Year : 2014

A SINGLE CELL APPROACH TO STUDYING ICECLC BEHAVIOUR IN PSEUDOMONAS

Reinhard Friedrich

UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Département de Microbiologie Fondamentale

A SINGLE CELL APPROACH TO STUDYING ICE*CLC* BEHAVIOUR IN PSEUDOMONAS

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Friedrich Fedor REINHARD

Master of Research de l'Université de Nottingham, UK

Jury

Prof. Christian Hardtke, président
Prof. Jan Roelof van der Meer, directeur de thèse
Prof. Philippe Moreillon, expert
Prof. Barth Smets, expert
Prof. John McKinney, expert

Lausanne 2014

2014

A SINGLE CELL APPROACH TO STUDYING ICE*CLC* BEHAVIOUR IN PSEUDOMONAS

Friedrich Fedor Reinhard

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président	Monsieur	Prof. Christian	**Hardtke**
Directeur de thèse	Monsieur	Prof. Jan Roelof	**Van der Meer**
Experts	Monsieur	Prof. Philippe	**Moreillon**
	Monsieur	Prof. Barth F.	**Smets**
	Monsieur	Prof. John	**McKinney**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Friedrich Fedor Reinhard

Master of Research Université de Nottingham, UK

intitulée

**A SINGLE CELL APPROACH TO STUDYING ICE*CLC* BEHAVIOUR
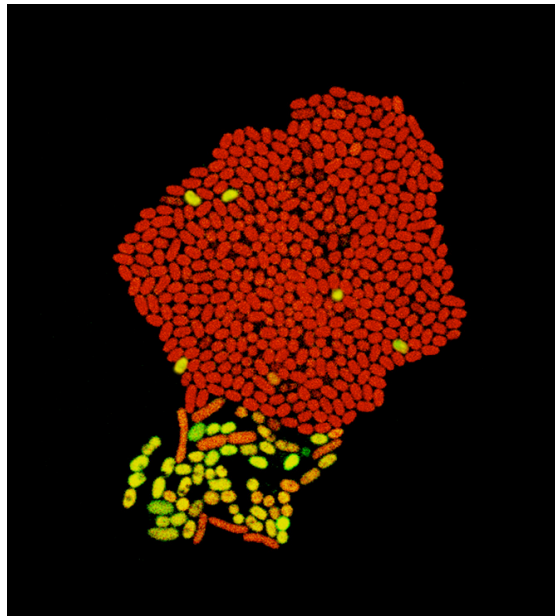IN PSEUDOMONAS**

Lausanne, le 6 janvier 2014

pour Le Doyen
de la Faculté de Biologie et de Médecine

Prof. Christian Hardtke

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

SUMMARY

Horizontal gene transfer (HGT) is one of the key processes in bacterial genome evolution. Bacterial genome evolution has attracted broad scientific interest for three main reasons. Firstly, most bacteria display rapid generation times and therefore represent good objects for studying molecular mechanisms leading to genetic variation. Secondly, ultrahigh throughput sequencing technology has led to an explosion in the number of bacterial genome sequences, the meta-analysis of which has revealed the important role of mobile DNA elements for genomic variation. And thirdly, evolutionary processes lead to the distribution of virulence factors, antibiotic or toxic compound resistance genes facilitating the fast generation of bacterial mutants compromising global public health.

In this dissertation HGT is studied with the aim to better understand how and why bacterial variants evolve so quickly, thereby contributing to the knowledge of how to counteract the downsides and make use of the benefits of adaptive processes. To achieve this aim this research is focused on the study of ICE*clc*, a model system representing key mediators in bacterial genome evolution called integrative and conjugative elements (ICEs). ICEs are widespread mobile genetic elements (MGEs) that occur in basically all bacterial genomes and are capable of transferring their own as well as additional DNA to new host species. ICE*clc* is an example of a prevalent ICE type that was originally detected in the chromosome of *Pseudomonas knackmussii* B13 but has since then been discovered in a wide variety of other bacterial species, underlining the success of this class of ICE to infiltrate different host species. The element bestows the host with the capacity to metabolize a number of unique aromatic carbon substrates such as 3-chlorobenzoate. It can transfer at population averaged frequencies of 1-5% per donor cell to a number of *Beta-* and *Gammaproteobacteria*.

Despite the findings of the existence of ICE, and despite insights in their genetic structure and variability, the auxiliary genes they carry, and the mechanisms that mediate ICE excision and transfer regulation, still little is known about their biology as "self-acting entities" and their interactions with the host cell. To shed light on these particular aspects of ICE behaviour, this thesis follows an innovative investigative approach that is based on single cell observation and analysis techniques. Indeed, since prokaryotic biology is still largely based on the idea of clonal populations of cells that all behave similarly, novel, single-cell studies are expected to reveal a hidden dimension in ICE biology.

In this thesis, a two-step strategy is chosen. First, investigative efforts are directed at the development and improvement of fluorescence microscopy-based observation and image

analysis techniques particularly suited to the study of ICE*clc* gene activities at the single cell level (Chapter 2 and Chapter 3). Equipped with this methodological know-how, the ICE*clc* lifestyle is then investigated at the single cell level (Chapters 4, Chapter 5).

Chapter 2. Accurately detecting subpopulation sizes in bimodal populations remains problematic. Yet subpopulation size represents a powerful parameter by which cellular heterogeneity under different environmental conditions can be compared. In Chapter 2 a simple, statistics-based method for the analysis of small subpopulation sizes is proposed for use in the free software environment R along with a step-by-step protocol for easy implementation. The approach is tested and validated on the test system of transfer competence in the bacterium *Pseudomonas*, which causes a very small proportion of cells in resting phase (3-5%) to develop competence for conjugative transfer of ICE*clc*. In addition, computational modelling is used to test the efficacy of the developed method. Results show that the proposed approaches help to minimize inconsistencies in subpopulation classification caused by manual threshold placements, thereby increasing sensitivity to detect subpopulation changes. Further it is shown that the proposed approaches allow for a standardization of subpopulation evaluation across different experimental set-ups.

Chapter 3. Chapter 3 deals with microcolony growth procedures, which have gained recent interest due to their usefulness in time-lapse imaging of single cell growth, cell aging, and biofilm studies. However, not many of such procedures have been described in sufficient technical detail to be easily reproduced. In Chapter 3, therefore, a simple step-by-step procedure is provided, which allows time-lapse imaging of bacterial cell division and cell tracking up to the stage of a mono-layered colony of a few hundred cells. The innovation of Chapter 3 lies in the design of an experimental set-up that permits the simultaneous tracking of single cells during microcolony growth whilst changing medium conditions. This proved crucial for further experiments on cell fate (Chapter 4) and cell age (Chapter 5) in single ICE*clc*-active cells in the later chapters.

Chapter 4. Transfer competence development has never been shown at the single cell level and neither has the question of cell fate in ICE transfer-competent cells ever been addressed. This is probably because of the generally low frequency of transfer competence in ICE and the lack of technological know-how to capture this. In Chapter 4 this thesis shows for the first time how individual bacterial donor cells can become competent for horizontal transfer of a mobile DNA (ICE*clc*), which surprisingly, both in the presence and absence of a non-ICE carrier recipient species, leads to strongly impaired reproductive capabilities of the ICE*clc* transfer competent (tc) donor cell and finally cell death. Results show that the ICE induces transfer competence in 3%–5% of cells in a population under non-growing conditions. Further,

it is shown that the factors determining tc cell fate are ICE*clc*-encoded. Mutations in ICE genes disrupting tc microcolony formation (although tc cells grow slower than non-tc cells and ultimately dye, they still retain some potential to divide although at a slower rate than non-tc cells) lead to 5-fold decreased ICE transfer rates. Based on the results of Chapter 4 it is hypothesized that the tc state may actually have been selected for because it allowed more efficient transfer of ICE*clc*, whereas the large fitness cost the ICE*clc* reproductive inhibition program would impose on the population is only kept at bay by confining ICE*clc* activity to a small proportion of cells. Indeed, computer simulations show that low proportions of cell death at non-growing phase pose insignificant fitness disadvantage due to genetic drift and may therefore not have been selected against. Given the low transfer frequencies of most ICE, the regulation by subpopulation differentiation as described in this chapter might be widespread.

Chapter 5. Apparent stochasticity (e.g., the distribution of RpoS among cells) may have underlying causes, such as, some cells being in a slightly different growth phase or having experienced more previous biochemical damage, and thus producing more RpoS. The specific hypotheses this thesis tests in Chapter 5 are whether ICE*clc* activation in individual cells are dependent on cell pole age, whether it is confined to specific cellular lineages, or if it occurs as a consequence of pre-existing detectable biochemical damage. It is also tested whether tc cell formation occurs spatially at random within a microcolony. Through the use of time-lapse microscopy as proposed in Chapter 3, the life history of tc-cells within microcolonies is followed. Results show that the age of the cell pole is unlikely to play a role in the decision of its fate to initiate the ICE*clc* tc program. Furthermore, it is shown that initiation of transfer competence is not the result of the physiological state of ancestor cells, or of a close relative (sister cell). In contrast, tc-cells show higher levels of reactive oxygen species and membrane damage than non-tc cells, but whether cause for or effect of ICE*clc* activation could not be discerned. It is shown that ICE*clc* activation occurs spatially randomly in a microcolony, which might additionally be important for maximizing the chances in a biofilm to contact potential recipients.

The principle contributions of this thesis lie in the establishment of novel analytical and observational solutions to single cell bacterial research, and their usage to demonstrate existing hypotheses of ICE*clc* activation and transfer competence at the single cell level. But perhaps most interestingly, this thesis leads to the unexpected discovery and characterization of a novel phenotype corresponding to a severe fitness cost in ICE transfer competent cells. This discovery undoubtedly poses many new exciting questions concerning the role of mobile DNA in manipulating host cell behavior and differentiation in order to optimize horizontal transmission.

RÉSUMÉ

Le transfert horizontal de gènes (THG) est l'un des processus clefs de l'évolution des génomes bactériens. L'intérêt de l'étude de l'évolution des génomes bactériens provient de trois raisons principales. Premièrement, la plupart des bactéries présentent un temps de génération rapide. Elles sont donc de bons objets pour l'étude des mécanismes moléculaires conduisant à la variation génétique. Deuxièmement, les nouvelles technologies de séquençage ont conduit à une explosion du nombre de séquences de génomes bactériens, dont la méta-analyse a révélé le rôle important des éléments d'ADN mobiles dans la variation génomique. Troisièmement, les processus d'évolution mènent à la distribution de facteurs de virulence, d'antibiotiques et de gènes de résistance à des composés toxiques, facilitant la génération rapide de mutants pouvant avoir un effet sur la santé publique.

Dans cette dissertation, le THG est étudié dans le but de mieux comprendre comment et pourquoi ces mutants sont formés si rapidement, donc dans le but de contribuer à comprendre comment empêcher les effets néfastes de ces processus adaptatifs. Pour réaliser cet objectif, la recherche a été axée sur l'étude d'un système appelé élément conjugatif et intégratif (ICE en anglais) et jouant un rôle majeur dans l'évolution de génomes bactériens, en s'intéressant ici à l'ICE*clc* comme modèle. Les ICEs sont des éléments génétiques mobiles très répandus, retrouvés dans (quasiment) tous les génomes bactériens et capables de se transférer, mais aussi de mobiliser d'autres éléments d'ADN à une cellule receveuse. ICE*clc* est un modèle d'étude d'ICEs qui a été originellement détecté dans le chromosome de *Pseudomonas knackmussii* B13, mais qui a depuis lors été découvert dans une grande variété d'espèces bactériennes, montrant sa capacité à s'intégrer avec succès dans différentes espèces hôtes. Cet élément confère à l'hôte la capacité de métaboliser différents composés aromatiques comme le 3-chlorobenzoate. Il peut transférer avec une fréquence de 1-5% par cellule donneuse à de nombreuses *Beta-* et *Gammaproteobacteria*.

Malgré la découverte de l'existence d'ICEs et les études de leurs structures génétiques et leur variabilité, des gènes auxiliaires qu'ils portent et des mécanismes qui permettent la régulation de l'excision et du transfert, peu de choses est connu sur leur biologie en temps qu'entité à part entière et leurs interactions avec la cellule hôte. Pour élucider ces aspects particuliers de l'ICE, cette thèse présente une approche innovante d'investigation basée sur l'observation à l'échelle de la cellule unique et des techniques d'analyse. En effet, puisque la biologie procaryotique est encore largement basée sur l'idée d'une population

clonale de cellules agissant de la même façon, l'étude à l'échelle de la cellule unique pourrait permettre de révéler une nouvelle dimension à la biologie des ICEs.

Dans cette thèse, une stratégie à deux étapes a été choisie. Dans les deux premiers chapitres de recherche (Chapitre 2 et Chapitre 3), les efforts sont axés sur le développement et l'amélioration d'observations en fluorescence par microscopie, et sur les techniques d'analyse d'images permettant l'étude de l'activité de gènes de l'ICE*clc* à l'échelle de la cellule unique. Une fois l'approche méthodologique établie, le style de vie de l'ICE*clc* est ensuite étudié à l'échelle de la cellule unique (Chapitre 4, Chapitre 5).

La détection de sous-populations dans une population bimodale reste problématique mais elle représente une puissante possibilité pour comparer l'hétérogénéité cellulaire dans différentes conditions environnementales. Dans le Chapitre 2, une méthode statistique simple pour l'analyse de petites sous-populations est proposée, en présentant un protocole étape par étape pour une mise en place facile et une utilisation dans le logiciel R. Cette approche a été testée et validée en testant la compétence de transfert chez *Pseudomonas,* qui présente en phase stationnaire une très petite sous-population (3-5%) de cellules capable de développer une compétence pour le transfert conjugatif de l'ICE*clc*. De plus, la modélisation est utilisée pour tester l'efficacité de la méthode développée. Les résultats montrent que les approches menées aident à minimiser les erreurs de classification des sous-populations causées par un placement manuel du seuil. Ceci permet d'augmenter la sensibilité de détection de changement de sous-populations. Enfin, ces approches permettent une standardisation de l'évaluation des sous-populations entre les expériences.

Le Chapitre 3 s'intéresse à la croissance en microcolonie, sujet présentant un gain d'intérêt récent pour sa possibilité de cribler rapidement et simultanément un large nombre de bactéries environnementales, de suivre la croissance en temps réel à l'échelle de la cellule unique, de suivre la sénescence bactérienne ou d'étudier les biofilms. Malheureusement, peu de ces approches ont été décrites avec suffisamment de détails techniques pour être reproduites facilement. Dans le Chapitre 3, une procédure simple étape par étape est présentée, qui permet l'imagerie en temps réelle de la division cellulaire et le suivi de cellules jusqu'à la formation de colonies en mono-couches de quelques centaines de cellules. L'innovation dans ce Chapitre 3 provient de la mise au point expérimentale, permettant le suivi de cellules uniques durant la croissance en microcolonie lors de changement de conditions de milieux. Ces innovations se sont révélées cruciales pour les expériences sur la destinée cellulaire (Chapitre 4) et l'âge des cellules (Chapitre 5) dans des cellules ICE*clc*-actives présentées dans les chapitres suivants.

Le développement des cellules « tranfert-compétentes » (tc) n'avait jamais été montré à l'échelle de la cellule unique et la question de la destinée cellulaire des cellules tc n'avait jamais été adressée. Ceci est surement dû au fait que la fréquence de cellules tc est faible et à l'absence de moyen techniques pour visualiser ce développement. Dans le Chapitre 4, cette thèse montre pour la première fois comment des cellules donneuses individuelles peuvent devenir compétentes pour le transfert horizontal d'ADN mobile (ICE*clc*) mais dont le développement conduit à une capacité reproductive fortement affectée allant jusqu'à la mort cellulaire. Ce processus est maintenu en présence comme en absence de receveuse ne présentant pas d'ICE. Les résultats montrent que l'ICE induit la transfert-compétence dans 3-5% des cellules d'une population en absence de croissance. De plus, il a été montré que les facteurs déterminant le devenir des cellules tc soit codé par l'ICE*clc*. Des mutations dans des gènes interrompant la formation de microcolonies de cellules tc (En effet, bien que les cellules tc croissent lentement et finissent par mourir, elles gardent une certaine possibilité de se diviser, avec une vitesse plus lente que les cellules non-tc) conduisent à une diminution par 5 du taux de transfert de l'ICE. En se basant sur les résultats du Chapitre 4, il est supposé que l'état de transfert-compétence a été sélectionné parce qu'il permet un meilleur transfert de l'ICE*clc*, tandis que le coût de fitness important (l'inhibition de reproduction) imposé par l'ICE*clc* à la population est contrôlé par le fait que l'ICE*clc* est actif seulement dans une petite proportion de cellules. En effet, des simulations *in silico* montrent qu'une faible proportion de mort cellulaire en absence de croissance ne pose qu'un désavantage de fitness insignifiant, à cause du drift génétique, assurant un passage au travers de la sélection naturelle. Etant donné la faible fréquence de transfert dans la plupart des ICEs, la régulation par différenciation en sous-population expliquée dans ce chapitre pourrait être courante.

La stochasticité apparente (par exemple la distribution de RpoS entre les cellules) peut avoir des causes sous-jacentes, comme des cellules présentant une légère différence de phase de croissance ou ayant expérimentées plus de dommages biochimiques précédemment, et donc produisant plus de RpoS. Les hypothèses que cette thèse teste dans le Chapitre 5 sont : Est-ce que l'activation de l'ICE*clc* dans des cellules individuelles est dépendante de l'âge de la cellule (du pôle cellulaire) ; est-ce que c'est confiné à une lignée cellulaire spécifique ; est-ce qu'elle est la conséquence de dommages biochimiques pré-existants. Il est également testé si le développement en cellules tc apparaît spatialement au hasard au sein de la microcolonie. En utilisant la microscopie en temps réel proposé en Chapitre 3, l'histoire de la vie des cellules tc au sein des microcolonies est suivi. Les résultats montrent que l'âge du pôle cellulaire ne semble pas jouer de rôle dans la décision d'initier le développement en tc. De plus, il est montré que l'initiation en tc n'est pas le résultat de l'état

physiologique des ascendants d'une cellule, ni des cellules proches (cellules sœurs). Au contraire, les cellules tc montrent des quantités de dérivés réactifs de l'oxygène et des dommages membranaires plus important que les cellules non-tc, sans qu'il n'ait pu être démontré que c'est la cause de l'activation de l'ICE*clc*. Il a été montré que l'activation de l'ICE*clc* a lieu au hasard au sein de la microcolonie, ce qui peut être important pour maximiser les chances d'être en contact avec des potentielles cellules receveuses au sein d'une matrice.

Les contributions principales de cette thèse sont l'établissement de solution pour l'observation et l'analyse en recherche bactérienne à l'échelle de la cellule unique, et son usage pour démontrer les hypothèses existantes de l'activation de l'ICEclc et de la compétence de transfert à l'échelle de la cellule unique. Peut-être plus intéressant encore, cette thèse conduit à des découvertes insoupçonnées et à la caractérisation d'un nouveau phénotype correspondant à un coût de fitness important pour les cellules transfert-compétentes. Cette découverte pose à n'en pas douter de nombreuses questions concernant le rôle den l'ADN mobile dans la manipulation du comportement de la cellule hôte et sa différentiation pour optimiser le transfert horizontal.

CHAPTER 1

General introduction

BACTERIAL BIODIVERSITY

Bacteria have existed for more than three billion years and represent the most ancient forms of life on earth (56). Bacteria also likely represent the most abundant and diverse life forms in the Earth's ecosystem (7). Collectively they constitute the bulk of biomass on earth (128). Their phylogenetic and metabolic diversity greatly exceeds that of plants and animals (7, 128). For example, there are more than 800,000 insect species of which each individual harbours millions to billions of bacteria (7). Assuming that 10% of all insect species harbour distinct microbial symbiont species (7, 46, 75), this aspect alone would increase the number of extant bacterial species by several orders of magnitude. As another illustrative example of bacterial diversity and abundance, Torsvik and co-workers (198) have estimated that a handful of soil from the top 10 cm of a beech forest contains billions of microbial organisms with more than 10,000 different bacterial strains. Further, on average, one cm^2 of leaf surface is estimated to contain more than 10 million bacteria (119). Other environments teeming with microbial life include the human body (gut, mouth, skin) (202, 203), plant roots (126), hypersaline lakes (88), thermal vents (158), hot springs (204, 217), arctic environments (31), soil crusts (44), deserts (189), glacial ice (182), air-borne, intercontinental dust (64), and ultra-high-pressure rocks and fluids of the deep subsurface (157, 226). The fact that microorganisms have occupied nearly every imaginable niche on Earth indeed reflects the enormous evolutionary potential of these organisms and their capacity to diversify and adapt to new environments and persist in the face of environmental insults.

As opposed to animal diversity, bacterial diversity is not well reflected in the morphology as can be readily visualized by eye (Figure 1). Rather, bacterial diversity is reflected in their biochemistry, which comprises a plethora of cellular metabolic pathways depending on the environment the bacteria are residing in. Bacterial biochemical activity often results in a geochemical transformation of the environment, either as a consequence of chemical reactions for energy requirements and respiration, or as a consequence of molecule release such as antibiotics, toxins, signalling molecules, or extracellular polymeric substances (proteins, nucleic acids and lipids). A classic example for bacterial geochemical transformation of the environment is one of today's most important biotechnological processes, the purification of wastewater. Microbial communities live here as biofilm flocks in activated sludge, where they transform sewage waste water into "clean" water (organic matter-reduced water) by efficiently transforming soluble organic waste into bacterial biomass (7).
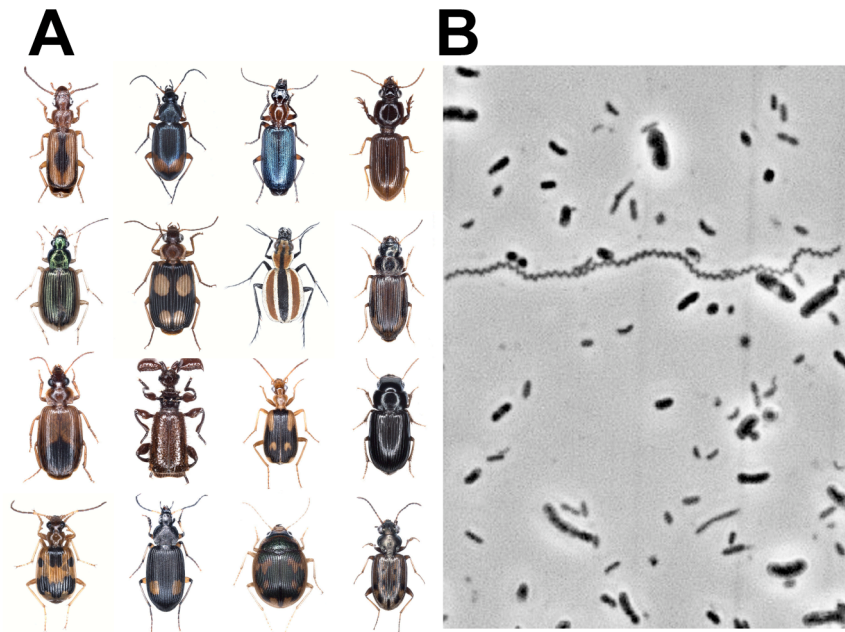
Figure 1. Examples of biodiversity in insects and bacteria. (A) Photographs of species of *Carabidae* (ground beetles) collected in the Kalahari semi-desert, BRinK field station, Kuzikus Wildlife Reserve, Namibia. Insects are typically classified on a visual basis by comparison of morphologies. [Adapted and modified with permission from Johanna Reinhard and David Schimrosczyk.] (B) Phase contrast microscopy image (400 x magnification) of a sample of activated sludge from a wastewater treatment plant in Lausanne, Switzerland. Only several simple morphologies of microbes can be distinguished visually in the activated sludge sample, yet several thousand different bacterial strains can be expected in such a sample (7). [Reinhard (unpublished).]

Hence, an understanding of how microorganisms attain metabolic diversity not only represents a highly interesting fundamental evolutionary research question, but should also bring knowledge towards ameliorating human and environmental health. For example, concrete answers to how bacteria develop resistance to antibiotics, how they become pathogenic, or how they rapidly adapt to degrade xenobiotic pollutants in the soil, might allow for targeted manipulation of these mechanisms to either inhibit (e.g. development of antibiotic resistance) or enhance (e.g. degradation of soil pollutants) microbial activity to suit human purposes (92, 205).

This thesis hopes to contribute to such promising cause by studying the integrative and conjugative element ICE*clc*. Studying ICE*clc* not only allows insight into one of the most important drivers of bacterial adaptation, the evolution of bacteria via horizontal gene transfer (HGT), but also represents an opportunity to investigate another mechanism of adaptation that has only recently, facilitated by the development of advanced technology of single cell observation and analysis, been revisited: Bacterial individuality. Thus, this introduction will concentrate on both ICE*clc*'s role in HGT, and bacterial individuality. In addition, a section is dedicated to single cell investigative methods, which have been crucial to the discoveries in this thesis.

HORIZONTAL GENE TRANSFER

Horizontal gene transfer. One of the most interesting questions in evolutionary microbiology is how bacteria can so quickly evolve unique features of their proteins, membranes, and genetic responses allowing them to survive sometimes rapid environmental change. Bacteria are great evolvers for several reasons. One good reason is that their generally short generation times and large population sizes allow for relatively quick rates at which *de novo* beneficial mutations may accumulate via subsequent rounds of natural selection and vertical transmission (DNA transfer to daughter cells via DNA replication and cell division) (76, 104, 153).

However, an even quicker way for bacteria to evolve is via HGT, the horizontal transfer of genetic information between different organisms (Figures 2, 3A) (76, 104, 153, 183). Estimates hold that some 20% of an average bacterial genome are of 'foreign' origin, whereas up to 80% of all bacterial and archaeal genes may have been engaged at some point during evolutionary time in horizontal transfer (101, 105). The consequence of this abundant gene exchange is that the taxonomic relationships between lineages (or evolutionary history) can almost be better described by a network than by a Darwinian phylogenetic tree (Figure 2) (183).

Importantly, HGT allows for the acquisition of pre-selected gene constellations where whole "blocks" of genetic information are acquired in single gene transfer events. The block-wise transfer of genetic information has two important outcomes: Firstly, it allows for the rapid community distribution of existing adaptive traits such as antibiotic resistance (20, 143), xenobiotic compound degradation (67, 69), pathogenicity and virulence (109), symbiosis, or heavy metal resistance (109). Secondly, HGT may fast-forward evolution of novel adaptive traits via so called "patchwork assembly", which is the bringing together of complementary, and pre-shaped (by natural selection) genetic building blocks (operons) from different microbial origins in the recipient cell resulting in novel biochemical pathways (80, 92, 141, 186, 205).

To illustrate evolution by HGT consider the two hypothetical real-case scenarios in Box 1. With regard to the first scenario, HGT is indeed notoriously known for its implication in the evolution of so-called "superbugs", which is the result of indiscriminate and inappropriate use of antibiotics in the treatments of patients and in the food industry (5, 8, 92, 135, 148, 183). Superbugs is a common term in the press and scientific literature for multi-drug resistant bacterial strains such as *Streptococcus pneumoniae*, *Staphylococcus aureus* and *Pseudomonas aeruginosa* that cause many lethal infections each year and whose spread is now considered a serious and increasing global public health risk (42, 183).

5

Box 1. Two scenarios of horizontal gene transfer.

Scenario 1:

A farmer feeds low doses of antibiotics to life-stock to reduce infection and enhance growth, thereby increasing economic return (181). The constant exposure of the local microbial community to antibiotics results in de novo evolution of antibiotic-resistant populations via successive rounds of random genetic mutations (8). The farmer visits a hospital but does not wash his hands or work-clothes prior doing so and thus facilitates cell-to-cell contact between hitch-hiking "farm bacteria" and locally adapted "hospital bacteria" (94, 102). Via HGT, the bacterial interaction between farm and hospital strains results in the exchange of genetic antimicrobial resistance traits. As a result, the progenitor of a potentially successful strain with novel but pre-selected antimicrobial resistance traits may appear within minutes, multi-resistant to both antibiotics as used on the farm as well as in the hospital (102). Selection pressure and the negative impact of the newly acquired genes on the host cell will determine if this strain will remain or perish in the hospital environment.

Scenario 2:

Toxic runoff at a military base contaminates a groundwater aquifer. The runoff contains chlorobenzene, a xenobiotic compound that is believed to have been alien to existing enzyme systems before its environmental release (186). Initially, chlorobenzene remains in the aquifer at high concentration but suddenly is degraded in a relatively short period of time (205). What happened? Facilitated by HGT, different gene-clusters from different bacterial origins recombined in a single host, which, as a result, was enabled to degrade chlorobenzene using it as a carbon source (205). Since the recombinant host gained a selective advantage due to its newly acquired catabolic potency, it colonized most of the chlorobenzene-contaminated area, thereby lowering the chlorobenzene concentration at the site.

The second scenario of Box 1 highlights HGT as a beneficial rather than a destructive quality from the human or environmental health perspective. It illustrates how bioremediation of contaminated groundwater or wastewater may be enhanced when biodegradative genes can be readily transferred to endogenous bacteria (18, 183, 195, 205).

Both scenarios, and the evidence associated to these, are examples showing that HGT represents a central process in the microbial world with implications for human health and environmental quality. A better understanding of the mechanisms and the regulation of HGT therefore will contribute to our knowledge towards a better control of the spread of antibiotic resistance and biodegradation genes (92, 183, 186).

What is known so far is that, although HGT can be achieved in several ways and involve the transmission of many different types of genetic substrates, all known mechanisms of HGT have the following steps in common (Figure 2) (183, 196):

1. An activation of the transfer process;
2. The selection of the recipient;
3. The uptake and successful entry of genetic material into the recipient;
4. The short-term establishment of genetic material in the recipient cell;
5. The long-term stable inheritance in the recipient cell.

Facilitating these steps are three main mechanisms of HGT:

1. Transformation, the uptake of DNA from the environment;
2. Transduction, the DNA transfer from one bacterium to another via bacterial viruses;
3. Conjugation, the transfer via cell-to-cell contact, now known to shuttle many different genetic fragments including plasmids, transposons, integrons, and integrative and conjugative elements (ICEs).

An additional potential transfer mechanism is based on nanotubes. Nanotubes are only recently discovered tubular protrusions composed of membrane components that can bridge between neighbouring cells and conduct the transfer of DNA and proteins (57). However, the abundance and relative importance of nanotubes in bacterial HGT needs still to be validated.

7
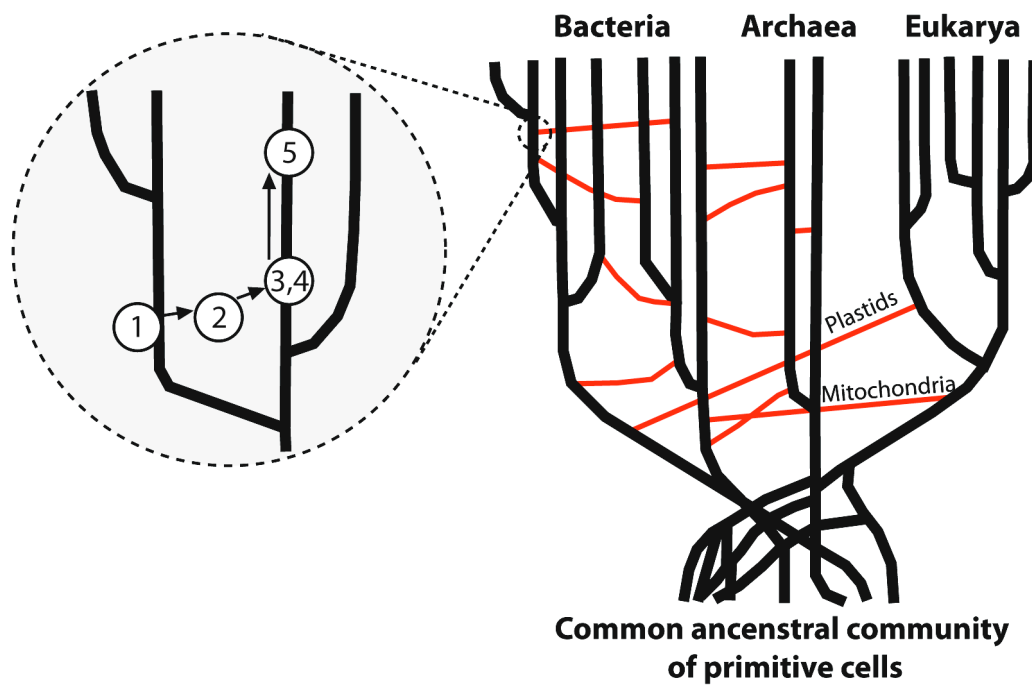
Figure 2. Different steps of horizontal gene transfer. Contributions of HGT (red lines) to the genome composition of different types of organisms is presented through a web connecting bifurcating branches that complicate, yet do not erase, the tree of life. Note that HGT can occur between closely related, but also distantly related organisms. The inset illustrates five steps that lead to the stable inheritance of a transferred gene in a new host: First, activation of the transfer process, second, selection of the recipient, third, uptake and successful entry of genetic material into the recipient, fourth, short-term establishment of genetic material in the recipient cell, and fifth, long-term stable inheritance in the recipient cell. [Adapted and modified from Smets and Barkay (183).]

Mobile genetic elements. In an attempt to understand the different mechanisms of HGT in bacterial genome evolution and to assess their respective impact on genome plasticity, a classification of the different DNA regions affected by HGT is valuable. Such classification is typically based on DNA sequence motifs. For example mobile genetic element (MGE) identification relies both on phylogenetic analysis of gene sequences to reveal topological inconsistencies between different gene families (97, 104), and on nucleotide compositional analysis to identify any gene that has a nucleotide pattern that differs significantly from the overall genome (for example GC content, cumulative GC skew, tetranucleotide frequencies or codon usage) (91) (Figure 3B). Applying both sequence-based analytical approaches on a rapidly increasing number of sequenced genomes (as of today, more than 2590 bacterial genomes have been sequenced, more than 190 eukaryotic genomes, and more than 3600 virus genomes (86)) creates a picture with two interesting facts:

Firstly, MGEs appear to be dominant features of most prokaryotic and eukaryotic genomes (97); for example, MGEs constitute 35% of the genome of *Escherichia coli* (216), 50% of the human genome (87), and about 80% of the maize genome (218).

Secondly, MGE structure, size and evolutionary origin seem to be quite diverse, and therefore can be categorized into different types of MGEs including both mobile and non-mobile elements. Non-mobile or "defective" MGEs are thought to be the result of having lost the mobility-genes over evolutionary time and thus represent a remnant marker of DNA pointing to past HGT activity. Mobile MGEs on the other hand represent active facilitators of ongoing HGT events. In total, mobile and defective MGEs span a variety of sometimes phylogenetically unrelated elements such as bacteriophages, IS elements, plasmids, transposons, conjugative transposons, integrons, and genomic islands (GEIs) (Figure 4) (29). It is now well established that MGE activity adds variability to genomes by one or several of the following ways (55, 79, 97):

1. By enhancing the potential for gene transfer between organisms via providing highly efficient mechanisms for the passage of genes into a recipient cell (for example via conjugative transfer);
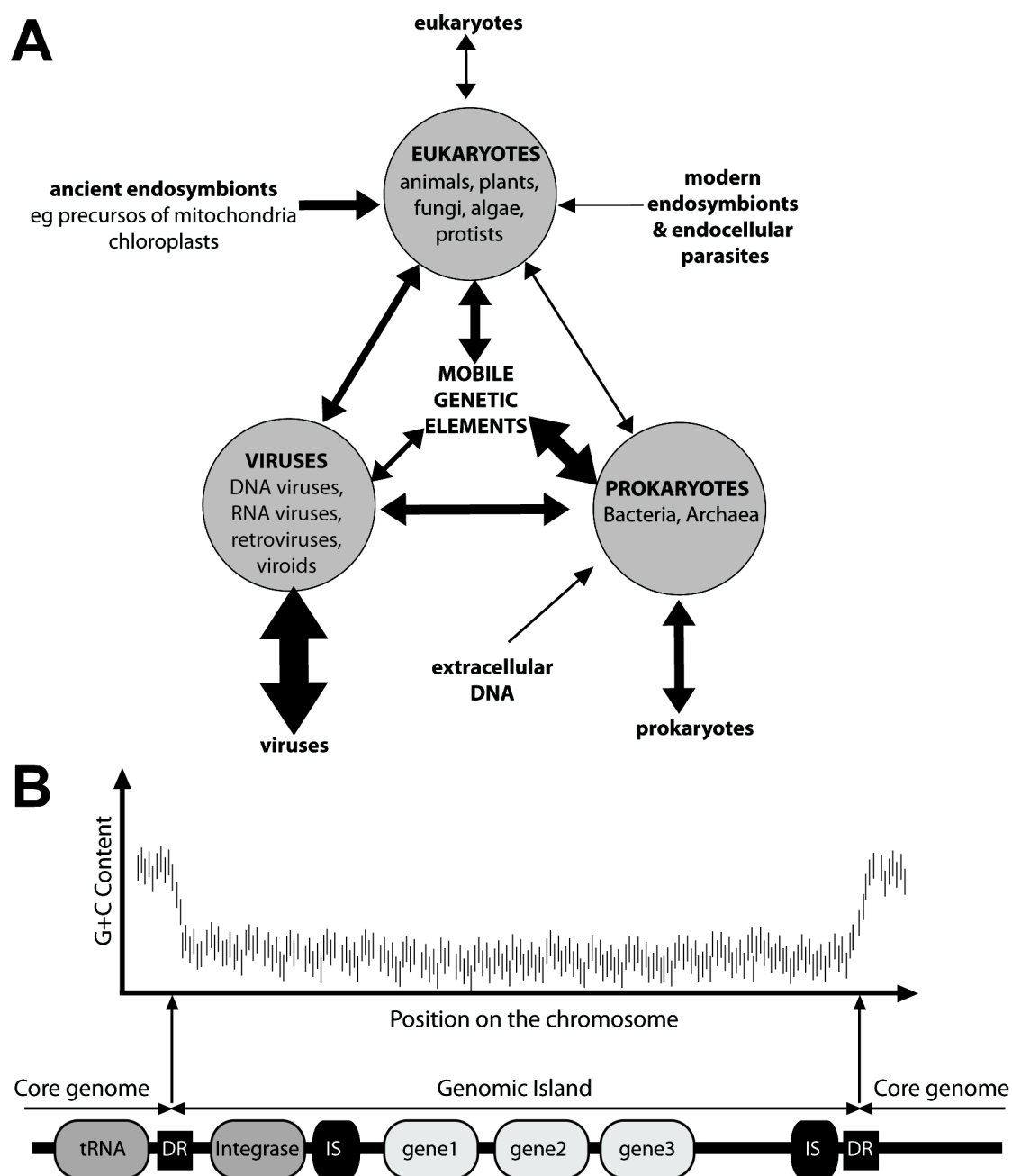
Figure 3. Impact of HGT and general characteristics of GEIs. (A) Contributions from HGT to the genome composition of different types of organisms. The thickness of each arrow indicates the predicted relative impact of each contribution. [Adapted and modified from Keese (97).] (B) GEIs are horizontally acquired, relatively large segments. GEI nucleotide characteristics (e.g. GC content) often differ from the rest due to their foreign origin. GEIs are often inserted at tRNA genes and flanked by direct repeats (DR). GEIs typically harbour genes encoding factors involved in genetic mobility, such as integrases, transposases and insertion sequences (IS). In addition GEIs usually carry several genes encoding traits that may increase bacterial adaptability or fitness under certain growth conditions. Depending on the encoded traits GEIs are named pathogenicity, symbiosis, metabolic, fitness or resistance islands. [Adapted and modified from Juhas *et al.* (92).]

2. By altering the function of genes in the vicinity of the insertion in the host genome via disruption or inactivation of genes at the site of insertion;

3. By contributing novel structural and functional genetic material;

4. By acting as primary vehicle for the spread and patchwork-assembly of antibiotic-resistance genes, for pathogenicity determinants, or for biodegradation pathways amongst bacteria.

GEIs. GEIs are a prominent subcategory of mobile and non-mobile MGEs and include a rather diverse group of elements called ICEs, integrated plasmids, non-replicative but excisable elements, and cryptic or damaged prophages. The defining features of GEIs include the following characteristics (Figure 3B) (55, 92):

1. They represent relatively large segments of DNA, usually between 10 and 200 kb;

2. They are often but not always inserted at tRNA genes;

3. They are flanked by 16–20-bp direct repeats (DR), suggesting a site-specific mode of integration of GEIs into the target site (174);

4. They harbour genes encoding integrases or factors related to plasmid conjugation systems or phages involved in GEI transfer;

5. They carry insertion elements or transposons, which may have been implicated in mobilizing genetic material onto or deleting DNA from the element;

6. They carry genes offering a selective advantage for host bacteria such as the provision of antibiotic resistance determinants, symbiosis factors, or metabolic pathways of aromatic compounds. Depending on the functions they provide to the host, GEIs are often described as pathogenicity, symbiosis, metabolic, fitness or resistance islands (Table 1) (55, 174).

Due to their capacity of shuttling host-beneficial genes and operons between species, GEIs allow for "fast forward evolution" via "patchwork assembly", or in other words, the import of pre-evolved and functional pathways from different chromosomal sources (55). GEIs must therefore be considered instrumental in shaping the bacterial genome. For example, GEIs may confer several traits to the host with potential adaptive benefits, such as antibiotic resistance (19, 143), iron scavenging (107), plant symbiosis (193) or aromatic compound degradation (71, 201).
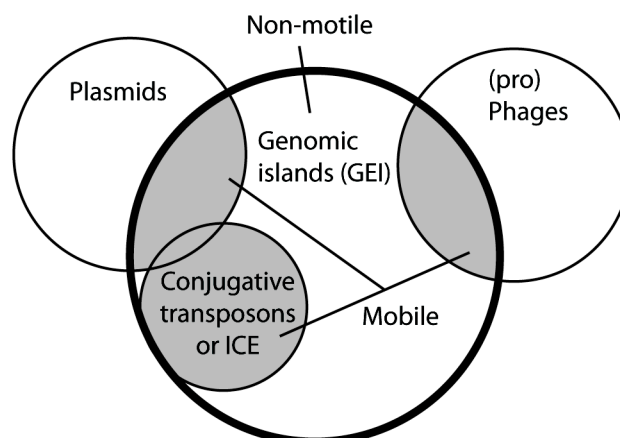
Figure 4. Variable types of GEIs. GEIs come in a large spectrum of varieties and encompass other categories of elements, such as conjugative transposons/ICEs, integrated plasmids, nonreplicative but excisable elements, and cryptic or damaged prophages. Grey-shaded areas point to self-mobile GEIs. [Adapted from Juhas *et al.* (92).]

Table 1. Key characteristics of experimentally described ICEs. [Adapted and modified from Wozniak and Waldor (223).]

ICE	Host	Size (kb)	Site of insertion	Notable phenotypes
SXT	*Vibrio cholerae*	99.5	*prfC*	Cm^R, SXT^R, Sm^R
R391	*Providencia rettgeri*	89	*prfC*	Hg^R, Kn^R
ICEBs1	*Bacillus subtilis*	20	tRNA^{Leu} gene	None known
PAPI-1	*Pseudomonas aeruginosa*	108	tRNA^{Lys} gene	Virulence factors and regulation of biofilm formation
ICE*clc*^{B13}	*Pseudomonas knackmussii*	103	tRNA^{Gly} gene	3-chlorobenzoic acid degradation
ICEMlSymR7A	*Mesorhizobium loti*	502	tRNA^{Phe} gene	Symbiosis with *Lotus corniculatus* involving nodulation and nitrogen fixation
ICE*Hin*1056	*Haemophilus influenzae*	59	tRNA^{Leu} gene	Tet^R, Cm^R, Amp^R
pSAM2	*Streptomyces ambofaciens*	10.9	tRNA^{Pro} gene	None known
Tn*916*	*Enterococcus faecalis*	18	AT-rich regions	Tet^R
CTnDOT	*Bacteroides spp.*	65	GTANNTTTTGC	Tet^R, Erm^R
TnGBS2	*Streptococcus agalactiae* (group B *Streptococcus*)	33.5	Intergenic regions upstream of σ^A promoters	None known

σ^A, RNA polymerase factor σ^A (also known as RpoD); Amp^R, ampicilin resistance; Cm^R, chloramphenicol resistance; Erm^R, erythromycin resistance; Hg^R, mercury resistance; Kn^R, kanamycin resistance; *prfC*, peptide chain release factor 3; Sm^R, streptomycin resistance; SXT^R, sulfamethoxazole and trimethoprim resistance; Tet^R, tetracycline resistance.

Although GEIs are defined as such by sharing many of the above listed features, it is important to note that the current classification of GEIs is not based on a unifying mode of functioning or lifestyle (92). For example GEI functions necessary for their maintenance, excision, transfer or integration might differ substantially between different GEIs. In fact, GEIs may be phylogenetically diverse and share multiple and parallel ancestries. For example GEIs might contain phylogenetically unrelated categories such as prophage-like elements, integrated plasmids, and integrative and conjugative elements (ICEs), the last of which are most relevant to the topic of this thesis (55, 92, 223).

ICEs. While some GEIs have only remnant DNA signatures left of the genes coding for mobility, other GEIs have retained full functionality and are still self-transmissible. Such mobile GEIs are commonly referred to as ICEs (92, 141, 223). Because of their mobility ICEs represent interesting model systems for studying "evolution in action". Indeed, key characteristics of ICE mobility have been experimentally inferred from a number of ICEs conferring different beneficial traits to different hosts (Table 1). Some of these include the antibiotic resistance-conferring ICEs ICE*SXT* in *Vibrio cholerae* (20), PAPI-1 in *Pseudomonas aeruginosa*, CTn*961* and CTn*DOT* in *Bacteroidetes*, and ICE*Hin*1056 in *Haemophilus influenza* (Mohd-Zain, 2004), the nitrogen fixation-conferring ICE*Ml*SymR7A in *Mesorhizobium loti* (164), and the chloroaromatic compound degradation-conferring ICE*clc* in *P. knackmussii* B13 (140, 177). From these model systems we know now that the ICE life-style includes several characteristic steps (Figure 5):

1. Integration into and excision from the host chromosome;
2. Conjugative transfer to a new host via cell-to-cell contact;
3. Maintenance in the host chromosome;
4. Careful regulation of both the transfer steps and the host-benefiting accessory features (antibiotic resistance, aromatic compound degradation etc.) (Figure 5).

Integration and excision from the host chromosome is perhaps the best characterized feature of the ICE life-style; it is mediated by a site-specific phage-like recombinase called integrase in conjunction with accessory proteins (recombination directionality factor/excisionase) that bind to regions surrounding the recombination site and help the integrase to position itself (28, 110, 113, 114, 162, 165, 176, 219). The integrase usually recognizes the direct repeat sequences *attL* and *attR*, which flank the element on either side (200). In the excision reaction, the *attL* and *attR* sites are thus recombined to form the *attB* (the
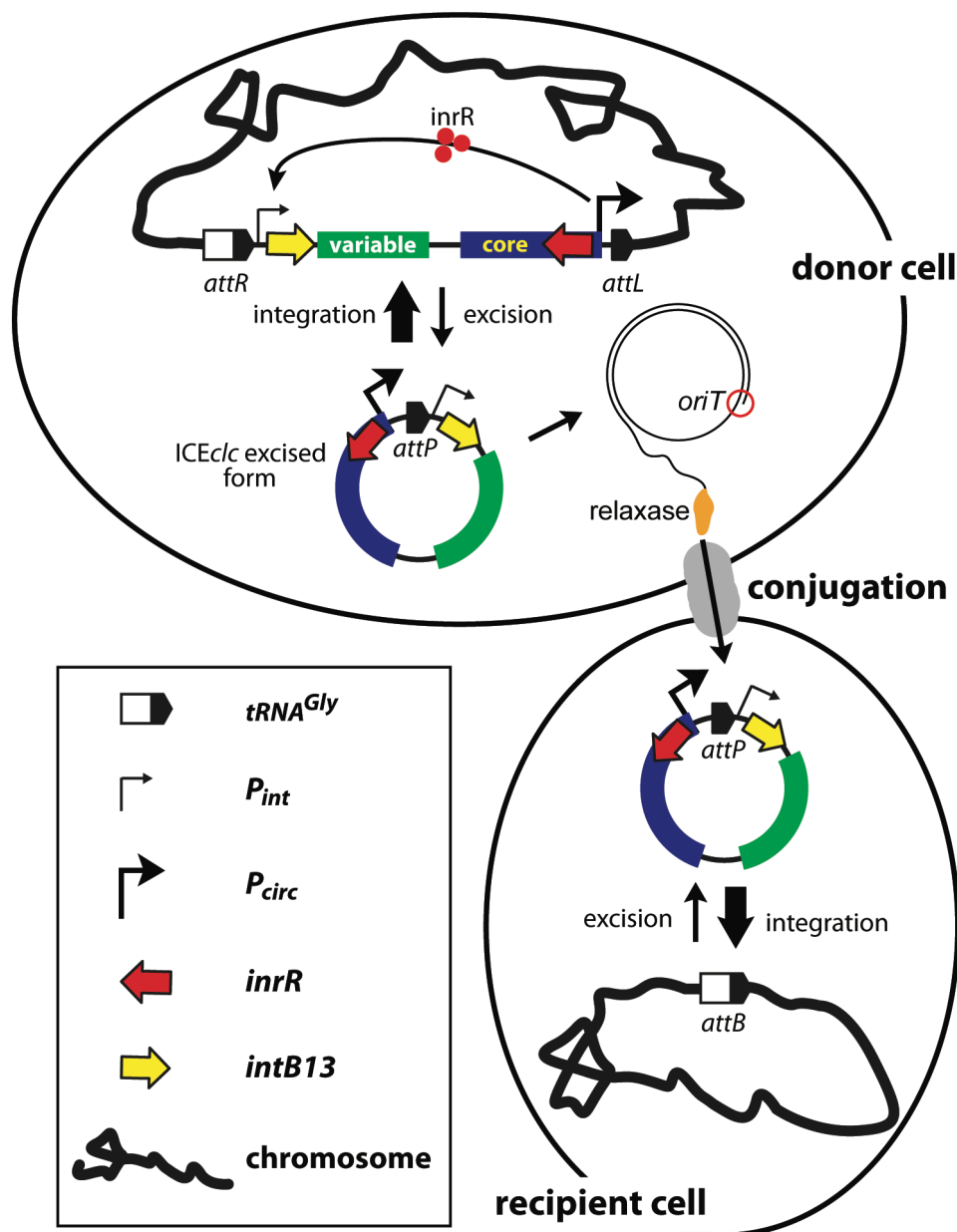
Figure 5. Schematic model of "lifestyle" of ICE*clc*. In the donor cell, ICE*clc* is integrated in a *tRNA^{Gly}* gene and the integrase gene (*intB13*) is under the weak promoter P$_{int}$. In stationary phase, the activation factor InrR induces the production of IntB13 that catalyzes the excision of the ICE*clc* to a circular form. Conjugation starts with DNA processing proteins forming a relaxosome at one or both of the origins of transfer (*oriT*). The single-strand DNA and relaxase complex is then supposed to be transferred into recipient cells through a type IV secretion system. In the recipient, the incoming DNA is recircularized and used as a template to reconstruct the second strand. Once in circular form, the strong constitutive promoter P$_{circ}$ is placed in front of *intB13*. This results in a temporary overexpression of IntB13 that mediates the chromosomal integration of ICE*clc* in the integration site *attB* of a *tRNA^{Gly}* gene. The return to an integrated state restores the control of the weak promoter P$_{int}$ over *intB13*. Meanwhile in the donor cell, the circularized ICE*clc* that is not transferred may be reintegrated into a *tRNA^{Gly}* gene. [Adapted and modified from Miyazaki *et al.* (141).]

chromosomal attachment or integration site) and *attP* (for the "phage attachment site", the leftover recombination site on the now excised ICE circular form) (Figure 5). Once liberated as excised circular form, the ICE molecule can then proceed to conjugative type transfer.

Conjugative type ICE transfer from donor to host cell is less well characterized, and current knowledge is mostly based on similarities to plasmid conjugative systems (32, 142). It is thought that upon a signal a relaxase covalently binds to and nicks the circular plasmid/ICE DNA at a specific DNA sequence pattern called the origin of transfer (*oriT*). This enables the unwinding of a single linear DNA strand (ssDNA) by helicase activity, while simultaneously the other strand is replicated by rolling circle replication. The relaxase remains bound to the displaced ssDNA and interacts with a coupling protein to recruit the nucleo-protein complex to a protein complex called the mating pore formation structure (Mpf). The Mpf is a plasmid/ICE-encoded protein structure that assembles at the cell surface upon transfer-initiation and forms a junction between the donor and recipient cell through which the relaxase-ssDNA nucleoprotein can be "actively pumped" by means of the coupling protein (32). Together, coupling protein and Mpf constitute the so-called type IV secretion system, the exact make-up of which seems to be a major distinguishing feature between different ICE families. For example, gene clusters involved in transfer of ICE*Hin*1056 are evolutionarily very distant from all previously described plasmid T4SS genes (91). Also the ICE elements pKLC102 and PAPI-1 from *P. aeruginosa*, and ICE*clc* from *P. knackmussii* B13 show little sequence homology to known plasmid conjugative systems (91)). Once in the new cell, the co-transferred relaxase participates to the recircularisation of the transferred ssDNA (175), which, in the case of an ICE, is then ready to integrate into the new host chromosome (Figure 5).

ICE transmission is governed by complex regulatory networks that can be activated and repressed by environmental stimuli. These signals influence the expression and activity of ICE-derived and host-derived factors to modulate ICE gene expression and ICE transfer. A variety of different regulatory mechanisms that control the transition from the integrated state to ICE excision and transfer are now beginning to become understood, which in several cases are reminiscent of bacteriophage behaviour (223). For example, activation and excision of the antibiotic resistance determinant ICE*SXT* of *Vibrio cholerae* is controlled by an SOS response pathway involving the master regulator SetR (20, 21). Another well-described ICE model system is the CTn*DOT* family (the ICEs CTn*961* and CTn*DOT*) in *Bacteroidetes*, which is well known for the dissemination of antibiotic resistances. In this ICE family sub-inhibitory concentrations of tetracycline increase their transfer frequencies resulting in an increase of tetracycline resistance in the local bacterial community (38, 39).
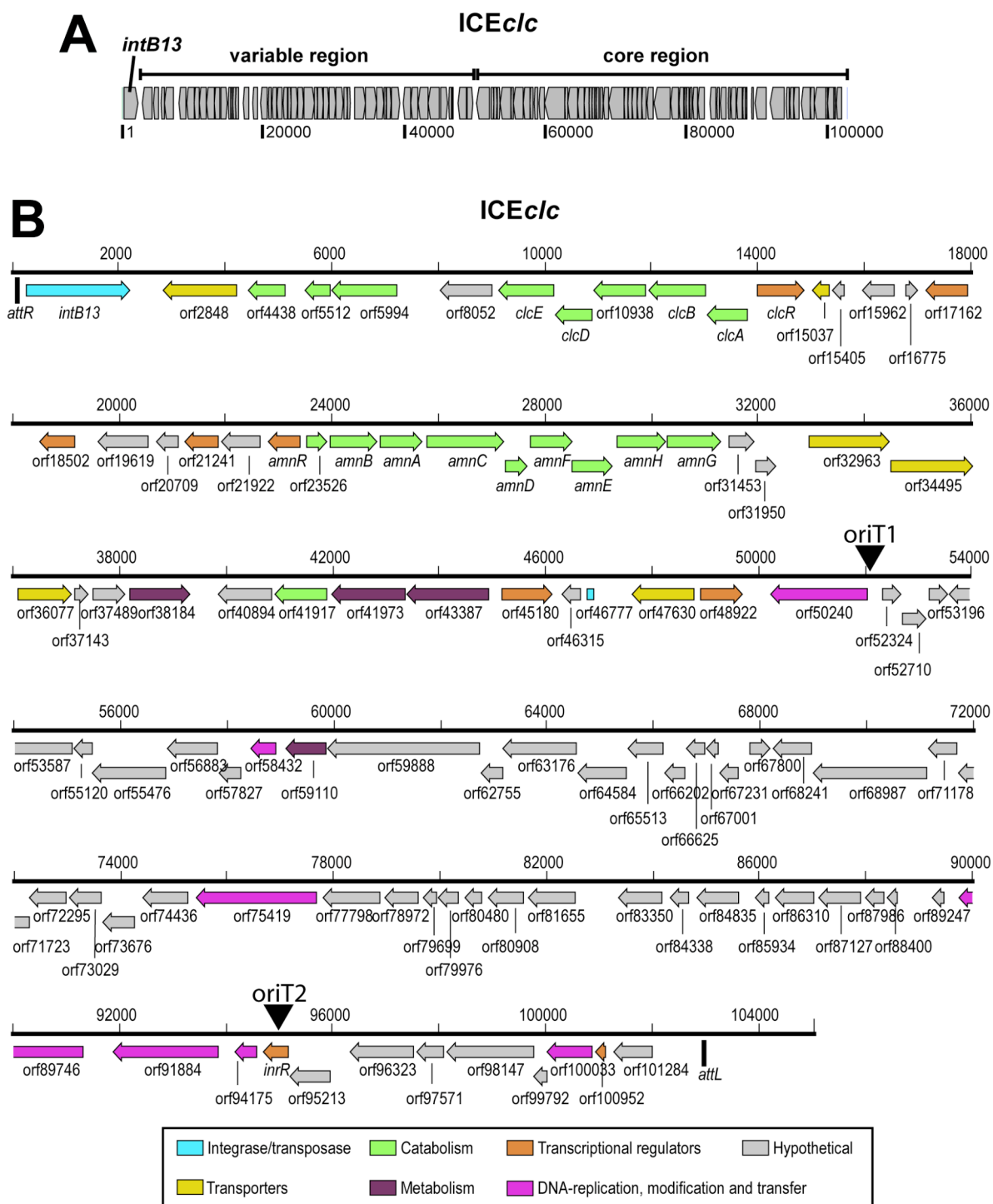
Figure 6. Schematic representation of the ICE*clc* genetic organization. (A) Overview of ICE*clc* genetic organization showing core and variable region. Gray pentagons point to ORFs. [Adapted and modified from Miyazaki *et al*. (141).] (B) Detailed ORF map of ICE*clc*. Arrows indicate orientation of putative ORFs. Colors correspond to predicted function. Thick vertical lines represent the border of ICE*clc*. [Adapted and modified from Miyazaki *et al*. (141).]

If chromosome replication or cell division occurs after ICE excision from the chromosome, the element may not be transmitted to progeny and as a result would be lost from the host cell. However, ICEs have evolved maintenance strategies counteracting this effect and ensuring that the frequency with which ICE-free cells arise stays low. For example, ICE*SXT* guarantees its maintenance in the original host cell population by a toxin-antitoxin 'addiction' module encoded by *mosAT* that is thought to kill any arising daughter cells without ICE*SXT* (222). The SOS response is also one of the triggers for excision and transfer of the 21-kb ICE*Bs1* element in *Bacillus subtilis* (25, 26). In contrast to ICE*SXT* ICE*Bs1* ensures its maintenance by DnaN-independent autonomous replication in the excised form (111). Maintenance of the *P. aeruginosa* PAPI-1 element is dependent on Soj (a ParA-analogous protein) and deletion of the *soj* gene leads to a rapid proliferation of cells without PAPI-1 in the population (162). ICE*clc* has a strong preference for reintegration after excision. It is suggested that this is because ICE*clc* excision leads to displacement of a strong promoter (P_{circ}) at the left end of the element in front of the integrase gene (Figure 5). A temporary overexpression of IntB13 occurs from the circular form and thus increases the favorability of integration in the new host chromosome (177).

In addition to the life-style-encoding core region ICEs generally also contain a "variable region" which encodes characteristics that benefit the host cell. This regions is referred to as variable because it represents the region that is least conserved among all ICEs, containing genes encoding a wide range of ICE-specific phenotypes, including resistance to antibiotics and heavy metals, the capacity to degrade aromatic compounds, and complex traits such as the ability to colonize a eukaryotic host, fix nitrogen, or promote virulence and biofilm formation.

ICE*clc*. Whereas previous studies on ICEs have been of cardinal importance to begin to understand the basic ICE-host control mechanisms, there is still a fundamental lack of understanding ICE-host behaviour at the single cell level, which is crucial for their evolutionary success. Recently, it could be shown by single cell analysis that activation of the P_{int}-promoter leading to excision of ICE*clc* in *Pseudomonas knackmussii* is the consequence of a bistable activation cascade occurring only in a 3-5% of cells in stationary phase (140, 178). Most of the genes implicated in ICE*clc* excision and transfer are only actively expressed in stationary phase cultures, suggesting they all respond to the bistable activation cascade (70, 140) and form a dedicated program to prepare cells for ICE*clc* transfer (i.e., a "transfer competent state"). ICE*clc* has a size of 103 kb and occurs in two copies in the chromosome of *P. knackmussii* B13, inserted at the 3'-end of genes for *tRNA*^{Gly} (71, 168). The element bestows the host with the capacity to metabolize a number of unique aromatic carbon substrates, among which 3-chlorobenzoate (3-CBA). Based on sequence comparisons to other genetic

17

elements ICE*clc* can be divided into three characteristic parts (Figure 6) (141). The first part is formed by the *intB13* integrase gene and is located directly near the *attR* end. This organization is similar as in other GEIs (see Figure 3 B). The second part consists of a ~50 kb region which is variable among related GEIs and which in case of ICE*clc* contains the accessory genes for chlorocatechol and 2-aminophenol metabolism. The third part comprises the remainder ~50 kb region until *attL*, and is mostly composed of genes with unknown functions, although some genes display homologies to conjugative processes or to phages. Interestingly, however, this region is highly conserved among a wide range of GEIs and uncharacterized genome regions in a variety of other bacteria. It was therefore hypothesized that this region represents the 'core' region of ICE*clc* and its relatives (Table 2), and encodes the functions necessary for the element's life-style, that is, excision, transfer and regulation of transfer (92, 141).

ICE*clc* can transfer at population averaged frequencies of 1-5% per donor cell to a number of *Beta-* and *Gammaproteobacteria*, such as *Pseudomonas putida*, *P. aeruginosa* or *Cupriavidus necator* (69, 177). Natural ICE*clc* variants have been discovered in a wide variety of other bacterial species (140), among which *Bordetella petrii* (109), *Burkholderia xenovorans* (71) or *Ralstonia* sp. strain JS705 (218), underlining the success of this class of ICE to infiltrate different host species. Not only ICE*clc*, but a variety of other ICE have self-transmission frequencies in the order of a few percent per donor cell or lower, suggesting that induction of a transfer-competent state in some cells in a clonal population of ICE-bearing donors is a common strategy (11, 13, 109, 162).

Table 2. Key characteristics of experimentally described GEIs/ICEs related to ICE*clc*. [Adapted and modified from Miyazaki *et al.* (141).]

ICE	Host	Size (kb)	Site of insertion	Notable phenotypes
ICE*clc*	*Pseudomonas knackmussii*	103	tRNA^{Gly} gene	Chlorocatechol and 2-aminophenol
GI1	*Bordetella petrii*	225	tRNA^{Gly} gene	Phtalate to protocatechuate metabolism
GI2	*Bordetella petrii*	143	tRNA^{Gly} gene	Aromatic compound degradation
GI3	*Bordetella petrii*	102	tRNA^{Gly} gene	Chlorocatechol degradation
GI6	*Bordetella petrii*	159	tRNA^{Gly} gene	None known
ICE*Hin*1056	*Haemophilus influenzae*	59	tRNA^{Gly} gene	Antibiotic, metal and antiseptic resistance
PAGI-2	*Pseudomonas aeruginosa*	105	tRNA^{Gly} gene	Complexation and transport of heavy metal
PAGI-3	*Pseudomonas aeruginosa*	103	tRNA^{Gly} gene	Metabolic functions and antibiotic resistance
PAPI-1	*Pseudomonas aeruginosa*	108	tRNA^{Gly} gene	Virulence factors and regulation of biofilm formation
CMGI-1	*Cupriviadus metallidurans*	109	tRNA^{Gly} gene	Heavy metal resistance
no name	*Burkholderia xenovorans*	124	tRNA^{Gly} gene	Chlorocatechol degradation

TIME LAPSE FLUORESCENCE MICROSCOPY FOR STUDYING MICROBIAL HETEROGENEITY

The study of phenotypic heterogeneity among bacteria, or bacterial individuality, requires by definition the observation at the single-cell level. Advances in single cell technology therefore, have recently boosted bacterial single-cell research and led to an awareness that average data obtained from traditional, population-based experiments often do not represent correctly the behavior, status or phenotype of single cells (Figure 7).

Microscopy has a long history as a powerful research tool for achieving single cell measurements; indeed, the first-ever images of differently shaped bacterial single cells were drawn by Antonie van Leeuwenhoek and published in 1638 (54, 112) (Figure 8A). Today, microscopy, and epifluorescence microscopy in particular, do not only distinguish between cells from differently shaped species but even between clonal cells of isogenic populations. This distinction between clonal cells is made possible by use of autofluorescent proteins or dyes (fluorophores) that label proteins of interest, follow expression from specific promoters, or stain specific nucleic acid sequences (98, 122).

In some epifluorescence microscopy applications the detection limits have been pushed so far that individual protein or mRNA molecules in the bacterial cell can be visualized and quantified (30, 40, 77, 116, 224). Furthermore, by employing differently coloured autofluorescent protein variants multiple parameters can be followed simultaneously and in the same cell (35, 121, 179). For example, in this thesis we measured differential gene expression in single *Pseudomonas putida* bacterial cells from two different promoters, the constitutively expressing P_{tac} and the inducible P_{int}. Their expression could be differentiated by cloning the gene for mCherry fluorescent protein downstream of P_{tac} and that for green fluorescent protein downstream of P_{int} (Reinhard, unpublished) (Figure 8 B).

In contrast to epifluorescence microscopy, flow cytometry of bacterial cells expressing autofluorescent proteins does not permit time-lapse imaging, but it gives that advantage of a high throughput acquisition of single-cell parameters (45, 146). In modern epifluorescence microscopes time-lapse imaging is greatly assisted by automated image acquisition. This, in combination with autofluorescent proteins that are generally non-toxic, permits observation in live cells and temporal imaging (time-lapse imaging) over long time-scales. Microscopes equipped with motorized stages can be programmed to take images of multiple regions of interest at set time-intervals with different light regimes (e.g., bright field, green fluorescence, red fluorescence). Some fully automated time-lapse set-ups contain autofocus systems to achieve consistently sharp images of bacterial cells throughout an experiment taking up to
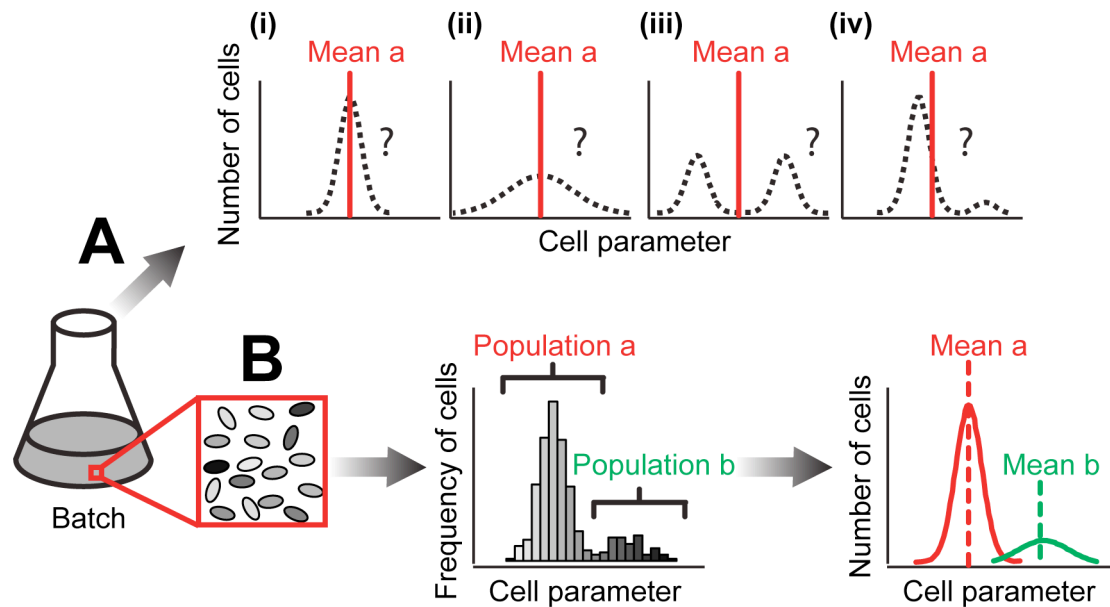
Figure 7. Subpopulation effects stay hidden within averaged data sets. Shades of grey represent cells with low (white), medium (gray), and high (black) levels of the measured parameter. (A) Batch culture experimental techniques measure the average of a parameter across an entire population, but equal population means may characterize very different populations as indicated by four different population structures (i-iv). (B) Single-cell analysis techniques such as flow cytometry or microscopy generate a more complete picture of the diversity that may be concealed by averaging. For example different population means (Mean a, Mean b) representing underlying subpopulations can be deduced from bimodal populations, which would not be possible by batch culture experimental techniques.

several days (4, 122, 212).

Taking full advantage of automated (or semi-automated) image acquisition quickly results in the generation of enormous data sets often containing measurements of multiple cell parameters from several cells each with hundred to thousands of time points. Such "big" amount of data needs image analysis software that automatically and in an unbiased way is able to extract relevant information. Typically, image analysis software is based on algorithms that in the first step segment an image into regions of interest (such as cells) based on the intensities of adjacent groups of pixels. These regions (cells) can then be rapidly classified according to many different criteria including their intensity, shape, size, co-localization with regions in other color channels, and countless more. Open source software exists where regions of interest can be tracked over time from time-lapse image stacks, with any number of quantitative phenotypes extracted at each time point (48, 73, 170, 187, 214). In this way the history of single cells within a growing population (such as a bacterial microcolony) can be traced back to one common ancestor allowing for the construction of phylogenetic lineage trees. These allow visualization and analysis of the effect of cell history and ancestry on specific cellular behavior (Figure 21) (48, 122). Movies constructed from time-lapse image stacks provide a direct view of genetic activities in individual cells (48, 122). Movies are not only aesthetically appealing, but also allow for the eye to pick out subtle patterns in individual cells that would be difficult to notice without visual interface (122).

Another important aspect in single cell analysis using time-lapse microscopy is the choice of incubation chambers that allow tracking of single cell growth whilst guaranteeing reproducible environmental conditions. Autofluorescent proteins need oxygen for fluorophore activation, and therefore a sufficiently high oxygen availability has to be ensured. Probably the simplest set-up involves in-house fabricated chambers that allow for two-dimensional microcolony growth between a coverslip and an agarose pad (23, 49, 59, 212). These systems are mostly limited to the growth conditions and sample size. Because they are sealed, medium conditions cannot be changed during the experiment and cells eventually 'pile up', limiting the length of time of observation. More sophisticated set-ups include flow chambers (4, 213) or capillary flow systems (36, 99, 120, 213, 225). Microfluidic chambers have been introduced to maintain continuous culturing conditions (15, 43). Some groups developed linear microfluidic chambers in which cells are confined to grow in a narrow groove. These devices facilitate analysis of cell lineages, as more closely related cells are closer together (213). Generally these systems, however, do not allow for the construction of phylogenetic lineage tress covering more than three generations (Figure 21).

A

B
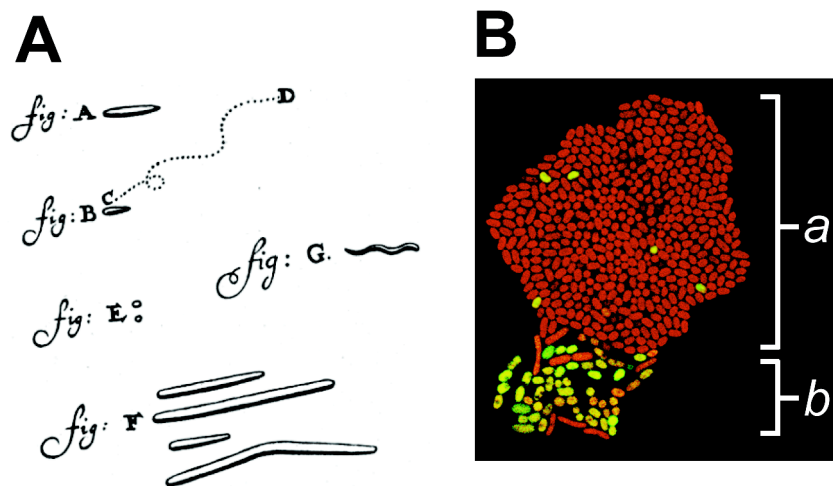
Figure 8. Examples showing how single cell microscopy may help to detect microbial diversity and bacterial individuality. (A) Diversity of bacteria from a human mouth as described by van Leeuwenhoek in a letter of 17 September 1683. *A*: motile Bacillus; *B*: *Selenomonas sputigena*, with *C...D* its path; *E*: Micrococci; *F*: *Leptothrix buccalis*; *G*: *Spirochaete*, probably *Spirochaeta buccalis*. [Adaped from Dobbell (54).] (B) Bacterial individuality in two (touching) microcolonies (*a* and *b*) of *P. putida* UWC1. Overlay of time-lapse confocal laser scanning microscopy images in excitation and emission wavelengths for EGFP and mCherry. An orange colour results where both marker proteins are expressed to a similar level. In this strain mCherry marker protein (red) is expressed from the constitutively expressing P_{tac}-promoter, while the EGFP marker protein (green) is expressed from the inducible P_{int}-promoter. [Reinhard (unpublished).]

BACTERIAL INDIVIDUALITY AND ADAPTATION

The viability of any organism critically depends on its ability to adapt to the environment. Bacteria have evolved global control sensory systems that allow cell physiological adjustment or movement to more favourable locations in response to an environmental change. Environmental changes might include changes in chemical composition of the environment (132), local temperature (131), or intensity of light (136). For example, bacteria harbour two-component systems that enable in a first step, the sensing of environmental change via membrane-bound sensor kinase proteins, to then in a second step, mount an appropriate cellular response via signal transduction pathways (132). Another global control sensory system, but mechanistically very different to two component systems, is catabolite repression, where the presence of the carbon source glucose can be sensed via absence of the trigger molecule cyclic AMP, and as a result catabolism of lower energy carbon substrates is shut down to save energy (68). A third global control sensory system in bacteria is the stringent response. Here, the alarmone ppGpp serves as a pathway trigger that can be thought of as a mechanism for adjusting the cell's biosynthetic machinery to amino acid limitation (129). Many other global control sensory systems exist, such as the heat shock response for assisting recovery from stress (82), the cold shock response for preventing ice formation in the cytoplasm (172), or quorum sensing for responding to the presence of other cells of the same species (17).

What combines all global response sensory systems is that in order to function they need to maintain a sensory component in an active, responsive state. Such maintenance is energetically costly however, and the fitness advantage of maintaining a sensory system becomes increasingly small the longer the environmental conditions remain constant (106). Furthermore, time-consuming readjustment to an adverse condition may in some cases, like the sudden presence of an antibiotic, fail to guarantee survival of the entire population (14). Not surprisingly therefore (at least for the evolutionary biologist), another strategy of adaptation has been found in bacteria, that in comparison to global response strategies, spares the metabolic costs of sensory machinery while ensuring immediate adaptation to changing conditions (106). This strategy is commonly referred to as a bet-hedging strategy, which functions on the basis of cell-to-cell heterogeneity also called bacterial individuality (185, 211). Heterogenic clonal populations, by chance, often include some individuals that are better adapted to survive future environmental change than others. Some of these are therefore likely to survive environmental change, thereby guaranteeing the survival of the clone. Importantly, the offspring of the survivors will again yield populations with cell-to-cell heterogeneity facing the same strategy of

pre-adaptation. However, unlike the global response, bet-hedging comes at the expense of sub-optimal net population growth due to a few slower growing or non-growing specialized phenotypes in a population (106). Modelling showed that bet-hedging is only evolutionary favourable as long as environments change rarely (106). However, as soon as frequent fluctuations in environmental conditions occur, global sense-response mechanisms are preferred (22, 106, 117).

So far, population heterogeneity has been found to play a role in the metabolism of carbon sources (139), persistence (15, 53, 188), biofilm formation decision (33), (34), (96), competence (212), sporulation (211), cell size (23), (6), phage development (133), (160), and horizontal gene transfer (140).

One of the earliest discoveries of population heterogeneity was the observation of variation in colony morphology arising from single bacterial clones (9). Later, in the 1940s, the phenomenon of persistence was discovered: an antibiotic susceptible population of bacteria containing an antibiotic resistant subpopulation, which generates equal fractions of persister and non-persister populations upon release of the antibiotic pressure (24). Seventeen years later, in a landmark study on the metabolism of lactose, Novick and Weiner (149) showed that the production of beta-galactosidase in individual cells was highly variable and random, with induction increasing the proportion of cells expressing the enzyme rather than increasing every cell's expression level equally.

Although all of these early studies presented important proof for the existence of population heterogeneity, they were hindered substantially by the lack of reliable single-cell assays of gene expression available at that time. Only recently, with the advent of gfp-reporter technology and time-lapse microscopy, have microbiologists achieved breakthroughs in elucidating the underlying mechanisms of population heterogeneity at single cell level. Recent hypotheses classify population heterogeneity into at least three types of underlying mechanisms (47): First, genetic mechanisms that are based on stochastic rearrangement or modification of DNA (phase variation). Second, epigenetic mechanisms that are based on the feedback architecture and molecular noise of genetic networks (gene expression noise). And third, phenotypic heterogeneity related to bacterial cell aging (4, 187). I will explain all three mechanisms of population heterogeneity formation more in detail below.

Population heterogeneity based on phase variation

As mentioned above, three main mechanisms are recognized that cause phenotypic heterogeneity: phase variation, noise and cell aging (47). Phase variation is a process causing temporary and reversible genetic heterogeneity in a population. It was first discovered in pathogenic bacteria (207). Pathogenic bacteria are often faced with a rapidly fluctuating, unpredictable and therefore challenging host environment. Such variable conditions can stem from the infection process that is dynamic over time, for example as a result of an inflammatory response or the production of antibodies, or they occur during transmission from one host to another (144). Phase variation causes spontaneous changes in the biosynthesis of surface exposed structures such as flagella, fimbriae, pili and outer lipopolysaccharides (LPS) (144, 207). Such structures are typically exposed to the immune system and play an important role as antigens during an immune response. Their phase variation characteristic makes sense as a strategy to evade the host immune system in that the resulting phenotypic heterogeneity in surface structures ensures that the risk of immune detection is spread among variable offspring, each of which has some chance of avoiding recognition (163).

Phase variation is achieved through changes in the DNA, either through the generation of novel genotypes via nucleotide sequence mutations, or through DNA modification via methylation (206). Common to all examples of phase variation are three main characteristics: high-frequency mutations at specific loci, reversibility, and prevalence for mutations in genes coding for surface-associated proteins (144). Underlying phase variation are at least four different mechanisms (47):

1. Slipped-strand mispairing (SSM);
2. Site specific recombination;
3. Epigenetic DNA modification;
4. Antigenic variation.

