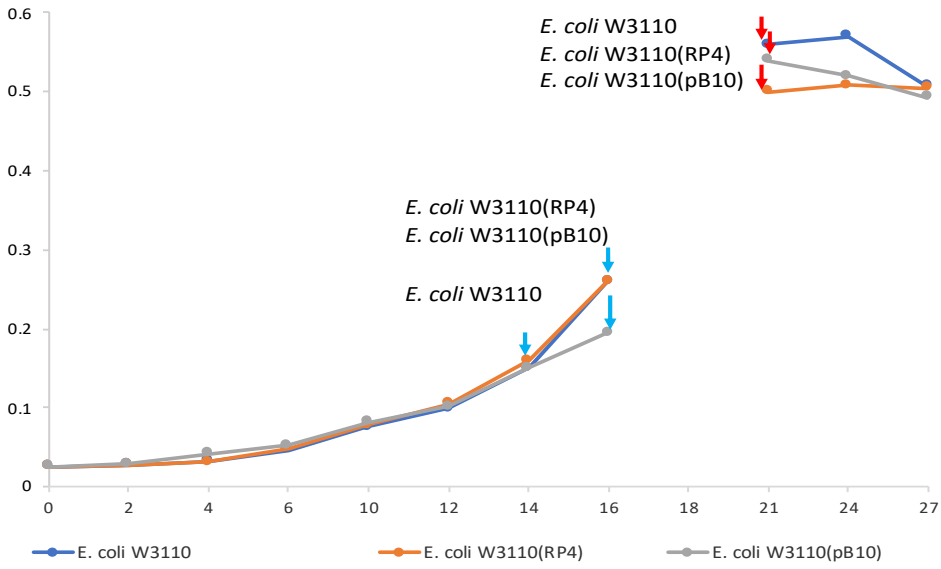
Reference growth curve with real-time sampling points

Strains	First time	OD_{600}	Second time	OD_{600}
E. coli W3110	7h		24h	0.45
E. coli W3110(pB10)	9h	0.065	26h	0.458

