Approche métagénomique pour l’étude de la dégradation de la quinoléine dans les sols
Jun Yuan

To cite this version:


HAL Id: tel-00955925
https://tel.archives-ouvertes.fr/tel-00955925
Submitted on 5 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
ECOLE CENTRALE DE LYON

THESE

pour obtenir le grade de

DOCTEUR

Spécialité : «Génomique microbienne environnementale»

préparée dans le laboratoire Ampère
dans le cadre de l’Ecole Doctorale «Électronique, Électrotechnique et Automatique»

par

Jun YUAN

Approche métagénomique pour l’étude de la
dégradation de la quinoléine dans les sols.

Soutenance prévue le 20 décembre 2012 devant la commission d'examen

JURY

Rapporteur Jean Luc Pernodet Directeur de Recherche CNRS-Université Paris-Sud
Rapporteur Christine Cagnon Maître de Conférences(HDR)-Université de Pau et des Pays de l'Adour
Examineur Pascale Bauda Professeur à l’Université de Lorraine
Encadrant Laure Franqueville Ingénieur de Recherche CNRS-Ecole Centrale de Lyon
Examineur Timothy M. Vogel Professeur -Université de Lyon
Directeur de thèse Pascal Simonet Directeur de Recherche CNRS-Ecole Centrale de Lyon
Acknowledgements

I am very glad to have this moment expressing my sincere thanks. First of all, I would like to thank the opportunity given by the Ecole centrale de Lyon and the China Scholarship Council who offered me the bursaries and enabled me to enroll for this degree.

I would like to thank my supervisors: Dr Pascal SIMONET, Research Director of Environmental Microbial Genomics Group in Laboratoire Ampère, UMR 5005 CNRS at the Ecole Centrale de Lyon and Dr Xiaojun ZHANG, Laboratory of Molecular Microbial Ecology and Ecogenomics, Shanghai Jiao Tong University. They awakened my interest in this topic and gave me continuous guidance, support, encouragement, and invaluable advices. I also would like to express my deepest gratitude to my co-supervisor, Dr Laure Franqueville, Environmental Microbial Genomics Group in Laboratoire Ampère, UMR 5005 CNRS at Ecole Centrale de Lyon who has provided me with valuable guidance in every stage of the writing of my thesis. Without her enlightening instruction, impressive kindness and patience, I could not have completed my thesis.

I appreciate the people from scientific comity: Jean Luc Pernodet, Christine Cagnon, and Pascale Bauda. Thanks for their time reading and evaluating my work.

During my Genefish project, I would like to thanks our group for their great contributions in this project including the former members: Laurent Philippot, Microbiologie des Sols-Géosols, UMRA111, CMSE, INRA, 21065 Dijon cedex, France and Jean-Claude Lazzaroni, Unité de Microbiologie, Adaptatation et Pathogénie, UMR5240, Université de Lyon, 69622 Villeurbanne cedex, France. Nathalie Lombard, Laure Franqueville, Samuel Jacquiod, Aurelie Faugier. I also would like to thanks Sébastien Cécillon, for his help in data analysis and GC/MS manipulation.

In my daily work, I have been extremely lucky and blessed with many cheerful and friendly colleagues, friends and non-academic staff: Timothy M. Vogel, Elisabeth Navarro, Samuel Jacquiod, Tom O. Delmont, Laurine Blanchard, Alban Mathieu, Joseph Nesme, Lorrie Macario, Jérémie Reboulet, Sébastien Cécillon, Sandrine Demanèche, Richard Barthollet, Monique Lacroix, Jean-Michel Monier,. Here, I would like to express my special thanks to Samuel Jacquiod and Tom O. Delmont for their selfless help especially during the early days when I joined the group.
Table of content

Acknowledgements .................................................................................................................... 1

Abbreviations and definitions .................................................................................................. 6

Résumé ....................................................................................................................................... 8

Abstract .................................................................................................................................... 10

General introduction ............................................................................................................... 12

Part I: quinoline biodegradation in soil microcosm ............................................................. 17

1. Bibliographical review of quinoline and its biodegradation ............................................... 17

   1.1 Quinoline biodegradation .......................................................................................... 18

       1.1.1 *Pseudomonas* sp. for quinoline biodegradation .............................................. 18

       1.1.2 *Rhodococcus* sp. for quinoline biodegradation ............................................... 20

       1.1.3 *Comamonas* sp. for quinoline biodegradation ................................................. 21

       1.1.4 *Thauera*.sp for quinoline biodegradation ........................................................ 22

1.2 Metabolic quinoline pathway ..................................................................................... 24

       1.2.1 Metabolic quinoline pathway by *Pseudomonas*.sp ......................................... 24

       1.2.2 Metabolic quinoline pathway by *Rhodococcus*.sp ........................................... 25

       1.2.3 Metabolic quinoline pathway under anaerobic/anoxic condition .................... 25

1.3 Genes involved in the quinoline biodegradation ........................................................ 26

2. Materials and methods ..................................................................................................... 30

   2.1 Quinoline microcosm ................................................................................................. 30

   2.2 RISA (rRNA intergenic spacer analysis) .................................................................... 31

   2.3 GC/MS analysis for quinoline biodegradation ........................................................... 32

3. Results .................................................................................................................................. 33

   3.1 Nucleotide BLAST of *ber* operon in the Rothamsted metagenome ............................. 33
3.2 RISA results ............................................................................................................... 34
3.3 Results of GC/MS analysis ........................................................................................ 40

4. Discussion and conclusion ........................................................................................... 42

PART II: in vitro development and use of Genefish to capture targeted DNA fragments47

1. Background ..................................................................................................................... 47

2. Bibliographical review of Lambda-Red homologous recombination and Genefish approach48

2.1 Lambda-Red homologous recombination ................................................................. 48

2.1.1 Introduction ............................................................................................................. 48

2.1.2. Lambda-Red recombination system .................................................................. 49

2.1.2.1 Overview of Lambda-Red recombination system .......................................... 49

2.1.2.2 Different vectors for the Lambda-Red system ................................................. 53

2.1.2.3 Substrates for the Lambda-Red system ............................................................ 56

2.1.2.3.1 Double-strand DNA recombination ......................................................... 57

2.1.2.3.2 Single-strand DNA recombination ............................................................ 59

2.2. The “Genefish” tool .................................................................................................. 62

2.2.1 General presentation of “Genefish” ....................................................................... 62

2.2.2 The suicide cassette ............................................................................................... 63

2.2.3 Homologous recombination with Lambda-Red system ........................................ 66

2.2.4 bcr operon ............................................................................................................. 66

3. Genefish application using bcr operon ......................................................................... 67

3.1. Capture plasmid and host strain construction ............................................................. 67

3.1.1 Materials: ............................................................................................................. 67

3.1.2 Methods: ................................................................................................................ 68

3.1.2.1 Highly conserved bcr fragment sequences determination ............................. 68
3.1.2.2 Capture plasmid construction ............................................................... 69
3.1.2.3 Host strain construction ........................................................................ 69
3.1.2.4 Escape rate test ..................................................................................... 70
3.1.3 Results ............................................................................................................. 71
3.1.3.1 Highly conserved \textit{bcr} fragment selected for capture plasmid construction .......................................................................................................................... 71
3.1.3.2 Capture plasmid construction ............................................................... 74
3.1.3.3 Host strain construction and escape rate test ........................................ 74
3.2. Genefish tool application .......................................................................................... 77
3.2.1 Materials .......................................................................................................... 77
3.2.2 Methods.......................................................................................................... 78
3.2.2.1 Co-electroporation ................................................................................ 78
3.2.2.2 Plasmid segregation .............................................................................. 79
3.2.3 Results ............................................................................................................ 80
3.2.3.1 Co-electroporation by using bcr c-d fragment ...................................... 80
3.2.3.2 Plasmid segregation by using pBAD35K7toxN-\textit{bcr}-d ....................... 82
3.2.3.3 Plasmid segregation by using pBAD35K7toxN-\textit{bcr}-a ....................... 83
3.3. Genefish improvements ............................................................................................ 84
3.3.1. Materials ......................................................................................................... 84
3.3.2 Methods.......................................................................................................... 85
3.3.2.1 Improvement of co-electroporation and plasmid segregation efficiency through culture time decrease ................................................................................. 85
3.3.2.2 Single-copy plasmid construction ........................................................ 86
Abbreviations and definitions

- **Δara**: Genotype of the bacteria means the bacteria lacks the gene *ara* and the function of arabinose metabolism. The gene in the chromosome of Genefish strain was deleted by genetic manipulation.

- **BAC**: Bacterial Artificial Chromosomes, BAC vectors were created from part of F’ plasmid and also were developed to hold much larger pieces of DNA than a plasmid can.

- **BLAST**: Basic Local Alignment Search Tool, this tool can compare nucleotide or amino acid sequences to DNA or protein data bases thus find out the similarity between sequences.

- **Bp/kb**: base pairs/kilobase pairs, as the units to describe (calculating the base pairs (A-T, C-G)) the size of DNA.

- **BGA**: Between Group Analysis, a supervised method for sample discrimination and class prediction based on the PCA.

- **Cb**: Carbenicillin, a bacteriolytic antibiotic belonging to the carboxypenicillin subgroup of the penicillins. It was used to screen pSIM6 containing colony in this study.

- **CFU**: Colony-Forming Units. Transfer diluted bacterial culture into medium plates by spreading; the colonies were formed from a single bacterial cell.

- **Cm**: Chloramphenicol, a prototypical broad-spectrum antibiotic, against a wide variety of Gram-positive and Gram-negative bacteria. It was used to screen capture plasmid or recombinants containing colony in this study.

- **DNA**: Deoxyribonucleic acid (DNA) molecular, the carrier of genetic information is biological macromolecules consisting of four nucleotides: A, G, C, T and deoxyribose connected by Phosphodiester bond.

- **ER**: Escape Rate=number of survival colony/total cell number (probability that the host strain escape from death after the toxic cassette induction)

- **GC/MS**: Gas chromatography–mass spectrometry, a method that combines the features of gas-liquid chromatography and mass spectrometry to identify different substances in sample. In this study, it was used to analyze the quinoline degradation in the soil microcosm.

- **IPTG**: Isopropyl β-D-1-thiogalactopyranoside, IPTG is a stable and highly-inducing chemical agent. In this study, it was used to induce the *relF* gene in Genefish capture plasmid.
- **Kn**: Kanamycin, an aminoglycoside antibiotic, commonly used to treat a wide variety of infection. It was used to screen our host strain in this study.

- **LB**: Luria-Bertani medium, a common medium for *E.coli*. Antibiotic can be added to select targeted bacteria.

- **mRNA**: messenger RNA, containing genetic information which also act as the template for protein synthesis.

- **PCA**: Principal Component Analysis, a mathematical procedure that converts a set of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. It can parse out the main factors from diverse system, simplify complex problems. In this study, PCA was used to analyze the main bacterial composition which responding to quinoline degradation.

- **PCR**: The polymerase chain reaction (PCR) is a rapid and specific *in vitro* technology for amplification of target DNA through several rounds of denaturation, annealing and extension. This technology is widely used in gene cloning, sequences analysis, regulation of gene expression and gene polymorphism studies.

- **RecBCD**: It is an enzyme complex (composed of three different sub-units RecB, C and D) of the *E.coli* bacterium that initiates *in vivo* homologous recombination.

- **RISA**: The rRNA intergenic spacer analysis (RISA) involves PCR amplification of the rRNA gene operon (the intergenic spacer region) between the small (16S rDNA) and large (23S rDNA) subunits which were used as bacterial species signature in phylogeny. These PCR products are compared through electrophoretic analysis for rapid examination of complex bacterial communities.

- **rRNA**: Ribosomal ribonucleic acid (rRNA) is one kind of RNA that binds to ribosomal proteins thus form the ribosome which is the factory for protein synthesis in all living cells.

- **SOB**: Super Optimal Broth, nutrient rich medium for *E.coli* culture.

- **ssDNA/dsDNA**: single stranded DNA/ double stranded DNA. Most of the DNA exists in double-helical structure, but after heat or alkali treatment it becomes single-stranded structured.
Résumé

Grâce au développement des technologies de métagénomique au cours des dix dernières années, il a été constaté que les micro-organismes représentent la plus grande ressource de diversité métabolique et génétique sur Terre. En effet, un gramme de sol contient $10^9$ cellules bactériennes et $10^3$-$10^4$ différentes espèces bactériennes. Certaines sont en mesure de réaliser des réactions enzymatiques conduisant à la dégradation complète de certains polluants toxiques pour l’environnement comme les composés organiques tels que la quinoléine. Cependant, l’immense réservoir de molécules et enzymes microbiennes n’a pas encore été exploité, car plus de 99% d’entre elles ne sont, pour l’instant, pas cultivables \textit{in vitro}.

Mon travail s’inscrit dans le cadre d’une collaboration entre l’Université SJTU (Shanghai Jiao Tong Université en Chine) et le groupe de G. M.E (Génomique Microbienne Environmentale) du laboratoire Ampère à l’Ecole Centrale de Lyon. Nos partenaires à l’Université SJTU ont construit un réacteur de dénitrification à l’échelle du laboratoire capable de dégrader la quinoléine en retirant la demande chimique en oxygène. Un nouvel outil appelé “Genefish” a été développé dans notre laboratoire comme une méthode alternative de la métagénomique pour aider à la découverte de nouveaux gènes d’intérêt industriel ou environnemental. A la suite des premiers travaux réalisés dans notre laboratoire, ma thèse présentée ici comporte deux parties.

Dans la première partie de ce travail, nous avons étudié le potentiel de dégradation de la quinoléine présente dans les bactéries d’un sol de référence largement étudié au laboratoire. Pour cela nous avons mis en place des expériences de microcosme qui visent à révéler la diversité potentielle des bactéries responsables de la dégradation de la quinoléine. Des analyses comparatives des profils RISA (Ribosomal Intergenic Spacer analysis) nous ont permis de mettre en évidence des changements dans la structure de la communauté des bactéries du sol incubé en conditions aérobie et anaérobie en présence de quinoléine. La dégradation de la quinoléine a été confirmée par technique de GC/MS (Gas Chromatography-Mass Spectrometry). Les travaux futurs seront de vérifier la communauté de bactéries responsables de la dégradation de quinoléine en utilisant la technique de NGS (Next Generation Sequencing).

Le deuxième objectif de ma thèse a été d’utiliser Genefish dont la finalité est de capturer des gènes ciblés (le gène \textit{bcr} qui serait responsable de la dégradation de quinoléine dans le réacteur de
nos partenaires) dans l'ADN métagénomique extrait du sol. Genefish consiste à élaborer une souche d'*E.coli* incluant un plasmide de capture permettant de pêcher les gènes recherchés dans un échantillon d’ADN métagénomique par recombinaison homologue. Le plasmide de capture comprend une cassette de deux gènes toxiques pour la souche qui activés par induction chimique vont permettre la sélection positive directe des clones recombinants, et deux sites multiples de clonage dans lesquels sont insérées les zones de recombinaison qui vont jouer le rôle d’hameçons. Nous avons testé la capacité de Genefish à capturer des produits PCR du gène *bcr*, l’efficacité de recombinaison reste faible à cause de la persistance de plusieurs copies du plasmide suicide dans la cellule après l’événement de recombinaison. Par conséquent, trois stratégies ont été essayées pour améliorer l’efficacité: la co-électroporation, la ségrégation de plasmide et la construction de plasmide suicide en mono-copie. Finalement, la stratégie de la ségrégation plasmidique fonctionne mais l'efficacité de recombinaison est encore trop faible peut-être due à l’incertitude des modèles de recombinaison homologue. Les travaux futurs se concentreront sur l'amélioration des fréquences de recombinaison par transfert de fragments du plasmide de capture dans le chromosome de la souche Genefish.

**Mots clés**

Métagénomique, gène *bcr*, quinoléine, microcosme, communauté des bactéries, RISA, GC/MS

Genefish, Lambda Red, recombinaison homologue, cassette toxique, taux d’échappement,
Abstract

As the development of metagenomic technologies in the past ten years, it is unuestionned that microorganisms encompass the largest resource of metabolic and genetic diversity in the world. Actually, one gramme of soil contains more than $10^9$ bacteria and $10^3$-$10^4$ species. Some of their members are able to carry out enzymatic reactions leading to the complete degradation of pollutants (such as quinoline). So, the biodegradation of some highly toxic or organic compounds by microorganisms will be a general trend for pollutant treatment. However, the huge reservoir of molecules and enzymes from microorganisms still need to be explored because more than 99% of microorganisms cannot be cultivated \textit{in vitro}.

My work was based on collaboration between the University SJTU and Ecole Centrale de Lyon. Our partners at the University SJTU have built a laboratory scale denitrification reactor which was capable of degrading quinoline by removing the chemical oxygen demand. A new tool called "Genefish" has been developed in our laboratory as an alternative method for metagenomics which aims to discover novel industrial or environmental genes of interest. Following the early work in our laboratory, my thesis is presented here in two parts.

In the first part, we set up a quinoline microcosm experiment both under aerobic and anaerobic condition using reference soil extensively studied in the laboratory at Ecole Centrale de LYON. This work aimed to reveal the potential bacterial diversity and even genes responsible for quinoline degradation. We used RISA(Ribosomal intergenic Spacer analysis) to analyze the bacteria community structure changes and GC/MS (Gas Chromatography-Mass Spectrometry) was also used to detect the quinoline degradation and reveal potential quinoline metabolic pathways under aerobic and anaerobic condition. Results showed great bacteria community structure changes and high quinoline degradation activity after the quinoline addition under aerobic condition. The future work is to investigate the bacteria community which may be responsible for quinoline degradation using the technique of NGS (Next Generation Sequencing).

The second object of my thesis was to use the Genefish tool to capture targeted genes
(the *bcr* gene responsible for the quinoline degradation in the wastewater treatment bioreactor) from the soil metagenome. The aim was to construct an *E.coli* strain containing a capture plasmid and Red system for capturing targeted genes from metagenomic DNA by homologous recombination. The capture plasmid includes a toxic cassette consisting of two suicide genes which can be activated by chemical induction, finally support the positive recombinants selection. It also contains two multiply cloning sites in which highly conserved sequences were inserted and works as the bait during recombination. We have tested the capacity of Genefish to capture the PCR products of *bcr* gene; the efficiency was low because of the persistence of several copies of the capture plasmid into the Genefish strain after recombination events. So, three strategies were tried to improve the recombination efficiency: co-electroporation, plasmid segregation and mono-copy capture plasmid construction. Finally, the strategy of plasmid segregation works but the recombination efficiency was still low maybe caused by the uncertain model of homologous recombination. The further research will focus on the transfer of the toxic cassette and homologous arms into the host strain chromosome, this new strategy will exclude the bad effect of low copy number capture plasmid, uncertain model of λ Red induced homologous recombination and the homologous arms site in the capture plasmid which are the most important factors influencing the homologous recombination efficiency in Genefish.

**Keywords**

Metagenomics, *bcr* gene, quinoline, microcosm, bacteria community, RISA, GC/MS

Genefish, lambda Red, homologous recombination, toxic cassette, Escape rate
General introduction

Thousand years ago, human knew little about microorganisms but they already use them to produce wine, vinegar and preserve food through appropriate methods. Antoni van Leeuwenhoek (1632-1723), the “father” of microbiology was the first to observe and describe single cell organisms through his handcrafted microscopes. However, in that era, his discovery didn’t catch enough attention and people lack the technology to recognize the tiny microorganisms. Two centuries later, the microbiology researches entered into physiology stage. The French microbiologist Louis Pasteur (1822-1895) considered that fermentation is caused by microorganisms, and he also proposed the “Germ theory of disease” and invented vaccine to cure rabies and anthrax. His works in microbiology greatly improved the development of medicine and fermentation industry. Start from the 20th century, microbiology developed quickly. In 1928, Alexander Fleming discovered antibiotic substance called penicillin, afterwords, more and more antibiotics have been screened after his achievements. In 1953, James Watson and Francis Crick discovered the double helix molecular structure of DNA which promoted the development of modern microbiology. Because of the cooperation, mutual promotion and interpenetration between microbiology, biochemistry, biophysics and molecular biology, rapid development of microbiology has been achieved not only in fundamental research but also in other applications such as agriculture, food industries, clean bio-energy production, medical treatment, etc.

Microorganisms appear everywhere and can live as individuals or clusters in the environment. They are not only ubiquitous, but also essential for all life in the world, because they represent the primary source of nutrients, and the primary recyclers of dead matter back to available organic form (Wooley, Godzik et al. 2010). Microorganisms are considered as the oldest life forms on earth, some of them have existed for billion years (Benzerara, Menguy et al. 2006). Scientists found microbial community that can survive in marine sediments with low rates of metabolism in the deep North Pacific Gyre for million years (D'Hondt, Spivack et al. 2009). Microorganisms exist in mild habitats such as the water we drink, the air we breathe. They can grow in extreme environments such as hypersaline (Ollivier, Caumette et al. 1994), high temperature (Ranieri and Boor 2009), low oxygen (Mesbah and Wiegel 2008), alkaline (Grant and Heaphy 2010), high pressure (Kato and Qureshi 1999), polluted sites such as mineral
water (Uroz, Calvaruso et al. 2009), etc. Because of their incredible adaption capacity, microorganisms attract extensive attention and are studied worldwide. They are gradually applied in every aspect of our life such as health care (Bixquert Jimenez 2009), environmental protection (Jose, Giridhar et al. 2011; Sasaki, Morita et al. 2011), agriculture (Bossio, Scow et al. 1998), industrial production of enzymes (Yamabhai, Buranabanyat et al. 2011), even biomass electricity and fuel generation (Juang, Yang et al. 2011; Wei, Liang et al. 2011). Although we know microorganisms have played central roles in Earth’s climatic, geological, geochemical and biological evolution (Xu 2006), but only a few of them can be cultivated in vitro.

One gram of soil may contain 10 billion microorganisms and thousands of species (Torsvik, Goksoyr et al. 1990), this makes soil the most biologically diverse ecosystem on Earth. The traditional culture-dependent approaches helped us to isolate some microorganisms but they cannot provide us the information on microbial communities of natural environments. These technical limitations block the people away from discovering the great amount and unknown reservoir of microbial resource. Over the last few decades, culture-independent approaches have significantly contributed to our understanding of microbial diversity and community composition in the environment (Zengler 2009). In order to exploit the huge genetic resource in uncultivable environmental bacteria, Handelsman et al (Handelsman, Rondon et al. 1998) proposed a novel method which is called metagenomics and it is now widely used. In this strategy, genomic DNA of microorganisms in environmental samples is extracted directly and cloned into plasmids or fosmids to form clone libraries (Fig.1). A clone library containing inserted heterologous genomic DNA then can be used to

1) Reveal microbial diversity (Handelsman 2004), gene diversity (Cowan, Meyer et al. 2005) in the environmental samples through high throughput sequencing.

2) Use fluorescence in situ hybridization (FISH) (Cottrell, Waidner et al. 2005; Ardura, Planes et al. 2011) to detect colonies containing target genes.

3) Select positive colonies possessing specific function (Uchiyama and Miyazaki 2009) according to their phenotypes-driven analysis.
Fig. 1: Principle of Metagenomic process: Genomic DNA extracted from soil or other environmental samples is cloned into plasmid such as fosmid thus forming the clone library. This clone library with inserted genomic DNA will be sequenced and analyzed for metabolism function, phylogeny and other researches.
However, metagenomic approaches still have a number of limitations that need to be resolved. The most difficult problem is the low resolution of microbial community when cloning environmental DNA samples of high complexity. Because of the fragments diversity, DNA from more dominant bacteria could be cloned and this usually causes difficulties in linking specific microbes (not predominant) to specific ecological functions, underestimating the real microbial diversity and even missing the rare community members (Warnecke and Hugenholtz 2007). According to the metagenomic process, the DNA fragments are digested and randomly cloned into vectors and this automatically decrease the opportunity to capture genes which are responsible for some specific functions. Even if metagenomic DNA libraries with million clones encompass the huge bacterial genomic diversity, this strategy still costs a lot and is not very flexible if we want to compare microbial community or specific genes in various environmental samples.