Slipped Strand Mispairing. A first mechanism of phase variation is termed slipped-strand mispairing (SSM) (144). SSM acts via short sequence repeats (SSRs) termed contingency loci that function as hot-spots for transient mispairing during either DNA replication or in the course of other processes that require DNA synthesis, such as DNA repair and recombination. Mispairing of complementary DNA strands at contingency loci can cause changes in SSR tract length, and subsequently result in a frame shift mutation abolishing appropriate gene expression (81) (Figure 9). Depending on whether the alteration occurs

upstream of or within the gene, SSM can either influence transcription or translation, respectively. A good example for transcriptional SSM regulation is the homopolymeric tract of cytidines (C-tract) in front of the promoter for fimbrial subunits in *Bordetella pertussis* (37) (Figure 9B). Here, the C-tract functions as a spacer that optimizes interaction between an activator, BvgA and RNA polymerase. A change from 17 repeats to 16 repeats disrupts BvgA-dependent fimbrial gene activation. Other similar examples of transcriptional SSM regulation include the adhesin encoding *nadA* gene in *N. meningitidis* (138) or the fimbriae-encoding *hif* gene in *Haermophilus influenza* (208). An example of translational SSM regulation can be seen in the expression of the adhesin *opa* genes in *N. gonorrhoeae* and *N. meningitidis*. Opa mediate bacterial adhesion to and invasion of host tissues by interacting with different classes of cellular receptors. Alterations of the reading frame where the initiation codon is out of frame with the mature molecule results in aberrant or truncated proteins (83). Interestingly, since the *opa* genes are present in multiple copies on the chromosome, effects of On/Off switching are combinatorial. For example, stochastic On/Off switching of 12 independent *opa* genes in *Neisseria gonorroehae* can potentially lead to a phenotypic variety of 4096 combinations (51, 81, 103, 147). A diverse arsenal of Opa receptors increases the likelihood of matching diverse receptor specificities, allowing bacterial populations to either sustain infection at a specific site or to adapt to new niches within the same or different hosts. Clonal variation through combinatorial On/Off switching is maybe best illustrated in the expression of a subset of LPS contingency loci in *H. influenza*. The genes here are required for the addition of the core sugars, glucose, galactose and sialic acid, to the conserved tri-heptose backbone of the LPS molecule (Figure 10) (144).
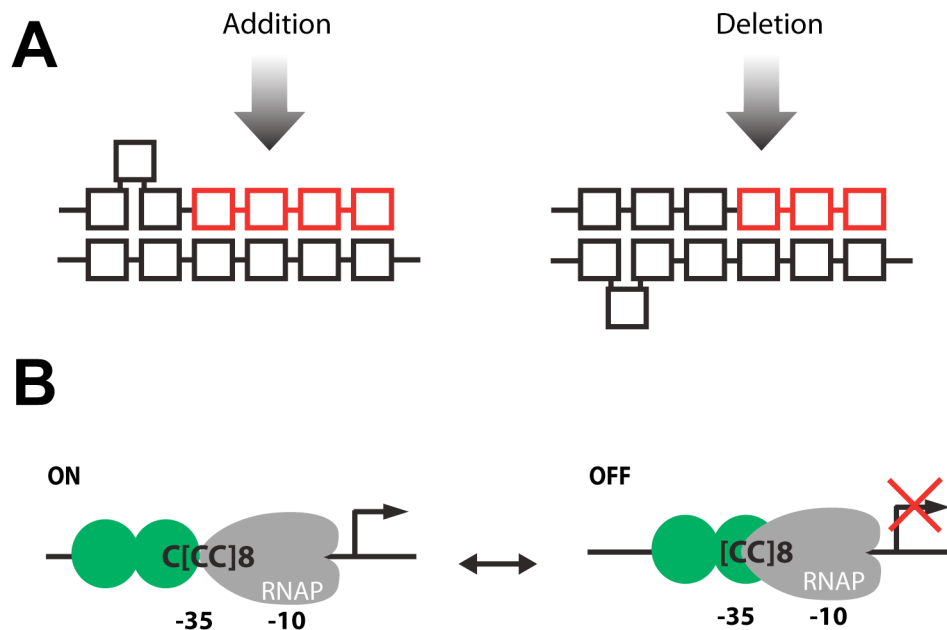
Figure 9. Slipped strand mispairing (SSM). (A) Diagram illustrating the addition or deletion of simple sequence DNA repeats through SSM. Each strand is represented by a single line and repeat units by open rectangles. DNA polymerase slippage gives rise to unpaired repeat units that generate bulges in DNA. If a bulge occurs in the synthesized strand (top strand) an addition of repeat length will follow. If the bulge occurs in the template strand (bottom strand) a deletion of repeat length will follow. [Adapted and modified from Moxon *et al.* (144).] (B) A change in C-tract length from 17 (ON) nucleotides to 16 (OFF) as a result of SSM disrupts BvgA (green circles) dependent activation of *fim* in *B. pertussis*. Promoters are depicted as -35, -10 with a black arrow indicating the direction of transcription. A red cross indicates no transcription. [Adapted and modified from van der Woude (206).]
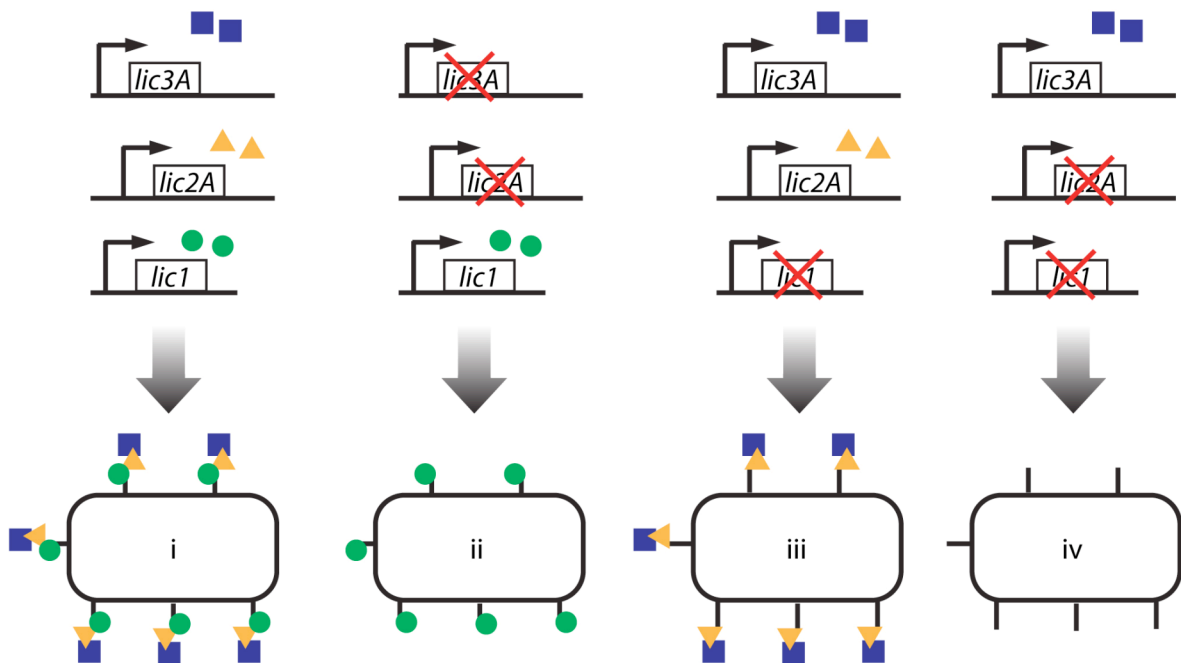
Figure 10. LPS phase variation based on SSM in *H. influenzae*. (i) - (iv) Independent expression of *lic2A*, *lic3A* and *NeuAc* results in combinatorial addition of galactose (yellow triangles), sialic acid (blue squares), and phosphorylcholine (green circles) at the bacterial cell surface, respectively. Boxes represent contingency loci within each gene. A red cross indicates no expression due to SSM. [Adapted and modified from Moxon *et al*. (144).]

Site-specific Recombination. Phase variation via site-specific recombination exhibits several defining features that highlight the difference of this mechanism to general recombination (47). Firstly, the area of homology is shorter than in general recombination. Secondly, the recombination event occurs between specific loci rather than anywhere within the region of homology. And thirdly, specific recombinases are required (206). In many cases, site-specific recombination alters the orientation of the promoter relative to its coding region. For example, expression of *fimA*, the major subunit of type 1 fimbriae in *E. coli,* phase varies as a result of the site-specific inversion of the *fimA* promoter (Figure 11A). The inversion is mediated by two site-specific recombinases, FimB and FimE that recognize two inverted repeats. In the "OFF" orientation, the promoter is incorrectly oriented for transcription of the *fimA* structural gene and hence fimbriae cannot be synthesized. However, inversion can switch back to the "ON" orientation due to the bidirectional recognition of the inverted repeats by FimB (100, 134). There exists at least one example where an invertible element leads to phase variation not through inversion of the promoter but through creation of a transcription terminator (Figure 11B). Depending on the orientation of the invertible element in the DNA of *Clostridium difficile*, an intrinsic transcription terminator in the 50 leader region of the cell wall gene *cwpV* is formed (62). Interestingly, there are also some reports in which genetic elements, through site-specific integration and excision, have been found to cause inheritable changes in the virulence of bacterial pathogens. For example in *Legionella pneumophila*, phase variation of LPS and other virulence factors correlates with the presence of a defective bacteriophage that alternates between an integrated and an extrachromosomal replicative form (81). Similarly, reversible inactivation of extracellular polysaccharide production by *Pseudoalteromonas atlantica* is caused by the integration and excision of IS*492*, a member of the IS*110* family of insertion elements, allowing this marine bacterium to alternate between free living and biofilm-forming stages (81).

DNA Modification. The third mechanism of phase variation is epigenetic DNA modification via methylation and occurs in the absence of a change in DNA sequence. DNA methylation is primarily known for its role in prokaryotes to monitor and regulate the contents of their genomes. For example, *E. coli* methylates its DNA in order to indicate replication status or direct DNA post-replicative mismatch repair. This functionality has been exploited in an elegant study that was first to visualize horizontal gene transfer of a plasmid (12). But the role of DNA methylation in phase variation serves a different purpose: That of changing expression of cell surface structures to ensure immune response evasion. This is achieved by involvement of differentially methylated sequences in the regulatory regions of the phase-varying gene or operon. Here, the methylation state affects the DNA binding of a regulatory protein that directly
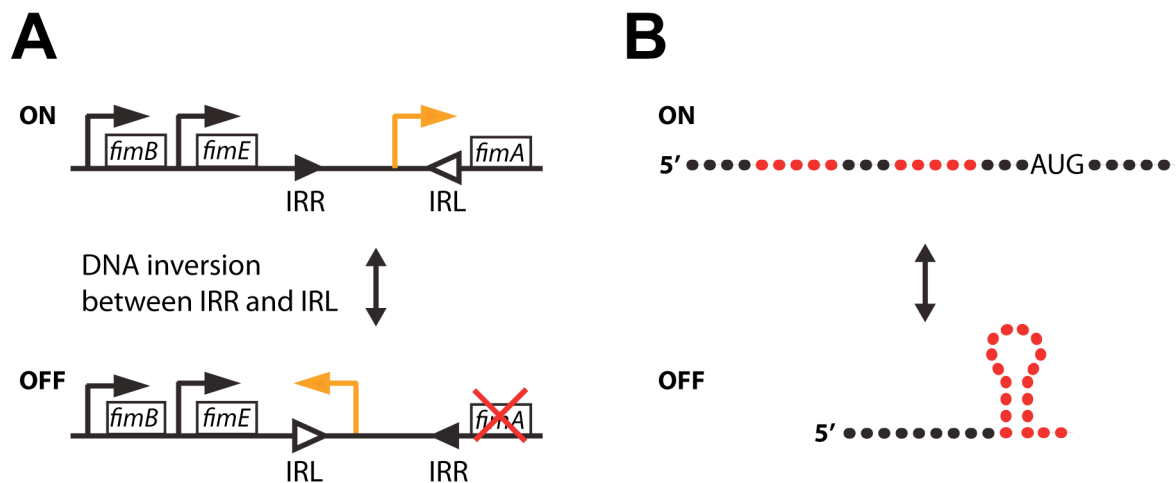
Figure 11. Phase variation mediated by site-specific recombination. (A) ON and OFF-switching of type 1 fimbrial expression in *E. coli* occurs as a result of *fimA* promoter re-orientation (orange) through DNA inversion mediated by site-specific recombinases *fimB* and *fimE* via recognition of flanking inverted repeats (IRR and IRL). Genes are shown as boxes, promoters as arrows, triangles as inverted repeats, and the red cross indicates no expression. [Adapted from van der Woude and Baumler (207).] (B) DNA inversion in the *cwpV* promoter of *C. difficile* forms an intrinsic terminator (bulge of aligned red dots) in the transcript of *cwpV* leading to the Off-phase by preventing transcript elongation. Transcipt is represented as dotted line, with red indicating sequence to be aligned. [Adapted and modified from van der Woude (206).]

regulates transcription. A recent example for such a mechanism is OxyR-dependent and Dam-dependent phase variation controlling expression of the family of *gtr* operons implicated in lipopolysaccharide modification in *S. enterica* serovars (207). Phase variation of *gtr* depends on differential occupation of the oxidative stress response protein OxyR at two binding sites in the regulatory region, which is dictated by the methylation state of two pairs of Dam target sequences (Figure 12). Another example of epigenetic phase variation is the expression of pyelonephritis-associated pili (pap) (84). Pap are cell surface structures in uropathogenic strains facilitating adhesion to host cell surfaces, especially in the urogenital tract. The regulatory region of the *papBA* operon contains six leucine responsive regulatory protein (Lrp) binding sites, and two deoxyadenosine methyltransferase (Dam) methylation sites (199). Alternating methylation of the proximal and distal methylation sites results in the freeing and occupation of Lrp from Lrp binding sites that are required for transcriptional activation. Switching in *pap* expression is hypothesized to be linked to cell division, only occurring when the DNA is transiently hemimethylated following DNA replication. There can be several non-identical homologs of *papBA* operons in one cell able to cross-regulate each other's switching frequency. Maintaining diversity in *pap*-like operons in one cell and promoting their sequential expression, potentially increases individual differences and therefore population heterogeneity.

Antigenic Variation. The fourth mechanism of phase variation is antigenic variation. Antigenic variation involves recombination of silent homologs to generate novel variants. For example, the *Neisseria gonorrhoea* genome contains multiple scattered copies of the silent and variable pili gene *pilS* (81, 83, 85, 137). The *pilS* coding region contains a number of semivariable and hypervariable regions, as well as a conserved region identical to *pilE*. This allows parts of *pilS* to be recombined by RecA with its distant homolog *pilE*, from where expression can take place due the presence of a promoter region (85). Thus, variability of *pilS* combined with subsequent gene shuffling to *pilE* can potentially generate millions of antigenic variants representing a powerful bet-hedging strategy to evade host immune response via population heterogeneity.
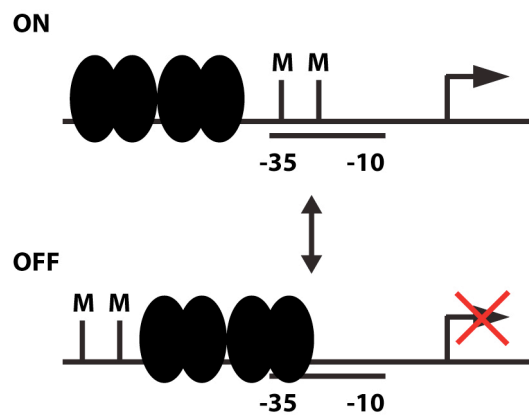
Figure 12. Phase variation through DNA modification in *S.enterica* serovars. Differential occupation of OxyR (black ovals) at two binding sites in the regulatory region is dictated by the methylation state of two pairs of Dam target sequences (M) leading to expression (ON) or non-expression (OFF) from the *gtr* operon. Promoters are depicted as -35, -10 with black arrows indicating direction of transcription. A red cross indicates no transcription. [Adapted and modified from van der Woude (206).]

Population heterogeneity based on noise

We have learnt above that population heterogeneity in phase variation is based on transient mutations or modification of DNA (206). However, population heterogeneity can also be the result of gene expression noise (10, 47, 58, 60, 66, 93, 122, 124, 125, 130, 167, 173, 185, 211). For example, and explained in more detail later, the induction of competence in *Bacillus subtilis* is probabilistically induced on the basis of noise (127, 184). Other examples, described in more detail later, are antibiotic persistence in *E.coli* (14, 15), lactose metabolism in *E.coli* (1, 156), lysis/lysogeny in phage lambda (154, 160), and sporulation in *B. subtilis* (210, 212).

Experimentally, noise in gene expression can be measured by coupling synthesis of autofluorescent proteins to promoters and record variability of fluorescence in single cells (49, 61, 142, 155). Such variability can either be determined from time-lapse observations that record the variation in fluorescence marker expression of a single-labelled strain over time (49, 61, 155), or, from single time-point observations that record the fluorescences of a double-labelled strain expressing two reporter genes variants under identical promoters (Figure 13) (60, 93, 142, 194, 209). The finite number effect plays an important role because low numbers of regulatory factors (between a few tens to a few hundred molecules per cell (93)), are often involved in promoter activation and gene transcription. The stochastic fluctuations in the biochemical reactions of these steps can therefore be large, leading to bursts of transcripts, predominantly through on and off RNA polymerase activity (66, 93). Resulting transcription rate variation is either amplified or buffered in the following translation step depending on protein lifetime and interval time between transcriptional bursts, causing variation in protein levels per individual cell. These variations in protein levels are known as "intrinsic" noise as its origin lies in the biochemical reactions of gene expression itself and independent of other factors. Intrinsic noise could therefore be described as the gene expression variation that would occur between cells if they were in precisely the same state (61, 194). The other ubiquitous component of gene expression noise is extrinsic noise (61, 167, 194). In fact, total gene expression noise is always made up of both, intrinsic and extrinsic noise (61, 167, 194). Extrinsic noise results either from global physiological variations in the cell, such as variations in the metabolic state of the cell, the cell cycle or cell age, or as a result of variability in upstream signal transduction (66).
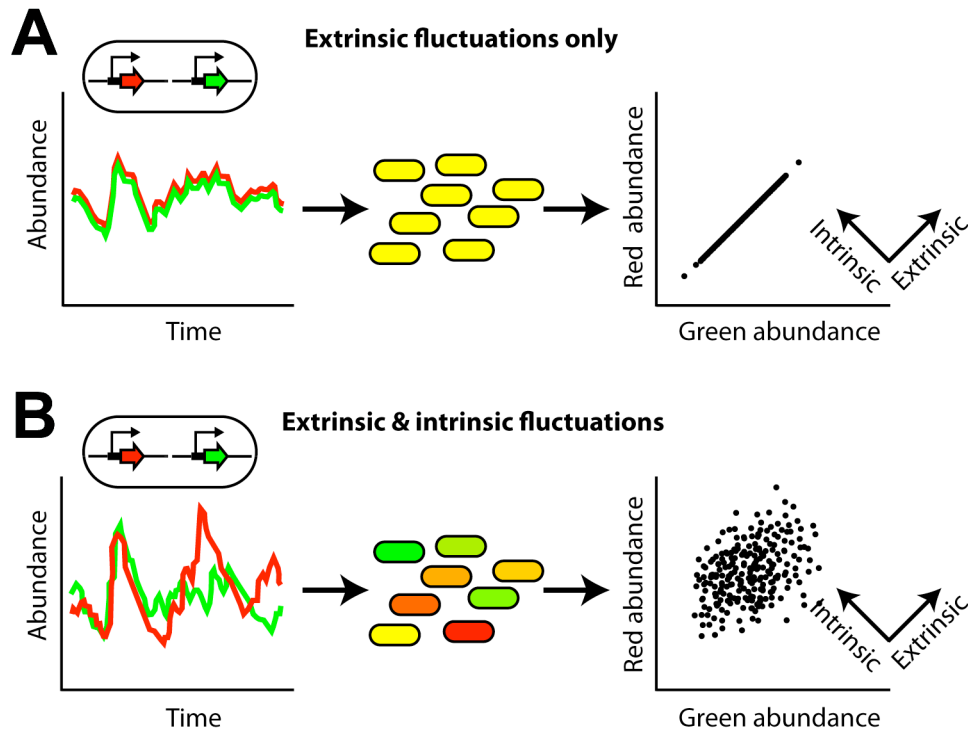
Figure 13. Measuring gene-intrinsic noise. Two almost identical genes, which encode red and green fluorescent proteins, are expressed from identical promoters, and are influenced identically by cell-specific factors, such as gene-regulatory signals. (A) The abundances of the two expressed proteins are perfectly correlated when stochasticity in the biochemical steps that is intrinsic to the process of gene expression (gene-intrinsic noise) is absent and the effects of intracellular heterogeneity are negligible. A scatter plot of protein abundance that was obtained from a 'snapshot' of a cell population contains points that are only on the diagonal. [Adapted and modified from Kaern *et al*. (93).] (B) Asynchronous protein abundances in the presence of gene-intrinsic noise are shown. Because the biochemical steps in the expression of the two genes are independent, gene-intrinsic noise causes the number of expressed proteins to differ, giving rise to a scatter plot that contains off-diagonal points. Evaluating the differences in expressed protein abundance within individual cells, and averaging these differences across a sufficiently large cell population can therefore provide a measure of the absolute magnitude of gene-intrinsic noise. [Adapted and modified from Kaern *et al*. (93).]
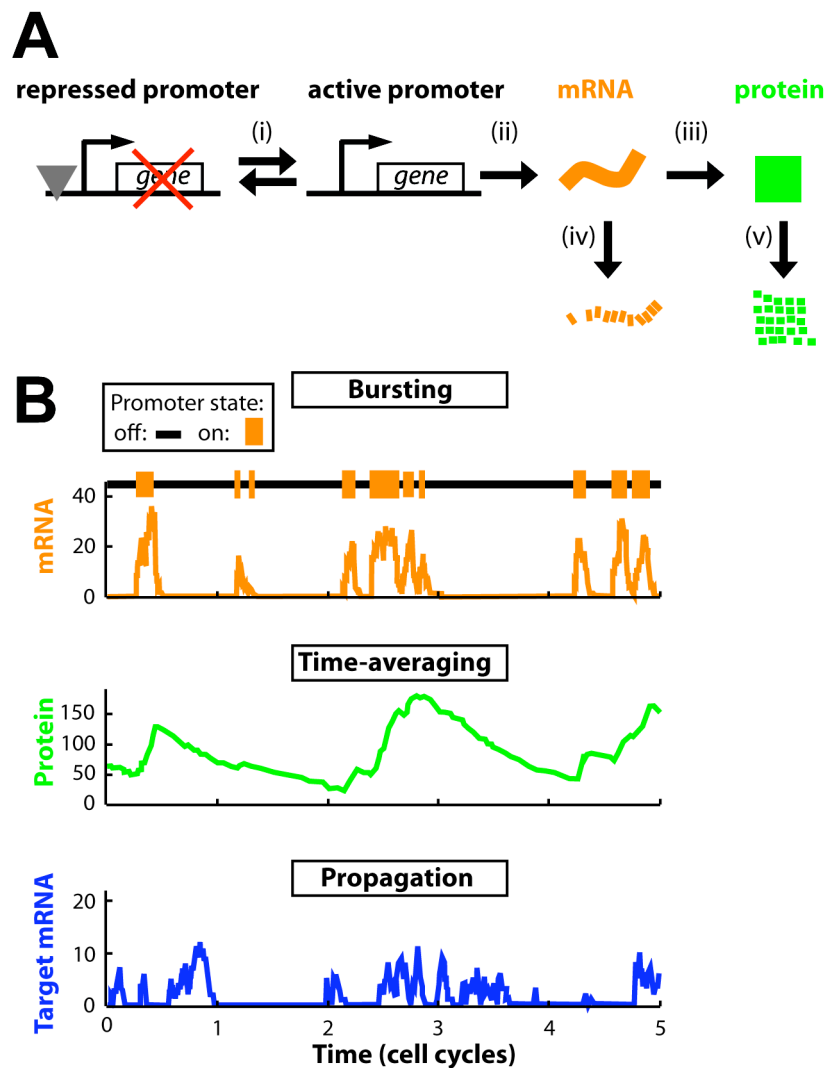
Figure 14. Gene expression noise is ubiquitous, and affects diverse systems at several levels. (A) A model of the expression of a single gene. Steps i-v represent several biochemical reactions, which are associated with mRNA and protein production, transitions between promoter states and the decay of mRNA and protein. These reactions involve binding and dissociation events that occur at random at the molecular level [adapted and modified from Kaern *et al.* (93).]. (B) Mechanisms that shape noise in gene expression. Noise is characterized by bursty expression of mRNA (top). Proteins typically have longer lifetimes than bursts, leading them to timeaverage or 'buffer' these bursts (middle). Finally, noise in one gene can propagate to generate further noise in the expression of downstream genes (bottom). [Adapted and modified from Eldar *et al.* (60).]

Elowitz and coworkers were among the first to experimentally distinguish between the contributions of each type of noise in *E.coli* (61). Hereto, they measured green or red fluorescence from identical promoters in individual, double-labelled cells concluding that uncorrelated fluctuations resulted from intrinsic noise, whereas correlated fluctuations reflected extrinsic noise (Figure 13) (61, 167).

While some of the measured fluorescence variability can be attributed to experimental error, a substantial part of measurable noise stems from random binding and dissociation events that occur between discrete reactants at the molecular level (1, 61, 95, 156, 167). Such biochemical noise tends to increase when the number of interacting reactants decreases, which is commonly known as the "finite number effect".

Generally, gene expression noise alone is not sufficient to cause heterogeneity at the phenotypic level. Often, phenotypic heterogeneity requires components such as feedback and non-linearity allowing the establishment of a bistable system with two stable expression patterns (27, 60, 125, 130).

While feedback may tip such a system from one stable state to another, non-linearity sets the sensitivity at which such a switch occurs. Examples of feedback designs that may favour bistability are the positive feedback loop (Figure 16) or the double negative feedback loop (Figure 15A) (58). In both network architectures an above-threshold fluctuation of gene expression results in the flipping of the gene from an "OFF" to an "ON" state (Figures 14B). The gene expression threshold is determined by non-linear activity levels of regulatory molecules, which can be caused either by cooperative binding or by molecular titration of these molecules. In cooperative binding the regulatory molecules become active only after assembly of identical subunits into multimers (27). In molecular titration regulatory molecules loose functionality when sequestered into inactive hetero-complexes by titrating molecules (27). In both cases a small fold change in input generates a larger fold-change in output (Figure 18C) (125).

However, it should be noted that not in all cases of population heterogeneity at the phenotypic level, must bistable or multi-stable switches between stable states serve as a basis. At least two studies showed that short-lived transcription factors and stochastic fluctuations in the expression of the transcription factors may be enough to create phenotypic diversity (152, 197).

The phage lambda bistable system. One of the best known phenotypes of a genetic switch system are the alternative lytic and lysogenic states of the bacterial virus lambda (154, 161). Although in this system the switch is only to a very minor extent influenced by biochemical noise and almost exclusively induced by external cues (e.g. UV light), the

lambda bistable system is a well-characterized example of a switch based on double negative feedback-loops. Bacteriophage λ infects *E. coli* with two outcomes: In the lytic pathway, it reproduces quickly, thereby killing the host and releasing phage particles. In the lysogenic pathway, it integrates into the host genome and remains dormant (154). Governing the alternative lytic and lysogenic states of phage lambda is a double negative feedback circuit that consists of two antagonistic transcriptional repressors CI and Cro (Figure 15A). At high concentrations the CI repressor inhibits transcription of the *cro* gene and other genes which are involved in lytic growth. Simultaneously CI activates its own transcription. As a result the phage is held in a dormant, lysogenic state (Figure 15B). *Vice verca*, high levels of the Cro repressor inhibit transcription of the *ci* gene. Further, Cro, too, activates its own transcription. As a result, genes involved in lytic growth are freely expressed (Figure 15C), phage DNA is replicated and progeny is produced. Thus, a feedback system with CI and Cro mutually inhibiting each others transcription while stimulating their own transcription, puts into place a threshold of repressor concentrations that, if surpassed, acts as a switch between lysogeny and the lytic state. Negative auto-regulation at high CI concentrations serves the additional purpose of preventing CI repressor concentration from raising too much. This ensures that the lysogen stays poised to respond to an inducing signal (such as UV irradiation) and that the lysogeny state continues to self-perpetuate. At the heart of the CI and Cro feedback systems lie recruiting reactions that involve interactions between molecules of repressor proteins, between repressor and RNA polymerase, and between repressor and DNA. In the lysogen state, a CI-DNA interaction in form of a repressor dimer occurs, preferentially at two of three adjacent and similar operator sequences (operator sites 1-3). Once the CI-dimer is bound to operator site 1, which has a higher binding affinity than operator site 2, this dimer recruits via protein-protein interaction another dimer that then also binds to DNA at operator site 2. In this way, a repressor protein complex is cooperatively formed, and able to recruit RNA polymerase to the CI promoter. Transcription of the *ci* gene can now follow, whereas that of the *cro* gene remains repressed (Figure 15B). However, when operator site 3 is occupied by CI, RNA polymerase is obstructed preventing transcription of the *ci* gene. CI binding to operator site 3 only occurs when its concentration reaches a high level. This negative autofeedback of CI is thought to have evolved to avoid locking the cell into the lysogenic state (154, 161). Operator site 3 can also be occupied by the Cro repressor (Figure 15C). Cro binds to operator site 3 when cleavage of CI occurs as a result of UV radiation and DNA damage. In this case the transcription of *cro* and other lytic genes is initiated. Later in the lytic program, Cro slows down transcription of its own gene by binding to operator sites 1 and 2 (Figure 15C) (154, 161).
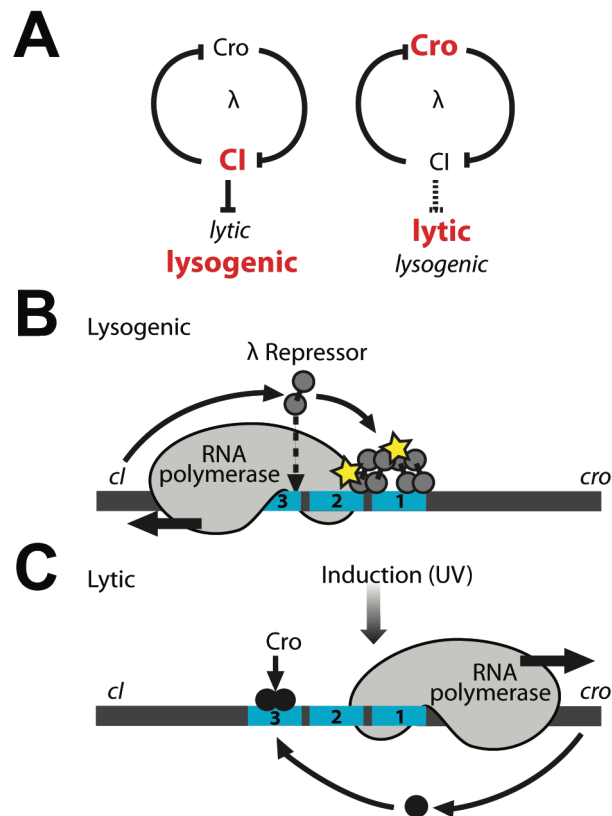
Figure 15. **The lambda switch.** (A) Classic example of the double -negative circuit governing the alternative lytic and lysogenic states of phage lambda. When the lambda repressor CI is at high levels it represses the gene for the Cro repressor and genes involved in lytic growth. Hence the phage is held in the dormant, lysogenic state. Conversely, when Cro is at high levels it represses the gene for CI under which condition genes involved in lytic growth are freely expressed. [Adapted and modified from Losick (125).] (B) In a lysogen, λ repressor preferentially occupies two adjacent operator sites, labeled 1 and 2. This preferential occupancy is determined by two factors: site 1 has the highest affinity for repressor, and a λ repressor dimer binds there cooperatively with another repressor dimer binding site 2. In this state, repressor activates transcription of its own gene (which proceeds leftward in the figure) as it represses transcription of the cro gene. With lower efficiency, λ repressor also binds the weak site 3 and thereby turns off transcription of its own gene. The yellow stars indicate protein-protein contacts of about equal strengths, one mediating cooperative binding of repressor dimers, the other mediating recruitment of RNA polymerase by repressor. (C) UV irradiation results in cleavage of repressor and the onset of transcription of cro and other lytic genes. Cro binds the same three operator sites, but in an order opposite that of repressor: it first binds site 3 and turns off expression of the repressor gene. Later in the lytic cycle, Cro decreases or stops transcription of its own gene by binding sites 1 and 2. [Adapted and modified from Ptashne (161).]

Competence. Bacterial competence development is presented here as an example to understand the interplay between key drivers of population heterogeneity including gene expression noise, nonlinearity-based thresholding, and feedback network architecture (Figure 16). It should be noted however that competence development cannot be regarded as a fate-sealing bistable system since the transition from the vegetative state to the competent state is a reversible process.

Competence development is the conditioning to take up extracellular DNA. Key to competence development in *Bacillis subtilis* is the master regulatory protein ComK, which at specific concentrations above a 'threshold' is able to induce its own transcription (210). A switch to high expression of *comK* results in both the activation and repression of genes involved in competence development, such as *comG* and *comS*, repectively, eventually leading to a stable competence state (Figure 16). However, the state of competence is only transient since ComK is targeted for degradation by the MecA/ClpP/ClpC complex of proteases. Although the MecA/ClpP/ClpC complex tends to degrade ComK, it is kept at bay by the presence of a competitor target peptide called ComS. Only when ComS levels decrease as a result of degradation and are not replenished due to ComK repression, nothing stops degradation of ComK to occur. As a result, ComK reaches sub-threshold concentrations and the cell transits back from competence to the vegetative state. (90, 191). As different members of the population switch between the "ON" and "OFF" competence state, a steady state in population heterogeneity results, which leads to competence for DNA uptake in ca. 10% of the cells of a population (90). While the spontaneous nature of the switch is thought to be driven by intrinsic, random fluctuations (noise) in ComK expression levels (192), the proportion of the competent cell fraction depends on the ComK concentration threshold level. This relationship can be attributed to the non-linear characteristic of ComK activity levels, which stem from both the cooperative nature of the ComK active state as a multimeric transcription factor (211), and ComK titration by the MecA/ClpP/ClpC complex of proteases which rapidly bind to and degrade ComK. Considering that DNA acquisition might represent risk and chance at the same time, the maintenance of a subpopulation competent for DNA uptake can be thought of as a typical bet-hedging strategy. While new DNA might serve as an advantageous blueprint for rapid adaptation to a changing environment, it potentially facilitates dangerous intake of foreign DNA such as viruses. By only committing part of a population to the DNA-uptake gamble, the population avoids risking extinction while still keeping open a backdoor to ameliorate its adaptive potential. The observed variability in the durations of competence in *B. subtilis* might reflect an adaptation to a broad range of extracellular DNA concentrations (60).
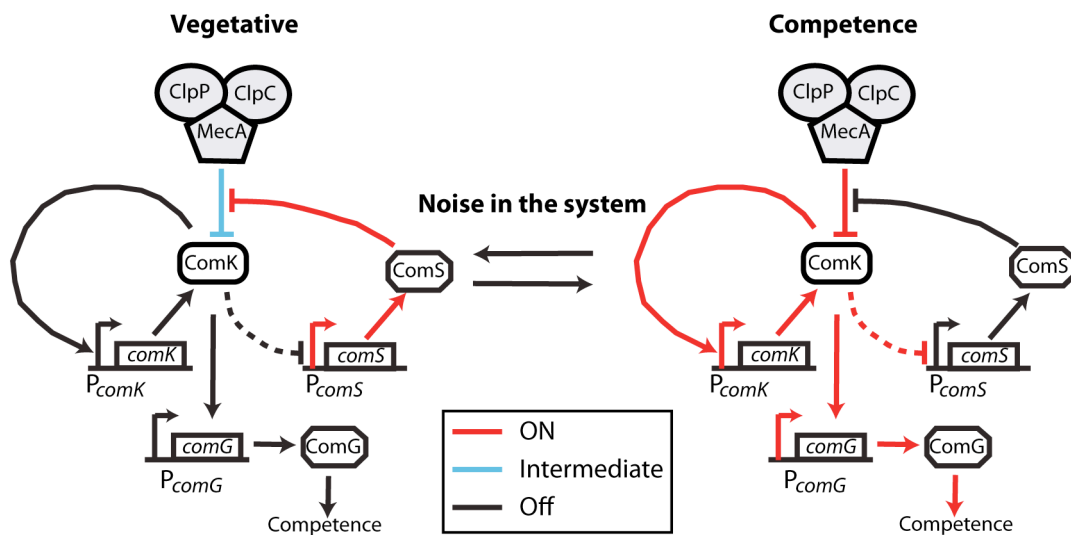
Figure 16. The DNA uptake competence cycle in *Bacillus subtilis*. Noise within the system drives the transition from the vegetative state to the competent state. [Adapted and modified from Johnston and Desplan (90).]

Sporulation. Bacterial competence development in *Bacillus subtilis* is an example of bistability based on a bi-directional cellular differentiation process: reversible transitions from the competent state to the non-competent state occur, yielding sub-populations of each state. In contrast, another type of bistability in *Bacillus subtilis*, sporulation, is based on a unidirectional and therefore irreversible differentiation process (185, 211). Once a cell has fully committed to sporulation, its fate is sealed; As part of a last resort survival strategy and triggered in response to mild nutrient limitation and high cell density, an elaborate developmental program is on the way that after the early stages inevitably culminates in the formation of a highly resistant endospore. The endospore is released by mother cell lysis and can remain dormant for many years while retaining the potential to germinate and resume growth once conditions become favourable again. Because not all cells in a population enter spore formation simultaneously, two distinct sub-populations result: a spore forming (ca. 20%) and a non-spore forming population (212). At the heart of this phenotypic bifurcation lies the master regulator Spo0A, which when phosphorylated is potentially able to initiate the sporulation cascade via direct control of the activity of more than 100 genes (49). Phosphorylation of Spo0A occurs via a multicomponent phosphorelay including the primary kinases KinA and KinB, which both feed phosphoryl groups into the relay, and two intermediate phosphotransferases, Spo0B and Spo0F, which subsequently transfer phosphoryl groups to Spo0A (Figure 17) (49). It is the net phosphate availability of this phosphorelay, that determines whether or not a high enough cellular concentration (threshold level) of phosphorylated Spo0A is reached to initiate sporulation (49). Cell-to-cell variability in phosphorelay phosphate availability represents thus a key driver for sporulation heterogeneity. Three main factors influence phosphorelay phosphate availability: expression levels of the phosphorelay genes KinA, KinB and SpoF, negative regulation of phosphorelay activity via dephosphorylation by the aspartyl-phosphate phosphatases RapA and Spo0E, and indirect positive feedback via SigH (49). In SigH-mediated positive feedback, Spo0A-P represses an unstable transcriptional repressor protein called AbrB, which is a repressor of the alternative sigma factor SigH. SigH is an activator of the transcription of *kinA*, *spo0F* and *spo0A* (49). Spo0A-P thus sets up a self-reinforcing cycle that stimulates both its synthesis and phosphorylation. The implication of this feedback loop is its potential to cause a rapid increase of Spo0A-P once a certain threshold phosphate charge has been reached. This would ultimately result in activation of the unidirectional sporulation program. Interestingly, this form of feedback helps, but is not essential to establish heterogeneous initiation of sporulation (49). The role of dephosphorylation in the phosphorelay is thought to provide the link between
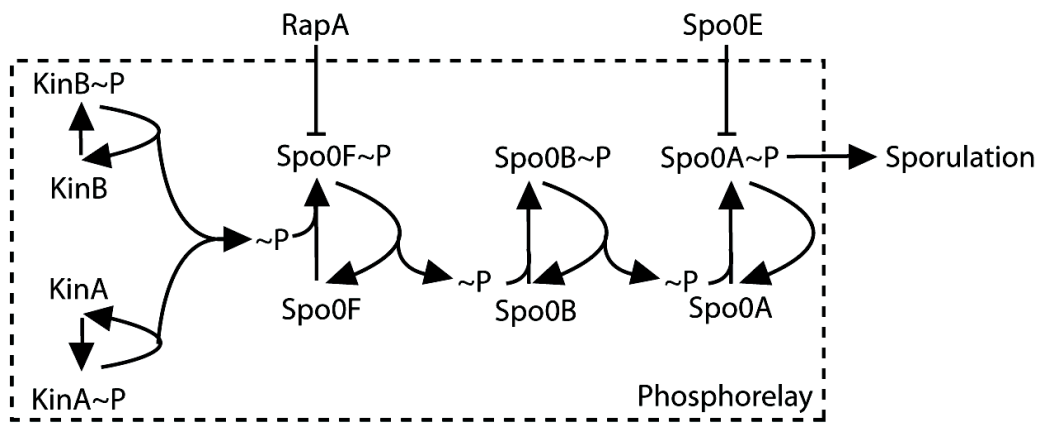
Figure 17. Main phosphotransfer routes within the *Bacillus subtili*s. [Adapted and modified from de Jong *et al*. (49).]

environmental cues and the timing of sporulation. For example, RapA phophatase activity is known to be influenced by growth rate, cell densities, and nutritional status of the cell (49).

The biological significance of sporulation bistability may be manifold. Most obviously, it may constitute a bet-hedging strategy to cope with an uncertain future of either a temporary shortage of nutrients or a prolonged period of famine (185, 211). While non-spore-forming cells are better adapted to a temporary shortage of nutrients because they can rapidly resume growth when nutrients become available again, spore formers are better adapted to a prolonged period of famine because they form dormant spores that can survive long periods of stress thus preserving the clonal lineage. The simultaneous presence of both cell types thus prepares the clonal population for different future environments (185, 211).

An additional benefit of sporulation heterogeneity might be optimisation of resources. As mother cells lyse to release endospores, nutrients are released that can stimulate diauxic growth of non-sporeformers. Thus, heterogeneity in the timing of spore formation allows utilization of these resources that would otherwise be lost (210). However, more importantly in this respect is that spore forming cells that are not yet irreversibly committed to sporulation, exhibit a behaviour referred to as cannibalism: spore forming cells produce extracellular toxins that cause extensive lysis (~70% of viable cells) in their non-spore forming siblings. The extracellular toxins are expressed in a Spo0A-dependent manner from two operons, *skf* (sporulating killing factor) operon and *sdf* (sporulating delay protein) operon, that also express immunity factors protecting spore forming cells from self-lysis by the toxins they produce (41, 78, 123). The resulting release of nutrients from lysed cells serves as a signal to prevent sporulating cells from continuing into the sporulation program. This allows sporulating cells that have not yet passed the point of no return in the otherwise unidirectional sporulation program, to quickly revert to the vegetative mode and resume growth should new nutrients arrive (41, 212). Another hypothesis is that nutrient release through cannibalism maintains a minimal nutrient concentration required for the completion of sporulation, a long lasting, energy-intensive process (about 7 hours at 37°C) (41).

Persistence. Another prominent example of population heterogeneity, similar to sporulation in that resistance to harsh environmental conditions is brought about by retardation of growth, is the bet-hedging strategy of bacterial persistence. In 1944 Bigger noted that treatment of *Staphylococcus aureus* with penicillin resulted in rapid killing of the majority of cells but with a small subpopulation of typically one per million cells (24) showing prolonged survival in the presence of the antibiotic (14, 24). Mutation as the primary cause of resistance was ruled out because when the resistant sub-population was re-cultured and subjected to treatment of antibiotic exposure as before, again a small fraction of survivors appeared.
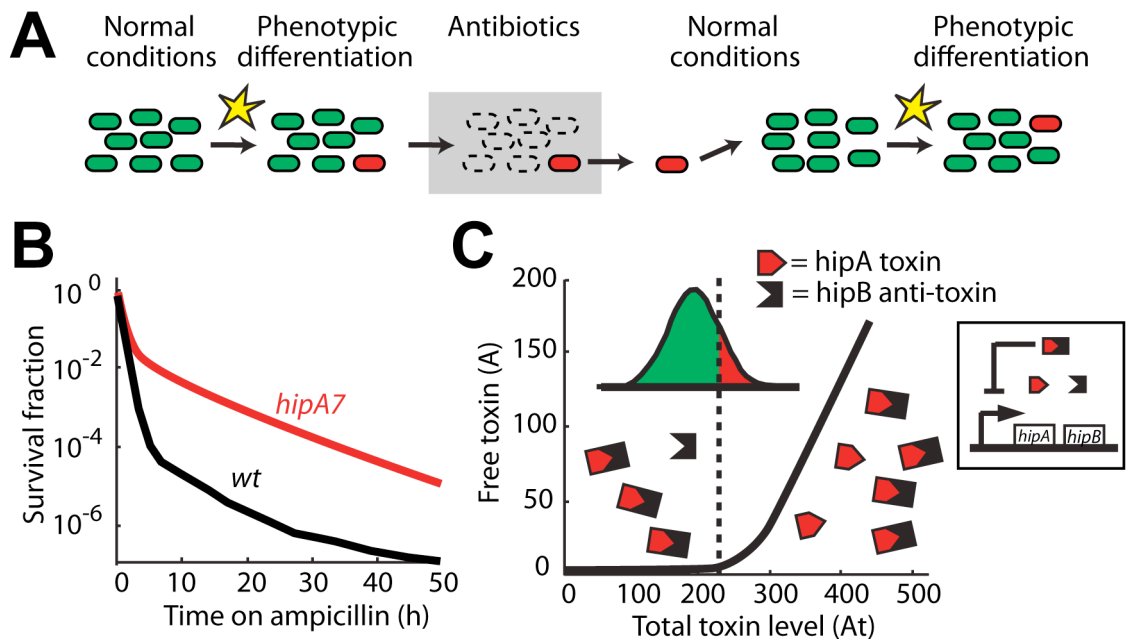
Figure 18. Stochastic bet-hedging in bacteria generates population diversity. (A) An external condition (e.g. starvation) triggers (yellow star) the phenotypic differentiation of the population into a small, stochastic subpopulation of dormant type 1 bacterial persisters (red). Dormant persisters then survive adverse conditions such as antibiotic treatment. Upon return to normal growth conditions, these cells divide and re-establish the population. Finally, a new persister subpopulation is determined. (B) Survival curve of wt *E. coli* (black) and high persistence mutant (red). At t = 0, overnight cultures were diluted into fresh medium containing ampicillin. Note that the initial killing of both strains is similar, indicating that the mechanistic interaction with the antibiotic is similar. Tailing is observed in both cases, but to a greater degree for the high persistence mutant, due to the higher fraction of dormant bacteria. [Adapted and modified from Balaban *et al*. (15).] (C) Threshold-based persistence. The smooth distribution of expression of the toxin (At) is transformed by the threshold (dotted line) into a bi-modal distribution of normally growing cells (green) and persistent cells (red). Inlet: Schematic organization of a typical toxin-antitoxin network motif. The toxin (HipA) is co-expressed with its antitoxin (HipB). Together the proteins form a complex that is not toxic and represses their expression. [Adapted and modified from Rotem *et al*. (171).]

Therefore, Bigger hypothesized that the observed antibiotic-tolerant fraction of cells must have been made up of genetically identical, but slow or non-growing cells, which he continued to refer to as persisters (24). Considering that antibiotics generally target processes associated with active growth such as cell wall biosynthesis, transcription, translation, or DNA replication, resistance by slow growth was indeed a plausible hypothesis (47), which was recently confirmed by a series of state-of-the art, single cell experiments by Balaban and coworkers (15). In these experiments an *E. coli hipA* mutant with a 1000-fold increased rate of persister cell formation (in comparison to the wild type) (145) was used. Together with a clever designed time-lapse imaging technology entailing specialized microfluidics incubation devices, individual persister cells could be observed, which displayed spontaneous entering and exiting of periods of reduced growth. Two types of persisters became evident. Type I persisters formed in response to a external triggers such as stationary phase (72), whereas, Type II persisters formed stochastically and continuously during population growth (72).

While the mechanism for type II persistence remains elusive at this point, type I persistence seems to be regulated by a threshold-based, titration-based mechanism involving the *hipBA* antitoxin toxin (AT) module (171). AT-modules typically consist of a co-expressed antitoxin and toxin pair, with the former neutralizing the toxic effect of the latter (74). AT-systems can be involved in many different cellular processes such as plasmid stabilisation, programmed cell death, modulation of the global levels of translation and replication during exposure to nutritional stress (74). In the *hipBA* AT-system, binding affinity of the toxin HipA to its antitoxin HipB causes an increased frequency of persistence. Consistent with the titration hypothesis that HipA and HipB neutralize each other, HipA was shown to be active only once a threshold of expression was surpassed, which in turn was dependent on the concentration of HipB. Although it was shown that free HipA toxin was responsible for initiating and maintaining the persistence state (Figure 18), its exact mechanism remains unclear (72). HipA interaction with other key regulatory proteins interfering with translation and transcription is likely.

Positive feedback, or double negative feedback, as mentioned above, are common components in bistable systems as they help maintain alternative stable states. In this respect it is interesting to note that HipA is thought to positively regulate its own expression. Another important factor in bistable systems is non-linearity, which helps to create a border between On and Off stable states. Heterodimer binding of the HipBA toxin antitoxin is innately ultrasensitive and therefore potentially represents an important component in bistability (27).

Lactose utilisation. Another classic example of population heterogeneity (showing that genetically identical cells in the same environment can exhibit different phenotypes) is the lactose utilisation system in *Escherichia coli* (115). The genes for lactose metabolism (*lac* genes) are fully expressed for every cell in a population under high concentrations of inducer. However, at moderate inducer concentrations, the *lac* genes are fully expressed in only a fraction of a population. This phenomenon of population heterogeneity was first discovered by Novick and Weiner in 1957, who hypothesized that switching from the uninduced state to the induced state must occur through a single rate limiting molecular process (149). Only recently, with the advent of sophisticated fluorescent marker technology, could this hypothesis be experimentally confirmed. Facilitated through molecular-level sensitivity in fluorescence microscopy coupled with time-lapse imaging, Choi and colleagues (40) were able to show that the "ON" switch of *lac* expression is triggered by stochastic dissociation of the LacI repressor from its DNA target sites (Figure 19). LacI is a tetrameric transcription factor that, upon binding to two *lac* operators, causes DNA loop formation in the promoter region of the *lac* operon (40). *Lac* gene transcription can only initiate if the loop structure disappears, which it can do in two ways. Either the LacI repressor dissociates from both operator sites (complete dissociation), or only from one operator site (partial dissociation) (40). Under high intracellular inducer concentration, it is always the complete dissociation of the LacI repressor that occurs. In this case, repressor dissociation is active and the repressor remains sequestered after dissociation by the inducer (40). Also, under high intracellular inducer concentrations a homogenous expression of the *lac* genes in the population is seen (40). In contrast, under low or intermediate intracellular inducer concentrations, LacI stochastically dissociates from the operators independently of the inducer. Under these conditions partial repressor dissociation occurs much more often than complete repressor dissociation (40). Furthermore, because in partial dissociation the repressor remains in proximity to the vacant operator binding site (as opposed to during complete dissociation) rebinding occurs rapidly (40). Consequently, little time for the generation of mRNA (a single copy) is available, only allowing for a small burst of protein (40). Although complete repressor dissociation under low or medium inducer concentrations occurs rarely, when it does, no quick rebinding of the repressor is possible. As a result multiple mRNAs can be transcribed resulting in a large burst of Lac proteins (40). Amongst the Lac proteins is also the LacY transporter, which allows influx of further inducer (40). As a result of increasing inducer concentrations, active displacement of LacI repressor from the *lac* operators occurs and thus a positive feedback loop is initiated that ultimately leads to a switch in phenotype from the Off to the On-state of lactose utilisation (Figure 19) (40).
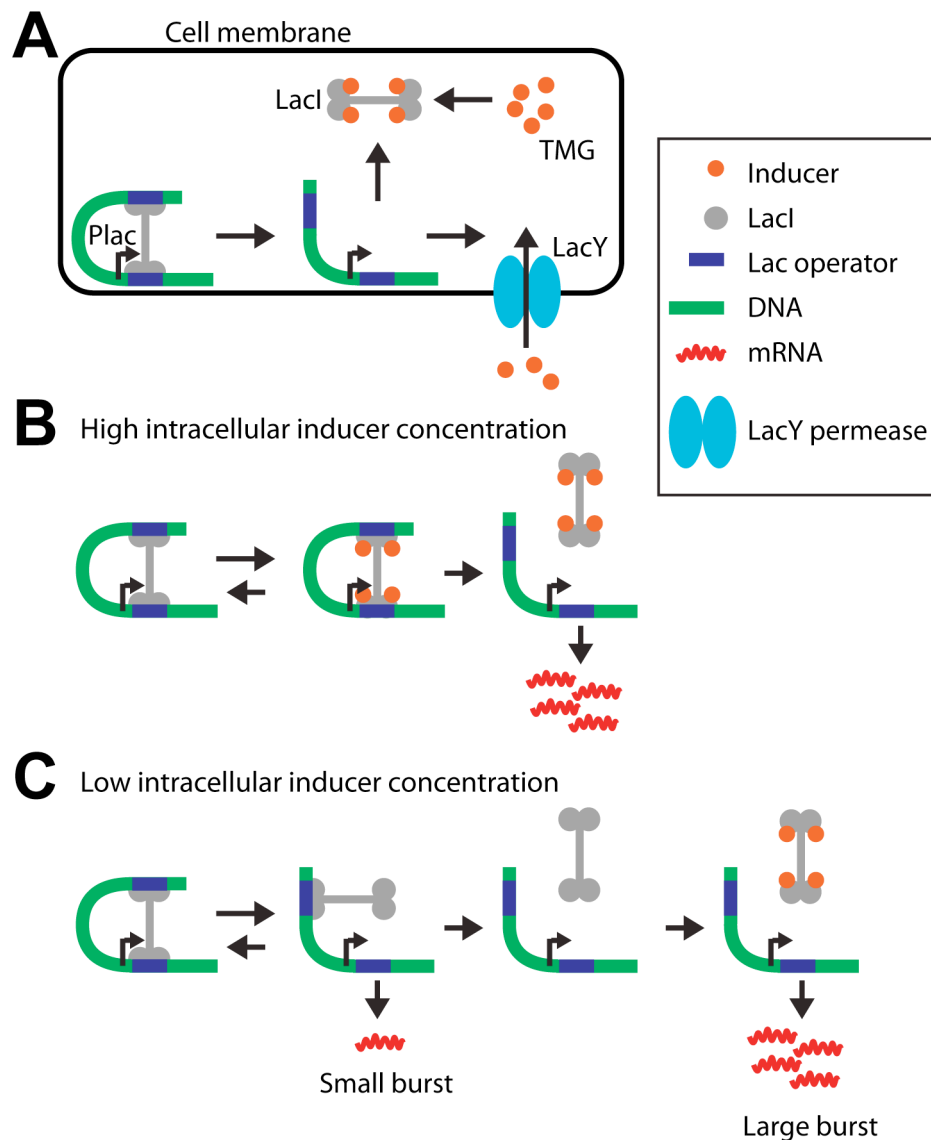
Figure 19. The expression of lactose permease in _E. coli_. (A) The repressor LacI and permease LacY form a positive feedback loop. Expression of permease increases the intracellular concentration of the inducer TMG, which causes dissociation of LacI from the promoter, leading to even more expression of permeases. Cells with a sufficient number of permeases will quickly reach a state of full induction, whereas cells with too few permeases will stay uninduced. (B) A high concentration of intracellular inducer can force dissociation of the repressor from its operators. (C) At low or intermediate concentrations of intracellular inducer, partial dissociation from one operator by the tetrameric LacI repressor is followed by a fast rebinding. Consequently, no more than one transcript is generated during such a brief dissociation event. However, the tetrameric repressor can dissociate from both operators stochastically and then be sequestered by the inducer so that it cannot rebind, leading to a large burst of expression. [Adapted and modified from Choi _et al_. (40).]

Population heterogeneity based on aging

Besides phase variation and biochemical noise, aging (or senescence), too, might play a role in driving bacterial individuality. Recent studies provided evidence that age affects cell size and timing of cell division (6, 23), growth rates (187, 190) and survival (213) and the localisation of proteins (159) and protein aggregates (118, 220). Hence, a picture emerges in which bacteria exhibit individuality for demographic (and therefore deterministic) rather than stochastic reasons.

Generally, aging can be defined as a decline in reproductive rate and survival due to deterioration of function with age (169). This interpretation is founded on the fact that structural damage accumulates in tissues, cells and subcellular structures as an inevitable consequence of vital activities (2). Modelling suggests that asymmetric differentiation between parental and progeny cells may readily evolve as a strategy to cope with structural damage (3, 63, 89, 215). In this case the damaged or aged structures are not equally propagated from parent to offspring (Figure 20C). Rather, the majority of damaged structures segregates to the parent, while the offspring receives newly synthesized structures and therefore emerges rejuvenated. Thus, aging can be seen as a division-of-labour strategy to cope with accumulating damage where parent fitness is sacrificed in favour of offspring fitness (150).

Indeed, multiple examples of asymmetrical aging exist in the microbial world. It has been shown that yeast (16) use cues acting as indicators of polarity to faithfully segregate damaging protein aggregates to the ageing cell lineage while leaving the other lineage rejuvenated and free of such damage (151). Although asymmetrical distribution of subscellular structures between cells during cell division is a common/known feature in many bacteria (180), a link between asymmetric differentiation and aging has so far only been shown in a few systems (4, 6, 23, 212). The first bacterium in which aging was demonstrated is the alphaproteobacterium *Caulobacter crescentus* (4). In this organism asymmetric division yields division yields two phenotypes: a stalked, chromosome replication-competent parent cell and an offspring replication-noncompetent swarmer cell. While the stalked cell remains attached to a solid surface by virtue of a polar and strongly adhesive holdfast, the swarmer cell is equipped with a polar flagellum allowing for dispersal and colonization of new territory (Figure 20A). Stalked cells give rise to new swarmer cells that eventually themselves differentiate into stalked cells. The *Caulobacter* life cycle thus entails a 'juvenile phase' in which swarmer cells must go through a period of differentiation before becoming capable of reproducing (108).
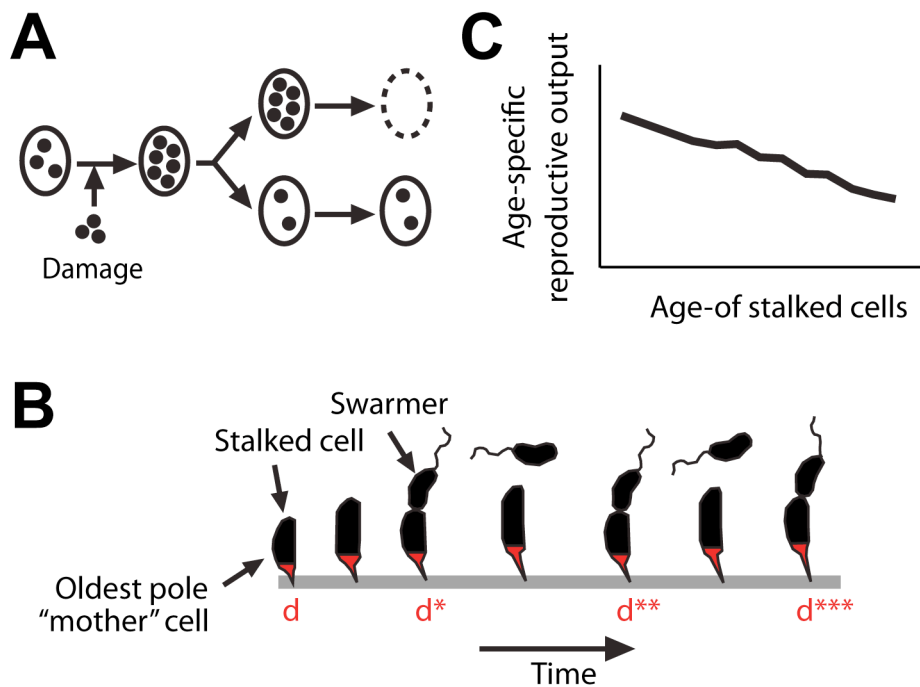
Figure 20. Bacterial Aging. (A) Concept of damage-related aging: damage is unequally distributed in the progeny of a dividing mother cell. If such damage is biased towards the older cell pole in each generation, cell-pole related aging occurs. Ovals, cells; black dots, damage; dotted oval, cell death. [Adapted and modified from Ackermann et al. (2).] (B) Schematic of the life cycle of *C. crescentus*. d, oldest pole mother cell; Number of red stars, number of times the old pole mother cell has witnessed a cell division. [Adapted and modified from Ackermann et al. (4).] (C) Schematic of reproductive output of *C. crescentus* (progeny produced per stalked cell and hour) as a function of age (4).

C. crescentus dimorphic life cycle strategy was elegantly exploited by Ackermann and colleagues to monitor hundreds of cell division events over long periods of time (130 divisions over a time-period of 300 hours) (4). In this case a flow chamber was prepared in which the swarmer cells were continuously flushed away by flowing medium, while stalked cells stayed attached to the chamber walls and thus could be monitored (Figure 20A). Combining this set-up with automated time lapse imaging, it became clear that with each division, stalked cells required progressively longer times to produce swarmer cells and hence were aging (Figure 20B).

But how about aging in symmetrically shaped cells? Recording of bacterial aging in morphologically (visually) symmetrically dividing cells such as rod-shaped *E. coli* is more tricky than in asymmetrically dividing cells such as *Caulobacter* (187). Because division in *E. coli* occurs at mid-cell, old and new cell poles can only be identified by looking back in division history with help of time-lapse movies (Figure 21A). The number of times an old pole has 'witnessed' cell division and therefore birth of a new cell pole during septation, can then be taken as a measure of cell age. Lineage trees are particular revealing when asking if a physical trait is inherited or related to cell pole age (Figure 21A). Alternatively, fluorescent marker fusion proteins can be employed that localize at cell poles. Although this technique might help in distinguishing old from new cell poles without the need of documenting growth history, it will be limited to age determination beyond only a few cell divisions (159).

The first demonstration of aging in visibly symmetrically dividing cells was reported by Stewart and colleagues (187). This group recorded the history and physical parameters of single cells in *E. coli* monolayer microcolonies. They found that cells of old pole age showed diminished growth rate, decreased offspring production and increased incidence of death in comparison to cells of young pole age. The discovery of aging in *E. coli* came as a surprise because it was widely accepted that organisms reproducing by binary fission were functionally immortal due to lack of visual asymmetry. However, this latter assumption does not hold true beyond visual perception since on a physiological level all bacterial cells can be considered inherently asymmetric for at least two reasons. Firstly, it has been shown that DNA segregates non-randomly to old or new cell poles (221). Secondly, any cell constituent with limited diffusion and a long half-life may be expected to accumulate at the old pole (187). For example, in *E.coli* it is known that components of the cell wall turn over slowly and are conserved in the poles where they are formed (50).
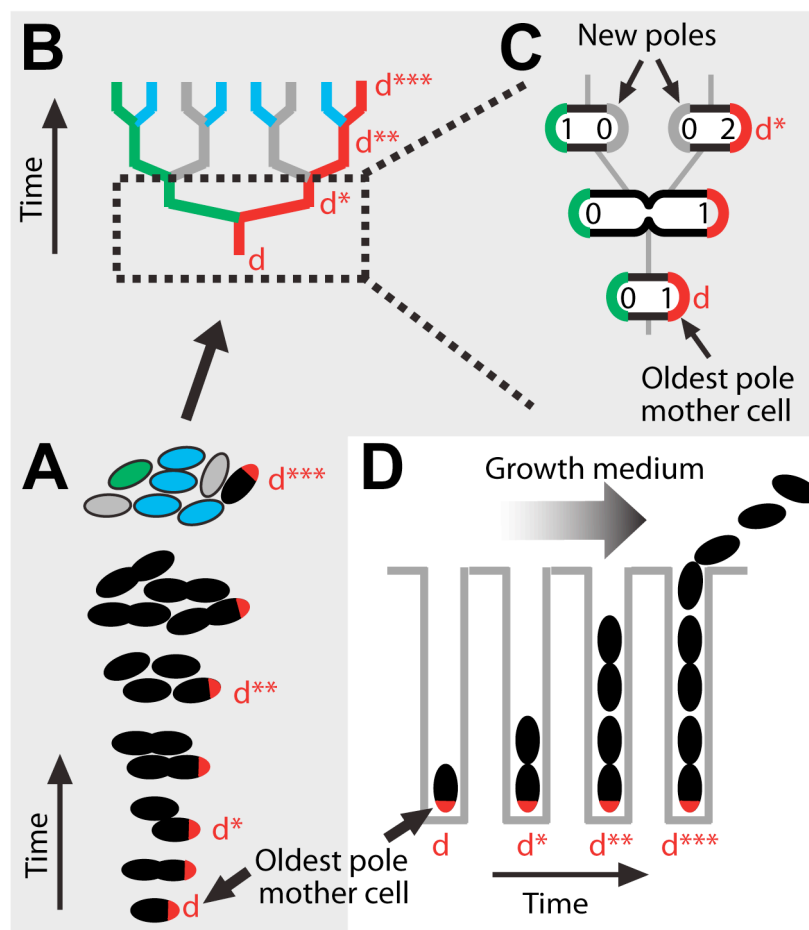
Figure 21. Schematic comparing microcolony- and microfluidics-based time-lapse observation systems. (A) Schematic of microcolony growth over time. (B) Lineage tree constructed from (A). Each cell division is represented by branching events in the lineage tree. (C) Magnification of dotted part of (B) showing assignment of cell-pole ages. Individuals at different ages are represented by rounded rectangles: Red pole, oldest pole; green pole, second oldest pole; grey pole, new pole; number within rectangle, pole age corresponding to the number of divisions the oldest pole has witnessed. (D) Schematic illustration of the microfluidic mother machine proposed by Wang and colleagues (213). The oldest pole mother cell is trapped at the end of the growth channel. Red pole, oldest cell pole; d, oldest pole mother cell; Number of red stars, number of times the old pole mother cell has witnessed a cell division.

Yet, another study on *E.coli* cell pole aging, this time by Wang and colleagues using a microfluidics device termed 'mother mashine' (213), as opposed to the microcolony growth procedure that was used by Stewart *et al.* (187) (see Figure 21), was not able to detect age related bias in growth rate. Rather, Wang and collegues saw a very robust mechanism of growth over hundreds of generations in cells inheriting the old pole (213). Rang *et al.* (166) reconciled the contradicting results of Wang *et al.* (213) and Stewart *et al.* (187) on basis of a population genetic model. They attributed the differences in aging in the two experiments to the differential experimental conditions used, respectively, claiming that insufficient extrinsic damaging agents were acting on the cells in Wang's experiment (166, 213); In comparison to the microcolony growth set-up from Stewart *et al.* (187), which could only track cells up to eight generations, Wang *et al.* (213) used a microfluidics-based system allowing for online tracking of single cells for up to 200 generations (Figure 21B). Although Wang and colleagues did not find growth-rate based aging (213), they still noted a higher incidence of death in old cells, which they attributed to the accumulation of some unidentified lethal agent, possibly protein aggregation (accumulating damages). Indeed, accumulation of protein aggregates in old cells of *E. coli* had been found previously by Lindner and colleagues (118). Another study showed bistable populations of *E. coli* consisting of damage-enriched and a damage-free population in stationary phase; whereas damage-free cells remained reproductively competent, damage-enriched cells eventually became non-culturable (52).

In another study of visibly symmetrically dividing rod cells, Bergmiller and colleagues showed that age in *Methylobacterium extorquens* influences cell size and the interdivision interval (23). In this species cells with young poles are smaller upon division and take longer to divide than cells with older poles and therefore contribute to cell size heterogeneity in the clonal population. This case showed that age differences between individuals represent regular, deterministic processes that affect biological properties and lead to population heterogeneity in another way than noise-based systems.

Another aging phenomenon was recently reported by Aldrich and co-workers (6), who focused on mycobacteria. They found that old-pole cells grew faster and showed longer birth lenghts than new-pole cells. As a result, physiologically distinct subpopulations showed differential susceptibility to antibiotics.

Thus, it can be concluded that while the exact molecular mechanisms leading to aging still remain unknown, more and more studies appear that demonstrate the role of aging in cell-to-cell heterogeneity. Aging might thus represent, next to stochastic mechanisms of cell-to-cell variability generation such as phase variation or biochemical noise-based heterogeneity, another important component upholding a bet-hedging survival strategy.

53

RESEARCH OBJECTIVES AND THESIS OUTLINE

In this dissertation HGT is studied with the general aim to better understand how and why bacteria evolve so quickly. In this way this work hopes to contribute to the knowledge of how to counteract the downsides or enhance the benefits of adaptive processes. To achieve this goal the research of this thesis is focused on the study of ICE*clc*, a model system representing key mediators in bacterial genome evolution.

Whereas previous studies on ICE*clc* have more focused on the molecular aspects of bistability formation (140, 142, 178), excision/integration (70), transfer (142) and regulation (140, 178), the principle research of this thesis was more directed at the establishment of single cell analysis tools and their usage to demonstrate the existence of a number of eloquent steps in transfer competence formation. Indeed, since prokaryotic biology in general is still largely based on the idea of clonal populations of cells which all behave the same, novel, single-cell studies could be expected to reveal a hidden dimension in ICE biology (47, 53, 122).

At the onset of this thesis there was fundamental lack of knowledge in single cell behaviour with regard to ICE*clc* and ICEs in general. It seems that the generally low ratios at which ICEs are transferred from donor to recipient and the methodological obstacle associated to this fact has disencouraged many researchers to investigate ICE-transfer at the single cell level (11, 13, 109, 162). However, such knowledge is crucial for the understanding of the evolutionary success of ICE since even in uniform environments, it cannot be assumed that ICEs behave similarly in all cells of a clonal population (47). More and more cases of bacterial individuality are being discovered in different fields such as prophage activation (154), sporulation (212), competence (90), persistence (15) and lactose metabolism (40), which points to the possibility that bacterial individuality might also play an important role in the regulation of HGT. Indeed, prior to the start of this thesis it could be shown that activation of the P_{int}-promoter leading to excision of ICE*clc* in *Pseudomonas knackmussii* occurs only in a 3-5% of cells (178), which was later suggested to be the result of a bistable activation cascade involving a dedicated program to prepare cells for ICE*clc* transfer (70, 140, 142). Not only ICE*clc*, but a variety of other ICE have self-transmission frequencies in the order of a few percent per donor cell or lower, suggesting that induction of a "transfer-competent" state in some cells in a clonal population of ICE-bearing donors is indeed a common strategy (11, 13, 109, 162).

Therefore, for this thesis, it seemed promising to try characterize differences between ICE-ON (transfer activated) versus ICE-OFF (transfer inactivated) cells at the single cell level, which had never been done before. One major obstacle, however, represented the low

proportion of ICE*clc*-ON cells and the rather weak promoter strengths of the available promoters (P$_{int}$-promoter, P$_{inr}$-promoter) that by fusion to genes of fluorescent marker proteins could indicate the ICE*clc*-ON state. These flaws had to be taken into account when designing single cell methodologies and strategies for this project. Therefore, the aims of this thesis comprise both methodological and fundamental research goals. The main goals of this thesis thus are:

1) **To establish a sound set of methods that can**
    a) **reliably characterize and quantify small subpopulation effects, and**
    b) **follow bacterial cellular growth and ICE*clc*-activation over time and as a function of varying growth conditions;**

2) **To perform fundamental research at the single cell level to**
    a) **provide direct proof for the "transfer competent state" of ICE*clc*-active cells,**
    b) **describe the physiology of transfer competent cells in wildtype and mutant backgrounds to quantify the cost of transfer competence in single cells and at the population level, and**
    c) **determine the genealogy of transfer competent cells in microcolonies to see if the potential for transfer competence is heritable or if there is a link between age and the development of transfer competence.**

The goals of this thesis are targeted in four research chapters. The first two research chapters (Chapter 2 and Chapter 3) are methodological in nature putting into place the tools needed for studying ICE*clc* at the single cell level. Equipped with these tools the following two chapters (Chapter 4 and Chapter 5) then target direct evidence for the state of transfer competence and subsequently aim to tease out differences between transfer competent versus non-transfer competent cells with regard to cell physiology, fitness, heredity, and cell age.

Outlook Chapter 2. A key challenge is the quantification and description of subpopulations. Recently, it could be shown by single cell analysis that activation of the (late) P$_{int}$-promoter leading to excision of ICE*clc* in *Pseudomonas knackmussii* is the consequence of a bistable activation cascade occurring only in part of the cells in stationary phase (140, 178). However, the fraction of ICE*clc* could not be determined in an accurate and reproducible way, which becomes important if it should be taken as a measure to compare different genetic

backgrounds or physiological conditions. In this chapter we revisit this problematic and propose a simple method for the quantification of subpopulations. Parts of this chapter were used for a paper by Minoia and colleagues (140).

Outlook Chapter 3. Time-lapse experiments require a delicate experimental set-up by which model organisms can be tracked, allowing sufficient extraction of qualitative and quantitative data. While many different set-ups have been proposed in the literature (4, 6, 15, 23, 187, 212, 213), only very few exist allowing manipulation of external factors during observation, amongst which are complicated micro-fluidics set ups (6, 15, 213) that cannot track genealogy of whole microcolonies. Here we provide a simple step-by-step procedure, which allows time-lapse imaging of bacterial cell division and cell tracking up to the stage of a monolayered microcolony of a few hundred cells whilst allowing for manipulation of environmental conditions. This set-up provides the basis for Chapter 4 and Chapter 5.

Outlook Chapter 4. Whereas recent studies on ICEs have been of cardinal importance to begin to understand the basic ICE-host control mechanisms, there is a fundamental lack of understanding with regard to ICE-host behaviour at the single cell level, which is crucial for their evolutionary success. We demonstrate here the existence of a transfer-competent state and further show that it not only leads to ICE*clc* transfer, but inevitably inhibits subsequent proliferation of competent donor cells and leads to lysis. We argue that ICE*clc*-induced donor cell division inhibition and lysis is a form of programmed cell-death, but which is fundamentally different from toxin-antitoxin addiction modules (65, 74, 222). We hypothesize that the transfer-competent state may actually have been selected for because it allowed more efficient transfer of ICE*clc*, whereas the large fitness cost the ICE*clc* reproductive inhibition program would impose on the population is only kept at bay by confining ICE*clc* activity to a small proportion of cells.

Outlook of Chapter 5. The objectives of the work presented in Chapter 5 were to study the appearance of tc cells as a function of their life history in microcolonies. In particular, we addressed the following questions: (i) Is tc cell formation dependent on cell pole age with older cells being more prone than younger cells? (ii) Does tc cell formation preferentially occur in specific lineages in a microcolony with, incidentally, lysing sister cells? (iii) Is tc cell formation a function of spatial organisation or density within a microcolony? (iv) Does tc cell formation occur as a consequence of pre-existing detectable biochemical damage?

Through the use of time-lapse microscopy we thus follow the life history of tc-cells within microcolonies. We demonstrate that the age of the cell pole is unlikely to play a role in initiating the ICE*clc* tc program. We also show that initiation of transfer competence is not the result of the physiological state of ancestor cells, or of a close relative (sister cell). In contrast,

tc-cells show higher levels of reactive oxygen species and membrane damage than non-tc cells, but whether this is causing or is the effect of ICE*clc* activation could not be discerned. We find that ICE*clc* activation occurs spatially randomly in a microcolony, which might additionally be important for maximising the chances in a biofilm to contact potential recipients.

REFERENCES

1.   **Acar, M., J. T. Mettetal, and A. van Oudenaarden.** 2008. Stochastic switching as a survival strategy in fluctuating environments. Nat Genet **40:**471-475.

2.   **Ackermann, M., L. Chao, C. T. Bergstrom, and M. Doebeli.** 2007. On the evolutionary origin of aging. Aging Cell **6:**235-244.

3.   **Ackermann, M., A. Schauerte, S. C. Stearns, and U. Jenal.** 2007. Experimental evolution of aging in a bacterium. BMC Evol Biol **7:**126.

4.   **Ackermann, M., S. C. Stearns, and U. Jenal.** 2003. Senescence in a bacterium with asymmetric division. Science **300:**1920.

5.   **Alanis, A. J.** 2005. Resistance to antibiotics: are we in the post-antibiotic era? Arch Med Res **36:**697-705.

6.   **Aldridge, B. B., M. Fernandez-Suarez, D. Heller, V. Ambravaneswaran, D. Irimia, M. Toner, and S. M. Fortune.** 2012. Asymmetry and aging of mycobacterial cells lead to variable growth and antibiotic susceptibility. Science **335:**100-104.

7.   **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev **59:**143-169.

8.   **Aminov, R. I., and M. Roderick, I.** 2007. Evolution and ecology of antibiotic resistance genes. FEMS Microbiol Lett **271:**147-161.

9.   **Andrewes, F. W.** 1922. Studies in group-agglutination. I. The *Salmonella* group and its antigenic structure. J Path Bacteriol **25:**505-521.

10.   **Avery, S. V.** 2006. Microbial cell individuality and the underlying sources of heterogeneity. Nat Rev Microbiol **4:**577-587.

11.   **Babic, A., M. B. Berkmen, C. A. Lee, and A. D. Grossman.** 2011. Efficient gene transfer in bacterial cell chains. MBio **2:**mBio.00027-11.

12.   **Babic, A., A. B. Lindner, M. Vulic, E. J. Stewart, and M. Radman.** 2008. Direct visualization of horizontal gene transfer. Science **319:**1533-1536.

13.   **Baharoglu, Z., and D. Mazel.** 2011. Vibrio cholerae triggers SOS and mutagenesis in response to a wide range of antibiotics, a route towards multi-resistance. Antimicrob Agents Chemother **55:**2438–2441.

14.   **Balaban, N. Q.** 2011. Persistence: mechanisms for triggering and enhancing phenotypic variability. Curr Opin Genet Dev **21:**768-775.

15. **Balaban, N. Q., J. Merrin, R. Chait, L. Kowalik, and S. Leibler.** 2004. Bacterial persistence as a phenotypic switch. Science **305:**1622-1625.

16. **Barker, M. G., and R. M. Walmsley.** 1999. Replicative ageing in the fission yeast Schizosaccharomyces pombe. Yeast **15:**1511-1518.

17. **Barnard, A. M. L., S. D. Bowden, T. Burr, S. J. Coulthurst, R. E. Monson, and G. P. C. Salmond.** 2007. Quorum sensing, virulence and secondary metabolite production in plant soft-rotting bacteria. Philos Trans R Soc Lond B Biol Sci **362:**1165-1183.

18. **Bathe, S., M. Lebuhn, J. W. Ellwart, S. Wuertz, and M. Hausner.** 2004. High phylogenetic diversity of transconjugants carrying plasmid pJP4 in an activated sludge-derived microbial community. FEMS Microbiol Lett **235:**215-219.

19. **Beaber, J. W., B. Hochhut, and M. K. Waldor.** 2002. Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. J Bacteriol **184:**4259-4269.

20. **Beaber, J. W., B. Hochhut, and M. K. Waldor.** 2004. SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature **427:**72-74.

21. **Beaber, J. W., and M. K. Waldor.** 2004. Identification of operators and promoters that control SXT conjugative transfer. J Bacteriol **186:**5945-5949.

22. **Beaumont, H. J., J. Gallie, C. Kost, G. C. Ferguson, and P. B. Rainey.** 2009. Experimental evolution of bet hedging. Nature **462:**90-93.

23. **Bergmiller, T., and M. Ackermann.** 2011. Pole age affects cell size and the timing of cell division in *Methylobacterium extorquens* AM1. J Bacteriol **193:**5216-5221.

24. **Bigger, J. W.** 1944. Treatment of staphylococcal infections with penicillin. Lancet **ii:**497 - 500.

25. **Bose, B., J. M. Auchtung, C. A. Lee, and A. D. Grossman.** 2008. A conserved anti-repressor controls horizontal gene transfer by proteolysis. Mol Microbiol **70:**570-582.

26. **Bose, B., and A. D. Grossman.** 2011. Regulation of horizontal gene transfer in Bacillus subtilis by activation of a conserved site-specific protease. J Bacteriol **193:**22-29.

27. **Buchler, N. E., and M. Louis.** 2008. Molecular titration and ultrasensitivity in regulatory networks. J Mol Biol **384:**1106-1119.

28. **Burrus, V., and M. K. Waldor.** 2003. Control of SXT integration and excision. J. Bacteriol. **185:**5045-5054.

29.   **Burrus, V., and M. K. Waldor.** 2004. Shaping bacterial genomes with integrative and conjugative elements. Res. Microbiol. **155:**376-386.

30.   **Cai, L., N. Friedman, and X. S. Xie.** 2006. Stochastic protein expression in individual cells at the single molecule level. Nature **440:**358-362.

31.   **Cary, S. C., I. R. McDonald, J. E. Barrett, and D. A. Cowan.** 2010. On the rocks: the microbiology of Antarctic Dry Valley soils. Nat Rev Microbiol **8:**129-138.

32.   **Cascales, E., and P. Christie.** 2003. The versatile bacterial type IV secretion systems. Nature Rev Microbiol **1:**137–149.

33.   **Chai, Y., F. Chu, R. Kolter, and R. Losick.** 2008. Bistability and biofilm formation in *Bacillus subtilis*. Mol Microbiol **67:**254-263.

34.   **Chai, Y., T. Norman, R. Kolter, and R. Losick.** 2010. An epigenetic switch governing daughter cell separation in Bacillus subtilis. Genes Dev **24:**754-765.