E. coli W3110(RP4)	8h	0.0473	25h	0.437

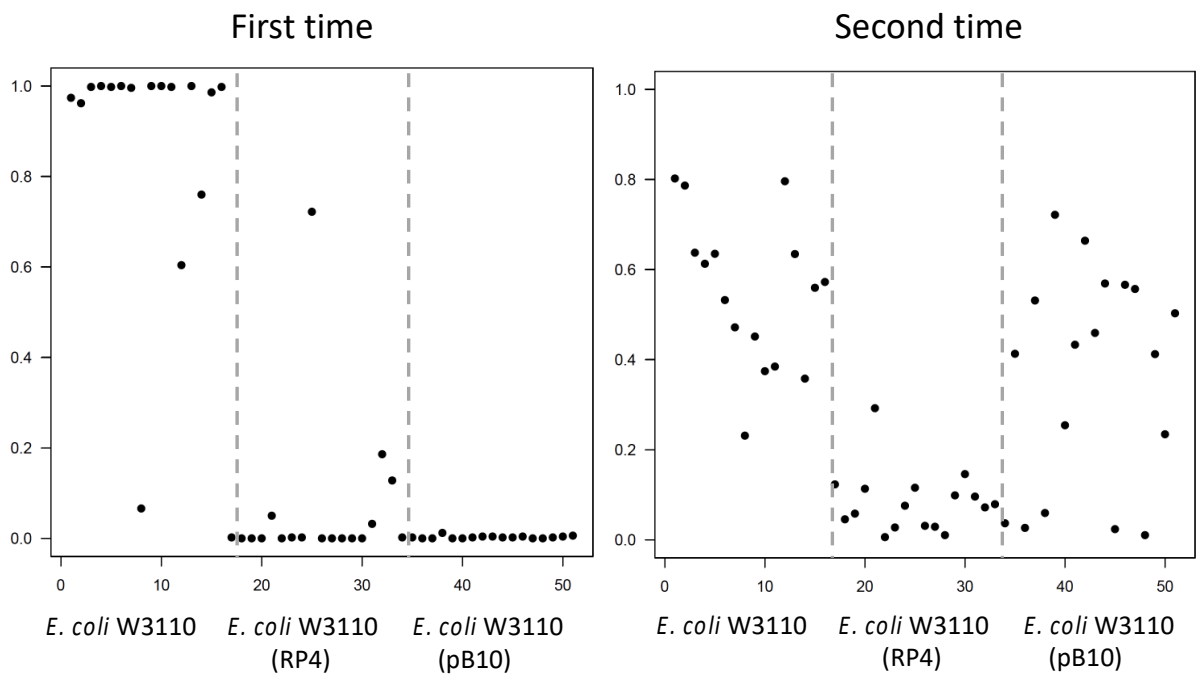


Point: Growth curve should be taken from real time.



Growth curve and sampling points in real-time

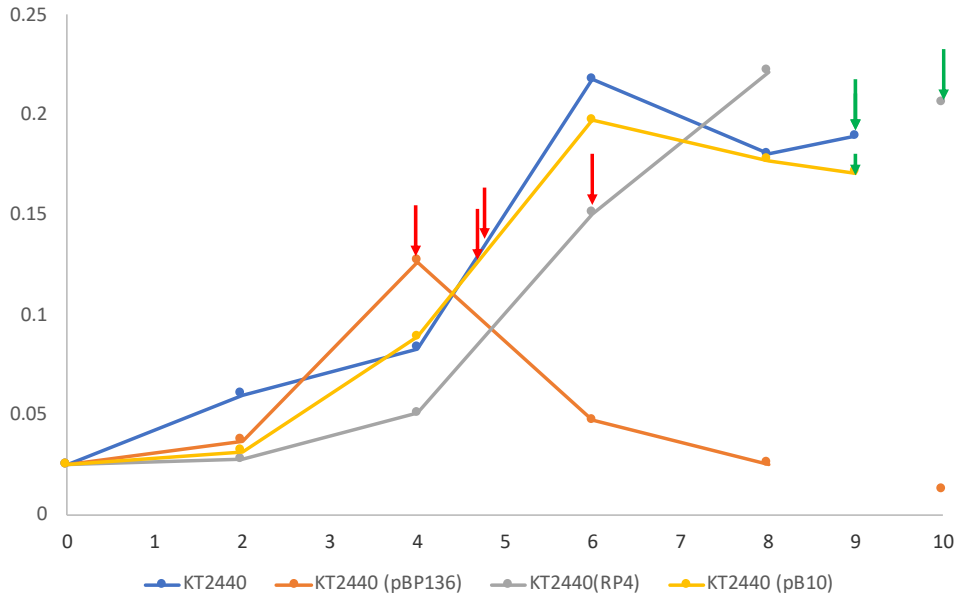
Strains	First time	OD_{600}	Second time	OD_{600}
E. coli W3110	7h		24h	0.45
E. coli W3110(pB10)	9h	0.065	26h	0.458
E. coli W3110(RP4)	8h	0.0473	25h	0.437



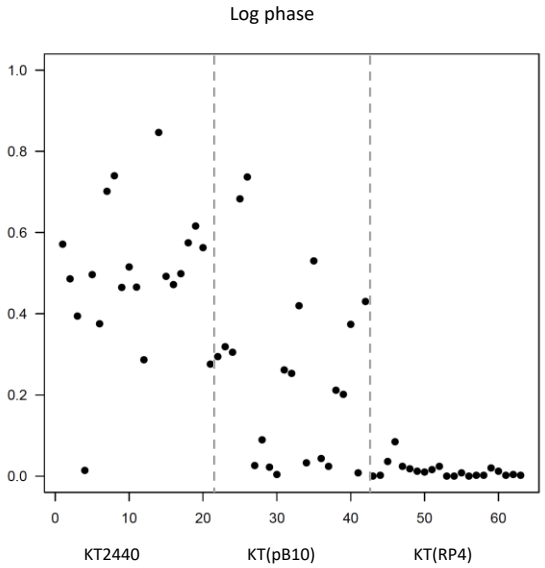
Conclusion:

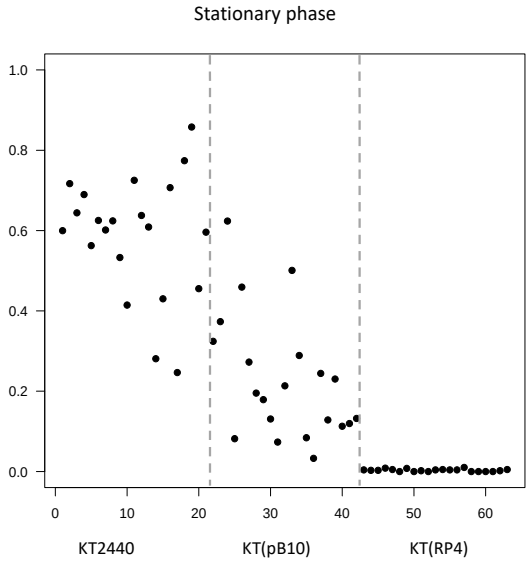
For *E. coli* W3110 strains, log phase sample can be distinguished

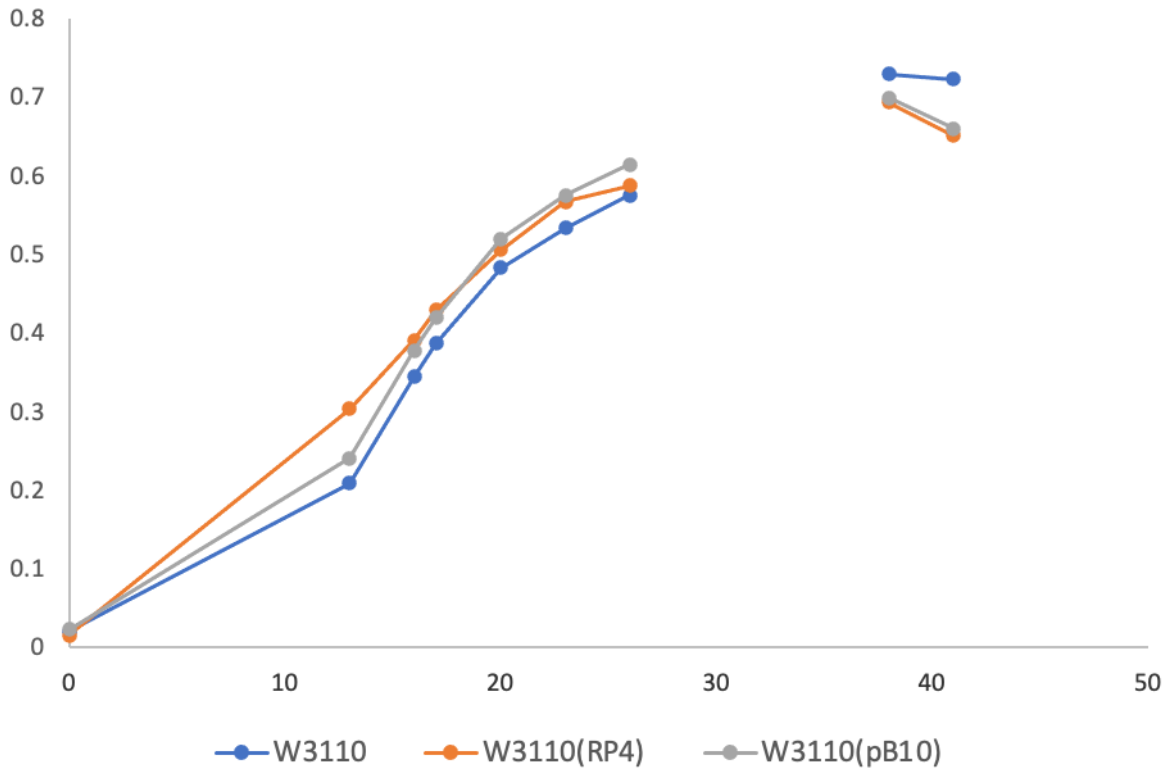
For CA10dm4RG strains, not so clearly, maybe because of late time points of 2 strains, need batch culture and take log phase sample in from same time-points.