Coking wastewater has large amount of toxic refractory organic content and without dealing with this direct emissions, the waste cause great pollution to the environment and also direct threat to human health. Coking wastewater treatment process has been long-term studied and the use of microbial biodegradation is the main method. In order to reveal the bacterial community structure during the quinoline biodegradation, our partners in Shanghai Jiao Tong University monitored the wastewater treatment process and constructed a laboratory scale denitrifying reactor which was estimated to degrade quinoline and remove the chemical oxygen demand. They found *Thauera*.sp enriched with Benzoyl-coenzyme A reductases operon was abundant in the reactor and they may be involved in the quinoline degradation. In our lab, soil of Rothamsted located at the Rothamsted experimental station in England was used as the resource for studying natural selection, biodiversity and the project of Metasoil. A metagenomic database (fosmid library of two million clones, pyro-sequenced) was generated from the soil of Rothamsted by former PhD student Tom O. Delmont (Delmont, Prestat et al. 2012). *bcr* operon sequences (coding for Benzoyl-coenzyme A reductases (BCRs)) were also found existing in this metagenomic database. In order to benefit from this resource, the first part of my work mainly focuses on enrichment and identification of the potential quinoline degrading bacterial community from the Rothamsted soil. Find out whether the *bcr* operon is essencial for quinoline degradation through microcosm and if there is *bcr* operon, we then can use traditional
metagenomic approach to capture relevant genes. Three objectives were included in this study:

1) Investigate the soil microbial community potential related to quinoline degradation under aerobic and anaerobic condition.

2) Compare the microbes responsible for quinoline degradation in this study with previous researches. This may help us to reveal the microbial diversity even new microbes related to quinoline degradation.

3) Find potential genes and even metabolic pathway of quinoline degradation under aerobic and anaerobic condition.

Overall, revealing the microbial diversity could lead to better understand the relationship between microorganisms and functions. Metagenomics is still a common tool to disclose the relationship between the genetic composition of environmental microorganisms and its community function. The function of interested genes in nature may help us to expand the capacity of wastes degradation and finally ameliorate the environment which has been increasingly polluted during the past decades. In order to capture target genes and fragments more efficiently and precisely from environmental metagenomes, a new tool “Genefish” was constructed in our group as a complementary method for metagenomics. Genefish is included in the European project Metaexplore which aims to discover and characterize new enzymes from environmental microbial community. The molecules of interest in this project are implicated in the biodegradation of recalcitrant (chitin, lignin) and xenobiotic halogenated compounds. In view of the initial results obtained in our group by Nathalie Lombard, Samuel Jacquiod and Laure Franqueville, the Genefish tool was constructed and presented promising performance.

The second part of this work was to develop and use the “Genefish” technology to capture \textit{bcr} operon. This work was in co-tutorial with Shanghai Jiao Tong University and EMG (Environmental Microbial Genomics) group of the Ampère Lyon laboratory (Lyon University). Two objectives were included in this study:

1) Genefish experiments with the genes of interest

Genefish system including the toxic cassette, \(\lambda\) Red system and the host strain have been validated by Nathalie Lombard (Thesis 2007), Laure Franqueville and Samuel Jacquiod (Thesis 2012). Here, we use \textit{bcr} operon as the target for Genefish application.

2) Genefish limitations and its improvements
Part I: quinoline biodegradation in soil microcosm

1. Bibliographical review of quinoline and its biodegradation

Quinoline (C₉H₇N) is an heterocyclic aromatic nitrogen organic compound with a double-ring structure that contains a benzene ring fused to pyridine at two adjacent carbon atoms which is usually found in coal tar, oil shale, chemical manufacturing plants and even groundwater near creosote wood preservation (Aislabie, Bej et al. 1990; Bai, Sun et al. 2010). Quinoline and its derivatives were widely used as some industrial intermediates, solvents, dyes, pharmaceuticals and pesticides (S.Fetzner 1998). Quinoline exhibits a particular chemical structure with a strong pharmacological property and is widely used as "parental" compounds to synthesize molecules with medical benefits, such as anti-malarial, antibiotics, anti-microbial, even anti-tumoral activities (Solomon and Lee 2011). Recently, Styrylquinoline derivatives have gained strong attention due to their activity as probable HIV integrase inhibitors (Normand-Bayle, Benard et al. 2005; Pommier, Johnson et al. 2005). Other studies show that some new 8-hydroxyquinoline derivatives possess antifungal and photosynthesis-inhibiting activity (Musiol, Jampilek et al. 2007; Musiol, Tabak et al. 2008).

However, coking wastewater contains large amount of quinoline, and its waste cause great pollution to the environment and also direct threat to human health. Due to its carcinogenic and mutagenic toxicity as well as its strong odor (Hirao, Shinohara et al. 1976; Barrick, Furlong et al. 1984; LaVoie, Shigematsu et al. 1984), waste materials containing quinoline must be eliminated before discharge. Thereby, complete degradation of such xenobiotic molecules represents a big challenge for environment preservation. Physical and chemical treatments for the removal of nitrogen organic compounds are usually based on hydrodenitrogenation and hydrodesulfurization (Sugaya 2001). Many microscale/nanoscale catalysts like phase-pure TiO₂ (Mao, Zhang et al. 2010), doped TiO₂, and WO₃ (Rockafellow 2009) were studied, Mao and Zhang (Mao, Zhang et al. 2010) use and develop P25 film in a recirculation reactor to remove quinoline from synthetic wastewater. However, this physical and chemical methods for quinoline degradation should be processed under high-temperature and high-pressure conditions, this also requires special equipment, extra cost and the wastewater after the treatment still need following operations to degrade quinoline completely.(Yu 1989).
Studies also showed that quinoline can be biodegraded under aerobic or anaerobic conditions by microorganisms such as *Pseudomonas* (Carl and Fetzner 2005; Bai, Sun et al. 2010), *Rhodococcus* (O’Loughlin 1996), *Desulfobacterium* (Licht 1997), *Moraxella* (Crescitelli 1950), *Nocardia* (Shukla 1987), *Burkholderia* (Jianlong, Xiangchun et al. 2002; Wang, Wu et al. 2004), *Comamonas* (Chen, Cui et al. 2003; Cui 2004), white rot fungus (Zhang 2007), and *Thauera* (Mao, Zhang et al. 2010).

1.1 Quinoline biodegradation

Quinoline is stable and slightly soluble in water, it is difficult to be degraded in natural environments which will tend to accumulation. Since the study of Antranikian (Antranikian, Vorgias et al. 2005), microorganisms have been considered as a clean resource for the application in future biotechnological processes such as toxic aromatic compounds degradation. As microorganisms encompass the largest resource of metabolic and genetic diversity, some of their members are able to carry out enzymatic reactions leading to the complete degradation of organic pollutants from contaminated sites (Kahng, Kukor et al. 2000; Oie, Albaugh et al. 2007). Here, we summarize the information of microorganisms, metabolic pathways and expected genes involved in quinoline biodegradation.

1.1.1 *Pseudomonas* sp. for quinoline biodegradation

*Pseudomonas* was described as the dominant polycyclic aromatic hydrocarbon-degrading bacteria the first round by Shukla (Shukla 1986) from sewage by enrichment with 0.03% quinoline in a phosphate salt medium culture. They found that isolated *Pseudomonas* sp. played an important role in quinoline degradation. It rapidly grew with a quinoline concentration of 0.01 to 0.015% (w/w) while it grew with a time delay at 0.025 to 0.03 % (w/w). Higher concentrations of quinoline were toxic and completely abolished the growth of this strain (Shukla 1986). In another study, two strains of *P aeruginosa* and *P putida* were isolated from oil contaminated and a creosote contaminated soil respectively and using quinoline as a sole source of carbon and nitrogen (Bai, Sun et al. 2010). *P aeruginosa QP* and *P putida QP* completely degraded quinoline within 24 hours and 12 hours respectively when they inoculated in phosphate salts medium containing 2.5 mM quinoline. However, these strains only degrade quinoline under aerobic
conditions, no quinoline degradation activity was demonstrated under denitrifying conditions, because this strain could not utilize nitrate as the electron acceptor during respiration. Another bacterial strain, BW003, isolated and identified as Pseudomonas sp. by Qinghua Sun (Sun, Bai et al. 2009; Bai, Sun et al. 2010) This strain can degrade 192 to 911 mg/L of quinoline with a removal rates ranging from 96% to 98% in 3 to 8 hours under optimum conditions at 30°C with pH 8 (Fig.2). Moreover, quinoline and its metabolic by-products can be eliminated from wastewater by controlling the C/N ratio using BW003. They compared the concentration variation of ammonia-N transformed from quinoline by BW003 by using two different media: MSM (Mainstream media) and MSM+glucose. Results show ammonia-N was much lower than that in MSM alone media, the author suggested that the extra ammonia-N was not utilized by BW003 when the available carbon source (quinoline and its metabolic products) was exhausted. However, in the MSM+glucose media, glucose supplied the carbon and energy source when the available carbon source was used up while the bacteria continues the synthesis of ammonia-N into the cells of bacteria or the transformation of ammonia-N into NO$_3^-$N. An immobilized cell bioreactor inoculated with Pseudomonas sp1 and Pseudomonas sp2 was successfully used to degrade benzene, naphthalene, phenanthrene, pyridine, quinoline, and isoquinoline in the wastewater discharging from a biomass gasification power-generation plant (Tian, Qian et al. 2006). After microbial adaptation phase, stable metabolic phase and high efficient reaction phase, the COD (chemical oxygen demand), and aromatic compounds in the bioreactor were removed efficiently.

**FIG.2**: Biodegradation of quinoline in different initial concentrations by the strain Pseudomonas sp. BW003 (Sun, Bai et al. 2009).
1.1.2 Rhodococcus sp. for quinoline biodegradation

Two aerobic gram-positive bacteria strains identified as Rhodococcus sp. Q1 and Rhodococcus sp. QL2 which were capable of utilizing quinoline as the dominant carbon, nitrogen as source of energy were isolated from soil (O’Loughlin 1996) and from activated sludge of a coke plant wastewater treatment process (Sun, Bai et al. 2009). However, 3.88mM of quinoline was toxic to Rhodococcus sp Q1. Results from Aislabie (Aislabie, Bej et al. 1990) indicated extra nitrogen sources like ammonium sulfate, potassium nitrate and potassium nitrite added to the substrate can stimulate the growth of Rhodococcus sp. Q1 and Rhodococcus sp. QL2. The efficiency of their promotional effects was as follows: (NH4)2SO4 >NH4NO3 >NaNO3 >Urea. Quinoline degradation by Rhodococcus sp. QL2 at different initial concentrations (37 °C, pH 8.0, and 150 rpm) is shown in Fig.3 (Zhu, Liu et al. 2008). Rhodococcus sp. QL has a strong degradation capacity of quinoline at low substrate concentration (60–120mg/L) within 5 and 7 h. 60–120mg/L was a quinoline concentration that detected in most contaminated soil and discharged wastewater(O’Loughlin 1996), so the strain has a great advantage of solving practical quinoline pollution. However, we still can see that Pseudomonas sp has a better capacity of quinoline degradation. Compared to Rhodococcus sp., Pseudomonas sp. BW003 can degrade quinoline at a concentration more than 900mg/L (Fig.2) while 680mg/L of the quinoline is toxic for Rhodococcus sp. (Fig.3).

![Graph](image.png)

Fig.3: Different initial concentrations of quinoline were used to test the capacity of quinoline degradation by Rhodococcus sp. QL2. C is quinoline concentration in medium, and C0 is initial quinoline concentration (Zhu, Liu et al. 2008).
1.1.3 Comamonas sp. for quinoline biodegradation

A Comamonas sp. has been isolated and identified as an aerobic strain which can degrade quinoline efficiently (Ulonska 1995; Sugaya 2001; Chen, Cui et al. 2003; Cui, Chen et al. 2003; Cui 2004). Sugaya (Sugaya 2001) used a strain Comamonas sp TKV3-2-1 to degrade quinoline in crude oil. Concentration changes of different nitrogen compounds (nitrogen concentration in quinoline, 2-Hydroxy-quinoline and ammonia) were recorded during the quinoline degradation process (Fig.4). Their results show that concentration of nitrogen derived from quinoline (CQn-N) in crude oil (7mM/L) disappeared within about 30 minutes and about 74% of the quinoline contained in crude oil was converted into 2-Hydroxy-quinoline. After 4 hours incubation, 2-Hydroxy-quinoline in water and crude oil were totally degraded; ammonia nitrogen concentration and total nitrogen concentration reached the highest level in the water phase. Over 90% of the quinoline metabolic products containing nitrogen were present as water-soluble substances and about 62% of them were accumulated as ammonia. Cui et al (Cui 2004) chose TOC (total organic carbon) to monitor the mineralization extent of quinoline by Comamonas sp. using quinoline as the sole carbon and energy resource. Fig 5 shows that the TOC decreased quickly at the beginning of quinoline degradation and nearly 70% of TOC was removed after 10 hour’s culture. The author suggested that this residual TOC could probably correspond to some persistent metabolites, and maybe also included soluble polymers from the biomass.

![FIG.4 Concentration changes of nitrogen component during quinoline-degradation reaction by Comamonas sp TKV3-2-1 in crude oil (Sugaya 2001).](image)

- ◆: Concentration of nitrogen derived from quinoline in crude oil (= CQn-N), ▲: concentration of nitrogen derived from 2-hydroxy-quinoline in crude oil [= (C2-OHQn-N) C], △: concentration of nitrogen derived from 2-hydroxy-quinoline in water phase [= (C2-OHQn-N) W], ◊: ammonia nitrogen concentration in water phase (= CNH4-N), ●: Total nitrogen concentration in water phase (= CT-N).
FIG. 5 TOC (total organic carbon) removal during quinoline biodegradation by Comamonas Q10 (Cui 2004). After 18 hours, quinoline was completely degraded and TOC also was removal with efficiency around 70%.

1.1.4 *Thauera.sp* for quinoline biodegradation

*Thauera.sp* was characterized by its ability to use aromatic hydrocarbons under anoxic conditions through a unique pathway. A laboratory-scale denitrifying reactor was adapted by using of seeding sludge of wastewater treatment plant and this reactor was estimated to degrade quinoline and remove the chemical oxygen demand (Bin Liu 2006). The reactor reached a steady state after a 6 weeks’ adaptation and the removal efficiencies for quinoline and COD on five consecutive days are shown in Fig. 6. DGGE (denaturing gradient gel electrophoresis) and Real Time-PCR results based on the sequences of 16s V3 region indicated that species of the genera *Thauera* and *Azoarcus* increased in abundance by about one order of magnitude during the period of adaptation. Microbial community of the reactor was dominated by *Thauera* related bacteria with an abundance of 74% and this suggested they might play an important role for quinoline and chemical oxygen demand removal under denitrifying conditions.

Three *thauera* strains *thauera* sp. 3-35, *thauera* sp.Q4 and *thauera* sp.Q20-C with identical 16S rRNA genes were isolated from a full-scale coking waste water treatment plant (Mao, Zhang et al. 2010). Their aromatic compound-degrading capacity was evaluated under both aerobic and anoxic conditions.
Fig. 6. Quinoline and chemical oxygen demand removal efficiencies over 5 consecutive days monitoring (Bin Liu 2006). Results showed that quinoline and chemical oxygen demand were steadily removed in the Lab-scale reactor.

After 4 days aerobic incubation with *Thauera* sp. 3-35, *Thauera* sp.Q4 and *Thauera* sp.Q20-C, except quinoline, almost all of the organic pollutants in the coking wastewater were completely degraded. This result demonstrated that these *Thauera* strains can degrade a wide spectrum of aromatic compounds under aerobic condition. However, they cannot degrade the same aromatic compounds under anoxic conditions. Comparing to aerobic conditions, most of the known *Thauera* species (including *T. aromatica* and *T. aminoaromatica* (Anders, Kaetzke et al. 1995; Mechichi, Stackebrandt et al. 2002) have more versatile aromatic compound-degrading ability under denitrification conditions. So, a denitrification condition seems to be necessary for the degradation of quinoline by *Thauera*.sp.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Condition</th>
<th>Quinoline Concentration</th>
<th>Time for completed degradation</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pseudomonas</em> sp. (Shukla 1986)</td>
<td>Oxic</td>
<td>10-15 mg/L</td>
<td>24h</td>
<td>30 °C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25-30 mg/L</td>
<td>Delay</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>30 mg/L</td>
<td>Toxic</td>
<td></td>
</tr>
<tr>
<td><em>Pseudomonas.aeruginosa</em> and <em>Pseudomonas.pseudola</em> (Aislabie, Bej et al. 1990)</td>
<td>Oxic</td>
<td>320 mg/L</td>
<td>24h</td>
<td>28 °C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>320 mg/L</td>
<td>12h</td>
<td></td>
</tr>
<tr>
<td><em>Pseudomonas</em> sp. BW003 (Bai, Sun et al. 2010)</td>
<td>Oxic</td>
<td>192 to 911 mg/L</td>
<td>3-8h</td>
<td>30 °C</td>
</tr>
<tr>
<td><em>Rhodococcus</em> sp. Q1 (O’Loughlin 1996)</td>
<td>Anoxic</td>
<td>60-120 mg/L</td>
<td>5-7h</td>
<td>35 °C</td>
</tr>
<tr>
<td><em>Rhodococcus</em> sp. QL2 (Zhu, Liu et al. 2008)</td>
<td>Anoxic</td>
<td>360 mg/L</td>
<td>12h</td>
<td>37 °C</td>
</tr>
<tr>
<td><em>Comamonas</em> sp TKV3-2-1(Sugaya 2001)</td>
<td>Anoxic</td>
<td>900 mg/L</td>
<td>0.5h</td>
<td>30 °C</td>
</tr>
<tr>
<td><em>Thauera</em>.sp (Bin Liu 2006)</td>
<td>Anoxic</td>
<td>40 mg/L</td>
<td>6 weeks</td>
<td>30 °C</td>
</tr>
</tbody>
</table>

Table1: Brief conclusion of quinoline biodegradation ability by different strain under different conditions. *Pseudomonas* sp. BW003 was the best strain to complete degradation of quinoline under oxic condition, while only *Thauera*.sp was reported to degrade quinoline under anoxic condition with a low efficiency.
1.2 Metabolic quinoline pathway

As we have mentioned before, microorganisms can degrade different kind of aromatic compounds that can be found in polluted wastes. Recently, two main possible pathways for the quinoline transformation were established from studies developed on *Pseudomonas* and *Rhodococcus* respectively. Grant and Al-Najjar (Grant and Al-Najjar 1976) suggested that the transformation starts with an hydroxylation at position 2 of the heterocyclic aromatic ring thus forming 2-hydroxyquinoline. Additional hydroxylation steps transform 2-hydroxyquinoline into 2, 6-dihydroxyquinoline and a trihydroxyquinoline. Based on this theory, other researches figured out the real pathway according to their experiments.

1.2.1 Metabolic quinoline pathway by *Pseudomonas*.sp

Aislabie et al (Aislabie, Bej et al. 1990) considered quinoline is first metabolized into 2-hydroxyquinoline by *Pseudomonas aeruginosa* QP and *P. putida* QP. Both strains cannot metabolize 2-methylquinoline because the methyl group at position 2 blocks the formation of hydroxyquinoline. Shukla et al (Shukla 1986) found that *Pseudomomas* sp. strain grown in quinoline media metabolized quinoline with a maximum accumulation of metabolites after 16 hours’ of culture. Four metabolite products were identified as 2-hydroxyquinoline, 8-hydroxycoumarin, 2, 8-dihydroxyquinoline and 2, 3-dihydroxyphenylpropionic acid. 2, 8-dihydroxyquinoline was oxidized very slowly while further incubation resulted in the complete disappearance of other three substrates. This implicated 2-hydroxyquinoline, 8-hydroxycoumarin and 2, 3-dihydroxyphenylpropionic acid which might be the intermediate molecules during quinoline degradation. Bai (Bai, Sun et al. 2010) revealed that most of the quinoline is first transformed into 2-hydroxyquinoline then 2-hydroxyquinoline was converted into 2, 8-dihydroxyquinoline which was immediately and rapidly transformed into 8-hydroxycoumarin and ammonium. Furthermore, the quinoline nitrogen transformation was studied and they demonstrated that the final product containing nitrogen after quinoline biodegradation was mainly NH$_3$–N. Appropriate C/N ratio in medium could improve the efficiency of quinoline degradation and reduce the yield of NH$_3$–N in substrates. So, a proposed pathway for the metabolism of quinoline by *Pseudomonas* sp. is presented in Fig. 7.
1.2.2 Metabolic quinoline pathway by \textit{Rhodococcus}.sp

Schwarz et al. (Schwarz, Bauder et al. 1989) used three bacterial strains to investigate the quinoline catabolism. They found 2-oxo-1, 2-dihydroquinoline, 8-hydroxy-2-oxo-1, 2-dihydroquinoline, 8-hydroxy coumarin, and 2, 3-dihydroxyphenylpropionic acid as intermediates of quinoline transformation conducted by \textit{Pseudomonas fluorescens} 3 and \textit{Pseudomonas putida} 86 which support the results of Shukla (Shukla 1986). However, other intermediates were found in the culture medium of a \textit{Rhodococcus} strain B1, such as 2-oxo-1, 2-dihydroquinoline, 6-hydroxy-2-oxo-1, 2-dihydroquinoline, and 5-hydroxy-6-(3-carboxy-3-oxo propenyl)-1H-2-pyridone. Fig.8 shows the proposed pathway for the transformation of quinoline by \textit{Rhodococcus} strain B1. Quinoline is degraded into 8-hydroxy coumarin and amonia after the third step by \textit{Pseudomonas} sp. While \textit{Rhodococcus} strain B1 need more steps to eliminate the nitrogen atom of quinoline and its pathway is still unknown.

1.2.3 Metabolic quinoline pathway under anaerobic/anoxic condition

Pereira et al. (Pereira 1987) studied the anaerobic transformation of quinoline in groundwater contaminated by wood-treatment chemicals under both field and laboratory conditions. Those groundwater samples contained quinoline, isoquinoline, methylquinolines, and acridines. The
results obtained illustrate the transformation of quinoline that occurred under anaerobic conditions starting with the hydroxylation at position 2 of the heterocyclic ring. This indicates that these anaerobic microorganisms use a similar quinoline transformation pathway than aerobic bacteria which has been described above. By using H$_2$O$^{18}$, their studies also showed that the oxygen required for the hydroxylation reaction is derived from water (Pereira 1988).

Wang et al. (Wang 1984; Wang, Suidan et al. 1984) acclimatized a methanogenic consortium with quinoline during 6 months under anaerobic conditions. They got 85% of the expected amount methane production, indicating that quinoline could be used as the sole resource under anaerobic conditions. However, no further intermediates were reported, and the pathway of quinoline metabolism under anaerobic conditions remains unknown. Brockman et al. (Brockman, Denovan et al. 1989) isolated two gram-negative bacteria from deep subsurface sediments. These two microorganisms mineralized quinoline under aerobic conditions and transformed quinoline to soluble intermediates under anaerobic conditions.

Liu et al. (Bin Liu 2006) used a laboratory-scale denitrifying reactor which was adapted from seeding sludge of wastewater treatment plant and was utilized to degrade quinoline and remove the chemical oxygen demand. After adaptation of *Thauera* and *Azoarcus* under anoxic and denitrifying conditions, quinoline and chemical oxygen demand removal was clearly noticeable suggesting that this consortium might play an important role.

### 1.3 Genes involved in the quinoline biodegradation

Microorganisms are able to transform or even mineralize many N-heterocyclic compounds of both natural and xenobiotic sources. Proposed pathways of bacterial quinoline degradation have been declared above. But so far, detailed knowledge of their gene organization and regulation is not yet available.