35.   **Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher.** 1994. Green fluorescent protein as a marker for gene expression. Science **263:**802-805.

36.   **Chao, T.-C., and A. Ros.** 2008. Microfluidic single-cell analysis of intracellular compounds. J R Soc Interface:S139-S150.

37.   **Chen, Q., K. B. Decker, P. E. Boucher, D. Hinton, and S. Stibitz.** 2010. Novel architectural features of *Bordetella pertussis* fimbrial subunit promoters and their activation by the global virulence regulator BvgA. Mol Microbiol **77:**1326-1340.

38.   **Cheng, Q., B. J. Paszkiet, N. B. Shoemaker, J. F. Gardner, and A. A. Salyers.** 2000. Integration and excision of a *Bacteroides* conjugative transposon, CTn*DOT*. J. Bacteriol. **182:**4035-4043.

39.   **Cheng, Q., Y. Sutanto, N. B. Shoemaker, J. F. Gardner, and A. A. Salyers.** 2001. Identification of genes required for excision of CTnDOT, a *Bacteroides* conjugative transposon. Mol. Microbiol **41:**625-632.

40.   **Choi, P. J., L. Cai, K. Frieda, and X. S. Xie.** 2008. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. Science **322:**442-446.

41.   **Claverys, J.-P., and L. S. Havarstein.** 2007. Cannibalism and fratricide: mechanisms and raisons d'etre. Nat Rev Microbiol **5:**219-229.

42.   **Collignon, P.** 2013. Superbugs in food: a severe public health concern. Lancet Infect Dis.

43.   **Cookson, S., N. Ostroff, W. L. Pang, D. Volfson, and J. Hasty.** 2005. Monitoring dynamics of single-cell gene expression over multiple cell cycles. Mol Syst Biol **1:**2005.0024.

44. **Csotonyi, J. T., J. Swiderski, E. Stackebrandt, and V. Yurkov.** 2010. A new environment for aerobic anoxygenic phototrophic bacteria: biological soil crusts. Environ Microbiol Rep **2:**651-656.

45. **Czechowska, K., D. R. Johnson, and J. R. van der Meer.** 2008. Use of flow cytometric methods for single-cell analysis in environmental microbiology. Curr Opin Microbiol **11:**205-212.

46. **Dasch, G. A., E. Weiss, and K. P. Chang.** 1984. Endosymbionts of insects, p. 811–833. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. The Williams & Wilkins Co., Baltimore.

47. **Davidson, C. J., and M. G. Surette.** 2008. Individuality in bacteria. Annu Rev Genet **42:**253-268.

48. **de Jong, I. G., K. Beilharz, O. P. Kuipers, and J. W. Veening.** 2011. Live cell Imaging of *Bacillus subtilis* and *Streptococcus pneumoniae* using automated time-lapse Microscopy. J Vis Exp.

49. **de Jong, I. G., J. W. Veening, and O. P. Kuipers.** 2010. Heterochronic phosphorelay gene expression as a source of heterogeneity in *Bacillus subtilis* spore formation. J Bacteriol **192:**2053-2067.

50. **de Pedro, M. A., K. D. Young, J.-V. Holtje, and H. Schwarz.** 2003. Branching of *Escherichia coli* cells arises from multiple sites of inert peptidoglycan. J Bacteriol **185:**1147-1152.

51. **Dehio, C., S. D. Gray-Owen, and T. F. Meyer.** 1998. The role of neisserial Opa proteins in interactions with host cells. Trends Microbiol **6:**489-495.

52. **Desnues, B., C. Cuny, G. Gregori, S. Dukan, H. Aguilaniu, and T. Nystrom.** 2003. Differential oxidative damage and expression of stress defence regulons in culturable and non-culturable *Escherichia coli* cells. EMBO Rep **4:**400-404.

53. **Dhar, N., and J. D. McKinney.** 2007. Microbial phenotypic heterogeneity and antibiotic tolerance. Curr Opin Microbiol **10:**30-38.

54. **Dobbel, C.** 1932. Antony van Leeuwenhoek and his 'little animals': being some account of the father of protozoology and bacteriology and his multifarious discoveries in these disciplines. Constable, London, UK.

55. **Dobrindt, U., B. Hochhut, U. Hentschel, and J. Hacker.** 2004. Genomic islands in pathogenic and environmental microorganisms. Nature Rev Microbiol **2:**414-424.

56.   **Doolittle, W. F.** 1999. Phylogenetic classification and the universal tree. Science **284:**2124-2129.

57.   **Dubey, G. P., and S. Ben-Yehuda.** 2011. Intercellular nanotubes mediate bacterial communication. Cell **144:**590-600.

58.   **Dubnau, D., and R. Losick.** 2006. Bistability in bacteria. Mol Microbiol **61:**564-572.

59.   **Eberhardt, A., L. J. Wu, J. Errington, W. Vollmer, and J. W. Veening.** 2009. Cellular localization of choline-utilization proteins in Streptococcus pneumoniae using novel fluorescent reporter systems. Mol Microbiol **74:**395-408.

60.   **Eldar, A., and M. B. Elowitz.** 2010. Functional roles for noise in genetic circuits. Nature **467:**167-173.

61.   **Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain.** 2002. Stochastic gene expression in a single cell. Science **297:**1183-1186.

62.   **Emerson, J. E., C. B. Reynolds, R. P. Fagan, H. A. Shaw, D. Goulding, and N. F. Fairweather.** 2009. A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein. Mol Microbiol **74:**541-556.

63.   **Erjavec, N., M. Cvijovic, E. Klipp, and T. Nystrom.** 2008. Selective benefits of damage partitioning in unicellular systems and its effects on aging. Proc Natl Acad Sci U S A **105:**18764-18769.

64.   **Favet, J., A. Lapanje, A. Giongo, S. Kennedy, Y. Y. Aung, A. Cattaneo, A. G. Davis-Richardson, C. T. Brown, R. Kort, H. J. Brumsack, B. Schnetger, A. Chappell, J. Kroijenga, A. Beck, K. Schwibbert, A. H. Mohamed, T. Kirchner, P. D. de Quadros, E. W. Triplett, W. J. Broughton, and A. A. Gorbushina.** 2013. Microbial hitchhikers on intercontinental dust: catching a lift in Chad. ISME J **7:**850-867.

65.   **Fozo, E. M., K. S. Makarova, S. A. Shabalina, N. Yutin, E. V. Koonin, and G. Storz.** 2010. Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. Nucleic Acids Res **38:**3743-3759.

66.   **Fraser, D., and M. Kaern.** 2009. A chance at survival: gene expression noise and phenotypic diversification strategies. Mol Microbiol **71:**1333-1340.

67.   **Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint.** 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol **3:**722-732.

68.   **Fujita, Y.** 2009. Carbon catabolite control of the metabolic network in Bacillus subtilis. Biosci Biotechnol Biochem **73:**245-259.

69. **Gaillard, M., N. Pernet, C. Vogne, O. Hagenbüchle, and J. R. van der Meer.** 2008. Host and invader impact of transfer of the *clc* genomic island into *Pseudomonas aeruginosa* PAO1. Proc Natl Acad Sci U S A **105:**7058-7063.

70. **Gaillard, M., N. Pradervand, M. Minoia, V. Sentchilo, D. R. Johnson, and J. R. van der Meer.** 2010. Transcriptome analysis of the mobile genome ICE*clc* in *Pseudomonas knackmussii* B13. BMC Microbiol **10:**153.

71. **Gaillard, M., T. Vallaeys, F. J. Vorholter, M. Minoia, C. Werlen, V. Sentchilo, A. Puhler, and J. R. van der Meer.** 2006. The *clc* element of *Pseudomonas* sp. strain B13, a genomic island with various catabolic properties. J Bacteriol **188:**1999-2013.

72. **Gefen, O., and N. Q. Balaban.** 2009. The importance of being persistent: heterogeneity of bacterial populations under antibiotic stress. FEMS Microbiol Rev **33:**704-717.

73. **Gefen, O., C. Gabay, M. Mumcuoglu, G. Engel, and N. Q. Balaban.** 2008. Single-cell protein induction dynamics reveals a period of vulnerability to antibiotics in persister bacteria. Proc Natl Acad Sci U S A **105:**6145-6149.

74. **Gerdes, K., S. K. Christensen, and A. Lobner-Olesen.** 2005. Prokaryotic toxin-antitoxin stress response loci. Nat Rev Microbiol **3:**371-382.

75. **Gil, R., A. Latorre, and A. Moya.** 2004. Bacterial endosymbionts of insects: insights from comparative genomics. Environ Microbiol **6:**1109-1122.

76. **Gogarten, J. P., and J. P. Townsend.** 2005. Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol **3:**679-682.

77. **Golding, I., J. Paulsson, S. M. Zawilski, and E. C. Cox.** 2005. Real-time kinetics of gene activity in individual bacteria. Cell **123:**1025-1036.

78. **González-Pastor, J. E., E. C. Hobbs, and R. Losick.** 2003. Cannibalism by sporulating bacteria. Science **301:**510-513.

79. **Hacker, J., and E. Carniel.** 2001. Ecological fitness, genomic islands and bacterial pathogenicity: A Darwinian view of the evolution of microbes. EMBO Rep. **2:**376-381.

80. **Hacker, J., and J. B. Kaper.** 2000. Pathogenicity islands and the evolution of microbes. Annu Rev Microbiol **54:**641-679.

81. **Hallet, B.** 2001. Playing Dr Jekyll and Mr Hyde: combined mechanisms of phase variation in bacteria. Curr Opin Microbiol **4:**570-581.

82. **Hecker, M., and U. Volker.** 1998. Non-specific, general and multiple stress resistance of growth-restricted Bacillus subtilis cells by the expression of the sigmaB regulon. Mol Microbiol **29:**1129-1136.

83. **Henderson, I. R., P. Owen, and J. P. Nataro.** 1999. Molecular switches--the ON and OFF of bacterial phase variation. Mol Microbiol **33:**919-932.

84. **Hernday, A., M. Krabbe, B. Braaten, and D. Low.** 2002. Self-perpetuating epigenetic pili switches in bacteria. Proc Natl Acad Sci U S A **99 Suppl 4:**16470-16476.

85. **Howell-Adams, B., and H. S. Seifert.** 2000. Molecular models accounting for the gene conversion reactions mediating gonococcal pilin antigenic variation. Mol Microbiol **37:**1146-1158.

86. http://www.ncbi.nlm.nih.gov/genome/browse/ 31Aug 2013, posting date. [Online.]

87. **International Human Genome Sequencing Consortium.** 2001. Initial sequencing and analysis of the human genome. Nature **409:**860-921.

88. **Jiang, H., H. Dong, G. Zhang, B. Yu, L. R. Chapman, and M. W. Fields.** 2006. Microbial diversity in water and sediment of Lake Chaka, an athalassohaline lake in northwestern China. Appl Environ Microbiol **72:**3832-3845.

89. **Johnson, L. R., and M. Mangel.** 2006. Life histories and the evolution of aging in bacteria and other single-celled organisms. Mech Ageing Dev **127:**786-793.

90. **Johnston, R. J., Jr., and C. Desplan.** 2010. Stochastic mechanisms of cell fate specification that yield random or robust outcomes. Annu Rev Cell Dev Biol **26:**689-719.

91. **Juhas, M., D. W. Crook, I. D. Dimopoulou, G. Lunter, R. M. Harding, D. J. Ferguson, and D. W. Hood.** 2007. Novel type IV secretion system involved in propagation of genomic islands. J Bacteriol **189:**761-771.

92. **Juhas, M., J. R. van der Meer, M. Gaillard, R. M. Harding, D. W. Hood, and D. W. Crook.** 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev **33:**376-393.

93. **Kaern, M., T. C. Elston, W. J. Blake, and J. J. Collins.** 2005. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet **6:**451-464.

94. **Kahn, N. A., R. Siddiqui, and H. Elsheikha.** 2010. Enemy within: strategies to kill 'superbugs' in hospitals. Int J Antimicrob Agents **36:**291.

95. **Kaufmann, B. B., and A. van Oudenaarden.** 2007. Stochastic gene expression: from single molecules to the proteome. Curr Opin Genet Dev **17:**107-112.

96. **Kearns, D. B., and R. Losick.** 2005. Cell population heterogeneity during growth of *Bacillus subtilis*. Genes Dev **19:**3083-3094.

97. **Keese, P.** 2008. Risks from GMOs due to horizontal gene transfer. Environ Biosafety Res **7:**123-149.

98. **Kentner, D., and V. Sourjik.** 2010. Use of fluorescence microscopy to study intracellular signaling in bacteria. Annu Rev Microbiol **64:**373-390.

99. **Kim, H. J., J. Q. Boedicker, and J. W. Choi.** 2008. Defined spatial structure stabilizes a synthetic multispecies bacterial community. Proc Natl Acad Sci U S A **105:**18188-18193.

100. **Klemm, P.** 1986. Two regulatory fim genes, fimB and fimE, control the phase variation of type 1 fimbriae in *Escherichia coli*. EMBO J **5:**1389-1393.

101. **Kloesges, T., O. Popa, W. Martin, and T. Dagan.** 2011. Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. Mol Biol Evol **28:**1057-1074.

102. **Kock, R., J. Harlizius, N. Bressan, R. Laerberg, L. H. Wieler, W. Witte, R. H. Deurenberg, A. Voss, K. Becker, and A. W. Friedrich.** 2009. Prevalence and molecular characteristics of methicillin-resistant Staphylococcus aureus (MRSA) among pigs on German farms and import of livestock-related MRSA into hospitals. Eur J Clin Microbiol Infect Dis **28:**1375-1382.

103. **Koomey, M.** 2001. Implications of molecular contacts and signaling initiated by *Neisseria gonorrhoeae*. Curr Opin Microbiol **4:**53-57.

104. **Koonin, E., K. Makarova, and L. Aravind.** 2001. Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol **55:**709–742.

105. **Koonin, E. V., and Y. I. Wolf.** 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res **36:**6688-6719.

106. **Kussell, E., and S. Leibler.** 2005. Phenotypic diversity, population growth, and information in fluctuating environments. Science **309:**2075-2078.

107. **Larbig, K., A. Christmann, A. Johann, J. Klockgether, T. Hartsch, and R. Merkl.** 2002. Gene islands integrated into tRNA(Gly) genes confer genome diversity on a *Pseudomonas aeruginosa* clone. J Bacteriol **184:**6665–6680.

108. **Lawler, M. L., and Y. V. Brun.** 2007. Advantages and mechanisms of polarity and cell shape determination in *Caulobacter crescentus*. Curr Opin Microbiol **10:**630-637.

109. **Lechner, M., K. Schmitt, S. Bauer, D. Hot, C. Hubans, E. Levillain, C. Locht, Y. Lemoine, and R. Gross.** 2009. Genomic island excisions in Bordetella petrii. BMC Microbiol **9:**141.

110. **Lee, C. A., J. M. Auchtung, R. E. Monson, and A. D. Grossman.** 2007. Identification and characterization of int (integrase), xis (excisionase) and chromosomal attachment sites of the integrative and conjugative element ICEBs1 of Bacillus subtilis. Mol Microbiol **66:**1356-1369.

111. **Lee, C. A., A. Babic, and A. D. Grossman.** 2010. Autonomous plasmid-like replication of a conjugative transposon. Mol Microbiol **75:**268-279.

112. **Leeuwenhoek, A. v.** 1939–1999. Collected letters, vol. 1-15. Swets and Zeitlinger, Amsterdam, The Netherlands.

113. **Lesic, B., S. Bach, J. M. Ghigo, U. Dobrindt, J. Hacker, and E. Carniel.** 2004. Excision of the high-pathogenicity island of Yersinia pseudotuberculosis requires the combined actions of its cognate integrase and Hef, a new recombination directionality factor. Mol Microbiol **52:**1337-1348.

114. **Lewis, J. A., and G. F. Hatfull.** 2001. Control of directionality in integrasemediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. Nucleic Acids Res **29:**2205-2216.

115. **Lewis, M.** 2005. The lac repressor. C R Biol **328:**521-548.

116. **Li, G.-W., and X. S. Xie.** 2011. Central dogma at the single-molecule level in living cells. Nature **475:**308-315.

117. **Libby, E., and P. B. Rainey.** 2011. Exclusion rules, bottlenecks and the evolution of stochastic phenotype switching. Proc Biol Sci **278:**3574-3583.

118. **Lindner, A. B., R. Madden, A. Demarez, E. J. Stewart, and F. Taddei.** 2008. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. Proc Natl Acad Sci U S A **105:**3076-3081.

119. **Lindow, S. E., and M. T. Brandl.** 2003. Microbiology of the phyllosphere. Appl Environ Microbiol **69:**1875-1883.

120. **Liu, P., and R. A. Mathies.** 2009. Integrated microfluidic systems for high-performance genetic analysis. Trends Biotechnol **27:**572-581.

121. **Livet, J., T. A. Weissman, H. Kang, R. W. Draft, J. Lu, R. A. Bennis, J. R. Sanes, and J. W. Lichtman.** 2007. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. Nature **450:**56-62.

122.   **Locke, J. C., and M. B. Elowitz.** 2009. Using movies to analyse gene circuit dynamics in single cells. Nat Rev Microbiol **7:**383-392.

123.   **Lopez, D., H. Vlamakis, R. Losick, and R. Kolter.** 2009. Cannibalism enhances biofilm development in *Bacillus subtilis*. Mol Microbiol **74:**609-618.

124.   **Lopez-Maury, L., S. Marguerat, and J. Bahler.** 2008. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat Rev Genet **9:**583-593.

125.   **Losick, R., and C. Desplan.** 2008. Stochasticity and cell fate. Science **320:**65-68.

126.   **Lundberg, D. S., S. L. Lebeis, S. H. Paredes, S. Yourstone, J. Gehring, S. Malfatti, J. Tremblay, A. Engelbrektson, V. Kunin, T. G. del Rio, R. C. Edgar, T. Eickhorst, R. E. Ley, P. Hugenholtz, S. G. Tringe, and J. L. Dangl.** 2012. Defining the core Arabidopsis thaliana root microbiome. Nature **488:**86-90.

127.   **Maamar, H., and D. Dubnau.** 2005. Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop. Mol Microbiol **56:**615-624.

128.   **Madigan, M. T., J. M. Martinko, D. Stahl, and D. P. Clark.** 2012. Brock Biology of Microorganisms, 13 ed, vol. 232. Pearson, London.

129.   **Magnusson, L. U., A. Farewell, and T. Nystrom.** 2005. ppGpp: a global regulator in Escherichia coli. Trends Microbiol **13:**236-242.

130.   **Maheshri, N., and E. K. O'Shea.** 2007. Living with noisy genes: how cells function reliably with inherent variability in gene expression. Annu Rev Biophys Biomol Struct **36:**413-434.

131.   **Martin, M., and D. de Mendoza.** 2013. Regulation of *Bacillus subtilis* DesK thermosensor by lipids. Biochem J **451:**269-275.

132.   **Mauriello, E. M.** 2013. Cell biology of bacterial sensory modules. Front Biosci **1:**928-943.

133.   **McAdams, H. H., and A. Arkin.** 1997. Stochastic mechanisms in gene expression. Proc Natl Acad Sci U S A **94:**814-819.

134.   **McClain, M. S., I. C. Blomfield, and B. I. Eisenstein.** 1991. Roles of fimB and fimE in site-specific DNA inversion associated with phase variation of type 1 fimbriae in *Escherichia coli*. J Bacteriol **173:**5308-5314.

135.   **McGowan, J. E., Jr., and F. C. Tenover.** 2004. Confronting bacterial resistance in healthcare settings: a crucial role for microbiologists. Nat Rev Microbiol **2:**251-258.

136. **McGrane, R. S., and G. A. Beattie.** 2013. Light Regulation of Swarming Motility in *Pseudomonas syringae* Integrates Signaling Pathways Mediated by a Bacteriophytochrome and a LOV Protein. MBio **4:**2150-7511.

137. **Mehr, I. J., and H. S. Seifert.** 1998. Differential roles of homologous recombination pathways in Neisseria gonorrhoeae pilin antigenic variation, DNA transformation and DNA repair. Mol Microbiol **30:**697-710.

138. **Metruccio, M. M. E., E. Pigozzi, D. Roncarati, F. Berlanda Scorza, N. Norais, S. A. Hill, V. Scarlato, and I. Delany.** 2009. A novel phase variation mechanism in the meningococcus driven by a ligand-responsive repressor and differential spacing of distal promoter elements. PLoS Pathog **5:**e1000710.

139. **Mettetal, J. T., D. Muzzey, J. M. Pedraza, E. M. Ozbudak, and A. van Oudenaarden.** 2006. Predicting stochastic gene expression dynamics in single cells. Proc Natl Acad Sci U S A **103:**7304-7309.

140. **Minoia, M., M. Gaillard, F. Reinhard, M. Stojanov, V. Sentchilo, and J. R. van der Meer.** 2008. Stochasticity and bistability in horizontal transfer control of a genomic island in *Pseudomonas*. Proc. Natl. Acad. Sci. U S A **105:**20792-20797.

141. **Miyazaki, R., M. Minoia, N. Pradervand, V. Sentchilo, S. Sulser, F. Reinhard, and J. R. van der Meer.** 2011. The *clc* element and related genomic islands in *Proteobacteria*. *In* A. P. Roberts and P. Mullany (ed.), Bacterial integrative mobile genetic elements. Landes Bioscience.

142. **Miyazaki, R., M. Minoia, N. Pradervand, S. Sulser, F. Reinhard, and J. R. van der Meer.** 2012. Cellular variability of RpoS expression underlies subpopulation activation of an integrative and conjugative element. PLoS Genet **8:**e1002818.

143. **Mohd-Zain, Z., S. L. Turner, A. M. Cerdeño-Tárraga, A. K. Lilley, T. J. Inzana, A. J. Duncan, R. M. Harding, D. W. Hood, T. E. Peto, and D. W. Crook.** 2004. Transferable antibiotic resistance elements in *Haemophilus influenzae* share a common evolutionary origin with a diverse family of syntenic genomic islands. J. Bacteriol. **186:**8114-8122.

144. **Moxon, R., C. Bayliss, and D. Hood.** 2006. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. Annu Rev Genet **40:**307-333.

145. **Moyed, H. S., and K. P. Bertrand.** 1983. hipA, a newly recognized gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis. J Bacteriol **155:**768-775.

146. **Muller, S., and G. Nebe-von-Caron.** 2010. Functional single-cell analyses: flow cytometry and cell sorting of microbial populations and communities. FEMS Microbiol Rev **34:**554-87.

147. **Nassif, X., C. Pujol, P. Morand, and E. Eugene.** 1999. Interactions of pathogenic Neisseria with host cells. Is it possible to assemble the puzzle? Mol Microbiol **32:**1124-1132.

148. **Nordmann, P., T. Naas, N. Fortineau, and L. Poirel.** 2007. Superbugs in the coming new decade; multidrug resistance and prospects for treatment of Staphylococcus aureus, Enterococcus spp. and Pseudomonas aeruginosa in 2010. Curr Opin Microbiol **10:**436-440.

149. **Novick, A., and M. Weiner.** 1957. Enzyme Induction as an All-or-None Phenomenon. Proc Natl Acad Sci U S A **43:**553-566.

150. **Nystrom, T.** 2007. A bacterial kind of aging. PLoS Genet **3:**e224.

151. **Nystrom, T.** 2011. Spatial protein quality control and the evolution of lineage-specific ageing. Philos Trans R Soc Lond B Biol Sci **366:**71-75.

152. **Ochab-Marcinek, A., and M. Tabaka.** 2010. Bimodal gene expression in noncooperative regulatory systems. Proc Natl Acad Sci U S A **107:**22096-22101.

153. **Ochman, H., J. G. Lawrence, and E. A. Groisman.** 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405:**299-304.

154. **Oppenheim, A. B., O. Kobiler, J. Stavans, D. L. Court, and S. Adhya.** 2005. Switches in bacteriophage lambda development. Annu Rev Genet **39:**409-429.

155. **Ozbudak, E. M., M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden.** 2002. Regulation of noise in the expression of a single gene. Nat Genet **31:**69-73.

156. **Ozbudak, E. M., M. Thattai, H. N. Lim, B. I. Shraiman, and A. Van Oudenaarden.** 2004. Multistability in the lactose utilization network of *Escherichia coli*. Nature **427:**737-740.

157. **Pace, N. R.** 1997. A molecular view of microbial diversity and the biosphere. Science **276:**734-740.

158. **Perner, M., J. M. Petersen, F. Zielinski, H. H. Gennerich, and R. Seifert.** 2010. Geochemical constraints on the diversity and activity of H2 -oxidizing microorganisms in diffuse hydrothermal fluids from a basalt- and an ultramafic-hosted vent. FEMS Microbiol Ecol **74:**55-71.

159. **Ping, L., B. Weiner, and N. Kleckner.** 2008. Tsr-GFP accumulates linearly with time at cell poles, and can be used to differentiate 'old' versus 'new' poles, in *Escherichia coli*. Mol Microbiol **69:**1427-1438.

160. **Ptashne, M.** 2006. Lambda's switch: lessons from a module swap. Curr Biol **16:**459-462.

161. **Ptashne, M.** 2011. Principles of a switch. Nature chemical biology **7:**484-487.

162. **Qiu, X., A. U. Gurkar, and S. Lory.** 2006. Interstrain transfer of the large pathogenicity island (PAPI-1) of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A **103:**19830-19835.

163. **Rainey, P. B., H. J. E. Beaumont, G. C. Ferguson, J. Gallie, C. Kost, E. Libby, and X.-X. Zhang.** 2011. The evolutionary emergence of stochastic phenotype switching in bacteria. Microb Cell Fact **10 Suppl 1:**S14.

164. **Ramsay, J. P., J. T. Sullivan, N. Jambari, C. A. Ortori, S. Heeb, P. Williams, D. A. Barrett, I. L. Lamont, and C. W. Ronson.** 2009. A LuxRI-family regulatory system controls excision and transfer of the Mesorhizobium loti strain R7A symbiosis island by activating expression of two conserved hypothetical genes. Mol Microbiol **73:**1141-1155.

165. **Ramsay, J. P., J. T. Sullivan, G. S. Stuart, I. L. Lamont, and C. W. Ronson.** 2006. Excision and transfer of the Mesorhizobium loti R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. Mol Microbiol **62:**723-734.

166. **Rang, C. U., A. Y. Peng, and L. Chao.** 2011. Temporal dynamics of bacterial aging and rejuvenation. Curr Biol **21:**1813-1816.

167. **Raser, J. M., and E. K. O'Shea.** 2005. Noise in gene expression: origins, consequences, and control. Science **309:**2010-2013.

168. **Ravatn, R., S. Studer, D. Springael, A. J. B. Zehnder, and J. R. van der Meer.** 1998. Chromosomal integration, tandem amplification, and deamplification in *Pseudomonas putida* F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from *Pseudomonas* sp. strain B13. J. Bacteriol. **180:**4360-4369.

169. **Rose, M.** 1991. Evolutionary biology of ageing. Oxford Univ. Press, Oxford.

170. **Rosenfeld, N., J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz.** 2005. Gene regulation at the single-cell level. Science **307:**1962-1965.

171. **Rotem, E., A. Loinger, I. Ronin, I. Levin-Reisman, C. Gabay, N. Shoresh, O. Biham, and N. Q. Balaban.** 2010. Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. Proc Natl Acad Sci U S A **107:**12541-12546.

172. **Russell, N. J.** 1990. Cold adaptation of microorganisms. Philos Trans R Soc Lond B Biol Sci **326:**595-611.

173. **Satory, D., A. J. Gordon, J. A. Halliday, and C. Herman.** 2011. Epigenetic switches: can infidelity govern fate in microbes? Curr Opin Microbiol **14:**212-217.

174. **Schmidt, H., and M. Hensel.** 2004. Pathogenicity islands in bacterial pathogenesis. Clin Microbiol Rev **17:**14-56.

175. **Schröder, G., and E. Lanka.** 2005. The mating pair formation system of conjugative plasmids—A versatile secretion machinery for transfer of proteins and DNA. Plasmid **54:**1-25.

176. **Schubert, S., S. Dufke, J. Sorsa, and J. Heesemann.** 2004. A novel integrative and conjugative element (ICE) of Escherichia coli: the putative progenitor of the Yersinia high-pathogenicity island. Mol Microbiol **51:**837-848.

177. **Sentchilo, V., K. Czechowska, N. Pradervand, M. Minoia, R. Miyazaki, and J. R. van der Meer.** 2009. Intracellular excision and reintegration dynamics of the ICE*clc* genomic island of *Pseudomonas knackmussii* sp. strain B13. Mol Microbiol **72:**1293-306.

178. **Sentchilo, V., R. Ravatn, C. Werlen, A. J. Zehnder, and J. R. van der Meer.** 2003. Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. Strain B13. Journal of Bacteriology **185:**4530-4538.

179. **Shaner, N. C., P. A. Steinbach, and R. Y. Tsien.** 2005. A guide to choosing fluorescent proteins. Nat Methods **2**.

180. **Shapiro, H. M.** 2000. Microbial analysis at the single-cell level: tasks and techniques. J Microbiol Methods **42:**3-16.

181. **Sibbald, B.** 2012. Farm-grown superbugs: while the world acts, Canada dawdles. CMAJ **184:**1553.

182. **Simon, C., A. Wiezer, A. W. Strittmatter, and R. Daniel.** 2009. Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. Appl Environ Microbiol **75:**7519-7526.

183. **Smets, B. F., and T. Barkay.** 2005. Horizontal gene transfer: perspectives at a crossroads of scientific disciplines. Nat Rev Microbiol **3:**675-678.

184. **Smits, W. K., C. C. Eschevins, K. A. Susanna, S. Bron, O. P. Kuipers, and L. W. Hamoen.** 2005. Stripping *Bacillus*: ComK auto-stimulation is responsible for the bistable response in competence development. Mol Microbiol **56:**604-614.

185. **Smits, W. K., O. P. Kuipers, and J. W. Veening.** 2006. Phenotypic variation in bacteria: the role of feedback regulation. Nat Rev Microbiol **4:**259-271.

186. **Springael, D., and E. M. Top.** 2004. Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. Trends Microbiol **12:**53-58.

187. **Stewart, E. J., R. Madden, G. Paul, and F. Taddei.** 2005. Aging and death in an organism that reproduces by morphologically symmetric division. PLoS Biol **3:**e45.

188. **Stewart, G. R., B. D. Robertson, and D. B. Young.** 2003. Tuberculosis: a problem with persistence. Nat Rev Microbiol **1:**97-105.

189. **Stomeo, F., A. Valverde, S. B. Pointing, C. P. McKay, K. A. Warren-Rhodes, M. I. Tuffin, M. Seely, and D. A. Cowan.** 2013. Hypolithic and soil microbial community assembly along an aridity gradient in the Namib Desert. Extremophiles **17:**329-337.

190. **Strovas, T. J., L. M. Sauter, X. Guo, and M. E. Lidstrom.** 2007. Cell-to-cell heterogeneity in growth rate and gene expression in *Methylobacterium extorquens* AM1. J Bacteriol **189:**7127-7133.

191. **Suel, G. M., J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz.** 2006. An excitable gene regulatory circuit induces transient cellular differentiation. Nature **440:**545-550.

192. **Suel, G. M., R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M. B. Elowitz.** 2007. Tunability and noise dependence in differentiation dynamics. Science **315:**1716-1719.

193. **Sullivan, J. T., J. R. Trzebiatowski, R. W. Cruickshank, J. Gouzy, S. D. Brown, R. M. Elliot, D. J. Fleetwood, N. G. McCallum, U. Rossbach, G. S. Stuart, J. E. Weaver, R. J. Webby, F. J. de Bruijn, and C. W. Ronson.** 2002. Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. J Bacteriol **184:**3086-3095.

194. **Swain, P. S., M. B. Elowitz, and E. D. Siggia.** 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci U S A **99:**12795-12800.

195. **Taghavi, S., T. Barac, B. Greenberg, B. Borremans, J. Vangronsveld, and D. van der Lelie.** 2005. Horizontal gene transfer to endogenous endophytic bacteria from poplar improves phytoremediation of toluene. Appl Environ Microbiol **71:**8500-8505.

196. **Thomas, C. M., and K. M. Nielsen.** 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol **3:**711-721.

197. **To, T.-L., and N. Maheshri.** 2010. Noise can induce bimodality in positive transcriptional feedback loops without bistability. Science **327:**1142-1145.

198. **Torsvik, V., J. Goksoyr, and F. L. Daae.** 1990. High diversity in DNA of soil bacteria. Appl Environ Microbiol **56:**782-787.

199. **Totsika, M., S. A. Beatson, N. Holden, and D. L. Gally.** 2008. Regulatory interplay between pap operons in uropathogenic *Escherichia coli*. Mol Microbiol **67:**996-1011.

200. **Toussaint, A., and C. Merlin.** 2002. Mobile elements as a combination of functional modules. Plasmid **47:**26-35.

201. **Toussaint, A., C. Merlin, S. Monchy, M. A. Benotmane, R. Leplae, M. Mergeay, and D. Springael.** 2003. The biphenyl- and 4-chlorobiphenyl-catabolic transposon Tn*4371*, a member of a new family of genomic islands related to IncP and Ti plasmids. Appl. Environ. Microbiol. **69:**4837-4845.

202. **Turnbaugh, P. J., R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon.** 2007. The human microbiome project. Nature **449:**804-810.

203. **Ursell, L. K., J. L. Metcalf, L. W. Parfrey, and R. Knight.** 2012. Defining the human microbiome. Nutr Rev **70 Suppl 1:**S38-S44.

204. **Valverde, A., M. Tuffin, and D. A. Cowan.** 2012. Biogeography of bacterial communities in hot springs: a focus on the actinobacteria. Extremophiles **16:**669-679.

205. **van der Meer, J. R., C. Werlen, S. F. Nishino, and J. C. Spain.** 1998. Evolution of a pathway for chlorobenzene metabolism leads to natural attenuation in contaminated groundwater. Appl Environ Microbiol **64:**4185-4193.

206. **van der Woude, M. W.** 2011. Phase variation: how to create and coordinate population diversity. Curr Opin Microbiol **14:**205-211.

207. **van der Woude, M. W., and A. J. Baumler.** 2004. Phase and antigenic variation in bacteria. Clin Microbiol Rev **17:**581-611.

208.   van Ham, S. M., L. van Alphen, F. R. Mooi, and J. P. van Putten. 1993. Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. Cell **73:**1187-1196.

209.   Veening, J. W., L. W. Hamoen, and O. P. Kuipers. 2005. Phosphatases modulate the bistable sporulation gene expression pattern in *Bacillus subtilis*. Mol Microbiol **56:**1481-1494.

210.   Veening, J. W., W. K. Smits, L. W. Hamoen, and O. P. Kuipers. 2006. Single cell analysis of gene expression patterns of competence development and initiation of sporulation in *Bacillus subtilis* grown on chemically defined media. J Appl Microbiol **101:**531-541.

211.   Veening, J. W., W. K. Smits, and O. P. Kuipers. 2008. Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol **62:**193-210.

212.   Veening, J. W., E. J. Stewart, T. W. Berngruber, F. Taddei, O. P. Kuipers, and L. W. Hamoen. 2008. Bet-hedging and epigenetic inheritance in bacterial cell development. Proc Natl Acad Sci U S A **105:**4393-4398.

213.   Wang, P., L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun. 2010. Robust growth of *Escherichia coli*. Curr Biol **20:**1099-1103.

214.   Wang, Q., J. Niemi, C. M. Tan, L. You, and M. West. 2010. Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy. Cytometry A **77:**101-110.

215.   Watve, M., S. Parab, P. Jogdand, and S. Keni. 2006. Aging may be a conditional strategic choice and not an inevitable outcome for bacteria. Proc Natl Acad Sci U S A **103:**14831-14835.

216.   Welch, R., V. Burland, G. I. Plunkett, P. Redford, P. Roesch, D. Rasko, E. Buckles, S. Liou, A. Boutin, J. Hackett, D. Stroud, G. Mayhew, D. Rose, S. Zhou, D. Schwartz, N. Perna, H. Mobley, M. Donnenberg, and F. Blattner. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc Natl Acad Sci U S A **99:**17020-17024.

217.   Wemheuer, B., R. Taube, P. Akyol, F. Wemheuer, and R. Daniel. 2013. Microbial diversity and biochemical potential encoded by thermal spring metagenomes derived from the Kamchatka Peninsula. Archaea **2013:**136714.

218.   Whitelaw, C., W. Barbazuk, G. Pertea, A. Chan, F. Cheung, Y. Lee, L. Zheng, S. van Heeringen, S. Karamycheva, J. Bennetzen, P. SanMiguel,

N. Lakey, J. Bedell, Y. Yuan, M. Budiman, A. Resnick, S. Van Aken, T. Utterback, S. Riedmuller, M. Williams, T. Feldblyum, K. Schubert, R. Beachy, C. Fraser, and J. Quackenbush. 2003. Enrichment of gene-coding sequences in maize by genome filtration. Science **302:**2118–2120.

219. **Wilde, C., D. Mazel, B. Hochhut, B. Middendorf, F. Le Roux, E. Carniel, U. Dobrindt, and J. Hacker.** 2008. Delineation of the recombination sites necessary for integration of pathogenicity islands II and III into the *Escherichia coli* 536 chromosome. Molecular Microbiology **68:**139-151.

220. **Winkler, J., A. Seybert, L. Konig, S. Pruggnaller, U. Haselmann, V. Sourjik, M. Weiss, A. S. Frangakis, A. Mogk, and B. Bukau.** 2010. Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. Embo J **29:**910-923.

221. **Woldringh, C. L.** 2005. Is *Escherichia coli* getting old? Bioessays **27:**770-774.

222. **Wozniak, R. A., and M. K. Waldor.** 2009. A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. PLoS Genet **5:**e1000439.

223. **Wozniak, R. A., and M. K. Waldor.** 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol **8:**5525-5563.

224. **Yu, J., J. Xiao, X. Ren, K. Lao, and X. S. Xie.** 2006. Probing Gene Expression in Live Cells, One Protein Molecule at a Time. Science **311:**1600-1603.

225. **Zare, R. N., and S. Kim.** 2010. Microfluidic platforms for single-cell analysis. Annu Rev Biomed Eng **15:**187-201.

226. **Zhang, G., H. Dong, Z. Xu, D. Zhao, and C. Zhang.** 2005. Microbial diversity in ultra-high-pressure rocks and fluids from the Chinese Continental Scientific Drilling Project in China. Appl Environ Microbiol **71:**3213-3227.

CHAPTER 2

Improved statistical analysis of low abundance phenomena in bimodal bacterial populations

ABSTRACT

Accurate detection of subpopulation size determinations in bimodal populations remains problematic yet it represents a powerful way by which cellular heterogeneity under different environmental conditions can be compared. So far, most studies have relied on qualitative descriptions of population distribution patterns, on population-independent descriptors, or on arbitrary placement of thresholds distinguishing biological ON from Off states. We found that all these methods fall short of accurately describing small population sizes in bimodal populations. Here we propose a simple, statistics-based method for the analysis of small subpopulation sizes for use in the free software environment R and test this method on real as well as simulated data.

Four so-called population splitting methods were designed with different algorithms that can estimate subpopulation sizes from bimodal populations. All four methods proved more sensitive than previously used methods when analyzing subpopulation sizes of transfer competent cells arising in populations of the bacterium $Pseudomonas\ knackmussii$ B13. The methods' resolving powers were further explored by bootstrapping and simulations. Two of the methods were not severely limited by the proportions of subpopulations they could estimate correctly, but the two others only allowed reliable subpopulation quantification when this amounted to less than 25% of the total population. In contrast, only one method was still sufficiently accurate with subpopulations smaller than 1 % of the total population.

This study proposes a number of rational approximations to quantifying small subpopulations and offers an easy-to-use protocol for their implementation in the open source statistical software environment R.

AUTHOR SUMMARY

Advances in microbiology have traditionally been based on population level studies, assuming that clonal populations consist of individual cells with on average the same behaviour. However, recent studies have shown important heterogeneity among cells in isogenic microbial populations, leading to a variety of clearly different bi- or multimodal phenotypes and even bistable phenotypic traits. Precise characterization of subpopulation differences is difficult, and to date there exists no universal protocol for the determination of small subpopulation sizes. The principal contribution of this study is a simple and practical statistical approximation to subpopulation quantification in bimodal populations. For this purpose we created a set of functions in the open source software environment *R* accompanied by step-by-step instructions for easy implementation. We develop and test our methods to study transfer competence development in the bacterium *Pseudomonas*, during which a very small proportion of cells in resting phase (3-5%) become prone for conjugative transfer of so-called Integrative and Conjugative Elements. In addition, we use computational modelling to test the efficacy of our methods to separate artificially constructed populations with different (known) proportions. Our approach helps to minimize inconsistencies in subpopulation classification, and increases sensitivity and accuracy of subpopulation size quantification.

INTRODUCTION

Advances in microbiology have traditionally been based on studies at the population level. Questions of how cells respond to their environment, interact with each other, or undergo complex processes such as cellular differentiation or gene expression have been mostly answered by inference from population-level data. Recent technological advances have facilitated the study of individual cells and led to new appreciation for the existence and importance of phenotypic heterogeneity (2, 14). There is no more doubt that gene expression is heterogeneous among cells in isogenic microbial populations and leads to physiological heterogeneity (18, 47, 51). In many cases distributions of physiological parameters among individual cells in populations show a small part, usually less than a few percent of the total, to be more than two-fold different from the population average (28, 50, 62). It is thought that the appearance or existence of small subpopulations with different phenotypes in a clonal population may be beneficial for its survival under adverse conditions (18, 59). As example, persistence to antibiotic toxicity in *Escherichia coli* is a very rare phenomenon (3), yet it is of great importance since it enables population survival and outgrowth when the antibiotic is removed. Growth to stationary phase of *B. subtilis* leads to the appearance of subpopulations with widely varying expression of glycolysis and gluconeogenesis enzymes that are thought to better enable stationary phase survival (18). In fact, an increasing number of phenotypic traits has been discovered that are not even homogenously distributed among all cells in a clonal bacterial population but rather lead to the formation of two (bimodal) distinct subpopulations. Current examples from microbiology include horizontal gene transfer activation in *Pseudomonas* (33, 34, 41), sporulation (19, 36), cannibalism (24), extracellular matrix formation (61), competence development(31, 48), and motility (13, 25) in *Bacillus subtilis*, the lysis-lysogeny switch of phage lambda (37), lactose utilization (35), the arabinose catabolic pathway (46), and chemotaxis in *E. coli* (26), quorum sensing-regulated bioluminescence in *Vibrio harveyi* (1), flagella expression in *Salmonella* typhi (6), or phase variation in a number of pathogens (55, 56). There is no reason not to assume that many more and diverse bimodal or even multimodal phenotypic differentiations in clonal bacterial populations would exist, and there is evidence that the extent of phenotypic variability is a selectable trait (47). Evidently, in order to better understand bimodal phenomena it is of critical importance to have accurate measurement and analysis tools for differentiating subpopulations within the total population. Most authors exploring bimodal phenomena have been relying on production of autofluorescent proteins to study critical promoters and regulatory events at the single cell level,

mainly because of the ease to detect expression of the reporter protein in individual cells (27, 47, 51). Such detection is typically performed by either epifluorescence microscopy and digital image analysis (9, 16, 18, 29, 51, 60) or by flow cytometry, if expression of the fluorescent reporter protein is sufficiently high (10, 30, 47, 49, 57, 58, 61). Measures of expression heterogeneity such as occurrence of bimodalities or subpopulation sizes, represent useful parameters to quantify phenotypic heterogeneity and its differences in mutants or as a result of growth conditions. However, the more one approaches very small subpopulation sizes (e.g., a few percent of the total) the more difficult it is to accurately detect and determine such events, and so far most methods do not take such low proportions into appropriate consideration. For example, subpopulation dynamics is often solely assessed in form of descriptive graphs that present the total distribution of fluorescence intensities for individuals. These included histograms (10, 17, 18, 30, 35, 46), cumulative distribution curves (CDFs) (4, 5, 44, 48, 53), normal quantile-quantile (Q-Q) plots (27, 42, 52) or percentile-percentile (P-P) plots (32). Although representations of total populations are useful for stating evident differences in distribution patterns between treatments, they tend to overlook more subtle differences which often need a quantitative approach. Quantification of subpopulation dynamics is generally done by addressing individual fluorescence values that fall within pre-defined boundaries of the total population. However, often these boundaries are determined independently of the nature of the distribution of the total population data. An example of this is when gating of clusters in flow cytometry is manually defined to identify subpopulation shifts (7, 49, 57, 61) or when threshold rules are based on background- or control fluorescence in fluorescence microscopy to determine "all-or-none" induction responses (1, 9, 13, 33, 58). A problematic with subpopulation quantification using pre-defined and distribution-independent thresholding is that such classification does not attempt to statistically approximate estimates for true, that is biologically relevant, subpopulations (since boundaries have nothing or little to do with the distribution of the data), but rather represents a pragmatic approach to achieve differentiation between treatments. Therefore, generally, such approach falls short of serving as a universal method for subpopulation quantification, especially when subpopulations overlap. One solution to this problem would entail a distribution-based approximation of the distinct subpopulations that is entirely independent of the experimental test system used (as long as the test system is sensitive enough), and the result of which could be expressed as a dimensionless quantity.

The aim of this study is to propose a methodology for quantifying small subpopulations (few percent) in bimodal populations. Our approach is based on a statistically valid approximation to finding the "true" subpopulation size in bimodal populations and expressing it

as a percentage of the total population size. The model system we use to develop our method is the bistable behaviour of the integrative and conjugative element called ICE*clc* of the bacterium *Pseudomonas knackmussii* B13 (22, 39, 40, 44). It was previously discovered that the promoter of the integrase gene (P_{int}) on ICE*clc* expresses under stationary-phase conditions in some 3% of cells in culture, specifically when they have been grown with 3-chlorobenzoate (3CBA) as sole carbon and energy source (44, 45). Cells that induce P_{int} are locked in a bistable state (33) and undergo a process of competence formation which enables ICE*clc* transfer (41). ICE*clc* behaviour was inferred from single-cell fluorescence measurements on strains carrying an additional single-copy transcriptional fusion between P_{int} and the gene for enhanced green fluorescent protein (eGFP) or mCherry. In first instance and because of the absence of clear bimodality, distribution-independent descriptors were used to describe P_{int} expression (44, 45). For this purpose, eGFP fluorescence intensities of at least one thousand imaged cells were ranked, from which the 95th percentile and the mean fluorescence intensity among the top five percent were calculated (44, 45). Alternatively, subpopulation sizes were determined from the 'breakpoint' in cumulatively ranked fluorescence values of thousands of individual imaged cells (34). Here we evaluate different methods for subpopulation characterization and propose a simple routine in the open source statistical software *R* that integrates some of the ideas of earlier studies (27, 33, 34). As these methods require population splitting (PS) into a large and small subpopulation (by use of a cutoff value) we call them PS methods. Our PS methods are particularly suitable for analysis of subpopulations of only a few percent of the total, which may otherwise be difficult to discern. A first data verification step is incorporated in the subroutine that summarizes data from different images to ensure that no outlier exposure errors or biases exist. The following steps then help to find the statistically most likely appropriate subpopulation size. We challenge PS methods in two ways; firstly, by measuring subpopulation sizes of ICE*clc* transfer competent cells of *P. knackmussii* B13 under different growth conditions, and secondly, by quantifying subpopulation sizes of computer-generated mixed populations.

RESULTS

Stationary phase induction of P$_{int}$-*egfp* in *P. knackmussii* B13. Single cell fluorescence can be quantified from a digital image with the help of image analysis software that recognizes cells as objects through thresholding of pixel intensities, and measures their average pixel fluorescence intensity (AGV). AGVs of all cells are typically plotted as histograms, CDFs, or as Q-Q plots. As noticed previously (33, 44), cells of *P. knackmussii* B13 P$_{int}$-*egfp* did not visibly fluoresce during exponential growth on 3CBA, whereas a small proportion of cells in the culture induced *egfp* in stationary phase (Figure 1A and B). This difference is reflected in the shapes of the histograms that can be constructed from the AGVs of cells grown under these conditions; in the histograms of Figure 1 both populations look similarly in that they follow the shape of a normal distribution. However, paying attention to detail, it can be seen that under stationary phase conditions, a small proportion of cells manifests as a far-stretched right-hand tail of the histogram (Figure 1B and lower boxplot), which under exponential phase conditions is missing (Figure 1A and lower boxplot). The eGFP expression of such cells could be considered as outliers, or they could comprise a separate subpopulation, in which case the distribution of the data would be bimodal. The distribution is visualized more clearly in a boxplot representation, where, under stationary phase conditions, the histogram upper tail corresponds to boxplot outliers (Figure 1B, 2A). A CDF shows this particular subpopulation of cells with high eGFP expression as a 'kink' (Figure 2B, also see (44)), while in a normal Q-Q plot two lines with different slopes can be seen (Figure 2D, also see (27)). In all representations it becomes apparent that there is a subpopulation of cells behaving differently, but the Q-Q plot representation indicates that the data are bimodal. On the other hand, mean values alone, as commonly used as a measure in averaged samples, would not have revealed the bimodal nature of the population.

Which parameters would best describe and quantify the subpopulation effect? Quantification of the extent of bimodality is particularly important when less evident differences in population responses occur or effects of e.g., mutations need to be interpreted. For example, previous analysis suggested that P$_{int}$-*egfp* is induced more strongly under stationary phase conditions when cells are pre-grown on 3CBA than on fructose (44). This interpretation was based on use of distribution-free analyses and parameters such as the 95th percentile, the boosted mean or the mean of the top 5% of the population in a CDF (33, 44). Although these methods have worked satisfactorily to conclude that cells that had grown on 3CBA were different from those grown on fructose (44), they did not provide a biological explanation for the

choice of the 95% percentile-AGV value. Other distribution-free parameters like the boosted mean (mean of AGV between 75 and 95th percentile) or mean of the top 5% AGV of the population also permitted statistical differentiation of eGFP expression from P_{int}-*egfp* in cultures of *P. knackmussii* B13 under different growth conditions, but did not allow calculation of the actual subpopulation size (44). Therefore, we decided to follow another approach that aimed to separate the bimodal data, which would allow the level of induction to be described in terms of the percentage of induced cells of the total population and mean AGV of induced cells. Because these methods rely on splitting of the population into large and small subpopulation, we refer to these methods as population splitting (PS) methods.

Quantile-quantile plot interpretation of bimodality. When plotting all AGV values in cumulative order as a function of their theoretically derived normally distributed ranking number, a so-called normal Q-Q plot, normally distributed AGV values among a population will become visible as a straight line (Figure 2C, also see (27, 32)), the slope of which corresponds to the standard deviation of the population. The median AGV in a normal Q-Q plot is found at the ranking number of 'zero' (Figure 2C-F). Deviations from a normal distribution will become visible in the normal Q-Q plot as deviations from the straight line (Figure 2D). Ideally, bimodal normally distributed subpopulations appear as two intersecting straight lines with different slopes (and therefore different standard deviations). Indeed, while AGV values of single cells in exponentially growing populations of *P. knackmussii* B13 cells expressing *egfp* from P_{int} were distributed along a single straight line (Figure 2C), AGVs from cells in stationary phase distributed in the diagram along two straight lines with different slopes (Figure 2D). Calculation of the size of the (eGFP inducing) smaller subpopulation would thus in essence consist of finding a statistically correct approximation of the point where the two straight lines would intersect and subsequent determination of the number of data points in each population. However, this proves difficult because it is impossible to determine *a priori* whether cells close to the intersection point would belong to one or the other subpopulation. Nevertheless, because of the large size of the 'eGFP uninduced' subpopulation (large subpopulation) compared to that of the eGFP inducing one (small subpopulation), a highly robust linear regression can be calculated for the large subpopulation on basis of a sub-sample of this subpopulation. We took this sub-sample as equivalent to the approximate interquartile range (IQR) (Figure 2E) of the large subpopulation. The large subpopulation IQR can be calculated from all AGV points between visually placed minimum and maximum AGVs (grey area: Figure 2E, Protocol supporting information (S)1, which can easily be estimated from a normal Q-Q plot.
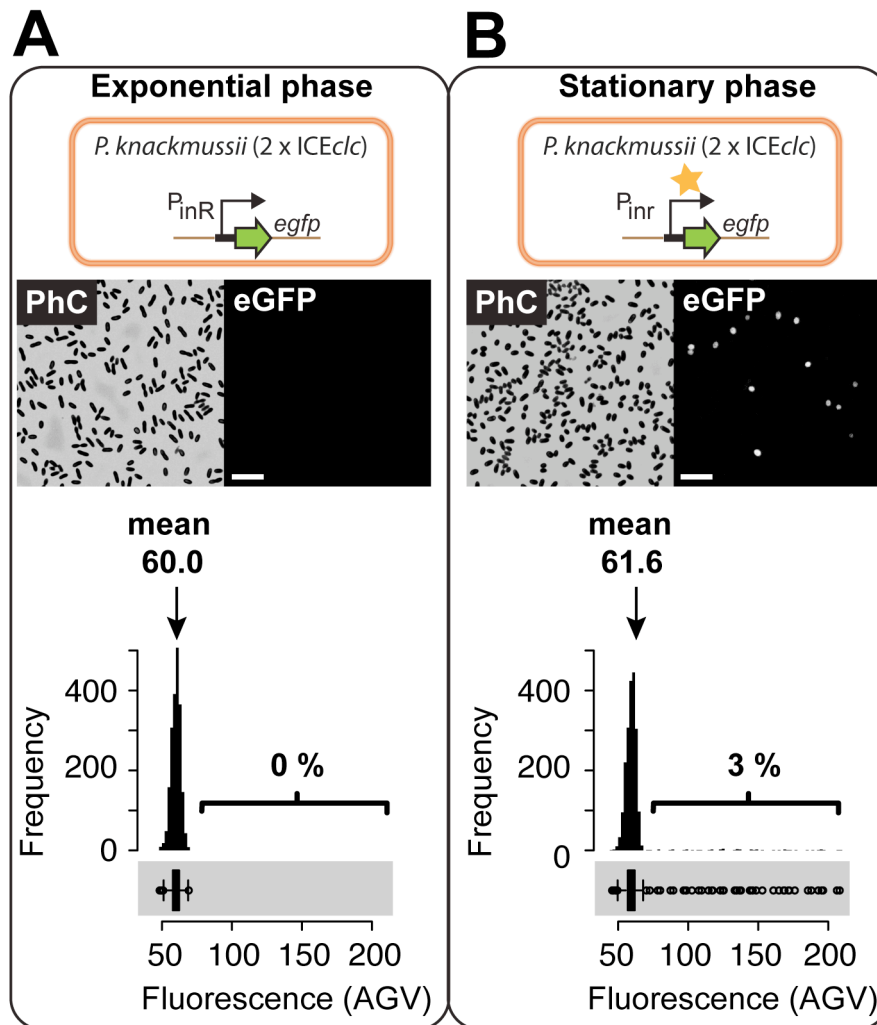
Figure 1. eGFP expression from a monocopy randomly inserted P$_{InR}$-*egfp* fusion in planctonic cells of *P. knackmussii* B13 grown in batch culture and sampled in exponential phase (A) or stationary phase (B). Micrographs show typical population differences of cells grown on 5 mM 3-chlorobenzoate (3CBA) under non-inducing (exponential phase) and inducing conditions (stationary phase), taken under eGFP illumination (right) and the corresponding image in phase contrast (PhC, left). The white bar in images corresponds to a scale of 10 μm. Graphs show fluorescence values (AGVs) measured from single cells represented as histograms and lateral boxplots (grey area below graph). Percentages correspond to calculated sizes of subpopulations statistically significantly expressing eGFP. Note that the calculated mean fluorescence values over the whole population are statistically significantly different if assuming both are normally distributed (P=0.00056, Welch 2-sample t-test).

Since the slope in a Q-Q plot corresponds to the standard deviation it can be used to calculate the upper cutoff value at the 1% confidence level assuming that the large subpopulation is normally distributed (Figure 1E): $cutoff = 2.576 \times SD + median$, where 2.576 is the constant of the quantile function of the normal distribution with probability 0.995, SD is the standard deviation of the large subpopulation and $median$ is the median of the large subpopulation. When applying such method, we calculated that 2.8% of cells in stationary phase cultures of *P. knackmussii* B13 P_{int}-*egfp* grown on 3CBA and 1.2% in cultures grown on fructose expressed *egfp* statistically different from the large subpopulation (Figure 2F, Table 1). The method, therefore, permitted calculation of subpopulations of proportionally low abundance ($\approx$ few percent of the total).

This method was termed *Default* in *R* to distinguish it from three other methods of subpopulation separation proposed in this study: *Manual*, *Boxplot1.5*, and *Boxplot3* (Protocol S1). *Manual* allows the user to manually distinguish large and small subpopulation by visually placing the cutoff value between the two subpopulations on a Q-Q plot (this can be done in *R* by use of the *locator(...)* function, which reads the position of the graphics cursor when the mouse button is pressed; see Protocol S2, Figure S3). Alternatively, the same procedure can also be done on a histogram, in which case the histogram peak-to-tail border has to be visually determined (Figure S3). Bates and collegues (4) deduced subpopulation size by determining the midpoints of histogram peaks. However, when comparing histogram mid-point determination versus histogram peak-to-tail border determination as means to define subpopulation we found the latter more reliable (Figure S3). A similar idea based on manual placement of population separation aids has been used previously (although without the use of interactive graphs), where visually placed tangents in a CDF plot were employed and approximate reading by eye determined the cutoff point between small and large subpopulation (34). The methods *Boxplot1.5* and *Boxplot3* both work simply by applying commonly used formulas for outlier detection in boxplots (12, 20); here we consider the upper tail outliers as part of the small subpopulation and represent them as a percentage of the whole population. *Boxplot1.5* uses the formula $cutoff_{mild} = Q_3 + IQR \times 1.5$, where Q_3 is the 3rd quartile of the data, IQR the interquartile range, and *cutoff* the lower limit for mild outlier determination. Similarly, *Boxplot3* uses the formula $cutoff_{extreme} = Q_3 + IQR \times 3$ for extreme outlier determination.
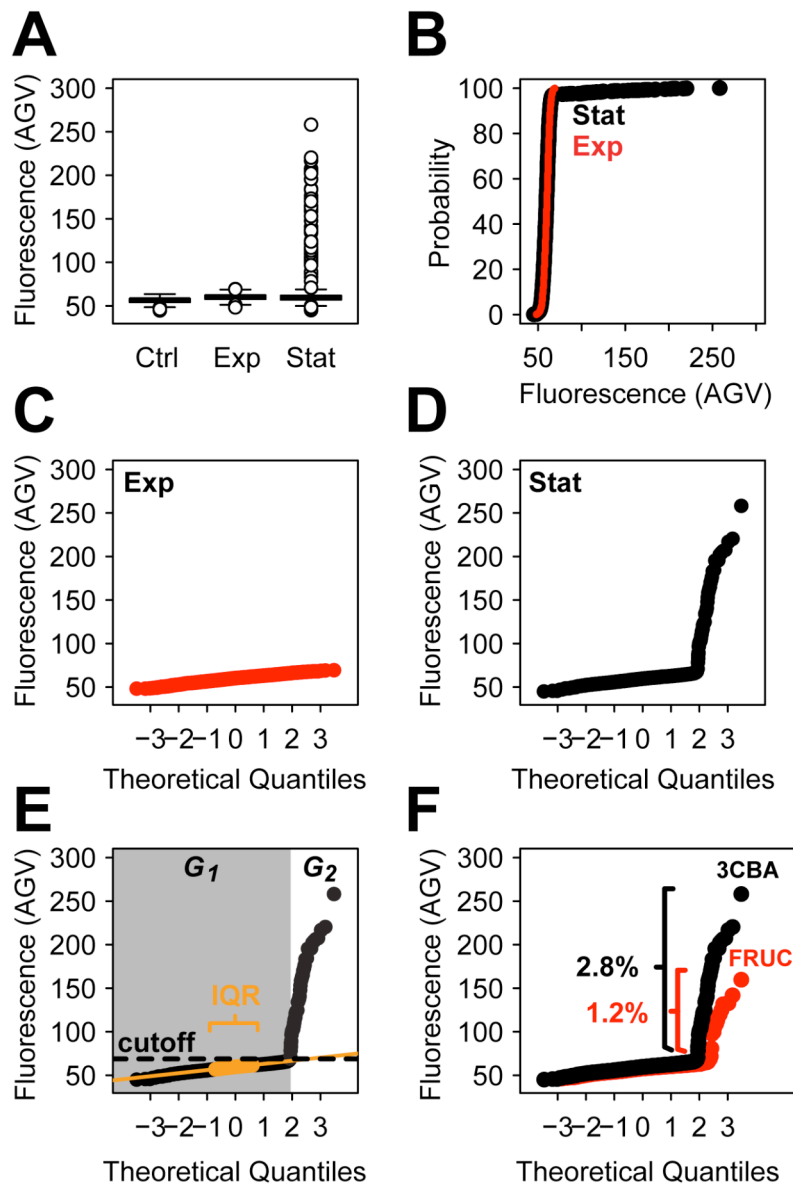
Figure 2. Distribution of eGFP fluorescence intensities (AGV) in cells of *P. knackmussii* B13 strain 1343 (single copy insertion of a P$_{Int}$-*egfp* fusion) taken at exponential or stationary phase after growth on 10 mM 3CBA. (A) Boxplot representation. Ctrl, wildtype *P. knackmussii* B13 without eGFP. (B) Cumulative distribution curve representation (CDF). Stat, stationary phase. Exp, exponential phase. (C, D) Same as (B) but as normal quantile-quantile (Q-Q) plot representations. (E) Same as (D) but data is categorized in large subpopulation (*G$_1$*, grey area) and small subpopulation (*G$_2$*, white area) below and above a cutoff line (cutoff), respectively. The placement of the cutoff line is determined via the slope of a regression line (orange line) fitted to the data points belonging to the interquartile range (IQR, orange) of the large subpopulation. (F) Distribution of eGFP fluorescence intensities in cells taken at inducing conditions (stationary phase) grown on either 10 mM 3CBA (CBA, black) or 10 mM fructose (Fruc, red). Percentages express subpopulation fractions of fructose and 3CBA induced cells (see further Table 1).

Table 1. Varying subpopulation sizes of ICE*clc* transfer competent cells in *P. knackmussii* B13-1343 P_{int}-*egfp* grown on different carbon sources.

Category	Carbon source[1]	% Subpopulation[2]	Significantly different category[3]
A	3CBA	4.7 ± 1.4	B*, C**, D**, E**, F*
B	Fructose	2.2 ± 0.4	A*, C*, D*, E*, F*
C	4-Hydroxybenzoate	0.6 ± 0.2	A**, B*
D	Anthranilate	0.3 ± 0.2	A**, B*, F*
E	Benzoate	0.1 ± 0.3	A**, B*, C*
F	Glucose	0.7 ± 0.5	A*, B*, D*

1) 10 mM of carbon source in minimal medium (see Methods).

2) Average ICE*clc* transfer competent subpopulation of cells (percent of total) determined from biological triplicates, expressing *egfp* from P_{int} ± standard deviation. Sampled 15 - 20 h after onset of stationary phase. Determined via *R* command *find.sub.pop(...)* in *Default* mode.

3) * and ** indicate significant differences at $P<0.05$ and $P<0.01$, respectively, as determined by the Welch Two Sample t-test.

Method comparison. To compare methods that relied on population splitting (PS) into large and small subpopulation (*Default*, *Manual*, *Boxplot1.5*, *Boxplot3*) to methods that did not (*Mean*, *Boosted Mean*, *95th percentile*, *Mean Top 5%*), we analyzed small subpopulation sizes of cells defined by eGFP expression from both the P_{int} and the P_{inR} promoters inserted in single copy in *P. knackmussii* B13 derivatives, and grown under different conditions (Figure 3, Table S4, S5). *P. knackmussii* cultures in 3CBA were typically growing exponentially between 8 and 20 h after inoculation, whereas stationary phase (i.e., cessation of growth) was reached after 24 h (Table S2). *P. knackmussii* cultures in fructose were typically growing exponentially between 20 and 40 h after inoculation, and reached stationary phase after 45 h (Table S2). In contrast, *P. knackmussii* cultures on glucose grew slightly faster and reached stationary phase after 12 h (Table S2). We further tested benzoate, 4-hydroxybenzoate and anthranilate (Table 1). Cultures on anthranilate grew much slower, with stationary phase reached after 50 h (Table S2). Analysis of all culture conditions indicated that growth on 3CBA elicited the strongest induction of P_{int} and P_{inR} promoters in comparison to the others (Table 1, Figure 3, Table S4). Further, PS methods indicate that a larger subpopulation of P_{int}-*egfp* expressing cells is formed on fructose in comparison to glucose, benzoate, and the other two aromatic compounds (Table 1, Figure 3A, Table S4). In contrast, with the exception of *Mean Top 5%*, non-PS methods failed to distinguish between 3CBA-, fructose- and glucose-grown induction (Figure 3B, Table S5). We therefore conclude that the PS methods are more sensitive to small but consistent changes in subpopulation sizes than non-PS methods.
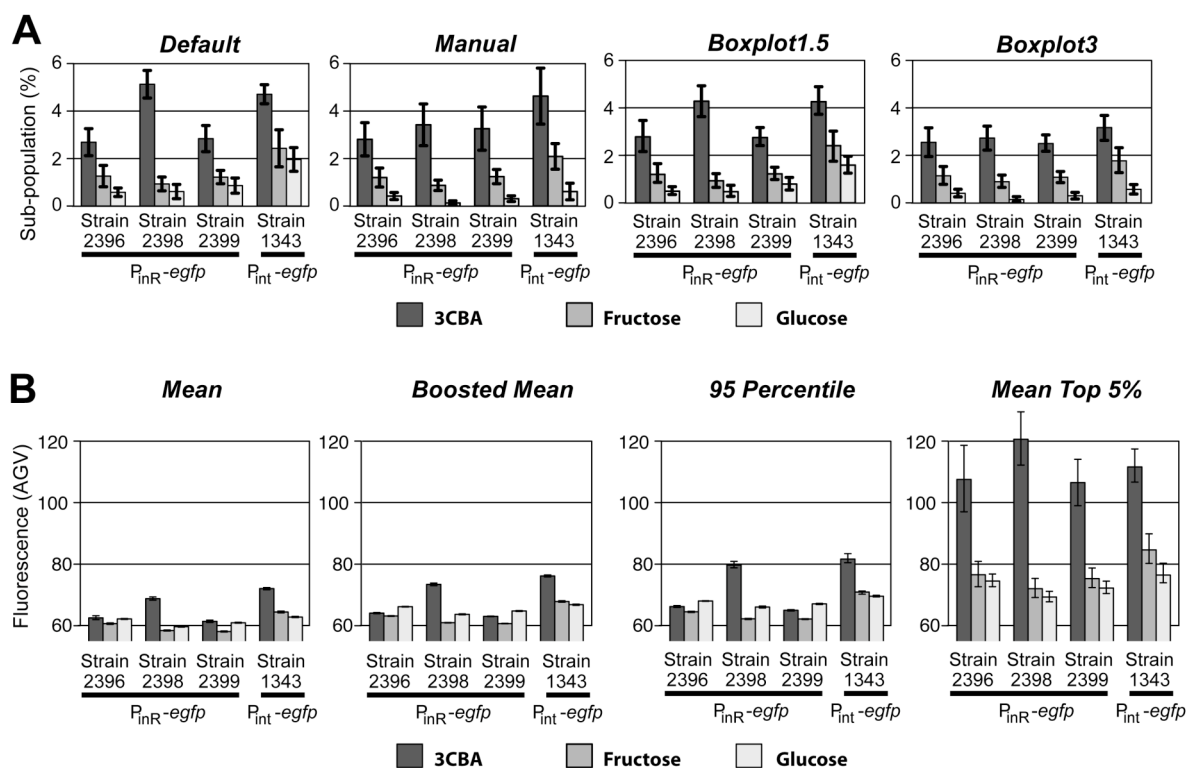
Figure 3. Different methods for quantification of subpopulation sizes of P$_{Int}$-*egfp* or P$_{inR}$-*egfp* expressing cells. (A) Output of four different PS methods for subpopulation size. For each method the same data set was used. (B) Same data as (A) but quantified via distribution-independent non-PS methods that do not determine subpopulation size. Error bars indicate the 95% confidence interval for re-sampled (bootstrapped) data. Dark grey bars: 3CBA grown cells; intermediate grey bars: fructose-grown cells; light grey bars: glucose-grown cells.

Method robustness analysis by bootstrapping. In order to assess the robustness and accuracy of estimating small subpopulation sizes using different PS methods, we tested each PS method separately on a number of slightly varying bimodal populations. For this purpose we used bootstrapping with re-sampling (with replacement) data from wet experiments followed by the PS method and calculation of 95% confidence intervals. Bootstrapping was carried out with 20 replicates for the manual PS methods *Default* and *Manual*, and 500 replicates for all other methods, PS and non-PS. The bootstrapping procedures were implemented in the *R* functions *get.ci(…)* and *get.ci.other(…)* (Protocol S1, 4) for PS and non-PS methods, respectively, both of which keep a record of the results after each replicate and calculate 95% confidence intervals (Figure 3, 4, 5). We compared eight different methods using the same data set including four PS (Figure 3A) and four non-PS methods (Figure 3B). Bootstrapping results indicate that, although less sensitive to small subpopulation changes, most non-PS methods are much more robust than PS methods; that is, they form smaller confidence intervals in response to random variations in data. An exception is the non-PS method *Mean Top 5%*, whose 95% confidence intervals look similar to those of the PS methods. Interestingly, *Mean Top 5%* is also the only non-PS method that confirmed a statistically significant eGFP fluorescence subpopulation change in *P. knackmussii* B13 P_{int}-*egfp*/P_{inR}-*egfp* grown on 3CBA versus grown on fructose or glucose (Table S4). However, *Mean Top 5%*, like all other non-PS methods but unlike most PS methods, failed to indicate a statistically significant difference between growth on fructose and growth on glucose (Table S5). The extreme robustness to random variation as seen in the methods *Mean*, *Boosted Mean*, and *95th Percentile*, might explain part of the reason why these methods fail to respond significantly to small changes in small subpopulations (Figure 3B, Table S5). On the other hand, PS methods *Default*, *Manual*, *Boxplot1.5*, and *Boxplot3*, showed comparably large confidence intervals, reflecting some inconsistency in separating small subpopulations from large subpopulations (Figure 3A, Table S4). Nevertheless, all PS methods distinguished between small subpopulation sizes of 3CBA-grown versus fructose-grown or glucose-grown *P. knackmussii* B13 P_{int}-*egfp*/ P_{inR}-*egfp*. Furthermore, PS methods *Manual*, *Boxplot1.5* and *Boxplot3.5* even showed significant differences between fructose-grown and glucose-grown *P. knackmussii* B13 P_{int}-*egfp*/ P_{inR}-*egfp*. Thus, our experiment showed that, while non-PS methods are generally more robust to overall variation in populations, they are also less sensitive to small subpopulation changes than PS methods.

Confidence interval calculation via bootstrapping may be particularly useful in cases where subpopulation determination is biased. As an example, subpopulation determination

according to the PS method *Manual* is inherently biased due to human subjectivity in placing the cutoff point on a Q-Q plot where subpopulations should be separated. This problematic can be diminished, however, by repeating the method several times on a resampled dataset (bootstrapping) and calculating the confidence interval. As another example case for the use of bootstrapping, normal Q-Q plot representation of P_{int}-*egfp* expression in *P. putida* UWC1 typically manifested as a curve (Figure 5) rather than the two lines of different slopes as seen in *P. knackmussii* B13 (Figure 2), which complicated the finding of the point of separation between subpopulations. However, re-applying PS methods on re-sampled datasets helped to define the confidence limits of the subpopulation determination itself (Figure 5). Another demonstration of such a case is shown in Figure S1, where a dataset that includes biases due to faulty data recording during image acquisitions is subjected to PS methods. This case also highlights the usefulness of summarizing single cell data as boxplots per image, which makes possible the filtering-out of image-inherent bias in a data set (Protocol S2).
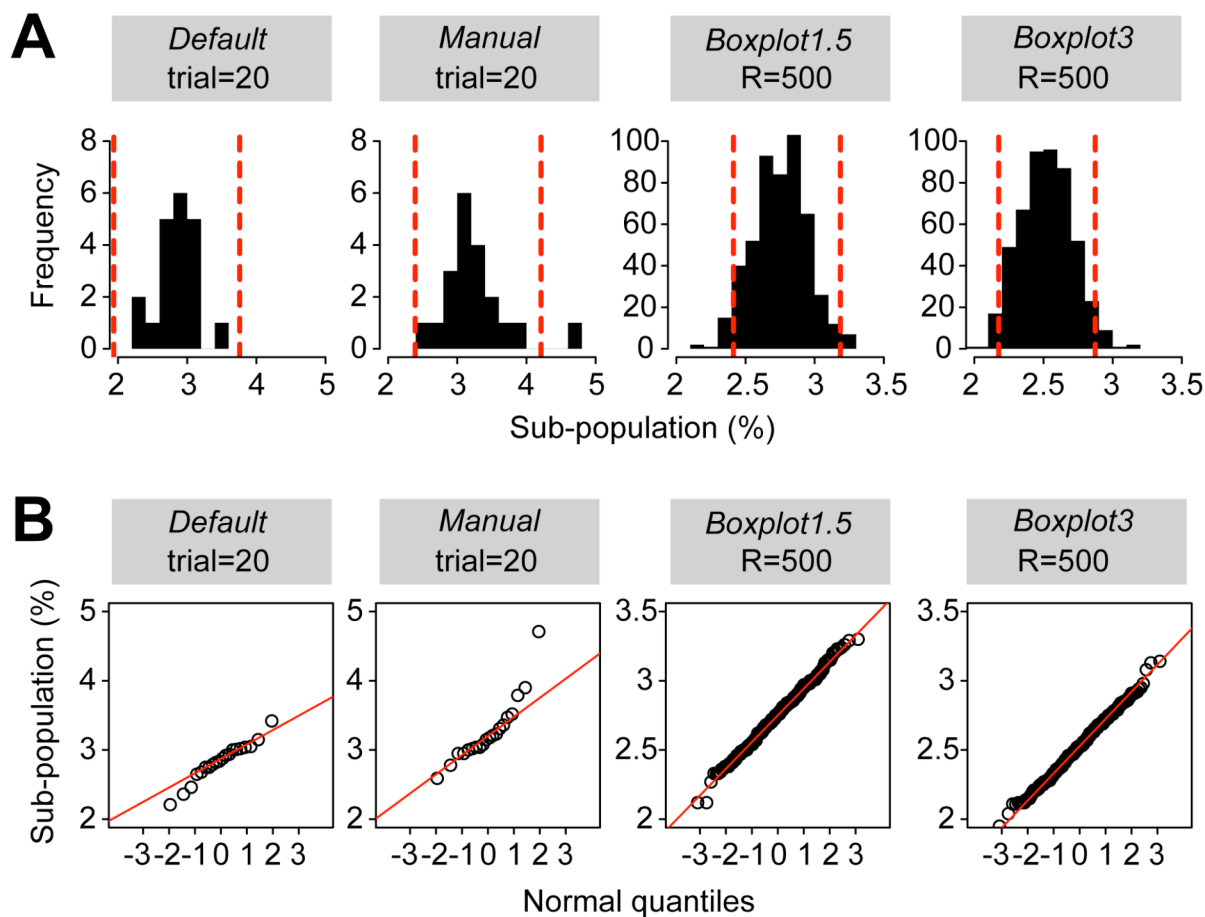
Figure 4. Effect of re-sampling methods of original data sets on the determination of confidence intervals for the subpopulation size of *egfp*-expressing cells in stationary phase cultures of *P. knackmussii* B13 strain 2399 (single copy P_{inR}-*egfp*) grown on 3CBA. (A) Bootstrapping of original data sets (re-sampling with replacement). Methods *Default* and *Manual* were repeated 20 times with manual intervention of the slope line determination. Methods *Boxplot1.5* and *Boxplot3* use 500 automatically re-sampled data sets. 95% confidence intervals (red, dotted lines) were calculated assuming a normal distribution of the results (mean $\pm$ SD $\times$ 1.96). (B) same data as in (A) but re-sampled subpopulation size determinations plotted as Q-Q plots. Note the normal distribution of the results.
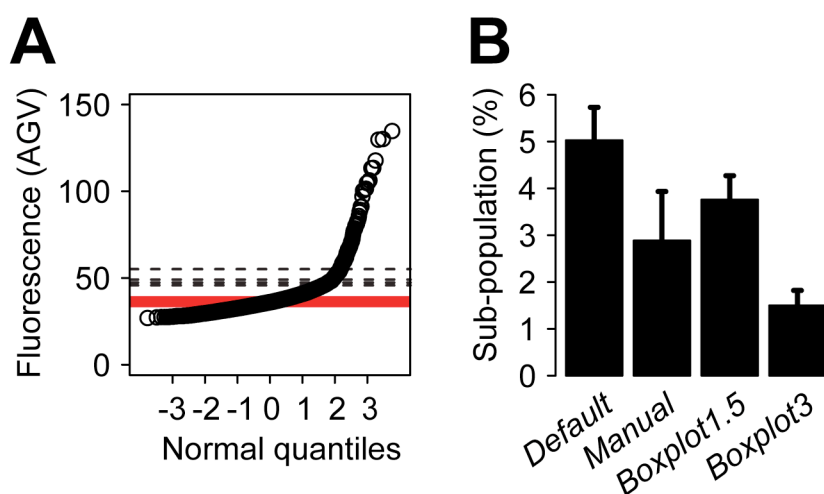
Figure 5. Example of a data set showing poorer Q-Q plot performance (smooth curve of distributed data points). (A) Q-Q plot of single cell eGFP fluorescence values obtained from *P. putida* UWC1-ICE*clc* P$_{int}$-*egfp* cells (strain 2508) grown on 3CBA to early stationary phase. Width of the red line corresponds the interquartile range of the fluorescence values. Dotted lines indicate threshold line placement for subpopulation calculation via methods *Boxplot3*, *Boxplot1.5*, *Default* and *Manual*. (B) Results from the four different subroutines on this data set. Error bars represent 95% confidence intervals on re-sampled data sets with bootstrapping (see Figure 4).

Computer simulations. Following the empirical bootstrap approach above, we wanted to test the performance of our proposed methods on a large variety of bimodal populations. Hereto we used computer simulations that not only allowed to treat large data sets but also had the added advantage that "true" subpopulation parameters were known before analysis. Thus, by comparing true and estimated subpopulation ratios, the accuracy of each PS method in estimating subpopulation proportions could be assessed, which we calculated according to: $100 \times \left[\left(S_{estimated} - S_{true} \right) / S_{true} \right]$, where $S_{estimated}$ is the estimated subpopulation size, and S_{true} is the true subpopulation size, both expressed as a percentage of the total population. In this way we could consider how the accuracy varies with different bimodal population paramters. In a first experiment we tested the accuracy of separating two subpopulations across a range of 64,000 simulated bimodal populations (Figure 6, 7, Video S1-S3). The populations were produced by mixing a single large subpopulation with a variety of smaller-sized subpopulations (Figure 6). To create the large subpopulation we used typical population parameters as found in non-induced populations of *P. knackmussii* B13 containing a P$_{inR}$-*egfp* fusion (Table S3). By plotting the method bias against true subpopulation size, true subpopulation standard deviation and/or true subpopulation mean, we now obtained an overview of the accuracy and robustness of the separation method, presented, for space reasons, either as selected representative 3D plots (Figure 7) or selected representative 2D plots (Figure 8). Only two of the four separation methods were tested in this way, *Boxplot1.5* and *Boxplot3*, since it would have been an almost impossible feat to test the other methods *Default* and *Manual* on an equally large number of datasets due to their requirement of a manual work-flow (mouse-clicking on an interactive graph). However, *Default* and *Manual* were still tested on a smaller scale including fewer simulated bimodal populations (Table 2). The simulation results show that *Boxplot1.5* and *Boxplot3* give less than 20% inaccuracy between the estimated and true population size over the entire span of tested subpopulation standard deviations (10-50) as long as the difference between means of the large and the small populations remains between 40 and 50 units, respectively (Figure 7, Figure 8, Video S1-S3, Table S6). Furthermore, the simulations indicated that separation becomes less accurate when the size of the small subpopulation is decreasing to below 1.1% or values in the small subpopulation become more diverse (i.e., higher standard deviation) (Figure 7, Figure 8, Video S1-S3, Table S6). Both methods also become rapidly unreliable when small subpopulation proportions become larger than 25% (Figure 7, Figure 8, Video S1-S3, Table S6), a feature also confirmed in another simulation experiment (Table 2). This is because outlier detection in boxplots beyond this point is not synonymous with bimodality anymore (Figure S2). However,

we found that the Q-Q plot- based PS methods *Manual* and *Default* could still be used to accurately determine subpopulations larger than 25%, since Q-Q plots show bimodality over a large range of subpopulation proportions (Table 2, Figure S2).