Strains	Log phase	Real OD	Stationary phase	Real OD
P. putid KT2440	5 h	0.148	9h	0.1895
P. putid KT2440(pB10)	5h	0.1353	9h	0.17
P. putid KT2440(RP4)	6h	0.151	10h	0.2058







Strains	First time	OD_{600}	Second time	OD_{600}
E. coli W3110	17 h		24h	0.45
E. coli W3110(pB10)	13 h	0.065	26h	0.458
E. coli W3110(RP4)	16 h	0.0473	25h	0.437

3.2 Results and discussion

3.3 Discussion

To checking generality of this method, more plasmids from different Incompatibility (Inc) groups will be introduced for Raman spectroscopy observation (Table 1). The research in our lab shows that various plasmids, including IncP-1, P-7, P-9, W group plasmids did not load fitness cost on *P. resinovorans* CA10dm4 host, it is insensitive to various plasmid. As a control, the *P. resinovorans* CA10dm4 will be introduced as another host to check if this property can be reflected by Raman spectrum.

Table 1. The host strains and plasmids used in Raman spectroscopy.

Host	Plasmid

	IncP-1	IncP-7	IncP-9	IncW
E. coli W3110	RP4	pCAR1	pWW0	R388
P. putida KT2440	pB10	Rms148	NAH7K2	
P. resinovorans CA10dm4	pB136	pDK1K		

IncP: Incompatibility groups are classified based on a *Pseudomonas* strain as a recipient, and IncP-1 corresponds to IncP in the *Escherichia coli* plasmid classification.

As table 1 shows, the combination between the two hosts *P. putida* KT2440, *E. coli* W3110 and plasmids pB10, RP4 were used for Raman spectroscopy and classification analysis. The combination of these two hosts and *P. resinovorans* CA10dm4 with the left plasmids will be used for Raman spectroscopy observation in my future work to predict an integral regular pattern of plasmid harboring with Raman spectrums.

Denoise with PCA

Huang C-K, Ando M, Hamaguchi H-o, Shigeto S. 2012. Disentangling Dynamic Changes of Multiple Cellular Components during the Yeast Cell Cycle by in Vivo Multivariate Raman Imaging. *Analytical Chemistry* 84: 5661-68

Raman spectroscopy also presents disadvantages. For instance, the Raman signals of certain compounds can be quite weak, making them difficult to detect or undetectable. The Raman signal of certain compounds can be composed of several peaks, or be unknown. Also, the

background of samples can interfere with the Raman signal of bacteria. The equipment can be quite costly, depending on the type of Raman spectroscope.

Reference

[1] Shintani M, Takahashi Y, Tokumaru H, Kadota K, Hara H, Miyakoshi M, Naito K, Yamane H, Nishida H, Nojiri H. Response of the Pseudomonas host chromosomal transcriptome to carriage of the IncP-7 plasmid pCAR1. 2010. *Environmental Microbiology*. 12(6): 1413–1426

[2] Takahashi Y, Shintani M, Takase N, Kazo Y, Kawamura F, Hara H, Nishida H, Okada K, Nojiri1 H. Modulation of primary cell function of host *Pseudomonas* bacteria by the conjugative plasmid pCAR1. 2015. *Environmental Microbiology* 17: 134-155

[3] Lee S, Choi H, Cha K, Chung H. Random forest as a potential multivariate method for near-infrared (NIR) spectroscopic analysis of complex mixture samples: Gasoline and naphtha. 2013. *Microchemical Journal* 110: 739–748.