*Pseudomonas putida* 86 is able to utilize quinoline as sole source of carbon, nitrogen and energy; its degradation proceeds via the 8-hydroxycoumarin pathway (Aislabie, Bej et al. 1990). The first step of this pathway, quinoline 2-oxidoreductase (Qor), a molybdo-iron/sulfur-flavoprotein which belongs to the xanthine oxidase family of molybdenum enzymes catalyzes the formation of 2-hydroxyquinoline (Blase, Bruntner et al. 1996). In the
second step, 2-hydroxyquinoline is hydroxylated at position C-8 by 1H-2- oxoquinoline 8-monoxygenase (OxoOR), a two-component nonheme iron oxygenase that is encoded by oxoO (oxygenase component) and oxoR (reductase component). The oxoO gene is localized approximately 15 kb upstream of oxoR on the chromosome of strain Pseudomonas putida 86; the qorMSL genes coding for Qor are situated in between (Rosche, Tshisuaka et al. 1997). Further metabolites of the pathway such as 8-Hydroxycoumarin and 2, 3-dihydroxyphenylpropionic acid were identified but their related metabolic enzymes, genetic organization and regulation are still unknown. Carl et al. (Carl, Arnold et al. 2004) characterized a gene cluster involved in quinoline degradation in strain Pseudomonas putida 86. They located the gene upstream oxoO encoding the oxygenase component of OxoOR and the gene oxoS coding for a XylS-type protein (a transcriptional activator). Several putative ORFs on the 7.3kb DNA segment between oxoO and qorMSL, and the 3.3kb fragment between qorl and oxoR were also identified (Fig.9).

Dot blot analysis with RNA probes suggested that ORF1, ORF2, ORF7, ORF8 and ORF9, as well as qorL, oxoO and oxoR are specifically transcribed in the presence of quinoline or 2-hydroxyquinoline. Transcription of these genes also requires the presence of the oxoS gene product: OxoS, a XylS-type protein which acts as a quinoline–dependent transcriptional activator. They also found that ORFs7 to -9 can start the transcription weakly in the absence of quinoline, and even in an OxoS’ mutant strain. Low level expression of ORF7 to -9 that is independent of quinoline suggests that they might be essential for bacterial housekeeping functions. The gene cluster is probably transcribed from several promoters, resulting in multiple overlapping polycistronic mRNAs (Carl, Arnold et al. 2004).

Latest researches by Carl and Fetzner (Carl and Fetzner 2005) showed that the gene cluster of P. putida 86 involved in quinoline degradation is transcribed from several quinoline-dependent promoters (PoxoO, Porf3, PqorM, and PoxoR) in the presence of oxoS (Fig.10). Internal promoters like Porf3, PqorM, and PoxoR ensure adequate expression of distal genes of the operon in long and unstable transcripts, they also may provide differential expression of genes within an
operon in response to different physiological states or growth conditions (Kaebenick, Dittmann et al. 2002).

![Genetic map of the qor region and the specific promoters of P.putida 86(Carl and Fetzner 2005).](image)

As we described before, *T. aromaticca* has been reported as a microorganism degrading quinoline and removing the chemical oxygen demand under anoxic conditions (Bin Liu 2006). However, the metabolic pathway and genetic regulations of quinoline degradation by *T. aromaticca* is still unknown. Nevertheless, the author suggested that the *bcr* operons encoding for Benzoyl-coenzyme A reductases might be involved in the quinoline degradation.

Benzoyl-coenzyme A reductases (BCRs) are known as key enzymes in the anaerobic degradation of many aromatic compounds such as benzene, toluene. Benzoyl-CoA (BCoA) serves as the substrate for BCRs, which dearomatize the aromatic ring by reduction yielding to cyclohexa-1, 5-diene-1-carboxyl-CoA (dienoyl-CoA) (Loffler, Kuntze et al. 2011). There are two different classes of BCRs, while both of them yield the identical dienoyl-CoA product using different electron activation processes.

Genes coding for class I subunits of BCRs are present in the genome of denitrifying bacteria (*Thauera, Azoarcus, Magnetospirillum* species) or bacteria displaying an anoxygenic photosynthesis (*Rhodopseudomonas, Rhodomicrobium* species)(Boll 2005). A class 1 BCR enzyme has so far only been isolated and characterized in the denitrifying, facultative anaerobe *Thauera aromaticca* strain K172 (Boll and Fuchs 1995). This enzyme is extremely oxygen-sensitive and composed of four genes as BcrABCD. The AD subunits bind to one ATP molecule each and are linked by a [4Fe-4S] cluster, whereas the BC subunits are proposed to bind two further [4Fe-4S] cluster and the benzoyl-CoA binding active site (Boll, Fuchs et al. 2000).

Recently, the class II BCRs was discovered in the obligate anaerobic Fe (III)-respiring *Geobacter.metallireducens* which uses low molecular aromatic growth substrates (Kung, Loffler et al. 2009). Different proteome analysis results suggested that a cluster of eight benzoate inducible genes coding for a class II BCRs complex referred to BamBCDEFGHI (Wischgoll,
Heintz et al. 2005; Heintz, Gallien et al. 2009). No ATP-binding motif was found in the BamB-I complex, suggesting that electron transfer to BCoA may be independent of ATP hydrolysis. Fig.11 shows the reaction of two different classes of BCRs that are involved in BCoA degradation. (A): Cluster of genes coding for BCR (bcrA-D) and other enzymes (had, oah, and dch) of the BCoA degradation pathway in *T. aromatica*. Note that BcrA-D couples ring reduction to a stoichiometric ATP hydrolysis (upper BCR reaction). (B): Cluster of genes coding for putative BCR (bamB-I) and another enzyme (bamA, a homolog for oah) involved in the BCoA degradation pathway in *G. metallireducens* and figure 12 shows the protein cluster structure of BCR (bcrA-D) and BCR (bamB-I) (Kung, Loffler et al. 2009).

![Fig.11. Enzymatic reactions of BCRs and organization of genes involved in BCoA degradation in *T. aromatica* and *G. metallireducens* (Kung, Loffler et al. 2009)](image)

![Fig.12. (Left): Composition of the class I BCR in *T. aromatica* (Loffler, Kuntze et al. 2011) (Right): Putative composition of the class II BCR in *G. metallireducens* (Loffler, Kuntze et al. 2011)](image)
2. Materials and methods

2.1 Quinoline microcosm

The soil for the microcosm was collected from an untreated control plot of Park Grass, Rothamsted, England, in June 2010. The Park Grass soil is an internationally recognized resource for about 150 years and has been designated as a reference soil for metagenomic studies (Delmont, Robe et al. 2011; Delmont, Robe et al. 2011; Delmont, Prestat et al. 2012). The quinoline microcosm under aerobic condition was performed at room temperature in a fume cupboard. 50g soil, 20ml double distilled H₂O and different dose of quinoline (0mg, 2mg, 10mg, 25mg and 50mg) were added into the microcosm bottles, each bottle was closed with cheese cloth. Each condition (for example 2mg quinoline added, aerobic condition), triplicates (three bottles of soil microcosm) were considered as a group and double distilled H₂O was replenished to its original content three days after the incubation and before sampling. One gram of quinoline microcosm samples from each bottle was collected and stored at -20°C for further use. Fig. 13 shows the design and sampling time points of the aerobic quinoline microcosm. Samples collected before adding the quinoline (T0) are considered as the negative control.

*Fig. 13: Quinoline soil microcosm under aerobic condition. The top diagram shows the microcosms group for each condition; triplicates for each condition and three samples are collected from each repeat. The bottom diagram shows the sampling time point of each condition. T0 is considered as the negative control which is collected before adding quinoline.*
For anaerobic condition construction, argon (Linde France) was injected into the bottles with a pressure of 2 bars for 2 minutes. The bottles were previously closed with a rubber plug, and two syringes were inserted into the plug, one syringe for the argon injection into the bottle (green arrow), meanwhile the second one (red arrow) for air expelling out of the bottle. After the injection, syringes were pulled out and the bottles were stored at 22°C. But in this case, before adding quinoline, microcosm was maintained for a 16 days period under anaerobic condition for adaptation. After this period, soil was sampled (T0) and quinoline was added. Samples collected from natural soil (T00) are considered as the negative control and T0 are considered as the negative control under anaerobic condition. Fig.14 shows the design and sampling points of the anaerobic quinoline microcosms.

Metagenomic DNA from aerobic and anaerobic microcosms was extracted and purified for the further RISA analysis (Annex 1).

2.2 RISA (rRNA intergenic spacer analysis)

In this study, the intergenic spacer between the small (16S) and large (23S) subunit rRNA genes of microcosms’ metagenomic DNA were amplified using primers (RISA-fw: 5’-TGCGGCTGGATCCCCCTCCTTT-3’, RISA-rv: 5’-CCGGGTTCCCCATTCGG-3’)(Normand P. 1996). The amplified products then were loaded and migrated in gel by using Agilent DNA 1000 Kit to analyze the bacterial community changes. Principal Component Analysis (PCA) and Between Group Analysis (BGA) were used to reveal the internal structure of data extracted from the gel by the software R (Annex 2).
2.3 GC/MS analysis for quinoline biodegradation

Soil samples from aerobic quinoline microcosm (T0, T1, T2, T3, T4 and T5) and samples from anaerobic quinoline microcosm (T00, T1, T3, T5 and T6) were analyzed by GC/MS. Triplicated samples were extracted, filtered (Mao, Zhang et al. 2010) in CH₂Cl₂ (HPLC grade) and then analyzed by using Agilent 6850 equipped with a DB-5 column (30-m length, 0.25-mm inner diameter) and a flame ionization detector, using the following method: the injector temperature was 280°C; the column of the GC was retained at 70°C for 3 min, and then increased to 280°C with an increment of 5°C/min; the temperature for the MS ion source was 200°C and electron energy was 70 eV.
3. Results

3.1 Nucleotide BLAST of bcr operon in the Rothamsted metagenome

In order to investigate the existence of bcr operon in the soil microcosm, nucleotide BLAST using the Rothamsted metagenomic DNA database (Thesis Tom O. Delmont) and bcr c, bcr a and bcr d from Thauera. aromaticica as the query sequences were done. Results are shown in table 2, sequences producing significant alignments in database then were taken as the query sequences to do a nucleotide BLAST in NCBI.

<table>
<thead>
<tr>
<th>reference</th>
<th>size</th>
<th>identities</th>
<th>Score(bits)</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>bcr c</td>
<td>929bp</td>
<td>229/267 85%</td>
<td>228</td>
<td>5e-58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>240/302 79%</td>
<td>100</td>
<td>4e-19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>117/140 83%</td>
<td>96</td>
<td>7e-18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>68/78 87%</td>
<td>76</td>
<td>6e-12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>52/57 91%</td>
<td>74</td>
<td>2e-11</td>
</tr>
<tr>
<td>bcr a</td>
<td>1298bp</td>
<td>291/337 86%</td>
<td>287</td>
<td>1e-75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>262/314 83%</td>
<td>210</td>
<td>2e-52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>139/155 89%</td>
<td>180</td>
<td>2e-43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>215/257 83%</td>
<td>170</td>
<td>2e-40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>164/192 85%</td>
<td>159</td>
<td>6e-37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>216/266 81%</td>
<td>123</td>
<td>5e-26</td>
</tr>
<tr>
<td>bcr d</td>
<td>799bp</td>
<td>308/362 85%</td>
<td>289</td>
<td>2e-76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>273/321 85%</td>
<td>256</td>
<td>3e-66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>118/137 86%</td>
<td>121</td>
<td>9e-26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>71/80 88%</td>
<td>88</td>
<td>2e-15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>58/63 92%</td>
<td>86</td>
<td>4e-15</td>
</tr>
</tbody>
</table>

Table 2: Nucleotide BLAST using bcr c, bcr a and bcr d gene as the query sequences and the metagenomic DNA from Rothamsted (60Gbp pyrosequencing) as the database.

All matched sequences (marked out in table 2) from the Rothamsted metagenomic DNA database were verified by nucleotide BLAST on NCBI. Rhodopseudomonas.palustris (9 sequences), Magnetospirillum.sp (5 sequences) and Thauera.sp (3 sequences) contain the target bcr c sequence. However, no target sequence matched with the sequence got from the Rothamsted database which has an E-value of 4e-19 compared to bcr c gene. BLAST using sequences that matched with bcr a gene also got target sequences from Rhodopseudomonas.palustris (13 sequences), Magnetospirillum.sp (11 sequences) and Thauera.sp (16 sequences). Other sequences from BLASTochloris.sp (1 sequence), Paracoccus.sp
(3 sequences), *Bradyrhizobium.sp* (3 sequences), *Acidovorax.sp* (3 sequences), *Pseudomonas.sp* (2 sequences), *Ensifer.sp* (1 sequence), and uncultured bacterium (8 sequences) were also detected. BLAST using sequences that matches with *bcr d* gene only got target sequences of *Rhodopseudomonas.palustris* (12 sequences) and *Thauera.sp* (13 sequences).

So far, in Rothamsted soil, *bcr* operon might primarily exist in *Rhodopseudomonas.palustris*, *Magnetospirillum.sp* and *Thauera.sp* as BLAST using *bcr c*, a fragments all showed positive hits in those strains and they were also reported to contain *bcr* type benzoyl-CoA reductase genes (Song and Ward 2005).

### 3.2 RISA results

rRNA intergenic spacer PCR products from each serie of quinoline soil microcosm samples (triplicates) were analyzed. Fig.15 shows the electrophoresis results of aerobic quinoline soil microcosm, repeats are not shown here but will be shown in Between Group Analysis.

![Electrophoresis results using rRNA intergenic spacer PCR products amplified from aerobic quinoline soil microcosm samples. The band in purple is the upper marker and the band in green is the lower marker. Lane 0-7 means the sampling time points already described above. T0 means the negative control of microcosm (original soil samples before adding quinoline).](image-url)
The gel profile clearly shows the band variety between each time point and each condition:

First, in the gel profile of 0mg quinoline aerobic microcosm, the band framed by red rectangle (size between 700-850 bp) became stronger at the time point T4. When quinoline was added into each soil microcosm, the same band appears earlier at time point T2. However, the band intensity increased in parallel to the quantity of quinoline present in the microcosm. Second, the band close to the upper marker exists in 0mg and 2mg quinoline microcosm while becomes unstable at the time points T2 to T5 of 10mg, 25mg and 50mg quinoline added microcosm. Third, compared with the negative control (0mg quinoline added), the profiles changed (band density) especially after 10 mg, 25mg and 50mg quinoline addition

Fig.16 shows the electrophoresis grams of anaerobic quinoline soil microcosm, repeats are not shown here but in Between Group Analysis later.

![Electrophoresis results using rRNA intergenic spacer PCR products amplified from anaerobic quinoline soil microcosm samples. The band in purple is the upper marker and the band in green is the lower marker. Lane 0-6 means the sampling time points already described above. T0 and T0’ means the negative control of microcosm (original soil samples before anaerobic condition and quinoline addition respectively).](image-url)
From the gel of negative control (anaerobic 0mg), the band framed by red rectangle also appeared in the anaerobic condition. However, when quinoline was added into the microcosm, the band first faded and then strengthened at time point T2 (10, 25, 50mg). The band framed by green rectangle was stronger in quinoline added soil sample and faded gradually during incubation. Starting from the time point T3, new (bands framed by blue rectangle) bands arised in the quinoline added (especially with 2mg quinoline) samples. The band closed to the upper marker (around 1000bp) existing in quinoline added microcosm was stable while it disappeared in negative control samples.

Normalized location and intensity of each RISA band under aerobic and anaerobic condition were used in the PCA and BGA analysis. The BGA results (Fig.17) below shows the bacterial community movements between each groups during the microcosm experiments. Under anaerobic condition, plots cannot be separated clearly, while under aerobic condition, Group NQ0 (No quinoline added), group NQ2 (2mg quinoline added) and the rest groups were separated clearly. So, quinoline could be the factor influencing the microbial community under anaerobic condition. Under aerobic condition, we still can find out the tendency of the microbial community shifts through the BGA results. The aerobic and anaerobic condition seems to be the main factors that influence the microbial community changes rather than quinoline addition.

Fig.17: BGA analysis of aerobic and anaerobic microcosm. group Q0,Q2,Q10,Q25,Q50 under anaerobic condition; group NQ0,NQ2,NQ10,NQ25,NQ50 under aerobic condition. The number means dose of quinoline (mg) added, N stands for oxygen needed condition, Q stands for quinoline.
The PCA and BGA analysis results of each series are shown below (Fig18, 19): while in each series under aerobic condition, although the plot cannot be directly separated but the tendency is clear (plots in middle left move toward the lower right). When quinoline was added into the microcosm, the tendency is clearer, especially in aerobic 10mg, 25mg and 50mg conditions. Plots can be separated into two groups: one group containing the plots of T0 and T1, the other group containing the plots T2 to T7. The same results also appeared in the anaerobic condition. One group includes T00 and the other group includes T0 to T7. From BGA results, the important concentration of quinoline added (10mg) shapes the bacterial community distinctly under both aerobic and anaerobic conditions. The time point that starts the great changes is T2 under aerobic condition and T0 under anaerobic condition.
Fig. 18: PCA and BGA analysis of bacterial community structure changes during the soil incubation with 0, 10 and 50 mg quinoline under aerobic condition. t0-t7 means sampling time point.
Fig. 19: PCA and BGA analysis of community structure changes during the soil incubation with 0, 10 and 50 mg of quinoline under anaerobic condition. t0-t7 means sampling time point.
3.3 Results of GC/MS analysis

The figures below show results obtained from the GC/MS analysis. The pure quinoline in solvent was tested to confirm the retention time (Fig.20). From the GC/MS profile, a highly contaminated molecule (reagent dissolved in syringe by dichloromethane) emerged and was identified as butylated hydroxytoluene. Fig.21 shows the GC/MS profile and the composition of the aerobic T1 (10mg quinoline added) soil sample. Quinoline and butylated hydroxytoluene were predominant, while no related metabolites previously mentioned in the chapter “metabolic pathway of quinoline” were detected.

![Fig.20: GC/MS results using quinoline with a dilution of 80 ppm. The peak appearing between 13-13.5 minutes is corresponding to quinoline and the peak appearing around 20-21 minutes is considered as a contamination](image)

![Fig.21: GC/MS results using aerobic soil sample T0 (left). Fig.21 (Right) shows us the GC/MS result using aerobic soil samples T1. Quinoline is detected with a retention time of 13.12 minutes.](image)

Here, we used 10mg quinoline added aerobic samples to find out the quinoline
biodegradation after GC/MS analysis. According to the GC/MS profile results, the quinoline peak area detected in each sample was automatically calculated. Results are shown below in tables 3 and 4, Fig 22 and 23. The histogram with error bar and P-value was drawn and calculated by using Microsoft Excel software. P-value below 0.05 means there are significant differences between each group and the P-value was also used for quinoline degradation analysis in this study.

<table>
<thead>
<tr>
<th>Aerobic samples</th>
<th>Area</th>
<th>Area%</th>
<th>Start time(min)</th>
<th>End time(min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T0-1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>T0-2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>T0-3</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>T1-1</td>
<td>7135575</td>
<td>63.93</td>
<td>13.045</td>
<td>13.409</td>
</tr>
<tr>
<td>T1-2</td>
<td>7452564</td>
<td>59.63</td>
<td>13.045</td>
<td>13.401</td>
</tr>
<tr>
<td>T1-3</td>
<td>5046638</td>
<td>51.11</td>
<td>13.053</td>
<td>13.409</td>
</tr>
<tr>
<td>T2-1</td>
<td>3357255</td>
<td>41.11</td>
<td>13.038</td>
<td>13.431</td>
</tr>
<tr>
<td>T2-2</td>
<td>3717641</td>
<td>44.6</td>
<td>13.053</td>
<td>13.401</td>
</tr>
<tr>
<td>T2-3</td>
<td>3270807</td>
<td>45.87</td>
<td>13.053</td>
<td>13.379</td>
</tr>
<tr>
<td>T3-1</td>
<td>744218</td>
<td>14.23</td>
<td>13.068</td>
<td>13.349</td>
</tr>
<tr>
<td>T3-2</td>
<td>1356890</td>
<td>16.1</td>
<td>13.060</td>
<td>13.416</td>
</tr>
<tr>
<td>T3-3</td>
<td>772650</td>
<td>11.54</td>
<td>13.075</td>
<td>13.334</td>
</tr>
<tr>
<td>T4-1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>T4-2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>T4-3</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>T5-1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>T5-2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>T5-3</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 3: quinoline peak area values in aerobic 10mg quinoline added samples.

Fig.22: Quinoline degradation test: samples collected from aerobic condition microcosms. 10mg of quinoline was completely degraded after the time point T4 under aerobic condition.
### Table 4: quinoline peak area value in anaerobic 10mg quinoline added samples.

<table>
<thead>
<tr>
<th>Anaerobic samples</th>
<th>Area</th>
<th>Area%</th>
<th>Start time</th>
<th>End time</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1-1</td>
<td>3732024</td>
<td>47.82</td>
<td>13.045</td>
<td>13.460</td>
</tr>
<tr>
<td>T1-2</td>
<td>4313605</td>
<td>51.15</td>
<td>13.045</td>
<td>13.357</td>
</tr>
<tr>
<td>T1-3</td>
<td>6187714</td>
<td>70.56</td>
<td>13.038</td>
<td>13.357</td>
</tr>
<tr>
<td>T3-1</td>
<td>3702209</td>
<td>52.8</td>
<td>13.045</td>
<td>13.387</td>
</tr>
<tr>
<td>T3-2</td>
<td>3345472</td>
<td>49.9</td>
<td>13.053</td>
<td>13.401</td>
</tr>
<tr>
<td>T3-3</td>
<td>4293936</td>
<td>54.38</td>
<td>13.045</td>
<td>13.475</td>
</tr>
<tr>
<td>T5-1</td>
<td>4726012</td>
<td>58.87</td>
<td>13.053</td>
<td>13.446</td>
</tr>
<tr>
<td>T5-2</td>
<td>4940835</td>
<td>59.22</td>
<td>13.045</td>
<td>13.423</td>
</tr>
<tr>
<td>T5-3</td>
<td>5120902</td>
<td>53.68</td>
<td>13.053</td>
<td>13.386</td>
</tr>
<tr>
<td>T6-1</td>
<td>4135664</td>
<td>44.42</td>
<td>13.045</td>
<td>13.349</td>
</tr>
<tr>
<td>T6-2</td>
<td>3613608</td>
<td>42.47</td>
<td>13.053</td>
<td>13.379</td>
</tr>
<tr>
<td>T6-3</td>
<td>4122499</td>
<td>44.81</td>
<td>13.045</td>
<td>13.409</td>
</tr>
</tbody>
</table>

**Fig.23**: quinoline degradation test: samples collected from anaerobic condition microcosm. No significant (P-value: 0.2 > 0.05) quinoline degradation was detected.

### 4. Discussion and conclusion

Coal tar remains the principal source of commercial quinoline (Gomtsyan, Bayburt et al. 2005), its waste materials containing quinoline does great damages to the environment and to humans. In order to find the quinoline degrading bacteria, former experiments usually incubate quinoline contaminated soil samples in quinoline containing phosphate salts medium (Shukla 1986).

To date, research has focused on the disclosing and recovery of microorganism community function in environmental samples because of its huge unrevealed genetic resources (Trevors, van Elsas et al. 1990; Eller, Kruger et al. 2005). Laboratory microcosms are commonly used to assess
potential ecological effects which may result from the introduction of a particular organism in a particular habitat by mimicking key field parameters in a laboratory situation (Bonkowski and Roy 2005). After a bio-adaptation strategy using contaminated soil, microcosms also can be applied to screen potential biodegradation microbial communities (Kools, Ferwerda et al. 2005).

The soil microorganism primarily consists of bacteria, actinomycetes, fungi, algae and protozoa. Currently, 16S rRNA gene sequence analysis and lots of technologies based on PCR such as random amplified polymorphic DNA (RAPD) (Mehta, Leite et al. 2001; Ponnumamy, Jose et al. 2011), amplified restriction fragment polymorphism (AFLP) (Duran, Gryzenhout et al. 2010; Ogunjobi, Fagade et al. 2010), restriction fragment length polymorphism (RFLP) (Tanti, Ray et al. 2011), terminal restriction fragment length polymorphism (T-RFLP) (Nithya and Pandian 2012), single-strand conformation polymorphism (SSCP) (Sen, Hamelin et al. 2008; Larentis and Alfreider 2011), denaturing gradient gel electrophoresis (DGGE) (Gillan, Speksnijder et al. 1998; Nicolaisen and Ramsing 2002) and temperature gradient gel electrophoresis (TGGE) (Gomez-Villalba, Calvo et al. 2006) have been widely applied to the study of microbial diversity.