With respect to the decreasing accuracy with decreasing small subpopulation sizes we conducted a second series of simulations dedicated to very small subpopulation sizes focussing on subpopulation proportions between 0.1 and 1.2% (Figure 9, 10, Video S4-S9, Table S7). Overall, *Boxplot3* manifested itself as the more robust and accurate method than *Boxplot1.5* for determining very small subpopulations. More specifically, population size estimations in *Boxplot3* were never less than 11% from the true population size (n=200000) over the entire span of percentage parameters tested, provided the mean difference was at least 67.8 units and standard deviation was set at 37.7 units (Figure 10, Table S7). By comparison, under the same conditions, estimations from *Boxplot1.5* were only within less than 20% from the true value when the tested subpopulation was larger than 1.1%, exponentially increasing to 352% where subpopulations were approaching 0.1% (Figure 10, Table S7).
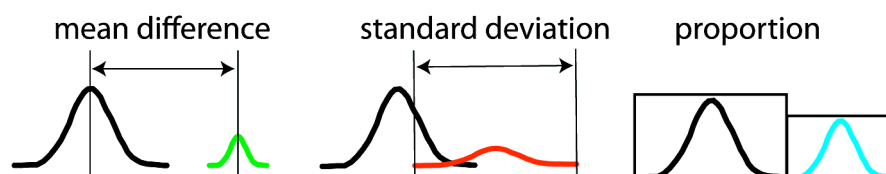
Figure 6. Scheme illustrating the three parameters, mean difference between large and small subpopulation, standard deviation of small subpopulation and proportion of small subpopulation. Parameters were changed in a computer simulation to create variations of mixed populations upon which the PS methods of subpopulation determination were tested (see Figure 7, 9). Black, large subpopulation. Colour, small subpopulation.

Figure 7 (on the right). 3D surfaces of simulation data showing the accuracy (*z* axis) in determining the subpopulation size using two different methods of population separation: *Boxplot1.5* and *Boxplot3*. Accuracy is shown as a function of different population mixtures (1600 per plot), with subpopulations either varying in mean differences (range: 2-137; n=40; *x* axis) and proportions (range: 0.1-40%; n=40; *y* axis) at a constant standard deviation (37.7) (A, B), or varying in mean differences (range: 2-137; n=40; *x* axis) and standard deviations (range: 10-50; *y* axis) at a constant proportion (3.2%) (C, D), or with varying standard deviations (range: 10-50; n=40; *x* axis) and proportions (range: 0.1-40%; n=40; *y* axis) at a constant mean difference (67.8) (E, F). Accuracy is taken as the difference between calculated and real subpopulation size and expressed as a percentage of the real subpopulation size. Accuracy therefore indicates the normalized deviance of the calculated subpopulation size from the real subpopulation size. A negative value indicates that the method underestimated the subpopulation size. A positive value indicates an overestimated result. A value of zero indicates perfect match between estimation and true value. A smooth surface of the same colour indicates a robust separation. NA, missing values.
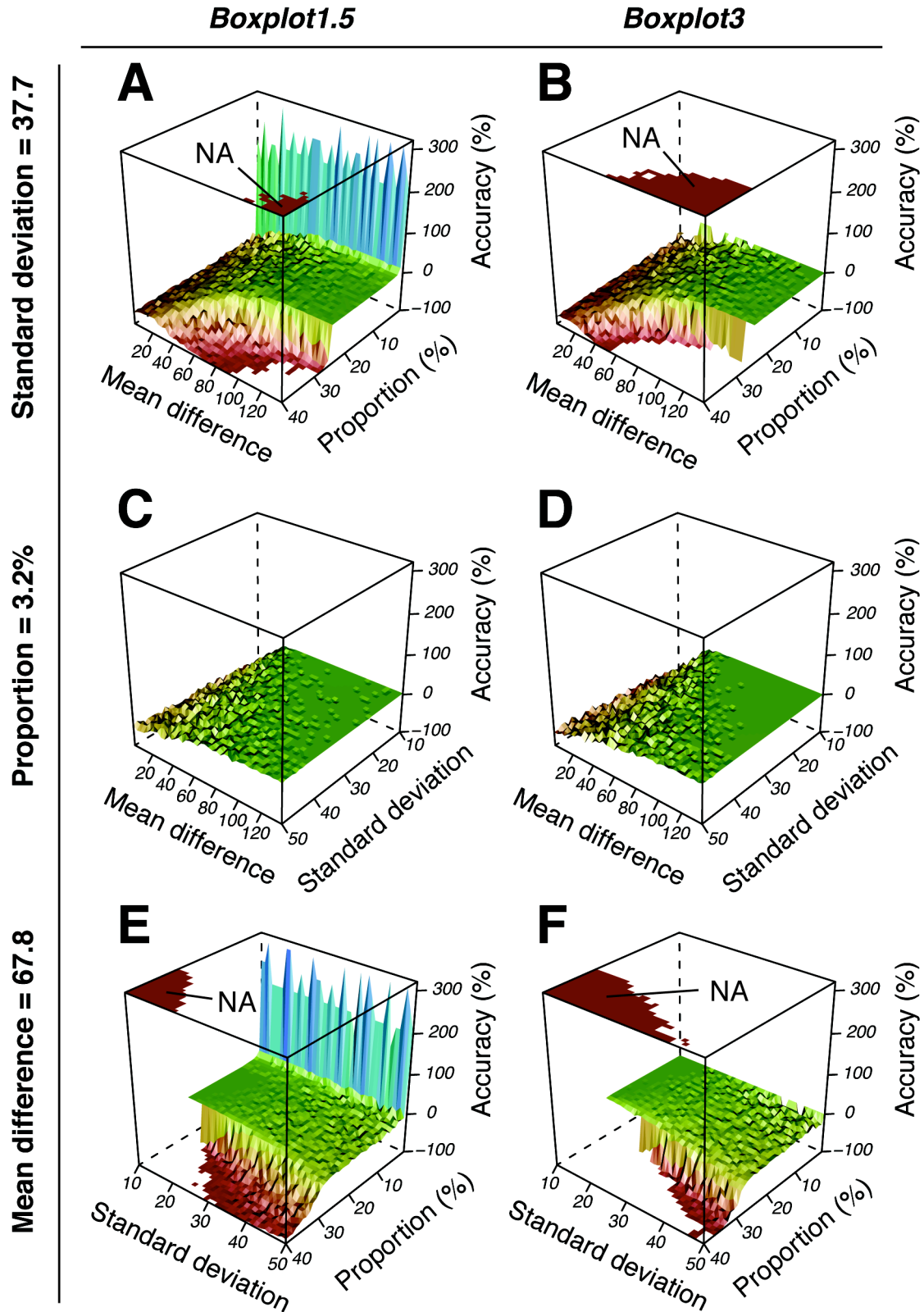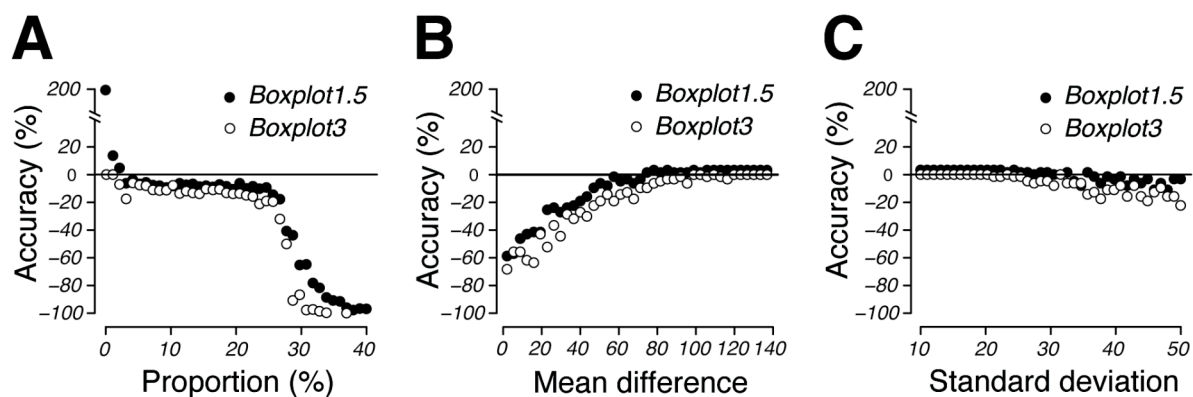
Figure 8. 2D representations of simulations shown in Figures 7A-F. Accuracy is shown as a function of subpopulation proportion (range: 0.1-40%; n=40) at a mean difference of 67.8 and a subpopulation standard deviation of 37.7 (A), as a function of mean difference (range: 2-137; n=40) at a subpopulation proportion of 3.2% and subpopulation standard deviation of 37.7 (B), or as a function of subpopulation standard deviation (range: 10-50; n=40) at a mean difference of 67.8 and a subpopulation proportion of 3.2% (C). Also see Table S6 for values of these graphs.

Figure 9 (following page). Same as Figure 7 but as a result from a simulation focusing on subpopulations with small proportions (range: 0.1-1.2%). Each surface is constructed from 250 data points, stemming from population separations of population mixtures with varying subpopulations with 15 different mean difference values (range: 2-137) and 15 different proportion values (range: 0.1-1.2%) at a constant standard deviation of 38.6. (A, B): Simulation was performed with population mixtures with n=2000. (C, D): Simulation was performed with population mixtures with n=20,000. (E, F) Simulation was performed with population mixtures with n=200, 000.
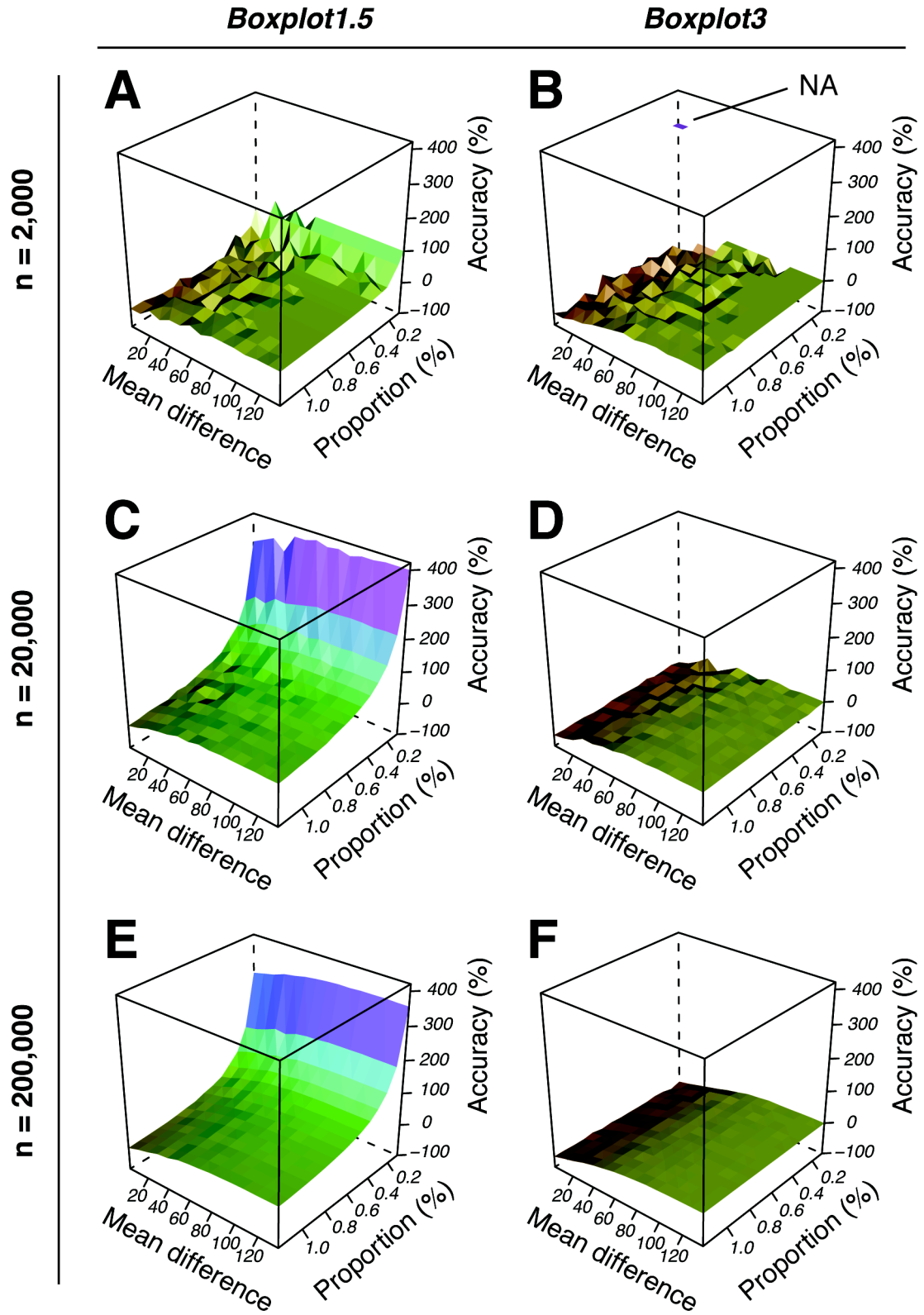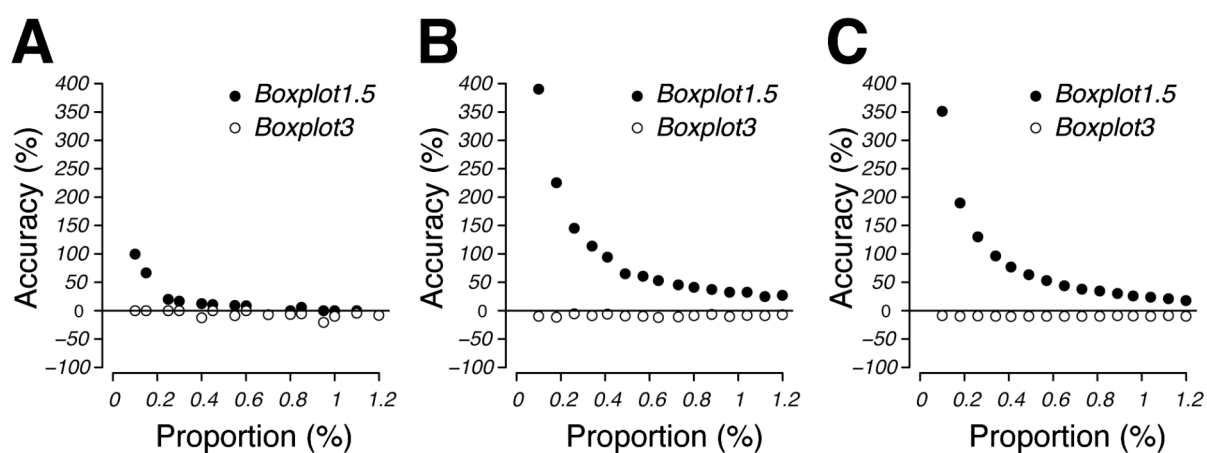
Figure 10. 2D representations of simulations shown in Figures 9A-F. Accuracy is shown as a function of subpopulation proportion (range: 0.1-1.2%; n=15) at a mean difference of 67.8 and a subpopulation standard deviation of 37.7. (A): Simulation was performed with population mixtures with n=2000. (B): Simulation was performed with population mixtures with n=20000. (C): Simulation was performed with population mixtures with n=200000. Also see Table S7 for values of these graphs.

Table 2. Comparison of the accuracy of population splitting (PS) methods in estimating subpopulations from simulated bimodal populations with different true subpopulation sizes.

True subpopulation size (% of total population)[1]	Estimated subpopulation size (% of total population)[2]			
	Default	*Manual*	*Boxplot1.5*	*Boxplot3*
1.00	1.75 ± 0.17	0.98 ± 0.03	1.17 ± 0.18	0.98 ± 0.06
3.00	3.80 ± 0.93	2.77 ± 0.20	2.92 ± 0.08	2.63 ± 0.21
6.00	5.80 ± 0.31	5.85 ± 0.26	5.57 ± 0.03	5.45 ± 0.05
9.00	8.80 ± 0.10	8.40 ± 0.30	8.28 ± 0.13	7.75 ± 0.35
12.00	11.35 ± 0.15	11.58 ± 0.35	10.97 ± 0.12	10.42 ± 0.13
15.00	14.25 ± 0.18	14.12 ± 0.13	13.85 ± 0.10	13.15 ± 0.33
18.00	17.02 ± 0.18	16.67 ± 0.28	16.52 ± 0.25	15.43 ± 0.28
20.00	18.88 ± 0.13	18.73 ± 0.45	17.83 ± 0.06	16.80 ± 0.26
30.00	27.50 ± 0.23	28.58 ± 0.38	12.47 ± 2.32	2.42 ± 1.06
40.00	37.13 ± 0.56	38.05 ± 0.26	1.93 ± 0.31	0.03 ± 0.03
50.00	46.02 ± 0.30	46.18 ± 0.60	0.30 ± 0.15	0.03 ± 0.03
60.00	54.32 ± 0.38	56.95 ± 0.28	0.10 ± 0.10	0.02 ± 0.03
70.00	63.02 ± 0.49	66.02 ± 0.21	0.03 ± 0.03	0.03 ± 0.03
80.00	70.88 ± 1.16	74.55 ± 0.17	0.05 ± 0.00	0.03 ± 0.03
90.00	75.40 ± 2.16	84.32 ± 1.08	0.17 ± 0.14	0.03 ± 0.03
92.00	75.64 ± 1.58	87.01 ± 0.38	0.25 ± 0.15	0.00 ± 0.00
95.00	71.42 ± 1.00	90.03 ± 0.58	0.35 ± 0.17	0.03 ± 0.03
98.00	50.59 ± 18.55	92.20 ± 0.10	0.32 ± 0.18	0.03 ± 0.03

1) True subpopulations were simulated using the *R* function *rnorm(…)* with a standard deviation of 37.7, a mean value of 127.3, and the number of observations corresponding to the subpopulation percentage to be tested from a total number of 2000 observations. Mean and standard deviation used for the simulations represent population parameters as obtained from fluorescence microscopy analysis of batch grown *P. knackmussii* B13 P$_m$-*egfp* in 3CBA (see Table S3).

2) Estimated subpopulation sizes (mean ± SD; 3 independent repetitions) were determined applying the PS methods on simulated bimodal populations using the *R* function *findsub(...)* (Protocol S2). A bimodal population was simulated by mixing two simulated populations, a real subpopulation and a second subpopulation. The second subpopulation was created using the *R* function *rnorm(…)* with a standard deviation of 3.9, a mean value of 63.0, and the number of observations depending on the sample size of real subpopulation to give a total of 2000 observations. Mean and standard deviation used for the simulations represent population parameters as obtained from fluorescence microscopy analysis of batch grown *P. knackmussii* B13 P$_m$-*egfp* in 3CBA (see Table S3).

DISCUSSION

Principal contribution of the study. The principal contribution of this study is a simple and practical statistical approximation to subpopulation quantification in bimodal populations. For this purpose we created a set of functions in the open source software environment *R* accompanied by a step-by-step instructional protocol for easy implementation (Protocol S1, 4).

Motivation of this study. The motivation to define methods of subpopulation quantification was twofold: firstly stemming from a need for a statistical tool do describe subpopulation sizes of ICE*clc* transfer competent cells in *Pseudomonas* in particular (21, 34, 41) and, secondly, to provide a more general set of tools for basic subpopulation quantification in single cell microbiology with easy implementation into existing image analysis work-flows.

Why try to distinguish between subpopulations ? Population-level parameters, such as the average cellular response, by definition will obscure biological detail that is noticeable in small subpopulations of cells. The task of determining the subpopulation sizes of ICE*clc*-transfer competent cells in *P. knackmussii B13* presents itself as a particularly challenging example. Firstly, this is because their proportions are typically small (3.3% of the total population; see Table S3) (33, 41); secondly, they commonly have an estimated mean expression value from the key P_{int}-promoter that is only twice as high as the mean of the non-active population (Table S3). Thirdly, the standard deviation of expression values in this subpopulation is ca. 10 times larger than that of the non-active population (Table S3). Together, this equates to subpopulations that are almost certainly overlapping and thus mixed to some degree, which makes it mathematically impossible to achieve "true" demarcation between subpopulations (32). Histograms of ICE*clc*-activity distributions typically resemble Gaussian curves with hardly noticeable tails extending to their right-hand sides (Figure 1). First, we speculated that such histograms are no suitable visualisations for manually placing subpopulation thresholds confidently and in a statistically acceptable way; even if a threshold was placed such that the histogram tail would be separated from the Gaussian curve, we questioned the reproducibility of such a placement due to a manual work-flow based on grounds of visual perception. Such an approach, we assumed, was likely be prone to bias (user arbitrariness) by subjective decision-making, therefore hindering reliable quantification of subpopulation changes. Indeed, Bates and colleagues (4, 5) offered a "manual" histogram-based approach earlier, which we noticed produced strong variability in subpopulation size determination of ICE*clc* transfer competent cells (Figure S3). Hence, we decided to improve

upon this by using Q-Q plot representations. These have the added advantage of showing two subpopulations, each with normally distributed data of different spread, as two straight lines of different slopes (see, e.g., Figure 2) (27). The point of demarcation between such subpopulations can be determined manually (as in the subroutine *Manual*). Yet, in cases where bimodal distribution patterns are less clear (e.g., Figure 5), we developed a method (named *Default*) that standardizes cutoff placement on grounds of the most reproducible part of the distribution pattern, that is, the part that is most robust to change by subpopulation effects. In a Q-Q plot this region conveniently corresponds to the lower (and longer) straight line, on which an interval of representative slope of that line should be easily definable. Nevertheless, under certain conditions *Manual* can be the more accurate tool (Table 2, Figure S3) and is especially useful in cases where the *Default* algorithm fails, for example in instances with datasets where the IQR of the larger subpopulation does not follow a Gaussian distribution (Figure S1). Generally, when subpopulation quantification becomes challenging/ambiguous, or risks to be influenced by subjective input from the user, it is good practice to apply quantification repeatedly on the same original but re-sampled (with replacement) data set. Importantly, both *Default* and *Manual* are not limited by the proportions of the tester subpopulation in order to produce quantitatively correct results (Table 2), in contrast to *Boxplot* methods.

Boxplot1.5 and Boxplot3 define subpopulations without prompting the user for input since their subpopulation classification is simply based on outlier detection as commonly used in boxplots (12, 20, 54). This latter trait was found especially useful when numerous data sets needed to be analysed as shown in the simulations of this study, where 64000 bimodal populations were analysed within ca. 10 hours (see Figure 7, 9). As expected, Boxplot methods respond differently than manual methods to changes in distribution patterns (Figure 5, Figure S1, S2). *Boxplot1.5* and *Boxplot3* only allowed for reliable subpopulation quantification where subpopulations amounted to less than 25% of the total population (Figure 7, 8, Table 2, Figure S2). This is not surprising, since it is well documented that under certain distribution scenarios boxplots fail to visualize bimodality (11). *Boxplot3* was by far more accurate than *Boxplot1.5* with subpopulations smaller than 1 % of the total population (Figure 9, 10, Video S4-S9, Table S7). In contrast, due to its more conservative classification of outliers, *Boxplot3* tends to underestimate subpopulation sizes in comparison to *Boxplot1.5* (Figure 8). At low subpopulation sizes, it is relevant to increase sample numbers such as can be obtained from flow cytometry experiments (> 20000; see Figure 9), or calculate confidence limits from bootstrappings (e.g., function *get.ci(...),* Figure 5, Protocol S2). In general, when comparing PS methods to existing methods of quantification, we found that PS methods were more sensitive

to small subpopulation effects, for example when assessing ICE*clc* activity in *P. knackmusssii* B13 under different growth conditions (Figure 3, Table S4).

Limitations of the proposed methods. The strength of the proposed methods in this paper is also their weakness; the determination of the percentile corresponding to the cutoff point/threshold between two subpopulations can only be approximated, and becomes more inaccurate as subpopulations overlap. On the one hand approximation allows to split a bimodal population into two and characterize the biologically relevant fraction in a subpopulation response. On the other hand such characterization becomes increasingly inaccurate in describing the biologically relevant fractions until it eventually fails completely as subpopulations overlap. There are only two alternatives to this dilemma, which are analyses that either avoid finding subpopulation-relevant cut-off percentiles altogether, or describe all percentiles in a population, without specifying one. As an example of the second approach are visualisations of qualitative changes of entire populations through comparisons of distribution patterns (27, 32, 44, 53). MacArthur (32) even proposed a way to quantify qualitative changes spanning the total range of percentiles, that is calculating percentage differences per individual percentile between treatment and control (Figure S4, Table S8) (32).

Other studies. Few studies in the microbiology literature specify the problematic of statistically exposing true subpopulations from bimodal populations. Rather, it seems that most studies content themselves with a categorisation of subpopulations via thresholds based on fluorescence background levels, negative controls lacking fluorescent marker, or manual gating of clusters in flow cytometry (1, 7, 9, 13, 33, 49, 57, 58, 61). The reasons might be twofold. Firstly, pragmatism, which argues that as long as an approach serves the purpose of quantification at a sufficiently high resolution it is good enough. Secondly, the problematic that statistical distributions of subpopulation behaviours overlap, causing a certain degree of subpopulation mixing, and therefore make a precise demarcation between subpopulations impossible.

Conclusion. To date there exists no universal protocol in the microbiology literature for the determination of small subpopulation sizes. Rather, many labs use their own in-house methods of subpopulation quantification. We see the advantage/novelty of our proposed methods in the attempt to statistically deduce subpopulation size from a qualitative assessment of the underlying bimodal distributional shape. We argue that a distribution shape-based approach is by definition (inherently) more true to finding the biologically relevant subpopulation than distribution-independent methods. Consequently, our approach should help firstly, to minimize inconsistencies in subpopulation classification caused by manual threshold

placements, and secondly, to increase sensitivity and accuracy to subpopulation changes. Thirdly, our method would help to standardize subpopulation evaluation across different experimental set-ups. Since subpopulation size as expressed as percentage of the total population is a dimension-less quantity, it is also independent of scales and units linked to the sensitivity of recording equipments and experimental set-ups. Therefore, subpopulation size expressed as a fraction of the total population represents a suitable parameter for comparisons across a wide range of different studies.

METHODS

Culture conditions. All strains in this work are listed in Table S1. All strains were batch-cultured in 200 ml Erlenmeyer flasks containing 30 ml liquid minimal at 30°C and with 200 rpm rotary shaking. Type 21C minimal medium (MM) (23) was supplemented with either 3-chlorobenzoate (3CBA, 10 mM), fructose (10 mM), glucose (10 mM), benzoate (10 mM), anthranilate (10 mM) or 4-hydroxybenzoate (10 mM) as sole carbon and energy source. Increase in culture turbidity at 600 nm was followed during growth to estimate the onset of the stationary phase and exact sampling times for epifluorescence microscopy (Table S2). Stationary phase samples (i.e., 10 to 30 hours after cessation of turbidity increase in batch culture) of three microliter were deposited on microscope glass slides, covered with a 0.17 mm cover slip and immediately imaged.

Promoter reporter gene fusions. To examine expression of the P_{int} promoter at single cell level we used previously constructed transcriptional fusions between P_{int} and promoterless *egfp* genes (33), that were inserted in single copy on the chromosome of a variety of *Pseudomonas* strains (Table S1) via mini-Tn5 delivery, and verified by antibiotic selection markers and specific PCR amplification.

Digital imaging. Fluorescence intensities of single cells with or without transcriptional fusions to the *egfp* gene were determined by digital imaging. Single cells were visualized at 1000-fold magnification under a Zeiss Axioscope2 upright epifluorescence microscope equipped with a Spot Xplorer 1.4 MPixel cooled CCD camera (Visitron Systems GmbH, Puchheim, Germany). Images were recorded with phase-contrast illumination (10 ms) and with the filter eGFP HQ470/40 for eGFP fluorescence (excitation wavelength 480 ± 20 nm, emission wavelength 520 ± 20 nm, 500 ms) (Chroma Technology Corp, VT, USA). Average intensity values (AGV) of each cell were determined from 16-bit stored TIF-images using the program METAVIEW (version 6.1r5, Visitron Systems GmbH) using the phase-contrast image as mask for outlining the cells in the eGFP channel. Data were exported to EXCEL (Microsoft Corporation, Redmond, Wash.) or *R* (38). At least 1000 cells were measured for each condition and at least six images were taken per condition or strain.

Programming in *R*. All statistical analysis and computations were processed in *R*. For PS methods, an approach was followed that assumed bimodality of the data (i.e., containing two subpopulations each with a normal distribution). The list of individual cellular AGVs was hereto transferred from EXCEL to a data text file, which was placed into an *R* work-folder. Data were processed according to different PS and non-PS methods in a subroutine

written in *R* named *findsub(...)* (Protocol S1, S2). Essentially, the setting *Default* in *findsub(...)* ranks data according to their AGV and plots the values against a theoretical normal distribution (the normal Q-Q plot) (Protocol S1) (27). Subsequently, the subroutine determines the median and a region around the median to produce the linear regression line for the larger subpopulation. A horizontal separator line is then automatically generated according to $cutoff = 2.576 \times slope + median$, where *cutoff* is the point at which the horizontal separator line is drawn, *slope* is slope of the linear regression line (and therefore the standard deviation of the large subpopulation), and *median* is the median of the data set (Protocol S1). All data points above the horizontal separator line are considered to belong to the smaller subpopulation. The subroutine in *Default* mode further allows manual setting of the range of the large population from which the median value is determined via mouse-clicking on an interactive graph (Protocol S1). Other PS modes of *findsub(...)* include the modes *Manual*, *Boxplot1.5* and *Boxplot3*. While *Manual* allows manual determination of the breakpoint between subpopulations via mouse-clicking on an interactive graph (Protocol S1, Figure S3), *Boxplot1.5* and *Boxplot3* use an outlier algorithm as calculated by the *R* function *boxplot(...)* (*R* graphics package) (Protocol S1). The argument *range* of the function *boxplot(...)* determines how far the plot whiskers extend out from the box beyond which outliers are identified. *Boxplot1.5* uses *range* = 1.5 and *Boxplot3* uses *range* = 3, corresponding to mild and extreme outlier detection, respectively (54). Finally, a fifth mode of *findsub(...)* is the mode *Other*. This mode calculates results according to four non-PS algorithms including the population mean (*Mean*), and the population-independent methods 95th percentile (*95th Percentile*), mean between the 75th and 95th percentile (*Boosted Mean*), and mean of the top 5% of a population (*Mean Top 5%*) (Protocol S1).

Finally, the subroutines *get.ci(...)* and *get.ci.other(...)* were written in *R* (Protocol S2), allowing to bootstrap PS and non-PS methods, respectively, for 95% confidence interval determination. Bootstrapping was carried out via random sampling with replacement of data sets with subsequent application of the method of choice with n repetitions (Protocol S1, S2). For confidence interval calculations with 20 repetitions (*Default, Manual*), a normal distribution of the bootstrapped results was assumed (Protocol S1): $CI_{upper/lower} = mean \pm SD \times 1.96$, where $CI_{upper/lower}$ is the upper or lower confidence interval, respectively, *mean* is the population mean and *SD* is standard deviation. For methods *Default* and *Manual* repetitions were limited to 20 because every calculation requires manual intervention on an interactive graph for the method to work.

For confidence interval calculations with 500 repetitions (all other methods), the *R* function *boot.ci(...)* from the *R* boot package (8, 15) was used with the percentile method of bootstrap confidence interval calculation.

Simulations and data presentation. Bimodal populations were simulated by mixing a large subpopulation with multiple, smaller subpopulations varying in standard deviation, mean and size, respectively. Large and small subpopulations were created with the function *rnorm(...)* of the *R* statistical package (38). Parameters for the creation of the large subpopulation were set to standard deviation *SD*=3.9, and mean *mean*=63, both of which were considered typical values for AGV data sets obtained from stationary phase batch cultures of ICE*clc*-harbouring *Pseudomonas* tagged with a P_{int}-*egfp* reporter and grown on 3CBA (Table S3). Size *N* of the mixed populations was set to 2000, 20000 or 200000. Parameters for the creation of subpopulations were set to all possible combinations of either 40 or 15 equidistantly spaced values for standard deviations, mean values, or population sizes, which in total yielded 40^3 (=64000) or 15^3 (=3375) different subpopulations, respectively. The ranges for 40 equidistantly spaced parameter values were set to 10 to 50 for standard deviations, 65 to 200 for mean values, and 0.1% to 40% of the total population for small subpopulation sizes. The ranges for 15 equidistantly spaced parameter values were set as above except for 0.1% to 1.2% for small subpopulation sizes. Small subpopulation determination was carried out according to the PS methods *Boxplot1.5* and *Boxplot3*. For code and script for the simulation of mixed populations and their separation using *Boxplot1.5* and *Boxplot 3* see Protocol S3. The *R* package "lattice" (43) was used for 3D visualizations of the data by use of the function *wireframe(...)*. The freeware IMAGEJ (version 1.440, USA) was used for creating movies of the visualisations (Video S1-S9).

ACKNOWLEDGEMENT

REFERENCES

1. **Anetzberger, C., T. Pirch, and K. Jung.** 2009. Heterogeneity in quorum sensing-regulated bioluminescence of *Vibrio harveyi*. Mol Microbiol **73:**267-277.

2. **Avery, S. V.** 2006. Microbial cell individuality and the underlying sources of heterogeneity. Nat Rev Microbiol **4:**577-587.

3. **Balaban, N. Q.** 2011. Persistence: mechanisms for triggering and enhancing phenotypic variability. Curr Opin Genet Dev **21:**768-775.

4. **Bates, D., J. Epstein, E. Boye, K. Fahrner, H. Berg, and N. Kleckner.** 2005. The *Escherichia coli* baby cell column: a novel cell synchronization method provides new insight into the bacterial cell cycle. Mol Microbiol **57:**380-391.

5. **Bates, D., and N. Kleckner.** 2005. Chromosome and Replisome Dynamics in *E. coli*: Loss of Sister Cohesion Triggers Global Chromosome Movement and Mediates Chromosome Segregation. Cell **121:**899-911.

6. **Bonifield, H. R., and K. T. Hughes.** 2003. Flagellar phase variation in *Salmonella enterica* is mediated by a posttranscriptional control mechanism. J. Bacteriol. **185:**3567–3574.

7. **Button, D. K., and B. R. Robertson.** 2001. Determination of DNA Content of Aquatic Bacteria by Flow Cytometry. Appl. Environ. Microbiol. **67:**1636-1645.

8. **Canty, A., and B. Ripley.** 2009. boot: Bootstrap R (S-Plus) functions. R package version 1.2-41.

9. **Chai, Y., F. Chu, R. Kolter, and R. Losick.** 2008. Bistability and biofilm formation in *Bacillus subtilis*. Mol Microbiol **67:**254-263.

10. **Chastanet, A., V. Vitkup, G. C. Yuan, T. M. Norman, J. S. Liu, and R. M. Losick.** 2010. Broadly heterogeneous activation of the master regulator for sporulation in *Bacillus subtilis*. Proc Natl Acad Sci U S A **107:**8486–8491.

11. **Choonpradub, C., and D. McNeil.** 2005. Can the Box Plot be Improved? Songklanakarin Journal of Science and Technology **27:**649-657.

12. **Cleveland, W. S.** 1994. The Elements of Graphing Data, revised ed. Murray Hill, New Jersey.

13.   **Cozy, L. M., and D. B. Kearns.** 2010. Gene position in a long operon governs motility development in *Bacillus subtilis*. Mol Microbiol **76:**273-285.

14.   **Davidson, C. J., and M. G. Surette.** 2008. Individuality in bacteria. Annu Rev Genet **42:**253-268.

15.   **Davison, A. C., and D. V. Hinkley.** 1997. Bootstrap methods and their applications. Cambridge University Press, Cambridge.

16.   **de Jong, I. G., K. Beilharz, O. P. Kuipers, and J. W. Veening.** 2011. Live cell Imaging of *Bacillus subtilis* and *Streptococcus pneumoniae* using automated time-lapse Microscopy. J Vis Exp.

17.   **de Jong, I. G., J. W. Veening, and O. P. Kuipers.** 2010. Heterochronic phosphorelay gene expression as a source of heterogeneity in *Bacillus subtilis* spore formation. J Bacteriol **192:**2053-2067.

18.   **de Jong, I. G., J. W. Veening, and O. P. Kuipers.** 2012. Single cell analysis of gene expression patterns during carbon starvation in *Bacillus subtilis* reveals large phenotypic variation. Environ Microbiol **14:**3110-3121.

19.   **Errington, J.** 2003. Regulation of endospore formation in *Bacillus subtilis*. Nat Rev Microbiol **1:**117–126.

20.   **Frigge, M., D. Hoaglin, and B. Iglewicz.** 1989. Some implementations of the boxplot. The American Statistician **43:**50-54.

21.   **Gaillard, M., N. Pradervand, M. Minoia, V. Sentchilo, D. R. Johnson, and J. R. van der Meer.** 2010. Transcriptome analysis of the mobile genome ICE*clc* in *Pseudomonas knackmussii* B13. BMC Microbiol **10:**153.

22.   **Gaillard, M., T. Vallaeys, F. J. Vorholter, M. Minoia, C. Werlen, V. Sentchilo, A. Puhler, and J. R. van der Meer.** 2006. The *clc* element of *Pseudomonas* sp. strain B13, a genomic island with various catabolic properties. J Bacteriol **188:**1999-2013.

23.   **Gerhardt, P., R. G. E. Murray, R. N. Costilow, E. W. Nester, W. A. Wood, N. R. Krieg, and G. Briggs Phillips (ed.).** 1981. Manual of methods for general bacteriology. American Society for Microbiology, Washington, D.C.

24.   **González-Pastor, J. E., E. C. Hobbs, and R. Losick.** 2003. Cannibalism by sporulating bacteria. Science **301:**510-513.

25.   **Kearns, D. B., and R. Losick.** 2005. Cell population heterogeneity during growth of *Bacillus subtilis*. Genes Dev **19:**3083-3094.

26. **Korobkova, E., T. Emonet, J. M. Vilar, T. S. Shimizu, and P. Cluzel.** 2004. From molecular noise to behavioural variability in a single bacterium. Nature **428:**574-578.

27. **Leveau, J. H., and S. E. Lindow.** 2001. Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. J Bacteriol **183:**6752-6762.

28. **Lidstrom, M. E., and M. C. Konopka.** 2010. The role of physiological heterogeneity in microbial population behavior. Nat Chem Biol **6:**705-712.

29. **Locke, J. C., and M. B. Elowitz.** 2009. Using movies to analyse gene circuit dynamics in single cells. Nat Rev Microbiol **7:**383-392.

30. **Lopez, D., H. Vlamakis, R. Losick, and R. Kolter.** 2009. Cannibalism enhances biofilm development in *Bacillus subtilis*. Mol Microbiol **74:**609-618.

31. **Maamar, H., and D. Dubnau.** 2005. Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop. Mol Microbiol **56:**615-624.

32. **MacArthur, B., D., R. Tare, S., C. P. Please, P. Prescott, and R. Oreffo, O., C.** 2006. A non-invasive method for *in situ* quantification of subpopulation behaviour in mixed cell culture. J. R. Soc. Interface **3:**63-69.

33. **Minoia, M., M. Gaillard, F. Reinhard, M. Stojanov, V. Sentchilo, and J. R. van der Meer.** 2008. Stochasticity and bistability in horizontal transfer control of a genomic island in *Pseudomonas*. Proc. Natl. Acad. Sci. U S A **105:**20792-20797.

34. **Miyazaki, R., M. Minoia, N. Pradervand, S. Sulser, F. Reinhard, and J. R. van der Meer.** 2012. Cellular variability of RpoS expression underlies subpopulation activation of an integrative and conjugative element. PLoS Genet **8:**e1002818.

35. **Ozbudak, E. M., M. Thattai, H. N. Lim, B. I. Shraiman, and A. Van Oudenaarden.** 2004. Multistability in the lactose utilization network of *Escherichia coli*. Nature **427:**737-740.

36. **Piggot, P. J., and D. W. Hilbert.** 2004. Sporulation of *Bacillus subtilis*. Curr Opin Microbiol **7:**579–586.

37. **Ptashne, M.** 2006. Lambda's switch: lessons from a module swap. Curr Biol **16:**459-462.

38. **R Development Core Team.** 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

39. **Ravatn, R., S. Studer, D. Springael, A. J. B. Zehnder, and J. R. van der Meer.** 1998. Chromosomal integration, tandem amplification, and deamplification in

Pseudomonas putida F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from *Pseudomonas* sp. strain B13. J. Bacteriol. **180:**4360-4369.

40. **Ravatn, R., S. Studer, A. J. B. Zehnder, and J. R. van der Meer.** 1998. Int-B13, an unusual site-specific recombinase of the bacteriophage P4 integrase family, is responsible for chromosomal insertion of the 105-kilobase *clc* element of *Pseudomonas* sp. strain B13. J. Bacteriol. **180:**5505-5514.

41. **Reinhard, F., R. Miyazaki, N. Pradervand, and J. R. van der Meer.** 2013. Cell differentiation to "mating bodies" induced by an integrating and conjugative element in free-living bacteria. Curr Biol **23:**255-259.

42. **Remus-Emsermann, M. N., and J. H. Leveau.** 2010. Linking environmental heterogeneity and reproductive success at single-cell resolution. ISME J **4:**215-222.

43. **Sarkar, D.** 2009. lattice: Lattice Graphics. R package version 0.17-26 http://CRAN.R-project.org/package=lattice.

44. **Sentchilo, V., R. Ravatn, C. Werlen, A. J. Zehnder, and J. R. van der Meer.** 2003. Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. Strain B13. Journal of Bacteriology **185:**4530-4538.

45. **Sentchilo, V., A. J. Zehnder, and J. R. van der Meer.** 2003. Characterization of two alternative promoters for integrase expression in the *clc* genomic island of *Pseudomonas* sp. strain B13. Molecular Microbiology **49:**93-104.

46. **Siegele, D. A., and J. C. Hu.** 1997. Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations. Proc Natl Acad Sci U S A **94:**8168-8172.

47. **Silander, O. K., N. Nikolic, A. Zaslaver, A. Bren, I. Kikoin, U. Alon, and M. Ackermann.** 2012. A genome-wide analysis of promoter-mediated phenotypic noise in *Escherichia coli*. PLoS Genet **8:**e1002443.

48. **Smits, W. K., C. C. Eschevins, K. A. Susanna, S. Bron, O. P. Kuipers, and L. W. Hamoen.** 2005. Stripping *Bacillus*: ComK auto-stimulation is responsible for the bistable response in competence development. Mol Microbiol **56:**604-614.

49. **Strovas, T. J., and M. E. Lidstrom.** 2009. Population heterogeneity in *Methylobacterium extorquens* AM1. Microbiology **155:**2040-2048.

50. **Strovas, T. J., L. M. Sauter, X. Guo, and M. E. Lidstrom.** 2007. Cell-to-cell heterogeneity in growth rate and gene expression in *Methylobacterium extorquens* AM1. J Bacteriol **189:**7127-7133.

51. Taniguchi, Y., P. J. Choi, G. W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science **329:**533-538.

52. Tecon, R., and J. H. Leveau. 2012. The mechanics of bacterial cluster formation on plant leaf surfaces as revealed by bioreporter technology. Environ Microbiol **14:**1325-1332.

53. Tecon, R., M. Wells, and J. R. van der Meer. 2006. A new green fluorescent protein-based bacterial biosensor for analysing phenanthrene fluxes. Environ. Microbiol. **8:**697-708.

54. Tukey, J. W. 1977. Exploratory data analysis. McGraw-Hill, Reading, MA.

55. van der Woude, M. W. 2011. Phase variation: how to create and coordinate population diversity. Curr Opin Microbiol **14:**205-211.

56. van der Woude, M. W., and A. J. Baumler. 2004. Phase and antigenic variation in bacteria. Clin Microbiol Rev **17:**581-611.

57. Veening, J. W., L. W. Hamoen, and O. P. Kuipers. 2005. Phosphatases modulate the bistable sporulation gene expression pattern in *Bacillus subtilis*. Mol Microbiol **56:**1481-1494.

58. Veening, J. W., W. K. Smits, L. W. Hamoen, and O. P. Kuipers. 2006. Single cell analysis of gene expression patterns of competence development and initiation of sporulation in *Bacillus subtilis* grown on chemically defined media. J Appl Microbiol **101:**531-541.

59. Veening, J. W., W. K. Smits, and O. P. Kuipers. 2008. Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol **62:**193-210.

60. Veening, J. W., E. J. Stewart, T. W. Berngruber, F. Taddei, O. P. Kuipers, and L. W. Hamoen. 2008. Bet-hedging and epigenetic inheritance in bacterial cell development. Proc Natl Acad Sci U S A **105:**4393-4398.

61. Vlamakis, H., C. Aguilar, R. Losick, and R. Kolter. 2008. Control of cell fate by the formation of an architecturally complex bacterial community. Genes Dev **22:**945-953.

62. Wang, P., L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun. 2010. Robust growth of *Escherichia coli*. Curr Biol **20:**1099-1103.

ANNEX: SUPPORTING INFORMATION

DESCRIPTION OF SUPPLEMENTARY FILES

Figure S1. Bias compromises detection of small subpopulations in bimodal data. This file contains a series of graphs that demonstrate the obstructive role of bias in estimating subpopulation size in bimodal data. The left row of graphs are based on a faulty data set with data originating from two images that have much lower fluorescence values as a result of a mistake during image acquisition. The right row of graphs represents the same data set but with the data from the biased images removed. This panel of graphs highlights the practicality of summarizing single cell data as boxplots per image, which makes it possible to find the source of bias in a data set.

Figure S2. Failure of the method *Boxplot1.5* and success of the method *Default* to accurately analyze a bimodal population that contains a large subpopulation (40% of the total population). In this file the failure of the method *Boxplot1.5* and the success of the method *Default* to accurately analyze a simulated bimodal population that contains a large subpopulation (40% of the total population) is demonstrated.

Figure S3. Accuracy of a hand-analysis method estimating small subpopulation sizes in simulated bimodal populations via mid-point determination of large subpopulation histogram peak. This file contains a graphical explanation of a hand-analysis method for subpopulation detection which uses visual determination of the mid-point of the large subpopulation peak in a histogram as a basis. A similar hand-analysis method has been proposed recently by Bates and colleagues {Bates, 2005 #45}. Further, this file contains a data-table showing the accuracy performance of the method on multiple simulated bimodal populations, and an annotated script which was used for the simulations in *R*.

Figure S4. The use of P-P plots for non-parametric and graphical response quantification. This file illustrates the concept of employing P-P plots for non-parametric and graphical response quantification {MacArthur #70}, using results obtained from the measurement of ICE*clc* activity in *P. knackmussii* B13 grown under different environmental conditions as an example data-set.

Table S1. Bacterial strains used in this work. In this file we provide a list with all bacterial strains used in this work.

Table S2. *Pseudomonas knackmussii* B13 growth in batch culture. This file contains a table listing timing of exponential growth and onset of stationary phase in batch cultures of *P. knackmussii* B13 and *P. putida* UWC (ICE*clc*) grown on different carbon substrates.

Table S3. Large and small subpopulation parameters of fluorescence data from promoter-*egfp* reporters for ICE*clc* activation in *Pseudomonas knackmussii* B13. This file contains a data table showing typical measured large and small subpopulation parameters of fluorescence data obtained from promoter-*egfp* reporters for ICE*clc* activation in *P. knackmussii* B13 after growth on 3CBA. These parameters were used as reference parameters for ICE*clc* activation when creating some of the simulated subpopulations in Figures 7, 8, 9 and 10, Table 2, and Figure S2, S3.

Table S4. Significance testing of subpopulation effects from ICE*clc* activation under different conditions quantified by different PS methods. This file contains a data table showing results from quantifications of small subpopulation effects by different PS methods. Results from this table are visualized in Figure 3A.

Table S5. Significance testing of subpopulation effects from ICE*clc* activation under different conditions quantified by different non-PS methods. This file contains a data table showing results from quantifications of small subpopulation effects by different PS methods. Results from this table are visualized in Figure 3B.

Table S6. Bias as a function of subpopulation proportion (range: 0.1-40%; n=40) at a mean difference of 67.8 and a subpopulation standard deviation of 37.7. This file contains a data table showing numerical data corresponding to Figure 8.

Table S7. Bias as function of subpopulation proportion (range: 0.1-1.2%; n=15) at a mean difference of 67.8 and a subpopulation standard deviation of 37.7. Data table corresponding to Figure 10.

Table S8. ICE*clc* activity-response in *Pseudomonas knackmussi* B13 P$_{int}$-*egfp* to pre-growth on different carbon sources, quantified over percentile range. Data correspond to Figure S4C.

Protocol S1. Description of *R* functions for quantification of low abundance phenomena in bimodal populations. This file provides a detailed description of the proposed *R* functions *findsub(...)* and *get.ci(...)* as tools for quantification of small subpopulation phenomena and method confidence interval calculation, respectively. We also show examples of graphical and command-line output from these functions.

Protocol S2. Scripts and functions for quantification of low abundance phenomena in bimodal populations. This file contains the proposed *R* scripts and functions for quantifying low abundance phenomena in bimodal populations. Comments within scripts and the README file serve as step-by-step guidance for the implementation of the relevant functions in *R*. An

example data set is included, allowing for a demonstration of the relevant functions while following the step-by-step procedure (not included in this thesis).

Protocol S3. Scripts and functions for generating simulated data. This file contains the *R* scripts and functions used for generating the simulated bimodal populations that were analyzed in this paper.

Video S1. Accuracy of subpopulation determination as quantified by the methods *Boxplot1.5* or *Boxplot3* from different simulated bimodal populations. This file contains a movie showing the results of *Boxplot1.5* and *Boxplot3* methods of subpopulation detection tested on simulated bimodal populations with varying subpopulation proportions, standard deviations and mean values (see Methods). Method accuracy is shown as the percentage bias between estimated and true subpopulation size (z-axis), and as a function of subpopulation standard deviation (x-axis) and subpopulation proportion (y-axis). The 40 different movie image frames show results for different simulated subpopulation mean values. A bias of zero indicates that estimated subpopulation size equals true subpopulation size. Negative or positive bias indicates under- or over-estimation of subpopulation size in comparison to true subpopulation size, respectively. Instances where the method fails to detect any subpopulation size are indicated as solidly coloured squares at the top surface of the co-ordinate system (also see NA annotations in Figure 7, 9).

Video S2. Accuracy of subpopulation determination as quantified by the methods *Boxplot1.5* or *Boxplot3* from different simulated bimodal populations. This file contains a movie showing the results of the *Boxplot1.5* and *Boxplot3* methods of subpopulation detection tested on simulated bimodal populations with varying subpopulation proportions, standard deviations and mean values (see Methods). Method accuracy is shown as the percentage bias between estimated and true subpopulation size (z-axis) and as a function of subpopulation mean (x-axis) and subpopulation proportion (y-axis). The 40 different movie image frames show results for different simulated subpopulation standard deviations. A bias of zero indicates that estimated subpopulation size equals true subpopulation size. Negative or positive bias indicates under- or over-estimation of subpopulation size in comparison to true subpopulation size, respectively. Instances where the method fails to detect any subpopulation size are indicated as solidly coloured squares at the top surface of the co-ordinate system (also see NA annotations in Figure 7, 9).

Video S3. Accuracy of subpopulation determination as quantified by the methods *Boxplot1.5* or *Boxplot3* from different simulated bimodal populations. This file contains a movie showing the results of the *Boxplot1.5* and *Boxplot3* methods of subpopulation detection tested

on simulated bimodal populations with varying subpopulation proportions, standard deviations and mean values (see Methods). Method accuracy is shown as the percentage bias between estimated and true subpopulation size (z-axis) and as a function of subpopulation mean (x-axis) and subpopulation standard deviation (y-axis). The 40 different movie image frames show results for different simulated subpopulation proportions. A bias of zero indicates that estimated subpopulation size equals true subpopulation size. Negative or positive bias indicates under- or over-estimation of subpopulation size in comparison to true subpopulation size, respectively. Instances where the method fails to detect any subpopulation size are indicated as solidly coloured squares at the top surface of the co-ordinate system (also see NA annotations in Figure 7, 9).

Video S4. *Boxplot1.5* method accuracy as tested on simulated bimodal populations with low subpopulation proportions (0.1-1.2%). This file contains a movie showing the results of the *Boxplot1.5* method of subpopulation detection tested on simulated bimodal populations of three different population sizes ($n=2\times10^3$, $n=2\times10^4$, and $n=2\times10^5$) with simulated small subpopulation proportions ranging between 0.1 and 1.2% of the large subpopulation (see Methods). Method accuracy is shown as the percentage bias between the estimated and the true subpopulation size (z-axis) and as a function of subpopulation standard deviation (x-axis) and subpopulation proportion (y-axis). Different movie panels show results for different total population sizes. The 15 different movie image frames show results for different simulated subpopulation mean values. A bias of zero indicates that the estimated subpopulation size equals the true subpopulation size. Negative or positive bias indicates under- or over-estimation of subpopulation size in comparison to true subpopulation size, respectively. Instances where the method fails to detect any subpopulation size are indicated as solidly coloured squares at the top surface of the co-ordinate system (also see NA annotations in Figure 7, 9).

Video S5. *Boxplot1.5* method accuracy as tested on simulated bimodal populations with low subpopulation proportions (0.1-1.2%). This file contains a movie showing the results of the *Boxplot1.5* method of subpopulation detection tested on simulated bimodal populations of three different population sizes ($n=2\times10^3$, $n=2\times10^4$, and $n=2\times10^5$) with simulated small subpopulation proportions ranging between 0.1 and 1.2% of the large subpopulation (see Methods). Method accuracy is shown as the percentage bias between the estimated and the true subpopulation size (z-axis) and as a function of subpopulation mean (x-axis) and subpopulation standard deviation (y-axis). Different movie panels show results for different total population sizes. The 15 different movie image frames show results for different simulated

subpopulation proportions. A bias of zero indicates that the estimated subpopulation size equals the true subpopulation size. Negative or positive bias indicates under- or over-estimation of subpopulation size in comparison to true subpopulation size, respectively. Instances where the method fails to detect any subpopulation size are indicated as solidly coloured squares at the top surface of the co-ordinate system (also see NA annotations in Figure 7, 9).

Video S6. *Boxplot1.5* method accuracy as tested on simulated bimodal populations with low subpopulation proportions (0.1-1.2%). This file contains a movie showing the results of the *Boxplot1.5* method of subpopulation detection tested on simulated bimodal populations of three different population sizes ($n=2\times10^3$, $n=2\times10^4$, and $n=2\times10^5$) with simulated small subpopulation proportions ranging between 0.1 and 1.2% of the large subpopulation (see Methods). Method accuracy is shown as the percentage bias between the estimated and the true subpopulation size (z-axis) and as a function of subpopulation mean (x-axis) and subpopulation proportion (y-axis). Different movie panels show results for different total population sizes. The 15 different movie image frames show results for different simulated subpopulation standard deviations. A bias of zero indicates that the estimated subpopulation size equals the true subpopulation size. Negative or positive bias indicates under- or over-estimation of subpopulation size in comparison to true subpopulation size, respectively. Instances where the method fails to detect any subpopulation size are indicated as solidly coloured squares at the top surface of the co-ordinate system (also see NA annotations in Figure 7, 9).

Video S7. *Boxplot3* method accuracy as tested on simulated bimodal populations with low subpopulation proportions (0.1-1.2%). This file contains a movie showing the results of the *Boxplot3* method of subpopulation detection tested on simulated bimodal populations of three different population sizes ($n=2\times10^3$, $n=2\times10^4$, and $n=2\times10^5$) with simulated small subpopulation proportions ranging between 0.1 and 1.2% of the large subpopulation (see Methods). Method accuracy is shown as the percentage bias between the estimated and the true subpopulation size (z-axis) and as a function of subpopulation standard deviation (x-axis) and subpopulation proportion (y-axis). Different movie panels show results for different total population sizes. The 15 different movie image frames show results for different simulated subpopulation mean values. A bias of zero indicates that the estimated subpopulation size equals the true subpopulation size. Negative or positive bias indicates under- or over-estimation of subpopulation size in comparison to true subpopulation size, respectively. Instances where the method fails to detect any subpopulation size are indicated as solidly

coloured squares at the top surface of the co-ordinate system (also see NA annotations in Figure 7, 9).

Video S8. *Boxplot3* method accuracy as tested on simulated bimodal populations with low subpopulation proportions (0.1-1.2%). This file contains a movie showing the results of the *Boxplot3* method of subpopulation detection tested on simulated bimodal populations of three different population sizes ($n=2\times10^3$, $n=2\times10^4$, and $n=2\times10^5$) with simulated small subpopulation proportions ranging between 0.1 and 1.2% of the large subpopulation (see Methods). Method accuracy is shown as the percentage bias between the estimated and the true subpopulation size (z-axis) and as a function of subpopulation mean (x-axis) and subpopulation standard deviation (y-axis). Different movie panels show results for different total population sizes. The 15 different movie image frames show results for different simulated subpopulation proportions. A bias of zero indicates that the estimated subpopulation size equals the true subpopulation size. Negative or positive bias indicates under- or over-estimation of subpopulation size in comparison to true subpopulation size, respectively. Instances where the method fails to detect any subpopulation size are indicated as solidly coloured squares at the top surface of the co-ordinate system (also see NA annotations in Figure 7, 9).

Video S9. *Boxplot3* method accuracy as tested on simulated bimodal populations with low subpopulation proportions (0.1-1.2%). This file contains a movie showing the results of the *Boxplot3* method of subpopulation detection tested on simulated bimodal populations of three different population sizes ($n=2\times10^3$, $n=2\times10^4$, and $n=2\times10^5$) with simulated small subpopulation proportions ranging between 0.1 and 1.2% of the large subpopulation (see Methods). Method accuracy is shown as the percentage bias between the estimated and the true subpopulation size (z-axis) and as a function of subpopulation mean (x-axis) and subpopulation proportion (y-axis). Different movie panels show results for different total population sizes. The 15 different movie image frames show results for different simulated subpopulation standard deviations. A bias of zero indicates that the estimated subpopulation size equals the true subpopulation size. Negative or positive bias indicates under- or over-estimation of subpopulation size in comparison to true subpopulation size, respectively. Instances where the method fails to detect any subpopulation size are indicated as solidly coloured squares at the top surface of the co-ordinate system (also see NA annotations in Figure 7, 9).

SUPPLEMENTARY FILES
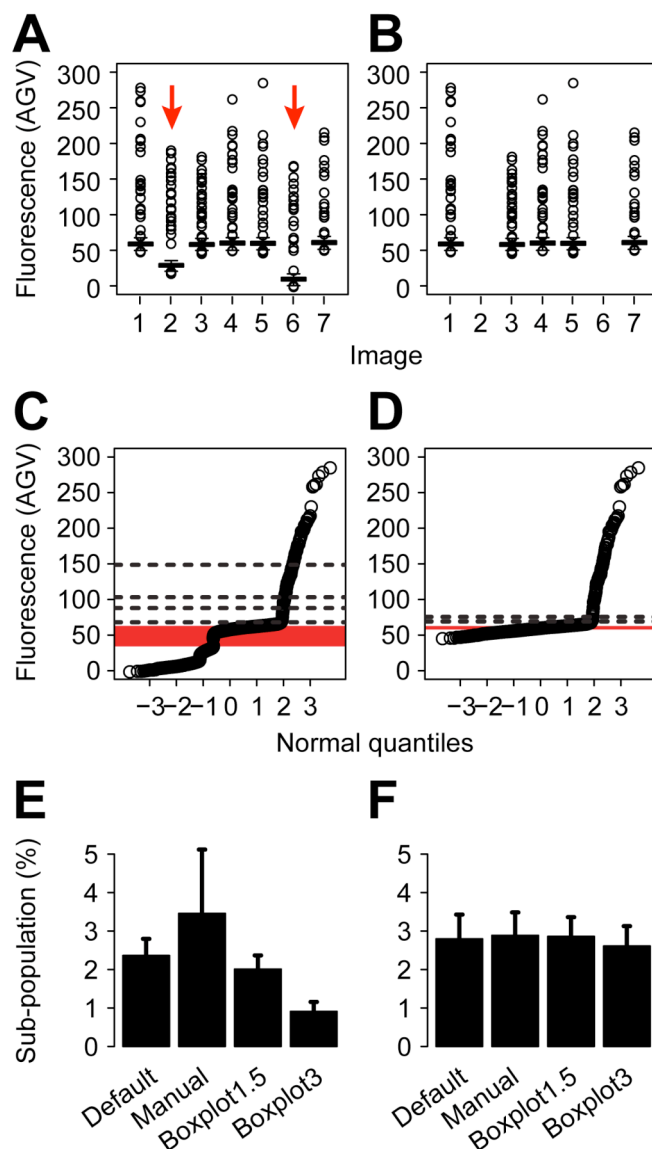
Figure S1

Figure S1 Bias compromises detection of small subpopulations in bimodal data. This file contains a series of graphs that demonstrate the obstructive role of bias in estimating subpopulation size in bimodal data. A), B) Box-plot representations of P_{int}-*egfp* expression in *P. putida* UWC1 for 7 different image sets (1-7). The left row of graphs are based on a faulty data set with data originating from two images that have much lower fluorescence values as a result of a mistake during image acquisition. The right row of graphs represents the same data set but with the data from the biased images removed. C), D) Q-Q plot representations of uncorrected and corrected data sets in A and B, respectively. E), F) Subpopulation size calculations using the uncorrected and corrected data sets from A and B, respectively. This panel of graphs highlights the practicality of summarizing single cell data as boxplots per image, which makes it possible to find the source of bias in a data set.
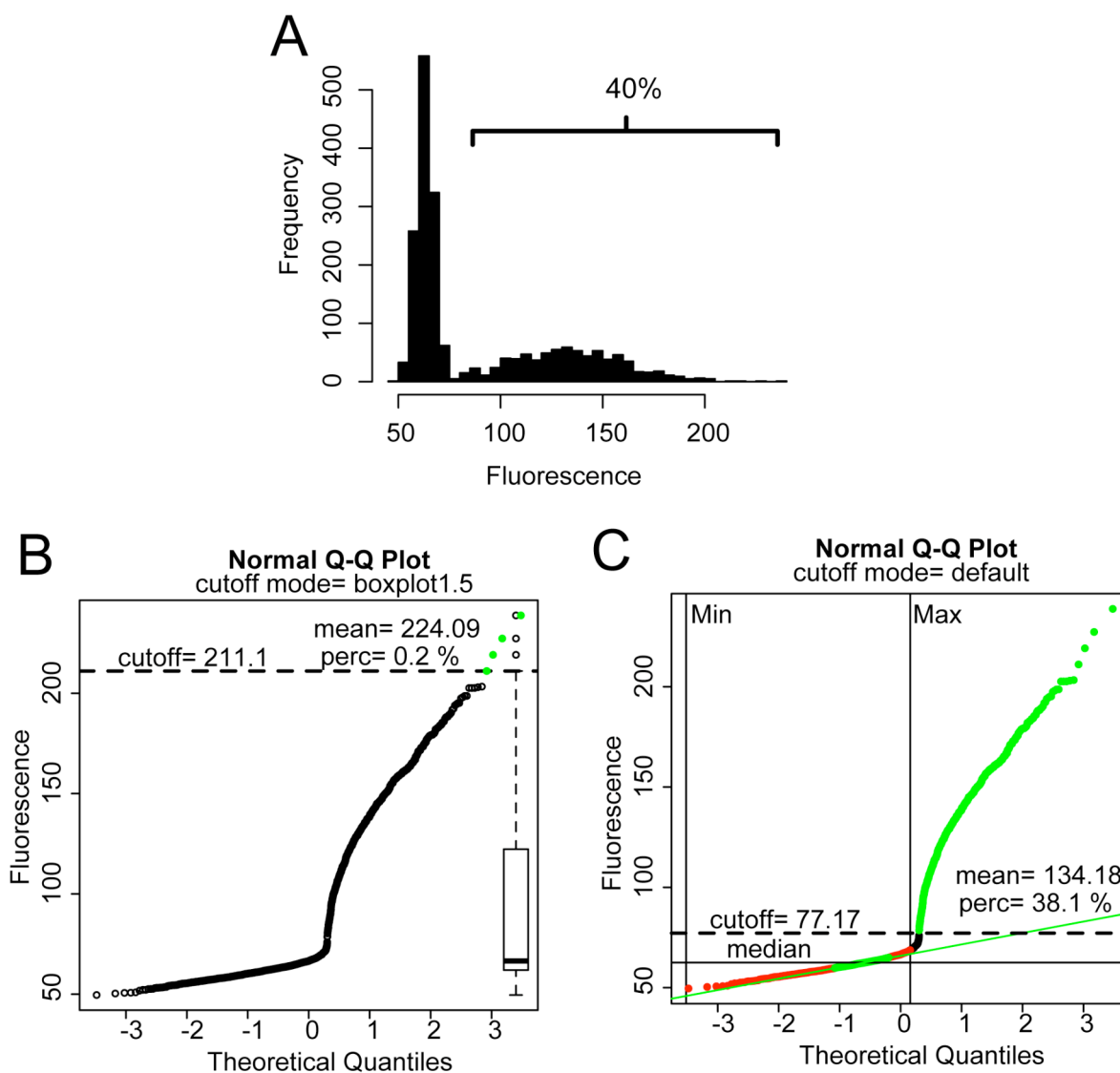
Figure S2

Figure S2. Failure of the method *Boxplot1.5* (B) and success of the method *Default* (C) to accurately analyze a simulated bimodal population (A; n=2000) containing a large subpopulation (40% of the total population). Bimodal population was created in *R* by mixing two simulated Gaussian subpopulations whose means (63.0, 127.3) and standard deviations (3.9, 37.7) represented typical population parameters as obtained from fluorescence microscopy analysis of batch grown *P.knackmussii* P_{int}-*egfp* in 3CBA (see Table S3). mean, estimated mean of subpopulation; perc, estimated percentage of subpopulation from total population; cutoff, point of separation between subpopulations; Min, estimated minimum of lower subpopulation; Max, estimated maximum of lower subpopulation; median, estimated median of lower subpopulation; Green points above cutoff, estimated subpopulation.

Figure S3
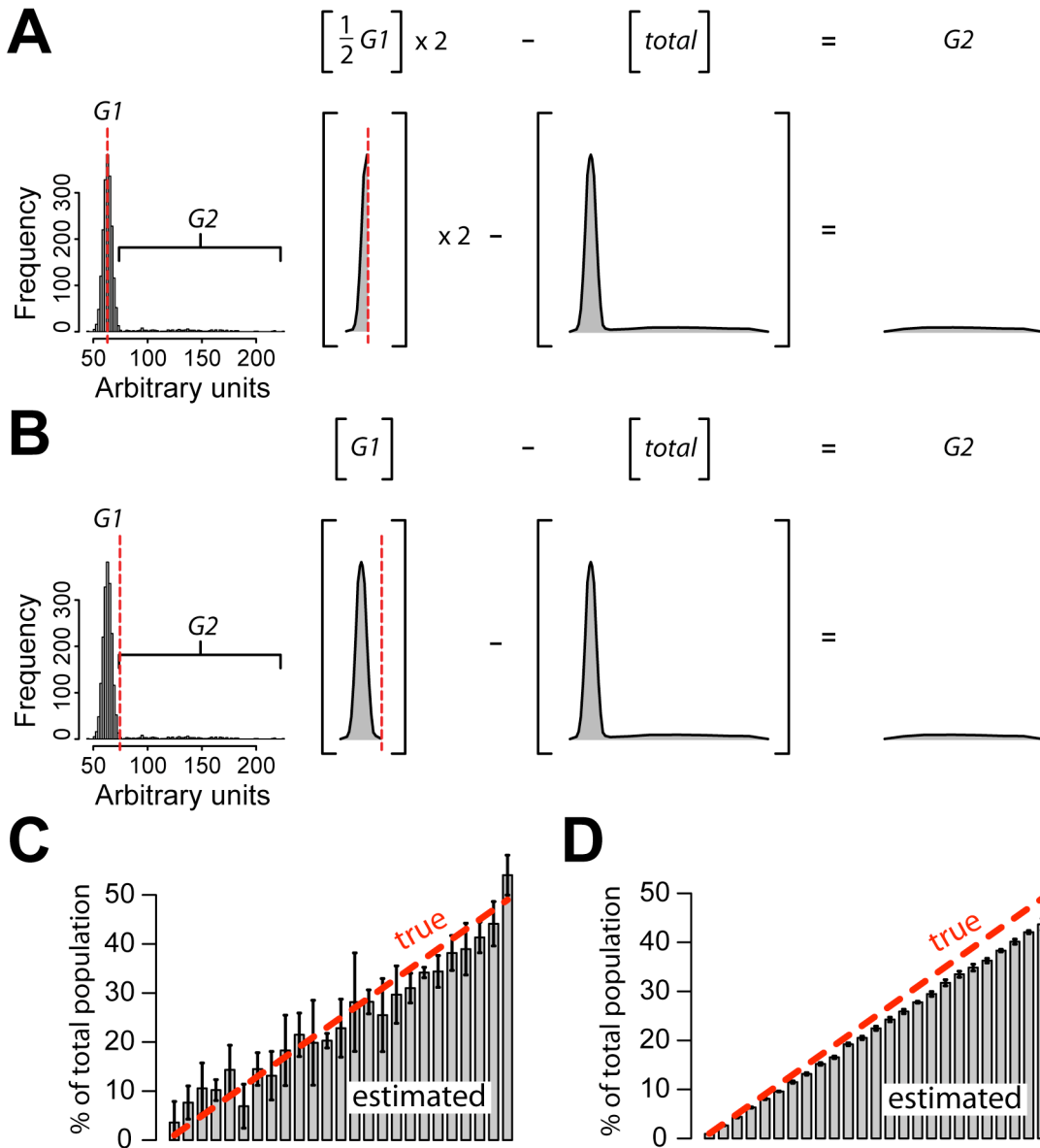
Figure S3. Estimation of the small subpopulation size (percentage of total population) in a bimodal population via visual determination of either the mid-point of the histogram peak (A, C), or alternatively the approximate border between histogram peak and histogram tail (B, D). (A) To calculate the percentage of the small subpopulation the number of observations in *G1* was subtracted from the total number of cells. The *G1* fraction was approximated by quantifying the "left" half of the peak mid-point (dotted line), and multiplying by 2. This approach is similar to the "hand-analysis method" proposed by Bates and colleagues [1] [2]. (B) Same as (A) but *G1* fraction was approximated by quantifying the "left" part of the peak-tail border (dotted line). This approach is similar to the *Manual* method as proposed in this paper (Protocol S1). Left panels in (A) and (B) show a histogram of a simulated bimodal population with standard deviations 3.9 and 37.7, mean values 63 and 127.3, and number of observations 1860 and 140, for *G1* and *G2* Gaussian fractions, respectively. Gaussian simulations were carried out by the *R* function *rnorm(...)*. (C) Simulation results of estimates of

a range of true subpopulation sizes (dotted line; 1-49%) using the method according to (A) and implemented in the *R* function *sim.midpoint(…)* (Figure S3 *R* code 1, see below). D) Same as (C) but method according to (B) implemented in the *R* function *sim.peaktail(…)* (Figure S3 *R* code 2, see below). Error bars denote SD (calculated from 5 independent repetitions).

Figure S3 Table. Accuracy of histogram-based manual method of subpopulations determination.

True subpopulation size (%)[2]	Peak mid-point determination[1]		Peak-tail border determination[1]	
	Estimated subpopulation size (%)[3]	Bias[4]	Estimated subpopulation size (%)[3]	Bias[4]
1.0	3.6 ± 4.3	260.0 ± 428.4	0.9 ± 0.0	-7.0 ± 4.5
3.0	7.7 ± 3.4	155.3 ± 112.8	2.7 ± 0.1	-11.3 ± 3.2
5.0	10.6 ± 5.2	111.2 ± 103.4	4.4 ± 0.2	-11.6 ± 4.6
7.0	10.2 ± 2.1	45.7 ± 30.4	6.3 ± 0.1	-9.7 ± 1.6
9.0	14.3 ± 5.0	59.3 ± 55.7	8.1 ± 0.2	-10.1 ± 1.7
11.0	6.9 ± 4.5	-37.1 ± 40.7	9.6 ± 0.1	-12.7 ± 0.6
13.0	14.5 ± 3.3	11.5 ± 25.6	11.5 ± 0.2	-11.5 ± 1.8
15.0	13.1 ± 5.0	-12.4 ± 33.0	13.2 ± 0.2	-12.2 ± 1.7
17.0	18.3 ± 7.2	7.7 ± 42.2	15.2 ± 0.3	-10.6 ± 1.6
19.0	21.5 ± 4.4	13.1 ± 23.3	16.6 ± 0.3	-12.9 ± 1.4
21.0	19.9 ± 8.7	-5.4 ± 41.2	19.2 ± 0.3	-8.5 ± 1.3
23.0	20.3 ± 1.5	-11.8 ± 6.5	20.5 ± 0.4	-10.8 ± 1.5
25.0	22.8 ± 5.9	-8.6 ± 23.6	22.5 ± 0.4	-10.1 ± 1.7
27.0	28.1 ± 10.0	4.2 ± 37.2	24.3 ± 0.5	-10.2 ± 1.7
29.0	28.2 ± 2.4	-2.8 ± 8.4	25.9 ± 0.4	-10.7 ± 1.6
31.0	25.5 ± 7.5	-17.7 ± 24.0	27.8 ± 0.2	-10.4 ± 0.7
33.0	29.7 ± 5.8	-10.1 ± 17.7	29.5 ± 0.5	-10.7 ± 1.5
35.0	31.0 ± 3.0	-11.4 ± 8.6	31.8 ± 0.6	-9.3 ± 1.8
37.0	34.2 ± 1.1	-7.6 ± 2.9	33.5 ± 0.6	-9.5 ± 1.7
39.0	34.4 ± 3.2	-11.8 ± 8.3	34.9 ± 0.7	-10.6 ± 1.8
41.0	38.2 ± 3.5	-6.9 ± 8.7	36.3 ± 0.5	-11.5 ± 1.1
43.0	39.0 ± 5.3	-9.4 ± 12.3	38.3 ± 0.2	-10.9 ± 0.6
45.0	41.3 ± 3.1	-8.1 ± 7.0	40.1 ± 0.5	-10.8 ± 1.2
47.0	44.1 ± 4.5	-6.1 ± 9.6	42.1 ± 0.3	-10.5 ± 0.6
49.0	54.0 ± 4.1	10.3 ± 8.3	43.7 ± 1.2	-10.9 ± 2.5

1) Methods of subpopulation detection (see Figure S3 Figure 1 A and B).

2) True subpopulations were simulated (see Figure S3 *R* code) using the *R* function *rnorm(…)* with a standard deviation of 37.7, a mean value of 127.3, and the number of observations corresponding to the subpopulation percentage to be tested from a total number of 2000 observations. Mean and standard deviation used were obtained from fluorescence microscopy analysis of batch grown *P. knackmussii* P$_m$-*egfp* in 3CBA (see Table S3).

3) Estimated subpopulations (mean ± SD; 5 independent repetitions) also shown in Figure S3 1C, D. Were determined by quantifying the larger subpopulation in a bimodal population and subtracting it from the total population. This was either done by peak mid-point determination or by peak-tail border determination (see Figure S3 Figure 1A,B for details) using the *R* function *sim.midpoint(…)* and *sim.peaktail(…)*, repectively (see Figure S3 *R* code 1, 2). The bimodal population was simulated by mixing two simulated populations; a true subpopulation and a second subpopulation. The second subpopulation was created using the *R* function *rnorm(…)* with a standard deviation of 3.9, a mean value of 63.0, and the number of observations depending on the sample size of true subpopulation to give a total of 2000 observations. Mean and standard deviation used were obtained from fluorescence microscopy analysis of batch grown *P. knackmussii* P$_m$-*egfp* in 3CBA (see Table S3).

4) Bias was calculated according to: (estimated subpopulation size – true subpopulation size)/true subpopulation size x 100.

Figure S3 *R* **code 1**. *R* code defining the function *sim.midpoint(…)*, which simulates bimodal populations and estimates subpopulation size according to histogram peak mid-point determination (see Figure S3 Figure 1A).

```
#vector of subpopulation sizes to be tested
my.true.fractions<-rep(seq(1,50,2),each=5)/100


#definition of function
sim.midpoint<- function(total.N = 2000, true.fractions = my.true.fractions) {
        true.size<-true.fractions*total.N
        sim.list<-list()
        for (i in 1:length(true.fractions)) {


                a<-rnorm(total.N-total.N*true.fractions[i],63,3.9)#subpopulation1
                b<-rnorm(total.N*true.fractions[i],127.3,37.7)#subpopulation2
                my.sim<-c(a,b)#mixed bimodal population

                sim.list[[i]] <- my.sim
                }


        my.results<-numeric()
        for (i in 1:length(sim.list)) {
                sim.index<-sim.list[[i]]
                hist(sim.list[[i]],breaks=100, col="grey", main=paste(i,"of",length(sim.list)))
                my.estimate <-locator(n=1)#mouse-click determination of histogram peak mid-point
                dev.off()
                my.results[i]<-length(sim.index)-length(sim.index[sim.index<my.estimate$x])*2
                }
        midpoint.data<<-cbind(my.results, true.size)/20
        midpoint.data
}
sim.midpoint()#execution of function
```

Figure S3 *R* **code 2**. *R code defining the function* sim.peaktail(...), which simulates bimodal populations and estimates subpopulation size according to histogram peak-tail border determination (see Figure S3 Figure 1A).

```
#vector of subpopulation sizes to be tested
my.true.fractions<-rep(seq(1,50,2),each=5)/100
#definition of function
sim.peaktail<- function(total.N = 2000, true.fractions = my.true.fractions ) {
        true.size<-true.fractions*total.N
        sim.list<-list()
        for (i in 1:length(true.fractions)) {

                a<-rnorm(total.N-total.N*true.fractions[i],63,3.9)#subpopulation1
                b<-rnorm(total.N*true.fractions[i],127.3,37.7)#subpopulation2
                my.sim<-c(a,b)#mixed bimodal population

                sim.list[[i]] <- my.sim
                }

        my.results<-numeric()
        for (i in 1:length(sim.list)) {
                sim.index<-sim.list[[i]]
                hist(x=sim.index,breaks=100, col="grey", main=paste(i,"of",length(sim.list)))
                my.estimate <-locator(n=1)#mouse-click determination of histogram peak-tail border
                dev.off()
                my.results[i]<-length(sim.index)-length(sim.index[sim.index<my.estimate$x])
                }
                peaktail.data<<-cbind(my.results, true.size)/20
                peaktail.data
}
sim.peaktail()#execution of function
```

Figure S3 References

1.   Bates D, Epstein J, Boye E, Fahrner K, Berg H, Kleckner N: **The** *Escherichia coli* **baby cell column: a novel cell synchronization method provides new insight into the bacterial cell cycle**. *Mol Microbiol* 2005, **57**(2):380-391.

2.   Bates D, Kleckner N: **Chromosome and Replisome Dynamics in** *E. coli*: **Loss of Sister Cohesion Triggers Global Chromosome Movement and Mediates Chromosome Segregation**. *Cell* 2005, **121**(6):899-911.