[4] Shintani M, Matsui K, Inoue J, Hosoyama A, Ohji S, Yamazoe A, Nojiri H, Kimbara K, Ohkuma M. Single-Cell Analyses Revealed Transfer Ranges of IncP-1, IncP-7, and IncP-9 Plasmids in a Soil Bacterial Community. 2014. *Applied and Environmental Microbiology* 80: 138–145

[5] Babic A, Lindner AB, Vulic M, Stewart EJ, Raman M. Direct visualization of horizontal gene transfer. 2008. *Science* 319: 1533-1536

Alvarez-Martinez CE, Christie PJ. 2009. Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* 73: 775-808

Ding H, Hynes MF. 2009. Plasmid transfer systems in the rhizobia. *Can J Microbiol* 55: 917 27

Gogarten JP, Townsend JP. 2005. Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* 3: 679-87

Heuer H, Smalla K. 2012. Plasmids foster diversification and adaptation of bacterial populations in soil. *FEMS Microbiol Rev* 36: 1083-104

Koraimann G, Wagner MA. 2014. Social behavior and decision making in bacterial conjugation. *Front Cell Infect Microbiol* 4: 54

Palonpon AF, Sodeoka M, Fujita K. 2013. Molecular imaging of live cells by Raman microscopy. *Curr Opin Chem Biol* 17: 708-15

Smith R, Wright KL, Ashton L. 2016. Raman spectroscopy: an evolving technique for live cell studies. *Analyst* 141: 3590-600

Thomas CM, Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer

between bacteria. *Nat Rev Microbiol* 3: 711-21

Top EM, Holben WE, Forney LJ. 1995. Characterization of diverse 2,4-dichlorophenoxyacetic acid-degradative plasmids isolated from soil by complementation. *Appl Environ Microbi*Heuer, H. and Smalla, K. (2012) 'Plasmids foster diversification and adaptation of bacterial populations in soil', *FEMS microbiology reviews*, 36(6), pp. 1083–1104.

Kumar, S. *et al.* (2020) 'Rapid detection of bacterial infection and viability assessment with high specificity and sensitivity using Raman microspectroscopy', *Analytical and Bioanalytical Chemistry*, 412(11), pp. 2505–2516. doi: 10.1007/s00216-020-02474-2.

Maquelin, K. *et al.* (2003) 'Prospective Study of the Performance of Vibrational Spectroscopies for Rapid Identification of Bacterial and Fungal Pathogens Recovered from Blood Cultures', *Journal of Clinical Microbiology*, 41(1), pp. 324–329. doi: 10.1128/JCM.41.1.324-329.2003.

Nakamoto, K. (2006) 'Infrared and R aman Spectra of Inorganic and Coordination Compounds', *Handbook of Vibrational Spectroscopy*.

Puppels, G. J. *et al.* (1990) 'Studying single living cells and chromosomes by confocal Raman microspectroscopy', *Nature*, 347(6290), pp. 301–303. doi: 10.1038/347301a0.

Rösch, P. *et al.* (2005) 'Chemotaxonomic Identification of Single Bacteria by Micro-Raman Spectroscopy: Application to Clean-Room-Relevant Biological Contaminations', *Applied and Environmental Microbiology*, 71(3), pp. 1626–1637. doi: 10.1128/AEM.71.3.1626-1637.2005.

Shintani, M. *et al.* (2014) 'Single-cell analyses revealed transfer ranges of IncP-1, IncP-7, and IncP-9 plasmids in a soil bacterial community', *Applied and environmental microbiology*, 80(1), pp. 138–145.

Takahashi, Y. *et al.* (2015) 'Modulation of primary cell function of host P seudomonas bacteria by the conjugative plasmid pCAR 1', *Environmental microbiology*, 17(1), pp. 134–155.

Trevors, J. T. (1986) 'Plasmid curing in bacteria', *FEMS Microbiology Reviews*, 1(3–4), pp. 149–157. doi: 10.1111/j.1574-6968.1986.tb01189.x.