The intergenic spacer region between the small (16S) and large (23S) rRNA genes were considered to be under less evolutionary selective pressure than the 16S and 23S rRNA coding genes and therefore provided an excellent tool for a finer identification at the species/strain level (Mora, Ricci et al. 2003; Panangala, van Santen et al. 2005). The rRNA intergenic spacer analysis (RISA) was used to study the genetic structure of the bacterial community pools from environment samples (Ranjard, Poly et al. 2000; Buchan, Alber et al. 2001). RISA is easy to perform, allows rapid examination at the composition of complex bacterial communities, and can also be performed without the use of specific and expensive equipment (Borneman and Triplett 1997; Acinas, Anton et al. 1999). Different fingerprints represent different genetic structures giving evidence that some populations are specific to a given location in addition to the common populations of all the microenvironments.

As described in the introduction, quinoline can be degraded under both aerobic and anaerobic condition. In this study, we used soil microcosm experiments which would give the opportunity to reveal the diversity of quinoline degrading bacteria and to investigate the presence of bcr operon containing bacteria. Meanwhile, if the bcr operon is enriched during soil
microcosm experiment, we can use the microcosm metagenomic DNA as substrate to test the capacity of the Genefish tool in capturing low percentage genes.

We used quinoline enriched microcosms under aerobic and anaerobic conditions to investigate the quinoline degrading bacterial diversity. After the RISA analysis, the community structure of quinoline microcosms changed after the quinoline was added. From the PCA and BGA results, addition of 2mg quinoline into the soil microcosm generates slight changes in bacterial community structure under anaerobic conditions while obvious changes under aerobic condition. 10mg, 25mg and 50mg of quinoline added in the soil sample shapes the community structure under aerobic condition. Nucleotide BLAST results showed that bcr type genes were related to *Rhodopseudomonas.palustris*, *Magnetospirillum.sp* and *Thauera.sp* in Rothamsted soil samples. *Rhodopseudomonas.palustris* is a purple photosynthetic bacterium widely distributed in nature such as swine waste lagoons, earthworm droppings, marine coastal sediments and pond water (Harwood 1992). It grows with or without oxygen and uses many alternative forms of inorganic electron donors, carbon and nitrogen. It degrades plant biomass and chlorinated pollutants and it generates hydrogen as a product of nitrogen fixation (Smith and Wilcox 1992; Roberts 2005). It is also reported to degrade p-coumarate under anaerobic condition (Yang, Horton et al. 2005), when growing photosynthetically in the absence of oxygen, *R. palustris* converts aromatic compounds to Ac-CoA, which is then used as carbon and energy source (Villa-Komaroff, Efstratiadis et al. 1978; Roberts and Burgess 2005). Pelletier and Harwood (Pelletier and Harwood 2000) have compared differences in the pathway of anaerobic benzoate degradation between *R. palustris* and *T. aromatica*. These two organisms have similar benzoyl-CoA reductase enzymes (BadDEFG in *R. palustris* (Egland and Harwood 1999)), but once benzoyl-CoA is reduced to a cyclohexadienecarboxyl-CoA intermediate, the next step in *T. aromatica* is an hydration of the ring, whereas the next probable step in *R. palustris* is a two-electron reduction to give cyclohex-1-ene-1-carboxyl-CoA as an intermediate (Song and Ward 2005). *Magnetospirillum* strains are microaerophiles thriving at the aerobic/anaerobic interface in natural ecosystems (Saari, Torma et al. 2011). Shinoda et al. (Shinoda, Akagi et al. 2005) isolated a phenol-degrading *Magnetospirillum* strain which represents a third denitrifying bacterial group capable of anaerobic utilization of aromatic compounds. In *Magnetospirillum* sp. strain TS-6, transcription of the bcr operons was observed in cells grown in aerobic condition on
benzoate, even though the Bcr is active only under anaerobic conditions (Kawaguchi, Shinoda et al. 2006).

Gas chromatography–mass spectrometry (GC/MS) is a method that combines the features of gas-liquid chromatography and mass spectrometry. It has been widely used in the separation and identification of complex components (Wittmann 2007), which have the high resolution of the GC and high sensitivity of the MS. It is a highly effective tool for the qualitative and quantitative analysis of drugs and metabolites in biological samples (Tuo, Yan et al. 2012). GC/MS analysis suggested the quinoline degradation in aerobic soil microcosms (10mg quinoline added) while no evident quinoline degradation functions were detected in anaerobic soil microcosms (10mg quinoline added). 10mg of quinoline were almost degraded after 28 days (T4) under aerobic condition. However, for the GC/MS results, we still need the negative control (sterilized soil sample with 10mg quinoline). Although there were slightly bacterial community structure changes (in RISA) under anaerobic adaptation but no quinoline was degraded after the GC/MS analysis. One hypothesis is: the anaerobic condition enriched the anaerobic bacteria and they grew using the nutrition (organic compounds, inorganic salts etc.) in the soil, finally leads to the community structure changes. Benzoyl-coenzyme A reductase was reported as an ATP dependence key enzyme of anaerobic aromatic metabolism (Boll and Fuchs 1995). So, for the quinoline degradation under anaerobic condition, bacteria usually need the energy captured from anaerobic respiration such as denitrification (Zumft 1997). The reason why no quinoline degradation was detected under anaerobic condition might be the lacking of nitrates in these soil microcosms. Adding extra nitrogen sources such as ammonium sulfate, potassium nitrate in soil microcosms may stimulate the growth of anaerobic bacteria and lead to quinoline degradation.

After the quinoline soil microcosm experiment in this study, it is confirmed that the soil microcosm strategy was a good method to analyze microbial diversity involved in contaminants biodegradation. Although quinoline degradation was detected under aerobic condition with the Rothamsted soil, but the principal bacterial responding community were still not yet confirmed. Further works should focus on 16S rRNA amplification, identification or even pyrosequencing of quinoline soil microcosm metagenome: this would help us to reveal (1) the relationship between quinoline degradation and relative microbial community (specific band shifts in the profile before and after quinoline addition into microcosms), (2) reveal the bacteria involved in quinoline
degradation and perhaps new bacteria not yet discovered through quinoline enrichment even in selective medium.
PART II: *in vitro* development and use of Genefish to capture targeted DNA fragments

1. Background

The “gene” was first considered as a crucial factor that cause an inheritable phenotype characteristic in organisms (Noble 2008). Since the development of molecular biology, gene is defined as a region of genomic sequence, corresponding to a unit of inheritance, which is composed of regulatory regions, transcribed regions, and or other functional sequence regions (Pearson 2006; Pennisi 2007). The genotype of an organism is the conclusive force to determine the phenotype of the organism. Because of the massive genotype diversity, our living world presents a colorful bio-ecology. Take microorganisms for example, they exist almost everywhere and some of them even can survive under extreme conditions such as hypersaline (Ollivier, Caumette et al. 1994), high temperature (Ranieri and Boor 2009), low oxygen (Mesbah and Wiegel 2008), alkaline (Grant and Heaphy 2010), high pressure (Kato and Qureshi 1999), polluted sites such as mineral water (Uroz, Calvaruso et al. 2009), etc. So, the specific genotype of those microorganisms raises high interest of scientists. However, relating genotypes to phenotypes of microorganisms is problematic because the interactions between genes, proteins and their physiological functions are extremely complex. Speculation indicated that only about 1% of microorganisms in nature can be cultured (Smith 1996) and the rest are still mysterious and needed to be revealed. As we mentioned in the general introduction, to date, metagenomic approaches afford to reveal the genetic diversity, bacterial diversity, and even some functional molecules responding to specific phenotypes. However, the defect of metagenomic approaches is also obvious, targeted genes or fragments were randomly captured which will dramatically decrease the capture efficiency. In order to capture target genes and fragments more efficiently and precisely from environmental metagenomes, a new tool “Genefish” was constructed in our group as a complementary method for metagenomics. The “Genefish” concept is based on the genetic recombination process catalyzed by the *λ* Red system and cloning of specific genes from a metagenome. Genefish is included in the European project Metaexplore which aims at
discovers and characterizes new enzymes from environmental microbial community. Interested enzymes encoding by new genes in this project are implicated in the biodegradation of recalcitrant (chitin, lignin) and xenobiotic halogenated compounds. Thanks to the effort of Nathalie Lombard, former PhD who built the capture-plasmid and counter-selection system, my assignment in this project was to continue the initial work achieved by Nathalie Lombard, Samuel Jacquiod and Laure Franqueville, to develop and to use the Genefish tool to capture relevant genes implicated in quinoline degradation.

2. Bibliographical review of Lambda-Red homologous recombination and Genefish approach

2.1 Lambda-Red homologous recombination

2.1.1 Introduction

Restriction enzymes and DNA ligases have been used as classical sub-cloning tools for more than 30 years (Arber and Linn 1969; Smith and Wilcox 1970; Roberts 1976; Kessler and Manta 1990). This method enables to efficiently get small DNA fragments in vitro (Kelly and Smith 1970; Villa-Komaroff, Efstratiadis et al. 1992; Roberts 2005). However, some genetic modification including gene deletion, gene insertion and large fragment cloning are difficult and even impossible to succeed by using this classical tool. Recently, a new in vitro technology called recombineering (recombination-mediated genetic engineering) (Ellis, Yu et al. 2001) has emerged, and by using this technology, genetic engineering could greatly be improved, and simplified for plasmid and strain construction. (Bian, Huang et al. 2012; Cottingham 2012; Marinelli, Hatfull et al. 2012). Recombineering is based on homologous recombination that can occur between two short homologous sequences of recipient and target fragments. To date, this strategy is commonly applied in genetic modification such as gene point mutation, deletion, insertion without the use of restriction enzymes and DNA ligases (Bird, Erler et al. 2012; Chai, Shan et al. 2012; Dai, Kim et al. 2012). Actually, it is catalyzed by the bacteriophage-encoded homologous recombination functions, such as the coliphage λ Red system (Yu, Ellis et al. 2000; Song, Dong et al. 2010;
Tuntufye 2011) and the RecET system from the Rac prophage (Cobb and Zhao; Zhang, Buchholz et al. 1998; Datta, Costantino et al. 2008; Swingle 2010; Cobb and Zhao 2012). Here, we are focusing on the λ Red system which is also applied in our study Genefish. The process of λ Red homologous recombination includes four steps: (1) Preparation of the host cells which contain the λ Red system (into their chromosome or on a plasmid). (2) Preparation of targeted exogenous linear DNA containing homologous sequences for recombineering. (3) Induction of λ Red system and electroporation of the target DNA into the host cells. (4) Selection or screening for the recombinants containing requested genetic changes. During the homologous recombination, nucleotide sequences are exchanged between two similar or identical DNA molecules resulting in new combinations of genetic material.

2.1.2. Lambda-Red recombination system

2.1.2.1 Overview of Lambda-Red recombination system

The Lambda-Red homologous recombination system primarily consists of three genes: gam, exo and bet. In general, linear DNA transformed into E. coli will be degraded by the powerful RecBCD nuclease (an exonuclease from E.coli that degrades ssDNA and dsDNA) (Dabert, Ehrlich et al. 1992; Boyd, Weiss et al. 2000; Singleton, Dillingham et al. 2004). However, the λ Gam protein encoded by the gene gam can inhibits the nuclease activities of RecBCD, preserving linear DNA and thereby allowing it to be taken as substrate for recombination (Karu, Sakaki et al. 1975; Murphy 1991). The λ Exo protein encoded by the gene exo is only required for dsDNA recombineering. It has a 5’ to 3’ exonuclease activity, Exo binds and remains bound to 5’ ends of dsDNA and start the degradation in 5’ to 3’ direction (Carter and Radding 1971; Cassuto and Radding 1971). This process forms dsDNA with a 3’ exposed dsDNA overhang where the Beta protein could further bind. The λ Beta protein encoded by the gene bet is a ssDNA binding protein that promotes annealing of two complementary DNA molecules. Beta binds to the 3’ overhang of the exposed dsDNA and protects the DNA from single-strand nuclease attack (Muniyappa and Radding 1986; Karakousis, Ye et al. 1998; Desai and Shankar 2003). Recombineering with linear dsDNA require all these three proteins but Exo is not necessary for
ssDNA recombination (Yu, Ellis et al. 2000; Ellis, Yu et al. 2001). Fig.24 shows the overview of λ Red recombination system for dsDNA recombineering.

The Gam protein inhibits the activity of the E.coli’s RecBCD nuclease thus allow the maintenance of the exogenous DNA after its transformation into the strain. Exo degrades dsDNA in 5’ to 3’ direction, leaving a 3’ overhang ds DNA. Beta binds to the single-stranded regions produced by Exo and facilitates recombination by promoting annealing to the homologous genomic target site(Sawitzke, Thomason et al. 2007). Fragment in blue stands for highly conserved sequences; fragment in yellow stands for exogenous target sequences; fragments in green and black stand for recipient sequences in the host strain.

Fig.24: Process of λ Red recombination system used for recombineering.
Although the lambda Red system was commonly used for genetic engineering (Brack, Eberle et al. 1976; Kopecka, Hillova et al. 1976; Pirrotta 1976; Martinsohn, Radman et al. 2008), the detailed mechanism by which lambda Red mediates double-stranded DNA recombination remains uncertain. Currently, three different Red-mediated dsDNA recombination mechanisms were proposed by Court et al (Court DL 2002), Poteete (Poteete 2008) and Mosberg et al (Mosberg, Lajoie et al. 2010) respectively.

The strand-annealing model for lambda Red-mediated dsDNA recombination proposed by Court et al. occurs during DNA replication (Fig. 25). The 3’ exposed end of dsDNA overhang anneals to its homologous target at the replication fork. This replication fork then backtracks and the former leading strand switches on to the degraded 5’ end dsDNA. The heterologous dsDNA blocks further replication from this fork. The other replication fork processing in the opposite direction around the circular bacterial chromosome will finish the rest of the replication task until the other 3’ overhang of dsDNA anneals to its complementary region and thus forming a crossover structure. This crossover structure then is resolved by unspecified \textit{E. coli} enzymes (Court DL 2002).

\begin{center}
\includegraphics[width=0.8\textwidth]{fig25.png}
\end{center}

\textit{Fig.25: Mechanism of strand-annealing model for lambda Red-mediated dsDNA recombination during DNA replication described by Court et al. Heterologous dsDNA is shown in yellow and homologous sequences are represented in blue at each end. (a) Beta promotes the annealing between the 3’ overhang and the complementary sequence in the lagging strand. (b) The annealing reaction blocks the replication proceeding. (c) While the replication in opposite direction still progresses until the replication fork arrives at the other homologous recognition sites. This crossover structure is resolved by unspecified \textit{E. coli} enzymes (Court DL 2002).}
Poteete et al (Poteete 2008) assume that the dsDNA recombination also occurs between target DNA (a lambda phage DNA they used) and a plasmid which does not have a second replication fork (plasmid replicate in one direction). In contrast to the Court mechanism, the two opposite direction replication forks are not formed in the mechanism proposed by Poteete. So, an alternate replisome invasion model was established by Poteete (Fig.26). The Exo 5’ to 3’ exonuclease activity degrades each end of the exogenous dsDNA homologous region into ssDNA overhangs. One of the 3’ overhangs first anneals to its complementary sequence on the replication fork. Distinct from the mechanism put forward by Court, the leading strand switches to 3’overhang which already annealed with its complementary sequence and serves the 3’overhang as the new template for synthesis (Poteete 2008). So the replication will continue through the exogenous dsDNA, and release the original lagging strand template. The author indicated that a second strand-switching event might occur at the other end of the incoming dsDNA to complete the recombination. However, in this mechanism, no more information were given on the endonuclease which is responsible for resolving the structure after the first template switching event and the formation of the new replication fork.

Fig.26: Replisome invasion model for lambda Red-mediated dsDNA recombination according to Poteete. Heterologous dsDNA of the plasmid is shown in yellow and homologous sequences are in blue at each ends. (a) Beta promotes the annealing between 3’ overhang and its complementary sequence in the plasmid. (b) The leading strand switches to the 3’ overhang of the Exo processed dsDNA and consider it as the new template for new synthesized strand. (c) The new replication fork travels through the heterologous dsDNA to continue the replication. (d) The replication fork arrives at the other homologous sequence site and the leading strand switches again on-to the original homologous sequence of the plasmid.
Mosberg et al (Mosberg, Lajoie et al. 2010) proposed a lambda Red mediated dsDNA recombination progression via a ssDNA intermediate. According to their theory, dsDNA molecules transformed into *E.coli* are completely dissociated to ssDNA by Exo. This dissociation starts at one end of the dsDNA and exposes two homologous regions. When the replication fork arrive at the homologous sequences, this ssDNA will anneals to its complementary targets and is inappropriately taken as part of the newly synthesized strand (Fig.27). No more information about the resolution of the mismatch structure between lagging strand and non-homologous sequences was mentioned by the authors.

![Fig.27: Mechanism of ssDNA intermediate model for lambda Red-mediated dsDNA recombination outlined by Mosberg et al (Mosberg, Lajoie et al. 2010)]. Heterologous dsDNA is represented in yellow and homologous sequences are represented in blue at each ends. (a) Exo completely dissociates the exogenous dsDNA and generates two ssDNA. (b) One ssDNA anneals to its complementary sequence at each end of the homologous sites. This ssDNA is inappropriately used as part of the newly synthesized strand.

### 2.1.2.2 Different vectors for the Lambda-Red system

For λ Red recombineering strategy, in order to obtain the highest recombineering efficiency, an appropriate expressed level of the λ Red recombination proteins is required. Limited expression of Gam protein couldn’t inhibit the RecBCD nuclease activities completely (Sergueev, Yu et al. 2001), a high expression of Gam could also cause plasmid instability (Silberstein Z 1990; Murphy 1991) and even cell death (Kirill Sergueev 2001). The λ Red enzymes could be expressed under the control of their own promoter or under heterologous regulated promoters. While endogenous phage promoter leads to tight regulation and coordinates expression thus
leading to a higher recombination frequency (Court DL 2002). Court et al (Yu, Ellis et al. 2000; Ellis, Yu et al. 2001; Court DL 2002) have developed and utilized a λ prophage Red system on the E.coli chromosome for recombineering application. In this system, they do some genetic modification of λ prophage including the deletion of lysis, replication and structural genes thus inactivate the λ prophage infectious cycle. However, the critical functions such as transcriptional control of the Red are retained (Fig.28).

![Diagram](image)

**Fig.28:** Schematic representation of λ phage constructs into E.coli chromosome used for recombineering (Sharan, Thomason et al. 2009)

Bacterial DNA is shown in blue bar, phage DNA in red bar and the region containing a deletion and/or a substitution is a white bar: (a) Genetic structure of λ prophage which is integrated into the bacterial chromosome: the complete λ prophage is flanked by its attachments site att, where integration with the bacterial chromosome occurred and is adjacent to the bacterial biotin genes, bioAB. The int and xis genes located in the pL operon and encode function products which can help phage DNA to integrate into or excise phage DNA out of the bacterial chromosome. There are two operons as shown with arrows and their promoters pL and pR in the prophage. The exo, bet, and gam genes are under the control of the temperature-sensitive repressor CI857. At temperatures less than 32°C (Yu, Ellis et al. 2000), CI repressor binds to the operators and prevents transcription of the pL operon and thereby the expression of Red. The expression of the Red system can be turned on easily at 42°C inducing the repressor CI857 denaturation and as a consequence, allows the Red genes transcription. When shifting back to low temperature, CI857 re-natures and recovers its initial
functions thus completely blocks the transcription of pL. When λ Red functions are turned on for a short time period, like 5 min, cells become more recombinogenic after linear DNA uptake (Yu, Ellis et al. 2000). The N gene encodes an anti-transcription terminator and prevents RNA polymerase termination of pL transcripts at terminators tL1, tL2, and tL3. The kil gene is adjacent to gam and when expressed for over 1 hour kills the host bacterial cell (Sergueev, Yu et al. 2001). The replication genes O and P of the λ prophage are in the pR operon. Cro functions as a partial repressor of the pL and pR operons when CI is inactive at 42°C. The lysis and structural genes, SRA-J, are shown located beyond their regulator Q (Sharan, Thomason et al. 2009).

(b) The defective prophages used for recombineering in DY380, SW102, and the HME strains (Lee, Yu et al. 2001; Warming S 2005): cro through the bioA prophage genes are deleted. The right att site is also deleted preventing any excision of the prophage. In DY380 and SW102 strains, those deleted genes are replaced by the tetracycline resistance cassette, tetRA.

(c) The mini-λ phage DNA (Court, Swaminathan et al. 2003) is shown with the lytic genes cro through J deletion. In different constructs of the mini λ, the cro-J region is replaced with various drug resistant cassettes. As we mentioned before in the genetic structure of λ prophage, mini- λ has kept att sites plus int and xis, allowing its integration and excision.

Recently, for the convenient use of λ Red systems, its critical elements were integrated into a set of different plasmids. For instance, Murphy et al (Murphy 1998; Murphy, Campellone et al. 2000) designed a plasmid possessing phage recombination functions under the control of an IPTG-inducible lac promoter. Later on, Datsenko (Datsenko and Wanner 2000) and Wang (Wang, Sarov et al. 2006) developed recombineering plasmids under the arabinose inducible pBAD promoter (Chao, Chiang et al. 2002; Terpe 2006). In this construction, the AraC protein is both a positive and a negative regulator (Yajima, Muto et al. 1993; Cobb and Zhao 2012). In the presence of arabinose, transcription from the pBAD promoter is turned on; in its absence, transcription occurs at very low levels (Lee, Francklyn et al. 1987). Other studies (Dodd, Shearwin et al. 2005; Donald L. Court 2006) show that the repression system of the λ Red machinery is strong exclusively under the controls of pL and pR promoters. So the ideal conditions is to ensure the presence of pL and pR promoters on all of the prophage constructs, whether carried on the bacterial chromosome, on low copy plasmids or under the control of a temperature sensitive gene such as cI857.
The pSIM vectors (Datta, Costantino et al. 2006) consist of the basic elements for its replication and other necessary elements for prophage induced homologous recombination. Fig.28 (d) shows the λ segment of the pSIM plasmids. The genes from cro to att and genes beyond tL3 including int and xis were removed. The red genes were directly connected to pL by a deletion that removes kil through N genes. The drug resistance marker specific of each pSIM plasmid replaces the rex gene (inhibit cell function without phage superinfection (L snyder 1989) ) adjacent to cI857. The basic features are conserved in all of these Red expression pSIM plasmids. This λ Red system was left under the control of native phage elements for optimal expression and regulation. The pL promoter drives gene expression and is controlled by the temperature sensitive but reversible λ CI857 repressor. In all, the cro gene is inactive to maximize pL expression, and the replication genes are absent or inactive to prevent lethal effects on the cell.

2.1.2.3 Substrates for the Lambda-Red system

Based on the strategy of λ Red system, different DNA substrates could be used. Exogenous double or single-strand linear DNA both need to contain two regions homologous to their target sequences in any case. The inserted region between two highly conserved sequences could either be a selective marker or a non-selective DNA (Roberts 2005). Yu et al. (Yu, Ellis et al. 2000) demonstrated that 30- to 50-bp homologies are able to result in recombination in vitro. For those exogenous DNA substrates constructions, we only need to add homology sequences (30- to 50-bp) at each 5’ end of PCR primers. These primers can amplify the exogenous DNA with highly homology conserved sequences easily at each end of PCR products and this PCR dependant DNA substrates preparation could substantially facilitates the λ Red system application because we can modify the homology sequences as we want.
2.1.2.3.1 Double-strand DNA recombination

ds DNA substrate can be generated by polymerase chain reaction (PCR) with a pair of primers containing homologous sequences to the target site (Arber and Linn 1969). Each primer, about 70 bases in length, contains 50 nucleotides at the 5’ end corresponding to the homologous targeted region and around 20 nucleotides at its 3’ end to amplify the exchanged sequences (Fig.29).