Figure S4
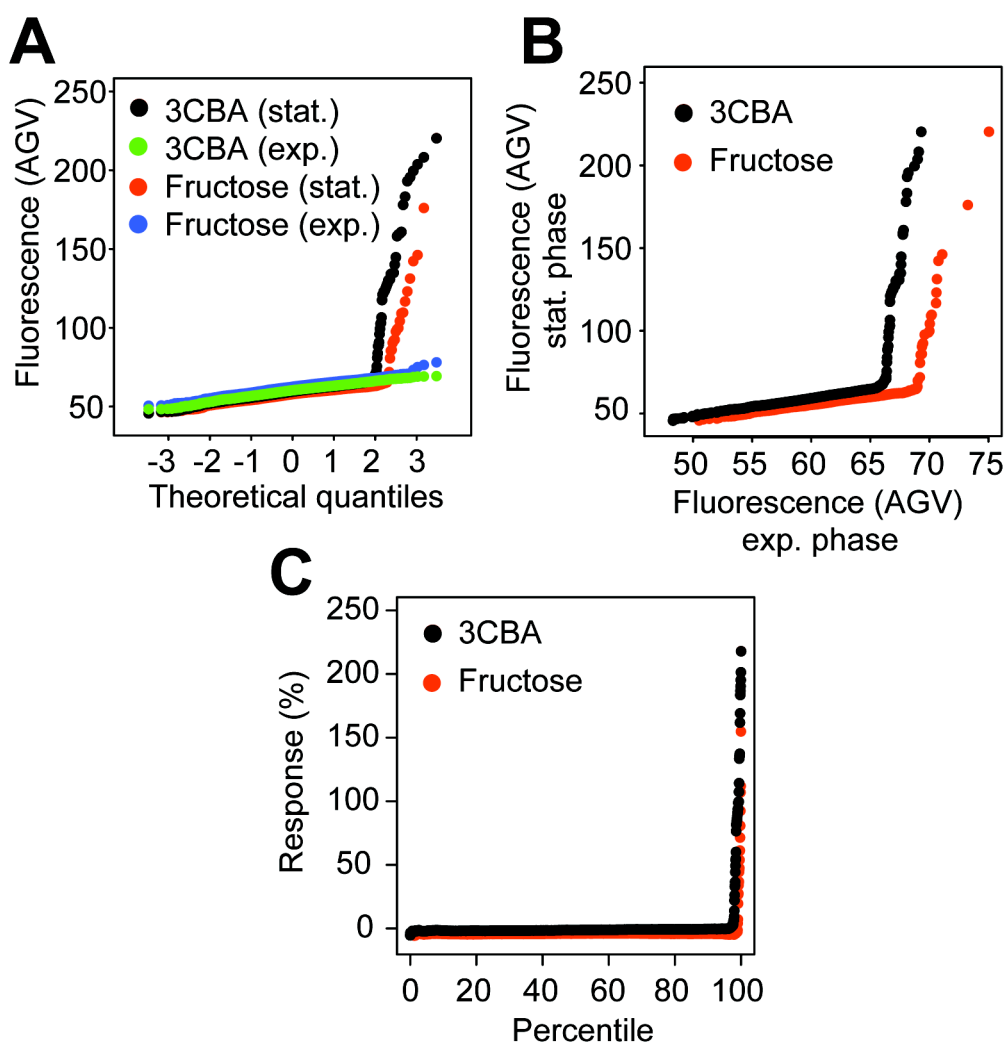
Figure S4. (A) Q-Q plot representation of single cell eGFP fluorescence from P$_{int}$-*egfp* in *P. knackmussii* B13 corresponding to ICE*clc* activation in response to different growth conditions. Black, stationary (stat.) phase after growth on 3CBA (5 mM); red, stat. phase after growth on fructose (5 mM); green, exponential (exp.) phase with growth on 3CBA (5 mM); blue, exp. phase with growth on fructose (5 mM). (B) P-P plot representation of data shown in (A) plotting ranked values of stat. phase fluorescence against ranked values of exp. phase fluorescence. (C) ICE*clc* activity-response in *Pseudomonas knackmussi* B13 to pre-growth on fructose or 3CBA, quantified over the full percentile range (also see Table S8). A similar quantification method was proposed by McArthur and collegues [1]. For each carbon source, response was calculated from the data sets shown in (A), and according to $100 \times \left[\left(P_k^{STAT} - P_k^{EXP} \right) \middle/ P_k^{EXP} \right]$, where P_k^{STAT} is the *k*th percentile at stat. phase and P_k^{Exp} is the *k*th percentile at exp. phase. Thus, the absence of any subpopulation activity in the exp. populations provide a negative control population, by which subpopulation activity in the stat. fractions may be assessed. Black, response to pre-growth on 3CBA (5mM). Red, response to pre-growth on fructose (5mM). In this

figure, a positive response corresponds to cell fluorescence stonger in stationary phase than in exponential phase, while a negative response corresponds to vice versa.

Figure S4 References

1.   MacArthur B, D., Tare R, S., Please CP, Prescott P, Oreffo R, O., C.: **A non-invasive method for** *in situ* **quantification of subpopulation behaviour in mixed cell culture**. *J R Soc Interface* 2006, **3**(6):63-69.

Table S1

Table S1. Strains used in this study.

Strain	Description	Reference
P. knackmussii B13-1343	P$_{in}$-*egfp*	[1]
P. knackmussii B13-2396		
P. knackmussii B13-2398	P$_{in}$-*egfp*	[2]
P. knackmussii B13-2399		
P. putida UWC-3408	ICE*clc*, P$_{in}$-*egfp*	[3]

Table S1 References

1.   Sentchilo V, Ravatn R, Werlen C, Zehnder AJ, van der Meer JR: **Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. Strain B13**. *Journal of Bacteriology* 2003, **185**(15):4530-4538.

2.   Minoia M, Gaillard M, Reinhard F, Stojanov M, Sentchilo V, van der Meer JR: **Stochasticity and bistability in horizontal transfer control of a genomic island in *Pseudomonas***. *Proc Natl Acad Sci U S A* 2008, **105**(52):20792-20797.

3.   Reinhard F, Miyazaki R, Pradervand N, van der Meer JR: **Cell differentiation to "mating bodies" induced by an integrating and conjugative element in free-living bacteria**. *Curr Biol* 2013, **23**(3):255-259.

<u>Table S2</u>

Table S2. *Pseudomonas knackmussi* B13 growth as determined via culture turbidity at 600 nm.

Strain	Carbon substrate	Exponential growth (h)[1]	Onset of stationary phase (h)[1]
P. knackmussi B13-1343	3-chlorobenzoate (5 mM)	8 - 20	24
	glucose (10 mM)	2 - 8	12
	fructose (10 mM)	20 - 40	45
	benzoate (10 mM)	4 - 10	12
	4-hydroxybenzoate (10 mM)	1 - 15	18
	anthranilate (5 mM)	40 - 48	50
P. putida UWC (ICE*clc*) - 2508	3-chlorobenzoate (5 mM)	30 - 50	60

1) Time after inocculation.

Table S3

Table S3. Reference subpopulation parameters calculated from single cell fluorescence values obtained from P_{inR}-*egfp* reporter in *P. knackmussii* B13.

Strain (promoter-*egfp* reporter)	Large subpopulation[1]		Small subpopulation[1]		Mean difference (ratio)[2,3]	Proportion (%)[4]
	SD[2]	Mean[2]	SD[2]	Mean[2]		
P. knackmussii B13-1343 (P_-*egfp*)	4.7	70.0	27.1	109.5	39.5 (1.6)	4.2
P. knackmussii B13-2396 (P_-*egfp*)	3.2	60.3	44.2	142.2	81.9 (2.4)	3.0
P. knackmussii B13-2397 (P_-*egfp*)	3.0	59.4	34.1	125.1	65.7 (2.1)	2.0
P. knackmussii B13-2398 (P_-*egfp*)	5.2	66.1	38.8	121.5	55.4 (1.8)	4.0
P. knackmussii B13-2399 (P_-*egfp*)	3.3	59.3	44.3	138.3	79.0 (2.3)	3.3
Mean population parameter·	**3.9**	**63.0**	**37.7**	**127.3**	**64.3 (2.0)**	**3.3**

1) Large and small subpopulation were separated according to the *Boxplot3* method. Fluorescence values of at least 2000 cells in stationary phase were investigated per strain after 25 hours of batch growth on 3CBA (5mM).

2) eGFP fluorescence (arbitrary units)

3) Difference between the means of large and small subpopulation (arbitrary units).

4) Proportion of small subpopulation as determined by the *Boxplot1.5* method.

5) Mean estimated population parameters as calculated from the four P_-*egfp* strains only. These were used as reference parameters for the creation of the bimodal simulations in this paper.

Table S4

Table S4. Proportions of subpopulations as calculated via four different PS methods from single cell eGFP fluorescence data.

Category[1]	Carbon source	Method[2]	Subpopulation size (%)[3]				Mean (± SD)	Welch Two Sample t-test[5]	Wilcoxon rank sum test[5]
			P_{inR}-egfp 23964	P_{inR^-}egfp 23984	P_{inR^-}egfp 23994	P_{int^-}egfp 13434			
A	3CBA	*Default*	2.7	5.1	2.8	4.7	3.8 ± 1.3	B*, C*	B*, C*
		Manual	2.8	3.4	3.3	4.6	3.5 ± 0.8	B**, C**	B*, C*
		Boxplot1.5	2.8	4.3	2.8	4.3	3.6 ± 0.9	B*, C**	B*, C**
		Boxplot3	2.6	2.7	2.5	3.2	2.8 ± 0.3	B**, C**	B*, C*
B	Fructose	*Default*	1.3	0.9	1.2	2.4	1.5 ± 0.7	A*	A*
		Manual	1.2	0.9	1.2	2.1	1.4 ± 0.5	A**, C*	A*, C*
		Boxplot1.5	1.2	0.9	1.2	2.4	1.4 ± 0.7	A*	A*
		Boxplot3	1.1	0.9	1.1	1.8	1.2 ± 0.4	A**, C*	A**, C*
C	Glucose	*Default*	0.6	0.6	0.9	2.0	1.0 ± 0.7	A*	A*
		Manual	0.4	0.1	0.3	0.6	0.4 ± 0.2	A**, B*	A*, B*
		Boxplot1.5	0.5	0.5	0.8	1.6	0.9 ± 0.5	A**	A**
		Boxplot3	0.4	0.1	0.3	0.6	0.4 ± 0.2	A**, B*	A*, B*

Data was obtained from P_{int}-egfp or P_{inR}-egfp expression in *P. knackmussii* B13 batch-grown to stationary phase with either 3CBA, fructose or glucose as carbon source (also see Figure 4 A).

1) Categories for significance testing

2) PS method used to determine subpopulation.

3) Subpopulation size (percent of total population)

4) Reporter-*egfp* construct and *P. knackmussii* B13 strain number.

5) Significance testing comparing the means between different categories of the same methods. * and ** indicate significant differences at P<0.05 and P<0.01, respectively.

Table S5

Table S5 Proportions of subpopulations as calculated via four different non-PS methods from single cell eGFP fluorescence data.

Category 1	Carbon source	Method[2]	eGFP fluorescence				Mean (± SD)	Welch Two Sample t-test[4]	Wilcoxon rank sum test[4]
			P_{inR}-*egfp* 23963	P_{inR}-*egfp* 23983	P_{inR}-*egfp* 23993	P_{int}-*egfp* 13433			
A	3CBA	*Mean*	62.5	68.8	61.4	72.1	66.2 ± 5.1		
		95 Percentile	66.2	79.7	65.0	81.7	73.1 ± 8.7		
		Mean top 5%	107.5	120.6	106.5	111.6	111.6 ± 6.4	B**, C**	B*, C*
		Boosted Mean	64.1	73.4	63.0	76.2	69.2 ± 6.6		
B	Fructose	*Mean*	60.7	58.4	58.1	64.4	66.2 ± 5.1		
		95 Percentile	64.5	62.2	62.2	70.9	65.0 ± 4.1		
		Mean top 5%	76.6	72.0	75.3	84.6	77.1 ± 5.4	A**	A*
		Boosted Mean	63.2	61.0	60.7	67.9	63.2 ± 3.3		
C	Glucose	*Mean*	62.2	59.7	61.0	62.8	61.4 ± 1.4		
		95 Percentile	68.0	66.1	67.1	69.6	67.7 ± 1.5		
		Mean top 5%	74.5	69.3	72.2	76.5	73.1 ± 3.1	A**	A*
		Boosted Mean	66.2	63.7	64.8	66.8	65.4 ± 1.4		

Data was obtained from P_{int}-*egfp* or P_{inR}-*egfp* expression in *P. knackmussii* B13 batch-grown to stationary phase with either 3CBA, fructose or glucose as carbon source (also see Figure 4 B).

1) Categories for significance testing

2) PS method used to determine subpopulation.

3) Reporter-*egfp* construct and *P. knackmussii* B13 strain number.

4) Significance testing comparing the means between different categories of the same methods. * and ** indicate significant differences at P<0.05 and P<0.01, respectively.

Table S6

Table S6A. Simulation results corresponding to Figure 8 with *Boxplot1.5* as method of subpopulation determination.

Figure 8A		Figure 8B		Figure 8C	
Proportion	Bias	Mean.diff	Bias	SD	Bias
0.10	200.00	2.00	-58.73	10.00	3.17
1.10	13.64	5.46	-57.14	11.03	3.17
2.10	4.76	8.92	-46.03	12.05	3.17
3.15	-6.35	12.38	-42.86	13.08	3.17
4.15	-3.61	15.85	-41.27	14.10	3.17
5.20	-6.73	19.31	-41.27	15.13	3.17
6.20	-5.65	22.77	-25.40	16.15	3.17
7.25	-7.59	26.23	-23.81	17.18	3.17
8.25	-7.88	29.69	-26.98	18.21	3.17
9.30	-9.14	33.15	-23.81	19.23	3.17
10.30	-8.25	36.62	-22.22	20.26	3.17
11.35	-6.17	40.08	-19.05	21.28	3.17
12.35	-7.29	43.54	-15.87	22.31	3.17
13.35	-6.74	47.00	-9.52	23.33	1.59
14.40	-8.68	50.46	-6.35	24.36	3.17
15.40	-8.12	53.92	-7.94	25.38	1.59
16.45	-8.81	57.38	-1.59	26.41	1.59
17.45	-7.16	60.85	-4.76	27.44	0.00
18.50	-8.38	64.31	-3.17	28.46	1.59
19.50	-10.51	67.77	-6.35	29.49	1.59
20.55	-6.33	71.23	-3.17	30.51	0.00
21.55	-9.98	74.69	1.59	31.54	0.00
22.60	-8.41	78.15	3.17	32.56	1.59
23.60	-10.17	81.62	1.59	33.59	-4.76
24.65	-9.13	85.08	3.17	34.62	-7.94
25.65	-14.42	88.54	1.59	35.64	1.59
26.70	-17.79	92.00	1.59	36.67	-1.59
27.70	-40.79	95.46	1.59	37.69	-6.35
28.70	-43.73	98.92	3.17	38.72	-1.59
29.75	-65.21	102.38	1.59	39.74	-3.17
30.75	-64.88	105.85	3.17	40.77	-1.59
31.80	-78.14	109.31	3.17	41.79	-7.94
32.80	-81.71	112.77	3.17	42.82	-3.17
33.85	-88.77	116.23	3.17	43.85	-6.35
34.85	-90.82	119.69	3.17	44.87	-3.17

137

Table S6A *Continued*

Figure 8A		Figure 8B		Figure 8C	
Proportion	Bias	Mean.diff	Bias	SD	Bias
35.90	-91.50	123.15	3.17	45.90	-11.11
36.90	-95.93	126.62	3.17	46.92	-6.35
37.95	-97.63	130.08	3.17	47.95	-11.11
38.95	-96.66	133.54	3.17	48.97	-3.17
40.00	-96.75	137.00	3.17	50.00	-3.17

Table S6B. Simulation results corresponding to Figure 8 with *Boxplot3* as method of subpopulation determination.

Figure 8A		Figure 8B		Figure 8C	
Proportion	Bias	Mean.diff	Bias	SD	Bias
0.10	0.00	2.00	-68.25	10.00	0.00
1.10	0.00	5.46	-55.56	11.03	0.00
2.10	-7.14	8.92	-55.56	12.05	0.00
3.15	-17.46	12.38	-61.90	13.08	0.00
4.15	-6.02	15.85	-63.49	14.10	0.00
5.20	-7.69	19.31	-42.86	15.13	0.00
6.20	-8.06	22.77	-52.38	16.15	0.00
7.25	-11.03	26.23	-36.51	17.18	0.00
8.25	-11.52	29.69	-44.44	18.21	0.00
9.30	-11.29	33.15	-28.57	19.23	0.00
10.30	-7.77	36.62	-31.75	20.26	0.00
11.35	-13.66	40.08	-26.98	21.28	-1.59
12.35	-11.74	43.54	-30.16	22.31	-1.59
13.35	-13.11	47.00	-22.22	23.33	0.00
14.40	-13.89	50.46	-19.05	24.36	-1.59
15.40	-10.39	53.92	-14.29	25.38	-1.59
16.45	-10.94	57.38	-19.05	26.41	-4.76
17.45	-10.60	60.85	-14.29	27.44	-6.35
18.50	-13.24	64.31	-12.70	28.46	-4.76
19.50	-13.85	67.77	-17.46	29.49	-4.76
20.55	-13.63	71.23	-9.52	30.51	-7.94
21.55	-15.08	74.69	-9.52	31.54	0.00
22.60	-15.93	78.15	-6.35	32.56	-6.35
23.60	-21.19	81.62	-4.76	33.59	-6.35
24.65	-19.07	85.08	-3.17	34.62	-6.35
25.65	-19.49	88.54	-3.17	35.64	-14.29
26.70	-32.02	92.00	-1.59	36.67	-12.70
27.70	-50.00	95.46	-6.35	37.69	-17.46
28.70	-90.77	98.92	0.00	38.72	-11.11
29.75	-86.72	102.38	0.00	39.74	-11.11
30.75	-97.72	105.85	-1.59	40.77	-7.94
31.80	-97.48	109.31	0.00	41.79	-15.87
32.80	-98.63	112.77	-1.59	42.82	-7.94
33.85	-99.70	116.23	-3.17	43.85	-15.87
34.85	NA	119.69	0.00	44.87	-19.05
35.90	NA	123.15	0.00	45.90	-12.70
36.90	-99.86	126.62	0.00	46.92	-9.52

Table S6B *Continued*

Figure 8A		Figure 8B		Figure 8C	
Proportion	Bias	Mean.diff	Bias	SD	Bias
37.95	NA	130.08	0.00	47.95	-15.87
38.95	NA	133.54	0.00	48.97	-15.87
40.00	NA	137.00	0.00	50.00	-22.22

Table S7

Table S7. Simulation results corresponding to Figure 10.

method	Figure 10A (n=2000)		Figure 10B (n=20000)		Figure 10C (n=200000)	
	Proportion	Bias	Proportion	Bias	Proportion	Bias
Boxplot1.5	0.10	100.00	0.10	390.00	0.10	351.50
	0.15	66.67	0.18	225.71	0.18	189.92
	0.25	20.00	0.26	145.10	0.26	130.16
	0.30	16.67	0.34	113.43	0.34	96.72
	0.40	12.50	0.41	93.90	0.41	76.81
	0.45	11.11	0.49	65.31	0.49	63.25
	0.55	9.09	0.57	60.53	0.57	52.71
	0.60	8.33	0.64	52.71	0.65	43.80
	0.70	-7.14	0.73	45.52	0.73	38.16
	0.80	0.00	0.80	40.99	0.81	34.57
	0.85	5.88	0.88	37.29	0.89	30.55
	0.95	0.00	0.96	32.81	0.96	25.57
	1.00	0.00	1.04	32.69	1.04	24.03
	1.10	0.00	1.12	24.55	1.12	21.10
	1.20	-8.33	1.20	27.08	1.20	17.92
Boxplot3	0.10	0.00	0.10	-10.00	0.10	-8.50
	0.15	0.00	0.18	-11.43	0.18	-9.80
	0.25	0.00	0.26	-5.88	0.26	-9.34
	0.30	0.00	0.34	-8.96	0.34	-9.84
	0.40	-12.50	0.41	-6.10	0.41	10.39
	0.45	0.00	0.49	-9.18	0.49	-9.64
	0.55	-9.09	0.57	-9.65	0.57	-9.54
	0.60	0.00	0.64	-12.40	0.65	-9.93
	0.70	-7.14	0.73	-11.03	0.73	-9.88
	0.80	-6.25	0.80	-8.70	0.81	-9.54
	0.85	-5.88	0.88	-6.78	0.89	-8.36
	0.95	-21.05	0.96	-10.42	0.96	-9.18
	10.00	-10.00	1.04	-8.17	1.04	-9.78
	1.10	-4.55	1.12	-8.48	1.12	-9.10
	1.20	-8.33	1.20	-7.08	1.20	-9.62

Table S8

Table S8. ICE*clc* activity-response in *Pseudomonas knackmussi* B13 P_{int}-*egfp* to pre-growth on different carbon sources, quantified over percentile range. Data correspond to Figure S4C.

Percentile[1]	Response (%)[2]	
	Fructose	3CBA
97.80%	-4.74	6.47
97.85%	-4.83	6.89
97.90%	-4.70	9.29
97.95%	-4.75	13.97
98.00%	-4.59	**22.02**
98.05%	-4.51	**26.27**
98.10%	-4.35	**32.97**
98.15%	-4.19	**33.72**
98.20%	-3.97	**36.64**
98.25%	-3.97	**44.32**
98.30%	-3.98	**49.31**
98.35%	-4.05	**53.88**
98.40%	-3.91	**54.74**
98.45%	-3.73	**60.10**
98.50%	-3.63	**76.48**
98.55%	-3.20	**81.58**
98.60%	-3.17	**82.25**
98.65%	-3.05	**84.14**
98.70%	-3.02	**84.27**
98.75%	-2.73	**85.43**
98.80%	-2.88	**86.73**
98.85%	-2.57	**88.33**
98.90%	-2.59	**89.67**
98.95%	-1.32	**91.11**
99.00%	4.07	**93.83**
99.05%	7.13	**94.03**
99.10%	19.60	**94.01**
99.15%	**26.68**	**98.82**
99.20%	**27.52**	**99.00**
99.25%	**34.04**	**99.61**
99.30%	**35.90**	**99.68**
99.35%	**36.79**	107.35
99.40%	**44.43**	114.22
99.45%	**46.14**	133.47

99.50%	47.41	135.01

Table S8 *Continued*

Percentile[1]	Response (%)[2]	
	Fructose	3CBA
99.55%	53.87	135.87
99.60%	60.95	137.32
99.65%	61.26	161.80
99.70%	71.40	169.08
99.75%	80.76	183.48
99.80%	92.43	186.68
99.85%	107.00	190.35
99.90%	111.90	195.29
99.95%	154.82	201.38
100.00%	217.95	217.84

1) Percentile range from 97.80-100%. Bold, first percentile value corresponding to response > 20%.

2) Positive values denote cell fluorescence brighter in stationary phase than in exponential phase, while a negative response corresponds to vice versa. Bold, response > 20%.

Protocol S1

Description of the *R* functions *findsub(…), get.ci(…),* and *get.ci.other(…)*

Different strategies of small subpopulation quantification and robustness testing were translated into the *R* [1] functions *findsub(...)* and *get.ci(...)*, respectively (for code see Protocol S2). In this way subpopulations could be computationally calculated following a simple protocol of commands (Protocol S1, Figure 1: *(5)-(8)*). The functions provide numerical as well as graphical output (Protocol S1, Figure 2-8). Functions used in *Default* or *Manual* mode were designed to work on basis of interactive graphs requiring mouse-clicking at the approximate locations of the large subpopulation minimum and maximum, or at the point of separation between large and small subpopulation, respectively (Protocol S1, Figure 2, 3). In contrast, functions in *Boxplot1.5* and *Boxplot3* modes were designed to run fully automatically once initiated (Protocol S1, Figure 4, 5). An additional mode *Other* was designed to employ other methods of characterization corresponding to the mean, 95th percentile, boosted mean, and the mean top five % of the total population, respectively (Protocol S1, Figure 6). Bootstrap confidence interval calculations may be performed by the functions *get.ci(...)* (Protocol S1, Figure 7, 8) and *get.ci.other(...)*. While the former calculates the confidence intervals of the methods in modes *Default*, *Manual*, *Boxplot1.5*, and *Boxplot3*, the latter calculates confidence intervals of *Other* modes. *get.ci(...)* and *get.ci.other(...)* both employ the *R* funtions *boot(...)* and *boot.ci(...)* included in the *R* package *boot* [2] [3], which perform re-sampling of original data set (with replacement) and subsequent method application. Shown below, are examples of *R* output of the functions *findsub(...)* and *get.ci(...)* as tested on an example data set (Protocol S2). The example data set was compiled from single cell eGFP fluorescence values of *Pseudomonas knackmussi* B13 strain 1343 grown in batch culture on minimal medium with 3-chlorobenzaote as carbon source (10 mM), sampled 10 h after reaching stationary phase.

```
(1)     rm(list=ls(all=TRUE))
(2)     setwd("/Users/My_Folder_Address")

(3)     example.data <- read.table("example.data.txt", header=TRUE)
(4)     source("FINDSUB_code.txt")

(5)     findsub(x=example.data$value, mode="default")
(6)     findsub(x=example.data$value, mode="manual")
(7)     findsub(x=example.data$value, mode="boxplot1.5")
(8)     findsub(x=example.data$value, mode="boxplot3")
(9)     findsub(x=example.data$value, mode="other")

(10)    get.ci(x=example.data$value, trials=20, mode="default")
(11)    get.ci(x=example.data$value, trials=20, mode="manual")
(12)    get.ci(x=example.data$value, R=500, mode="boxplot1.5")
(13)    get.ci(x=example.data$value, R=500, mode="boxplot3")

(14)    get.ci.other(x=example.data$value, R=500, mode="mean")
(15)    get.ci.other(x=example.data$value, R=500, mode="perc.95")
(16)    get.ci.other(x=example.data$value, R=500, mode="top.5")
(17)    get.ci.other(x=example.data$value, R=500, mode="boost")
```

Protocol S1 Figure 1. *R* functions. *(1)* Clears *R* working environment. *(2)* Sets the directory of *R* working environment. *(3)* Imports example data set as an object in *R*. *(4)* Imports function code for subpopulation analysis in *R*. *(5)-(8)* Four different modes of subpopulation analyses where x is assigned the data to be analysed (in this case *example.data*). Modes *Default* and *Manual* represent interactive methods requiring mouse-clicking on a graph. Modes *Boxplot1.5* and *Boxplot3* run automatically once initiated. Mode *Default* requires two mouse-clicks on the approximate minimum and maximum of the large subpopulation, respectively. Mode *Manual* only requires one mouse-click on the approximate point of separation of the to subpopulations. *(9)* Mode *other* uses other methods of characterization namely mean, 95th percentile, boosted mean, and mean top five. *(6)-(13)* A function for each mode to determine confidence intervals reflecting consistency (reproducibility) of the method. Automatic modes *Boxplot1.5* and *Boxplot3* employ bootstrapping with R replicates. The Default setting of R is set to 500 but can be changed. Similarly, interactive modes *Default* and *Manual* employ resampled trials of x repetitions, with the Default set to 20 trials. *(14)-(17)* Confidence interval determination for mean, 95th percentile, boosted mean, and mean top five are based on bootstrapping with R = 500 rounds.

A

```
> find.sub.pop(x=example.data$value, mode="default")

  mode cutoff n.sub-pop n.total %.sub-pop mean.sub-pop
1 default  69.29      72    2593      2.78       140.24
```

B

Normal Q-Q Plot
cutoff mode= default

C

Histogram of subpop.

D

**Normal Q-Q Plot
of subpop.**
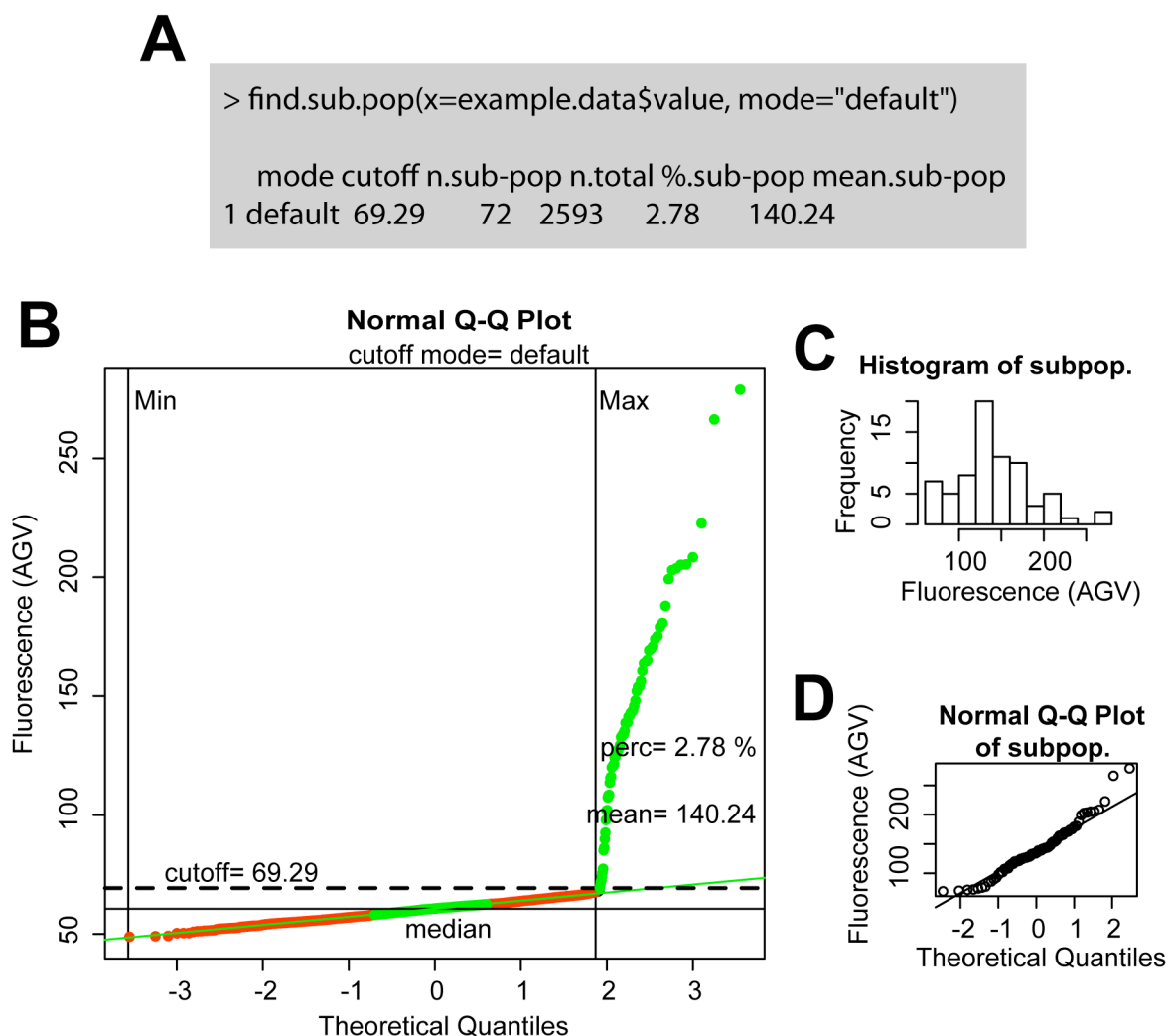
Protocol S1 Figure 2. Example of *R* output of the function *findsub(...)* with mode set to *Default* and *example.data* used as data set. (A) Command line output stating the cutoff value at which the small subpopulation was separated from the large subpopulation (cutoff), the size n of the small subpopulation (n.sub-pop) and the total population (n.total), the percentage of the small subpopulation (perc.sub-pop), and the mean value of the small subpopulation (mean.sub-pop). (B) Graphical output highlighting the manually placed minimum (Min) and maximum (Max) thresholds that encompass the estimated large subpopulation (vertical lines), its median (lower horizontal line), the points belonging to the IQR within the large subpopulation (green, vertical), the regression line fitted to the IQR points (green line), the cutoff value separating large and small subpopulation (dotted horizontal line), and the points belonging to the small subpopulation (green, vertical). (C) Histogram graphical output of the estimated small subsubpopulation. (D) QQ plot of the estimated small subpopulation.

A

```
> find.sub.pop(x=example.data$value, mode="manual")

  mode cutoff n.sub-pop n.total %.sub-pop mean.sub-pop
1 manual   69       72    2593      2.78       140.24
```

B

Normal Q-Q Plot
cutoff mode= manual

mean= 140.24
cutoff= 69   perc= 2.78 %

Fluorescence (AGV)
Theoretical Quantiles

C Histogram of subpop.

Frequency
Fluorescence (AGV)

D Normal Q-Q Plot of subpop.

Fluorescence (AGV)
Theoretical Quantiles
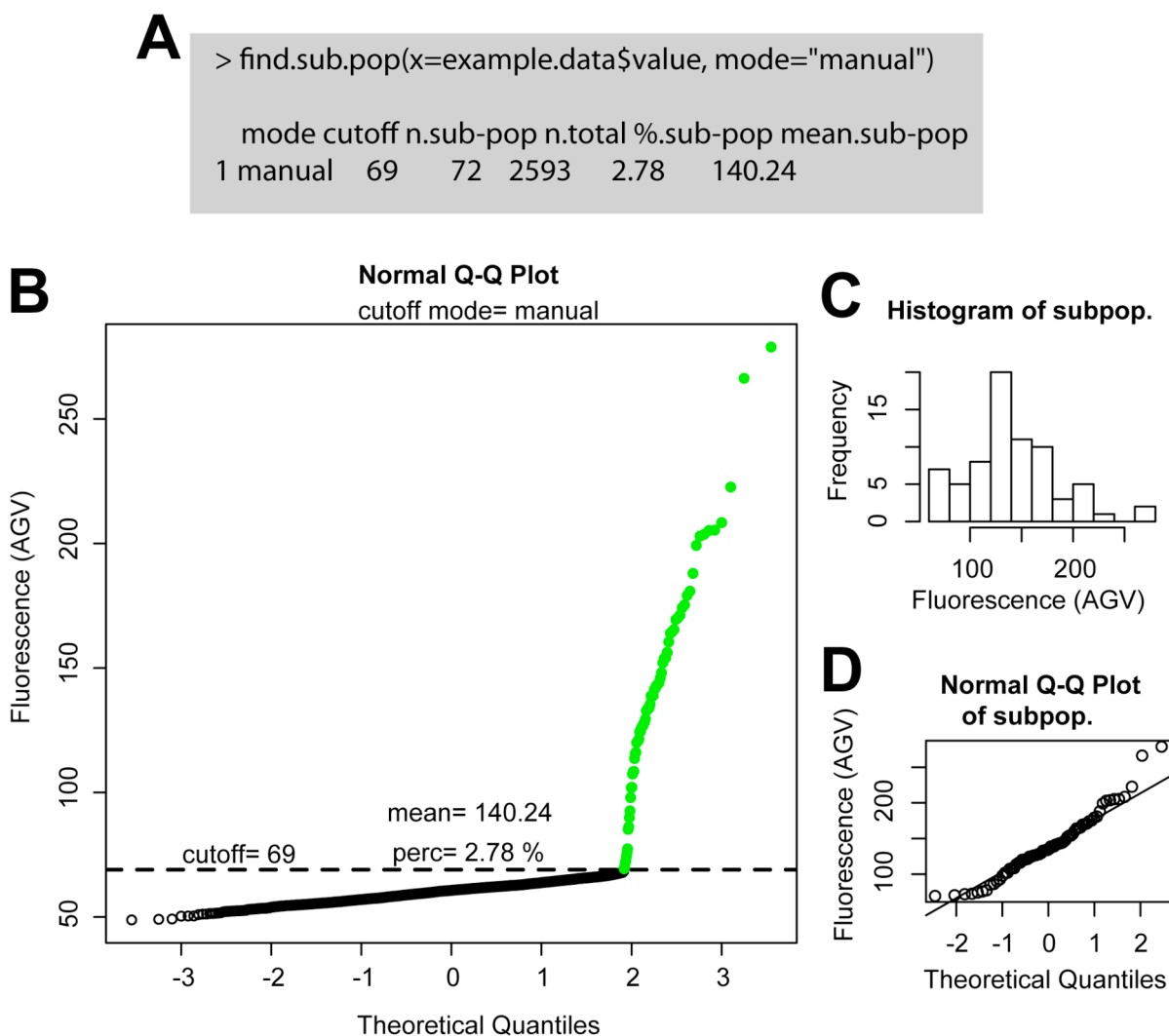
Protocol S1 Figure 3. Example of *R* output of the function *findsub(...)* with mode set to *Manual* and *example.data* used as data set. (A) Command line output stating the cutoff value at which the small subpopulation was separated from the large subpopulation (cutoff), the size of the small subpopulation (n.sub-pop) and the total population (n.total), the percentage of the small subpopulation (perc.sub-pop), and the mean value of the small subpopulation (mean.sub-pop). (B) Graphical output highlighting the manually placed cutoff separating large and small subpopulation (dotted horizontal line), and the points categorized as small subpopulation (green). (C) Histogram of the small estimated subpopulation. (D) QQ plot of the small estimated subpopulation.

A

> find.sub.pop(x=example.data$value, mode="boxplot1.5")

```
      mode cutoff n.sub-pop n.total %.sub-pop mean.sub-pop
1 boxplot1.5  69.39        72    2593      2.78       140.24
```

B

Normal Q-Q Plot
cutoff mode= boxplot1.5

mean= 140.24

cutoff= 69.39   perc= 2.78 %

Fluorescence (AGV)

Theoretical Quantiles

C

Histogram of subpop.

Frequency

100   200
Fluorescence (AGV)

D

Normal Q-Q Plot of subpop.

Fluorescence (AGV)

-2 -1 0 1 2
Theoretical Quantiles
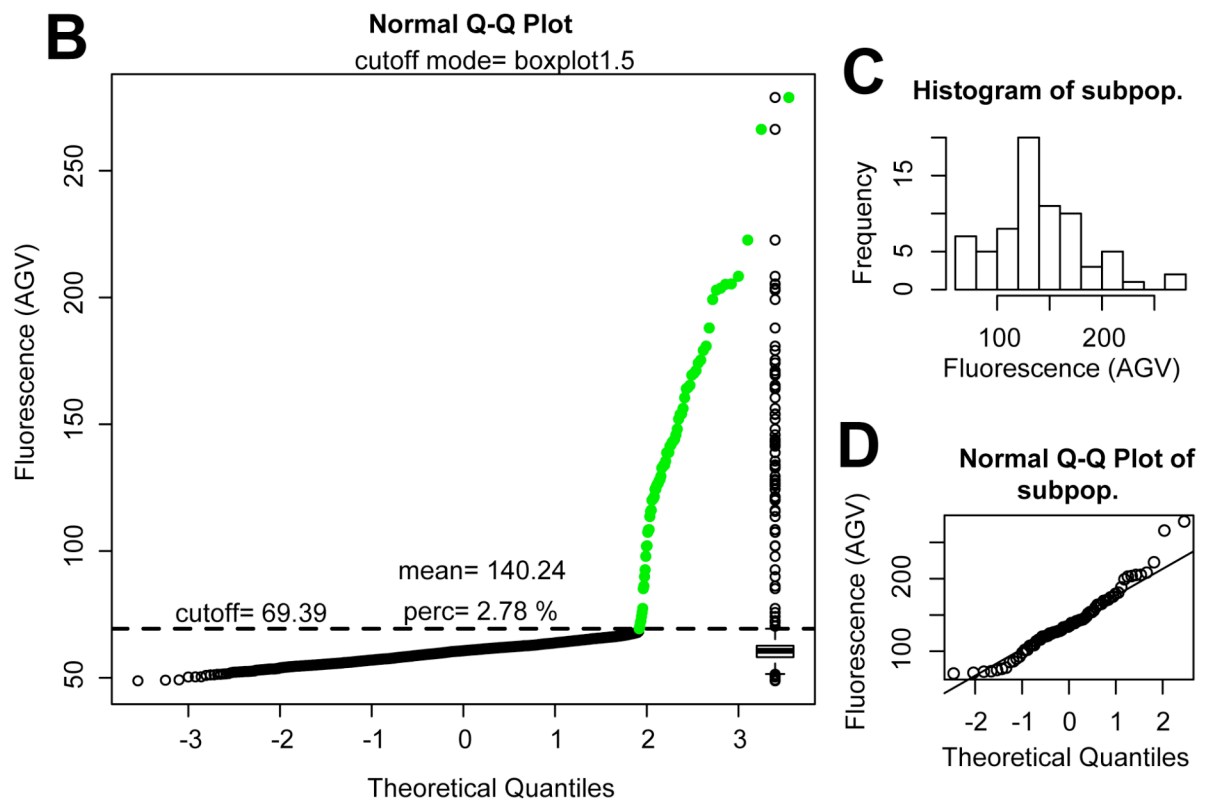
Protocol S1 3 Figure 4. Same as Figure 2 but with mode set to *Boxplot1.5*. Note that the cutoff value in this case is determined by how far the upper whisker of the boxplot extends, which is no more than 1.5 times the interquartile range from the box (upper quartile).

A

> find.sub.pop(x=example.data$value, mode="boxplot3")

```
   mode cutoff n.sub-pop n.total %.sub-pop mean.sub-pop
1 boxplot3 75.77      66    2593     2.55      146.41
```

B

Normal Q-Q Plot
cutoff mode= boxplot3

C

Histogram of subpop.

D

Normal Q-Q Plot of subpop.
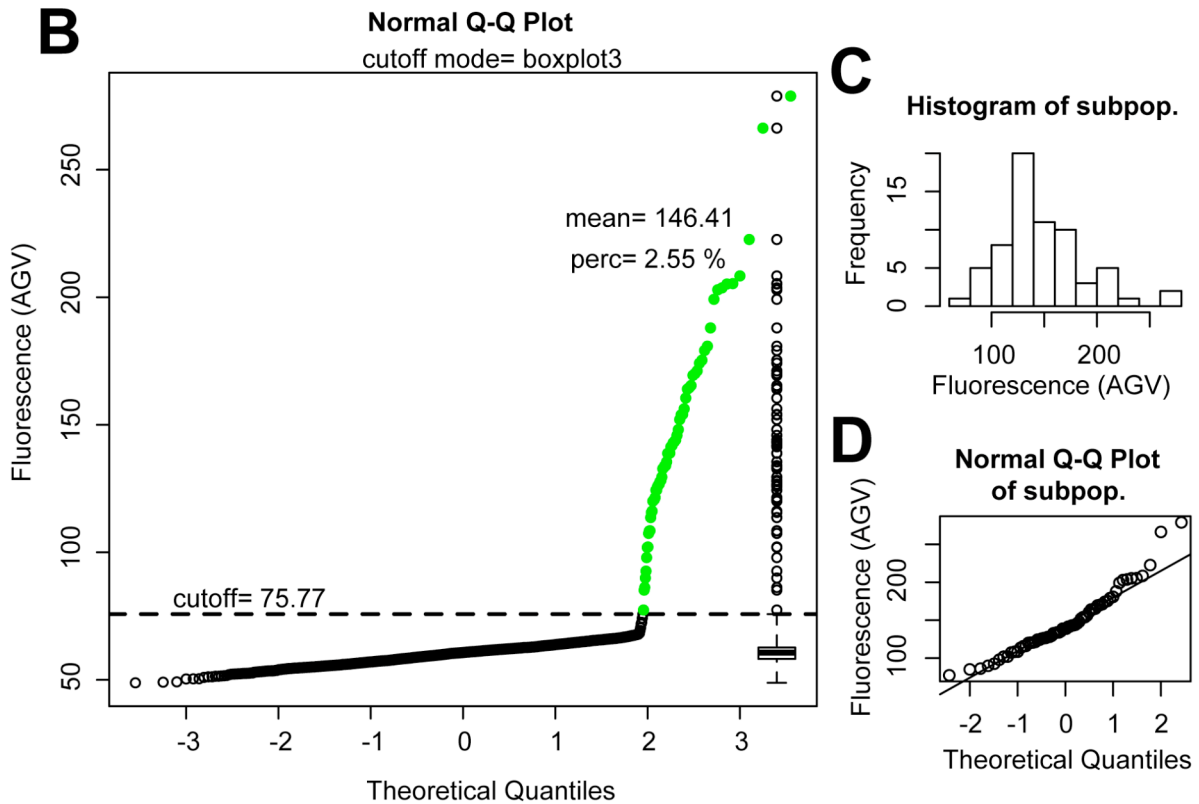
Protocol S1 Figure 5. Same as Figure 2 but with mode set to *Boxplot3*. Note that the cutoff value in this case is determined by how far the upper whisker of the boxplot extends, which is no more than 3 times the interquartile range from the box (upper quartile).

> find.sub.pop(x=example.data$value, mode="other")

```
  mean perc.95 perc.99  boost   top5
  62.5   66.2   146.0   64.1   107.5
```

Protocol S1 Figure 6. Same Figure 1 but with mode set to *Other*. Mode *Other* uses other methods of characterization namely mean (mean), 95th percentile (perc.95), 99th percentile (perc.99), boosted mean (boost), and mean top five (top5) of the total population, respectively.

A

```
> get.ci(x=example.data$value, trials=20, mode="default")
$quantiles
  2.5%   97.5%
2.25900 3.28275
$sd
[1] 0.2983887
$mean
[1] 2.796
$ci.95
[1] 2.211169 3.380831
$trials
 [1] 2.93 2.66 2.97 2.82 2.82 3.09 2.66 2.31 3.12 3.05
2.85 2.28 2.89 3.43 2.74 2.89 2.24 2.85 2.47 2.85
```

B mode=default

Histogram of my.replicate **Normal Q-Q Plot**
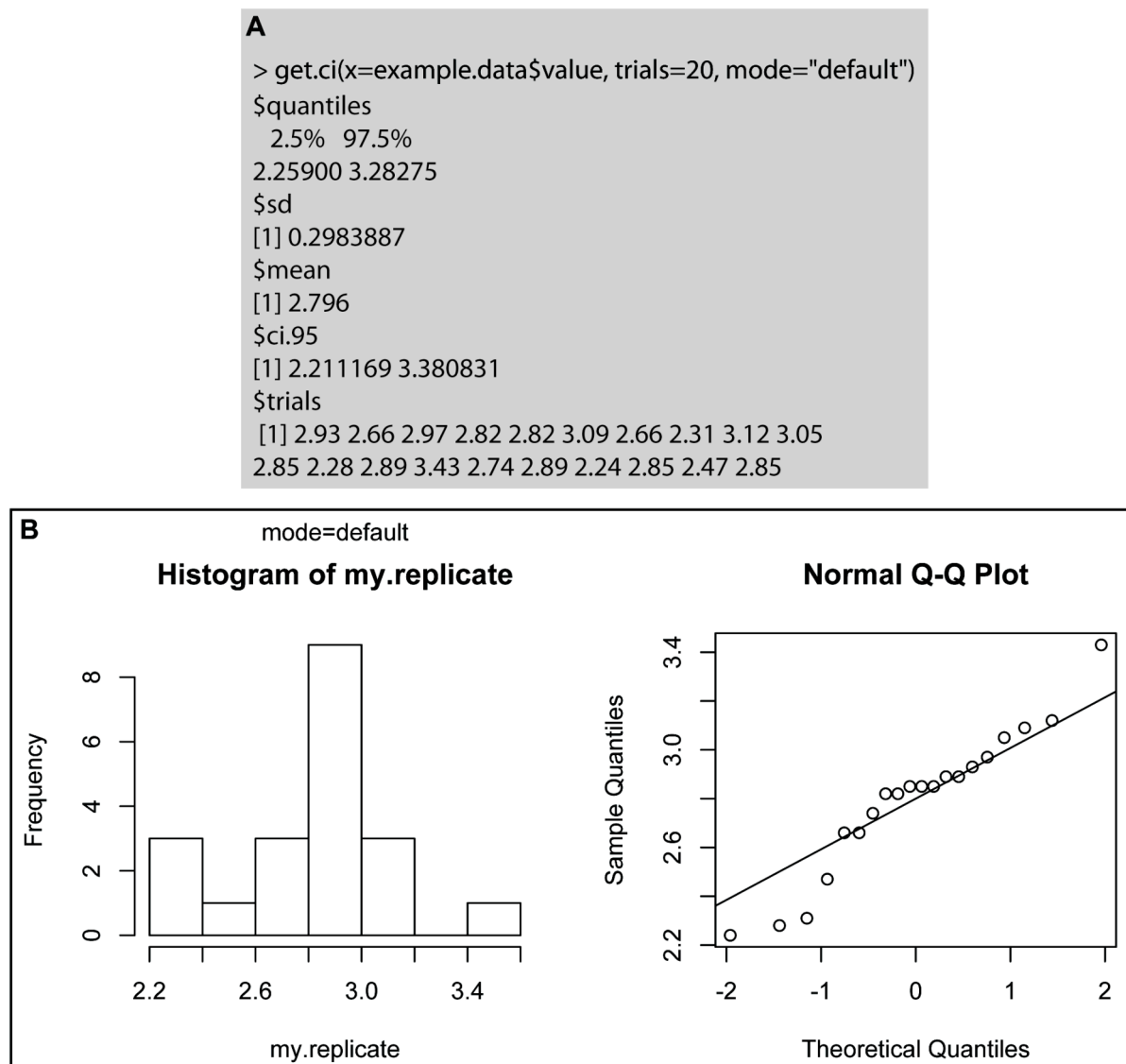
Protocol S1 Figure 7. Example of *R* output of the function *get.ci(...)* with mode set to *Default* and *example.data.txt* used as data set. (A) Command line output stating the 2.5 and 97.5 percentiles of the trial results ($quantiles), the standard deviation ($sd), the mean ($mean), and the 95% confidence interval ($ci.95). The latter was calculated according to $CI_{upper/lower} = mean \pm SD \times 1.96$, where $CI_{upper/lower}$ is the upper or lower confidence interval, *mean* is the population mean, and *SD* is the standard deviation of the boostrap results. Also shown are the individual results of applying *get.ci(...)* with mode set to *Default* on the re-sampled dataset ($trials). (B) Histogram graphical output (left) and QQ-plot graphical output (right) of the bootstrap results.

A

```
> get.ci(x=example.data$value, R=500, mode="boxplot1.5")
                   results
R              500.000000
Result           2.780000
Normal.low       2.148359
Normal.high      3.385401
Basic.low        2.090000
Basic.high       3.400000
Percent.low      2.160000
Percent.high     3.470000
```
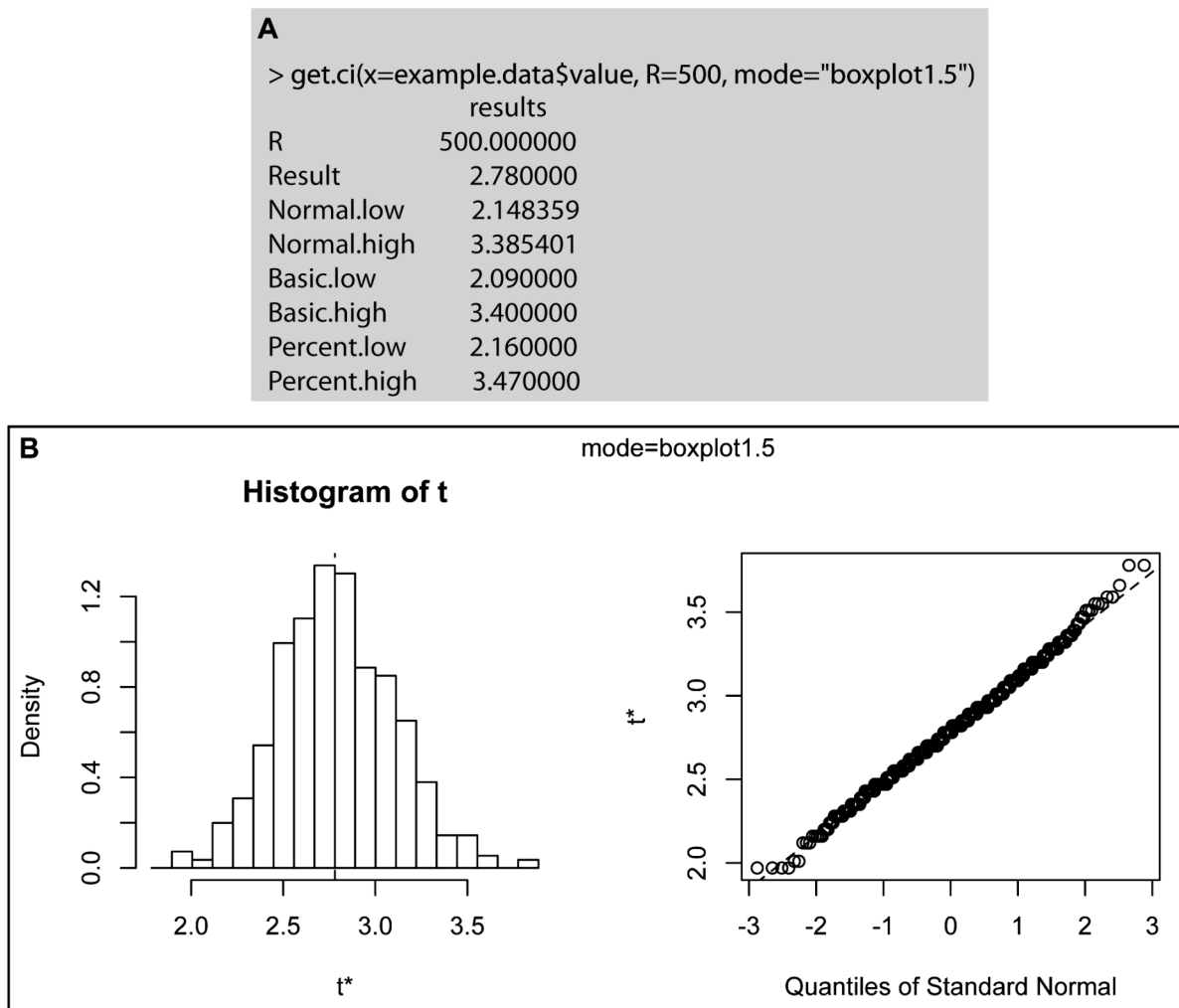
B   mode=boxplot1.5

Protocol S1 Figure 8. Example of *R* output of the function *get.ci(...)* with mode set to *Boxplot1.5* and *example.data.txt* used as data set. (A) Command line output stating the number of bootstraps (R), the result of applying *get.ci(...)* with mode *Boxplot1.5* on the original dataset (Result), and confidence intervals according to different algorithms including the basic bootstrap method (Basic), the studentized bootstrap method (Normal) and the bootstrap percentile method (Percent). (B) Histogram graphical output (left) and QQ-plot graphical output (right) of the bootstrap results.

References

1.   R Development Core Team: **R: A Language and Environment for Statistical Computing**. *R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0, URL* [*http://wwwR-projectorg*](http://wwwR-projectorg) 2009.

2.   Canty A, Ripley B: **boot: Bootstrap R (S-Plus) functions**. *R package version 12-41* 2009.

3.   Davison AC, Hinkley DV: **Bootstrap methods and their applications**. Cambridge: Cambridge University Press; 1997.

<u>Protocol S2</u>

Protocol S2: step-by-step script for *R*

#.. Step 1: Create a clean R environment.

rm(list=ls(all=TRUE))

#.. Step 2: Set the R work folder address. You need to create or choose a folder on your computer and enter its address into the command below. Tip: if you mouse-drag this folder into the setwd("...") location in this document, its address will be copied there saving you laborious address typing.

setwd("/Users/fedor/Desktop/R.test/FINDSUB") # sets work folder address

#.. Step 3: Before you continue, make sure to place into the newly created work folder the files FINDSUB_code.txt and example.data.txt and if you want to analyze our own data, also another file named my.data.txt (see step 4). Now import into R the data text file called "example.data.txt" via the read.table() command as shown below. This data set is required for demonstrative purposes of this script, but it is also needed when your own data is analyzed. Generally, to be recognized by the R commands of this script, the data to be imported from the text file needs to be organized in a specific way in the text file: a first line called the "header" consisting of three, space delimited words named "value", "image" and "name". Below these words then, are placed three columns of the corresponding data (see the file examle.data.txt). The category "value" contains the cell specific parameter under investigation, such as cellular fluorescence values. The category "image" describes from which image or sample the "value" originates from. The category "name" is the name of the experiment.

example.data <- read.table("example.data.txt", header=TRUE) # imports example.data.txt

#.. Step 4: Import into R the textfile called "my.data.txt". This file contains your own data set. Create a text file called "my.data.txt" and place it together with the files "examle.data.txt", "FINDSUB_code", and "script" into your R workfolder. This text file should be organized exactly as the file "example.data.txt" with the same column names ("value", "image" and "name") but

with your own data now replacing the example data. Tip: organize your data in Microsoft Excel (Microsoft Corporation, Redmond, Wash.) and then copy-paste it to a blank text file called "my.data.txt". Missing values should be replaced by "NA". As an example of how your data should be organized have a look at the file "example.data.txt".

```
my.data <- read.table("my.data.txt", header=TRUE) # imports my.data.txt
```

```
#.. Step 5: Quality control: Look at your data image by image (example data). To look at your own data replace "example.data" with "my.data.txt".
```

```
boxplot(example.data$value~example.data$image) # draws boxplots
```

```
#.. Step 6: Import the textfile called "FINDSUB_code.txt". This file contains the functions that are required for data analysis with this script. Furthermore, it reads in the dataset "example.data" for demonstrative purposes. Both, the file "FINDSUB_code.txt" and "examle.data.txt" need to be present in the R work folder.
```

```
source("FINDSUB_code.txt") # imports code
```

```
#.. Step 7: Data analysis in four different modes using the example data set "example.data$value", where "default" and "manual" modes are interactive and require mouse-clicking on a QQ plot graph, whereas "boxplot1.5" and "boxplot3" modes are fully automatic. "default" mode requires two mouse-clicks on the approximate start and end of the approximate main-population, respectively. "manual" mode only requires one mouse-cklicking on the approximate point of separation between main- and sub-population. Mode "other" uses other methods of characterization namely mean, 95th percentile, boosted mean, and mean top five. To look at your own data replace "example.data" with "my.data.txt".
```

```
findsub(x=example.data$value, mode="default")
findsub(x=example.data$value, mode="manual")
findsub(x=example.data$value, mode="boxplot1.5")
findsub(x=example.data$value, mode="boxplot3")
```

```
findsub(x=example.data$value, mode="other")
```

#.. Step 8: A function for each mode to determine confidence intervals reflecting consistency (reproducability) of the method. Automatic modes "boxplot1.5" and "boxplot3" employ bootstrapping with R rounds. The default setting of R is set to 500 but can be changed to any number. Similarly, interactive modes "default" and "manual" employ resampled trials of a number of repetitions, with the default set to 20 trials. To look at your own data replace "example.data" with "my.data.txt".

```
get.ci(x=example.data$value, trials=20, mode="default")
get.ci(x=example.data$value, trials=20, mode="manual")
get.ci(x=example.data$value, R=500, mode="boxplot1.5")
get.ci(x=example.data$value, R=500, mode="boxplot3")
```

#.. Step 9: Same as Step 8 but for the "other" methods: mean, 95th percentile, boosted mean, and mean top five.

```
get.ci.other(x=example.data$value, R=500, mode="mean")
get.ci.other(x=example.data$value, R=500, mode="perc.95")
get.ci.other(x=example.data$value, R=500, mode="top.5")
get.ci.other(x=example.data$value, R=500, mode="boost")
```

Protocol S2: *R* code

The following code defines the functions findsub(), get.ci(), and get.ci.other(). findsub() quantifies small sub-population effects according to the methods "default", "manual", "boxplot1.5", "boxplot3" or "other", which can be chosen via the argument mode. The function get.ci() bootstraps "default", "manual", "boxplot1.5", or "boxplot3", which can be chosen via the argument mode. Arguments R and trials determine the number of bootstraps for the methods "boxplot1.5" or "boxplot3" and "default" or "manual", repectively. The function get.ci.other() bootstraps population mean (mean), 95th percentile (perc.95), booseted mean (boost), population mean top 5% (top.5). Argument R determines the number of bootstraps.

```
############ findsub function
findsub <- function (x = example.data$value, mode = "default") {
        graphics.off()
        dev.off(!1)
        if (mode == "default") {
                c <-as.data.frame(qqnorm(x,ylab="Fluorescence (AGV)"))
                mtext(3,text= paste ("cutoff mode=",mode))
                my.p1<-locator(n=1)
                abline(v=my.p1$x)
                text("Min", x= my.p1$x+0.25, y=max(c$y)-2*max(c$y)/100)
                my.p2<-locator(n=1)
                abline(v=my.p2$x)
                text("Max", x= my.p2$x+0.25, y=max(c$y)-2*max(c$y)/100)

                sub.c <- subset(c, x>=my.p1$x & x<=my.p2$x)
                points(sub.c, pch = 21, bg="red", col="red", cex = 1)

                upper.y <- quantile(sub.c$y,0.75)
                lower.y <- quantile(sub.c$y,0.25)
                median.y <- median (sub.c$y)

                core <- subset(sub.c, y>=lower.y & y<=upper.y)
                points(core, pch = 21, bg="green", col="green", cex = 1)
                abline(h=median.y, col="black")
                text("median", x= -2.5, y=median.y+2*max(c$y)/100)
```

```
lm.core <- lm(core$y~core$x)
abline(lm.core, col="green")


slope <- coefficients(lm.core)[[2]]
cut.off<-round(slope*qnorm(0.995)+median.y,2)
abline(h=cut.off, col="black", lwd=2, lty=2)
text(paste("cutoff=",cut.off), x= -2.5, y=cut.off+2*max(c$y)/100)


subpop <<- subset(c, y > cut.off)
points(subpop, pch = 21, bg="green", col="green", cex = 1)


sub.perc <<- round(length(subpop$y)/length(c$y)*100,2)
sub.mean <- round(mean(subpop$y),2)


text(paste("perc=",sub.perc,"%"),x=0,y=cut.off+2*max(c$y)/100,col="black")
text(paste("mean=",sub.mean),x=0,y=cut.off+8*max(c$y)/100,col="black")


if (length(subpop$x) > 1){
        dev.new(width=6,height=3)
        par(mfrow=c(1,2))
        hist(subpop$y, xlab="Fluorescence (AGV)")
        qqnorm(subpop$y, ylab="Fluorescence (AGV)")
        qqline(subpop$y)
}



        my.data.frame <-
data.frame(mode,cut.off,length(subpop$y),length(c$y),sub.perc,sub.mean, row.names=NULL)
        names(my.data.frame)<-c("mode","cutoff","n.sub-pop","n.total","%.sub-
pop","mean.sub-pop")
        my.data.frame


} else if (mode == "manual") {


        c <-as.data.frame(qqnorm(x,ylab="Fluorescence (AGV)"))
        mtext(3,text= paste ("cutoff mode=",mode))
        my.p1<-locator(n=1)
        #abline(v=my.p1$x)
```

```
abline(h=my.p1$y, col="black", lwd=2, lty=2)


cut.off<-round(my.p1$y,2)
text(paste("cutoff=",cut.off), x= -2.5, y=cut.off+2*max(c$y)/100)


subpop <<- subset(c, y > cut.off)
points(subpop, pch = 21, bg="green", col="green", cex = 1)


sub.perc <<- round(length(subpop$y)/length(c$y)*100,2)
sub.mean <- round(mean(subpop$y),2)


text(paste("perc=",sub.perc,"%"),x=0,y=cut.off+2*max(c$y)/100,col="black")
text(paste("mean=",sub.mean),x=0,y=cut.off+8*max(c$y)/100,col="black")
mtext(3,text= paste ("cutoff mode=",mode))


if (length(subpop$x) > 1){
        dev.new(width=6,height=3)
        par(mfrow=c(1,2))
        hist(subpop$y, xlab="Fluorescence (AGV)")
        qqnorm(subpop$y, ylab="Fluorescence (AGV)")
        qqline(subpop$y)
}


        my.data.frame <-
data.frame(mode,cut.off,length(subpop$y),length(c$y),sub.perc,sub.mean, row.names=NULL)
        names(my.data.frame)<-c("mode","cutoff","n.sub-pop","n.total","%.sub-
pop","mean.sub-pop")
        my.data.frame


    } else if (mode == "boxplot1.5") {


        c <-as.data.frame(qqnorm(x,ylab="Fluorescence (AGV)"))


        cut.off<-round(boxplot(c$y,plot=T, range=1.5, add=T, h=F, at=3.4)$stats[5],2)
        abline(h=cut.off, col="black", lwd=2, lty=2)
        text(paste("cutoff=",cut.off), x= -2.5, y=cut.off+2*max(c$y)/100)


        subpop <<- subset(c, y > cut.off)
```

158

```
points(subpop, pch = 21, bg="green", col="green", cex = 1)


sub.perc <<- round(length(subpop$y)/length(c$y)*100,2)
sub.mean <- round(mean(subpop$y),2)


text(paste("perc=",sub.perc,"%"),x=0,y=cut.off+2*max(c$y)/100,col="black")
text(paste("mean=",sub.mean),x=0,y=cut.off+8*max(c$y)/100,col="black")
mtext(3,text= paste ("cutoff mode=",mode))


dev.new(width=6,height=3)
par(mfrow=c(1,2))
hist(subpop$y, xlab="Fluorescence (AGV)")
qqnorm(subpop$y, ylab="Fluorescence (AGV)")
qqline(subpop$y)


my.data.frame <-
data.frame(mode,cut.off,length(subpop$y),length(c$y),sub.perc,sub.mean, row.names=NULL)
        names(my.data.frame)<-c("mode","cutoff","n.sub-pop","n.total","%.sub-
pop","mean.sub-pop")
        my.data.frame


} else if (mode == "boxplot3") {


c <-as.data.frame(qqnorm(x,ylab="Fluorescence (AGV)"))


cut.off<-round(boxplot(c$y,plot=T, range=3, add=T, h=F, at=3.4)$stats[5],2)
abline(h=cut.off, col="black", lwd=2, lty=2)
text(paste("cutoff=",cut.off), x= -2.5, y=cut.off+2*max(c$y)/100)


subpop <<- subset(c, y > cut.off)
points(subpop, pch = 21, bg="green", col="green", cex = 1)


sub.perc <<- round(length(subpop$y)/length(c$y)*100,2)
sub.mean <- round(mean(subpop$y),2)


text(paste("perc=",sub.perc,"%"),x=0,y=cut.off+2*max(c$y)/100,col="black")
text(paste("mean=",sub.mean),x=0,y=cut.off+8*max(c$y)/100,col="black")
mtext(3,text= paste ("cutoff mode=",mode))
```

```
            dev.new(width=6,height=3)

            par(mfrow=c(1,2))

            hist(subpop$y, xlab="Fluorescence (AGV)")

            qqnorm(subpop$y, ylab="Fluorescence (AGV)")

            qqline(subpop$y)


            my.data.frame <-
data.frame(mode,cut.off,length(subpop$y),length(c$y),sub.perc,sub.mean, row.names=NULL)
            names(my.data.frame)<-c("mode","cutoff","n.sub-pop","n.total","%.sub-
pop","mean.sub-pop")
            my.data.frame


      } else if (mode == "other") {

            q.75 <- quantile(x,.75)

            q.95 <- quantile(x,.95)

            my.mean <- mean(x)

            my.perc.95 <- quantile(x,.95)

            my.perc.99 <- quantile(x,.99)

            my.boost <- mean(x[x > q.75 & x < q.95])

            my.top5 <- mean(x[x > q.95])


            my.data.frame <- data.frame(my.mean,my.perc.95,my.perc.99,my.boost,my.top5,
row.names=NULL)
            names(my.data.frame)<-c("mean","perc.95","perc.99","boost","top5")
            my.data.frame


      }
}



############ load library for bootstrap functions

library("boot")

############ bootstrap functions for population splitting methods
############ get.ci
```

```
fun.boxplot1.5 <- function (x,ind) {

        cut.off<-round(boxplot(x[ind] ,plot=F, range=1.5, add=T, h=F, at=3.4)$stats[5],2)

        subpop <- x[ind][x[ind] > cut.off]

        sub.perc <- round(length(subpop)/length(x[ind])*100,2)

        sub.perc

        }


fun.boxplot3 <- function (x,ind) {

        cut.off<-round(boxplot(x[ind] ,plot=F, range=3, add=T, h=F, at=3.4)$stats[5],2)

        subpop <- x[ind][x[ind] > cut.off]

        sub.perc <- round(length(subpop)/length(x[ind])*100,2)

        sub.perc

        }



get.ci <- function (x = example.data$value, mode="default",trials=20, R=500) {

        if (mode == "boxplot1.5") {

                my.boot<-boot(x,fun.boxplot1.5, R=R)

                graphics.off()

                dev.off(!1)

                dev.new(width=8,height=4)

                plot(my.boot)

                mtext("mode=boxplot1.5", line=3)

                my.boot.ci<-boot.ci(my.boot,type=c("norm","basic","perc"))

                names<-
c("R","Result","Normal.low","Normal.high","Basic.low","Basic.high","Percent.low","Percent.high")

                results<-
c(my.boot.ci$R,my.boot.ci$t0,my.boot.ci$normal[2],my.boot.ci$normal[3],my.boot.ci$basic[4],my.boot.ci
$basic[5],my.boot.ci$percent[4],my.boot.ci$percent[5])

                data.frame(row.names=names,results)


        } else if (mode == "boxplot3") {

                my.boot<-boot(x,fun.boxplot3, R=R)

                graphics.off()

                dev.off(!1)

                dev.new(width=8,height=4)

                plot(my.boot)
```

161

```
                mtext("mode=boxplot3", line=3)

                my.boot.ci<-boot.ci(my.boot,type=c("norm","basic","perc"))

                names<-
c("R","Result","Normal.low","Normal.high","Basic.low","Basic.high","Percent.low","Percent.high")

                results<-
c(my.boot.ci$R,my.boot.ci$t0,my.boot.ci$normal[2],my.boot.ci$normal[3],my.boot.ci$basic[4],my.boot.ci
$basic[5],my.boot.ci$percent[4],my.boot.ci$percent[5])

                data.frame(row.names=names,results)


        } else if (mode == "manual") {

                my.replicate <<- replicate(trials,findsub(sample(x,size=length(x),replace=T),
mode="manual")[1,5])

                graphics.off()

                dev.off(!1)

                dev.new(width=8,height=4)

                par(mfrow=c(1,2))

                hist(my.replicate)

                mtext("mode=manual", line=3)

                qqnorm(my.replicate)

                qqline(my.replicate)

                list(

                quantiles=quantile(my.replicate,probs = c(0.025, 0.975)),

                sd=sd(my.replicate),

                mean=mean(my.replicate),

                ci.95=c(mean(my.replicate)-
qnorm(0.975)*sd(my.replicate),mean(my.replicate)+qnorm(0.975)*sd(my.replicate)),

                trials=my.replicate)


        } else if (mode == "default") {

                my.replicate <<- replicate(trials,findsub(sample(x,size=length(x),replace=T),
mode="default")[1,5])

                graphics.off()

                dev.off(!1)

                dev.new(width=8,height=4)

                par(mfrow=c(1,2))

                hist(my.replicate)

                mtext("mode=default", line=3)

                qqnorm(my.replicate)
```

```
            qqline(my.replicate)
            list(
            quantiles=quantile(my.replicate,probs = c(0.025, 0.975)),
            sd=sd(my.replicate),
            mean=mean(my.replicate),
            ci.95=c(mean(my.replicate)-
qnorm(0.975)*sd(my.replicate),mean(my.replicate)+qnorm(0.975)*sd(my.replicate)),
            trials=my.replicate)


        }
}




############ bootstrap functions for non-population splitting methods
############ get.ci.other

fun.mean <- function (x,ind) {
        mean(x[ind])
        }

fun.perc95 <- function (x,ind) {
        quantile(x[ind],.95)
        }

fun.top5 <- function (x,ind) {
        q.95 <- quantile(x[ind],.95)
        mean(x[ind][x[ind] > q.95])
        }

fun.boost <- function (x,ind) {
        q.75 <- quantile(x[ind],.75)
        q.95 <- quantile(x[ind],.95)
        mean(x[ind][x[ind] > q.75 & x[ind] < q.95])
        }


get.ci.other <- function(x, mode, R) {
```

```
if (mode == "mean") {
        my.boot<-boot(x,fun.mean, R=R)
        graphics.off()
        dev.off(!1)
        dev.new(width=8,height=4)
        plot(my.boot)
        mtext("mode=mean", line=3)
        my.boot.ci<<-boot.ci(my.boot,type=c("norm","basic","perc"))
        names<-
c("R","Result","Normal.low","Normal.high","Basic.low","Basic.high","Percent.low","Percent.high")
        results<-
c(my.boot.ci$R,my.boot.ci$t0,my.boot.ci$normal[2],my.boot.ci$normal[3],my.boot.ci$basic[4],my.boot.ci
$basic[5],my.boot.ci$percent[4],my.boot.ci$percent[5])
        data.frame(row.names=names,results)


} else if (mode == "perc.95") {
        my.boot<-boot(x,fun.perc95, R=R)
        graphics.off()
        dev.off(!1)
        dev.new(width=8,height=4)
        plot(my.boot)
        mtext("mode=perc.95", line=3)
        my.boot.ci<<-boot.ci(my.boot,type=c("norm","basic","perc"))
        names<-
c("R","Result","Normal.low","Normal.high","Basic.low","Basic.high","Percent.low","Percent.high")
        results<-
c(my.boot.ci$R,my.boot.ci$t0,my.boot.ci$normal[2],my.boot.ci$normal[3],my.boot.ci$basic[4],my.boot.ci
$basic[5],my.boot.ci$percent[4],my.boot.ci$percent[5])
        data.frame(row.names=names,results)


} else if (mode == "top.5") {
        my.boot<-boot(x,fun.top5, R=R)
        graphics.off()
        dev.off(!1)
        dev.new(width=8,height=4)
        plot(my.boot)
        mtext("mode=top.5", line=3)
```

```
                my.boot.ci<<-boot.ci(my.boot,type=c("norm","basic","perc"))
                names<-
c("R","Result","Normal.low","Normal.high","Basic.low","Basic.high","Percent.low","Percent.high")
                results<-
c(my.boot.ci$R,my.boot.ci$t0,my.boot.ci$normal[2],my.boot.ci$normal[3],my.boot.ci$basic[4],my.boot.ci
$basic[5],my.boot.ci$percent[4],my.boot.ci$percent[5])
                data.frame(row.names=names,results)


        } else if (mode == "boost") {
                my.boot<-boot(x,fun.boost, R=R)
                graphics.off()
                dev.off(!1)
                dev.new(width=8,height=4)
                plot(my.boot)
                mtext("mode=boost", line=3)
                my.boot.ci<<-boot.ci(my.boot,type=c("norm","basic","perc"))
                names<-
c("R","Result","Normal.low","Normal.high","Basic.low","Basic.high","Percent.low","Percent.high")
                results<-
c(my.boot.ci$R,my.boot.ci$t0,my.boot.ci$normal[2],my.boot.ci$normal[3],my.boot.ci$basic[4],my.boot.ci
$basic[5],my.boot.ci$percent[4],my.boot.ci$percent[5])
                data.frame(row.names=names,results)
        }
}



############ define example data set
example.data <- read.table("example.data.txt", header=TRUE)
############
##.. Single cell eGFP fluorescence values of Pseudomonas knackmussi B13-1343 grown in batch
culture on minimal medium with 3-chlorobenzaote (10 mM) to stationary phase, sampled 10 h after
reaching stationary phase.
```

Protocol S3

Protocol S3: Script for using the function *popsim()*

```
# create a clean R environment.
rm(list=ls(all=TRUE))

# set work folder address
setwd("/Users/fedor/Desktop/R.test/POPSIM")

# import function code
source("POPSIM_code.txt")

## run simulation
popsim (range=1.5, length.main=2000, mean.main=63, sd.main=3.9, runs=5, perc.low=0.1,
perc.high=1.2, mean.low=65, mean.high=200, sd.low=10, sd.high=50)

## popsim argument description
# range:              define the method of sub-population detection: range = 1.5 (Boxplot1.5) or
range = 3 (Boxplot3)
# length.main:        define the total population length
# mean.main:          define the main populaion mean
# sd.main:            define the main population standard deviation
# runs:               define the number of different small sub-population parameters to be tested per
category
# perc.low:           define the smallest small sub-population to be tested (% of total population)
# perc.high:          define the largest small sub-population to be tested (% of total population)
# mean.low:           define the smalles small sub-population mean to be tested
# mean.high:          define the largest small sub-population mean to be tested
# sd.low:             define the smallest small sub-population standard deviation to be tested
# sd.high:            define the largest small sub-population standard deviation to be tested



## look at the raw data
run.param
```

run.param column description

lengths: length n parameters for simulations of small sub-populations

means: mean value parameters for simulations of small sub-populations

sds: standard deviation parameters for simulations of small sub-populations

msr: mean of simulated small sub-population

mse: mean of estimated small sub-population

m.diff: difference between simulated and estimated means

psr: percentage of simulated small sub-population

pse: percentage of estimated small sub-population

p.diff: difference between simulated and estimated percentage

sdsr: standard deviation of simulated small sub-population

sdse: standard deviation of estimated small sub-population

sd.diff: difference between simulated and estimated standard deviation

my.range: range argument setting as used in the R function boxplot()

lenght.mix: length n of total population

Protocol S3: *R* code for the function

The following code defines the function popsim(), which identifies sub-populations from simulated bimodal populations using the methods Boxplot1.5 (argument range = 1.5) or Boxplot3 (argument range = 3).

```
#------------------------------------------
   # 1) functions get.results and get.sub
#------------------------------------------

# the function get.results mixes main- and sub-population, separates them according to either
boxplot1.5 or boxplot3 method, and compares the result with the simulated sub-population
get.results <- function(u,v,w,range) {

        pop.sub.real <<- rnorm(u,v,w) # simulate small sub-population
        pop.mix <<- c(pop.sub.real, sample(pop.main,length(pop.main)-length(pop.sub.real))) # mix
small sub-population and main population
        pop.sub.exp <<- get.sub(range=range) # estimate small sub-popualtion using either the
Boxplot1.5 or the Boxplot3 method

        if ((length(pop.sub.exp)==0)==F) {

                sdsr <<- sd(pop.sub.real) # standard deviation of simulated small sub-population
                sdse <<- sd(pop.sub.exp)  # standard deviation of estimated small sub-population
                sd.diff <<- sd(pop.sub.exp)-sd(pop.sub.real) # difference between simulated and
estimated standard deviation

                psr <<- length(pop.sub.real)/length(pop.mix)*100 # percentage of simulated small sub-
population
                pse <<- length(pop.sub.exp)/length(pop.mix)*100 #  percentage of estimated small sub-
population
                p.diff <<- length(pop.sub.exp)/length(pop.mix)*100-
length(pop.sub.real)/length(pop.mix)*100 # difference between simulated and estimated percentage

                msr <<- mean(pop.sub.real) # mean of simulated small sub-population
                mse <<- mean(pop.sub.exp) #  mean of estimated small sub-population
```

168

```
        m.diff <<- mean(pop.sub.exp)-mean(pop.sub.real) # difference between simulated and
estimated means


        my.range <<- range
        length.mix <<-length(pop.mix)


    } else {
        sdsr <<- sd(pop.sub.real)
        sdse <<- NA
        sd.diff <<- NA


        psr <<- length(pop.sub.real)/length(pop.mix)*100
        pse <<- NA
        p.diff <<- NA


        msr <<- mean(pop.sub.real)
        mse <<- NA
        m.diff <<- NA
        sep.diff <<- NA


        my.range <<- range
        length.mix <<- length(pop.mix)
    }
}


#-------------------------------------------


# the function get.sub allows to define the methods Boxplot1.5 (range = 1.5) and Boxplot3 (range = 3)
get.sub <- function(range = range) {
        cut.off <- boxplot(pop.mix,plot=F, range=range)$stats[5]
        pop.mix [pop.mix > cut.off]
        }




#-------------------------------------------
  # 2) final simulation function
#-------------------------------------------
# use range = 1.5 for Boxplot1.5, or range = 3 for Boxplot3
```

```
popsim <- function (range=1.5, length.main=2000, mean.main=63, sd.main=3.9, runs=10,
perc.low=0.1, perc.high=1.2, mean.low=65, mean.high=200, sd.low=10, sd.high=50) {


        #------- setting population parameters ---------
        # creation of main population
        pop.main <<- rnorm (length.main, mean.main, sd.main)


        # determination of small sub-population parameters
        # determination of the sub-population sizes to be tested
        range.perc <- seq(from=perc.low,to=perc.high,length.out=runs)
        range.length <- range.perc*length.main/100


        # determiation of small sub-population means to be tested
        range.mean <- seq(from=mean.low,to=mean.high,length.out=runs)


        # calculation of the mean differences (mean.diff) between small sub-population and main
population
        range.meandiff <- range.mean - mean.main


        # determiation of small sub-population standard deviations to be tested
        range.sd <- seq(from=sd.low,to=sd.high,length.out=runs)


        # creating dataframe with all small sub-population parameters
        lengths <- rep(range.length,each=runs^2)
        means <- rep(rep(range.mean,each=runs),runs)
        sds <- rep(range.sd,runs^2)
        run.param <- data.frame(lengths=lengths, means=means, sds=sds)
        #-------------------------------------------


        for(i in 1:runs^3){
                        get.results (u=run.param[i,][[1]],v=run.param[i,][[2]],w=run.param[i,][[3]],
range=range)


                        run.param$msr[i] <- msr
                        run.param$mse[i] <- mse
                        run.param$m.diff[i] <- m.diff


                        run.param$psr[i] <- psr
```

```
        run.param$pse[i] <- pse
        run.param$p.diff[i] <- p.diff

        run.param$sdsr[i] <- sdsr
        run.param$sdse[i] <- sdse
        run.param$sd.diff[i] <- sd.diff

        run.param$my.range[i] <- my.range
        run.param$length.mix [i] <- length.mix
    }
    run.param<<-run.param
}
```

CHAPTER 3

Microcolony growth procedures

ABSTRACT

Microcolony growth procedures have gained recent interest for the following reasons: rapid simultaneous screening for growth of large numbers of bacteria from environmental samples, time-lapse imaging of single cell growth, cell aging or biofilm studies. Not many of such procedures have been described in sufficient technical detail to be easily reproduced. Here we provide a simple step-by-step procedure, which allows time-lapse imaging of bacterial cell division and cell tracking up to the stage of a monolayered colony of a few hundred cells.

INTRODUCTION

For a variety of reasons it can be interesting to study microbial growth at the micro- rather than at the macroscale as is typically done in shaken flask suspended cultures, bioreactors or the like. Numerous studies on bacterial biofilms or colonies have demonstrated conclusively that cells at different spatial locations within the closely packed cellular structure of a biofilm or colony face different growth conditions and can display different activity and phenotypes(15). More recently, the idea of phenotypic identity of individual bacteria among clonal populations has been challenged by the discovery and deeper understanding of bistable switches underlying e.g., competence development or sporulation (11), horizontal gene transfer (2, 12), by cell aging (14), differential killing and resistance development (3, 5), or by cooperative behaviour (1) and bet-hedging (16). Phenotypic differentiation was also shown to occur among bacteria degrading hydrocarbons in bioreactors (17), suggesting that many more such phenomena exist. To study individual cellular behaviour in populations one needs different tools that allow single cell distinction, analysis of cellular genealogies, and easy analysis of individual gene expression. Major advances in the use of autofluorescent proteins, fluorescent staining techniques or *in vivo* specific protein labelling techniques have enabled single cell study of gene expression and subcellular protein localization patterns (2, 6). Microscale growth experiments on the other hand are very helpful to set up the conditions under which to study individual cell behaviour. Combinations of single cell tracking in microscale growth experiments may well be used to monitor the reaction time of cells to signals and stimuli, such as light, poisons, nutrients, chemical agents, signalling compounds and so on.

Microscale experiments might also be interesting from the perspective of studying growth on traces of carbon compounds that otherwise are too toxic to be applied in large quantities. Microscale growth has already been shown to offer excellent possibilities for large library screening or for screening growth characteristics of large numbers of environmental isolates simultaneously in micro-Petri dishes (9). The procedure we describe below is an example of how colony formation from a single cell up to a few hundred cells can be monitored in real time by microscopic observation, which may be used for further advanced biodegradation studies as well.

GENERAL PROCEDURE STRATEGY AND OUTLINE

Microcolony growth procedures are designed as such that direct microscopic observations are possible in real time conditions without disturbance of the specimen. Several examples of microcolony growth assays exist and we will briefly mention a few of them, while concentrating our technical description on only one. Bacterial biofilm growth assays could actually be considered the first such strategy for real-time microcolony observation, in which the multilayered structure is analyzed during a period of days up to two weeks, as a function of cell type, gene mutations or surface material. In combination with time-lapse imaging it is possible to follow cellular division from single cells to a single-layered microcolony, from which individual cell division rates and cell pole age for every individual in the microcolony can be calculated.

A microcolony growth procedure has to start with the choice of material to grow the structure on. As mentioned, continuous flow biofilm growth experiments typically start with a closed chamber with glass cover slips on both sides (4, 15). Bacterial cells can be inoculated at low density, some of which will attach to the surface, and because of the constant carbon feed from one direction of the flow cell, will develop to a mature biofilm on the surface in a matter of days (Fig. 1). The formation and structure of the biofilm can be examined microscopically from underneath one of the glass cover slips, ideally by confocal laser scanning microscopy, in order to produce multilayered focussed images. Biofilm studies profit greatly from the use of fluorescently-tagged cells, which permit easier visualization of individual cells in a complex biofilm structure and allow the *in vivo* study of activity of specific cellular promoters in and during biofilm growth. Procedures for continuous flow biofilm formation have been described extensively elsewhere and need not be repeated here (4, 7). To the best of our knowledge, no study has yet tracked the origin and fate of every single cell in a mature biofilm by time-lapse imaging, but theoretically this should be possible.

Other microcolony growth procedures have used porous materials to inoculate cells onto and follow the microcolony development in time. Such porous materials can be bacterial filters or porous aluminium oxides that can be moulded to include nanocavities (9). The porous surfaces can then be placed on a medium surface, from which nutrients will diffuse through the support to the cells. Porous nanocavity plates have the advantage that growth of thousands of microcolonies can be followed simultaneously, because they are physically separated from each other. This permits growth analysis, for example, of separated individual bacterial cells in communities, of phenotypic variation upon stress (8), or of large libraries of mutants.
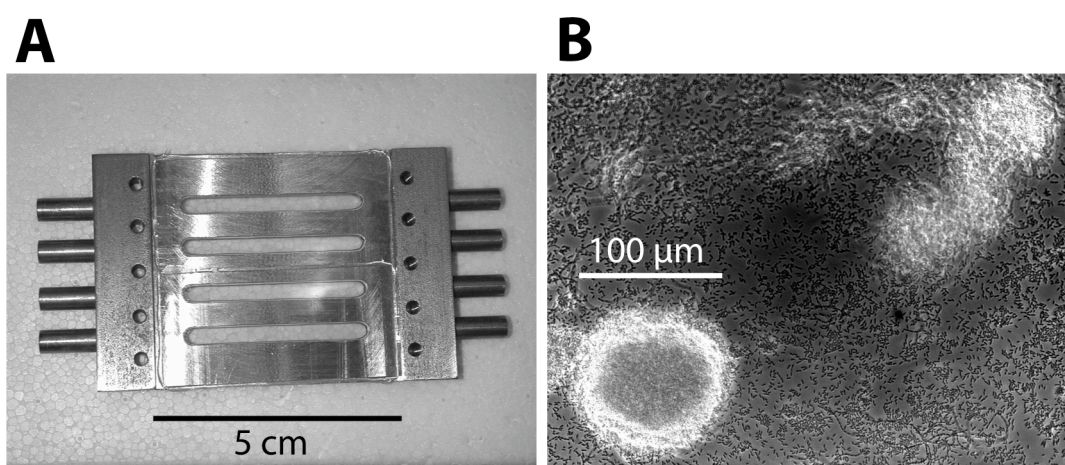
Figure 1. Example of a biofilm flow microcolony growth setup. (A) A four-channel flow chamber, covered on both sides with cover slips. (B) A heterogeneous biofilm of *Pseudomonas knackmussii* B13 on 10 mM 3-chlorobenzoate developed after 5 days flow, imaged at 40 × phase contrast.

Microcolony assays have also been carried out in spatially confined microfluidics structures, in which cells are provided with nutrients by diffusion from an agar surface below, but are physically confined to a microstructure on top (3). The microstructure forces the cells to grow within the spatial boundaries defined by the structure, for example, a capillary. In this case cells will divide and arrange themselves in longitudinal direction, which can be used to examine cell arrest, mutant arisal and phenotypic variation, such as upon antibiotic addition (3).

Finally, microcolonies can be grown on the surface of a semi-solidified gel matrix that provides nutrients to the cells, but is so thin that light passes through and cell division can be followed directly by microscopy. In this procedure, cell division is physically constrained by the presence of the gel surface on one and a glass cover slip on the other side. Even so, in our experience two-to three layered microcolonies will develop eventually, that are not hindered in the *z*-direction by the cover slip. Examples of this procedure include recent studies that have used methods of single cell tracking (2, 14, 16). The crux in the procedure is to form a flat gel surface onto which to inoculate the cells; then cover the surface but still allowing oxygen (for aerobic bacteria) to penetrate, and to prevent drying out of the surface and the cells. Glass slides with small concave inclusions and silicon surrounding ridges have hereto been deployed (2, 14).

The procedure we describe in more detail below was inspired on earlier gel matrix assays, but with some modifications. These modifications, we find, easily allow further online manipulation of growth conditions or cell staining during the course of an experiment without interfering with the original structure of the microcolonies and the positions of the individual cells therein. Our procedure also allows long-term (up to one week) online-monitoring of the cells in the microcolony, without desiccation or oxygen depletion throughout the length of the experiment. The protocol describes the creation of a flat and thin nutrient surface on which cells can be inoculated and are constrained by a glass cover slip, but with the difference that the other side of the gel matrix is exposed to a headspace (Fig. 2). The system is then placed within an air tight microscope observation chamber for incubation and time-lapse imaging. We termed this protocol the *pancake* method because it involves the turning of a gel patch, similar to the flipping of a pancake (see Fig. 2).

Since the gel patch is inoculated with a diluted cell suspension, flipping of the gel patch results in the entrapment of isolated single cells at the glass-gel interface (Fig. 2A). Nutrients reach the cells by diffusion from the gel, whereas the gel layer is so thin (1 mm) that oxygen can diffuse freely to the cell layer that forms in between the gel and the glass (Fig. 2B). Further substrate can be added on top of the gel without interfering with the original positions of the cells or developing microcolonies on the other side. Substrate addition can be done either in

form of a liquid solution, or as a volatile released into the headspace within the closed chamber (in the case of, e.g., naphthalene or toluene). Additional advantage is that the cells can be specifically stained after defined incubation periods, by applying the dye again on top of the gel surface, after it will diffuse through the gel and stain the cells. Examples of useful stains include the Live/Dead reagent (Invitrogen, Molecular Probes).

Not unsurprisingly, cells grow into microcolonies even without extra added carbon source to the gel, because of carbon traces in the gel that cannot be completely removed. Final average colony sizes even without any carbon added reach up to a few hundred cells, but this size is inversely proportional to the density of the inoculated cell suspension. The concentration of cells in the inoculum can thus easily be adjusted to obtain monolayered microcolonies (Fig. 3), which is an important prerequisite when single cells are to be investigated with microscopes that cannot resolve along the z-axis (like confocal laser scanning microscopy, CSLM), or in experiments with prolonged incubation periods. In our hands, we find that at above 0.1 mM extra carbon added to the gel, colony sizes will start to increase as a result of the specific extra carbon source.

We generally assemble the whole setup in a commercially available *perfusion chamber system* from H. Sauer (Germany, see below, Fig. 2), which comes with a rigid metal cast POC chamber, round coverslips and silicon or Teflon spacer rings. The POC chamber allows both closed and open configurations, and several add-on options for nutrient-flow circuits exist. But in theory any airtight system (to prevent desiccation of the gel patch) consisting of two glass cover slip layers and a spacer-ring(s) should work as long as it fits into the microscope table.

We used the pancake method to full satisfaction in combination with three different microscopes with different set-ups, a Zeiss Axioscope upright epifluorescence microscope, a Leica inverted epifluorescence microscope, and an inverted CLSM from Leica. CLSM has the obvious advantage that multi-layered cell structures can be focused more properly, but has the disadvantage that quantification of autofluorescent protein signals is more difficult (in case this is important). In the upright microscope-POC chamber configuration, the distance of the chamber to the condensor is too large to enable proper phase-contrast imaging, but the epifluorescence imaging mode is unrestricted.
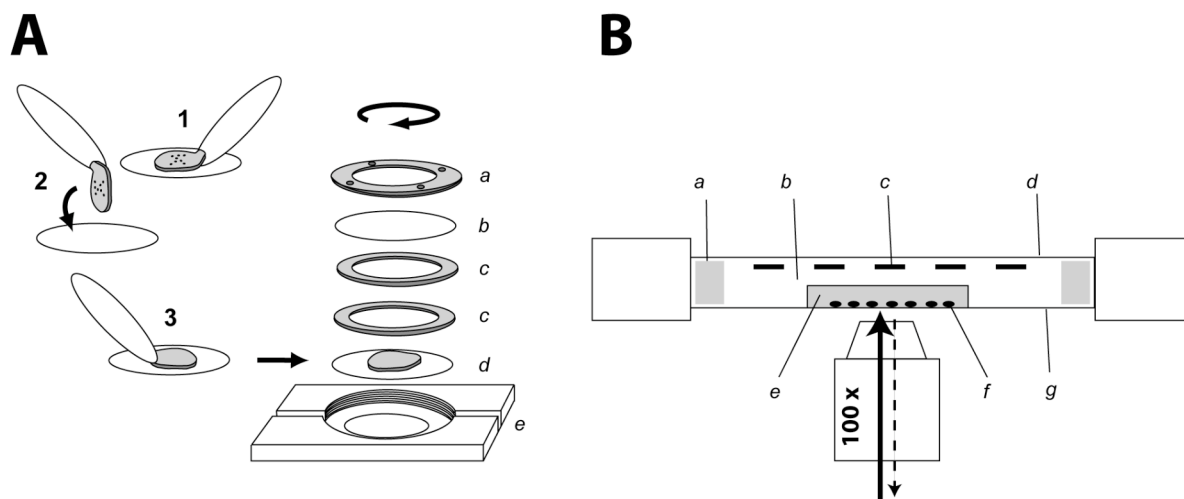
Figure 2. Schematic outline of the pancake method and perfusion incubation chamber assembly. (A) POC assembly parts and gel patch flipping. a: locking ring, b: upper coverslip, c: silicon ring, d: lower coverslip, cells and gel patch, e: perfusion chamber. (B) Sideview of POC chamber in inverted microscope configuration. a: silicon ring, b: headspace, c: condensation, d: upper coverslip, e: gel patch, f: cells, g: lower coverslip, h: objective.
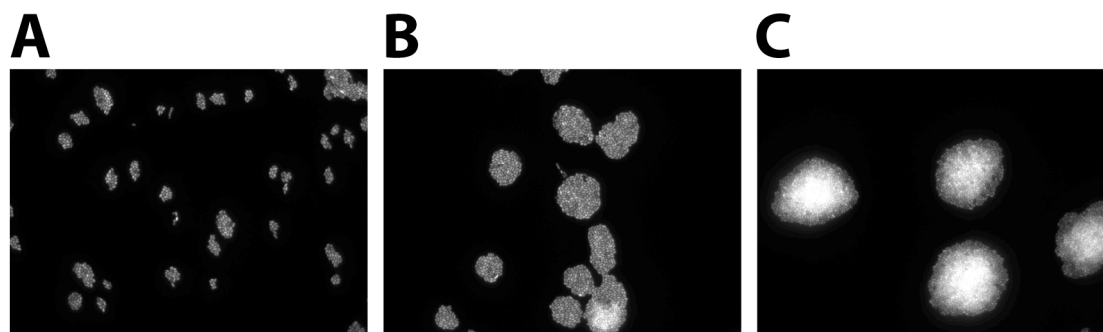
Figure 3. Microcolony sizes of *Pseudomonas putida* UWC1-ICE*clc* equipped with a constitutively expressed mCherry protein after 144 h incubation at 20°C on agarose surfaces supplemented with 0.1 mM 3-chlorobenzoate in the pancake method. Initial inoculation densities: $\approx 10^4$ cells per µl (A), ≈ 1000 cells per µl (B), ≈ 100 cells per µl (C). Images taken in mCherry epifluorescence and upright mode with a Zeiss Axioscope at 1000 × magnification.

ADVANTAGES AND DISADVANTAGES

We have experienced the following advantages of our modified procedure to prepare gel surface growth in a closed microscope chamber. First, because of the system containment and upside-down configuration, the gel surface is less prone to shrinking in the z direction because of desiccation. We found that in the 'cells-on-top' configuration, an agarose surface (even in a completely closed chamber with moisture) shrinks with about 1-2 µm per day, which makes automated imaging more difficult except if z-stacks are recorded. In the upside-down configuration, any shrinking of the agar surface does not influence the automated focusing of cells directly close to the glass cover slip (Fig. 2B). Secondly, the use of the microscope chamber with round cover slips and upside-down configuration permits the proper functioning of even short working distance lenses, which are basically limited to the thickness of the cover slip (0.17 mm) at high magnification. Thirdly, in upside-down configuration on an inverse microscope, the fluorescence light path does not travel through the agarose nor is it obstructed by condensation on the upper glass cover. Turning the POC chamber around for observation on an upright microscope has the small risk that the gel patch will fall down or loosen because of gravity. Condensation cannot form between the agarose patch and bottom coverslip. Sequential carbon substrates or staining dyes can be applied on the gel surface without affecting or disrupting the positions of single cells or the microcolony structure. In combination with CLSM multilayered colonies can be explored. The method is suitable for single-cell tracking, time-lapse imaging, and single-cell fluorescence imaging over periods of one week. Finally, the microscope chambers can be positioned exactly on the microscope stage, but also removed and incubated separately, if multiple such systems need to be analysed and imaged.

The pancake method does have some drawbacks. Turing around the gel patch requires a bit of practise. As an alternative the second cover slip can be placed directly on the inoculated cell surface, and then turned around, but we find that in that case removing the support cover slip for the gel (which is now on the top side) is not unproblematic and can either break or will again lift the whole patch from the lower glass cover. Condensation forms on the inside of the upper coverslip obstructing somewhat bright-field and phase contrast views, both of which rely on light passing through the whole chamber from the opposing light source into the lens. Another disadvantage of the gel patch protocol might be accumulation of (toxic) metabolites and change of pH over time in the gel-surface provided to the cells.

181

PROCEDURE

Protocol: Preparation of the gel patch. Prepare the POC Chamber and the cover slips by autoclaving and subsequent drying in a sterile flow hood. Separate the cover slips during autoclaving by inserting a tooth pick between them. Install a round cover slip (42 mm diameter, 0.17 mm thickness) at the bottom of the POC chamber and place a 1 mm Silicon ring on top (Fig. 2). For optimal sterile conditions carry out the whole procedure in a laminar air flow bench.

1. Place a drop of 150 μl molten liquid 1% agarose medium maintained at 50°C on the bottom glass cover slip.
2. Immediately place another round cover slip on top of the agarose medium droplet. Press softly until the droplet is flat and the top cover slip touches the Silicon separation ring. Let the agarose solidify.
3. Remove the upper glass slide and the Silicon ring to expose the agarose surface.
4. Prepare the bacterial suspension to inoculate on the surface. The cell density in this suspension may depend on the research question, but we generally start with a suspension that has $\approx 5 \cdot 10^6$ cells per ml. This corresponds approximately to a 100-fold dilution of an overnight culture on 4 mM fructose. Collect the cells by centrifugation at low speed and wash once in carbon free minimal medium to remove traces of organic carbon or possible toxic secondary metabolites. Centrifuge again to collect cells and resuspend in the medium to the final desired cell density.
5. Pipette a volume of 10 μl of cell suspension on the planar agarose surface, distribute the droplet evenly by tilting in a circular motion and allow it to dry at room temperature in the sterile flow hood. This typically takes five to ten minutes. Stop when a water film is no longer visible. Avoid drying too long or else the agarose will dry out.
6. Now take another coverslip to lift the agarose patch and place it upside down onto a new round cover slip for the POC chamber. This requires some practise but is most easily performed by slightly inserting the edge of a cover slip at an angle between the agarose and the support glass surface until the agarose makes enough contact with the new coverslip to be lifted (Fig. 2A: steps 1-3). The cells will now be facing down.

Note: As an alternative, you could add a new cover to the side of the gel patch with the cells, turn the system around and remove the back support cover slip. We find, however, that this is more difficult because the back cover often sticks better to the agarose.

7. Insert the new support coverslip with the gel patch into the POC chamber. Now place two Silicon rings (each of 1 mm thickness), finally a fresh round glass cover and close the system with the metal ring.

Note: Instead of placing the gel patch in a closed configuration, the POC chamber can also be operated in open mode, which makes subsequent manipulation of substrates or dyes more easy. The open POC-mode requires, however, an inverted microscope set-up.

8. The chamber is now ready to be mounted on a microscope. Growth temperature can be adjusted to ambient temperature of the room in which the microscope is operated, or set to a specific temperature if an incubation chamber on the microscope table is used.

Note: As an alternative, the POC chamber can be removed from the microscope table for external incubation in a small temperature incubator. In this case, one cannot make time-lapse imaging and finding back the exact position may be more difficult.

9. It may be fairly difficult at first to spot and focus single cells at the agarose-cover slip interface. It is helpful to find the edge of the gel patch and use this area to focus, because often a higher concentration of bacteria can be found here. From there on move towards the centre of the patch and focus on an isolated single cell. To make sure to focus on a growing individual cell, wait until the first cell division before setting up the automated time-lapse function.

10. For time-lapse microscopy take images at regular time intervals. The shorter the time interval the more accurate every cell can be followed. Cells will start to divide synchronously but after a few divisions will lose this synchronicity. Typically we start with images taken every 10-15 minutes. Manual readjustments may be needed to remain in focus over longer periods. Depending on programming abilities and automated *x-y* stages, multiple positions can be monitored over time.

11. We use Metamorph or Metaview (Molecular Devices) to produce image stacks and track cells, but any imaging program can be of help here. ImageJ is a public domain software for image analysis (rsb.info.nih.gov/ij/).

Optional protocol: Cell staining and new carbon substrates. We find that microcolonies grown to stationary phase on gel patches can be reactivated by the addition of

new substrate. For this purpose, open of the POC chamber and pipette a 10 µl droplet of a 100 mM carbon substrate solution on the top of the gel patch. Within 15 minutes the substrate has reached the cells by diffusion.

In a similar manner, dyes can be added by opening the POC chamber and applying a 10 µl drop of the dye solution (e.g., Live-Dead® stain from Invitrogen) on the top of gel patch. Dyes will diffuse through the cell and stain the cells after ≈ 15 minutes. Staining is complete after 1-2 hours. Do not add more than 10 µl at once, or else microcolonies can disassemble at the glass-gel interface.

MATERIALS

Chemicals and solutions. Growth medium of choice for your favourite bacterium. For *Pseudomonas*, *Burkholderia* and *Cupriavidus* we use Pseudomonas defined medium (Gerhard et *al*., 1981). The basic medium contains per litre: 1 g NH_4Cl, 3.49 g $Na_2HPO_4.2H_2O$, 2.77 g KH_2PO_4, has a pH of 6.8 and is sterilized by autoclaving. Carbon substrates are added according to the experimental setup. Afterwards per L of medium 20 ml of Hutner's vitamin free mineral base (filter sterilized), and 2 ml of a 500-fold diluted filter-sterile vitamin solution are added. Hutner's vitamin free mineral base contains per litre: 10 g NTA, dissolve and neutralize with ca. 6 g KOH, 14.45 g $MgSO_4 \cdot 7H_2O$, 3.33 g $CaCl_2 \cdot 2H_2O$, 9.74 mg $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, 99 mg $FeSO_4 \cdot 7H_2O$, and 50 ml of Metals 44 solution. Hutners is sterilized by filtration and stored 4°C. Metals 44 solution contains per 100 ml: 387 mg $Na_4EDTA \cdot 4H_2O$, 1.095 g $ZnSO_4 \cdot 7H_2O$, 914 mg $FeSO_4 \cdot 7H_2O$, 154 mg $MnSO_4 \cdot H_2O$, 39.2 mg $CuSO_4 \cdot 5H_2O$, 24.8 mg $Co(NO_3)_2 \cdot 6H_2O$, 17.7 mg $Na_2B_4O_7 \cdot 10H_2O$, a few drops of 6 N H_2SO_4, is sterilized by filtration and stored at 4°C. Antibiotics can be added upon desire and according to the characteristics of the bacterial strain used.

Agarose medium solution is prepared by dissolving 1.0% (*w/v*) ultra-pure agarose into the selected growth medium. Autoclave the growth medium with the agarose, cool down to 50°C before pouring the gel patch.

POC chamber. We use a POC set-up (Helmut Saur Laborbedarf, D-72734 Reutlingen, Germany). Silicon rings come with the system. Teflon® rings can be custom manufactured to resist organic volatile contaminants if used as growth substrate. The system can be operated with round coverslips, 42 or 25 mm ø (Helmut Saur).

Microscope. Epifluorescence microscope (upright or inverse) with appropriate filters in case fluorescent markers are used. Confocal laser scanning microscope for analysis of multiple cell layers. Magnifications: 63 × or 100 × 10. Preferentially use objectives with long working distances.

185

DATA EXAMPLES

Example 1. Quantification of microcolony GFP expression. In this example we used a *Pseudomonas putida* strain constitutively expressing mCherry from the *tac* promoter, and *egfp* from the integrase promoter *intB13p* (12). Microcolonies were grown on gel patches with 0.1 mM 3-chlorobenzoate to study induction of the P_{int} promoter in stationary phase. These conditions are known to result in bifurcation of the population and leading to around 2-5% of cells expressing P_{int} (13). Microcolonies were grown from single cells in the pancake protocol for up to one week. After 24 hours cells in the colony stopped dividing and the cells entered stationary phase. The overlay image (Fig. 4) shows *egfp* expression in a number of individual cells in a microcolony of cells that all express mCherry at 48 hours after start of the experiment.

Example 2. Cell pole tracking using time-lapse fluorescence microscopy. Cell pole tracking is a technique that is used to study bacterial cell aging effects. In this example we used *Cupriavidus necator* strain JMP134, a strain which degrades 2,4-dichlorophenoxyacetic acid (10), but which was grown here on diluted nutrient broth agarose gel patches. Microcolony development was followed here by phase-contrast microscopy with images taken every 10 minutes. The three images show manual tracking of young cell poles appearing in the first, second and third generation of cell division over the time course of seven generations (8.71 h, Fig. 5).
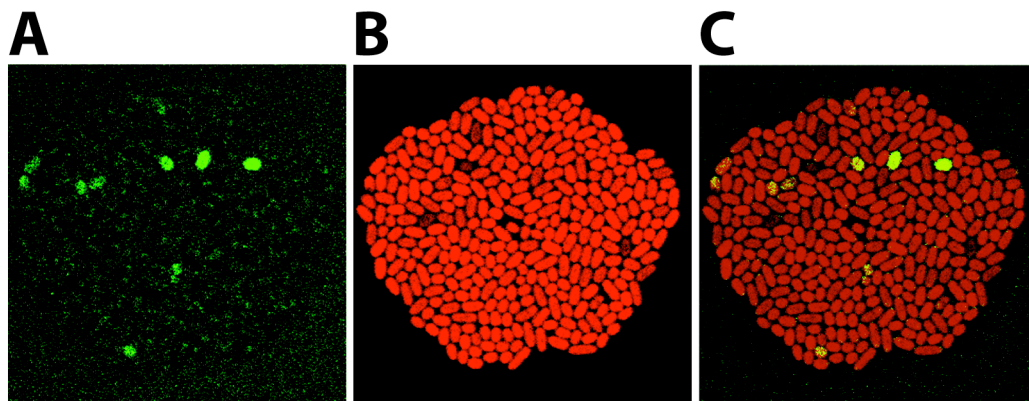
Figure 4. Example of a microcolony of *Pseudomonas putida* UWC1-IC*Eclc* constitutively expressing mCherry and inducing egfp from *intB13p*, formed on agarose surfaces with 0.1 mM 3-chlorobenzoate after 48 h. (A) Confocal laser scanning image in excitation and emission wavelengths for egfp at 1000 × magnification. (B) Idem for mCherry. (C) Overlay of A and B. Note that only a small proportion of cells express egfp from *intB13p*.
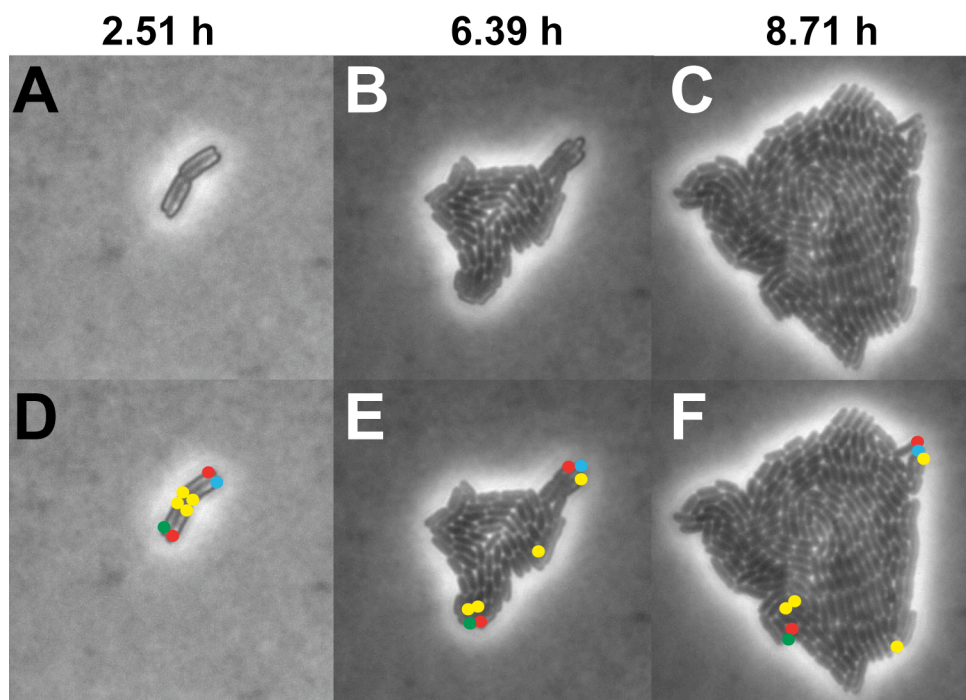
Figure 5. Cell pole tracking in individuals within a developing microcolony of *Cupriavidus necator* on diluted agarose medium in the pancake method. (A), Four-cell stage after 2.51 h. (B) and (C) Multiple cell stages after 6.39 and 8.71 h, respectively. (D) to (F), Corresponding marker overlays to track poles. Ancestral cell poles: green and blue. Second generation new cell poles: red. Third generation cell poles, yellow.

OUTLOOK AND RESEARCH NEEDS

There is a clear new interest in single cell microbiology, phenotypic variation and its underlying causes. As outlined above this requires sets of methods to visualize and record the fates of individual bacteria during growth. Although a number of simple methods to grow and follow individual bacterial growth exist currently, many more will be deployed by combining microstructuring and –fluidics techniques with bacteria growth. Advances in computer systems and microscopy will doubtless permit next generation data recordings, enabling further dissection and understanding of microbial life and development at the real microscale level.

REFERENCES

1. Ackermann, M., B. Stecher, N. E. Freed, P. Songhet, W. D. Hardt, and M. Doebeli. 2008. Self-destructive cooperation mediated by phenotypic noise. Nature **454:**987-990.

2. Babic, A., A. B. Lindner, M. Vulic, E. J. Stewart, and M. Radman. 2008. Direct visualization of horizontal gene transfer. Science **319:**1533-1536.

3. Balaban, N. Q., J. Merrin, R. Chait, L. Kowalik, and S. Leibler. 2004. Bacterial persistence as a phenotypic switch. Science **305:**1622-1625.

4. Christensen, B. B., C. Sternberg, J. B. Andersen, R. J. Palmer, Jr., A. T. Nielsen, M. Givskov, and S. Molin. 1999. Molecular tools for study of biofilm physiology. Methods Enzymol **310:**20-42.

5. Dhar, N., and J. D. McKinney. 2007. Microbial phenotypic heterogeneity and antibiotic tolerance. Curr Opin Microbiol **10:**30-38.

6. Glaser, P., M. E. Sharpe, B. Raether, M. Perego, K. Ohlsen, and J. Errington. 1997. Dynamic, mitotic-like behavior of a bacterial protein required for accurate chromosome partitioning. Genes Dev **11:**1160-1168.

7. Heydorn, A., A. T. Nielsen, M. Hentzer, C. Sternberg, M. Givskov, B. K. Ersboll, and S. Molin. 2000. Quantification of biofilm structures by the novel computer program COMSTAT. Microbiology **146 (Pt 10):**2395-2407.

8. Ingham, C. J., M. Beerthuyzen, and J. van Hylckama Vlieg. 2008. Population heterogeneity of Lactobacillus plantarum WCFS1 microcolonies in response to and recovery from acid stress. Appl Environ Microbiol **74:**7750-7758.

9. Ingham, C. J., A. Sprenkels, J. Bomer, D. Molenaar, A. van den Berg, J. E. van Hylckama Vlieg, and W. M. de Vos. 2007. The micro-Petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms. Proc. Natl. Acad. Sci. U. S. A. **104:**18217-18222.

10. Laemmli, C. M., J. H. J. Leveau, A. J. B. Zehnder, and J. R. van der Meer. 2000. Characterization of a second *tfd* gene cluster for chlorophenol and chlorocatechol metabolism on plasmid pJP4 in *Ralstonia eutropha* JMP134(pJP4). J. Bacteriol. **182:**4165-4172.

11. Losick, R., and C. Desplan. 2008. Stochasticity and cell fate. Science **320:**65-68.

12.   **Minoia, M., M. Gaillard, F. Reinhard, M. Stojanov, V. Sentchilo, and J. R. van der Meer.** 2008. Stochasticity and bistability in horizontal transfer control of a genomic island in *Pseudomonas*. Proc. Natl. Acad. Sci. U S A **105:**20792-20797.

13.   **Sentchilo, V., R. Ravatn, C. Werlen, A. J. Zehnder, and J. R. van der Meer.** 2003. Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. Strain B13. Journal of Bacteriology **185:**4530-4538.

14.   **Stewart, E. J., R. Madden, G. Paul, and F. Taddei.** 2005. Aging and death in an organism that reproduces by morphologically symmetric division. PLoS Biol **3:**e45.

15.   **Tolker-Nielsen, T., and S. Molin.** 2000. Spatial Organization of Microbial Biofilm Communities. Microb Ecol **40:**75-84.

16.   **Veening, J. W., E. J. Stewart, T. W. Berngruber, F. Taddei, O. P. Kuipers, and L. W. Hamoen.** 2008. Bet-hedging and epigenetic inheritance in bacterial cell development. Proc Natl Acad Sci U S A **105:**4393-4398.

17.   **Wiacek, C., S. Muller, and D. Benndorf.** 2006. A cytomic approach reveals population heterogeneity of Cupriavidus necator in response to harmful phenol concentrations. Proteomics **6:**5983-5994.

CHAPTER 4

Host cell differentiation to primitive 'mating bodies' induced

by an integrating and conjugative element in

free-living bacteria

SUMMARY

Lateral gene transfer (LGT) is one of the most important processes leading to prokaryotic genome innovation (21). LGT is typically associated with conjugative plasmids and bacteriophages, but recently, a new class of mobile DNA known as Integrating and Conjugative Elements (ICE) was discovered, which is abundant and widespread among bacterial genomes (4, 9, 11, 13, 17, 28). ICE are thought to be hybrids of plasmids and bacteriophages, but very little is known about the factors that control vertical and horizontal transmission. By studying at single cell level the behavior of a prevalent ICE-type in the genus *Pseudomonas* as model (9, 18), we uncover for the first time the remarkable and novel way in which the ICE orchestrates specific host cell differentiation to ensure horizontal transmission. We find that the ICE induces a state of transfer-competence in 3-5% of cells in a population under non-growing conditions. Through mutant analysis, transfer experiments and heterologous expression we show how ICE factors control the development of transfer-competent cells into specific assemblies that we name 'mating bodies'. Interestingly, cells in mating bodies undergo fewer and slower division than non-transfer competent cells, and eventually lyse. On the other hand, mutations in ICE genes disrupting mating body formation lead to fivefold decreased ICE transfer rates. Hence, by confining the transfer competent state to a small proportion of the population, ICE horizontal transmission is achieved with little cost in terms of vertical transmission. Given low transfer frequencies of most ICE (1, 14, 22) we anticipate regulation by subpopulation differentiation to be widespread.

RESULTS AND DISCUSSION

ICE*clc* differentiates a small proportion of host cells into transfer competent cells. ICE are normally integrated and stably maintained in the chromosome of their bacterial host. However, in order to spread to other hosts, they must excise by site-specific recombination, conjugate to a new recipient cell and reintegrate at one or more unique insertion sites (2, 3, 15, 22, 24). Previous results using a prevalent model ICE named ICE*clc* (after its propensity to code for a metabolic pathway for c̲h̲l̲oroc̲atechol degradation) in the bacterium *Pseudomonas knackmussii* B13 showed that expression of the ICE*clc* integrase is limited to a small fraction of cells (18, 19, 25), suggesting that only these cells are capable of transmitting the ICE to new recipients.

In order to demonstrate the existence of the presumed dedicated 'transfer competent' (tc-) cells in a population, we followed ICE*clc* transfer at individual cell level (Supplementary Experimental Procedures). As donor for ICE*clc* we used *P. knackmussii* B13, labeled with a single copy fluorescent reporter gene (*egfp*) fusion to the promoter of the *intB13* integrase gene (P_{int}). Intact integrase is essential for ICE*clc* excision and transfer (18, 20), and expression of eGFP from P_{int} would thus indicate the initiation of the ICE horizontal transmission process at single cell level. *P. knackmussii* B13 cells were additionally labeled with a single copy constitutively expressed P_{tac}-*mCherry* fusion to differentiate them from recipient (Table S1). As recipient for ICE*clc* we use *Pseudomonas putida* UWC1 containing a single-copy engineered conditional trap that upon ICE*clc* insertion leads to strong eGFP production (Figure 1A) (24). Time-lapse epifluorescence microscopy of mixtures of donor and recipient cells on agarose nutrient surface showed indeed appearance of brightly fluorescent recipient cells, indicating ICE*clc* transfer (Figure 1B, arrow *a*). Brightly fluorescent recipient cells (n=78 scored events) arose exclusively upon observable physical contact to a donor cell that had expressed P_{int}-*egfp*. No ICE*clc* transfer was observed with donor strains with deletions in the *intB13* gene or in the left attachment site *attL*, which cannot excise ICE*clc*. This showed, therefore, that ICE*clc* can only transfer from specialized cells (tc-cells) that have somehow differentiated from the rest of the population. The tc-state can be detected by expression of reporter protein from an extra copy P_{int} promoter (Figure S1), but transfer likely depends on expression of a large number of other genes on ICE*clc* (8, 18) and signals from the host cell (19). These results thus confirmed previously formulated hypotheses (18) that P_{int}-expressing cells are indeed those and only those that are capable of transferring ICE*clc.*
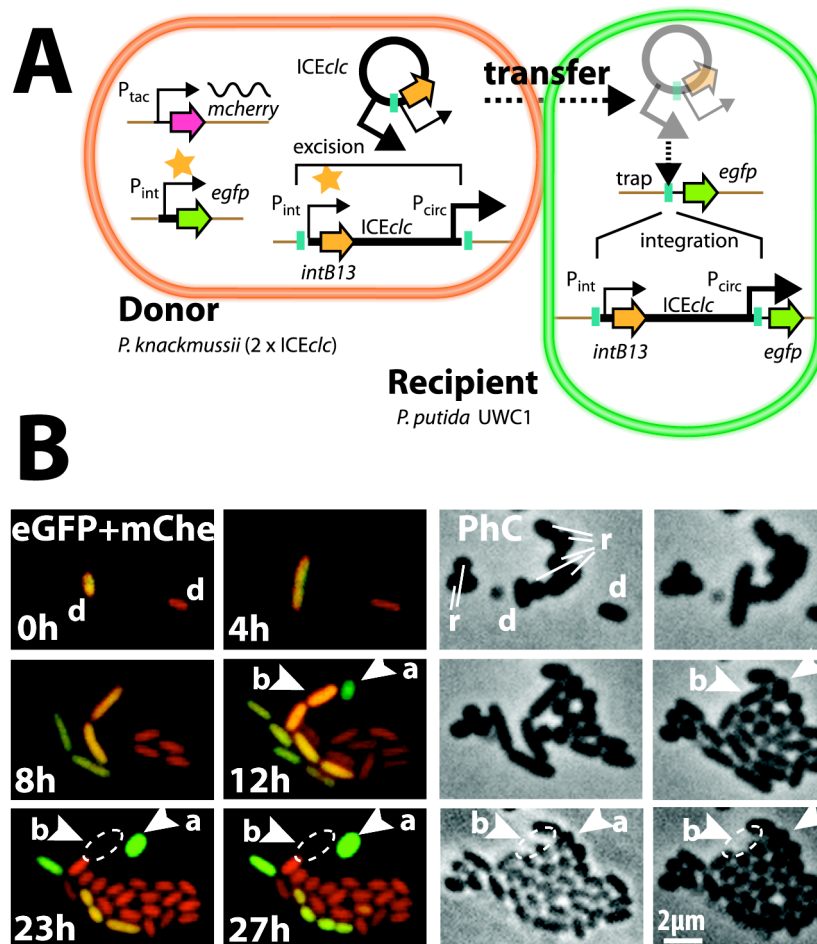
Figure 1. ICE*clc* transfer and competence formation. (**A**) Schematic principle of ICE*clc* transfer and explanation of the single-copy engineered conditional trap in the recipient cell (24). The star at the P_{int}-promoter indicates it being expressed only in transfer competent (tc-) cells. (**B**) Time-lapse epifluorescence (eGFP and mCherry overlay) and phase-contrast (PhC) imaging of ICE*clc* transfer from *P. knackmussii* B13 donors (*d*) to *P. putida* UWC1 (*r*, both labeled as indicated in **a**). Tc-donor cells visible in orange (overlay of eGFP from P_{int} and red from mCherry). Non-tc donor cells fluoresce red only. Note how a brightly eGFP expressing transconjugant of *P. putida* becomes visible from t = 12 h onwards (arrow marked *a*) whereas the corresponding B13 donor cell lyses (arrow marker *b*). See also Figure S1.

Transfer competent donor cells face severe growth impairments. Intriguingly, tc-donor cells displayed different morphology than non tc-donor cells in the time-lapse transfer experiment (Figure 1B). To test this observation further, *P. knackmussii* B13 cells were collected from a liquid culture under stationary phase conditions and deposited on the surface of a nutrient agarose. Tc-cells in this case consistently formed smaller microcolonies than cells that showed no expression of the P_{int}-*egfp* marker (Figure 2A, B). Tc-cells showed aberrant cell shapes, divided more slowly, accumulated more biochemical damage (i.e., reactive oxygen species and membrane damage, Figure S2A, B), and lysed more frequently than cells in microcolonies from non-tc cells (Figure 2A, B). In fact, up to 50% of tc-cells from *P. knackmussii* B13 placed on nutrient surface were unable to divide at all (Figure 2B). As expected, a small proportion of cells in microcolonies from non-tc starter cells developed transfer competence once the colony entered stationary phase, as indicated by their eGFP fluorescence from P_{int} (Figure 2B, 50 h). Also such newly formed tc-cells within microcolonies displayed slower division and higher lysis probability compared to neighbouring non-tc cells when presented with fresh nutrients (Figure 2C). Similar formation of small microcolonies from tc-cells and their growth inhibition was detected in three other species, into which a single ICE*clc* copy was inserted via conjugation, notably *P. putida* UWC1 and two *Pseudomonas aeruginosa* strains (Figure 2B). Hence, specific small microcolony formation from tc-cells and their reproductive inhibition is not host-dependent, is not an artifact of the labeling techniques (Supplementary Experimental Procedures), and occurs independently of the presence of recipient cells without ICE*clc*. Collectively, these experiments demonstrated that ICE*clc* tc-cells arise species-independently and reproducibly at low frequency (~10^{-2} per cell) under non-growing conditions (stationary phase, Figure 2A). Tc-cells can divide to form a small microcolony when new growth substrate becomes available, but their cell division is specifically and severely inhibited, and tc-cells eventually lyse (Figure 2A, B). Reversion of the tc-state was not detected in time-lapse experiments (n~100 tc-starter cells), but may occur at lower frequencies given that ICE*clc* excision and reintegration in different chromosomal attachment sites can be detected by PCR in suspended cell cultures (24).
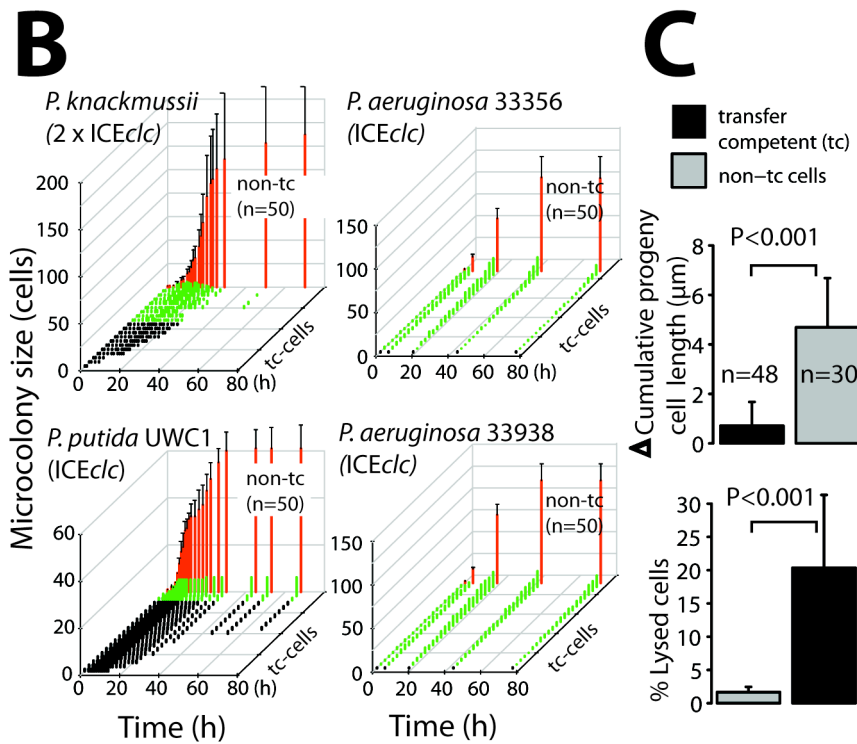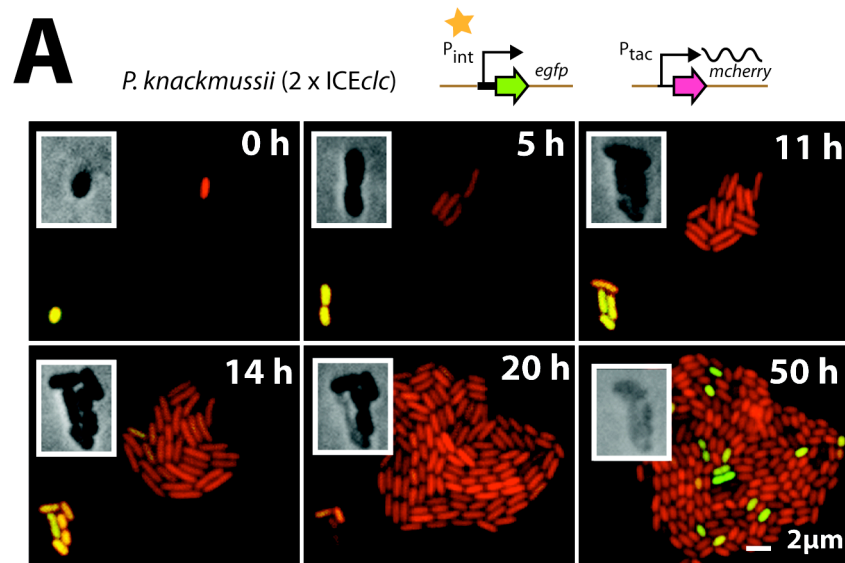
Figure 2 (on the left). Formation of specific transfer colony morphotypes of ICE*clc* tc-cells. (A) Limited division of a tc-cell (orange-green cell at time 0) into a transfer colony morphotype (TCM) compared a regular microcolony formed from a non-tc cell of *P. knackmussii* B13 (labeled as indicated on top image). Inset shows corresponding phase-contrast micrographs of the TCM, highlighting decay and cell lysis after 20 h. Note how a small proportion of cells in regular microcolony randomly again develops transfer competence at t = 50 h in stationary phase (green cells). **(B)** Quantification of TCM sizes achieved from ICE*clc* tc- compared to non-tc cells taken from stationary phase suspended culture and deposited on nutrient surface. Green bars: tc-cells dividing more than once. Black bars: no cell division. Red bars, average microcolony growth from non-tc cells. Disappearing lineages indicate cell lysis. ICE*clc* donors: *P. knackmussii* B13 (2 copies), *P. putida* UWC1, *P. aeruginosa* ATCC33356 and *P. aeruginosa* ATCC33938 (all one ICE*clc* copy)(7). **(C)** Cumulative length of progeny of newly formed ICE*clc* tc-cells of *P. knackmussii* B13 in regular microcolonies compared to neighbours during a 6.8 h period, after renewed addition of 4 mM glucose (70 h). Mean percentage of lysis among tc- versus non-tc B13 cells across 22 examined microcolonies. Average size: 175 ± 78 cells, average proportion of 13 ± 6 tc-cells, during the time window of 51-75 h in stationary phase and after addition of glucose at time 70 h. Error bars denote s.d. See also Figure S2.

ICE*clc* controls process of tc-cell microcolony formation. In order to demonstrate that ICE*clc* controls the development of tc-cell microcolonies (TCM) we measured the proportion of TCM arising from tc-cells in strains with intact or mutant ICE*clc*. Since *P. knackmussii* B13 carries two ICE*clc* copies we used hereto *P. putida* UWC1 with a single integrated ICE*clc* (Figure 3). The proportion of TCM developing from stationary phase *P. putida* UWC1 (ICE*clc*) cells placed on nutrient surface, scored by epifluorescence microscopy after 120 h incubation was almost 0.7% (Figure 3C). This proportion is slightly lower than the proportion of tc-cells identified by P_{int}-*egfp* expression, because some of those do not develop into TCM (Figure 2B, table S2). In contrast, *P. putida* UWC1 without ICE*clc*, but with a mini-Tn*7*-delivered single copy of the *clc*-genes, enabling growth on the same nutrient surface (Supplementary Experimental Procedures) produced only 0.02% TCM (P<0.05, Figure 3C). *P. putida* UWC1 carrying a mutant ICE*clc* with a transposon insertion in *orf18502*, causing silencing of expression of ICE*clc* transfer genes in a region strongly conserved among a wide variety of ICE (8, 18) (Figure 3A), produced a proportion of TCM non-significantly different from strain UWC1 without ICE*clc* (0.06%, P>0.05, Figure 3C). In contrast, deletions in *intB13* or in *attL* that prevent ICE excision showed TCM proportions similar to wild-type ICE*clc* (P>0.05, Figure 3C, Figure S2C). This indicated that development of tc-cells into TCM is due to the presence and expression of ICE*clc* genes. It also indicates that slower growth of tc-cells is not due to ICE*clc* excision or any other activity of the IntB13 integrase, as a result of which cells

could no longer metabolize growth medium components (Supplementary Experimental Procedures).

Formation of tc-cell microcolonies is intimately linked to ICE*clc* transfer. To identify which ICE*clc* factors are responsible for the cell growth inhibition and TCM formation, we cloned ICE*clc* gene combinations under control of the LacI^q/P$_{lac}$ system on a plasmid in *P. putida* UWC1 without ICE*clc* (Supplementary Experimental Procedures, Table S1). Of various combinations a particular 6-kb DNA fragment from within the conserved ICE*clc* region (18) invoked strong growth inhibition of *P. putida* UWC1 cultures (Figure 3A). Subsequent subcloning of individual and combinations of gene fragments under LacI^q/P$_{lac}$ expression control indicated that growth inhibition was due to the presence of a small open reading frame on ICE*clc* which we named *shi* (Japanese for *dead*) (Figure 3A, B). Expression of *shi* was necessary for culture growth inhibition, but in addition required the 5'-region of an upstream gene with partial homology to *parA* (Figure 3A, B). Microscopic observations showed that population growth inhibition was due to formation of slow growing cells with abnormal long morphologies. *P. putida* UWC1 cells in which the *parA-shi* region was expressed produced small microcolonies on nutrient surface (Figure S3A) with cell morphologies and cellular damage similar to what was observed for tc-cells within TCM for both *P. knackmussii* and *P. putida* UWC1 with ICE*clc* (Figure 3C). This suggested that expression of the *parA-shi* region in a host without ICE*clc* reproduced at least partly the development of tc-cells into TCM.

Tailored deletion of either *parA*^{ICEclc} or *shi* on ICE*clc* in *P. putida* UWC1 indeed significantly reduced the proportion of TCM compared to wild-type (P<0.01, Figure 3D). Interestingly, also ICE*clc* transfer rates from *P. putida* UWC1 with the *parA*^{ICEclc} or *shi* deletion were up to fivefold lower than of the wild-type (P<0.01, Figure 3E). In contrast, *P. putida* UWC1-ICE*clc-Dshi* produced the same proportion of P$_{int}$-*egfp* expressing cells in stationary phase as wild-type ICE*clc* in *P. putida* (3.6±1.0 vs 3.8±0.5, p=0.94). This indicates that tc-cells in *P. putida* UWC1-ICE*clc-Dshi* indeed arise, but cannot develop into TCM. Since we could score TCM only when consisting of at least two cells, this would explain the observed reduced proportion of TCM in this strain (Figure 3D). The formation of TCM is thus intimately linked to more effective ICE*clc* transfer.
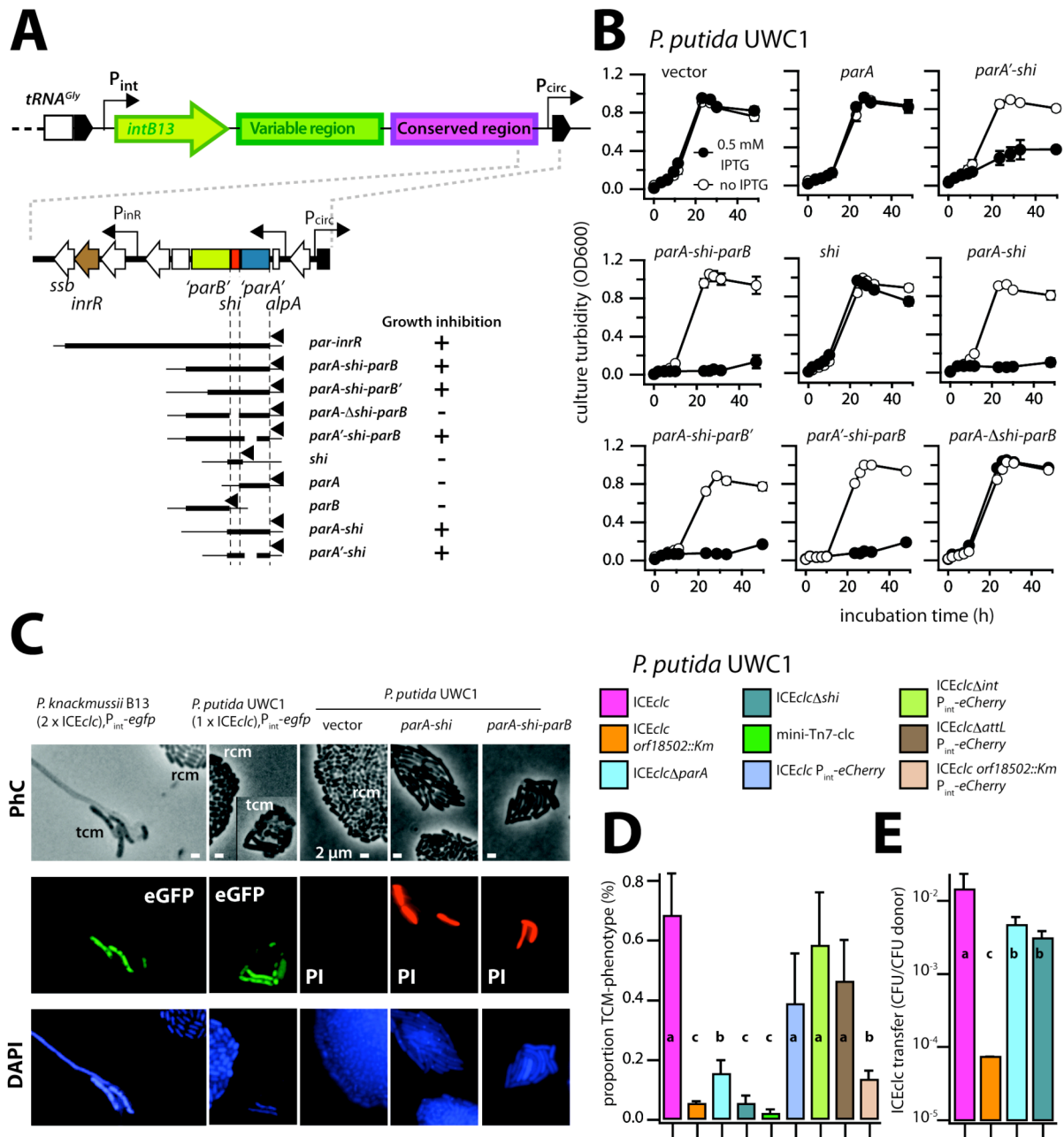
Figure 3. ICE*clc* dependency of TCM formation in *P. putida* UWC1. (**A**) Indication of the relevant gene region on ICE*clc* implicated in tc-cell microcolony (TCM) formation. Fragments introduced by plasmid into *P. putida* UWC1 (without ICE*clc*) are depicted. (**B**) Effect on population growth of different cloned ICE*clc* fragments after IPTG induction (black symbols) from the heterologous P$_{lac}$-promoter (triangles in panel **a**). (**C**) Small microcolony formation and cellular differentiation induced in *P. putida* UWC1 without ICE*clc* but with the *parA-shi-parB*^{ICEclc} locus (20 h after IPTG induction), compared to TCM of *P. knackmussii* B13 (P$_{int}$-*egfp*) and *P. putida* UWC1 (ICE*clc*, P$_{int}$-*egfp*). rcm, microcolony from non-tc cells. PI, propidium iodide. DAPI, 4',6'-diamino-2-phenylindole. (**D**) Proportions of TCM (Figure 2A) per total number of microcolonies scored after 120 h on agarose surface (stationary phase) with 0.1 mM 3-chlorobenzoate. (**E**) ICE*clc* transfer frequencies in donor-recipient filter matings. Letters indicate significance groups in ANOVA followed by Tukey's post-hoc testing (P<0.05). See also Figure S3.

ICE*clc* control is unlike plasmid toxin-antitoxin systems. The implication of ICE*clc* encoded factors in inhibition of proliferation of tc-cells is reminiscent of but distinctly different from plasmid and other ICE-encoded toxin-antitoxin systems that inhibit plasmid- or ICE-free daughter cells to divide (6, 10, 27). First of all, the *parA*[ICE*clc*] and *shi* gene products do not display any significant similarity to any of the known seven classes of TA-systems from plasmids, ICE and chromosomes (6). The *shi* open reading frame encodes a small hypothetical protein (85 amino acids) which is highly conserved among ICE similar to ICE*clc*. Very low but discernable homology of Shi is found to a eukaryotic voltage-gated calcium channel domain suggesting it may influence bacterial membrane potential (Figure S3B). Secondly, the process of cell division inhibition within TCM is not the result of ICE*clc* loss or ICE excision. To demonstrate this, we measured the eGFP reporter signal intensity from P_{int} in individual tc-cells and their offspring within TCM. Since P_{int}-expression is completely silent in absence of ICE*clc,* daughter cells that would have lost ICE*clc* would no longer produce new eGFP and their fluorescence would diminish by 50% after every cell division (23). Contrary to this, eGFP fluorescence did not diminish in tc-cell lineages of *P. knackmussii* B13 (n=8 TCM; Figure 4), nor in *P. putida* UWC1 (ICE*clc*) without or with deletions in *intB13* or *attL* (Figure S4). Inhibition of tc-cell division is thus not a consequence of loss of ICE*clc* and is therefore unlike classical toxin-antitoxin systems.

TCM formation, a tradeoff between horizontal and vertical transmission. In conclusion, our results show that ICE*clc* transfer occurs exclusively from tc-cells, the development of which is initiated in non-growth stationary phase conditions. The phenomenon of transfer competence is so far only known from *Enterococcus* spp. where plasmid-free recipients induce plasmid conjugal transfer from potential donors by peptide pheromones (5, 26), without actually showing cellular differentiation of conjugating cells. Subpopulation competence development also occurs in early stationary phase *B. subtilis* cells, but this involves the process of DNA uptake (natural transformation) rather than conjugation (16). In contrast, ICE*clc* differentiated tc-cells can divide into specific microcolonies (TCM) when cells are presented with new nutrients, although many tc-cells do not divide at all and eventually lyse within TCM (Figure 2A). TCM formation is controlled by ICE*clc* factors *shi* and *parA*, and tc-cells within TCM show distinctly different morphologies than cells in regular microcolonies in which ICE*clc* is not active. Because mutants in which transfer competence is initiated but which do not form TCM display fivefold lower transfer rates (deletion of *shi* on ICE*clc*, Figure 3E), we conclude that the formation of TCM provides a selective advantage to the success of ICE*clc* horizontal transmission. We hypothesize that even though transfer
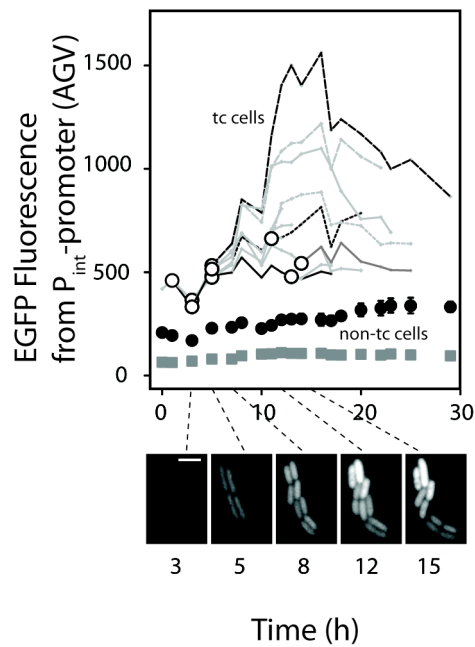
Figure 4. eGFP fluorescence from the P$_{Int}$ promoter in tc-cells of *P. knackmussii* B13 and siblings. Black circles, mean eGFP fluorescence in non-tc cells (n=500 at each time point). Grey squares, average background fluorescence of the epifluorescence image. Open circles, cell division events. See also Figure S4.

201

competence is initiated in stationary phase such cells are nutrient starved and tc-cells need nutrients to complete the transfer process. In addition, the few divisions that tc-cells undergo to form a TCM may serve to increase the chance to contact a suitable recipient cell on a surface (Figure 1B, Figure S3C). TCM may in that perspective be regarded as primitive 'mating bodies'. We show that ICE*clc* horizontal transmission leads to sacrifice of tc-donor cells, but by confining the transfer competent state to a small proportion of cells (3-5%) fitness loss at the population level becomes undetectable (table S3) (7). Slower growth and lysis of tc-donor cells may be a consequence of specific high demand for resources to produce the conjugative system or of excessive oxidative damage (Figure S2B). Lysis will also locally release nutrients from which transconjugants might profit. ICE*clc* horizontal/vertical transmission tradeoff is thus maintained at the proportion of cells which enter transfer competence (Figure S3D). Such subpopulation control may also explain the system of so-called 'fertility inhibition' that conjugative plasmids use to repress transfer and avoid strong fitness cost to the host (12). Since ICE*clc* relatives are widespread in proteobacterial genomes and have a very conserved core region including the *parA-shi* genes (13, 18), this suggests that they employ a similar mechanism of horizontal transmission. The mechanism displayed by ICE*clc* presents a new concept on how mobile DNA can control host cell behavior and differentiation in order to ensure horizontal transmission.

EXPERIMENTAL PROCEDURES

See supporting information.

ACKNOWLEDGEMENTS

REFERENCES

1. **Babic, A., M. B. Berkmen, C. A. Lee, and A. D. Grossman.** 2011. Efficient gene transfer in bacterial cell chains. MBio **2:**mBio.00027-11.

2. **Beaber, J. W., B. Hochhut, and M. K. Waldor.** 2004. SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature **427:**72-74.

3. **Bose, B., J. M. Auchtung, C. A. Lee, and A. D. Grossman.** 2008. A conserved anti-repressor controls horizontal gene transfer by proteolysis. Mol Microbiol **70:**570-582.

4. **Burrus, V., and M. K. Waldor.** 2004. Shaping bacterial genomes with integrative and conjugative elements. Res. Microbiol. **155:**376-386.

5. **Chatterjee, A., C. M. Johnson, C.-C. Shu, Y. N. Kaznessis, D. Ramkrishna, G. M. Dunny, and W.-S. Hua.** 2011. Convergent transcription confers a bistable switch in Enterococcus faecalis conjugation. Proc Natl Acad Sci U S A **108:**9721–9726.

6. **Fozo, E. M., K. S. Makarova, S. A. Shabalina, N. Yutin, E. V. Koonin, and G. Storz.** 2010. Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. Nucleic Acids Res **38:**3743-3759.

7. **Gaillard, M., N. Pernet, C. Vogne, O. Hagenbüchle, and J. R. van der Meer.** 2008. Host and invader impact of transfer of the *clc* genomic island into *Pseudomonas aeruginosa* PAO1. Proc Natl Acad Sci U S A **105:**7058-7063.

8. **Gaillard, M., N. Pradervand, M. Minoia, V. Sentchilo, D. R. Johnson, and J. R. van der Meer.** 2010. Transcriptome analysis of the mobile genome ICE*clc* in *Pseudomonas knackmussii* B13. BMC Microbiol **10:**153.

9. **Gaillard, M., T. Vallaeys, F. J. Vorholter, M. Minoia, C. Werlen, V. Sentchilo, A. Puhler, and J. R. van der Meer.** 2006. The *clc* element of *Pseudomonas* sp. strain B13, a genomic island with various catabolic properties. J Bacteriol **188:**1999-2013.

10. **Gerdes, K., S. K. Christensen, and A. Lobner-Olesen.** 2005. Prokaryotic toxin-antitoxin stress response loci. Nat Rev Microbiol **3:**371-382.

11. **Gross, R., C. A. Guzman, M. Sebaihia, V. A. dos Santos, D. H. Pieper, R. Koebnik, M. Lechner, D. Bartels, J. Buhrmester, J. V. Choudhuri, T. Ebensen, L. Gaigalat, S. Herrmann, A. N. Khachane, C. Larisch, S. Link, B. Linke, F. Meyer, S. Mormann, D. Nakunst, C. Ruckert, S. Schneiker-**

Bekel, K. Schulze, F. J. Vorholter, T. Yevsa, J. T. Engle, W. E. Goldman, A. Puhler, U. B. Gobel, A. Goesmann, H. Blocker, O. Kaiser, and R. Martinez-Arias. 2008. The missing link: Bordetella petrii is endowed with both the metabolic versatility of environmental bacteria and virulence traits of pathogenic Bordetellae. BMC Genomics **9:**449.

12.   Haft, R. J., J. E. Mittler, and B. Traxler. 2009. Competition favours reduced cost of plasmids to host bacteria. ISME J **3:**761-9.

13.   Juhas, M., J. R. van der Meer, M. Gaillard, R. M. Harding, D. W. Hood, and D. W. Crook. 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev **33:**376-393.

14.   Lechner, M., K. Schmitt, S. Bauer, D. Hot, C. Hubans, E. Levillain, C. Locht, Y. Lemoine, and R. Gross. 2009. Genomic island excisions in Bordetella petrii. BMC Microbiol **9:**141.

15.   Lee, C. A., A. Babic, and A. D. Grossman. 2010. Autonomous plasmid-like replication of a conjugative transposon. Mol Microbiol **75:**268-279.

16.   Maamar, H., A. Raj, and D. Dubnau. 2007. Noise in gene expression determines cell fate in Bacillus subtilis. Science **317:**526-529.

17.   Mathee, K., G. Narasimhan, C. Valdes, X. Qiu, J. M. Matewish, M. Koehrsen, A. Rokas, C. N. Yandava, R. Engels, E. Zeng, R. Olavarietta, M. Doud, R. S. Smith, P. Montgomery, J. R. White, P. A. Godfrey, C. Kodira, B. Birren, J. E. Galagan, and S. Lory. 2008. Dynamics of Pseudomonas aeruginosa genome evolution. Proc Natl Acad Sci U S A **105:**3100-3105.

18.   Minoia, M., M. Gaillard, F. Reinhard, M. Stojanov, V. Sentchilo, and J. R. van der Meer. 2008. Stochasticity and bistability in horizontal transfer control of a genomic island in *Pseudomonas*. Proc. Natl. Acad. Sci. U S A **105:**20792-20797.

19.   Miyazaki, R., M. Minoia, N. Pradervand, S. Sulser, F. Reinhard, and J. R. van der Meer. 2012. Cellular variability of RpoS expression underlies subpopulation activation of an integrative and conjugative element. PLoS Genet **8:**e1002818.

20.   Miyazaki, R., and J. R. van der Meer. 2011. A dual functional origin of transfer in the ICE*clc* genomic island of *Pseudomonas knackmussii* B13. Mol Microbiol **79:**743-758.

21.   Ochman, H., J. G. Lawrence, and E. A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405:**299-304.

22.   **Qiu, X., A. U. Gurkar, and S. Lory.** 2006. Interstrain transfer of the large pathogenicity island (PAPI-1) of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A **103:**19830-19835.

23.   **Remus-Emsermann, M. N., and J. H. Leveau.** 2010. Linking environmental heterogeneity and reproductive success at single-cell resolution. ISME J **4:**215-222.

24.   **Sentchilo, V., K. Czechowska, N. Pradervand, M. Minoia, R. Miyazaki, and J. R. van der Meer.** 2009. Intracellular excision and reintegration dynamics of the ICE*clc* genomic island of *Pseudomonas knackmussii* sp. strain B13. Mol Microbiol **72:**1293-306.

25.   **Sentchilo, V., R. Ravatn, C. Werlen, A. J. Zehnder, and J. R. van der Meer.** 2003. Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. Strain B13. Journal of Bacteriology **185:**4530-4538.

26.   **Suzuki, A., M. Mori, Y. Sakagami, A. Isogai, M. Fujino, C. Kitada, R. A. Craig, and D. B. Clewell.** 1984. Isolation and structure of bacterial sex pheromone, cPD1. Science **226:**849-850.

27.   **Wozniak, R. A., and M. K. Waldor.** 2009. A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. PLoS Genet **5:**e1000439.

28.   **Wozniak, R. A., and M. K. Waldor.** 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol **8:**5525-5563.

ANNEX: SUPPORTING INFORMATION

INVENTORY

Figure S1. Calibration data showing the thresholding necessary to identify transfer competent cells on the basis of the intensity of the eGFP signal from the P_{int}-promoter.

Figure S2. Background data showing the thresholding necessary to identify tc-cell microcolonies on the basis of propidium iodide and reactive oxygen species staining. Further background to show tc-cell microcolony sizes in *attL* and *intB13* mutants of ICE*clc.*

Figure S3. Background data and conceptual schemes to illustrate the formation of tc-cell microcolonies in the process of ICE*clc* horizontal and vertical transmission.

Figure S4. Further data to show the intensity of the P_{int}-reporter signal in mother and daughter cells of tc-cell microcolonies of *attL* and *intB13* mutants of ICE*clc*.

Table S1. All strains and plasmids used in the study.

Table S2. Calibration data to link the proportion of tc-cell microcolonies with the number of P_{int}-reporter gene expressing cells within tc-cell microcolonies.

Table S3. Model predictions to illustrate the fate of a mutant without cell growth inhibition mechanism in tc-cell microcolonies.

Experimental procedures. Strains and media, microcolony growth procedures, epifluorescence and time lapse microscopy, ICE*clc* transfer experiments, staining techniques, identification of tc-cells, specific cloning strategies, background on growth on nutrient medium, background on thresholding on the basis of reporter gene expression, statistical procedures.

FIGURES AND TABLES
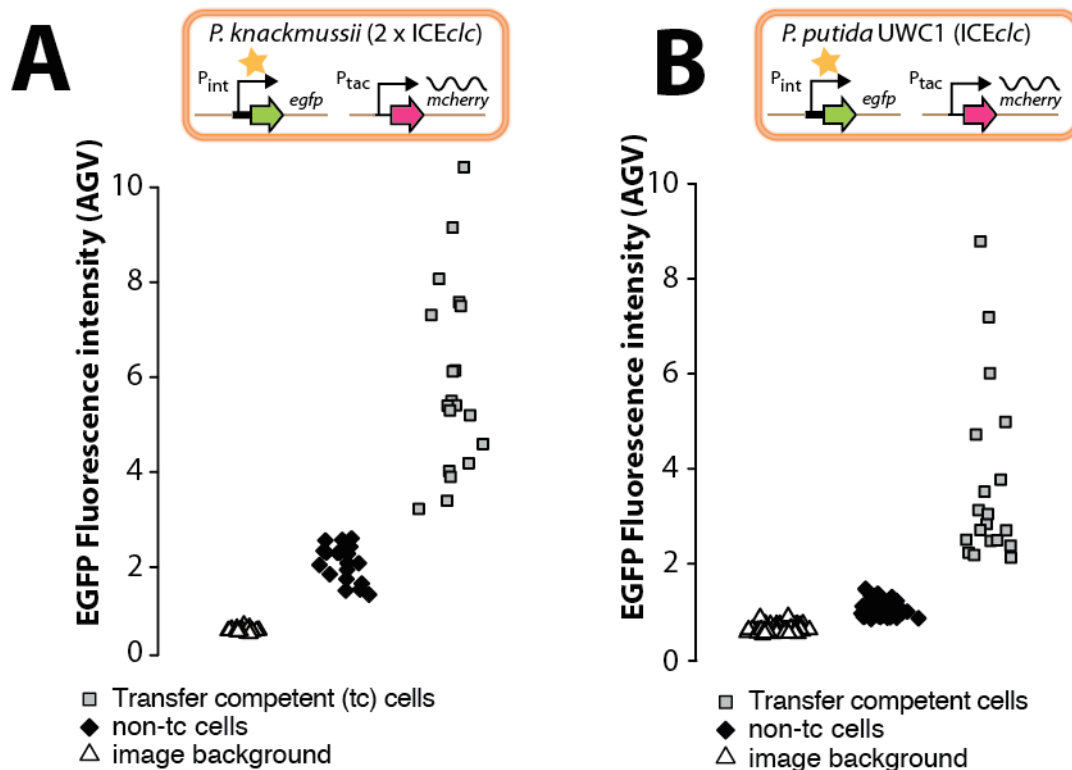
Figure S1, related to Figure 1. Fluorescence Intensity Levels from P$_{int}$ in ICEclc Transfer Competent (tc) and Non-tc Cells. (A) eGFP fluorescence in transfer competent cells (open squares, n=19), compared to non-tc cells in case of *P. knackmussii* B13 (black diamonds, n=19) and background of the nutrient surface (open triangles), directly after inoculation on nutrient agarose. (B) as (A), but for *P. putida* UWC1 (ICEclc) (n=19, ICEclc transfer competent cells, n=53 non-tc cells, n=93 background determination).
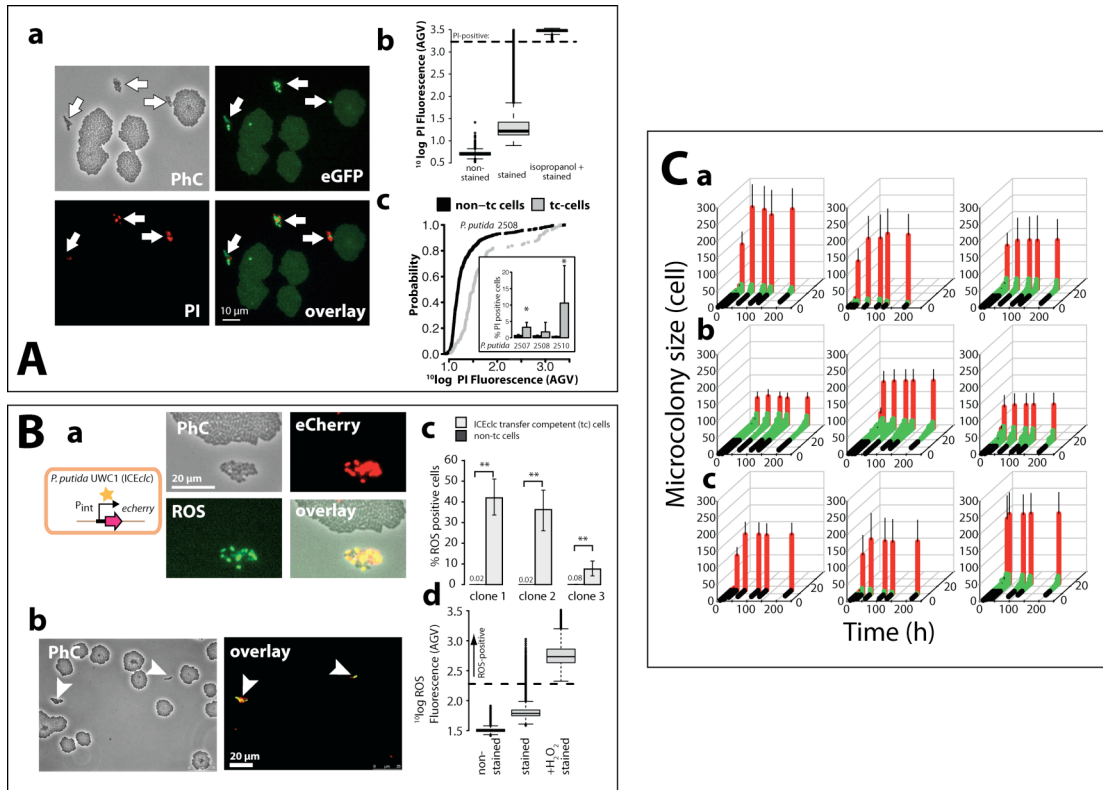
Figure S2, related to Figure 2. Preferential Cell Damage in Microcolonies Formed from ICE*clc* Transfer Competent Cells. (A) Propidium iodide (PI) stains. **a,** Transfer colony morphotypes (TCM, arrows) formed from ICE*clc* transfer competent (tc) starter cells between regular microcolonies from non-tc starter cells of *P. putida* UWC1 (ICE*clc*) scored after 120 h development on agarose surface with 0.1 mM 3-CBA. Note how PI-staining (indicative for membrane damage) invariably correlates to TCM. Images show phase contrast (PhC), eGFP and PI fluorescence (artificially colored green and red, respectively), and in overlay of both. **b,** Quantification of red fluorescence in individual cells in suspended stationary phase culture on 3-CBA of *P. putida* UWC1 (ICE*clc*, P_{int}-*egfp*) either non-stained, stained with PI or treated with isopropanol and then stained. Dotted line indicates minimum PI fluorescence of dead cells. **c,** PI fluorescence in individual cells in suspended stationary phase culture on 3-CBA of three clones of *P. putida* UWC1 (ICE*clc*, P_{int}-*egfp*) as a function of P_{int}-*egfp* expression. (B) Increased Reactive Oxygen Stress (ROS) in ICE*clc* Transfer Competent Cells. **a** ROS staining of an TCM of *P. putida* UWC1 (ICE*clc*) formed on agarose surface with 0.1 mM 3-CBA after 120 h. Relevant images show phase-contrast (PhC), eCherry expression from P_{int}, H_2DFCDA-stain (ROS), and overlay. Relevant labeling details of the ICE*clc* host strain indicated schematically. **b,** Image overview showing exclusive strong ROS staining of TCM. **c,** Percentage of ROS-positive among ICE*clc* transfer competent cells (identified on the basis of their eCherry expression from P_{int}) and ICE*clc* non-tc cells stained in stationary phase suspended batch cultures (three clones with different positions of the single copy P_{int}-*eCherry* fusion). **d,** Thresholding on stationary phase suspended batch grown cells above which ROS-staining is considered positive. Asterisks indicate significance at the P<0.001 level, while error bars represent 95% confidence

intervals. (C) Effect of *attL* and *intB13* deletions in ICE*clc* on formation of TCM. **a**, Microcolony sizes from transfer competent (tc, green and black bars) or non-tc starter cells (red bars) of three independent P$_{int}$-*eCherry* clones of *P. putida* UWC1 (ICE*clc,* D*attL*) mutant. Z-axis, colony sizes over time in individual transfer competent starter cells. Red bars, mean colony sizes over time of 50 colonies from non-tc starter cells. **b** as **a**, but for three independent P$_{int}$-*eCherry* clones of *P. putida* UWC1 (ICE*clc*, D*intB13*). **c**, as **a**, but for three independent P$_{int}$-*eCherry* clones of *P. putida* UWC1 (ICE*clc*).
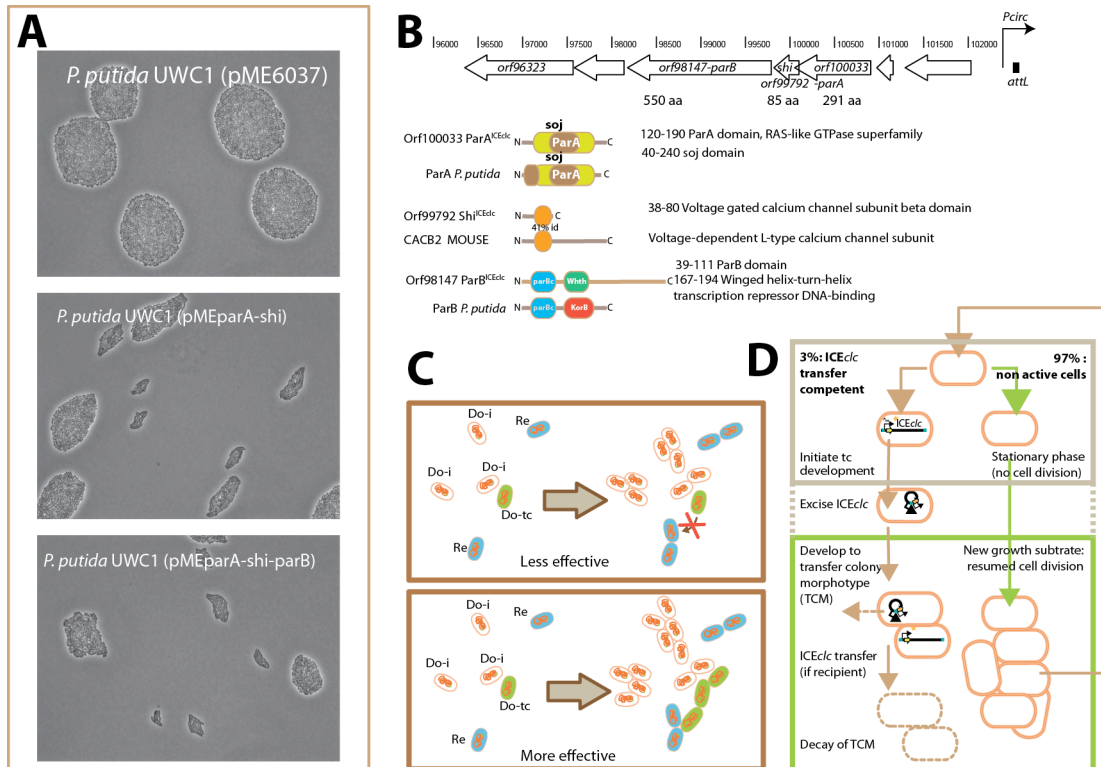
Figure S3, related to Figure 3. Transfer competent cell microcolonies. (A) Small microcolony formation induced in *P. putida* UWC1 without ICE*clc*. Phase-contrast micrographs (400 x magnification) of colonies of *P. putida* UWC1 strains on nutrient surface after 8 h induction with IPTG. Top image: *P. putida* UWC1 with the empty pME6032 vector. Middle image: *P. putida* UWC1 with the cloned *parA-shi* region from ICE*clc*. Bottom image: *P. putida* UWC1 with the cloned *parA-shi-parB* region from ICE*clc*. For cellular details and staining, see Figure 3d in main text. **(B)** The *parA-shi-parB* region of ICE*clc*. Top part shows location of *orf98147 (parB*-like) , *shi (orf99792)* and *orf100033* (*parA*-like) on ICE*clc* nearby the left end (*attL*) and the outwards facing P$_{circ}$-promoter. Predicted sizes in amino acids listed below. Lower part shows predicted Conserved Domains and Interpro domains of Orf98147, Orf100033 gene products and Shi, compared to ParB and ParA proteins of *P. putida*, and CACB2 of Mouse, respectively. GenBank Accession number of ICE*clc*: AJ617740. **(C)** Conceptual idea of the effect of having a primitive mating body as opposed to a single transfer-competent (tc) cells in the chance of contacting potential recipient cells. Upper panel: scenario for a tc-cell that is not capable of forming a small colony (e.g., *shi* mutation). Lower panel: tc-cell that can divide a few times despite producing aberrant cell morphologies. Do-i, inactive donor cells. Do-a, tc-cell (here depicted in green to represent the expression of the P$_{int}$-*egfp* fusion. Re, recipient cell. Arrow indicates provision of nutrients to restart cell division after stationary phase. **(D)** Model of transfer competence and cell reproductive inhibition pathway induced by ICE*clc*. Note that the exact timing of ICE*clc* excision was not determined here and may occur both during stationary phase or upon renewed nutrient stimulation.