Fig.29: ds DNA substrates for recombination generated by PCR: homologous region (in blue bar) and non-homologous sequences which usually contain selective marker (usually antibiotic resistance gene).
When the ds DNA including selection marker is used (antibiotic resistance genes) for recombination, positive recombinants can easily be screened through the specific phenotype. However, not all exogenous ds DNA contain selective marker, for recombinants selection in this case, two steps counter-selection method or \( loxP, FRT \) (Flp recognition target) dependant method are helpful (Fig.30) (Tischer, Smith et al. 2010). In these two-steps counter-selection processes, two homologous recombination phases were performed consecutively. Two different types of ds DNA substrate containing identical highly conserved sequences are essential. The first recombination event occurs with ds DNA possessing a selection marker and positive recombinants can directly be screened. Then the selected recombinants are taken as objects to be replaced in the second round recombination using initial targeted but non selective marker ds DNA. The phenotype changes before and after recombination indicates the gene replacement in the targeted region.

Fig.30 A two-step recombination process by using non-selective marker ds DNA. Green bar are target sequences to be replaced by non-selective marker ds DNA (yellow bar). In the first step, a selective marker containing fragment was inserted (red bar). In the second step, the target sequence of recombinants that was already selected after the first step recombination was replaced by the expected fragment (yellow bar). Positive recombinants can be screened by the counter-selection of the selective marker.
In another method (Datsenko and Wanner 2000), a selective marker flanked by two $\text{loxP}$ or $\text{FRT}$ sites at each end is linked to a non-selective DNA fragment which contains two homologous sequences corresponding to the target DNA site. The $\text{loxP}$ and $\text{FRT}$ sites are regions of 34 base pairs (bp) in length and consist of two 13 bp inverted repeats separated by an 8 bp spacer region (Hoess, Wierzbicki et al. 1986; Datsenko and Wanner 2000). The FLP recombinase binds to the palindromic sequences, whereas the 8 bp spacer corresponds to the site of DNA break, exchange and ligation (Lacroix, Giovannini et al. 2011). Target DNA deletion produced by Cre-lox or Flp-FRT recombination is determined by the orientation and location of flanking $\text{loxP}$ or $\text{FRT}$ sites. When the two $\text{loxP}$ or $\text{FRT}$ sites are in the same orientation, Cre or Flp recombinase protein expressed by the host strain can induce a site-specific recombination thus deleting the DNA sequences between two $\text{loxP}$ or $\text{FRT}$ sites respectively (Fig. 31). However, after the deletion of the selective marker, a single $\text{loxP}$ or $\text{FRT}$ sites will remain.

![Fig.31: a site-specific dependent recombination system applied to modify targeted sequence by using the loxP or FRT sites. Green bars are the target sequences to be replaced by the non-selective marker ds DNA (yellow bar). The short bar in pink can be loxP or FRT sites which participate in the Cre-Lox or Flp-FRT recombination respectively.](image)

2.1.2.3.2 Single-strand DNA recombination

Recombineering can also be performed using synthetic oligonucleotides or short denatured PCR products (Swaminathan S 2001; Costantino and Court 2003; Yang and Sharan 2003; Maresca, Erler et al. 2010). Single-strand DNA can be used to create point mutation through homologous recombination and this method is highly efficient and do not require any selection
step. In contrast, an optimal condition that primarily contains two variables can stimulate the process of recombination when using single-strand DNA. The first variable is the use of a “lagging-strand” oligo nucleotide, a single strand corresponding in sequence to the DNA chain that is replicated discontinuously (Kim, Dallmann et al. 1996; Ellis, Yu et al. 2001; Costantino and Court 2003). Ellis (Ellis, Yu et al. 2001) found a strand bias in recombination levels when using two complementary ss DNAs and a strand bias of 4 to 10 fold was seen in each of four independent experiments using lagging strand. The other important variable is to avoid the *E. coli* methyl-directed mismatch repair (MMR) system: preventing MMR increases recombination frequency about 100 fold (Li, Costantino et al. 2003). Yang and Sharan (Yang and Sharan 2003) developed a more generally useful “hit and fix” two-step recombineering approach which can be used to generate and find subtle changes in BAC DNA and, in this approach, recombinants can be screened by colony hybridization method (Fig.32). A set of long primers containing 80 bases of homologous and 20 bases of target site are used to generate the DNA substrates. This strategy consists of two steps called “hit” and “fix” respectively. “Hit” targeting DNA is first inserted into the wild type BAC. Recombinants can be identified by specific “hit” hybridization probe. Positive recombinants were used as the target to perform the “fix” step. “Fix” targeting DNA including a point mutation is recombined into the target sites thus generating recombinants with the expected point mutation.
Fig.32: Two-step “hit & fix” method to create point mutations using single stranded short PCR product or oligonucleotides as targeting vector (Sharan, Thomason et al. 2009).

(a): 180 bases of PCR products containing 80 bases of homologous at each end and 20 unique bases in between. The 20 unique bases are used as the target to design probes for colony hybridization. The PCR products can be denatured to get single-stranded oligonucleotides used as targeting substrates.

(b): Process of point mutation (G to A) by using “hit & fix” method. Step1: The “Hit” PCR amplified targeting DNA is used to replace 20 nucleotides (green bar) with 20 heterologous nucleotides (yellow bar). In step 2, the 20 nucleotides (yellow bar) are replaced by the “Fix” targeting DNA (green bar) except for the expected mutation. Recombinants can be identified by colony hybridization or PCR by using primers specific to the recombinant DNA.
2.2. The “Genefish” tool

2.2.1 General presentation of “Genefish”

The Genefish tool that we describe here can be considered as a complementary method for metagenomics. This approach aims to directly recover specific genes or DNA fragments from a metagenome. The “Genefish” project is based on the genetic recombination process catalyzed by the λ Red system as we described above and cloning of specific genes from the metagenome in situ. In order to achieve this objective, genetic engineered system has been constructed expanding the application of λ Red system. This new system called “Genefish” is composed of three molecular units. The first part consists in the capture plasmid which contains two short highly conserved DNA fragments cloned from the gene of interest (the specific gene we want to clone from a metagenome). This capture plasmid also encompasses a two suicide gene relF and colE3 cassette allowing positive selection of the intended clones. The mechanism of this positive selection will be introduced in a latter chapter. The second part is made of the λ Red system as we described before, and in our project, the λ Red system is integrated in a plasmid pSIM6. The system is strictly controlled by the temperature sensitive promoter cI857. The third part corresponds to the host strain which was improved after genetic modification and is taken as recipient for Genefish process (Fig. 33).
2.2.2 The suicide cassette

The capture plasmid, called pBAD35K7toxN, contains a suicide cassette composed of two toxic genes. After their induction, genes relF and colE3 express lethal proteins in the E.coli cell and lead to cell death (Fig. 34).

**Fig. 34:** The plasmid pBAD35K7 tox: red arrow indicates relF gene, and blue arrow represents the colE3 gene; the tick marks represent the 2 multi-cloning sites devoted to the insertion of the two short target sequences; cmR is Chloramphenicol resistance gene(Source: Lombard. N. Thesis 2007)

relF gene encodes a polypeptide of 51 amino acids which have 40% homology to the hok and gef gene products (Gerdes, Bech et al. 1986; Poulsen, Larsen et al. 1989). Induction of relF gene leads to the collapse of the cell membrane potential, arrest of respiration, changes in morphology and cause cell death (Gerdes, Bech et al. 1986; Gerdes, Poulsen et al. 1990). In our plasmid, relF is under the control of a synthetic lac promoter P_{A1-03/04}(Jensen, Ramos et al. 1993). lacIq gene in the chromosome of our host strain is a mutation of lacI gene. The specific mutation on the promoter of lacI gene results in higher level expression of lac repressor (de Boer and Glickman 1998). The expressed repressor protein LacIq inhibits the promoter P_{A1-03/04} thus shutting down the expression of the relF gene. However, LacIq can be blocked by the presence of IPTG at the final concentration of 500μg/ml (Knudsen, Saadbye et al. 1995), this will release the promoter P_{A1-03/04} allowing the expression of the relF gene.

ColE3 gene encodes a RNase colicin E3 that specifically induces the cleavage of a single phosphodiester bond in the E. coli 16S ribosomal RNA (Zarivach, Ben-Zeev et al. 2002).This cleavage results in a complete inactivation of the ribosome and inhibition of the protein synthesis thus finally leading to cell death. The expression of colE3 is under the control of P_{bad} promoter
which is regulated by the protein AraC, encoded by the *araC* gene. The AraC protein demonstrates both negative and positive regulations to *colE3* gene and this mechanism also called “the light switch mechanism” because the arms switch the system on and off (Harmer, Wu et al. 2001). In the absence of L (+) arabinose, the dimeric AraC protein binds to two sites (one is *araO2* site and the other is *araI1* site) near the promoter P_{bad} and forms a loop that prevents RNA polymerase binding to the *ara* operon promoter, and thereby blocks transcription (Guzman, Belin et al. 1995; Sa-Nogueira and Mota 1997). When arabinose is present, it breaks the *araO2-araI1* loop and leads to occupancy at *araI1-araI2* site. This process finally activates the P_{bad} promoter and the ColE3 expression starts leading to the cell death (Lobell and Schleif 1990). Further researches showed that the P_{bad} promoter is highly induced by L (+) arabinose at a concentration of 1% (mass/volume) in the culture media (Qiu 2008). However, in the absence of L(+)arabinose, the *colE3* gene still has a basal expression (Lee 1980; Lee, Francklyn et al. 1987). This low concentration of colicin still can kill a majority of cells and thus makes the system inefficient. So, in order to counteract the remaining colicine, genetic modification of the host strain was done. The gene *immE3* encoding the protein ImmE3 has been inserted into the host strain chromosome. This immunity E3 protein binds to colicin E3 and inhibits its RNase activity (Masaki and Ohta 1982; Lau, Rowsome et al. 1984; Yajima, Muto et al. 1993). ImmE3 expression is under the control of a constitutive promoter P_{bla}. In normal living cells, ImmE3 binds to the colicin molecules thus enables the cells growing. However, when *colE3* gene is induced by 1% (mass/volume) L (+) arabinose, a huge quantity of colicine is expressed. The quantity of ImmE3 is not sufficient to deactivate the toxic activity of colicin thus leading to cell death. The figure 35 and figure 36 show how these two toxic genes are processing in the host cell. These constructions were accomplished by former PhD student Nathalie Lombard (thesis 2007).
Fig. 35: Regulation of the toxic cassette on pBAD35K7tox plasmid before its induction. AraC protein (orange) binds the $P_{bad}$ promoter of the colE3 gene (dark blue) thus decreasing the amount of ColE3. Some of the ColE3 proteins that are expressed will be suppressed by ImmE3 (yellow). LacIq gene product (green) inhibits the $P_{lac}$ promoter of relF gene.

Fig. 36: Regulation of the toxic cassette on pBAD35K7tox plasmid after its induction. IPTG (black) inhibits LacIq thus releases the $P_{lac}$ promoter of the relF gene leading to the cell death. L (+) arabinose (purple) at the concentration of 1% (mass/volume) greatly induces the colE3 expression and also causes the cell death.
2.2.3 Homologous recombination with Lambda-Red system

Court et al (Datta, Costantino et al. 2006) created a low copy number, Red expression vector (pSIM6) with the DNA replication origin of pSC101. This plasmid contains an ampicillin resistance drug marker and occurs in 16 copies per cell (Hashimoto-Gotoh, Franklin et al. 1981) (Fig.37). Genes *gam*, *beta* and *exo* encode proteins Gamma, Beta and Exo respectively. These three genes are under the control of the temperature sensitive cI857 repressor (Yu, Ellis et al. 2000). When temperature is increased up to 42°C, expression of these three genes starts and enhances the occurrence of homologous recombination. The temperature sensitive replicons of pSIM6 permits curing of plasmids by growing the host strains at 37°C once the expected recombination is confirmed (Datta, Costantino et al. 2006; Sharan, Thomason et al. 2009). It is preferable to use 37°C as the P<sub>L</sub> operon remains mostly repressed by CI857 at this temperature, blocking unwanted Red expression into the host strain (Datta, Costantino et al. 2006).

![pSIM6 plasmid](image)

**Fig.37:** pSIM6 plasmid: three genes *gam*, *bet* and *exo* are under the control of the repressor cI857. When the temperature is increased up to 42°C, translation of these three genes will start and enhances the occurrence of homologous recombination.

*Source: Datta, Court, Costantino (Datta, Costantino et al. 2006) and N. Lombard.*

2.2.4 *bcr* operon

More and more man-made organic compounds and the associated aromatic molecules wastes cause great pressures for environmental conservation and health caring. Thereby, complete degradation of such xenobiotic molecules represents a big challenge. Benzoyl-coenzyme A is established as a central intermediate in the anaerobic biodegradation
pathway of many aromatic compounds such as benzene, toluene, benzoate, aniline and etc (Spormann and Widdel 2000; Carmona, Zamarro et al. 2009). Bacterial Benzoyl-coenzyme A reductases (BCRs) are described as key enzymes in the benzoyl-CoA degradation pathway. Genes coding for Bcr α, β, γ, δ subunits of BCRs are present in the genomes of denitrifying bacteria (*Thauera, Azoarcus, Magnetospirillum* species) or bacteria displaying an anoxygenic photosynthesis (*Rhodopseudomonas, Rhodobacterium* species)(Boll 2005). However, this kind of BCR enzyme has so far, only been isolated and characterized in the denitrifying, facultative anaerobe *Thauera aromatic* strain K172(Boll and Fuchs 1995). So, *Thauera aromatic* strain could be taken as the model for bcr operon and support the Genefish application.

The bcr operon in *Thauera* sp is composed of four genes bcr c, bcr b, bcr a and bcr d respectively (Fig.38). This bcr operon then was used as the template for for Genefish application.

![Fig.38: Structure of the bcr operon in Thauera.sp. Four genes bcr c, b, a, d are shown with different colors. The bcr operon (gi|19471177|emb|AJ224949.2) is 4576bp in length.](image)

3. Genefish application using bcr operon

3.1. Capture plasmid and host strain construction

3.1.1 Materials:

Plasmid pBAD35K7toxN (constructed by Nathalie Lombard) was used for capture plasmid construction, the host strain *E.coli* 14102 (F-lac-3350, galT22, IN (rrnD-rrnE) 1, rpsL179Δtrp::[KnR-immE3_lacIq] ΔaraBAD, ΔnarGnarH) used in this study was constructed and supplied by our collaborator Jean-Claude Lazzaroni (Unité de Microbiologie, Adaptation et pathogenie,UMR5240, Université de LYON,69622 Villeurbanne cedex,France). Metagenomic DNA from a quinoline degradation bioreactor (partner in Shanghai Jiao Tong University, China) was used as the template for highly conserved bcr fragment amplification. Table 5 and 6 present
primers (invitrogen) used for target fragment amplification.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence of primer 5’ to 3’</th>
<th>Tm(℃)</th>
<th>Length of PCR products</th>
</tr>
</thead>
<tbody>
<tr>
<td>bcr c fw-BglII</td>
<td>GAAGATCTTCGCCTACATGCCGTAT</td>
<td>60.5</td>
<td>270bp</td>
</tr>
<tr>
<td>bcr c rv-PvuI</td>
<td>ATCGATCGAACATCCGCTTCCACAT</td>
<td>60.4</td>
<td></td>
</tr>
<tr>
<td>bcr a fw-AvrII</td>
<td>ATCCTAGGAAGTCGAACGCATCAA</td>
<td>59.5</td>
<td>276bp</td>
</tr>
<tr>
<td>bcr a rv-HpaI</td>
<td>CGTGAACATGTGTGACTGACTTCG</td>
<td>59.5</td>
<td></td>
</tr>
<tr>
<td>bcr d fw-AvrII</td>
<td>TCCTAGGTGACGACACTGCAGGCA</td>
<td>63.2</td>
<td>277bp</td>
</tr>
<tr>
<td>bcr d rv-HpaI</td>
<td>CGTGAACATGTGTGACTGACTTCG</td>
<td>62.2</td>
<td></td>
</tr>
<tr>
<td>bcr c fw1</td>
<td>GCTTTACGAGGACCTTGATTTC</td>
<td>53.9</td>
<td>4134bp</td>
</tr>
<tr>
<td>bcr d rv1</td>
<td>CATCTGTAGGTTTGCAGCTTTG</td>
<td>53.5</td>
<td></td>
</tr>
<tr>
<td>bcr fw s</td>
<td>CGCAAGGTATCGGCTACAT</td>
<td>54.8</td>
<td>4084bp</td>
</tr>
<tr>
<td>bcr rv s</td>
<td>TGGCTGGTATCGGCTTAGGT</td>
<td>54.9</td>
<td></td>
</tr>
</tbody>
</table>

Table.5: Primers designed to amplify highly conserved sequences of bcr c; bcr a; bcr d; whole bcr operon based on the sequence of AJ224949.2 (NCBI). Red bases are the restriction enzyme recognition sites and green bases are the protection bases added to improve the digestion efficiency. Primers for whole bcr operon amplification: Primer bcr c fw1 and bcr d rv1 are used to amplify the whole bcr operon from metagenomic DNA samples. Then primer bcr fw s and bcr rv s were used to amplify shorter bcr operon taking the amplified bcr operon as the template for Genefish application.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence of primer 5’ to 3’</th>
<th>Tm(℃)</th>
<th>Length of PCR products</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1 fw2</td>
<td>TGAATGATGATGCGCTGCAAG</td>
<td>52.3</td>
<td>523 bp</td>
</tr>
<tr>
<td>r1 rv1</td>
<td>CAGGGCAGGGTGGTAAATA</td>
<td>53.9</td>
<td></td>
</tr>
<tr>
<td>r2 fw2</td>
<td>AAGAGATTACGGCGAGCCA</td>
<td>56.1</td>
<td>415 bp</td>
</tr>
<tr>
<td>r2 rv1</td>
<td>CAGGGCAGGGTGGTAAATA</td>
<td>56.5</td>
<td></td>
</tr>
</tbody>
</table>

Table.6: Primers designed to amplify homologous sequences inserted into the capture plasmid. All these primers targeted on the capture plasmid.

3.1.2 Methods:

3.1.2.1 Highly conserved bcr fragment sequences determination

Here, we take bcr c and bcr a as a group, bcr c and bcr d as the other group (group as the bait for recombination) to construct the capture plasmid. In order to get homologous sequences of bcr c, bcr a and bcr d gene, all sequences from different species of Thauera.sp (genera Thauera was dominant in metagenomic DNA supplied by our partner in Shanghai) were collected from NCBI and were aligned using the software ClustalX. Selected highly conserved bcr c, bcr a and bcr d were amplified by PCR while adding specific restriction enzyme sites corresponding to the multi-cloning site on plasmid pBAD35K7toxN. The PCR thermal program was set as 95℃ for 5 min, followed by 30 cycles at 95℃ for 30 s, 60℃ for 30 s, and 72℃ for 30 s, followed by a final extension performed at 72℃ for 5 min and then kept at 15℃. PCR products were verified,
digested by Bstz17I and StuI and purified by illustra GFX PCR DNA and Gel Band Purification Kit before the ligation step (GE Healthcare).

### 3.1.2.2 Capture plasmid construction

The first step of capture plasmid construction was a double enzyme digestion of bcr c fragment and pBAD35K7toxN by using BglII and PvuI restriction enzyme (Fermentas). Conditions for bcr c and pBAD35K7toxN digestions were as followed: 50 μl reaction in a DNase-free 500 μl tube containing 1 μl of BglII (10U/μl) and 1μl of PvuI (10U/μl), 5μl of purified bcr c fragment(6γng/μl) or 5μl of pBAD35K7toxN plasmid (supplied by Laure Franqueville,70ng/μl), 10 μl of 10x Tango Buffer and 33μl of double distilled H₂O. The reaction was performed in a 37°C water bath and afterwards the enzymes inactivated at 80°C for 20min. After the digestion reaction, products were verified using agarose gel electrophoresis. The 6360bp fragment of digested capture plasmid then was cut out of the gel and purified for further ligation use. However, the bcr c fragment length was about 270 bp and there were no significant changes before and after the digestion. So, the digested bcr c fragments were used directly for the next step of ligation.

Ligation between the digested pBAD35K7toxN plasmid and bcr c fragment using T4 DNA ligase (Fermentas) was performed as followed: 25μl ligation reaction contains 50 ng of vector and a 3-fold molar excess of insert (6 ng), 2.5μl of 10-x concentrated ligase buffer, 1μl T4 DNA ligase (1U/μl), and double distilled H₂O. The ligation was performed at 22°C for 3 hours in water bath.

Then highly conserved sequence bcr a or bcr d fragment (flanked with restriction enzyme sites) was inserted into the upstream multiple cloning site of the plasmid pBAD35K7toxN-bcrc thus forming two capture plasmids called pBAD35K7toxN-bcrc-a and pBAD35K7toxN-bcrc-d.

### 3.1.2.3 Host strain construction

The λ Red system is indispensable for our Genefish strategy. In this study, we constructed two types of host strain: one contains plasmid pSIM6 and the other contains both capture plasmid and pSIM6. When we use the former for Genefish application, capture plasmid and target exogenous DNA should be transformed into the host strain together for controlling the copy number of capture plasmid. While only target exogenous DNA was electroporated into the latter host strain, this could simplify the Genefish process.
For the host strain containing plasmid pSIM6: plasmid pSIM6 was transformed into *E.coli* 14102 competent cells. We added carbenicillin (50µg/ml) in culture medium as the selection marker for the pSIM6 transformants at 29°C.

For the host strain containing both capture plasmid and pSIM6: the constructed capture plasmid was transformed into pSIM6 containing host strain and the transformants were selected under the pressure of carbenicillin (50µg/ml) and chloramphenicol (50µg/ml).

### 3.1.2.4 Escape rate test

According to the Genefish design, only recombinants (in which toxic cassette has been replaced by targeted fragment) could survive and be screened. However, uncertain factors may cause mutation or deletion into the toxic gene plasmid sequences and this enable the strain escaping from death after induction. In the Genefish strategy, we define the escape rate as a probability of the host strain to escape from death after the toxic cassette induction. The relationship between total colony, escape rate and recombinant rate are shown in figure 39, by using the strain with a lower escape rate for Genefish application, higher recombinant ratio in survival colony would be screened.

![Fig.39: Relationship between total colonies, escape rate and recombinant rate. When the escape rate is high, false positive recombinants ratio in survival total colony would increase and this is undesirable results. Only when we use the strain that has the lowest escape rate, recombinants can predominant in total colony.](image-url)
3.1.3 Results

3.1.3.1 Highly conserved bcr fragment selected for capture plasmid construction

Eight bcr c gene sequences were collected from NCBI and the alignment results (table 7, Fig.40) showed a high similarity with few bases difference between those sequences. As the bcr c gene has 929 bp in length, part of the alignment results are presented below. These alignment results showed a conserved bcr c gene between 6380 bp and 6740 bp of original sequence AJ224949.2.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>62468033</td>
</tr>
<tr>
<td>gi</td>
<td>62468031</td>
</tr>
<tr>
<td>gi</td>
<td>62468029</td>
</tr>
<tr>
<td>gi</td>
<td>62468027</td>
</tr>
<tr>
<td>gi</td>
<td>62468025</td>
</tr>
<tr>
<td>gi</td>
<td>62468023</td>
</tr>
<tr>
<td>gi</td>
<td>62468021</td>
</tr>
<tr>
<td>gi</td>
<td>19471177</td>
</tr>
</tbody>
</table>

Table 7: Eight sequences of bcr c gene got from NCBI. The last sequence with reference AJ224949.2 is a whole operon involved in the benzoyl-CoA degradation pathway.