211
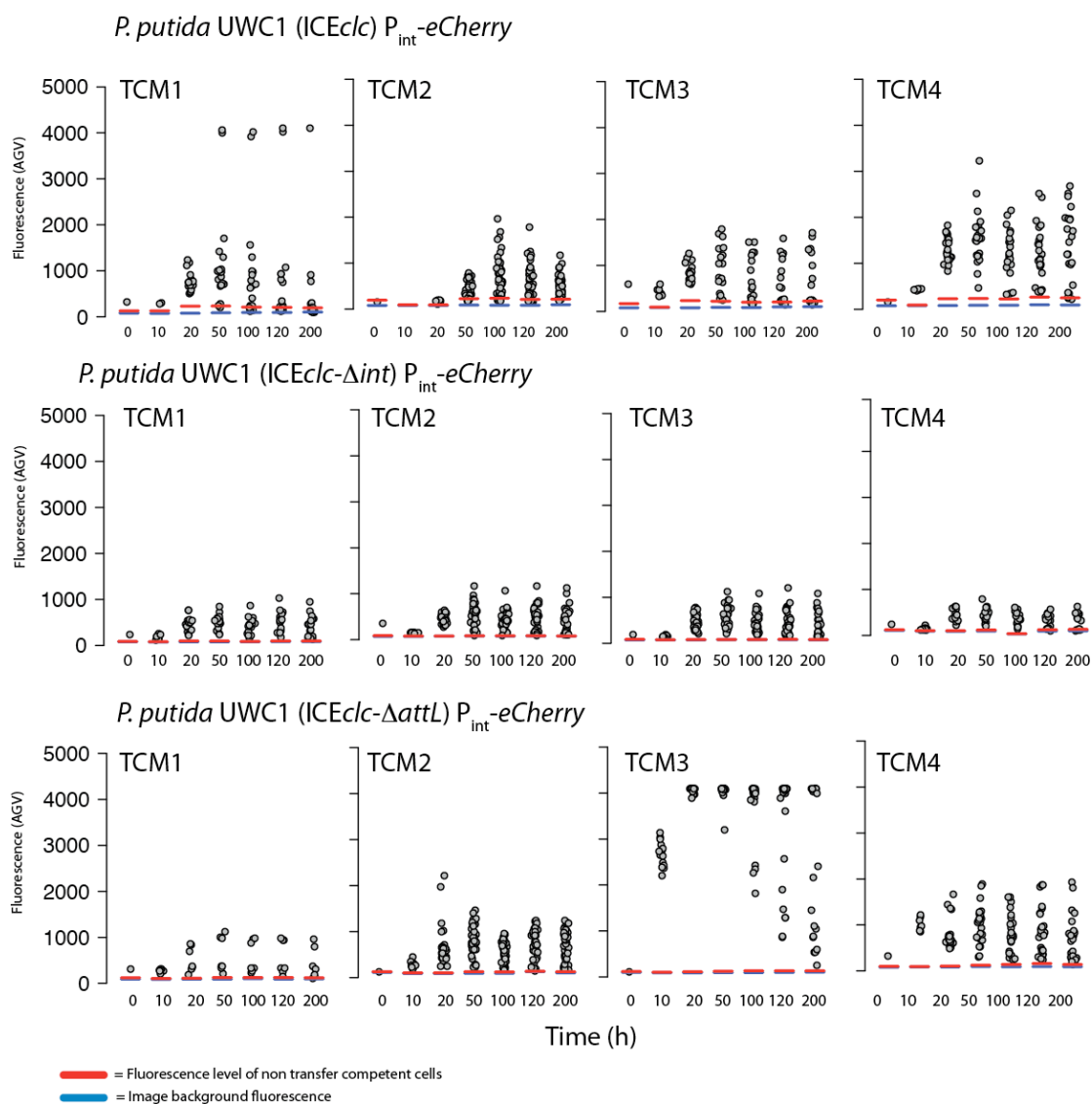
Figure S4, Related to Figure 4. eCherry Expression from Pint in Siblings Developing from ICE*clc* Transfer-Competent or Non-tc starter Cells of *P. putida* UWC1 (ICE*clc*) with or without Deletions of *intB13* or *attL.* (Top) Individual fluorescence intensity of cells within four different TCM over time for *P. putida* UWC1 (ICE*clc*) Pint-*eCherry.* (Middle) As a but for *P. putida* UWC1 (ICE*clc-int*). (Bottom) As a but for *P. putida* UWC1 (ICE*clc-attL*).

Table S1. Bacterial strains used in the study.

Strain and designation	Description	Lab collection strain number	Reference
P. knackmussii B13	Two (non-tandem) copies of ICE*clc* integrated in different *tRNAGly*-genes	78	(12)
P. putida UWC1	Plasmid-less derivative of *P. putida* mt-2 (TOL), used as recipient for ICE*clc*, Rif^r	1291	(4)
P. knackmussii B13 (P$_{in}$-*egfp*)	Single copy mini-Tn5 insertion with P$_{in}$-*egfp* fusion. Three clones with different insertion position. Km^r	1343-1345	(11)
P. putida UWC1 (jimX-*egfp*)	Single copy mini-Tn5 insertion with jimX-*egfp* artificial ICE*clc* integration site leading to strong eGFP production upon integration, Km^r	2756	(9).
P. putida UWCGC	Derivative of strain 1291 carrying a single copy mini-Tn7 insertion with Gm^r and P$_{tac}$-*mCherry*.	2744	(6)
P. putida UWC1 (ICE*clc*)	One copy of ICE*clc* integrated in *tRNAGly*-gene #3, obtained via conjugation with B13 and strain 1291	2610	(9).
P. putida UWC1 (ICE*clc*, P$_{in}$-*egfp*)	Single copy mini-Tn5 insertion with P$_{in}$-*egfp* fusion in strain 2610. Three clones with different insertion position, Km^r	2507, 2508, 2510	This work
P. knackmussii B13 (P$_{in}$-*egfp*, P$_{tac}$-*mCherry*)	Single copy mini-Tn7 insertion with P$_{tac}$-*mCherry* into strain 1343, used for constitutive labeling, Km^r, Gm^r	2586	This work
P. putida UWC1 (ICE*clc*, P$_{in}$-*egfp*, P$_{tac}$-*mCherry*)	Single copy mini-Tn7 insertion with P$_{tac}$-*mCherry* into strain 2507, used for constitutive labeling, Km^r, Gm^r	2590	This work
P. putida UWC1 (ICE*clc*)	One copy of ICE*clc* integrated in *tRNAGly*-gene #6, obtained via conjugation with B13 and strain 1291	2738	(6)
P. putida UWC1 (ICE*clc*-D*int*)	Knockout of *intB13* gene on ICE*clc* in strain 2738. No excision or transfer of ICE*clc*. Km^r	2785	(6)
P. putida UWC1 (ICE*clc*-D*attL*)	Knockout of *attL* on ICE*clc* in strain 2738. No excision or transfer of ICE*clc*. Km^r	2786	This work
P. aeruginosa ATCC33356	One copy of ICE*clc* into *P. aeruginosa*	2854	(1)

(ICE*clc*, P-*egfp*)	ATCC33356, further equipped with single copy mini-Tn5 insertion containing P-*egfp*. Km		This work
P. aeruginosa ATCC33938 (ICE*clc*, P-*egfp*)	One copy of ICE*clc* into *P. aeruginosa* ATCC33938, further equipped with single copy mini-Tn5 insertion containing P-*egfp*. Km	2869	(1) This work
P. putida UWC1 (ICE*clc*-*orf18502::Km*)	One copy of ICE*clc* with mini-Tn5-Km insertion into *orf18502*, in *P. putida* UWC1. No more transfer of ICE*clc*.	2961	This work.
P. putida UWC1 (mini-Tn7-*clc*)	Single copy chromosomal insertion of a mini-Tn*7*(*clcRABDE*) into *P. putida* UWC1. Allows growth on 3-chlorobenzoate.	3227	This work
P. putida UWC1 (ICE*clc*-D*attL,* P-*eCherry*)	Single copy mini-Tn5 insertion of a P-*eCherry* cassette into strain 2786. Three clones with different insertion position. Km	3403-3405	This work
P. putida UWC1 (ICE*clc*-D*int,* P-*eCherry*)	Single copy mini-Tn5 insertion of a P-*eCherry* cassette into strain 2785. Three clones with different insertion position. Km	3406-3408	This work
P. putida UWC1 (ICE*clc*, P-*eCherry*)	Single copy mini-Tn5 insertion of a P-*eCherry* cassette into strain 2738. Three clones with different insertion position. Km	3414-3416	This work
P. putida UWC1 (ICE*clc*-*orf18502::Km,* P-*eCherry*)	Single copy mini-Tn5 insertion of a P-*eCherry* cassette into strain 2961. Three clones with different insertion position. Km	3406-3408	This work
P. putida UWC1 (ICE*clc*-D*parA*)	Deletion of *parA* (orf100033) on ICE*clc* in strain 2738. Km	3509	This work
P. putida UWC1 (ICE*clc*-D*shi*)	Deletion of *shi* (orf99792) on ICE*clc* in strain 2738. Km	4095	This work
Plasmid	**Description**	**Lab collection number**	**Reference**
pME6032	pVS1-p15A ori shuttle vector carrying *lacI*-P; Tc	3448	(3)
pTCB177	pUC28 derivative carrying 6.4 kb *Eco*RI	1324	(11)

	fragment of ICE*clc*		
pME-EE6.4	pME6032 derivative carrying 6.4 kb *Eco*RI fragment from pTCB177 with *par-inrR* region	3977	This work
pMEparAsB	pME6032 derivative carrying full length *parA-shi-parB* locus	4039	This work
pMEparAsB'	pMEparAsB derivative lacking 980-bp 3' part of *parB*	4094	This work
pMEparADsB	pME6032 derivative carrying *parA* and *parB*; *shi* deleted	4041	This work
pMEparA'sB	pMEparAsB derivative having internal 330-bp deletion in *parA*	4039	This work
pMEshi	pME6032 derivative carrying *shi*	4072	This work
pMEparA	pME6032 derivative carrying *parA*	4151	This work
pMEparB	pME6032 derivative carrying *parB*	4150	This work
pMEparAs	pME6032 derivative carrying *parA* and *shi*	4149	This work
pMEparA's	pMEparAs derivative having internal 330-bp deletion in *parA*	4194	This work

Supplementary Table 2. Correlation between TCMs and ICE*clc* activity in different strains of *Pseudomonas putida* with ICE*clc*.

Strain	Strain Number	Background	Reporter	Total TCM·	% TCM with active P
P. putida	3403	ICE*clc* D*attL*	P-*eCherry*	11	91
	3404	ICE*clc* D*attL*	P-*eCherry*	3	67
	3405	ICE*clc* D*attL*	P-*eCherry*	6	100
P. putida	3406	ICE*clc* D*intB13*	P-*eCherry*	48	96
	3407	ICE*clc* D*intB13*	P-*eCherry*	22	100
	3408	ICE*clc* D*intB13*	P-*eCherry*	11	91
P. putida	3414	ICE*clc*	P-*eCherry*	32	94
	3415	ICE*clc*	P-*eCherry*	18	100
	3416	ICE*clc*	P-*eCherry*	26	92
P. putida	2507	ICE*clc*	P-*egfp*	11	91
P. putida	2508	ICE*clc*	P-*egfp*	12	100
P. putida	2510	ICE*clc*	P-*egfp*	8	100
P. putida	2590	ICE*clc*	P-*egfp*, P-*mCherry*	34	88

a) Strains were grown for 120 h on agarose medium with 0.1 mM 3-chlorobenzoate. TCMs were identified under phase contrast microscopy on basis of small size. Absence or presence of fluorescence from P-*eCherry* or P-*egfp* was then scored on individual cells.

Table S3. Modeling of development of reproductive inhibition independent mutants in populations of ICE*clc* containing hosts.

Parameters						
Probability of transfer competent	0.03	0.03	0.03	0.03	0.20	0.25
Mut Rate (per generation)	$10^{\cdot}$	$10^{\cdot}$	$10^{\cdot}$	$10^{\cdot}$	$10^{\cdot}$	$10^{\cdot}$
Nr mutants (start)	10	0	0	10	1	1
Cycle	End Ratio (mutant/WT) After Cycle					
1	$3.75 \cdot 10^{-4}$	$1.75 \cdot 10^{-4}$	$9.56 \cdot 10^{-6}$	$2.00 \cdot 10^{-4}$	$2.00 \cdot 10^{-4}$	$2.00 \cdot 10^{-4}$
2	$5.13 \cdot 10^{-4}$	$3.09 \cdot 10^{-4}$	$9.25 \cdot 10^{-6}$	$2.04 \cdot 10^{-4}$	$2.47 \cdot 10^{-4}$	$2.64 \cdot 10^{-4}$
3	$6.46 \cdot 10^{-4}$	$4.43 \cdot 10^{-4}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$3.51 \cdot 10^{-4}$
4	$7.81 \cdot 10^{-4}$	$5.78 \cdot 10^{-4}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$4.39 \cdot 10^{-4}$
5	$9.17 \cdot 10^{-4}$	$7.14 \cdot 10^{-4}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$5.27 \cdot 10^{-4}$
6	$1.05 \cdot 10^{-3}$	$8.49 \cdot 10^{-4}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$6.14 \cdot 10^{-4}$
7	$1.26 \cdot 10^{-3}$	$9.84 \cdot 10^{-4}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$7.90 \cdot 10^{-4}$
8	$1.46 \cdot 10^{-3}$	$1.19 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$9.66 \cdot 10^{-4}$
9	$1.66 \cdot 10^{-3}$	$1.39 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$1.23 \cdot 10^{-3}$
10	$1.87 \cdot 10^{-3}$	$1.59 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$1.58 \cdot 10^{-3}$
11	$2.07 \cdot 10^{-3}$	$1.80 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$2.02 \cdot 10^{-3}$
12	$2.27 \cdot 10^{-3}$	$2.00 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$2.64 \cdot 10^{-3}$
13	$2.48 \cdot 10^{-3}$	$2.20 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$3.43 \cdot 10^{-3}$
14	$2.68 \cdot 10^{-3}$	$2.41 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$4.49 \cdot 10^{-3}$
15	$2.88 \cdot 10^{-3}$	$2.61 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$5.90 \cdot 10^{-3}$
16	$3.09 \cdot 10^{-3}$	$2.82 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$7.85 \cdot 10^{-3}$
17	$3.29 \cdot 10^{-3}$	$3.02 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$1.04 \cdot 10^{-2}$
18	$3.56 \cdot 10^{-3}$	$3.22 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$1.38 \cdot 10^{-2}$
19	$3.84 \cdot 10^{-3}$	$3.49 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$1.84 \cdot 10^{-2}$
20	$4.11 \cdot 10^{-3}$	$3.77 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$2.45 \cdot 10^{-2}$
21	$4.38 \cdot 10^{-3}$	$4.04 \cdot 10^{-3}$	$9.20 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$2.49 \cdot 10^{-4}$	$3.26 \cdot 10^{-2}$

Simulation parameters

N cells at start: 50,000; Generation time ICEclc transfer competent cells: 2 h; Generation time non-active cells: 1 h; Nr. Generations per cycle: 10; Death rate of ICEclc transfer competent cells: 0.4 (1/h); Total amount of substrate per cycle: 10^{-4} g /mL; Amount of carbon per cell: $2 \cdot 10^{-12}$ g; Biomass yield: 0.3 g/g; Ks: $2 \cdot 10^{-5}$ g/ml; μ(max): 0.693 (1/h); μ(max, mutant): 0.693 (1/h); Population dilution factor per cycle: 0.001

EXPERIMENTAL PROCEDURES

Strains and Culture Media. All strains used in this work are listed in table S1. ICE*clc* has a size of 103 kb and occurs in two copies in the chromosome of *P. knackmussii* B13, inserted at the 3'-end of genes for *tRNA*^{Gly} . (2, 7)The element provides the host with the capacity to metabolize a number of unique aromatic carbon substrates, among which 3-chlorobenzoate (3-CBA). *P. putida* UWC1 (ICE*clc*) contains one copy of the ICE*clc* element integrated in *tRNA*^{Gly} (9). *P. aeruginosa* ATCC33356-2608 and *P. aeruginosa* ATCC33938-2606 are wild-type strains in which one copy of ICE*clc* was integrated via conjugative transfer from *P. knackmussii* B13 as donor (1). Strains were grown from a single well-separated colony in five ml of type 21C minimal medium (MM) (10) containing 4 mM 3-CBA until stationary phase (24 – 96 h) at 30°C and at 180 rpm agitation, before being reinoculated onto agarose nutrient surfaces.

ICE*clc* mutations. Deletion mutants in a variety of ICE*clc* genes and regions (i.e., *intB13*, *attL*, *orf100033-parA*^{ICEclc}, *orf99792-shi*) were produced by double recombination as described previously (6) or by Tn*5* mutagenesis and selection (*orf18502*). Mutants were verified by PCR analysis and DNA sequencing. Orf numbering is according to GenBank entry AJ617740 (ICE*clc*).

Promoter Reporter Gene Fusions. Transfer competent cells were visualized by fluorescent reporter protein expression from a single-copy chromosomal transcriptional fusion of promoterless *egfp* or *eCherry* genes to P$_{int}$, the promoter in front of the *intB13* gene in the integrated state (5, 10). P$_{int}$-fusions were inserted in single copy on the chromosome of a variety of *Pseudomonas* strains (table S1) via mini-Tn*5* delivery, and verified by antibiotic selection markers and specific PCR amplification. To improve visualization strains were additionally tagged with mCherry constitutively expressed from a single copy P$_{tac}$-*mCherry* mini-Tn*7* insertion (13). For all mini-Tn*5* insertions at least three independent clones with different insertion positions were purified and used for subsequent time-lapse microscopy experiments (table S1). *P. putida* UWC1 recipients with the conditional insertional trap (Figure 1A) were described previously (9).

Microcolony Growth Procedure. To follow the real-time fate, elongation and division of single individual ICE*clc* transfer competent (tc) and non-tc *Pseudomonas* cells we grow ICE*clc* donor strains in presence or absence of recipients from single cells to microcolonies on agarose nutrient surface in a closed sterile microscopy chamber as described previously (8).

ICE*clc* Transfer. Transfer rates of ICE*clc* wild-type and mutants used *P. knackmussii* B13 or *P. putida* UWC1 (ICE*clc*) as donors and *P. putida* UWC1-Gm (gentamycin resistance) as recipient in 48 h filter matings, as described previously (6, 9). Presence of ICE*clc* was verified by PCR using tailored primers (6, 9).

Population Growth Inhibition. The effect of individual or combinations of ICE*clc* genes on growth of *P. putida* UWC1 (without ICE*clc*) was tested in suspended culture and in microcolonies formed on agarose patches. Different combinations of ICE*clc* genes (Supplementary Experimental Procedures, table S1) were cloned in the broad-host range vector pME6032 under IPTG derepressable LacI^q-P$_{lac}$ control (3). Two sets of triplicate suspended cultures were inoculated at a culture turbidity (measured at 600 nm) of ~0.001 in type 21C MM-medium with 10 mM fructose, with or without 0.5 mM IPTG, after which the culture growth was continued to be monitored for 40 h. For microcolony experiments cells of *P. putida* UWC1 with pME6032-based plasmids were allowed to go through 3 cell doublings, after which 3 µl of a 50 mM IPTG stock solution was dropped on the back of a 150 µl agarose patch. Microcolony development was continued to be imaged during another 8 h, after which the cells were stained with PI and, after another 2 h, with DAPI (as described above).

Microscopy. Cells, microcolonies and the brightness of fluorescent reporters were imaged with a Leica AF 6000 epifluorescence microscope (Leica Microsystems Heidelberg GmbH). Images were taken with a DFC320 monochrome camera (Leica Microsystems GmbH), a 100/1.30 oil immersion lens (HCX PL FLUOTAR; Leica) or a 40/0.60 (HCX PL Fluotar L40) lens for measuring microcolony morphotypes, at an exposure time of 800 ms. Images were digitally recorded as 16-bit TIFF-files using the Leica AF6000 software, and analysed using METAMORPH (version 6.1r5, Visitron Systems, Germany). Images for display were artificially colored 'red' (for eCherry, mCherry and PI) or 'green' (for eGFP or H$_2$DCFDA), stored as 8 bit files using the Leica LAS software, and then auto-leveled and cropped to the final resolution and image size using ADOBE PHOTOSHOP CS4 (Adobe Inc.).

Time-Lapse Experiments. Time-lapse experiments were facilitated by the use of a mark-and-find motorized stage (Märzhäuser Wetzlar, Germany). Regions containing ICE*clc* tc- and non-tc cells to follow over time were identified at the start of the experiment by recording individual cell eGFP or eCherry fluorescence values expressed from the single copy P$_{int}$ – fusion. The cutoff for being identified as ICE*clc* transfer competent cell based on P$_{int}$–*egfp* or –*eCherry* expression was taken as a value higher than a fluorescence value of Q3 (upper quartile) + IQR (interquartile range) x 3, (Figure S1). Microcolony formation of both types of cells was followed until stationary phase and until new tc-cells appeared in the microcolonies formed from ICE*clc*

non-tc starter cells. Between 10 and 20 regions per agarose patch were followed in time. The total number of cells and the proportion of cells expression reporter from P_{int} per microcolony were counted at each time point.

Transfer competent cell microcolony (TCM) Quantification. In addition to quantifying the proportion of cells expressing the P_{int}-*egfp* reporter fusion, we measured the proportion of TCM among all microcolonies produced on agarose surface as an indication for the cellular differentiation process conferred by ICE*clc*. TCMs were identified on gel patches incubated for 120 h, stained first with propidium iodide (PI) and subsequently –with a one hour interval, with 4',6' diamino-2-phenylindole·2HCl (DAPI). To score as TCM it consisted of at least two cells, had started from cells expressing P_{int}-*egfp* and/or contained clustered PI-positive cells and representing an outlier low-end area-size (upper quartile - interquartile range x 1.5) in comparison to neighbouring colonies (Figure S2, table S2). The proportion of TCM among all microcolonies was quantified from 50 randomly placed and imaged regions within an area spanning a typical gel patch. Total biomass and median colony size (in area units) per image was determined on thresholded DAPI stained images in METAMORPH. The number of microcolonies per image was then calculated by dividing the total biomass area divided by the median colony size, which was summed up across all investigated images to give the total number of microcolonies. The frequency of TCMs was then calculated as the total number of positively scored TCM divided by the calculated total number of microcolonies over all images per patch (Table S2).

Staining Procedures. For staining developed microcolonies we pipetted a droplet of 3 µl of staining solution to the back (non-inoculated) side of a 150 µl agarose patch and incubated the patch for one hour in the dark before microscopy imaging. As staining solutions we used PI (final concentration 30 µM), H_2DCFDA (20 µg/ml), or DAPI (20 µg/ml). In the case of staining suspended cultures, one µl of PI stock solution (20 mM) was added to 500 µl cell suspended and incubated for 30 min in the dark. For H_2DCFDA staining, one µl of a 10 mM stock solution in DMSO was added to 500 µl cell suspension and the suspension was incubated for two hours at 22°C in the dark before microscopy examination. As a positive control for reactive oxygen species, cells were treated with H_2O_2 (final concentration in suspension of 80 µM) for two hours in the dark, after which they were stained using H_2DCFDA.

Construction of pME6032-derived plasmids with *parA-shi-parB* fragments. A 6.4 kb EcoRI fragment of pTCB177 was subcloned into pME6032, generating plasmid pME-EE6.4. To generate pME-parA, pME-shi and pME-parB, the three genes *parA^{ICEclc}*, *shi* and *parB^{ICEclc}* were separately amplified from ICE*clc* by PCR, cloned into pUC18-

miniTn7-Lac, and from there subcloned into the *Sac*I-*Kpn*I sites on pME6032. In the same way the full length *parA-shi-parB* region was PCR amplified and cloned into pME6032. The resultant plasmid pME-parAsB was digested by *Stu*I and self-ligated to generate pME-parAsB', which lacked the 3' 980-bp of *parB*. On the pUC18-miniTn7-Lac derivative plasmid the full length *parA-shi-parB* region was digested by *Sph*I and self-ligated. This creates an 330-bp internal deletion in *parA*. The fragment was recovered by *Sac*I-*Kpn*I digestion and subcloned into pME6032, generating pME-parA'sB. Plasmids pME-parAsB and pME-parA'sB were then digested by *Sal*I-*Xho*I and self-ligated to generate pME-parAs and pMEparA's, respectively, that lack *parB*.

Thresholding reporter gene intensities from P_{int} for identification of cells as transfer competent. Transfer competent (tc) cells were identified at the beginning of microcolony growth on agarose nutrient surface, or within newly formed microcolonies during stationary phase on the basis of the per-cell fluorescence value of eGFP or eCherry expressed from a single copy chromosomally located fusion to the P_{int}-promoter (Figure S1). Strains were precultured in aqueous suspension (batch culture) on 4 mM 3-chlorobenzoate (3-CBA) until stationary phase, after which the culture was diluted and individual cells were deposited on a fresh solid agarose medium surface with 0.1 mM 3-CBA in an approximate density of 1 cell per 70 μm^2, (8). Tc cells arising in stationary phase have activated the reporter from P_{int} and are visible as brighter green cells on the agarose surface. A random region on the agarose patch (50 × 70 μm^2) was chosen and single cells were focused under phase contrast at 1000 × magnification. An image was then recorded of eGFP or eCherry fluorescence and quantified using a METAMORPH subroutine. The cutoff for being identified as ICE*clc* tc-cell based on P_{int}–*egfp* or –*eCherry* expression was taken as a value higher than a fluorescence value of Q3 (upper quartile) + IQR (interquartile range) x 3 (interquantile range) in a boxplot representation of all data. If the image contained at least one cell with brighter than average eGFP or eCherry fluorescence it was assumed to represent an ICE*clc*-tc starter cell and the region was saved for the duration of the time-lapse investigation. In a similar way, a randomly chosen non-bright cell was identified and followed as an ICE*clc* non-tc control cell. Twenty such regions were followed for each experiment.

To identify tc cells within microcolonies developed from non-tc starter cells we imaged and quantified the eGFP or eCherry fluorescence intensity of all individual cells within microcolonies during stationary phase. These were analyzed via the subroutine and method as outlined above for single starter cells.

Biochemical damage occurring in transfer competent cells exposed to new growth substrate. A large proportion of cells (up to 100%) in transfer colony morphotypes (TCM) developed from ICE*clc*-tc starter cells stained positive with propidium iodide (PI), suggesting those cells to carry severely compromised membranes permitting PI entry and binding to DNA (Figure S2A). Positive staining with PI was identified on suspended cells as having a PI fluorescence higher than the minimum fluorescence of stained cells and was compared to cells first treated with isopropanol and then PI stained (Figure S2A,c and d). In contrast, microcolonies formed from non-tc starter cells in stationary phase displayed less than 1% PI-positive cells, and the occurrence of staining PI-positive was not significantly ($P>0.05$) correlated to expression of eGFP from P_{int} at stationary phase in the microcolony (Figure S2A).

TCM from ICE*clc*-tc cells also stained very brightly with dichlorodihydrofluorescein diacetate (H_2DCFDA, Figure S2B), which stains cells with higher intracellular concentrations of reactive oxygen species (ROS) and is a further potential indicator for cell damage. A significantly larger proportion ($P<0.001$) of ICE*clc*-tc cells in stationary phase suspended culture also stained H_2DCFDA-positive (15.6 ±1.5 %, n=47, three independent clones, Figure S2B). These results therefore indicated that cells in which ICE*clc* is activated accumulate a variety of biochemical damage during stationary phase, and continue to do so when provided with new growth substrate.

Quantification of the proportion of transfer competent cells via their propensity to form aberrant transfer colony morphotypes (TCM). Importantly, two methods were used to identify the proportion of ICE*clc*-tc cells in a population: (i) expression of reporter protein from P_{int} and (ii) poor reproduction and, consequently, consistent formation of smaller microcolonies with a high percentage of damaged cells in comparison to microcolonies starting from a non-tc cell (Figure 2). Since smaller microcolonies could possibly also form because of defects in use of 3-CBA in the agarose medium (for example, as a result of excision and loss of ICE*clc*) or because of toxicity of 3-CBA metabolites on sensitive cells, we first calibrated co-occurrence of TCM and reporter induction. Notably, in strains with ICE*clc* having deleted *intB13* or *attL* abolishing ICE*clc* excision, TCM still formed at the same proportion (Figure 2C). This showed that TCM are not a consequence of loss of ICE*clc* in individual cells. As explained above, staining with PI above threshold correlated in 100% of the observed microcolonies with expression of eGFP from P_{int} (in TCM with at least two cells formed, Figure 2A). Since up to half of all ICE*clc*-tc starter cells do not grow at all on fresh medium surface (e.g., Figure 2b), the proportion of TCM is lower than that obtained by scoring P_{int}-*egfp*

223

expression. The background score for TCM in strains without ICE*clc* but which could grow on the same medium (i.e., *P. putida* UWC1 mini-Tn7-*clc*) was 0.05%. This probably reflects the proportion of naturally dead and damaged cells during manipulation from stationary phase culture to fresh agarose medium. Significantly, TCM also stained very brightly with H_2DCFDA (Figure S2C), but this was not used for TCM quantification.

Verification of fluorescent labeling and insertion artifacts. In order to ensure that induction of transfer competence or cell death in tc-cells was not due to labeling artifacts, we scored both proportions of TCM in labeled and non-labeled strains (Figure 3, compared *P. putida* UWC1 (ICE*clc*) and strain having P_{int}-*eCherry* fusion). In all cases of mini-transposon labeling we analyzed three independent clones with different insertion position (Table S1), without any noticeable effect on proportions of TCM or expression of reporter protein from P_{int}. strains and of cells that express reporter protein from P_{int} was independent of the reporter used (eGFP or eCherry, note Figure 3), or of the insertion position of the mini-Tn5 used to deliver the single copy P_{int}-fusion on the chromosome of the test strain. In order to exclude specific effects of eGFP we used the same expression from P_{int} using eCherry. The same results were recorded for three clones with different insertional position of the P_{int}-*egfp* or P_{int}-*eCherry* fusion in *P. knackmussii* B13 and in *P. putida* UWC1 (ICE*clc*) (not shown).

Modeling of effect of mutations abolishing cell reproductive inhibition process. One could imagine that spontaneous mutants lacking the TCM process could arise, but modeling of simple serial batch growth suggests that at the current proportion of 3% tc-cells mutants without cell killing appearing at rates of 10^{-8} per generation are unlikely to take over before a few hundreds of generations (Table S3). Only if a large fraction of tc-cells (>20%) would arise in a stationary phase populations such mutants would rapidly outcompete the wild-type in serial batch transfer growth experiments.

Statistics. Data analysis and statistics were performed using the program R (http://cran.r-project.org). 95% Confidence intervals on fluorescence intensities of subpopulations was calculated via 500 bootstrapping cycles (using the R algorithms 'quantiles' and 'boot.ci'). The threshold fluorescence for being assigned a PI-positive cell was identified as the minimum PI fluorescence of isopropanol treated stained cells (Figure S2A). The threshold for being categorized as an ROS-positive cell was identified as the minimum fluorescence among H_2O_2-treated and H_2DCFDA-stained cells (Figure S2B). Statistical significance was tested at the $P<0.001$ (two asterisks) or $P<0.01$ level (one asterisk). Measurement differences in other experiments were tested using ANOVA followed by a Tukey post hoc test.

REFERENCES

1.   **Gaillard, M., N. Pernet, C. Vogne, O. Hagenbüchle, and J. R. van der Meer.** 2008. Host and invader impact of transfer of the *clc* genomic island into *Pseudomonas aeruginosa* PAO1. Proc Natl Acad Sci U S A **105:**7058-7063.

2.   **Gaillard, M., T. Vallaeys, F. J. Vorholter, M. Minoia, C. Werlen, V. Sentchilo, A. Puhler, and J. R. van der Meer.** 2006. The *clc* element of *Pseudomonas* sp. strain B13, a genomic island with various catabolic properties. J Bacteriol **188:**1999-2013.

3.   **Klockgether, J., A. Munder, J. Neugebauer, C. F. Davenport, F. Stanke, K. D. Larbig, S. Heeb, U. Schock, T. M. Pohl, L. Wiehlmann, and B. Tummler.** 2010. Genome diversity of Pseudomonas aeruginosa PAO1 laboratory strains. J Bacteriol **192:**1113-1121.

4.   **McClure, N. C., A. J. Weightman, and J. C. Fry.** 1989. Survival of *Pseudomonas putida* UWC1 containing cloned catabolic genes in a model activated-sludge unit. Appl Environ Microbiol **55:**2627-2634.

5.   **Minoia, M., M. Gaillard, F. Reinhard, M. Stojanov, V. Sentchilo, and J. R. van der Meer.** 2008. Stochasticity and bistability in horizontal transfer control of a genomic island in *Pseudomonas*. Proc. Natl. Acad. Sci. U S A **105:**20792-20797.

6.   **Miyazaki, R., and J. R. van der Meer.** 2011. A dual functional origin of transfer in the ICE*clc* genomic island of *Pseudomonas knackmussii* B13. Mol Microbiol **79:**743-758.

7.   **Ravatn, R., S. Studer, D. Springael, A. J. B. Zehnder, and J. R. van der Meer.** 1998. Chromosomal integration, tandem amplification, and deamplification in *Pseudomonas putida* F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from *Pseudomonas* sp. strain B13. J. Bacteriol. **180:**4360-4369.

8.   **Reinhard, F., and J. R. van der Meer.** 2010. Microcolony growth assays p. 3562-3570. *In* K. N. Timmis, V. de Lorenzo, T. McGenity, and J. R. van der Meer (ed.), Handbook of Hydrocarbon and Lipid Microbiology, vol. 5. Springer Verlag.

9.   **Sentchilo, V., K. Czechowska, N. Pradervand, M. Minoia, R. Miyazaki, and J. R. van der Meer.** 2009. Intracellular excision and reintegration dynamics of the ICE*clc* genomic island of *Pseudomonas knackmussii* sp. strain B13. Mol Microbiol **72:**1293-306.

10. **Sentchilo, V., R. Ravatn, C. Werlen, A. J. Zehnder, and J. R. van der Meer.** 2003. Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. Strain B13. Journal of Bacteriology **185:**4530-4538.

11. **Sentchilo, V., A. J. Zehnder, and J. R. van der Meer.** 2003. Characterization of two alternative promoters for integrase expression in the *clc* genomic island of *Pseudomonas* sp. strain B13. Molecular Microbiology **49:**93-104.

12. **Stolz, A., H. J. Busse, and P. Kampfer.** 2007. *Pseudomonas knackmussii* sp. nov. Int J Syst Evol Microbiol **57:**572-576.

13. **Tecon, R., O. Binggeli, and J. R. van der Meer.** 2009. Double-tagged fluorescent bacterial bioreporter for the study of polycyclic aromatic hydrocarbon diffusion and bioavailability. Environ Microbiol **11:**2271-2283.

CHAPTER 5

Life history analysis of integrative and conjugative element activation in growing microcolonies of *Pseudomonas*

ABSTRACT

ICE*clc* is an Integrative and Conjugating Element in the bacterium *Pseudomonas knackmussii.* ICE*clc* is vertically inherited through co-replication with the host-chromosome during cell division. Horizontal transmission is initiated in a small (~3%) subpopulation of cells during stationary phase, which become transfer competent (tc) and finally arrest cell growth. Here, we investigate the hypothesis whether pre-existing biochemical damage or cell age are factors predetermining cells in which the ICE*clc* transfer program is activated. Through the use of time-lapse microscopy, we follow the life history of tc-cells within microcolonies. We demonstrate that the age of the cell pole is unlikely to play a role in the decision of its fate to initiate the ICE*clc* tc program. Furthermore, we show that initiation of transfer competence is not the result of the physiological state of ancestor cells, or of a close relative (sister cell). In contrast, tc-cells show higher levels of reactive oxygen species and membrane damage than non-tc cells, but whether cause for or effect of ICE*clc* activation could not be discerned. ICE*clc* activation occurs spatially randomly in a microcolony, which might additionally be important for maximising the chances in a biofilm to contact potential recipients.

INTRODUCTION

Clonal bacterial populations display heterogeneous characters that are not due to sporadic mutations in individuals as a result of horizontally acquired DNA, loss of genetic information or rare and spontaneous point mutations (8). Common causes of such phenotypic heterogeneity include phase variation, asynchronized cell cycles or gene expression noise (8). In addition, there are some reports that suggest cell aging to play a role in the occurrence of phenotypic heterogeneity (1, 8). Cell age was shown to affect cell size and timing of cell division (3, 5), but also growth rates (27, 28) and cell survival (30). Furthermore, it has been shown in *Escherichia coli* that the sub-cellular localisation of proteins (17) and protein aggregates can differ among individuals in a clonal population as a result of cell age (13, 31). Cell age in symmetrically dividing bacterial species is frequently measured as the relative age (in numbers of divisions) of the poles of an individual cell in a population (27). This concept is based on the hypothesis that cell constituents with limited diffusion and a long half-life eventually accumulate at the old pole, resulting in a physiological asymmetry between the old and new poles (9). A distinction between new and old cell poles is thus necessary. Whereas new poles are produced at mid-cell during division by binary fission, old poles remain at the cell ends. The number of times an old pole has "witnessed" the birth of a new pole can then be taken as a measure for cell age. Cell pole births and age determinations can be inferred from tracking of dividing individual cells in e.g., microcolonies or microfluidics devices, by time-lapse microscopy (14). Tracking of every individual cell during formation of a microcolony also allows to determine their interrelatedness and genealogy, which can be represented as a lineage tree (5, 27, 29).

Individual cell tracking methods are particularly suitable to follow phenotypic heterogeneity events within isogenic populations, when coupled to autofluorescent protein expression from promoters of interest (26, 27, 29). Expression of the autofluorescent protein can then be linked to parameters like birth time, cell length, growth rate, cell pole age and physiological state (5, 27, 29). As an example, Veening *et al.* (29) used the *spoIIA-* and *abrB* promoters fused to *gfp* to show that sporulation in *Bacillus subtilis* cells within a microcolony was stochastic and not related to cell pole age, although it was inherited epigenetically as it preferentially followed cell lineages (29). In contrast, Bergmiller *et al.* showed that growth rate and cell size in *Methylobacterium extorquens* AM1 depend on cell pole age (5).

The objectives of the underlying study were to study bistable activation of the ICE*clc* element in *Pseudomonas knackmussii* B13 and *Pseudomonas putida.* ICE*clc* is an integrative and conjugating element with a size of 103 kb which enables its host to utilize the

unique carbon and energy source 3-chlorobenzoate (3-CBA) (10). ICE*clc* is integrated directly downstream of a gene for tRNA^{Gly} (25), but can become active in 3-5% of cells in a population during stationary phase, inducing them to become transfer competent (15, 22). Transfer competent (tc) cells excise ICE*clc* and can transfer the element to recipient cells (22). The molecular decisions that lead to ICE*clc* becoming activated in particular individual cells but not in others are not well understood. Recent findings showed the role of the stationary phase sigma factor in controlling the proportion of tc cells in a population and suggested that cells with on average higher amounts of RpoS are more prone to become active ICE*clc* donors (16). Intriguingly, the proportion of tc cells arising in culture is dependent on the carbon substrate that cells have used, being highest (3-5%) upon growth with 3-CBA but low upon growth with succinate or glucose (<1%) (10, 15). In analogy to activation of prophages by SOS response (18) one possible hypothesis for ICE*clc* excision is that cells in which the ICE is activated are already predetermined by biochemical damage that may be the result of cell age or other (4, 6, 7). Also, specific signalling effects at high cell density could possibly influence the arisal of tc cells (19-21), although previous experiments showed no effect of culture exposure to UV or chemicals, or of high and low cell density on the appearance of the excised ICE*clc* form (25).

The objectives of the work presented here were thus to study the appearance of tc cells as a function of their life history in microcolonies. In particular, we examined the following hypotheses: (i) tc cell formation is dependent on cell pole age with older cells being more prone than younger cells. (ii) tc cell formation preferentially occurs in specific lineages in a microcolony with, incidentally, lysing sister cells. (iii) tc cell formation is a function of spatial organisation or density within a microcolony. (iv) tc cell formation occurs as a consequence of pre-existing detectable biochemical damage. To study hypotheses (i)-(iii) we used a microcolony growth set-up as previously reported (23) with *P. knackmussii* B13 cells that were labelled by two fluorescent reporter markers. One of those consists of a constitutively expressed mCherry protein, the other of an enhanced green fluorescent protein (eGFP), which is brought under control of the *intB13* integrase promoter (P_{int}) and therefore expressed only in tc cells (22, 25). Life history of individual cells in the microcolonies was recorded by cell tracking on time-lapse image series taken by epifluorescence microscopy (5, 14, 24). Hypothesis (iv) was examined by staining for cell damage and reactive oxygen species (ROS) formation in individual cells grown in suspended batch culture on 3-CBA to stationary phase of *P. knackmussii* B13 or *P. putida* UWC1 (ICE*clc*).

RESULTS

Cell pole age distribution in microcolonies of *P. knackmussii* B13 cells. In order to determine whether the occurrence of tc cells is correlated to their cell pole age we determined the cell pole age distributions of in total 811 tc-cells of *P. knackmussii* B13 among 69 microcolonies in stationary phase from time-lapse imaging (Table 1, supporting information (SI) Raw data). B13 tc-cells were identified on the basis of the level of eGFP expression from a single copy chromosomal fusion to the P_{int}-promoter (Figure 1B, SI Figure 1). On average, 6.8 % of cells in a microcolony grown to stationary phase developed tc (Figure 1B, Table 2). Also on average, cells in such microcolonies went through seven cell divisions, producing between 160 and 200 cells (Table 2). The theoretical cell pole age distribution for cells in microcolonies that go through seven divisions follows an exponential decay, resulting in 50% of cells with cell pole age 1, 25% of cells with cell pole age 2, etc. (see Table 1). Chi-square testing indicated no statistically significant difference between the expected and the observed cell pole age distribution of ICE*clc* tc-cells pooled from in total 69 microcolonies (Table 1, Goodness-of-fit test: $P = 0.643$; Fisher's exact test: $P = 0.899$). ICE*clc* tc cells are thus as likely to appear in a cell with pole age 1 (youngest generation) as with pole age 7 (oldest generation). Given that the largest number of tc-cells appears among cells with a cell pole age of 1 suggests that prolonging the number of generations to grow a microcolony would not change the correlation of cell pole age to ICE*clc* formation. This indicates that there is no likely relation between the occurrence of transfer competence and cell pole age.

Lineage dependent tc cell occurrence. In order to test whether the arisal of tc cells is dependent on events occurring within a particular cell lineage, we tracked individual cells during the formation of three microcolonies of the P_{int}-*egfp* labeled *P. knackmussii* B13 cells (Figure 1A, SI Figure 2, 3). The three colonies were selected on a visual basis, being monolayered only and with sufficient tc-cells in stationary phase. Statistically speaking the three selected microcolonies may have had an overrepresentation of the number of tc-cells compared to a larger set of microcolonies (Table 2), but the cell pole age distributions were again not significantly different from the assumed theoretical distribution (Table 3). Lineaging of mother to daughter cells showed that cell division in the microcolonies becomes unsynchronized after 2-4 divisions. tc-cells mostly appeared in the final generation of a lineage, but some lineages stopped dividing sooner than others and, therefore, tc-cells appear at different moments in the microcolony (Figure 2A, C, E, Fig. S3). Once initiated, most but not all tc-cells did not further divide (Figure 1A). Between 7 and 12 % of individual cells in the three observed microcolonies lyse (Table 2; arrow in Figure 1A), with lysis sometimes occurring

before tc cell appearance (Figure 2C, E). Lineages with lysed cells occurred randomly (Figure 1A, Table 10). Interestingly, some of such cells developed a small concentrated green spot before lysis and consistently such cells did not divide any longer once the green spot was visible (Figure 3). ICE*clc* tc-cells did not systematically appear in cell lineages that stopped cell division earlier (Figure 2B, D, F). Furthermore, we tested if sister cells of tc-cells are more likely to become transfer competent themselves (Table 6), or if families of cells derived from the same ancestor are more likely to contain tc-cells (Table 9, 11). Both hypotheses were not supported by statistical analysis (Table 6, 9, 11).

Lineage dependent occurrence of cell lysis. In order to determine whether there was a correlation between cell lysis and cell pole age, we repeated the analysis of above for lysed cells and their position in the lineage tree. Statistical analysis did not support the hypothesis that cell lysis occurrence was related to cell pole age (Table 4). We then tested whether there was a correlation between the occurrence of lysed cells and the occurrence of tc-cells within the same microcolony, the hypothesis being that there would be a higher probability of having a lysed (lineage) sister cell when becoming a tc-cell. Comparing the observed frequency to a randomized probability of having lysing sister cells by Chi-square testing did again not show any significant difference (Table 5), meaning that tc formation is not related to nor the consequence of having a sister cell which is more prone to lyse. We further asked the question whether observed lysis could have been epigenetically inherited and tested if lysis occurred preferentially within clusters of closely related cells, which was not the case (Table 10).

Spatially dependent formation of tc-cells or lineaging within microcolonies. Next, we tested whether tc formation or cell lysis occurred specifically within certain lineages that form spatially within a microcolony. Interestingly, also lineages of B13 cells form in longitudinal direction within a microcolony (Figure 1C) as previously observed in *E. coli* (14, 24). To test whether tc-cells or cell lysis appear spatially random within a microcolony we divided the mature microcolony in random blocks of ~17 cells each, counted the occurrence of tc- and lysed cells within that block (Figure 1D, E), and tested in Chi-square or Fischer's exact test whether this occurrence follows Poisson distribution or not. Results showed that both tc- and lysed cells appear spatially at random within a microcolony (Table 7, 8).

Occurrence of biochemical damage in individual cells. In order to discern whether tc-cells arise as a result of some pre-existing or accumulated biochemical damage, we measured the timing of birth of tc-cells compared to non-tc cells in their final cell division, assuming that when mother cells leading to tc-cells would have been previously damaged they

would not have divided as fast as the others. Results, however, showed that for the three microcolonies the birth time of tc-cells was not different than for non-tc cells (Figure 2B, D, F). We further tested whether tc cells on average display more membrane damage or higher ROS levels than in non-tc cells. As this staining is difficult to perform on growing microcolonies because the dyes may influence the further life trajectory of the cells, we relied on staining of stationary phase suspended cultures. *P. knackmussii* B13 P$_{int}$-*egfp* stationary phase cultures on 3-CBA displayed less than 1% PI- positive cells. Being identified as PI-positive (i.e., having PI fluorescence as expected from isopropanol treated cells) was not significantly (P>0.05) correlated to being identified as a tc-cell (SI Table 1). In contrast, tc-cells showed higher levels of PI fluorescence than non-tc cells (Figure 4, P<0.001). *P. putida* UWC1 (ICE*clc*) P$_{int}$-*egfp* stationary phase cultures showed a significantly larger proportion of PI-positives among tc-cells than among non-tc cells (P<0.001, Figure 4, Fig. S9, S10, Table S1). Also, a significantly larger proportion of tc-cells than of non-tc cells in stationary phase 3-CBA grown cultures of *P. putida* UWC1 (ICE*clc*) P$_{int}$-*mCherry* stained ROS-positive (P<0.001, Figure 4, Figure S7, Figure S8, Table S1).

Figure 1 (next page). ICE*clc* tc Formation and Lysis during Microcolony Growth. (A) Lineage tree of 157 cells of microcolony 1A. Arrows point to cells (tree tips) identified as ICE*clc* tc-cells. Crosses point to lysed cells (shorter tree tips). Capital letters (*A*, *B*, …, *H*) indicate lineage clusters derived from the descendents of the eight initial cells after approximately 3 generations (see Table 9, 10). Small letters (*a*, *b*, …, *p*) name lineages sub-clusters derived from the descendents of the 16 initial cells after approximately 4 generations (see Table 11). Different colours and patterns above the tree uniquely identify lineages and sub-lineages after 70 h of growth as represent Ted in panel (C). Yellow stars indicate time of appearance of tc cells. Note non-linear time scale in panel A. (B) Phase contrast visualisation (left) and overlay of green and red fluorescence (right) of microcolony 1A after 70 h of growth. Note absence of cell shapes of lysed cells. Arrows shows an example of a lysed cell position. White bar indicates 2 μm. Non-tc cells appear as red because of constitutive mCherry expression from the P$_{tac}$ promoter. Tc-cells appear as yellow because of constitutive mCherry expression from the P$_{tac}$ promoter in combination with expression of eGFP from the P$_{int}$ promoter (diagram). Note absence of fluorescence in lysed cells (arrow). (C) Region map of microcolony 1A after 70 h showing location of lineages (colours) and sub-lineages (colours with or without pattern) as defined in (A). (D) Region map of microcolony 1A after 70 h showing randomly defined areas with occurrences of lysed cells (black dots). (E) Region map of microcolony 1A after 70 h showing randomly defined areas with occurrences of ICE*clc* tc-cells (green dots).
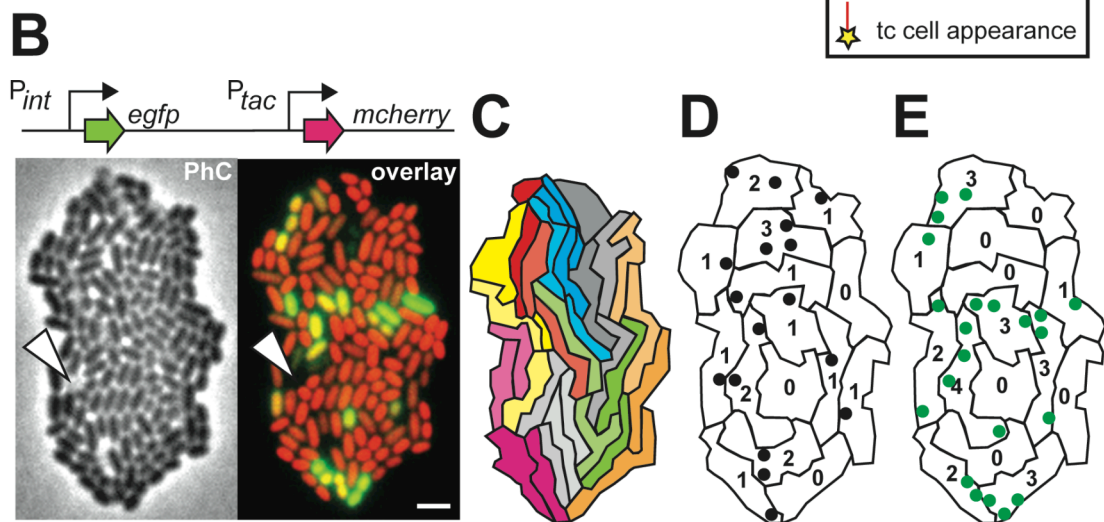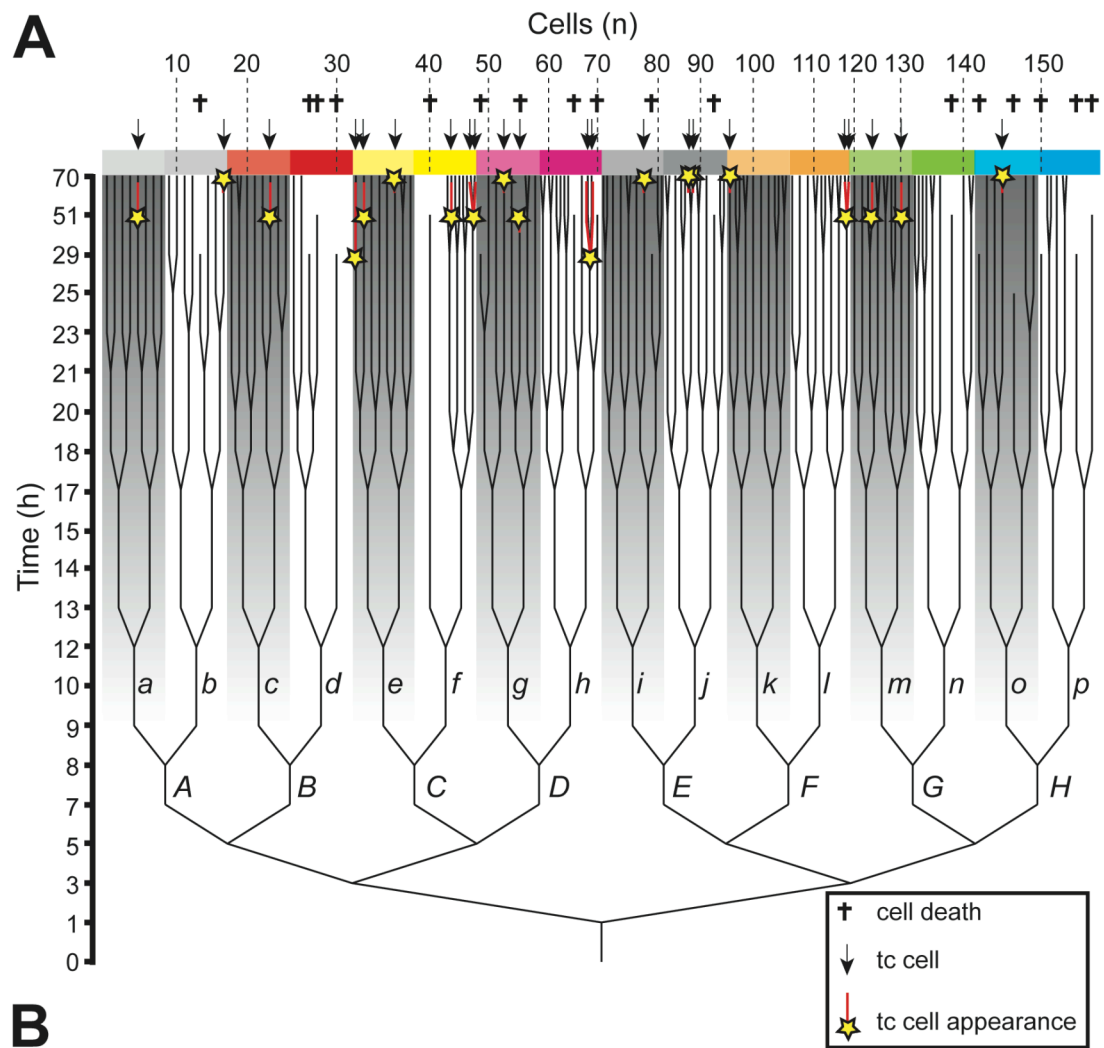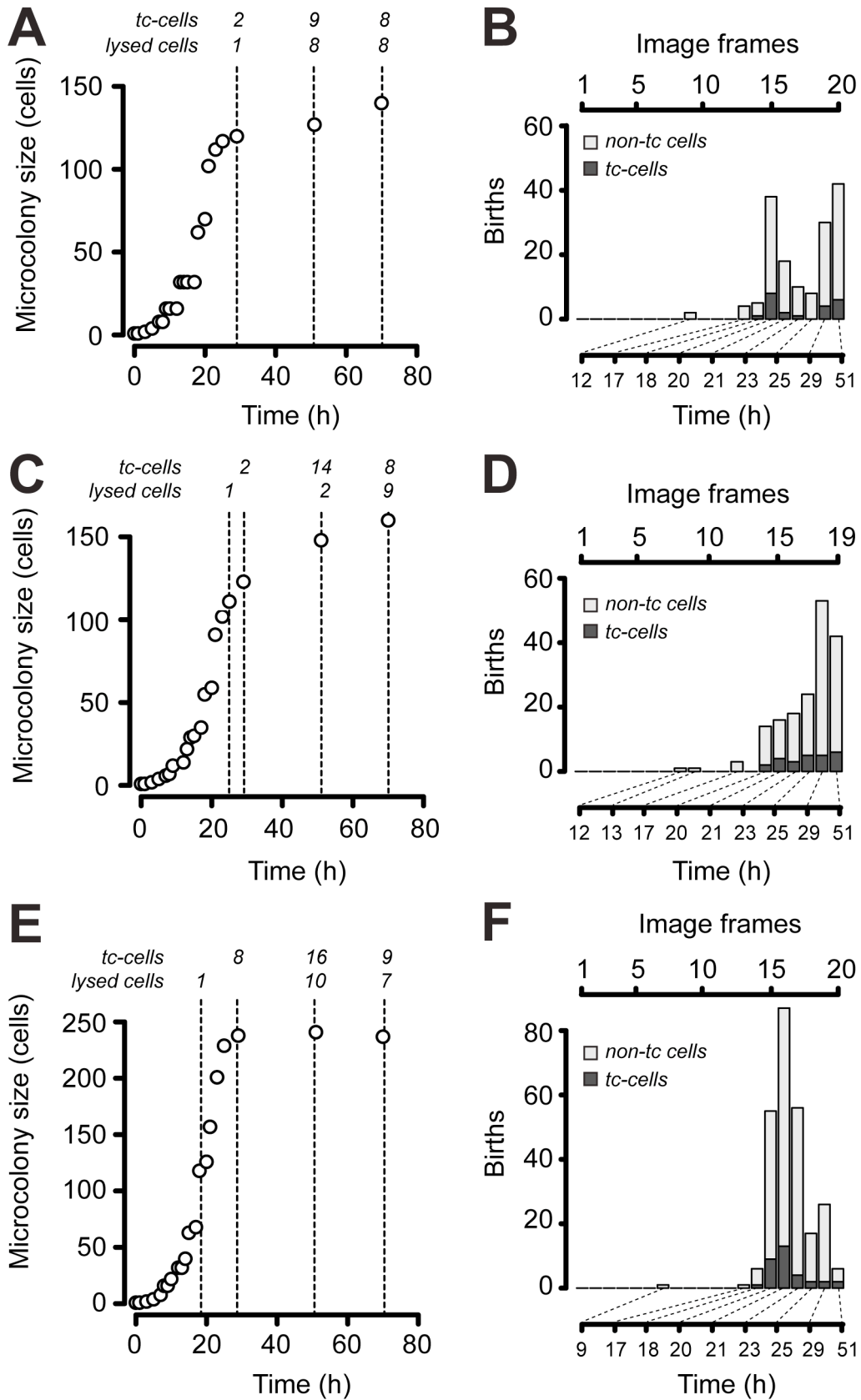
Figure 2 (previous page). Microcolony growth curves and birth times of last generation tc- versus non-tc cells in a microcolony. (A, C, E) Growth curves of microcolonies 1A, 7A and 17F, respectively. Cell lysis events (*lysed*) and number of ICE*clc* tc-cells (*TC*) are shown after 29, 51 and 70 h of growth, respectively (dotted line). (B, D, F) Birth time frequencies of last-generation tc- and non-tc cells in microcolonies 1A, 7A and 17F, respectively. Note how birth-time frequencies of tc- and non-tc cells are similarly distributed, suggesting a normal growth history of ancestors of tc-cells.
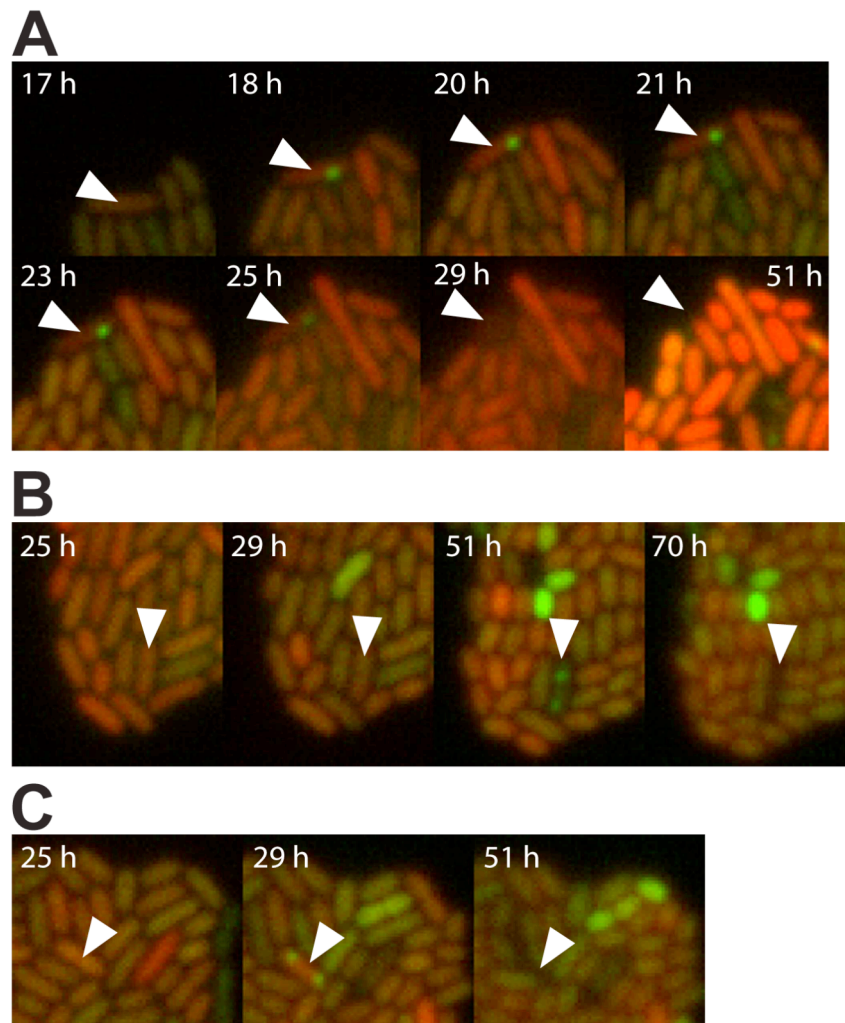
Figure 3. Three examples of cell lysis with preceding green spot development (arrow, A, B, C). Non-tc cells appear as red because of constitutive mCherry expression from the P_{tac} promoter. Tc-cells appear as yellow/green because of constitutive mCherry expression in combination with expression of eGFP from the P_{int} promoter. Note early green spot development (arrow) and subsequent lysis in cells before appearance of ICE*clc* tc-cells in A and C.

A

total number of cells

tc-cells	122		
unpaired		122	
non-tc cells			4414

PI positive (%)

tc-cells	0		
unpaired		0	
non-tc cells			0.5

P. knackmussii **1343**

B

total number of cells

tc-cells	601	111	47
unpaired	601	111	47
non-tc cells	5103	4364	3875

PI positive (%)

tc-cells	3.2	1.8	10.6
unpaired	1.2	0.9	0
non-tc cells	0.7	0.6	0.4

P. putida **2507 2508 2510**

C

total number of cells

tc-cells	136	91	225
unpaired	136	91	225
non-tc cells	4707	5139	8021

ROS positive (%)

tc-cells	41.9	36.3	7.6
unpaired	2.2	0.0	0.0
non-tc cells	0.0	0.0	0.1

P. putida **3014 3015 3016**

Figure 4 (on the left). Significance testing of proportions of PI-positive cells detected among tc-cells (i.e., eGFP positive cells) or non-tc cells in stationary phase suspended batch cultures by resampling data sets via unpairing PI and eGFP fluorescence or ROS and mCherry fluorescence on individual objects. A) *P. knackmussii* B13-1343 cells (n=4536) (single copy insertion of P_{int}-*egfp*). Paired data sets for tc cells (black) or non-tc cells (white). Gray, unpaired resampled data set of tc-cells. Data points above dotted threshold line represent PI-positive cells, proportions of which are presented in the grey box as percentage of all investigated cells. * indicates that the populations are statistically different ($P <$ 0.001) as tested with Student-t test and Wilcoxon test. Note how unpairing results in non-significant differences of tc cell populations compared to non-tc cells. B) Similar but for *P. putida* UWC1-2507 (5704 cells), -2508 (4475 cells) or -2510 (3922 cells), all carrying one copy of ICE*clc* and a single copy insertion of P_{int}-*egfp*. For thresholding in these strains see SI Figures 9 and 10. C) As B) but for *P. putida* UWC1-3014 (4843 cells), -3015 (5230 cells) or -3016 (8246 cells), all carrying one copy of ICE*clc* and a single copy insertion of P_{int}-*eCherry*. For thresholding in these strains see SI Figures 7 and 8. (*) indicates that the populations are statistically different ($P <$ 0.001) only as with Student-t test but not with Wilcoxon test.

Table 1. Cell age[a] distribution in ICE*clc* transfer competent cells of *P. knackmussii* B13.

Age[a]	Observed[b]	Expected[c]	Residuals[d]	P values[e]
1	406	405.50	0.50	0.643, 0.889
2	200	202.75	-2.75	
3	99	101.38	-2.38	
4	54	50.69	3.31	
5	33	25.34	7.66	
6	10	12.67	-2.67	
≥ 7	9	12.67	-3.67	

a) Cell age is referred to as the number of cell divisions the older pole of a cell has experienced during 70 h of microcolony growth.

b) Observed age frequency in ICE*clc* tc-cells (n = 811) from multiple independent microcolonies (n = 69) of similar size (mean = 173 cells, sd = 43 cells) equivalent to 7 generations. tc-cells were identified if their eGFP fluorescence intensity values exceeded a threshold value (X) determined according to $X = Q3 + IQR * 3$, where $Q3$ is the upper quartile and IQR is the interquartile range of eGFP fluorescence intensity values of all cells of a microcolony (see SI Figure 1).

c) The expected age frequency in 811 tc-cells from 69 microcolonies calculated from the theoretical age frequency of cells in a single microcolony with 7 generations under the assumption of equal division rates for each cell and no cell death. We used the formula $E_x = P_x * 811$, where E_x is the expected frequency of tc-cells with age x (x = 1, 2, …,7), and P_x is the probability of occurrence with age x. P_x equals $1/(2^x)$, assuming exponential growth of cells without lysis.

d) Residuals were calculated by subtraction of expected from observed frequencies.

e) *P*-values of the 'goodness-of-fit' test (first value) and Fisher's exact test (second value), performed with observed versus expected values as shown in the table. f) Maximum cell age was 8 cell divisions in some (n = 3) microcolonies.

Table 2. Proportion of ICE*clc* transfer competent and lysed cells of *P. knackmussii* B13 under microcolony growth conditions.

Microcolony	Number of cells[a] after 70 h	tc cells (%)[b]	Lysed cells (%)[c]
1A	140	22 (15.7)	17 (12.1)
7A	160	25 (15.6)	12 (7.5)
17F	237	33 (13.9)	18 (7.6)
Multiple (n = 69)	168 ± 46.3[d]	(6.8 ± 3.3)[d]	

a) Microcolony size determined as the number of visible cells after 70 h of growth on agarose surface (see Figure 1, SI Figure 2-6).

b) Number and percentage of ICE*clc* tc-cells in a microcolony after 70 h.

c) Number and percentage of lysed cells in the microcolony between 0 and 70 h. (see Figures 1A, B, D; SI Figure 2A, C, D; SI Figure 3A, C, D).

d) Average ± standard deviation

Table 3. Cell pole age distribution among tc cells of *P. knackmussii* B13 in three selected microcolonies.

Microcolony	AGE[a]	All cells Observed[b]	tc cells Observed[c]	Expected[d]	Residuals[e]	P values[f]
1A	1	80	8	11.21	-3.21	0.278, 0.666
	2	38	8	5.32	2.68	
	3	17	2	2.38	-0.38	
	4	13	1	1.82	-0.82	
	5+6+7[g]	3+4+2 = 9	1+1+1 = 3	1.26	1.74	
7A	1	89	11	12.94	-1.94	0.400, 0.866
	2	38	8	5.52	2.48	
	3	25	3	3.63	-0.63	
	4	9	0	1.31	-1.31	
	5+6+7+8[g]	4+4+2+1 = 11	2+1+0+0 = 3	1.60	1.40	
17A	1	127	20	16.44	3.56	0.776, 0.931
	2	65	7	8.41	-1.41	
	3	32	3	4.14	-1.14	
	4	15	1	1.94	-0.94	
	5+6+7+8[g]	7+5+2+2 = 16	1+1+0+0 = 2	2.07	-0.07	

a) Cell age is referred to as the number of cell divisions the older pole of a cell has experienced during 70 h of microcolony growth.

b) Observed age frequency in all cells of a microcolony.

c) Observed age frequency in ICE*clc* tc-cells of a microcolony.

d) The expected age frequency in TC cells of a microcolony was assumed to be similar to the observed age frequency in all cells of that microcolony. It was therefore calculated using the formula $E_x = P_x * n_{tc}$, where, per microcolony, E_x is the expected frequency of TC cells with age x (x = 1,2, …), P_x is the probability of occurrences of cells with age x, and n_{tc} is the total number of TC cells. P_x was determined according to $P_x = O_x/n_{tot}$, where, per microcolony, O_x is the observed frequency of cells with age x, and n_{tot} is the total number of cells including lysed cells.

e) Residuals were calculated by subtraction of expected from observed frequencies.

f) *P*-values of the goodness-of-fit (first value) and the Fisher's exact test (second value).

g) Categories are combined to achieve an expected frequency of > 1 as required for reliable goodness-of-fit testing.

Table 4. Cell age distribution in lysed cells of *P. knackmussii* B13 among three selected microcolonies.

Microcolony	AGE[a]	All cells Observed[b]	Lysed cells Observed[c]	Expected[d]	Residuals[e]	P values[f]
1A	1	80	12	8.66	3.34	0.242, 0.562
	2	3 8	1	4.11	-3.11	
	3	17	1	1.84	-0.84	
	4+5+6+7[g]	13+3+4+2 = 22	1+0+1+1 = 3	2.38	0.62	
7A	1	8 9	7	6.21	0.79	0.840, 1.000
	2	3 8	2	2.65	-0.65	
	3	25	1	1.74	-0.74	
	4+5+6+7+8[g]	9+4+4+2+1 = 20	2+0+0+0+0 = 2	1.40	0.60	
17A	1	127	8	8.96	-0.96	0.715, 1.000
	2	65	5	4.59	0.41	
	3	32	3	2.26	0.74	
	4	15	0	1.06	-1.06	
	5+6+7+8[g]	7+5+2+2 = 16	1+1+0+0 = 2	1.13	0.87	

Legend, see Table 3.

Table 5. Lysis in sister cells of tc cells of *P. knackmussii* B13.

Microcolony	Nr. lysed cells[a]	Observed[b]	Expected[e]	Confidence interval[d]
1A	17	0	2.38	0.00, 5.00
7A	12	0	1.95	0.00, 5.00
17F	18	0	2.26	0.00, 6.00

a) Lysis was classified as such if a cell showed reduced mCherry fluorescence intensity and disintegrated cell shape (see Figure 1B).

b) Observed frequency of cell lysis in sister cells of tc-cells. Sister cells were defined as the offspring of a dividing mother cell.

c) Expected frequency of cell lysis in sister cells of tc-cells. The expected frequency of sister cell lysis E_{ls} was calculated by $E_{ls} = P_{ls} * n_{tc}$, where P_{ls} is the probability of lysis and n_{tc} is the number of tc-cells in a microcolony. P_{ls} was obtained from the ratio of lysed cells divided by the total number of cells in a microcolony.

d) 95% Confidence intervals were calculated according to the bootstrap percentile method by bootstrap simulation of sister cell lysis (R = 5000).

Table 6. ICE*clc* transfer competence in sister cells of tc cells of *P. knackmussii* B13.

Microcolony	Nr. tc-cells	Observed	Expected	Confidence interval
1A	22	4	3.08	0.00, 6.00
7A	25	2	4.56	1.00, 8.00
17F	3 3	2	4.02	1.00, 8.00

Legend, see Table 5.

Table 7. Spatial occurrence of tc cells within a microcolony.

Microcolony	Occurrence[a]	Observed[b]	Expected[c]	Residuals[d]	P values[e]
1A	0	6	3.46	2.54	0.042, 0.408
	1	2	5.08	-3.08	
	2	2	3.72	-1.72	
	≥ 3	5	2.74	2.26	
7A	0	3	3.91	-0.91	0.445, 0.892
	1	8	5.74	2.26	
	2	2	4.22	-2.22	
	3	3	2.07	0.93	
	≥ 4	1	1.05	-0.05	
17F	0	7	6.07	0.93	0.244, 0.825
	1	5	8.34	-3.34	
	2	8	5.74	2.26	
	3	4	2.63	1.37	
	≥ 4	0	1.22	-1.22	

a) Occurrences of ICEclc tc-cells within equally sized, randomly placed areas across a microcolony (see Figure 1 and SI Figure 2, 3).

b) Observed frequency of occurrences.

c) Expected frequency of occurrences were obtained assuming a Poisson distribution with a mean (λ) value of $\lambda = n_{tc}/k$ occurrences, where n_{tc} is the total number of tc-cells within a microcolony and k is the number of sampling areas across that colony.

d) Residuals were calculated by subtraction of expected from observed frequencies.

e) P values of the goodness-of-fit test (first value) and Fisher's exact test (second value).

Table 8. Spatial occurrence of cell lysis within a microcolony of *P. knackmussii* B13.

Microcolony	Occurrence[a]	Observed[b]	Expected[c]	Residuals[d]	P values[e]
1A	0	3	4.83	-1.83	0.353, 0.763
	1	8	5.47	2.53	
	2	3	3.10	-0.10	
	≥ 3	1	1.60	-0.60	
7A	0	8	8.39	-0.39	0.986, 1.000
	1	6	5.92	0.08	
	≥ 2	3	2.68	0.32	
17F	0	9	11.34	-2.34	0.353, 0.669
	1	12	8.50	3.50	
	≥ 2	3	4.16	-1.16	

Legend, see Table 7

Table 9. ICE*clc* transfer competence across 8 lineage clusters (A-H) in a microcolony of *P. knackmussii* B13.

Microcolony	Cluster·(cells)[a]	Observed[b]	Expected[c]	Residuals[d]	*P* values[e]
1A	*A* (16)	2	2.24	-0.24	0.459, 0.966
	B (14)	1	1.96	-0.96	
	C (18)	6	2.52	3.48	
	D (22)	4	3.08	0.92	
	E (24)	3	3.36	-0.36	
	F (25)	3	3.50	-0.50	
	G (23)	2	3.22	-1.22	
	H (15)	1	2.10	-1.10	
7A	*A* (31)	5	4.51	0.49	0.469, 0.857
	B (24)	7	3.49	3.51	
	C (27)	5	3.92	1.08	
	D (31)	2	4.51	-2.51	
	E (21)	2	3.05	-1.05	
	F (17)	1	2.47	-1.47	
	G (9)	1	1.31	-0.31	
	H (12)	2	1.74	0.26	
17F	*A* (33)	5	4.27	0.73	0.457, 0.766
	B (25)	0	3.24	-3.24	
	C (35)	6	4.53	1.47	
	D (32)	5	4.14	0.86	
	E (32)	6	4.14	1.86	
	F (33)	2	4.27	-2.27	
	G (35)	6	4.53	1.47	
	H (30)	3	3.88	-0.88	

a) Clusters (*A*, *B*, …, *H*) are groups of related cells in a microcolony that are defined in the lineage trees of Figures 1, SI Figure 2, 3. Further legend details, see Table 7-8.

Table 10. Cell lysis across 8 lineage clusters (A-H) in a microcolony of *P. knackmussii* B13.