Fig.40: Alignment of the eight bcr c gene of Thauera sp collected in the table 7. The left titles are the reference of each sequence and the bases are in different colors. The homologous sequences of bcr c gene start from the 6380bp site of the sequence AJ224949.2.
For the *bcr a* gene, nine sequences were collected from NCBI and alignment results (table 8 and Fig.41) also showed that *bcr a* gene was highly conserved between 9770 bp and 10150 bp of original sequence AJ224949.2.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>62468127</td>
</tr>
<tr>
<td>gi</td>
<td>62468125</td>
</tr>
<tr>
<td>gi</td>
<td>62468123</td>
</tr>
<tr>
<td>gi</td>
<td>62468121</td>
</tr>
<tr>
<td>gi</td>
<td>62468119</td>
</tr>
<tr>
<td>gi</td>
<td>62468117</td>
</tr>
<tr>
<td>gi</td>
<td>62468115</td>
</tr>
<tr>
<td>gi</td>
<td>62468113</td>
</tr>
<tr>
<td>gi</td>
<td>19471177</td>
</tr>
</tbody>
</table>

Table 8: Nine sequences of *bcr a* gene get from NCBI. The last sequence with a reference AJ224949.2 is a whole operon involved in the benzoyl-CoA degradation pathway.

Fig.41: Alignment of the nine *bcr a* genes of *Thauera.sp* collected in the table 8. The left titles are the reference of each sequence and the bases are in different colors. The homologous sequences of *bcr a* gene start from the 8860bp site of the sequence AJ224949.2.
For the \textit{bcr} d gene, nine sequences were collected from NCBI and alignment results (table 9 and Fig.42) also showed that \textit{bcr} d gene was conserved between 10160 bp and 10520 bp of original sequence AJ224949.2.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>62468143</td>
</tr>
<tr>
<td>gi</td>
<td>62468141</td>
</tr>
<tr>
<td>gi</td>
<td>62468139</td>
</tr>
<tr>
<td>gi</td>
<td>62468137</td>
</tr>
<tr>
<td>gi</td>
<td>62468135</td>
</tr>
<tr>
<td>gi</td>
<td>62468133</td>
</tr>
<tr>
<td>gi</td>
<td>62468131</td>
</tr>
<tr>
<td>gi</td>
<td>62468129</td>
</tr>
<tr>
<td>gi</td>
<td>19471177</td>
</tr>
</tbody>
</table>

\textit{Table 9: Nine sequences of bcr d genes got from NCBI. The last sequence with a reference AJ224949.2 is a whole operon involved in the benzoyl-CoA degradation pathway.}

\textit{Fig.42: Alignment of the nine bcr d genes of Thauera.sp collected from table 9. The left titles are the reference of each sequence and the bases are in different colors. The homologous sequences of bcr d genes start from the 10160 bp site of the sequence AJ224949.2.}

According to the alignment results, the most conserved sequences of \textit{bcr} c, \textit{bcr} a and \textit{bcr} d genes from \textit{Thauera} sp. were confirmed. Six restriction enzyme digestion sites in multi-cloning sites were screened on the conserved \textit{bcr} c, \textit{bcr} a and \textit{bcr} d gene sequences. Results showed \textit{bcr} d gene contains a StuI and two PstI restriction enzyme sites and conserved \textit{bcr} c and \textit{bcr} a do not
show those restriction enzyme sites mentioned above. Finally, BglII and PvuI restriction enzyme were chosen for the downstream multiple cloning site digestion. AvrII and HpaI restriction enzyme for the upstream multiple cloning site digestion. All sequences were amplified by PCR using primers described in table 5.

3.1.3.2 Capture plasmid construction

According to the capture plasmid construction method (Fig. 43), two capture plasmids pBAD35K7toxN-bcrc-a and pBAD35K7toxN-bcrc-d were constructed and the inserted highly conserved sequences were verified by sequencing.

Fig. 43: Overview of the capture plasmids construction. Our original plasmid pBAD35K7toxN was used as the substrate for capture plasmids construction. Highly conserved sequence bcr c was first inserted into the downstream multiple cloning site. The highly conserved sequence bcr a or bcr d then was inserted to the constructed plasmid pBAD35K7toxN-bcrc respectively leading to the construction of two types of capture plasmid for Genefish application.

3.1.3.3 Host strain construction and escape rate test
Constructed capture plasmid were transformed into host strain *E.coli* 14102. Three colonies of the constructed host strain *E.coli* 14102 pBAD35K7toxN-brcr-a-9,-11,-13 and pBAD35K7toxN-brcr-d-1,-3,-7 were selected for escape rate test (protocol see annex 3). According to the escape rate results, colony pBAD35K7toxN-brcr-a-13 and pBAD35K7toxN-brcr-d-7 were chosen as the best strain for further Genefish explorations.

<table>
<thead>
<tr>
<th>ER Ara</th>
<th>ER Ara</th>
<th>ER Ara</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBAD35K7toxN-brcr-a-9</td>
<td>2.6E-07</td>
<td>pBAD35K7toxN-brcr-d-1</td>
</tr>
<tr>
<td>pBAD35K7toxN-brcr-a-11</td>
<td>4.0E-07</td>
<td>pBAD35K7toxN-brcr-d-3</td>
</tr>
<tr>
<td>pBAD35K7toxN-brcr-a-13</td>
<td>9.3E-07</td>
<td>pBAD35K7toxN-brcr-d-7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ER IPTG</th>
<th>ER IPTG</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBAD35K7toxN-brcr-a-9</td>
<td>9.5E-05</td>
</tr>
<tr>
<td>pBAD35K7toxN-brcr-a-11</td>
<td>1.3E-04</td>
</tr>
<tr>
<td>pBAD35K7toxN-brcr-a-13</td>
<td>6.7E-05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ER Ara and IPTG</th>
<th>ER Ara and IPTG</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBAD35K7toxN-brcr-a-9</td>
<td>1.3E-08</td>
</tr>
<tr>
<td>pBAD35K7toxN-brcr-a-11</td>
<td>2.5E-09</td>
</tr>
<tr>
<td>pBAD35K7toxN-brcr-a-13</td>
<td>2.3E-08</td>
</tr>
<tr>
<td>Dilutions</td>
<td>-4</td>
</tr>
<tr>
<td>-----------</td>
<td>----</td>
</tr>
<tr>
<td>9(CFU/plate)</td>
<td>740</td>
</tr>
<tr>
<td>1ml dilution 0</td>
<td>7.40E+07</td>
</tr>
<tr>
<td>Mean</td>
<td>7.50E+07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dilutions</th>
<th>-4</th>
<th>-5</th>
<th>-6</th>
<th>Dilutions</th>
<th>-4</th>
<th>-5</th>
<th>-6</th>
<th>Dilutions</th>
<th>-4</th>
<th>-5</th>
<th>-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>11(CFU/plate)</td>
<td>113</td>
<td>19</td>
<td>4</td>
<td>3(CFU/plate)</td>
<td>15</td>
<td>9</td>
<td>2</td>
<td>7(CFU/plate)</td>
<td>175</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>1ml dilution 0</td>
<td>1.10E+07</td>
<td>1.90E+07</td>
<td>4.00E+07</td>
<td>1ml dilution 0</td>
<td>1.50E+06</td>
<td>9.00E+07</td>
<td>7.90E+08</td>
<td>1ml dilution 0</td>
<td>1.70E+07</td>
<td>1.20E+07</td>
<td>1.30E+07</td>
</tr>
<tr>
<td>Mean</td>
<td>2.30E+07</td>
<td>Mean</td>
<td>1.00E+07</td>
<td>Mean</td>
<td>2.90E+07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dilutions</th>
<th>0</th>
<th>-1</th>
<th>Dilutions</th>
<th>0</th>
<th>-1</th>
<th>Dilutions</th>
<th>0</th>
<th>-1</th>
<th>Dilutions</th>
<th>0</th>
<th>-1</th>
<th>Dilutions</th>
<th>0</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>9(CFU/plate)</td>
<td>0</td>
<td>11</td>
<td>11(CFU/plate)</td>
<td>8</td>
<td>13(CFU/plate)</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ml dilution 0</td>
<td>2.00E+01</td>
<td>1ml dilution 0</td>
<td>1.60E+02</td>
<td>1ml dilution 0</td>
<td>4.00E+01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.00E+01</td>
<td>Mean</td>
<td>1.60E+02</td>
<td>Mean</td>
<td>4.00E+01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dilutions</th>
<th>0</th>
<th>-1</th>
<th>Dilutions</th>
<th>0</th>
<th>-1</th>
<th>Dilutions</th>
<th>0</th>
<th>-1</th>
<th>Dilutions</th>
<th>0</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(CFU/plate)</td>
<td>0</td>
<td>10</td>
<td>3(CFU/plate)</td>
<td>0</td>
<td>1</td>
<td>7(CFU/plate)</td>
<td>0</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ml dilution 0</td>
<td>2.00E+02</td>
<td>1ml dilution 0</td>
<td>2.00E+01</td>
<td>1ml dilution 0</td>
<td>8.00E+01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.00E+02</td>
<td>Mean</td>
<td>2.00E+01</td>
<td>Mean</td>
<td>8.00E+01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dilutions</th>
<th>-1</th>
<th>-2</th>
<th>Dilutions</th>
<th>-1</th>
<th>-2</th>
<th>Dilutions</th>
<th>-1</th>
<th>-2</th>
<th>Dilutions</th>
<th>-1</th>
<th>-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>9(CFU/plate)</td>
<td>104</td>
<td>4</td>
<td>11(CFU/plate)</td>
<td>760</td>
<td>33</td>
<td>13(CFU/plate)</td>
<td>28</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ml dilution 0</td>
<td>1.00E+04</td>
<td>4.00E+03</td>
<td>1ml dilution 0</td>
<td>7.60E+04</td>
<td>3.30E+04</td>
<td>1ml dilution 0</td>
<td>2.80E+03</td>
<td>3.00E+03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>7.20E+03</td>
<td>Mean</td>
<td>5.40E+04</td>
<td>Mean</td>
<td>2.90E+03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dilutions</th>
<th>-1</th>
<th>-2</th>
<th>Dilutions</th>
<th>-1</th>
<th>-2</th>
<th>Dilutions</th>
<th>-1</th>
<th>-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(CFU/plate)</td>
<td>496</td>
<td>55</td>
<td>3(CFU/plate)</td>
<td>392</td>
<td>47</td>
<td>7(CFU/plate)</td>
<td>264</td>
<td>40</td>
</tr>
<tr>
<td>1ml dilution 0</td>
<td>4.90E+04</td>
<td>5.50E+04</td>
<td>1ml dilution 0</td>
<td>3.90E+04</td>
<td>4.70E+04</td>
<td>1ml dilution 0</td>
<td>2.60E+04</td>
<td>4.00E+04</td>
</tr>
<tr>
<td>Mean</td>
<td>5.20E+04</td>
<td>Mean</td>
<td>4.30E+04</td>
<td>Mean</td>
<td>3.30E+04</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dilutions</th>
<th>0</th>
<th>Dilutions</th>
<th>0</th>
<th>Dilutions</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>9(CFU/plate)</td>
<td>0</td>
<td>11(CFU/plate)</td>
<td>0</td>
<td>13(CFU/plate)</td>
<td>0</td>
</tr>
<tr>
<td>1ml dilution 0</td>
<td>1ml dilution 0</td>
<td>1ml dilution 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.00E+00</td>
<td>Mean</td>
<td>1.00E+00</td>
<td>Mean</td>
<td>1.00E+00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dilutions</th>
<th>0</th>
<th>Dilutions</th>
<th>0</th>
<th>Dilutions</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(CFU/plate)</td>
<td>0</td>
<td>3(CFU/plate)</td>
<td>0</td>
<td>7(CFU/plate)</td>
<td>0</td>
</tr>
<tr>
<td>1ml dilution 0</td>
<td>1ml dilution 0</td>
<td>1ml dilution 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.00E+00</td>
<td>Mean</td>
<td>1.00E+00</td>
<td>Mean</td>
<td>1.00E+00</td>
</tr>
</tbody>
</table>

Table 10: Escape rate (ER) results of E. coli strain 14102, pBAD35K7toxN-bcrc-a-9,-11,-13 and pBAD35K7toxN-bcrc-d-1,-3,-7 in table a; table below shows the counting results in the test.
3.2. Genefish tool application

3.2.1 Materials

*E. coli* 14102 pBAD35K7toxN-bcrc-a-13 and pBAD35K7toxN-bcrc-d-7 were chosen as the host strain for the Genefish application. Two different types of target *bcr* genes (*bcr c-a* 3617 bp and *bcr c-d* 4084 bp) were used as the donor of targeted fragments to test the Genefish capacity for capturing different length of target fragment. Primers R1rv1 (5’-CAGGGCAGGGTCGTTAAATA-3’) and R2fw2 (5’-AAGAGATTCGCGCAGACCA-3’) were used to verify positive recombinants. This pair of primers was used to amplify the region located between conserved targeted sequences of the capture plasmid. PCR products of positive recombinants and pBAD35K7toxN-bcrc-d were 4290 bp and 4945 bp in length respectively. In order to confirm the insertion of targeted exogenous fragment in the recombinants, seven groups of primers were designed targeted on the *bcr* operon (Table 11).

<table>
<thead>
<tr>
<th>Region</th>
<th>Primers</th>
<th>Sequence of primer 5' to 3'</th>
<th>Annealing temperature(°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoom 1</td>
<td>rct fw 1-2</td>
<td>GCAGACCAAAACGATCTCA</td>
<td>57.7</td>
</tr>
<tr>
<td>rct rv 1-2</td>
<td>CCAGGTAGTCTCTTCAGCATC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoom 2</td>
<td>rct fw 2-2</td>
<td>GGTCGAAGGAGCACAAACCAG</td>
<td>57.9</td>
</tr>
<tr>
<td>rct rv 2-2</td>
<td>TTGATGACTTCAGGGTTGTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoom 3</td>
<td>rct fw 3-2</td>
<td>TGCCCTACATCTCGTTCAA</td>
<td>57.4</td>
</tr>
<tr>
<td>rct rv 3-2</td>
<td>CGCCGAGTTCTTCAGGATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoom 4</td>
<td>rct fw 4-2</td>
<td>CGCCACGAATAACAGTGC</td>
<td>58.6</td>
</tr>
<tr>
<td>rct rv 4-2</td>
<td>CAAGGTGCGGTCTCGATGAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoom 5</td>
<td>rct fw 5-2</td>
<td>CTCGATCAAGGAGCTGCAACA</td>
<td>59.5</td>
</tr>
<tr>
<td>rct rv 5-2</td>
<td>TTCACCGATCTGCAAGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoom 6</td>
<td>rct fw 6-2</td>
<td>GAAGAGGCGGACAAGCTTTT</td>
<td>59.5</td>
</tr>
<tr>
<td>rct rv 6-2</td>
<td>TTGATGCGGATCGACTTGGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoom 7</td>
<td>rct fw 7-2</td>
<td>AACATGGGTCTGCACGAAC</td>
<td>58.7</td>
</tr>
<tr>
<td>rct rv 7-2</td>
<td>TTGACCTACCGGAAGCAGTG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 11 and fig.44: PCR primers designed to confirm positive recombinants.*

**pBAD35K7toxN-sens 2-bcr c-a**
5788 bp
3.2.2 Methods:

According to the Genefish strategy, former PhD student Samuel Jacquiod did some tests (Thesis defence 11/2012). However, results showed recombinants cannot be easily screened (low number of recovered recombinants) because each cell contains 15-30 copies of capture plasmid. Because of this, only the host cell contains all recombined capture plasmids can survive after the toxic cassette induction otherwise the toxicity expressed by non-recombined plasmid could still kill the cell. To deal with this situation is to strictly control the copy number between capture plasmid and host strain during the homologous recombination. Two methods called co-electroporation and plasmid segregation were used in this study.

3.2.2.1 Co-electroporation

The aim of this strategy was to simultaneously transform the capture plasmid and the target DNA fragment into pSIM6 containing host strain. During the transformation, mole ratio between capture plasmid and the host strain were strictly controlled as 1:1. Fig.45 shows the process of co-electroporation.

![Co-electroporation process](image)

*Fig.45: co-electroporation process used in our study. The host strain containing pSIM6 is induced at 42°C for 15minin. Then, the mix of capture plasmid and target DNA is co-electroporated into the competent cells by controlling their mole ratio.*
3.2.2.2 Plasmid segregation

Firstly, target DNA fragments were transformed into the host strain containing pSIM6 and the capture plasmid. Then, after the transformation, total plasmids (reservoir of recombined or non-recombined plasmids) were extracted. In the last step, plasmids mixture were transformed into plasmid-free host strain while controlling the molar ratio (1:1) between plasmid mixture (copy number) and competent cells (number) (Fig.46).

Fig.46: Plasmid segregation process. Target DNA is electroporated into capture plasmid and pSIM6 containing competent cells. The plasmid pool of transformants are extracted and used as the substrate for a second run of electroporation.
3.2.3 Results.

3.2.3.1 Co-electroporation by using bcr c-d fragment

Results obtained with co-electroporation (annex 4) showed that the total number of colony including transformed plasmid decreased when more target DNA were used (table 12-a). In the dilution of $10^{-2}$, only 16 colonies of ratio 1:1:5 grew in the CM containing medium. The total number of survival colony decrease when using the ratio 1:1:10 for co-electroporation (6 colonies and no colony grew in the dilution of $10^{-2}$ and $10^{-3}$). So, decreasing the quantity of DNA fragment for the co-electroporation may generate more colonies and this will also eliminate the risk of electro sparks during the electroporation. Only one colony grew using the Lambda Red activated competent cells while there was also one colony growing in the negative control (table 12-b).

Further co-electroporation experiments using the ratio 1:1:0.5, 1:1:1, 1:1:2, 1:0.5:0.5, 1:0.5:1 and 1:0.5:2. Table 7 shows the recombination experiment results.

<table>
<thead>
<tr>
<th>Mole ratio (strain:plasmid:DNA)</th>
<th>Total colony in each dilution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>1:1:0 (No CM, with Kn)</td>
<td>++</td>
</tr>
<tr>
<td>1:1:5 (No CM, with Kn)</td>
<td>++</td>
</tr>
<tr>
<td>1:1:10 (No CM, with Kn)</td>
<td>++</td>
</tr>
<tr>
<td>1:1:5 (CM with Kn)</td>
<td>16</td>
</tr>
<tr>
<td>1:1:10 (CM with Kn)</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mole ratio (strain:plasmid:DNA)</th>
<th>Recombinants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>42°C (activated)</td>
</tr>
<tr>
<td>1:1:5 (Cm+IPTG+Ara)</td>
<td>0</td>
</tr>
<tr>
<td>1:1:10 (Cm+IPTG+Ara)</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 12: total number of colonies (a) and potential recombinants (b) after co-electroporation. Colony screened in the medium without CM stand for competent cells survival after electroporation. While colony screened with CM means capture plasmid was well transformed. However, among these colonies, only 1 colony may be recombinant (table b)
Total colony after toxic cassette induction

<table>
<thead>
<tr>
<th>Copy control (cell: plasmid:DNA)</th>
<th>Repeat 1(colony)</th>
<th>Repeat 2(colony)</th>
<th>Repeat 3(colony)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1:5</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:1:10</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:1:0.5</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1:1:1</td>
<td>9</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>1:1:2</td>
<td>7</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1:0.5:0.5</td>
<td>3</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>1:0.5:1</td>
<td>0</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>1:0.5:2</td>
<td>6</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Total colony</td>
<td>1</td>
<td>26</td>
<td>37</td>
</tr>
</tbody>
</table>

*Table 13: co-electroporation results using bcr c-d fragments. All these transformants will be verified in the following step.*

![Gel electrophoresis of colony PCR products screened after selection. M and + stand for 1kb plus DNA marker and positive control (PCR products of plasmid pBAD35K7toxN-bcrc-d).](image)

After PCR verification (Fig.47 and 48), none of grown colonies were positive recombinants. The size of PCR products showed multiple differences in length (table 14). As we have mentioned, the positive recombinants should have a PCR products around 4290bp long. These results told us that most of the capture plasmid deleted the toxic cassette to escape from the death thus generate the PCR products with different size.

<table>
<thead>
<tr>
<th>Mole ratio (cell:plasmid:DNA)</th>
<th>No PCR products</th>
<th>PCR Products 1-1000bp</th>
<th>1000-2000bp</th>
<th>2000-3000bp</th>
<th>3000-4000bp</th>
<th>4000bp</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1:5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>1:1:10</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>co-electroporation with ratio 1:1:0.5, 1:1:1, 1:1:2, 1:0.5:0.5, 1:0.5:1, 1:0.5:2</td>
<td>3</td>
<td>13</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td></td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>16</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td>37</td>
</tr>
</tbody>
</table>

*Table 14: distribution of PCR amplified fragments. Positive recombinants have a 4290bp PCR product while all those transformants have shorter PCR products patterns.*
3.2.3.2 Plasmid segregation by using pBAD35K7toxN-bcrc-d

At last, more than 674 colonies (table 15) were obtained (annex 5) and the 94 colonies from mole ratio condition 1:0.5(1:2) and 1:0.5(1:4) were selected for recombinants verification using colony PCR (Fig 49). The mole ratio (1:2) means the strain: plasmid (extracted from condition of 1:0.5 in the first round electroporation). Mole ratio 1:0.5(1:2) contains two parts, the first part 1:0.5 means mole ratio between strain and plasmid. Then the plasmid was extracted and taken as the source of recombinant plasmid for the second round of electroporation.

<table>
<thead>
<tr>
<th>Mole ratio (strain : DNA)</th>
<th>Mole ratio (strain : Plasmid)</th>
<th>colony</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:0.5</td>
<td>1:0.5</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>1:1</td>
<td>++</td>
</tr>
<tr>
<td>1:1</td>
<td>1:0.5</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>1:1</td>
<td>++</td>
</tr>
<tr>
<td>1:2</td>
<td>1:0.5</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>1:1</td>
<td>207</td>
</tr>
<tr>
<td>1:4</td>
<td>1:0.5</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>1:1</td>
<td>261</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>674++</td>
</tr>
</tbody>
</table>

Table 15: transformants screened using plasmid segregation protocol. More than 674 colonies were captured, + mean colonies contaminated. 94 colonies in green were selected for colony PCR verification.

Fig. 49: Colony PCR products gel electrophoresis, M and + mean 1kb plus DNA marker and positive control (pBAD35K7toxN-bcrc-d) respectively. PCR products of positive recombinants and pBAD35K7toxN-bcrc-d were 4290 bp and 4945 bp in length.
Colonies 77 and 94 may be positive recombinants as their PCR products are close to the expected size. So far, the plasmid segregation seems to work better than co-electroporation because more transformants were screened and this gave more opportunity to screen positive recombinants. Primer group 1-4 (PCR products room 1-4, see Fig. 44) were used to amplify the target sequence in potential recombinants 77 and 94 (Fig.50).

Fig.50: Specific PCR used to amplify captured bcr operon fragment. Three groups of primers were used and + stands for the positive PCR products size of captured colony pBAD35K7toxN-bcrc-a D8. 77 and 94 were supposed recombinants candidates. Lane M and – are 1kb plus DNA marker and negative control respectively. After the gel electrophoresis, colony 77 and 94 did not show predicted band signal.

Former colony PCR results showed colonies 77 and 94 have positive band signal and they were considered as positive recombinants, but no predicted band signal were detected using bcr operon targeted primers (group 1-4). Finally, colonies 77 and 94 were false positive recombinants.

3.2.3.3 Plasmid segregation by using pBAD35K7toxN-bcrc-a

330 transformants were screened (table 16) and these colonies were first verified by colony PCR.

<table>
<thead>
<tr>
<th>Copy control (strain : DNA)</th>
<th>Copy control (strain :Plasmid)</th>
<th>colony</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:0.5</td>
<td>1:0.5</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>1:1</td>
<td>47</td>
</tr>
<tr>
<td>1:1</td>
<td>1:0.5</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>1:1</td>
<td>77</td>
</tr>
<tr>
<td>1:2</td>
<td>1:0.5</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>1:1</td>
<td>76</td>
</tr>
<tr>
<td>1:4</td>
<td>1:0.5</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>1:1</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>330</td>
</tr>
</tbody>
</table>

Table 16: colonies screened after plasmid segregation using plasmid pBAD35K7toxN-bcrc-a.
Because too many colonies were verified by gel electrophoresis, only recombinant candidates were presented here in Fig 51.

Colony D8 was chosen and verified by primers group 1-4 (table 10). Results shown colony D8 has all positive signals in the agarose gel. All amplified four fragments were sequenced and compared with the whole \textit{bcr} operon and the alignment results show they match 99.9%. There are few mismatches but this indicated colony D8 was a recombinant by using the new tool Genefish with pBAD35K7toxN-bcrc-a.

3.3. Genefish improvements

From the results of both co-electroporation and plasmid segregation, false positive recombinants account for a large proportion in survival total colony. All transformants captured using pBAD35K7toxN-bcrc-d were false positive and only one positive recombinant was captured in 330 transformants. Among those false positive recombinants, toxic cassette in capture plasmid was deleted in varying degrees which were testified by PCR products sequencing (thesis Samuel Jacquiod 11/2012). It is unusual to get so many false positive recombinants even if the strain we used has the lowest escape rate.

3.3.1. Materials

Capture plasmids pBAD35K7toxN-bcrc-a and pBAD35K7toxN-bcrc-d, host strain \textit{E.coli} 14102 containing plasmid pSIM6 and In-Fusion HD Cloning Kit (Clontech) were used.
Single-copy number plasmid pJEL250 (Masai, Kaziro et al. 1983; Valentin-Hansen 1986) was supplied by Jean-Claude Lazzaroni (Fig.52). The pJEL250 contains the ORI region from plasmid R1, the gene encoding the replication initiation protein (repA1), genes parAB (Bagdasarian, Lurz et al. 1981) coding for ParA-ParB (partitioning proteins), and other copy number controlling genes copAE. All those genes are under the control of the thermosensitive CI857 repressor, each cell contains one copy of pJEL250 when the culture temperature is less than 37°C. Primer Bstz-pJEL250 fw: tgtatacGCTGCGGTAAGTCGCATAAA, StuI-pJEL250 rv: aaggcctAGATCGTTACCCGCAA ACA were used for single copy number capture plasmid construction through classical digestion and ligation. Primer Infusion fw: TTGGAACCTCTTTACGTGGCTGCGGTAAGTGCATAAAAA and primer Infusion rv: GGCAGATCTAGGCCTAGATCGTTACCCGCAAACA were used to amplify targeted sequences on plasmid pJEL250. Each 5’ end of primer contains 15bp (underlined) homologous to one end of digested capture plasmid.

Fig.52: presentation of the single-copy number plasmid pJEL250. All copy number controlling genes are marked out in arrows (From Laure Franqueville).

3.3.2 Methods:

3.3.2.1 Improvement of co-electroporation and plasmid segregation efficiency through culture time decrease

According to the process of co-electropration and plasmid segregation (fig.53), long culture time of the host strain may be the factor causing the toxic cassette deletion. Especially during the process of large amount capture plasmid preparation. This is a key step because the prepared capture plasmid will be used as the substrate for co-electroporation and plasmid segregation. If
the capture plasmid delete toxic cassette during the culture, transformed plasmids will cause the emergence of false positive recombinants.

**Fig.53: Over-view of the Genefish process in this study. Host strain with capture plasmid is chosen according to its low escape rate. Then this strain is cultivated more than 10 hours for plasmid extraction. The extracted capture plasmid (marked into square) will be used as the substrate for co-electroporation and plasmid segregation.**

During plasmid preparation process, starter culture (overnight host strain pre-cultured) was taken as the inoculum for large scale culture of host strain as quick as possible. But after long time culture, cells consumed the nutriments from the medium and accumulated secondary metabolites especially in the starter culture. Moreover, toxic cassette has a basal expression and toxicity in the culture medium will accumulate during strain growth. In order to resolve this conflict, host strain needs to regulate the balance between growth and toxic protein accumulation. Thus deleting the toxic cassette of capture plasmid seems to be an effective strategy.

Decreasing the culture time may also decrease or even avoid the false positive recombinants in transformants. Starter culture step was canceled but enriched medium SOB was used to improve the growth of the host strain. Capture plasmids pBAD35K7toxN-bcrc-a and pBAD35K7toxN-bcrc-d were freshly prepared through this new strategy. The host strain was transformed again by using those freshly prepared capture plasmid to improved co-electroporation and plasmid segregation.

### 3.3.2.2 Single-copy plasmid construction

As described before, the capture plasmid in this study has 15-30 numbers of copies per cell. This argues the co-electroporation and plasmid segregation strategy. However, the fundamental solution would be to construct a single-copy capture plasmid. Single-copy capture plasmid could simplify the experiments and also increase the recombinant detection rate. In this single-copy...
plasmid construction, the plasmid should contain the toxic cassette, the highly conserved sequences from capture plasmid and the system responding to single-copy number plasmid phenotype. In order to construct a single-copy plasmid, the ORI of our capture plasmid should be replaced by the single-copy regulation system including the ORI region from pJEL250 (Masai, Kaziro et al. 1983; Valentin-Hansen 1986). Finally, three strategies were employed to construct single-copy number capture plasmid.

3.3.2.2.1 Classical digestion and ligation

In this strategy, it is needed to construct the vector and PCR fragment with the same restriction enzyme sites at each end. As the plasmid pJEL250 contains the BsaAI restriction enzyme site in the expected PCR fragment (Fig.54). So, Bstz17I and StuI were chosen for restriction enzyme digestion. The double enzyme digestion of plasmid pBAD35K7toxN-bcrc-a was done by using Bstz17I and StuI. 50µl digestion reactor contains 0.5µl of StuI (10U/µl), 1µl of Bstz17I (5U/µl), 10µl of capture plasmid pBAD35K7toxN-bcrc-a (110ng/µl), 5µl of 10×NE buffer 4 (Biolabs) and 33.5µl of double distilled H2O. The reaction was performed in a 37°C water bath and inactivates the enzymes at 80°C for 20min.

For PCR amplification of the expected fragment, primer Bstz17I-pJEL250 fw and primer stuI-pJEL250 rv were used. The PCR thermal program was set as 95°C for 5 min, followed by 30 cycles at 95°C for 30 s, 60°C for 30 s, and 72°C for 5min, followed by a final extension performed at 72°C for 5 min and then kept at 15°C. PCR products were verified, digested by Bstz17I and StuI and purified by illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) before the ligation step. 25µl ligation reaction system contains 5 µl 5x ligation buffer, 4 µl linearized vector (14.4ng/µl), 6.2 µl PCR fragment (23.5ng/µl), 2µl T4 ligase enzyme (5U/µl), 2µl PEG4000 and 5.8µl double distilled H2O. Chemical calcium chloride transformation method was tested.
3.3.2.2 In-Fusion HD Cloning Kit method

In-Fusion HD Cloning Kit (Clontech) was involved in this new capture plasmid construction (Fig.55). This kit is designed for fast, directional cloning of one or more fragments of DNA into a vector. The cornerstone of In-Fusion Cloning technology which is Clontech’s proprietary In-Fusion Enzyme, which fuses DNA fragments e.g. PCR-generated sequences and linearized vectors, efficiently and precisely by recognizing a 15 bp overlap at their ends. This 15 bp overlap can be engineered by designing primers for amplification of the expected sequences. In-Fusion HD Kits offer increased cloning efficiency, especially for longer fragments, short oligonucleotides, and multiple fragments.

The double enzyme digestion of plasmid pBAD35K7toxN-brc-c-a was done by using BsaI and Stul. 50μl digestion reaction contains 0.5μl of Stul (10U/μl), 1μl of BsaI (5U/μl), 10μl of capture plasmid pBAD35K7toxN-brc-c-a (110ng/μl), 5μl of 10×NE buffer 4(Biolabs) and 33.5μl of double distilled H2O. The reaction was performed in a 37°C water bath and enzymes were inactivated at 80°C for 20min. The digested products then were verified by agarose gel electrophoresis. The targeted digested fragments contained two fragments, one was about 1146bp and the other was about 5732bp. The 5732bp fragment then was extracted from agarose gel and purified by Nucleo Spin Extract II (Macherey N).
Primer Infu fw and Infu rv were used to amplify targeted pJEL250 fragment flanked by 15bp homologous sequences corresponding to the capture plasmid at each end. The PCR thermal program was set as follow: 95°C for 5 min, followed by 30 cycles at 95°C for 30 s, 60°C for 30 s, and 72°C for 5 min, followed by a final extension performed at 72°C for 5 min and then kept at 15°C. Advantage HD Polymerase Mix (Clontech) was used for the PCR reaction. The PCR product were verified and purified by NucleoSpin Extract II kit (Macherey-Nagel). Concentration of purified PCR and capture plasmid products was tested by using Qubit dsDNA BR Assay Kit (Invitrogen). 25µl ligation reaction system contains 5 µl 5x ligation Buffer, 4 µl linearized vector (14.4ng/µl), 6.2 µl PCR fragment (23.5ng/µl), 2µl T4 ligase enzyme (5U/µl), 2µl PEG4000 and 5.8µl double distilled H2O. Chemical calcium chloride transformation method was tested.

![Diagram](image)

**Fig.55:** schema presenting the structure of capture plasmid pBAD35K7toxN-brc-c-a and pJEL250 and the In-Fusion cloning. The sequences between cmR and brc c gene contain three unique restriction enzyme digestion sites BsaAI, BstZ17I and StuI. While three BsaAI digestion sites exist in targeted fragment of pJEL250. So, BstZ17I and StuI sites were chosen for In-Fusion cloning method.
3.3.3 Results:

3.3.3.1 Improved co-electroporation

The results of improved co-electroporation are shown in table 17. The transformants containing capture plasmid pBAD35K7toxN-bcrc-a and pBAD35K7toxN-bcrc-d in the ratio of 1:1:1 (strain: plasmid: DNA) was about 3.0x10^3 and 9.0x10^3 respectively. While in the control
condition 1:1(strain:plasmid), about 9.5x10^3 and 3.0x10^4 transformants containing capture plasmid pBAD35K7toxN-bcrc-a and pBAD35K7toxN-bcrc-d respectively were screened. Comparing the transformants screened by Cm (50μg/ml) between the condition 1:1 and 1:1:1, when exogenous DNA was added for recombination, 2/3 of transformants were lost. The total number of competent cells after the transformation was about 2.88 \times 10^6 and 4.34 \times 10^6 respectively. Although there was no colony growing in the plates after the induction of toxic genes, this strategy was a great improvement as it decreased the false positive recombinants proportion.

<table>
<thead>
<tr>
<th>Mole ratio</th>
<th>With Cm (50μg/ml)</th>
<th>With kn(25μg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10^2</td>
<td>10^3</td>
</tr>
<tr>
<td>1:1(strain: plasmid)</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>1:1:1(strain: plasmid: DNA)</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

Table17: results of co-electroporation using improved pBAD35K7toxN-bcrc-a and pBAD35K7toxN-bcrc-d plasmid preparation.

3.3.3.2 Improved plasmid segregation

For the plasmid segregation, newly prepared capture plasmid was transformed into the host strain containing pSIM6. After the liquid activation and escape rate test, colony pBAD35K7toxN-bcrc-a a1 (escape rate 2.4E-08) and pBAD35K7toxN-bcrc-d b3 (escape rate 1.1E-09) were chosen for plasmid segregation.

The mole ratio (1:1, 1:2 and 1:4) between competent cells and targeted fragment were used in the first round electroporation, each condition repeated 4 times to enrich the plasmid. For the second round of electroporation, the plasmid mixture (both recombinants and non-recombinants) was transformed into E.coli 14102, host strain with a mole ratio 1:1 and 1:2(competent cells:
plasmid). Results are shown in table 18 and 54 colonies were obtained.

<table>
<thead>
<tr>
<th>Mole ratio (strain : DNA)</th>
<th>Mole ratio (strain : Plasmid)</th>
<th>Colony</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1</td>
<td>1:1</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>1:2</td>
<td>10</td>
</tr>
<tr>
<td>1:2</td>
<td>1:1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1:2</td>
<td>6</td>
</tr>
<tr>
<td>1:4</td>
<td>1:1</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>1:2</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>54</td>
</tr>
</tbody>
</table>

*Table 18: results from improved plasmid segregation using pBAD35K7toxN-bcrc-a.*

Condition 1:2:1 (green) was taken as the control for total transformants and total cells counting after the electroporation. For this condition, $2.3 \times 10^7$ cells were screened under kanamycin selection and only 600 transformants contained the capture plasmid (screened under chloramphenicol selection). Among those 600 transformants, only 3 colonies escaped from death after the induction of the toxic cassette with IPTG and arabinose.

All 54 transformants were verified by colony PCR, two colonies might be positive recombinants because they generated the same PCR products size comparing to recombinant D8 (sequenced positive recombinant). But after the specific PCR which aimed at amplifying recombined *bcr* fragment, they were found as cheaters (Fig.57). Similar results also obtained after improved plasmid segregation, 64 transformants were screened. Three colonies showed the same colony PCR products size comparing to the positive recombinant. But after PCR verification, none of them were recombinants (gel not shown).

PCR products zoom 1 of colony a2 and a3 were sent for sequencing. Sequences were aligned to the original capture plasmid pBAD35K7toxN-bcrc-a, and results showed that the sequences between 3000 bp and 3900 bp of the capture plasmid were deleted. This deleted region contains the *relF* and *rrnB* gene. After deletion, colony escaped from death and generated the similar positive colony PCR products size.
Fig. 57: two colonies a2, a3 may be positive recombinants according to their colony PCR products (left), a4 is taken as a negative control. After the specific PCR which uses three group primers to amplify inserted bcr operon, none of them are positive recombinants. + is the positive control (pBAD35K7toxN-bcrc-a recombinant colony D8).

Because too much dsDNA transformed into competent cells usually cause sparks during electroporation and lead low survival rate of host strain. In previous electroporation, not too much target dsDNA was transformed for recombination, maximal dose was about 60ng. However, more dsDNA electroporated into the competent cells could increase the rate of recombination. So, when large amount dsDNA were used for electroporation, other cuvettes from cell projects were employed to avoid the sparks.

In the plasmid segregation experiments, 250 ng of target dsDNA brc-a gene were transformed into its corresponding competent cells. Here, a negative control: electroporation without target dsDNA was done. Negative control plasmids were also transformed into E.coli competent cells during the second round of electroporation. After the selection, 29 colonies and 15 colonies grew in the selection plates respectively (table 19). 15 colonies obtained in the negative control, while no targeted fragments were transformed in this control which means those colonies contain capture plasmid deleted from their toxic cassette. This also indicated the culture time (around 10 hours) for capture plasmid and pSIM6 containing strain was still too long. The other 29 colonies were verified by colony PCR, and results showed that 2 positive recombinants were captured (Fig.58 and 59). bcr c-d gene was also taken as the substrates for plasmid segregation, but no recombinants were captured in 26 captured colonies.

<table>
<thead>
<tr>
<th>Repeats</th>
<th>First</th>
<th>Second</th>
<th>Third</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid segregation(colony)</td>
<td>12</td>
<td>12</td>
<td>5</td>
<td>29</td>
</tr>
<tr>
<td>Negative control(colony)</td>
<td>12</td>
<td>3</td>
<td>0</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 19: colonies screened after the improved plasmid segregation using pBAD35K7toxN-bcrc-a

93
Fig. 58: gel electrophoresis of colony PCR products using plasmid segregation. M and + mean the 1kb plus DNA marker and positive control (colony recombinant D8). 1 and 2 show the positive recombinants.

Fig. 59: colonies verification by bcr operon specific PCR. Zoom 1, 2 and 3 are fragments amplified using group 1, 2 and 3 primers. After this verification, we can see colonies 1 and 2 are positive recombinants as they show the positive signal of bcr operon products.
3.3.3.3 single-copy capture plasmid construction

Classic ligation and In-Fusion method were tested to construct the single-copy plasmid

After three repeat, no expected colony grows in the plates using the In-Fusion and classical ligation samples (table 20).

<table>
<thead>
<tr>
<th>quantity</th>
<th>Positive control</th>
<th>Capture plasmid</th>
<th>Plasmid pJEL250</th>
<th>In-Fusion sample</th>
<th>In-Fusion sample</th>
<th>Classic ligation sample</th>
<th>Classic ligation sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colony</td>
<td>1µl</td>
<td>1ng</td>
<td>1ng</td>
<td>2µl</td>
<td>4µl</td>
<td>2µl</td>
<td>4µl</td>
</tr>
<tr>
<td></td>
<td>92</td>
<td>124</td>
<td>106</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>(CFU)</td>
<td>51</td>
<td>168</td>
<td>47</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>72</td>
<td>127</td>
<td>68</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

*Table 20: results of classical ligation and novel In-Fusion method after the chemical transformation.*

Although the In-Fusion HD Cloning Kit enables us to fusion targeted linearized vectors and PCR products, but the In-Fusion Kit seems works not very well with our substrates. The reason might be the long target fragment used in this study. Long insert decreases the efficiency of ligation and might also decrease the efficiency of the In-Fusion process.

In order to explain this low In-Fusion efficiency, we tried to confirm the existence of expected single-copy plasmid in the In-Fusion products. The transformation efficiency of our host strain was also investigated. Primers R1fw2 (TGAATGATGTAGCCGTAAG), R2rv1 (CAGGGCAGGGTCGTTAAATA) were used to amplify the inserted target fragment of In-Fusion products. If there was expected single-copy plasmid, the PCR products should be longer than using template of pBAD35K7toxN-bcrc-a. But no PCR products were detected using In-Fusion products which mean the In-Fusion reaction didn’t work. Transformation efficiency of our host strain was tested by using plasmid pBAD35K7toxN-bcrc-a, and the optimized efficiency was $10^6$ cfu/µg. But the protocol of In-Fusion kit requires a transformation efficiency of at least $10^8$ cfu/µg.
4. Conclusions and Discussion

Genefish tool aims to capture targeted gene or DNA fragment from metagenomic DNA. This tool depends on the λ Red recombination system and toxic cassette induction which will cause phenotype difference between recombinants and non-recombinant cells. In the frame of the development of Genefish, it is essential to test each unit involved in the tool such as (1) the genetic modification of host strain; (2) escape rate of the host strain containing capture plasmid; (3) efficiency of λ Red system inducing homologous recombination when using different sizes of targeted sequence or highly conserved sequences.

First studies developed by Nathalie Lombard during her PhD (Thesis 2007), were focused on the construction of the toxic cassette system which contains two suicide genes (\textit{colE3} and \textit{relF}) and the validation of host strain that already described above (section host strain construction and toxic cassette). Then Laure Franqueville tested the efficiency of λ Red system for homologous recombination: the pSIM6 was chosen to achieve chromosome recombination. The results showed homologous recombination efficiency was around $10^{-4}$ (recombinants/competent cell) (supplied by Laure Franqueville).

So, the Genefish system was established and the aim of my work was to test the capacity and limitations of Genefish. The capture plasmid containing two highly conserved sequences was constructed and transformed into the host strain containing the pSIM6 plasmid. Targeted exogenous DNA was transformed into the best host strain (with the lowest escape rate) containing both the capture plasmid and pSIM6. However, this direct transformation strategy (original Genefish strategy) was not suitable because the capture plasmid has about 15 copies in each cell and the λ Red system was not sufficient to recombine all plasmid. So, the residual capture plasmid in the transformants would lead to cell death after the toxic cassette induction. We finally lost the recombinants and also failed to capture the targeted fragments.

Strategy co-electroporation and plasmid segregation were applied to overcome this problem. However, only one recombinant in 330 obtained transformants was screened. The recombination efficiency was too low and there were too many false positive recombinants deleting their toxic cassette spontaneously. Analysis indicated the large number of false positive recombinants would be caused during the capture plasmid preparation.
New strategy was figured out to decrease the culture time at each culture step and the capture plasmid was re-extracted from the culture (culture time was decrease from 10 hours to 7 hours and 30 minutes). After the improvement, promising results after co-electroporation and plasmid segregation experiments showed that most of false positive recombinants were eliminated. No recombinant (either false positive or positive) grew using co-electroporation strategy. When using plasmid segregation strategy, less false positive recombinants grew but two recombinants in 29 obtained transformants were screened. The positive recombinant ratio was improved from $1/330 \approx 0.3\%$ to $2/29 \approx 7\%$ (positive recombinant/total transformants) after decreasing the culture time for the Genefish application.

Although positive recombinants rose by controlling mole ratio between cells and capture plasmids, but the process was too complicated to manage. Three strategies (classical digestion and ligation, In-Fusion, two In-Fusion steps), two transformation protocols (electroporation and chemical transformation) were applied to construct the single-copy number capture plasmid. However, no expected single-copy number capture plasmid was constructed. The reason might be the low transformation efficiency of our host strain, the optimized efficiency got was $10^6 \text{cfu/\mu g}$; while the protocol of In-Fusion kit required a transformation efficiency at least $10^8 \text{cfu/\mu g}$.

Anyway, Genefish system was constructed, and up to 7% (2 positive recombinants in 29 total colonies) succeeds in capturing \textit{bcr c-a} sequences \textit{in vitro}. Positive recombinants ratio was increased by decreasing the culture time at each culture step, using co-electroporation and plasmid segregation. However, the homologous recombination efficiency is still too low ($10^9$); only 2 recombinants were screened using about $10^9$ competent cells. This recombinant’s capture efficiency was still too low to capture targeted fragments in environment samples.

Why the homologous recombination efficiency is so low? This was caused by the low copy number capture plasmid. Although we counted the mole ratio between host strain and plasmid, but the number of strain cell cannot be precisely defined. We used empirical value $10^9$ cells in 1ml (OD600=1) \textit{E.coli} culture for Genefish application. Hence, inaccurate number of host strain cells may probably affects the Genefish efficiency. However, series of mole ratio tests showed no significant difference. So, three reasons could be summarized here: 1) the plasmid-plasmid system for homologous recombination, 2) uncertain model of \textit{\lambda} Red homologous recombination, 3) inappropriate capture plasmid structure for recombination.
The λ Red homologous recombination system has been broadly used in bacterial genetic engineering, such as the chromosome modification (Mizoguchi, Tanaka-Masuda et al. 2007; Feher, Karcagi et al. 2008; Lesic 2008; Yamamoto, Izumiya et al. 2009), plasmid modification (Yosef, Bloushtain et al. 2004; Thomason, Costantino et al. 2007; Song, Dong et al. 2010), and bacterial artificial chromosome modification (Court, Swaminathan et al. 2003; Zhang and Huang 2003; Warming S 2005; Hollenback, Lyman et al. 2011). Two optimal λ Red system used for those modifications are summarized by Donald L.Court (Sharan, Thomason et al. 2009): For bacterial chromosome or BAC DNA modification, mobile recombineering systems like the pSIM vectors, mini-λ or the replication-defective λ phage (λTetR) can be introduced into the bacterial cells to be engineered. When high or low copy plasmids have to be manipulated, bacterial strains containing the prophage recombination system on the chromosome (e.g DY380 and its derivatives like SW102) should be used. However, in this study, the mobile recombineering system pSIM6 was applied to modify the low copy capture plasmid. This plasmid-plasmid mode (recombination on plasmid and was induced by mobile recombineering systems pSIM6) maybe the reason caused low recombination efficiency. The recombination efficiency using pSIM6 plasmid to modify bacterial chromosome has been validated by Laure Franqueville already whiles the plasmid-plasmid recombination efficiency in this study remained too low. In order to exclude the effect of low copy capture plasmid in our toxic cassette system selection, a target fragment containing antibiotic resistance operon (like tetracycline) could be used to test the plasmid-plasmid recombination efficiency. During the recombinants selection step, two rounds selection could be used to confirm whether the low recombination results were caused by the plasmid-plasmid system or by the low copy number of capture plasmid. The first round selection: add tetracycline in mediums (only recombinants containing the tet gene can survive) to select recombinants. This would reveal the homologous recombination efficiency by using “plasmid-plasmid system”. Second round selection: re-select recombinants that have been captured during the first round by IPTG, Ara and tet containing mediums. If we get high efficiency after first round selection and low efficiency after the second round selection, this means toxic cassette selection system in plasmid-plasmid recombination system is not suitable and the single-copy number capture plasmid will be the potential solution. If we get low efficiency in first round selection, this means “plasmid-plasmid system” maybe not suitable for
Although the $\lambda$ Red recombination system is commonly used, the detailed mechanism by which lambda Red mediates double-stranded DNA recombination remains uncertain. The general theory of $\lambda$ Red recombination occurs during the DNA replication. Because DNA polymerase is able to add free nucleotides only to the 3' end of the newly forming strand, this results in elongation of the new strand in a 5'-3' direction. The leading strand is the template strand for new complementary strand synthesis in 5' to 3' direction corresponding to the movement of the replication fork. The lagging strand is the complementary strand of leading strand in the original sequence, because the movement of replication fork is opposite to the working orientation of DNA polymerase, replication of the lagging strand is slower than that of the leading strand. On the lagging strand, primase "reads" the DNA and adds RNA to it in short, separated segments, RNA primers are removed by 5'-3' exonuclease activity of polymerase I and replaces the RNA nucleotides with DNA nucleotides (Nosikov, Braga et al. 1976). DNA ligase joins the fragments together at last, while the homologous recombination especially the 3’ overhang of target dsDNA will invade the complementary site on lagging strand.

So far, there are three $\lambda$ Red homologous recombination mechanisms which are already described in the introduction chapter. The Court mechanism (Fig.25) generates crossover junctions after the homologous recombination and this structure would be resolved by unspecified *E. coli* enzymes (Court DL 2002). Poteete et al (Poteete 2008) proposed another mechanism in which one recombinant DNA is generated but one copy of original chromosome or plasmid would remain (Fig.26). Recently, J.A.Mosberg et al (Mosberg, Lajoie et al. 2010) proposed a simpler $\lambda$ Red homologous recombination mechanism, in which dsDNA recombination mediated by Lambda Red proceeds via a ssDNA intermediate (Fig.27). However, one original copy of chromosome or plasmid still remains. So, if the $\lambda$ Red homologous recombination mechanism is exactly as Poteete or J.A.Mosberg said, original capture plasmid will be retained after recombination. After toxic cassette induction, transformants will be killed even if they recombined. This may be the critical reason of low homologous recombinants recovery. So, one copy of capture system (on the plasmid or chromosome) in each cell may solve the problem. But, construct a single copy plasmid containing capture plasmid in this study has been proved difficult (in chapter “Genefish improvements”).
An other hypothesis could be considered: capture plasmid used in this study is 6880 bp taking pBAD35K7toxN-bcrc-d for example (Fig.60). The two highly conserved sequences bcr c and bcr d are closed to the ORI region of the plasmid. The bcr c fragment is just separated in few base pairs from the plasmid replication ORI site which means the replication fork will pass and finish the replication of the bcr c region within a short time. This inappropriate site of highly conserved sequences could decrease the invasion opportunities of exogenous DNA thus close the window for homologous recombination. The bcr d fragment is located upstream (2 kb away) bcr c, if homologous recombination occurs, the region containing bcr c, ORI, cmR and bcr d will be replicated while the rest (more than 4 kb) will be replaced by targeted sequence. This recombined capture plasmid may not been successfully resolved because of the instable capture plasmid structure (the size of the plasmid is about 6.9 kb while 4 kb need to be replaced).

**Fig.60: structure of capture plasmid pBAD35K7toxN-bcrc-d.**
Red arrow shows the location of high conserved sequence bcr c and the region marked in brackets is the replication region during homologous recombination
5. Perspectives

Although we have encountered some problems using the Genefish tool to capture targeted \textit{bcr} fragments, but there are still some solutions. As mentioned during the discussion chapter, we can first test the plasmid-plasmid recombination efficiency using \textit{\lambda} Red system. In any case, transfer the region containing highly conserved sequences and the toxic cassette of capture plasmid on the bacterial chromosome would solve many problems that we have encountered, such as the plasmid-plasmid recombination system, copy number of plasmid, highly conserved sequences too close to the ORI etc. So, the next step is to fix the toxic cassette on an appropriate bacterial chromosome region and test for chromosome recombination.

The ultimate aim of Genefish tool is to capture targeted fragment from metagenomic DNA, while in this study, we still cannot get enough recombinants using pure PCR target fragments. Even if we can reach the recombination efficiency at \(10^{-4}\) using PCR products, we still need to pay attention on the accessibility of targeted fragment in metagenomic DNA. In metagenomic DNA, the targeted fragment may be diluted in million or even more times, and this could dramatically decrease the Genefish efficiency. During the metagenomic DNA extraction, genomes will be cut into 2-3 kb fragment randomly by bead beating and this will also decrease the percentage of optimized target fragment (two highly conserved sequences present at each end of fragment). Other conditions for Genefish application still need to be tested, such as the host strain improvements: knock out a set of exonucleases (RecJ, ExoI, ExoVII, ExoX, and Exo) genes which were already testified improving the homologous lambda Red recombineering (Mosberg, Gregg et al. 2012); recombination efficiency using different size of target fragments; optimize GC\% of targeted sequences; size of highly conserved sequences; and similarity level between donor and recipient DNA etc.
References:


in Chile." Fungal Biol 114(9): 746-752.


aromatic compounds under denitrifying conditions, and description of Thauera phenylacetica sp. nov., Thauera aminoaromaticasp. nov., and Azorarcus buckelii sp. nov." Arch Microbiol 178(1): 26-35.


Zhang, X. M. and J. D. Huang (2003). "Combination of overlapping bacterial artificial chromosomes by a
Zhang, X. Y. (2007). "Studies on quinoline biodegradation by a white rot fungus (Pleurotus ostreatus BP) in
liquid and solid state substrates." Fresenius Environmental Bulletin 16(8).
533-616.
Annex 1: metagenomic DNA extraction

Each aerobic and anaerobic microcosm was composed of 120 samples, and the protocols for DNA extraction are recorded as follow:

- Weight 0.3 g soil samples into 2ml lysing Matrix tubes, add 600µl extraction buffer (50% w/w solution of CTAB 10 % (w/w) and NaCl 0.7M, 50% (w/w) 240mM K$_2$HPO$_4$/KH$_2$PO$_4$ buffer) and 600µl Phenol: Chloroform: Isoamylalcohol (25:24:1Roth ref: A156.2) solution.
- Transfer those tubes into FastPrep-24 instrument (MP Biomedicals), Bead beating with a FastPrep speed of 5.5M for 30 seconds.
- Centrifuge at 4°C for 10 minutes at speed of 11,000 rpm.
- Transfer the supernatant into 2ml eppendorf tubes and add 500µl of Chloroform: Isoamylalcohol (24:1) solution. Vortex well and centrifuge them at 4°C for 5 minutes at speed of 11,000 rpm.
- Transfer the supernatant into 2ml eppendorf tubes, add 1/10 volume NaCl(5M) and 2 fold volume of 100% ethanol(-20°C). Then store tubes at 4°C overnight.
- Centrifuge at 4°C for 30 minutes at speed of 12,000 rpm, discard supernatant and add 100µl ethanol (70%).
- Centrifuge at 4°C for 5 minutes at speed of 12,000 rpm. Discard supernatant carefully and add 100µl ethanol (70%).
- Centrifuge at 4°C for 5 minutes at speed of 12,000 rpm. Discard supernatant carefully with pipette.
- Dry the DNA at room temperature for 15 minutes
- Add 50µl of ddH2O to dissolve the pellet
- Purification with illustra GFX PCR DNA and Gel Band purification kit.
- Add 500µl of capture buffer type 3, mix thoroughly and transfer the mix to assembled GFX Microspin Column and collection tube.
- Centrifuge for 30 seconds at speed of 16,000 rpm, discard flow through and add 500µl of wash buffer type 1
- Centrifuge for 30 seconds at speed of 16,000 rpm. Discard collection tube and transfer GFX Microspin Column to a 1.5ml eppendorf tube.
- Add 50µl of elution buffer type 6, centrifuge for 60 seconds at speed of 16,000 rpm.
Those extracted DNA can be stored at -20°C.

Concentrations of extracted DNA samples were quantified by Qubit® dsDNA BR Assay Kit.
Annex 2: protocols for RISA analysis on Agilent 2100 bioanalyzer

For RISA, Agilent DNA 1000 Kit was used. The relevant protocols are shown below:

1) Preparing the Gel-Dye Mix
   - Allow the DNA dye concentrate (blue) and DNA gel matrix (red) to equilibrate to room temperature for 30 minutes.
   - Vortex the blue-capped DNA dye concentrate (blue) for 10 seconds and spin down. Make sure the DMSO is completely thawed.
   - Pipette 25\textmu l of the blue capped dye concentrate (blue) into a red-capped DNA gel matrix vial (red). Store the dye concentrate at 4 °C in the dark again.
   - Cap the tube, vortex for 10 seconds. Visually inspect proper mixing of gel and dye.
   - Transfer the gel-dye mix to the top receptacle of a spin filter.
   - Place the spin filter in a microcentrifuge and spin for 15 minutes at room temperature at 2240 g ± 20 % (for Eppendorf microcentrifuge, this corresponds to 6000 rpm).
   - Discard the filter and put the mix at 4°C or for RISA use.

2) Setting up the Chip Priming Station
   - Replace the syringe:
     - Unscrew the old syringe from the lid of the chip priming station.
     - Release the old syringe from the clip. Discard the old syringe.
     - Remove the plastic cap of the new syringe and insert it into the clip.
     - Slide it into the hole of the luer lock adapter and screw it tightly to the chip priming station.
   - Adjust the base plate:
     - Open the chip priming station by pulling the latch.
     - Using a screwdriver, open the screw at the underside of the base plate.
     - Lift the base plate and insert it again in position C. Retighten the screw.
   - Adjust the syringe clip:
     - Release the lever of the clip and slide it down to the lowest position.

3) Setting up the Bioanalyzer
   - Adjust the chip selector:
Open the lid of the bioanalyzer and make sure that the electrode cartridge is inserted in the instrument. If not, open the latch, remove the pressure cartridge and insert the electrode cartridge.

Remove any remaining chip and adjust the chip selector to position (1).

4) Starting the 2100 Expert Software

- Go to desktop and double-click the following icon.
- If more than one instrument is connected to the PC, select the instrument we want to use in the tree view.

5) Loading the Gel-Dye Mix

- Allow the gel-dye mix to equilibrate to room temperature for 30 minutes before use. Protect the gel-dye mix from light during this time.
- Take a new DNA chip out of its sealed bag and place the chip on the chip priming station.
- Pipette 9.0μl of the gel-dye mix at the bottom of the well marked.
- Set the timer to 60 seconds, make sure that the plunger is positioned at 1 ml and then close the chip priming station. The lock of the latch will click when the Priming Station is closed correctly.
- Press the plunger of the syringe down until it is held by the clip.
- Wait for exactly 60 seconds and then release the plunger with the clip release mechanism.
- Visually inspect that the plunger moves back at least to the 0.3 ml mark.
- Wait for 5 seconds, and then slowly pull back the plunger to the 1 ml position.
- Open the chip priming station.
- Pipette 9.0 μl of the gel-dye mix in each of the wells marked.

6) Loading the Marker

- Pipette 5 μl of green-capped DNA marker (green) into the well marked with the ladder symbol and into each of the 12 sample wells.

7) Loading the Ladder and the Samples

- Pipette 1 μl of the yellow-capped DNA ladder (yellow) in the well marked with the ladder symbol.
- In each of the 12 sample wells pipette 1 μl of sample (used wells) or 1 μl of deionized water (unused wells).

8) Inserting a Chip in the Agilent 2100 Bioanalyzer
Open the lid of the Agilent 2100 bioanalyzer.

Check that the electrode cartridge is inserted properly and the chip selector is in position (1).
Refer to “Setting up the Bioanalyzer” on page 8 for details.

Place the chip carefully into the receptacle. The chip fits only one way.

Carefully close the lid. The electrodes in the cartridge fit into the wells of the chip.

The 2100 expert software screen shows that we have inserted a chip and closed the lid by displaying the chip icon at the top left of Instrument context.

9) Starting the Chip Run

In the Instrument context, select the appropriate assay from the Assay menu.

Accept the current File Prefix or modify it. Data will be saved automatically to a file with a name using the prefix we have just entered. At this time we can also customize the file storage location and the number of samples that will be analyzed.

Click the Start button in the upper right of the window to start the chip run. The incoming raw signals are displayed in the Instrument context.

To enter sample information like sample names and comments, select the Data File link that is highlighted in blue or go to the Assay context and select the Chip Summary tab. Complete the sample name table.

To review the raw signal trace, return to the Instrument context

After the chip run is finished, remove the chip from the receptacle of the bioanalyzer

10) Cleaning Electrodes after a DNA 1000 Chip Run.

Slowly fill one of the wells of the electrode cleaner with 350 µl deionized analysis-grade water.

Open the lid and place electrode cleaner in the Agilent 2100 bioanalyzer.

Close the lid and leave it closed for about 10 seconds.

Open the lid and remove the electrode cleaner.

Wait another 10 seconds to allow the water on the electrodes to evaporate before closing the lid.
Annex 3: Escape rate test

Escape rate tests are process in two steps: first selection in liquid induction media and second selection in solid induction medium plates. The pre-selection process called liquid activation. Liquid activation was a preliminary selection of potential colonies for escape rate test while escape rate test was used to calculate the real strain escaping rate from death. This rate means the ratio of colonies in a total 10ml culture escaping from death. The CFU were recorded on plates with different types of medium which were already used in the liquid activation test.

Target strains were streak cultivated on plates containing 20ml LB agar and 20µl chloramphenicol (CM: 50mg/ml) at 37°C. At least 24 colonies of each target strains were chosen for the liquid activation. Those colonies were cultured under four different mediums in 96 well plates: medium (1) LB broth medium with chloramphenicol (50µg/ml), medium (2) LB broth medium containing 1% arabinose (w/v), medium (3) LB broth medium containing IPTG (500µg/ml), medium (4) LB broth medium containing 1% arabinose (w/v) and IPTG (500µg/ml) respectively. These four kinds of medium all contain 50µg/ml chloramphenicol. The protocol for liquid activation is described below:

- LB broth induction media preparation: calculate the total volume of medium for the culture, each well we need 100µl of medium. 47 total colonies means 48 well are needed and one well is the blank control. So, 6ml of each kind medium is enough: 6ml LB broth medium with 6µl chloramphenicol(CM:50mg/ml); 5.4ml LB broth medium with 6µl chloramphenicol(CM: 50mg/ml) and 600µl of arabinose(10%,w/v); 6ml LB broth medium with 6µl chloramphenicol(CM: 50mg/ml) and 15µl IPTG(200mg/ml); 5.4ml LB broth medium with 6µl chloramphenicol(CM:50mg/ml), 15µl IPTG(200mg/ml) and 600µl of arabinose(10%,w/v).

- Starter culture of target colony is inoculated in CM (50µg/ml) containing LB broth medium. 100µl of medium is transferred into each well from A1-to-D12 of the 96 well plates. 47 colonies are inoculated into each well respectively and the last well (D12) is the blank control. Incubate this 96 well plate on a shaker at 37°C and at speed of 150 rpm for about 2 hours until turbidity appears in each well.

- During this time, prepare two other 96 well plates for the liquid activation. A1-D12 of the first plate containing 100µl LB broth medium (1) and E1-H12 containing 100µl LB broth
medium (2). A1-D12 of plate number 2 contains 100μl LB broth medium (3) and E1-H12 contains 100μl LB broth medium (4).

- Transfer the starter culture into the prepared 96 well plate 1 and 2 respectively. Make sure that each colony starter culture is accurately transferred into their corresponding wells. Incubate these two plates on a shaker at 37°C and at speed of 150 rpm for an overnight culture.
- Store the starter culture after 8 hours’ culture: add 100μl of 50% glycerol (25% final concentration) into each well of starter culture, mix gently and store the plate at -20°C.

Results of liquid activation: Fig 61 shows the 96 well plates used for the liquid activation. Plate one (left) shows that colonies in wells A1 to D11 grow very well while no bacteria grows in the blank control well (D12). However, there were bacteria growing in wells F8 and G8, this means these two strains have escaped from death induced by 1% arabinose. In the same way, well A1-A3, A6, A7, B2, B4, B7, B10 and C5 of plate two have escaped from death induced by 200μg/ml IPTG. No colony can survive under the induction of both 1% arabinose and 200μg/ml IPTG (E1-H11 of plate 2). So, for the further escape rate test, those colonies mentioned above cannot be chosen although they cannot survive under both toxic genes induction. At last, three colonies of each strain, colony pBAD35K7toxN-brcr-a-9,-11and -13(well B1) and pBAD35K7toxN-brcr-d-1,-3 and -7 were selected for escape rate test.

---

**Fig.61: liquid activation results of strain 14102, pBAD35K7toxN-brcr-a and pBAD35K7toxN-brcr-d.** Wells in yellow show growing strain; 1, 2, 3 and 4 means the type of culture medium used in those wells.

Second step of escape rate test for pBAD35K7toxN-brcr-a and pBAD35K7toxN-brcr-d: Three colonies of pBAD35K7toxN-brcr-a-9,-11,-13 and other three colonies of pBAD35K7toxN-brcr-d-1,-3,-7 were selected for escape rate test. Each of these 6 colonies was streak cultivated in plate containing 20ml LB agar and 20μl chloramphenicol (CM: 50mg/ml).

All plates were cultivated at 37°C for overnight, one single colony was picked up from each plate and cultivated in 50ml tube containing 15minl LB broth medium and 15μl chloramphenicol (CM: 50mg/ml). All tubes were fixed on a shaker at 37°C at speed of 200rpm for 7 hours culture.
LB agar medium preparation:

a) For the total colony number counting in culture: 50ml LB agar medium with 50µl chloramphenicol (CM: 50mg/ml), mix well then pour into three 10cm Petri plates.

b) For the colony number escaping from death (induced by arabinose): 27 ml LB agar medium with 30µl chloramphenicol (CM: 50mg/ml) and 3ml of filtered 10% arabinose (m/v). Mix well then pour into one 10cm Petri plates.

c) For the number of colony escaping from death (induced by IPTG): 50ml LB agar medium with 50µl chloramphenicol (CM: 50mg/ml) and 125µl IPTG (200mg/ml). Mix well then pour into three 10cm Petri plates.

d) For the number of colony escaping from death (induced both by IPTG and arabinose): 27 ml LB agar medium with 30µl chloramphenicol (CM: 50mg/ml), 3ml of filtered 10% arabinose (m/v) and 75µl IPTG (200mg/ml). Mix well then pour into two 15cm Petri plates.

Preparation of culture samples:

15minl culture was separated into two parts, one part contains 10ml culture for the escape rate test and the other part with 5ml culture is used for the strain storage. For the strain storage, protocol was described before and all these six strains were stored at -80°C.

Process for the escape rate test on 10ml culture: centrifuge at 4°C for 10min at speed of 4000g. Carefully discard the supernatant, add 800µl of 0.8 % (m/v) NaCl, mix well and transfer the solution into a DNase-free 1.5ml eppendorf tube. Adjust the volume of the solution to 1ml at last.

Serial 10-fold dilution of each solution: $10^{-6}$ dilution was used to calculate the total number of cells in original culture. For each diluted sample, 6 DNase-free 1.5ml eppendorf tubes were pre-filled with 900 µl of 0.8 %( m/v) NaCl. This solution was considered as the original dilution ($10^0$). At each step, 100 µl of the previous dilution is added to 900 µl of 0.8 %( m/v) NaCl.

For the total colony number, pour 100µl of dilution $10^{-4}$, $10^{-5}$, $10^{-6}$ of each strain into three 10cm Petri plates containing medium (1).

Escape rate when using arabinose, pour 450µl of dilution $10^0$, $10^{-1}$ of each strain into two 10cm Petri plates containing medium (2).

Escape rate when using IPTG, pour 100µl of dilution $10^{-1}$, $10^{-2}$, $10^{-3}$ of each strain into three 10cm Petri plates containing medium (3).
Escape rate when using IPTG and arabinose, pour 450µl of dilution $10^0$ of each strain into two 15cm Petri plates containing medium (4).

All plates were then transferred into incubator at 37°C for overnight culture.
Annex 4: Protocol for the co-electroporation

Lambda Red activation:

✓ Host strain (14102, *E.coli* with plasmid pSIM6) is streak cultivated in 20ml LB agar medium with 20μl carbenicillin (50mg/ml) at 29°C.

✓ Pick up a single colony and cultivate in 100ml LB broth medium with 100μl carbenicillin (50mg/ml) at 29°C at speed 200 rpm. Measure the OD600 from time to time.

✓ When the OD600 is between 0.4-0.6, warm up 50ml of culture into water bath at 42°C and the other half culture at 29°C at speed 100rpm for 15 minutes. This process at 42°C is used to activate the λ Red system and the culture at 29°C is taken as the negative control.

Competent cells preparation:

✓ The protocol for competent cells preparation is the same as before. At last, each 50ml culture is prepared at the end as 10 tubes of competent cells. This means each tube contains 4X10^9 cells.

✓ Calculation: calculate the quantity of capture plasmid needed for the recombination. The plasmid is 6880 bp long and the average molar mass of each base is 330, so the molar mass of capture plasmid pBAD35K7toxN-bcrc-d is 4540800. The formula to calculate the quantity of plasmid (ng): (Quantity (ng) x 10^{-9} x 6.02x10^{23})/4540800=4.0x10^{9}(number of competent cells). According to this, 30ng of plasmid for the co-electroporation was enough. From the same calculation 18ng of *bcr operon* give the same copy number between the plasmid and the host strain.

Electroporation:

✓ Positive control: electroporate plasmid pBAD35K7toxN-bcrc-d alone into the lambda red activated competent cells with a ratio of 1:1. With this positive control, the quantity of competent cells was estimated. For the homologous recombination test, electroporate both plasmid and target *bcr operon* into the lambda red activated competent cells with the ratio 1:1:5 and 1:1:10. The negative control: electroporate both plasmid and target *bcr operon* into the non-activated lambda red competent cells with the ratio 1:1:5 and 1:1:10.

Plate preparation

✓ Two mediums were prepared to test the electroporation efficiency. LB agar medium with kanamycin (final concentration 25μg/ml) for the total number of cells after the electroporation and LB agar medium with CM (final concentration 50μg/ml) to calculate the total number of
pBAD35K7toxN-bcrc-d transformed cell. For the homologous recombinants detection, the LB agar medium contains CM, arabinose, IPTG with a final concentration of 50\(\mu\)g/ml, 1 \%(m/m) and 500\(\mu\)g/ml respectively.

After the transformation, all plates were cultivated at 37°C for overnight culture until the colonies were visible.
Annex 5: Protocol for plasmid segregation

First time electroporation:

Culture of host strains *E.coli* 14102 pBAD35K7oxN-bcrc-d+pSIM6 on 20ml LB agar medium with 20μl CM(50mg/ml) and 20μl carbenicillin (CB 50mg/ml) at 29°C for about 30 hours.

Liquid culture of the host strain: Pick up a single colony and do a starter culture with 10ml LB broth medium which contains 10μl CM and 10μl CB at 29°C and at speed 200rpm. After overnight culture, transfer 1ml of starter culture into 50ml LB broth medium which contains 50μl CM and 50μl CB and culture at 29°C at speed 200rpm. Measure the OD600 from time to time after 3 hours of culture until reaches OD600 of 0.4-0.6.

The protocol for host strain preparation was the same as before with a 42°C, 15min induction. 50ml culture was prepared as competent cells into 10 tubes.

The mole ratio between host strain and target DNA fragment for the first round electroporation were: 1:0.5, 1:1, 1:2 and 1:4. Two tubes of competent cells were used for each ratio condition. After the electroporation, 1ml of transformed cells was cultivated at 37°C for one hour.

NucleoSpin® Plasmid kit (Macherey Nagel) was used for the plasmid extraction.

The extracted plasmid mixture was quantified by using Qubit® dsDNA BR Assay Kit (Invitrogen). Finally, the plasmid mixture has a concentration of 40ng/μl for each ratio.

Second round of electroporation:

Host strain *E.coli* 14102, was streaked cultivated in 20ml LB agar medium with 10μl Kn (50mg/ml) at 37°C.

Pick up a single colony and cultivated in 50ml LB broth medium with 25μl Kn (50mg/ml) at 37°C and at speed of 200rpm. 50ml of this culture was then prepared as competent cells into 10 tubes. The ratio between host strain and plasmid mixture for the second round of electroporation were 1:0.5 and 1:1. For the condition 1:0.5, 20ng of plasmid mixture were used. All transformants were screened using the selective mediums which were already described before.