Microcolony	Cluster (cells)[a]	Observed[b]	Expected[c]	Residuals[d]	P values[e]
1A	*A* (16)	1	1.73	-0.73	0.050, 0.584
	B (14)	3	1.52	1.48	
	C (18)	1	1.95	-0.95	
	D (22)	4	2.38	1.62	
	E (24)	2	2.60	-0.60	
	F (25)	0	2.71	-2.71	
	G (23)	1	2.49	-1.49	
	H (15)	5	1.62	3.38	
7A	*A* (31)	0	2.16	-2.16	0.265, 0.614
	B (24)	2	1.67	0.33	
	C (27)	3	1.88	1.12	
	D (31)	0	2.16	-2.16	
	E (21)	2	1.47	0.53	
	F (17)	2	1.19	0.81	
	G + H (9 + 12)[f]	3	1.47	1.53	
17F	*A* (33)	3	2.33	0.67	0.527, 0.923
	B (25)	2	1.76	0.24	
	C (35)	2	2.47	-0.47	
	D (32)	3	2.26	0.74	
	E (32)	5	2.26	2.74	
	F (33)	1	2.33	-1.33	
	G (35)	1	2.47	-1.47	
	H (30)	1	2.12	-1.12	

f) Categories are combined to achieve an expected frequency of > 1 as required for reliable goodness-of-fit testing.

Further legend details, see Table 9.

Table 11. ICE*clc* transfer competence across 16 lineage clusters (a-p) in a microcolony of *P. knackmussii* B13.

Microcolony	Sub-cluster (cells)[a]	Observed[b]	Expected[c]	Residuals[d]	P values[e]
1A	*a* (8)	1	1.12	-0.12	0.946, 0.995
	b (8)	1	1.12	-0.12	
	c (9)	1	1.26	-0.26	
	d + e (5 + 9)[f]	3	1.96	1.04	
	f (9)	3	1.26	1.74	
	g (9)	2	1.26	0.74	
	h (13)	2	1.82	0.18	
	i (11)	1	1.54	-0.54	
	j (13)	2	1.82	0.18	
	k (12)	1	1.68	-0.68	
	l (13)	2	1.82	0.18	
	m (12)	2	1.68	0.32	
	n (11)	0	1.54	-1.54	
	o + p (7 + 8)[f]	1	2.10	-1.10	
7A	*a* (15)	1	2.18	-1.18	0.253, 0.930
	b (16)	4	2.33	1.67	
	c (13)	3	1.89	1.11	
	d (11)	4	1.60	2.40	
	e (12)	0	1.74	-1.74	
	f (15)	5	2.18	2.82	
	g (15)	1	2.18	-1.18	
	h (16)	1	2.33	-1.33	
	l (11)	2	1.60	0.40	
	j (10)	0	1.45	-1.45	
	k (10)	0	1.45	-1.45	
	l (7)	1	1.02	-0.02	
	m (8)	1	1.16	-0.16	
	n +o + p (1 + 4 + 8)[f]	2	1.89	0.11	
17F	*a* (17)	2	2.20	-0.20	0.882, 0.996
	b (16)	3	2.07	0.93	
	c (9)	0	1.16	-1.16	
	d (16)	0	2.07	-2.07	
	e (16)	3	2.07	0.93	
	f (19)	3	2.46	0.54	
	g (16)	2	2.07	-0.07	
	h (16)	3	2.07	0.93	
	i (16)	3	2.07	0.93	
	j (16)	3	2.07	0.93	
	k (17)	1	2.20	-1.20	
	l (16)	1	2.07	-1.07	
	m (19)	2	2.46	-0.46	

Table 11-*Continued*

Microcolony	Cluster (cells)[a]	Observed[b]	Expected[c]	Residuals[d]	P values[e]
	n (16)	4	2.07	1.93	
	o (14)	1	1.81	-0.81	
	p (16)	2	2.07	-0.07	

Legend, see Table 10

DISCUSSION

One of the mysteries in ICE gene transfer among bacteria is the mechanism that controls the typically low frequency by which ICE*clc* becomes excised in clonally identical populations of donor cells. Indeed, our previous results on ICE*clc* in *P. knackmussii* B13 using stable fluorescent reporter gene fusions at single-cell level indicated that 3-5% of cells in stationary phase suspended culture after growth on 3CBA as sole carbon and energy source measurably express the P_{inR} and P_{int} promoters (15, 25). More recently, we could demonstrate that cells that activate P_{int} are those that initiate a program for transfer competence, through which they become able to transfer ICE*clc* (22). In addition, we showed that the proportion of cells expressing P_{inR} and P_{int} depends on RpoS levels, with cells having higher than population-average RpoS levels showing a higher likelihood of activating the tc-cascade (16). These results thus suggested that the decision to activate or not P_{inR} and P_{int} in an individual cell is a stochastic process. However, apparent stochasticity (e.g., the distribution of RpoS among cells) may have underlying causes, such as, some cells being in a slightly different growth phase or having experienced more previous biochemical damage, and thus producing more RpoS. The specific hypotheses we thus wanted to test in this work were whether ICE*clc* activation in individual cells was dependent on cell pole age, whether it was confined to specific cellular lineages, or occurred as a consequence of pre-existing detectable biochemical damage. We also tested whether tc cell formation occurs spatial randomly within a microcolony. In essence, none of the hypotheses is supported by the experimental observations, from which we conclude that ICE*clc* activation is a random process depending on stochastic variation among individual cells.

Most of the experiments here were based on time-lapse imaging of individual cells in growing microcolonies. This has the advantage of tracing the behaviour and life history of all cells throughout exponential and stationary phases, when the colony stops growing and ICE*clc* tc-formation is initiated. On the other hand, to avoid that the microcolony became multilayered and individual cells could no longer be traced accurately, we had to restrict the amount of carbon available to cells. As a consequence of this, only a limited number of generations (7-8) could be followed before the colony entered stationary phase. Therefore, there is a relatively low number of cells with 'old' poles within the demographic distribution of the clonal population in the colony. Overall, however, there was no statistical support to conclude that old cells would be more prone to activating ICE*clc* (Table 1). In fact, the finding that ICE*clc* activation events occur as frequently among cells with poles of 1 generation than in any other category suggests that cell age is not a determining factor for the activation process, and that designing

the experiment differently to observe more older cells (e.g., by a mother machine (30)) would not have changed the outcome.

Along similar lines we tested whether occurrence of tc-cell activation is confined to specific cell lineages that develop over time, but there was no statistical evidence to support that hypothesis (Tables 9-11). This suggests that in case of ICE*clc* there is no pre-existing epigenetic modification, as was hypothesized for the case of sporulation in *Bacillus subtilis* (29), that would determine groups of cells from particular lineages to all activate ICE*clc*. Interestingly, although the timing of tc-cell appearance within the microcolony was relatively coherent (i.e., within 29 to 70 h, at the onset of stationary phase, Figure 2), their position was statistically speaking random (Table 7). One could imagine that positionally random activation has a selective advantage for a conjugative element that attempts to gain access to new host cells (Figure 5). Totally random appearance of ICE*clc* tc-cells within a microcolony is in major contrast to other examples of bistability phenomena. For example, *B. subtilis* cells expressing sporulation promoters (P_{spoIIA}-*gfp* and P_{abrB}-*gfp*) cluster rather in particular cell lineages (29).

Finally, we tested whether ICE*clc* tc-cells appear as a consequence of previously experienced biochemical damage. One indication for this could be the frequency of cell lysis in sister cells of tc-cells. However, there was no statistical evidence for increased frequency of tc-cell appearance among sisters of lysed cells (Table 5). Also, some tc-cells still divided after initiation of the tc-state but their timing of birth was no different than non-tc cells. This would indicate that on average tc-cells are not more damaged than non-tc cells. The only indication for a possible pre-existing biochemical damage came from staining of stationary phase suspended cultures. This showed that tc-cells on average stain brighter with PI than non-tc cells. At least for *P. putida* UWC1 carrying ICE*clc* the tc-cells accumulate more ROS than non-tc cells. Now since these stainings were performed on a single time sample, we cannot discern whether tc-initiation occurs because of damage (indicative higher PI and ROS staining), or whether damage occurs as a consequence of tc-initiation. To differentiate between those might be an interesting avenue to pursue specifically by using tools that would allow identification over time of the tc-state of cells and of reactive oxygen species presence. One possibility here would be reporter gene fusions to promoters in *P. knackmussii* or *P. putida* reactive to oxygen radicals, which, however, so far have not been identified.

The hypothesis we have entertained recently is that ICE*clc* transfer competence is initiated in stationary phase cells, but that actual ICE transfer will need subsequent energy for cells (22). In this respect it is interesting that microcolonies in stationary phase have several incidences of lysing (non-tc) cells, the released nutrients of which could possibly benefit tc-cells. Since ICE*clc* activation and lysis occurs spatially randomly throughout the microcolony

"biofilm" a small amount of nutrients would become available, which might not be sufficient to renew cell division but might provide enough energy to stimulate ICE*clc* transfer. For example, in cannibalism in *B. subtilis*, it is thought that cell lysis provides food to the non-cannibalized fraction of cells delaying the one-way process to the formation of the spores (12). Finally, spatially random occurrence of ICE*clc* tc-cells might have been selected for because it would allow a higher chance of contacting possible recipient cells and thus potentially more efficient ICE transfer. This, together with the postulated beneficial effects on transfer of outgrowth of tc-cells to small groups of tc-cells (22), would then add to the strategy of ICE*clc* to maximize its horizontal transmission rates from a proportionally small number of tc-cells arising in the population strains. It is also possible that random lysis throughout a biofilm might somehow play a significant role in overall biofilm stability and structure, perhaps through the release of structural, and signalling molecules from lysed cells, or the formation of gaps, which would add to the "porosity" of a biofilm potentially affecting medium flow and nutrient circulation. This could then be regarded as self-destructive cooperation strategy, whereby the death of one cell helps others, benefiting public good, in this case a favourable biofilm environment (2).
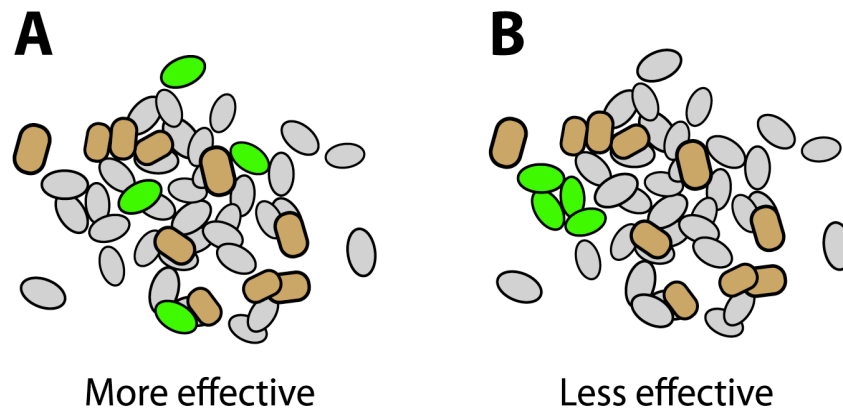
Figure 5. Conceptual idea showing transfer probability as a function of spatial arrangement of transfer competent cells. Higher probability of donor-to-recipient cell-to-cell contact (and therefore more effective transfer) when donor cells appear randomly distributed throughout the donor microcolony (A) as opposed to when they appear clustered (B). Green, transfer competent cells. Grey, donor strain microcolony. Brown, recipient cells.

MATERIALS AND METHODS

Strains and Culture Media. The strains used in this work were derivatives of *P. knackmussii* B13 and *P. putida* UWC1 (ICE*clc*), which carry single copy transposon insertions with reporter gene fusions between the P_{int}-promoter and *egfp* or *mCherry* (22). Strains were revived by streaking from -80°C glycerol stocks onto nutrient agar, where necessary supplemented with kanamycin at 50 µg/ml. Single colonies were picked and cultured in five ml of type 21C minimal medium (MM) (11) containing 4 mM 3-CBA until stationary phase (24 – 96 h) at 30°C and at 180 rpm agitation. Stationary phase cultures were recovered and diluted before being inoculated onto agarose nutrient surfaces to initiate microcolony growth. Agarose nutrient patches included 0.1 mM 3-CBA as sole (added) carbon and energy source, but cells also utilize traces of available organic carbon source in the agarose.

Microcolony Growth Procedure. To initiate and follow microcolony development in real-time we used a perfusion chamber setup as described previously, in which the microcolony forms in between the surface of a glass slide and the nutrient surface (22, 23). With the used culture dilutions, medium components and concentrations, microcolonies from non-tc starter cells developed to stationary phase in about 24 hours and reached a size of 160-200 cells.

Microscopy. Cells in microcolonies were examined with an inverted Leica AF6000 epifluorescence microscope (Leica Microsystems Heidelberg GmbH). Images were taken with a DFC320 monochrome camera (Leica Microsystems GmbH) and a 100/1.30 oil immersion lens (HCX PL FLUOTAR; Leica), at an exposure time of 800 ms. Filters used for eGFP and ROS-stain was GFP BP470/40, and for PI and mCherry Y3 BP535/50 (Leica). Images were digitally recorded as 16-bit TIFF-files using the Leica AF6000 software.

Time-lapse experiments were facilitated by the use of a mark-and-find motorized stage (Märzhäuser Wetzlar, Germany). A random region on the agarose patch was chosen and single cells were focused under phase contrast at 1000 × magnification. Microcolony formation was then followed with images taken at roughly 1-h intervals until 70 h after start of the experiment and until cells appeared in the microcolonies which had activated the P_{int}-promoter (22). Individual cells were considered to have activated ICE*clc* when the fluorescence level of the reporter protein expressed from P_{int} was higher than a value of the Q3 (upper quartile) + 3 times the interquartile range (SI Figure 1, (22)).

Image Analyses. Original images consisted of 16-bit TIFF-files, which were used for fluorescence intensity measurements in METAMORPH (version 6.1r5, Visitron Systems, Germany). METAMORPH was also used for manual counting of cells to determine microcolony

sizes, for artificial colouring of 'red' (mCherry) or 'green' (for eGFP), for creating overlays of 'red' (mCherry) and 'green' (for eGFP), for creating time-lapse image stacks, alignment of these stacks and subsequent creation of time-lapse movies and for conversion to 8 bit files for subsequent processing. 8 Bit images were cropped to the final resolution and image size using ADOBE PHOTOSHOP CS4 (Adobe Inc.). mCherry fluorescence images for lineage tree analyses were prepared in GRAPHIC CONVERTER X (version 6.0.4 1386, Lemke Software GmbH, Germany) by subsequently carrying out inversion, changing resolution from 300 ppi to 600 ppi, and manually adjusting levels.

Lineage Tree Analyses. Time-lapse microscopy pictures were aligned using METAMORPH to create image stacks. Aligned images were then analyzed with the MATLAB-based program SCHNITZCELL (24) with extended MATLAB scripts by Bergmiller (5) to extract cell pole age. SCHNITZCELL was used to segment the images, identify cells, and track cells over consecutive images. Time-lapse videos of growing microcolonies were created from aligned image stacks in METAMORPH, which were used in parallel to the SCHNITZCELL processing. This manual frame-by-frame cell tracking was employed to verify correct cell lineage mapping.

Staining. To stain for membrane damage we used propidium iodide (PI) at a final concentration of 30 μM. As a general ROS marker, cells were stained with dichlorodihydrofluorescein diacetate (H2DCFDA) at 20 μg/ml. Hereto, one μl of an appropriate stock solution (PI, 20 mM; H2DCFDA, 10 mM in DMSO) was added to 500 μl cell suspension and incubated for 30 min (PI) or 2 h (H2DCFDA) in the dark, after which cells were examined by epifluorescence microscopy.

The threshold fluorescence for being assigned a PI-positive cell was identified as the minimum PI fluorescence of isopropanol treated stained cells (22). The threshold for being categorized as a ROS-positive cell was identified as the minimum fluorescence among H_2O_2-treated and H2DCFDA-stained cells as described in SI Figures 7-10 and Reinhard and colleagues(22).

Calculations. Data analysis and statistics were performed using the program R (Ihaka and Gentleman, 1996; http://cran.r-project.org).

ACKNOWLEDGEMENTS

REFERENCES

1. **Ackermann, M., S. C. Stearns, and U. Jenal.** 2003. Senescence in a bacterium with asymmetric division. Science **300:**1920.

2. **Ackermann, M., B. Stecher, N. E. Freed, P. Songhet, W. D. Hardt, and M. Doebeli.** 2008. Self-destructive cooperation mediated by phenotypic noise. Nature **454:**987-90.

3. **Aldridge, B. B., M. Fernandez-Suarez, D. Heller, V. Ambravaneswaran, D. Irimia, M. Toner, and S. M. Fortune.** 2012. Asymmetry and aging of mycobacterial cells lead to variable growth and antibiotic susceptibility. Science **335:**100-104.

4. **Beaber, J. W., B. Hochhut, and M. K. Waldor.** 2004. SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature **427:**72-74.

5. **Bergmiller, T., and M. Ackermann.** 2011. Pole age affects cell size and the timing of cell division in *Methylobacterium extorquens* AM1. J Bacteriol **193:**5216-5221.

6. **Bose, B., J. M. Auchtung, C. A. Lee, and A. D. Grossman.** 2008. A conserved anti-repressor controls horizontal gene transfer by proteolysis. Mol Microbiol **70:**570-852.

7. **Bose, B., and A. D. Grossman.** 2011. Regulation of horizontal gene transfer in Bacillus subtilis by activation of a conserved site-specific protease. J Bacteriol **193:**22-29.

8. **Davidson, C. J., and M. G. Surette.** 2008. Individuality in bacteria. Annu Rev Genet **42:**253-68.

9. **de Pedro, M. A., J. C. Quintela, J. V. Holtje, and H. Schwarz.** 1997. Murein segregation in Escherichia coli. J Bacteriol **179:**2823-2834.

10. **Gaillard, M., T. Vallaeys, F. J. Vorholter, M. Minoia, C. Werlen, V. Sentchilo, A. Puhler, and J. R. van der Meer.** 2006. The *clc* element of *Pseudomonas* sp. strain B13, a genomic island with various catabolic properties. J Bacteriol **188:**1999-2013.

11. **Gerhardt, P., R. G. E. Murray, R. N. Costilow, E. W. Nester, W. A. Wood, N. R. Krieg, and G. Briggs Phillips (ed.).** 1981. Manual of methods for general bacteriology. American Society for Microbiology, Washington, D.C.

12. **González-Pastor, J. E., E. C. Hobbs, and R. Losick.** 2003. Cannibalism by sporulating bacteria. Science **301:**510-513.

13. **Lindner, A. B., R. Madden, A. Demarez, E. J. Stewart, and F. Taddei.** 2008. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. Proc Natl Acad Sci U S A **105:**3076-3081.

14. **Locke, J. C., and M. B. Elowitz.** 2009. Using movies to analyse gene circuit dynamics in single cells. Nat Rev Microbiol **7:**383-92.

15. **Minoia, M., M. Gaillard, F. Reinhard, M. Stojanov, V. Sentchilo, and J. R. van der Meer.** 2008. Stochasticity and bistability in horizontal transfer control of a genomic island in *Pseudomonas*. Proc. Natl. Acad. Sci. U S A **105:**20792-7.

16. **Miyazaki, R., M. Minoia, N. Pradervand, S. Sulser, F. Reinhard, and J. R. van der Meer.** 2012. Cellular variability of RpoS expression underlies subpopulation activation of an integrative and conjugative element. PLoS Genet **8:**e1002818.

17. **Ping, L., B. Weiner, and N. Kleckner.** 2008. Tsr-GFP accumulates linearly with time at cell poles, and can be used to differentiate 'old' versus 'new' poles, in *Escherichia coli*. Mol Microbiol **69:**1427-1438.

18. **Ptashne, M.** 2011. Principles of a switch. Nature chemical biology **7:**484-487.

19. **Ramsay, J. P., A. S. Major, V. M. Komarovsky, J. T. Sullivan, R. L. Dy, M. F. Hynes, G. P. Salmond, and C. W. Ronson.** 2013. A widely conserved molecular switch controls quorum sensing and symbiosis island transfer in Mesorhizobium loti through expression of a novel antiactivator. Mol Microbiol **87:**1-13.

20. **Ramsay, J. P., J. T. Sullivan, N. Jambari, C. A. Ortori, S. Heeb, P. Williams, D. A. Barrett, I. L. Lamont, and C. W. Ronson.** 2009. A LuxRI-family regulatory system controls excision and transfer of the Mesorhizobium loti strain R7A symbiosis island by activating expression of two conserved hypothetical genes. Mol Microbiol **73:**1141-1155.

21. **Ramsay, J. P., J. T. Sullivan, G. S. Stuart, I. L. Lamont, and C. W. Ronson.** 2006. Excision and transfer of the Mesorhizobium loti R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. Mol Microbiol **62:**723-734.

22. **Reinhard, F., R. Miyazaki, N. Pradervand, and J. R. van der Meer.** 2013. Cell differentiation to "mating bodies" induced by an integrating and conjugative element in free-living bacteria. Curr Biol **23:**255-9.

23. **Reinhard, F., and J. R. van der Meer.** 2010. Microcolony growth assays p. 3562-3570. *In* K. N. Timmis, V. de Lorenzo, T. McGenity, and J. R. van der Meer (ed.), Handbook of Hydrocarbon and Lipid Microbiology, vol. 5. Springer Verlag.

24. **Rosenfeld, N., J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz.** 2005. Gene regulation at the single-cell level. Science **307:**1962-1965.

25. **Sentchilo, V., R. Ravatn, C. Werlen, A. J. Zehnder, and J. R. van der Meer.** 2003. Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. Strain B13. Journal of Bacteriology **185:**4530-8.

26. **Silander, O. K., N. Nikolic, A. Zaslaver, A. Bren, I. Kikoin, U. Alon, and M. Ackermann.** 2012. A genome-wide analysis of promoter-mediated phenotypic noise in *Escherichia coli*. PLoS Genet **8:**e1002443.

27. **Stewart, E. J., R. Madden, G. Paul, and F. Taddei.** 2005. Aging and death in an organism that reproduces by morphologically symmetric division. PLoS Biol **3:**e45.

28. **Veening, J. W., W. K. Smits, and O. P. Kuipers.** 2008. Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol **62:**193-210.

29. **Veening, J. W., E. J. Stewart, T. W. Berngruber, F. Taddei, O. P. Kuipers, and L. W. Hamoen.** 2008. Bet-hedging and epigenetic inheritance in bacterial cell development. Proc Natl Acad Sci U S A **105:**4393-8.

30. **Wang, P., L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun.** 2010. Robust growth of *Escherichia coli*. Curr Biol **20:**1099-103.

31. **Winkler, J., A. Seybert, L. Konig, S. Pruggnaller, U. Haselmann, V. Sourjik, M. Weiss, A. S. Frangakis, A. Mogk, and B. Bukau.** 2010. Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. Embo J **29:**910-923.

257

ANNEX: SUPPORTING INFORMATION
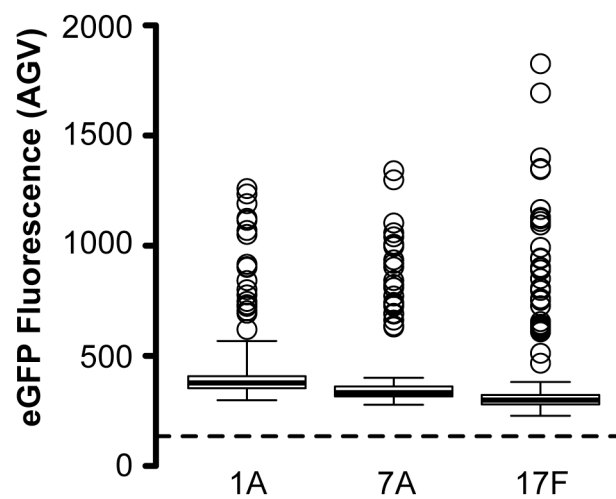
SI Figure 1. Classification of ICE*clc* transfer competent cells. Boxplots summarize eGFP fluorescence intensities in cells of microcolonies 1A (n = 157), 7A (n = 172) and 17F (n = 255). White circles represent outliers and correspond to the eGFP intensities of ICE*clc* transfer competent cells. Outliers were classified as values higher than a value of the Q3 (upper quartile) + 3 times the interquartile range. The dotted line represents background maximum fluorescence.
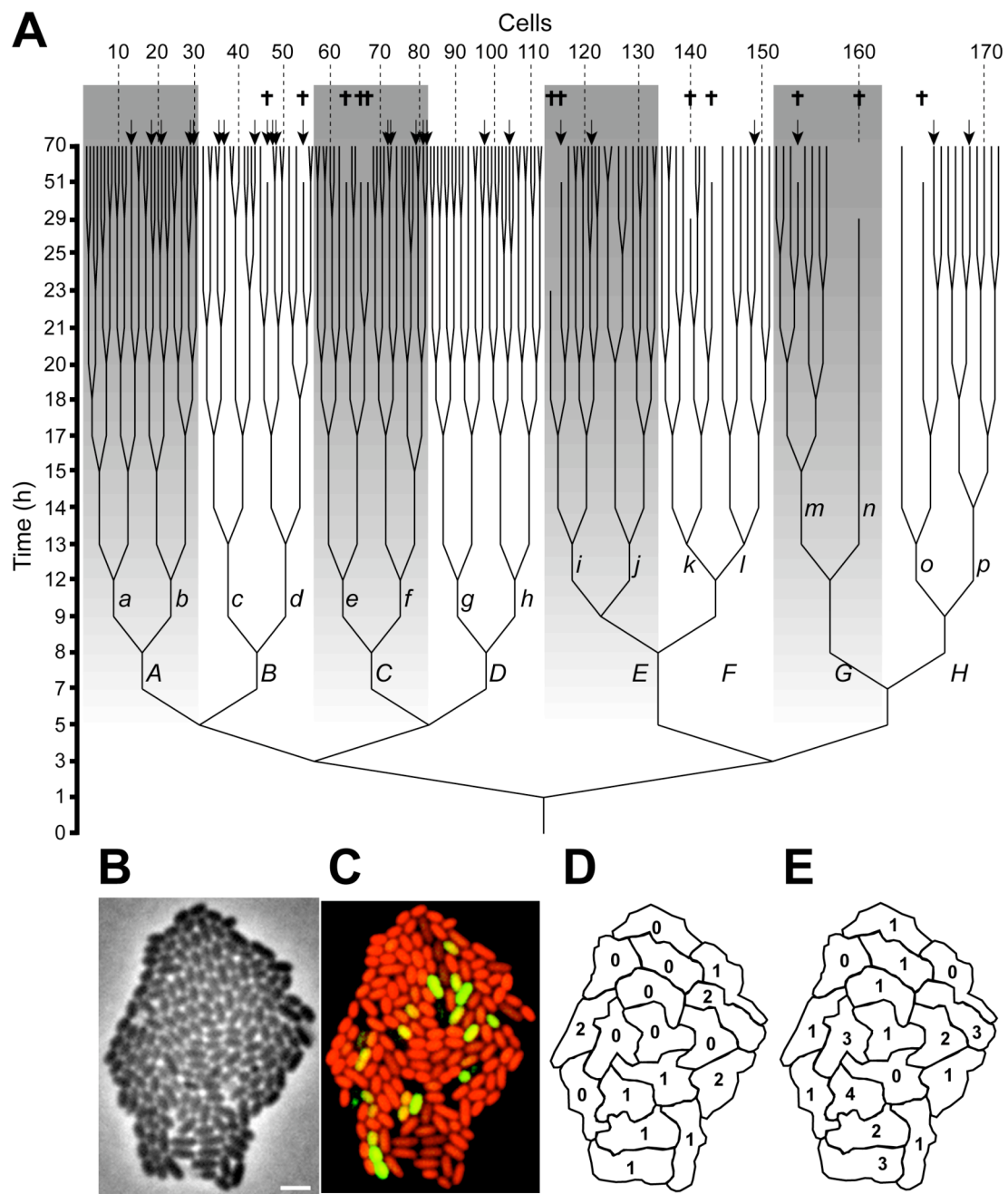
SI Figure 2 (previous page). Random and Independent Appearance of ICE*clc* Transfer Competence and Lysis during late Microcolony Growth. (A) Lineage tree of 172 cells of microcolony 7A. Arrows point to cells (tree tips) with above-threshold eGFP fluorescence intensities (SI Figure 1) and classified as ICE*clc* transfer competent (TC) cells. Crosses point to lysed cells (shorter tree tips). Capital letters (*A*, *B*, …, *H*) name lineage tree clusters (see Table 9, 10). Small letters (*a*, *b*, …, *p*) name lineage tree sub-clusters (see Table 11). (B) Phase contrast visualisation of microcolony 7A after 70 hours of growth. Note absence of cell shapes of lysed cells. White bar indicates 2 µm. (C) Overlay of green and red fluorescence of microcolony 7A after 70 hours of growth. Non-TC cells appear as red because of constitutive mCherry expression from the P_{tac} promoter while below-threshold (see SI Figure 1) expression of eGFP from the P_{int} promoter. TC cells appear as yellow because of constitutive mCherry expression from the P_{tac} promoter while above-threshold (see SI Figure 1) expression of eGFP from the P_{int} promoter. Note absence of fluorescence in lysed cells. (D) Region map of microcolony 7A after 70 hours of growth showing randomly defined areas with respective occurrences of lysed cells. (E) Region map of microcolony 7A after 70 hours of growth showing randomly defined areas with respective occurrences of ICE*clc* competent cells.
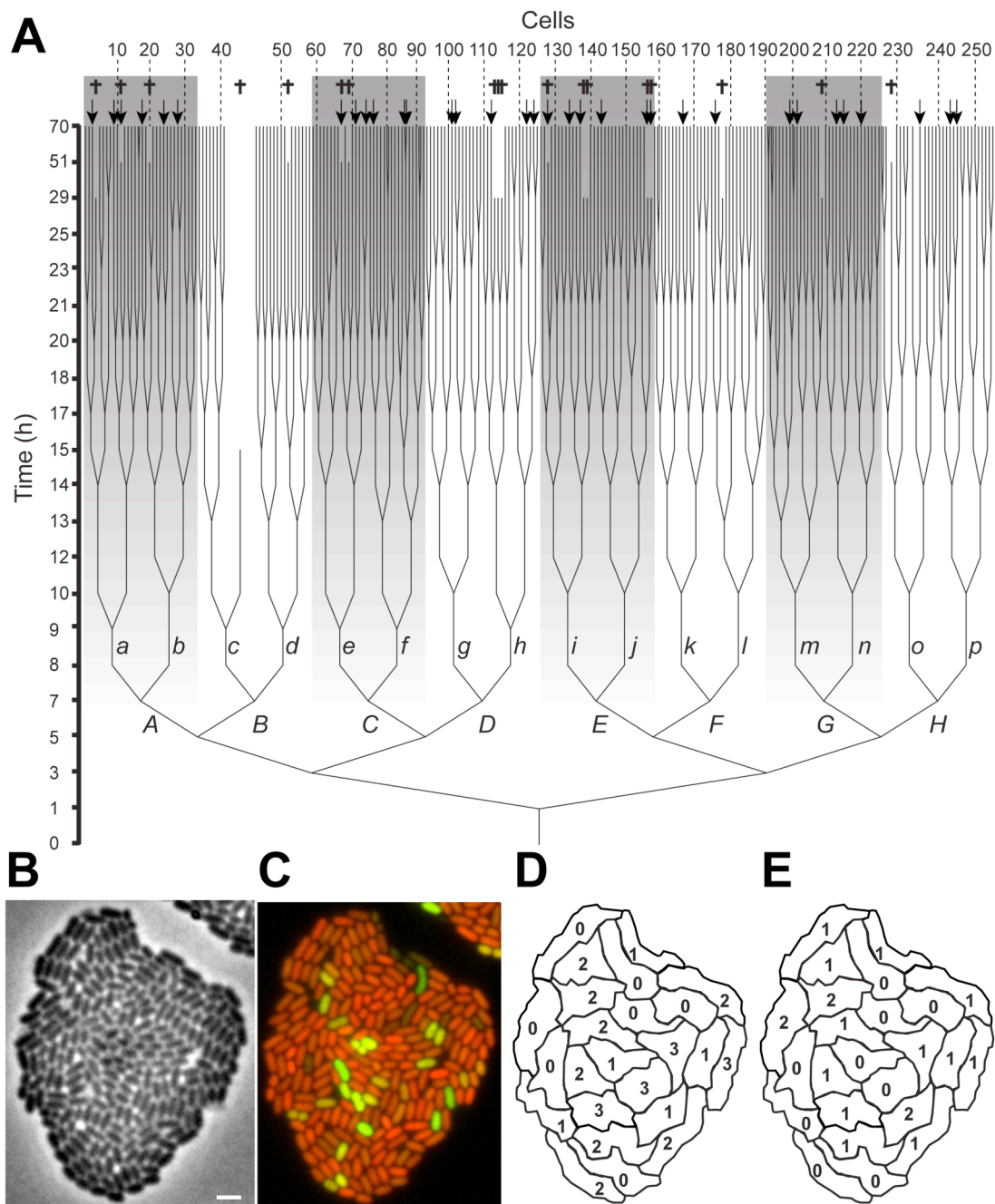
SI Figure 3 (previous page). Random and Independent Appearance of ICE*clc* Transfer Competence and Lysis during late Microcolony Growth. (A) Lineage tree of 255 cells of microcolony 17F. Further legend description, see SI Figure 2.
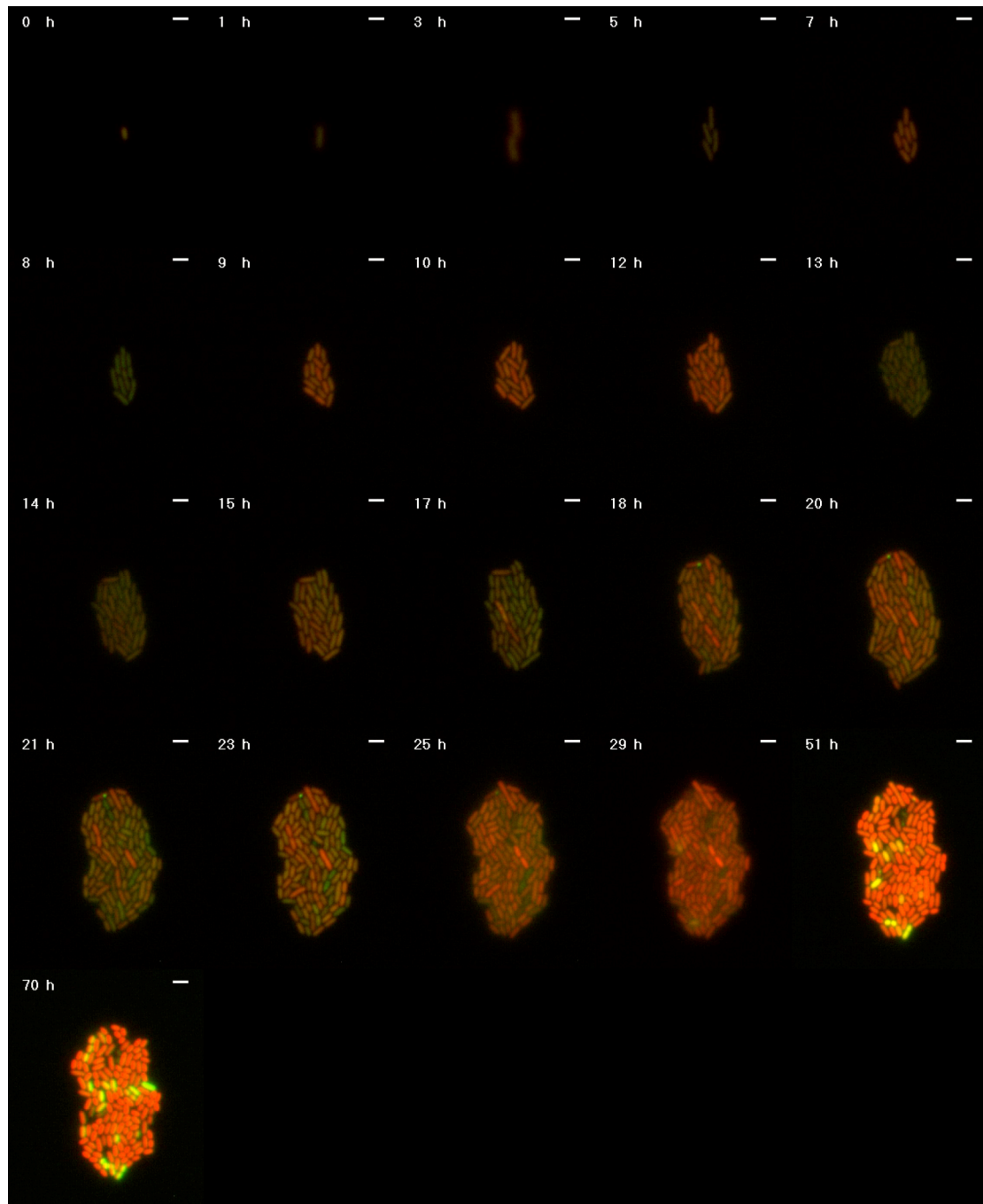
SI Figure 4. Growing *Pseudomonas knackmussii* B13 microcolony 1A. This film shows 70 hours (21 frames) of the growth of a *P. knackmussii* B13 microcolony harbouring the P_{int}-*egfp* and P_{tac}-*egfp* constructs. For all frames, images were taken at the time as indicated at the top left corner. The complete lineage history of the entire microcolony from the single initial cell to all descendents at 70 hours has been tracked and recorded, allowing pole ages and cell fates to be assigned to every cell. The data derived from this microcolony were used for Figure 1, and Tables 2-11. White bar indicates 2 μm.
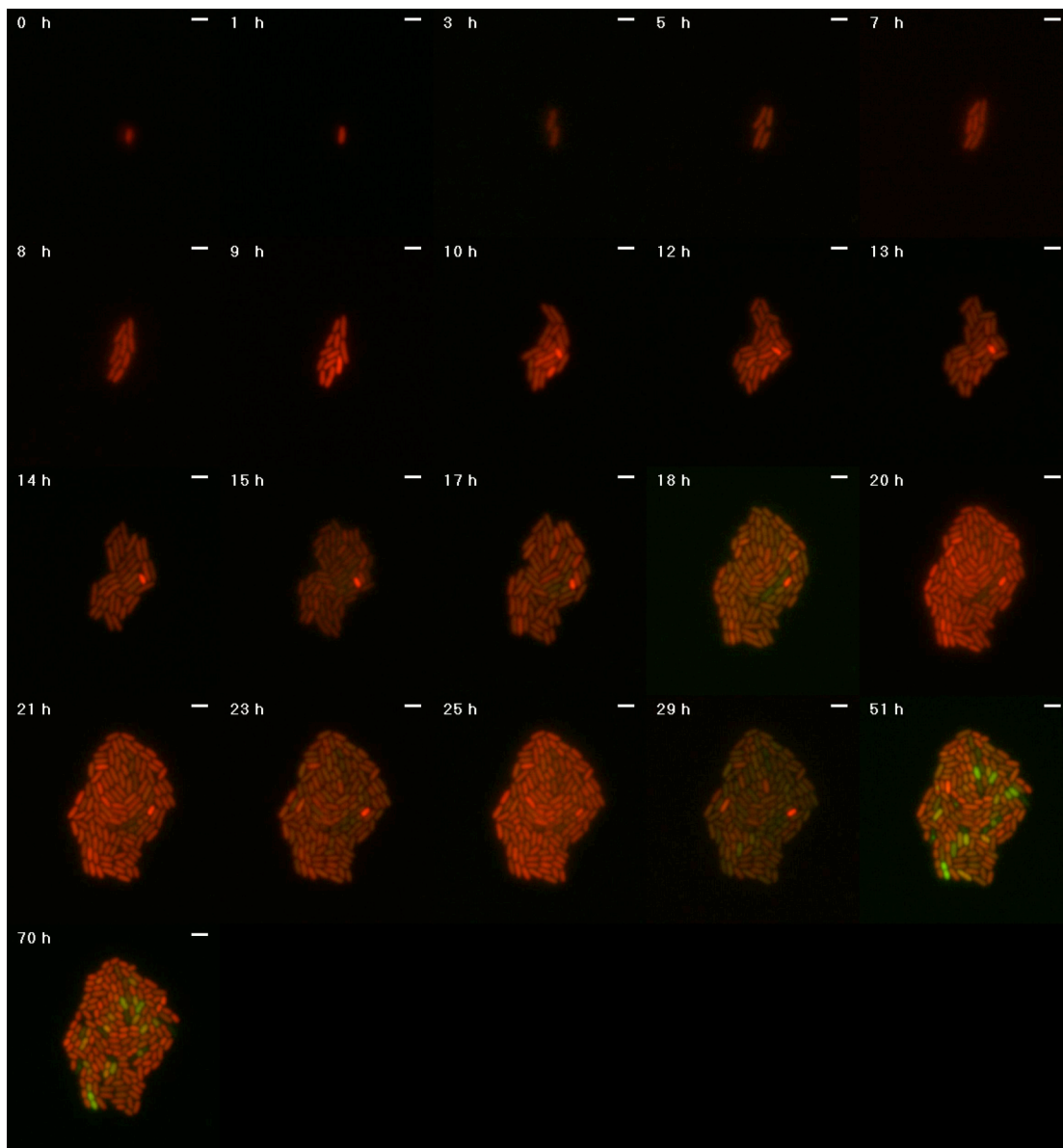
SI Figure 5. Growing *Pseudomonas knackmussii* B13 microcolony 7A. The data derived from this microcolony were used for SI Figure 2, and Tables 2-11. White bar indicates 2 μm. Further legend description, see SI Figure 4.
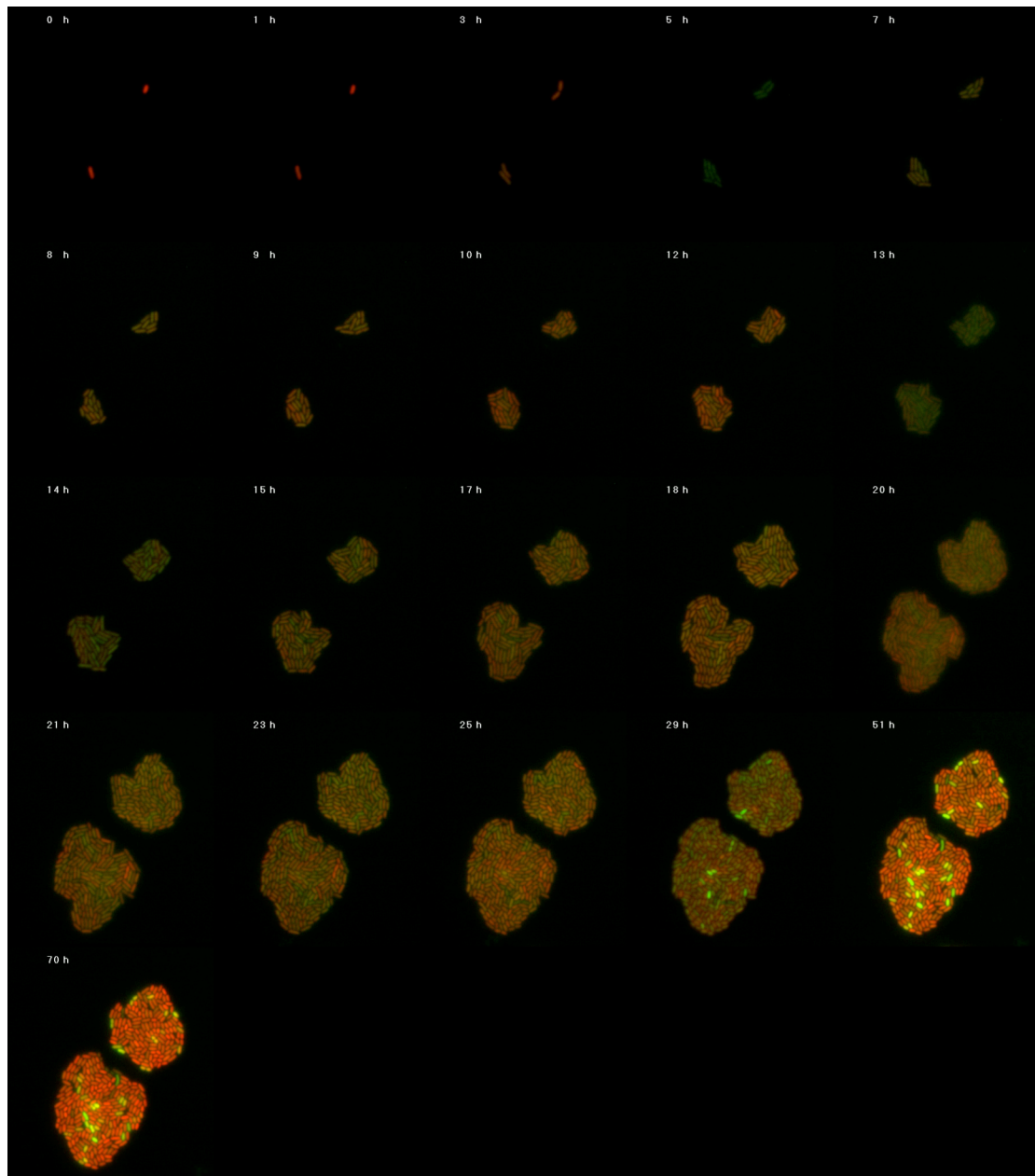
SI Figure 6. Growing *Pseudomonas knackmussii* B13 microcolony 71F (top). The data derived from this microcolony were used for SI Figure 3, and Tables 2-11. Further legend description, see SI Figure 4.
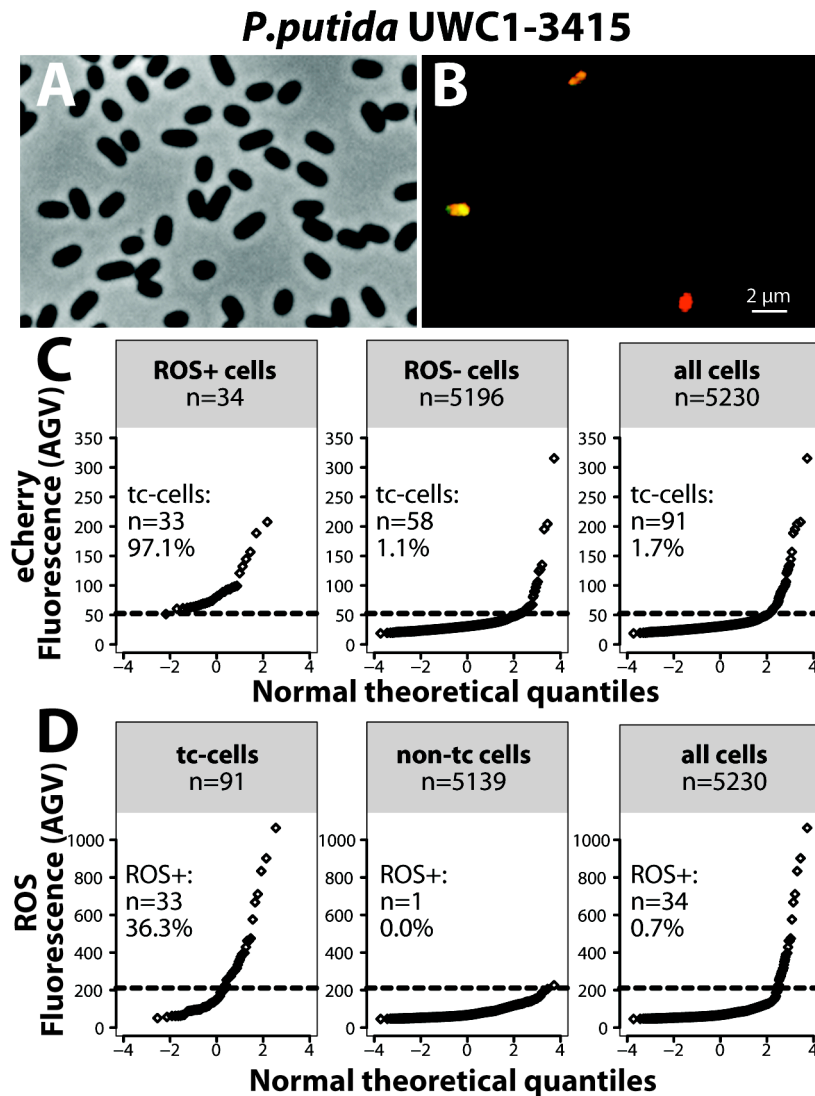
P.putida UWC1-3415

SI Figure 7. Thresholding of ROS- and tc-cells in stationary phase suspended batch culture of *P. putida* UWC1-3415. A) Phase contrast image of individual cells. B) Corresponding overlay image of eCherry (red) and ROS-channels (green), highlighting active or positive cells. C) Theoretical quantile plotting of eCherry fluorescence values across all cells in a population organized in increasing order. The line at ~210 distinguishes cells that are considered having expressed P_{int}-*echerry* above normal and thus are considered tc-cells. Panels distinguish the proportion of tc-cells among those that are ROS-positive. D) Similar as C, but for ROS fluorescence intensity. Note that the ROS-positive threshold is placed on the basis of the lower outliers of H_2O_2-treated cells {Reinhard, 2013}.
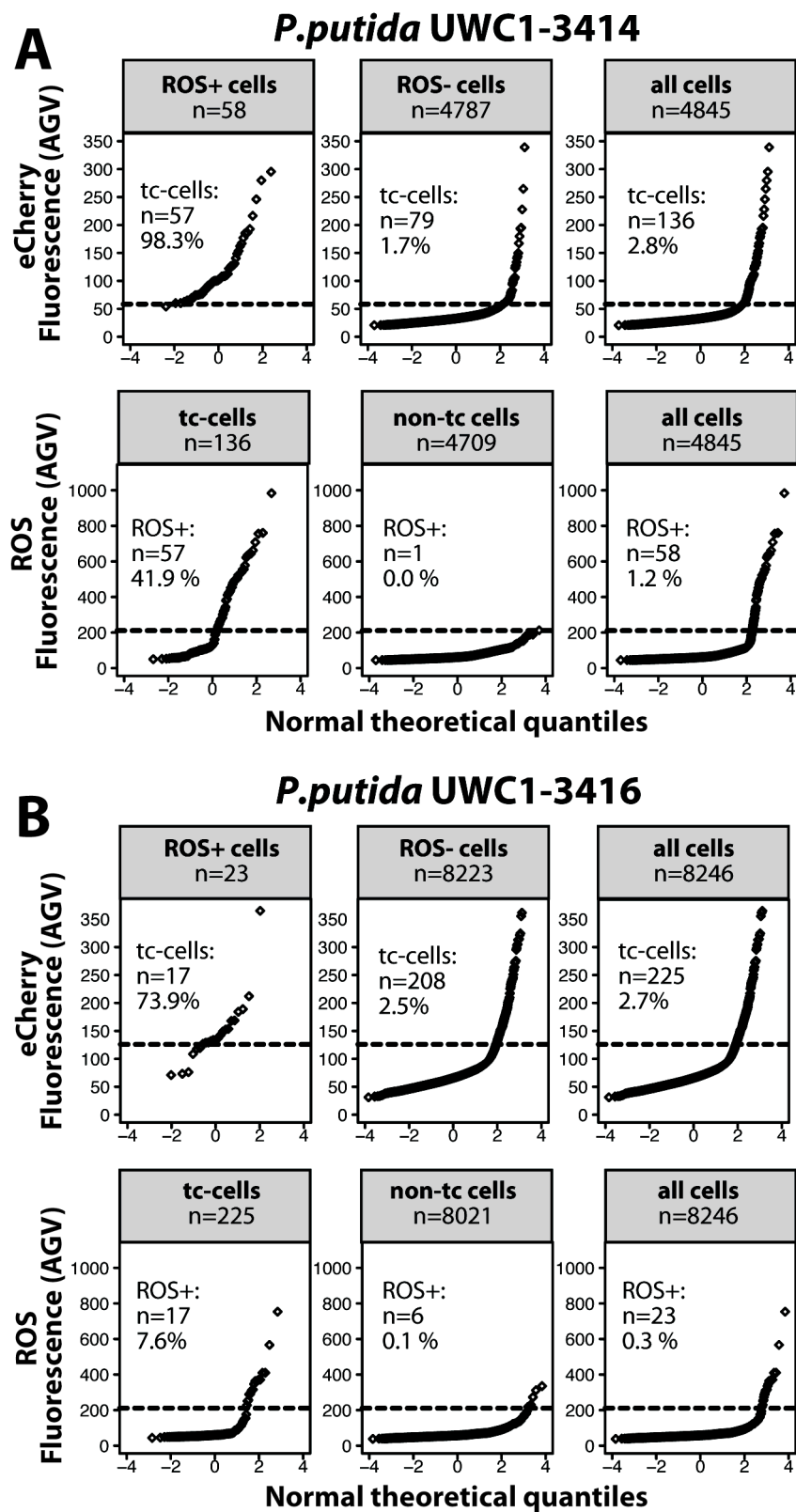
SI Figure 8. Thresholding of ROS- and tc-cells in stationary phase suspended batch culture of *P. putida* UWC1-3414 and *P. putida* UWC1-3416. Same as SI Figure 7 C and D but for strains *P. putida* UWC1-3414 and *P. putida* UWC1-3416.
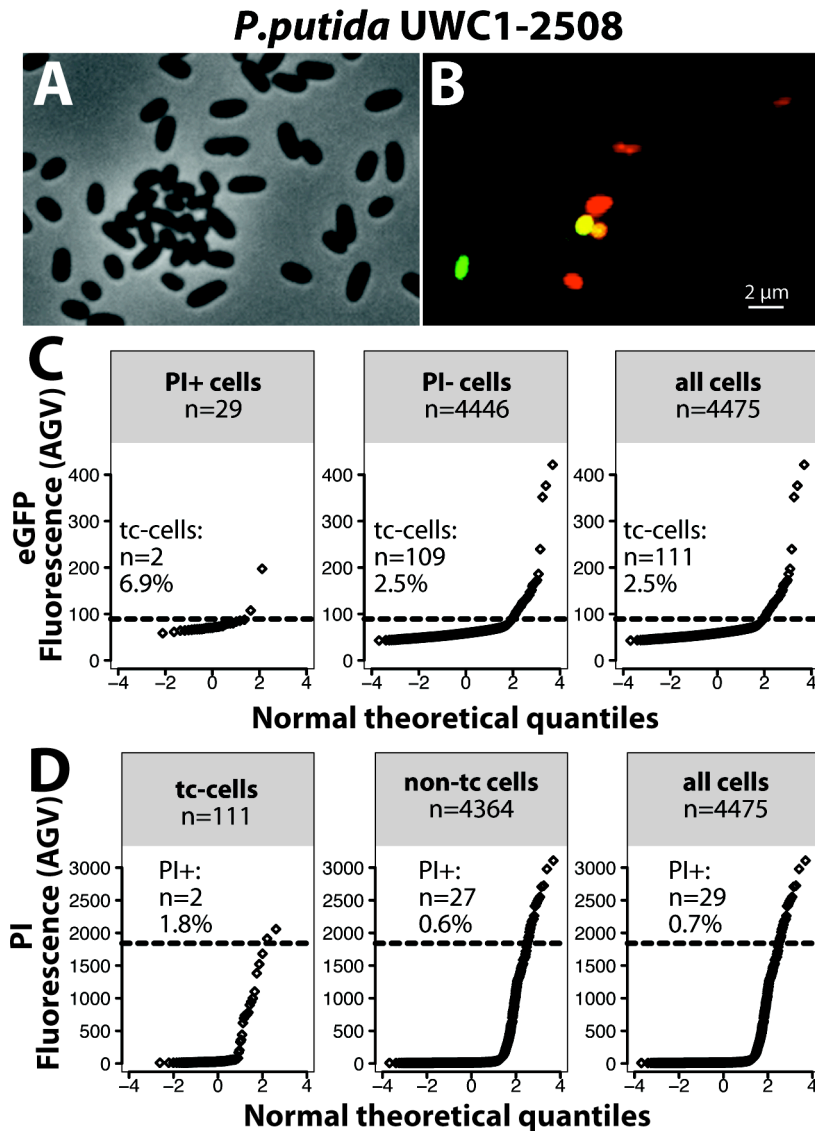
SI Figure 9. Thresholding of PI- and ICE*clc* active cells in stationary phase suspended batch culture of *P. putida* UWC1-2508. A) Phase contrast image of individual cells. B) corresponding overlay image of eGFP (green) and PI-channels (red), highlighting active or positive cells. C) Theoretical quantile plotting of eGFP fluorescence values across all cells in a population organized in increasing order. The line at ~95 distinguishes cells that are considered having expressed P_{int}-*egfp* above normal and thus classify as tc-cells. Panels distinguish the proportion of tc-cells among those that are PI-positive. D) Similar as C, but for PI fluorescence intensity. Note that the PI-positive threshold is placed on the basis of the lower outliers of isopropanol-killed cells {Reinhard, 2013}.
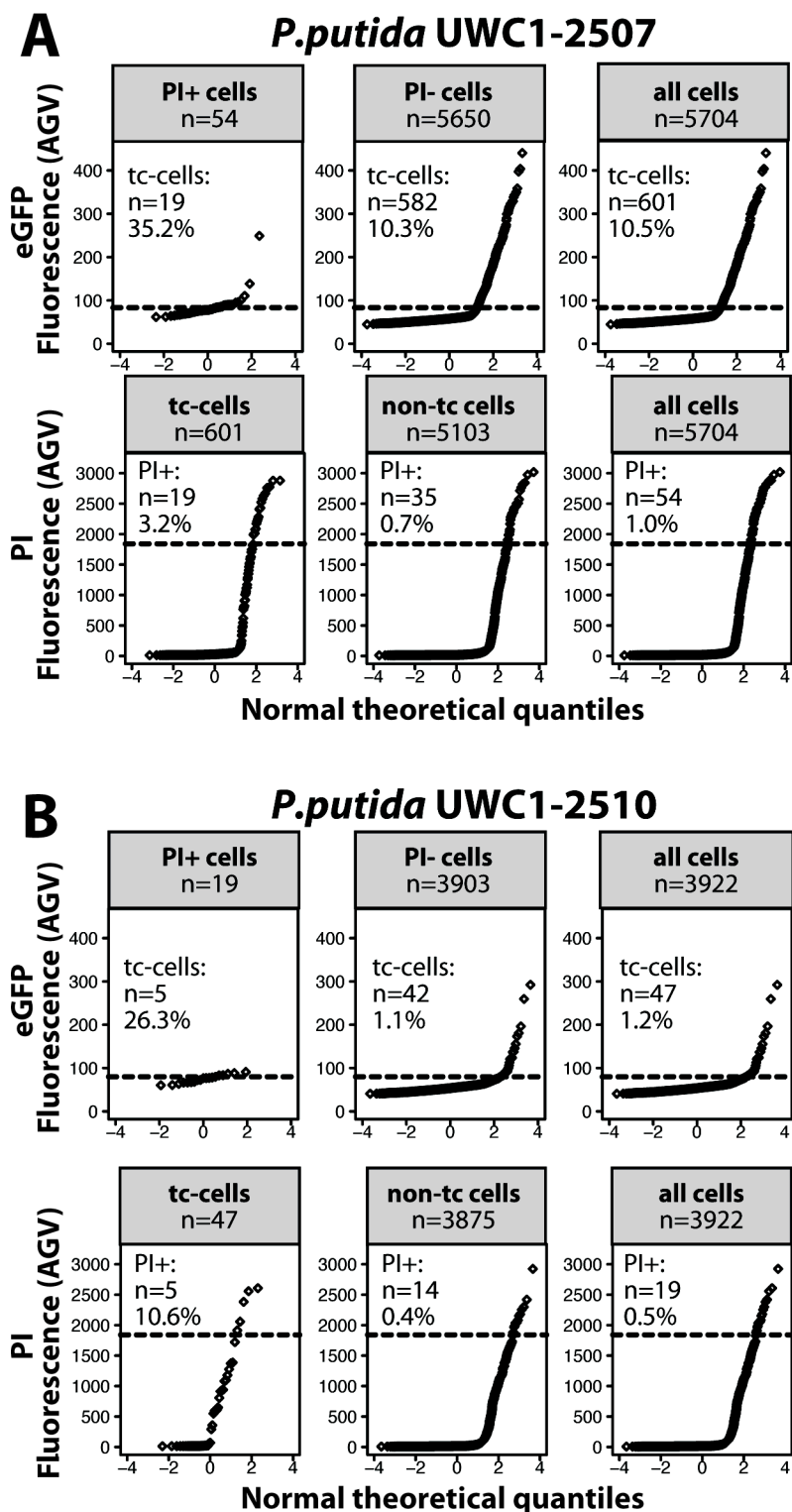
SI Figure 10. Thresholding of PI- and tc-cells in stationary phase suspended batch culture of *P. putida* UWC1-2507 and *P. putida* UWC1-2510. Same as SI Figure 9 C and D but for strains *P. putida* UWC1-3407 and *P. putida* UWC1-2510.

SI Table 1. Table for Chi-square test of independence.

Strain		tc-cells, ROS+	tc-cells, ROS-	non-tc cells, ROS+	non-tc cells, ROS-		p-value	X²
P.putida UWC1-3414	P-*eCherry*	57	79	1	4708	<	2.2×10	1925.851*
P.putida UWC1-3415	P-*eCherry*	33	58	1	5138	<	2.2×10	1762.98*
P.putida UWC1-3416	P-*eCherry*	17	208	6	8015	<	2.2×10	413.8541*
		tc-cells, PI+	tc-cells, PI-	non-tc cells, PI+	non-tc cells, PI-			
P.putida UWC1-2507	P-*egfp*	19	582	35	5068	<	1.2×10	32.5473*
P.putida UWC1-2508	P-*egfp*	2	109	27	4337		0.35	0.8745
P.putida UWC1-2510	P-*egfp*	5	42	14	3861	<	2.2×10	81.5317*
P.knackmussii B13-1343	P-*egfp*	0	122	21	4393		0.93	0.0077

* significant at the 1 % level.

CHAPTER 6

General discussion

At the start of thesis it was suspected that ICE*clc* activation is the consequence of a bistable decision. This was inferred from the finding that the expression of ICE*clc*'s integrase gene, an essential factor facilitating the element's excision from the chromosome, is detectable only in a small fraction (3%) of stationary phase cells (31). Later we discovered that the proportion of integrase-expressing cells is influenced by gene expression noise and can increase to 20% in cells carrying a second gene copy of the chromosomally located stress/starvation sigma factor gene *rpoS* (26). True for all bistable biological phenomena is that these cannot be explained by studying average population behaviour. Instead, the detection requires observation and analysis at the single cell level, which can be achieved through fluorescence microscopy in combination with production of GFP variants expressed from key bistable promoters (18, 31). A major goal of this thesis was to develop fluorescence microscopy-based imaging techniques to describe the fate and life history of single ICE*clc*-active cells (in comparison to non-ICE*clc*-active cells). To realize this goal, two methodologies of single-cell investigation were needed. First, an analytical tool was required that could reliably distinguish ICE*clc*-active from ICE*clc*-inactive cells on the basis of single cell fluorescence values (Chapter 2). Second, an observation set-up was needed that would allow the tracking of individual cellular growth over time (Chapter 3). Equipped with these tools, investigations on fate (Chapter 4) and life history (Chapter 5) of single ICE*clc*-active cells could then be carried out yielding new and unexpected insights into ICE*clc* behaviour. Below, I summarize the main findings and significance of each chapter and put them into broader perspective taking the most recent research in the field into consideration. Further, I provide ideas how the contributions of each chapter could provide a basis for future studies. Where relevant I also provide an evolutionary perspective on our results.

Chapter 2. The study of bistability by genetic manipulation of key regulator genes, or change of environmental conditions, requires first and foremost a reliable method of response quantification. Underlying bistability phenomena are bimodal population structures. These can be described best when the subpopulations involved are disentangled and characterized separately. However, this becomes increasingly challenging the more subpopulations of bimodal populations tend to overlap. To date, only limited, detailed documentation exists specifying such undertaking (4, 5, 23). Our aim in Chapter 2 was therefore, to develop tools for subpopulation detection with special reference to the proportionality effect seen in ICE*clc* activity (18). As an outcome, we presented several statistical methods for use in the free statistical software environment *R*, all of which could reliably approximate population parameters of overlapping subpopulations in bimodal populations. The methods were primarily

developed as a pragmatic, standardized solution to distinguish ICE*clc*-active from ICE*clc* non-active cell fractions on basis of fluorescence data from promoter-reporter gene fusions, but in essence could be applied to any bimodal population data.

Comparison of the different proposed methods, applied on real data sets, showed that there can be several ways to correctly quantify subpopulation sizes and, hence, bistability effects. For example, in a comparison of ICE*clc* activation during growth on 3CBA versus fructose and glucose in *P. knackmussii* B13, all of the proposed methods in this chapter quantified subpopulation sizes more accurately than population independent methods which had been employed previously (24, 31) such as the 95th percentile, or the mean top five percent of a population. Some of the novel proposed methods were simply based on outlier detection as is commonly used in boxplots (35). Boxplot-based outlier detection proved particularly useful when a large number of data sets needed to be analyzed in a short period of time because the subpopulation determination procedure in this case did not require user input as an extra manual step. Further, because of the widespread popularity of boxplots as a general descriptive statistical tool (9), such method is easily comprehensible. Importantly, however, we found that boxplot-based outlier determination was inaccurate under certain subpopulation constellations; typically where boxplot representations failed to indicate bimodality (8). Indeed, simulations revealed that inaccuracies in boxplot estimations generally appeared as soon as subpopulations constitute more than 25% of the total population. In these instances alternative methods based on visual assessment of Q-Q plot visualisations were found to be more reliable, which, however, require manual user input and therefore represent a more time-consuming procedure. Also, we found that manual tools are prone to bias by user input. However, such error could be statistically accounted for by calculation of confidence intervals through repetitive assessments of the same but re-sampled data sets (bootstrapping). Overall, in most cases of small subpopulation detection (<25%) any of the developed analysis methods proved sufficiently accurate as tested on a wide range of simulated bimodal populations. Indeed, we successfully employed our boxplot outlier method earlier to quantify proportions of ICE*clc* activity (3%) using P_{inR}- and P_{int} reporter gene fusions (24). Other examples of experimental phenotype characterisation using this method concern the mutants shown in Supporting Information (SI) Table 1. This table shows the effect of disrupting ICE*clc*'s integrase *intB13* or its left attachment site *attL*, both of which were previously shown to have no effect on ICE*clc* integrase promoter activity (29). The table also shows that a mutation in orf18502, a potential regulator of ICE*clc* transfer genes in a wide variety of ICE (18, 24), abolished ICE*clc* integrase promoter activity (29). Thirdly, SI Table 1 shows an augmentation of the proportion of cells with

ICE*clc* activity as a result of an additional gene for the global transcription factor RpoS added to the host genome. The latter was used to describe how the expression level of the global transcription factor RpoS in individual cells across a population can modulate the frequency of cells activating excision of the ICE*clc* element (26).

When studying ICE*clc* bistability in future research, be it as a function of genetic modification of key regulator genes, host genome environment, or environmental condition, the detection of ICE*clc*-active proportions as a measure of phenotype characterisation will remain invaluable. To achieve this task, a uniform algorithm, rigorously tested, and with sound statistical/mathematical founding, will help to standardize subpopulation estimation and make possible the direct comparison between intra- and inter-research group results. Together the results of Chapter 2 provide new approaches to subpopulation quantification that are well suited as a basis for future bistability phenotype characterisations.

Chapter 3. The innovation of Chapter 3 of this thesis lies in the design of a methodology that allows the tracking of single cells during microcolony growth whilst allowing change in medium conditions. This innovation provided the basis for experiments on cell fate (Chapter 4) and cell age (Chapter 5) in single ICE*clc*-active cells. Although similar systems had already existed at the start of this work, few of these were described in sufficient technical detail to be easily reproduced, or were unsuitable for studying the fate of ICE*clc* activated cells. This was primarily due to a sealed sample environment in these set-ups, which would disallow medium conditions to be manipulated during an experiment (10, 14, 34, 36). Our research, however, required the possibility of adding liquid suspensions such as dissolved carbon source or of diluted dyes (ROS, PI) to the growth medium during the course of an experiment, in order to draw conclusions on the effect of physiological growth difference on ICE*clc* activation (see Chapter 4). Crucially, since another objective was to analyze the life history of each cell (Chapter 5), any online manipulation of culture conditions needed to be designed in such a way that the original structural integrity of the observed microcolony was to be preserved during medium manipulation; an unpredictable positional change of single cells within the microcolony would have rendered cell tracking, and therefore lineage tree calculation very complicated if not impossible. Alternatively, a capillary flow set-up such as used by Balaban and colleagues (3) and Wang and colleagues (37) would have elegantly resolved the problem of maintaining positional information of growing single cells whilst changing growth medium. Still, lineage-genealogy in such capillary flow system is generally limited since relatives are pushed out of the observation channels after a few divisions already. A true alternative, however, would have represented a capillary flow system such as designed in the McKinney lab (11). Here, the

growth of microcolonies can be followed on a glass surface covered with a transparent semi-permeable membrane through which nutrients can diffuse from a liquid medium solution (11). We, however, remained with the simpler and cheaper system of a microcolony growth procedure (10, 14, 34, 36), but which we modified substantially. The pancake-like turning of a gel patch (hence termed "pancake" method), inoculated with a diluted cell suspension on one side, allowed the entrapment of isolated single cells at the interface between observation glass and gel patch. Further, since the "top" side of the gel patch in this case is exposed to a headspace, nutrient or dye suspensions such as ROS- or PI stains can be applied form this side as droplets onto the patch surface, and without interfering with the original positions of the cells on the other side. Also, oxygen can diffuse freely from the headspace to the cell layer that forms in between the gel and the glass. Furthermore, since the whole system is placed within an air-tight observation chamber, nutrient patches are prevented from drying out over long periods of time, lasting up to several weeks depending on how often the system is opened and if the gel patch is re-hydrated. At the same time, volatile nutrients can be released into the headspace within the closed chamber, such as naphthalene or toluene, which are then absorbed by the gel patch and diffuse towards the cells.

In the future, the pan-cake set-up may be employed also to study the role of ICE*clc* activation in more complex, multi-layered biofilms. For this purpose, it is especially useful that our set-up was originally designed for confocal laser scanning microscopy (CLSM), which mostly represent inverted microscopes. For example Seoane and collegues used our set-up in combination with CLSM for studying plasmid invasion (32). A downfall, however, is that waste products are not removed in our system, the accumulation of which might influence gene expression patterns.

Chapter 4. In Chapter 4 we provide evidence for the existence of a transfer-competent (tc) ICE*clc* state, which ultimately leads to reproductive inhibition of tc-state cells. Key to this unexpected finding was the use of single cell observation and analysis tools as described in Chapters 2 and 3 of this thesis, respectively. The result of reproductive inhibition in tc cells was particularly exciting as it represented a new phenomenon in horizontal transfer and raised many questions regarding the possible mechanism of reproductive inhibition in transfer competent cells: could ICE*clc* somehow sense cellular damage and try to escape from the host cell as a last resort strategy? Did reproductive inhibition in transfer competent cells represent a consequence of ICE*clc* activation rather than a starting signal, and was this perhaps caused by costly expression of transfer genes, an elevated integrase activity, ICE*clc* excision, or a mating apparatus that damages the host cell membrane? Was it possible that host cell death resulted

from the killing effect of a toxin-antitoxin (TA) addiction system to ensure ICE stability? Although some answers could be provided to the many questions, a complete elucidation of the mechanism was not possible within time frame of this thesis (but will hopefully provide a few years of exciting single cell research in this field). Nevertheless, we could show for the first time how individual bacterial donor cells can become competent for horizontal transfer of a mobile DNA (ICE*clc*), which, both in the presence and absence of a non-ICE carrier recipient species, leads to strongly impaired reproductive capabilities of the ICE*clc* transfer competent (tc) donor cell and finally cell death. Interestingly, the effect was similar in nature and magnitude in four different *Pseudomonas* species that we used as ICE*clc* donors, but was absent in a host strain without ICE*clc*, which showed that reproductive inhibition is not host dependent and requires the presence of ICE*clc*. Previous results suggested that ICE*clc* excision can take place in stationary phase cultures (24), but from the experiments presented in this chapter we cannot exclude that ICE*clc* transfer only occurs once cells receive sufficient nutrients for cell division to commence. Simultaneously, whereas the reproductive inhibition system initiates in non-dividing cells, it only restricts cell division once substrates are again sufficiently abundant for growth. tc cells accumulated significantly more reactive oxygen species (ROS) even when non-dividing. Once tc cells received enough carbon to resume division, membrane damage and cell death (exemplified by propidium iodide-staining) rapidly occurred often resulting in lysis. Since rehabilitation of tc cells to full reproductive mode was not detected, these traits strongly suggested that development of the tc state is a terminal phenotype. This, however, is in apparent contradiction to the earlier finding that ICE*clc* can excise and re-integrate at different chromosomal locations (30). An explanation to this paradox might be, since ICE*clc* re-integration dynamics was detected by use of PCR on populations, that ICE*clc* activation must be possible without invoking a terminal phenotype, but at frequencies lower than what can be observed microscopically (30). Importantly, we could show that the process of reproductive inhibition is dependent on the presence of ICE*clc* and on expression of the excision and transfer functions (18), but we could rule out that ICE*clc* excision or transfer itself are required, which also rules out the possibility that tc cell inhibition and lysis is due to a temporary overexpression of the ICE*clc*'s integrase as was shown previously in *Escherichia coli* (28). Two possibilities remained. Either ICE*clc* induces two independent 'pathways' in transfer competent cells during stationary phase: (i) the enzymes and factors for ICE*clc* excision and transfer and (ii) a reproductive inhibition system. Or reproductive inhibition is a consequence of horizontal gene transfer itself, that is, of the costly energy requirements associated with expressing a transfer apparatus, and/or of a transfer apparatus that damages the host cell membrane, perhaps in a

manner analogous to the holins that can facilitate phage release (38). Interestingly, two loci, *shi* and *parA^ICEclc^*, were identified within ICE*clc* that appear to slow down growth, prolong shape of tc cells, and ensure an elevated level of horizontal transfer (29). This conclusion was derived from two lines of evidence. Firstly, lac-induced expression of *shi* and *parA^ICEclc^* in ICE*clc*-free host invoked strong growth inhibition during batch growth (29). Microscopic investigation of these cultures indeed revealed slow growing cells with aberrant shapes and cellular damage as identified by PI staining (29). Such cells also formed similar small microcolonies (TCM) as have been observed from tc cells in wild-type (29). Secondly, *parA^ICEclc^* or *shi* mutants showed a significantly lower proportion of TCM compared to wild-type (29). Interestingly, *parA^ICEclc^* and *shi* mutants also showed significantly lower ICE*clc* transfer rates compared to wildtype (5 times lower) (29). This suggested that differentiation of a single tc cell into TCM must be significantly augmenting ICE*clc* transfer rates. However, more extensive phenotypic descriptions of *shi* and *parA^ICEclc^* mutants are needed. This may be achieved via time-lapse microscopy at the single cell level. At least in *shi* mutants such investigation could be focussing on the ICE*clc*-active proportion, since these mutants were shown to exhibit the same proportion of P$_{int}$-*eGFP*-expressing cells as wildtype (29). Also, ROS stains may be performed at the single cell level to see if ROS staining in mutants is reduced with regard to the integrase-active cell proportion, and *vice versa*, if ROS staining is augmented at the population level when *shi* and *parA^ICEclc^* are expressed from plasmid vectors in ICE*clc* free host cells. Further, reporter gene fusions with the promoter of *shi* and *parA^ICEclc^* may show if this promoter is indeed synchronized with P$_{int}$ (and P$_{inR}$), and therefore underlies the same bistable switch, which in part is regulated by RpoS (26). An idea on the potential mode of action of *shi* and *parA^ICEclc^* might be gained from sequence comparisons. Both genes are highly conserved amongst the core regions of proteobacterial ICE*clc* relatives (21, 24, 25), which suggests that generally in ICEs a similar mechanism of horizontal transmission is employed. The *parA^ICEclc^* is encoding a domain similar to the ATPase domain of *ParA*/Soj and might interfere somehow with the cell division proteins, as suggested for ParA/ParB of *Caulobacter*, in which ParB modulates ATPase activity of ParA, and ParA-ADP is supposed to act in cytokinesis (13). The *shi* open reading frame encodes a small hypothetical protein that possesses very low but discernible homology to a eukaryotic voltage-gated calcium channel domain, suggesting that it may influence bacterial membrane potential.

Another important conclusion from this chapter is that ICE*clc* mediated reproductive inhibition must be unlike plasmid addiction systems. Plasmids prevent reproduction of plasmid-free daughter cells that accidentally arise upon improper partitioning of replicated plasmid DNAs during cell division, via TA addiction modules (16, 19, 39). In this case, daughter cells

without plasmid cannot produce the antitoxin while the more stable toxin remains present inhibiting cell division. Recently, also ICE*SXT* was shown to operate a TA system (*mosAT*) to prevent ICE-free cells to proliferate, which can arise when cells with excised ICE would divide (16, 19, 39). Our results with ICE*clc* suggest a different path. By single cell eGFP intensity quantification we demonstrated that the integrase promoter of ICE*clc* remains active in transfer competent starter cells and their siblings. Since P_{int} was not induced in the absence of ICE*clc* this indicates that such daughter cells had not lost ICE*clc*, yet they arrested further cell division and lysed. Moreover, mutations of the ICE*clc* integrase gene *intB13* and left attachment site *attL* abolished ICE*clc* excision but still led to continued P_{int}-activation and cell division arrest in transfer competent cells. We therefore conclude that the process of reproductive inhibition by ICE*clc* must be fundamentally different from TA addiction modules, which inhibit plasmid- (or ICE-) free daughter cells to reproduce (16, 19, 39). Noteworthy, ICE*clc* does not code for any TA system significantly similar to the known seven classes of TA-systems from plasmids, ICE and chromosomes (16). Interestingly, this also tells us that even if a spontaneous mutation would occur that immobilized ICE*clc*, cell death could still occur. In this way programmed cell death systems might have been imported and stabilized in bacterial genomes via mobile DNA (15).

Another key question remains if the observed reproduction inhibition is somehow contributing to, rather than resulting from, ICE*clc* activation. Although the evidence of this chapter supports a hypothesis that ICE*clc* is activated independently of the observed growth inhibition, it cannot be excluded that bacterial damage acts in conjunction with RpoS to generate an ICE-inducing stimulus, which then leads to further damage of the host cell (see Chapter 5). It has been shown previously that expression of ICE*clc*'s integrase is largely dependent upon the stress/starvation sigma factor RpoS, which becomes active at the onset of stationary phase. Other studies, both of ICEs and of phages, have revealed that lateral gene transfer can be induced by agents that generate cellular damage, such as treatments that induce the bacterial SOS response to DNA damage (7). Such "rats leaving the sinking ship" hypothesis is attractive in a sense that ICE*clc* regulates its activity to stage an escape from a cell signifying limited survival potential, however, it is in contrast to our finding that ICE*clc*-free strains do not show TCM formation or increased levels of ROS staining indicative for cellular damage. Furthermore, previous work had found no link between different stresses and stimulation of ICE*clc*'s integrase promoter, including UV radiation, heat shock, or chemical stress, and thus excluded ICE*clc* activation as part of the classical SOS response (31).

Finally, the findings of this chapter raise interesting evolutionary questions regarding the ICEclc host-symbiont relationship. Clearly, ICEs rely on the host for DNA replication and maintenance, and therefore their fitness is tightly coupled to that of the host cell (20). Consequently, it can be safely assumed that increasing host cell fitness would also increase ICE fitness and *vice versa*, and it is therefore no surprise to see that many ICEs follow a strategy of contributing beneficial genes to the host cell genome (12, 22). For example ICE*SXT* of *Vibrio cholera* and ICE*Hin*1056 of *Haemophilus influenza* confer antibiotic resistance to their hosts. ICEclc confers the ability to metabolize chloro-aromatic compounds such as 3-CBA, and 2-aminophenol. It was further shown that under non-selective growth conditions (e.g. growth with succinate or acetoin), where there is no implication of ICEclc encoded metabolic pathways, there seems no strong selective pressure against maintaining ICEclc (18). The low impact of ICEclc on host fitness under non-selective conditions was assumed to be due to the fact that ICEclc is integrated in the host chromosome with regeneration of the integration site (17), and is silent during exponential phase (18).

Another strategy for an ICE to increase its fitness is to swap hosts by transferring to another host that potentially benefits more from ICE genes than the current host does. For example, Springael et al. (33) showed that *in situ* horizontal transfer of ICEclc from the inoculum to contaminant bacteria in the reactors was involved in the establishment of novel 3-chlorobenzoate degrading populations that were more competitive under the defined reactor conditions than the inoculum strain. Thus, while ICE-encoded beneficial genes and a low energetic impact on host metabolism under non-selective growth conditions ensure an optimal vertical transmission of the ICE, the ability to horizontally transfer to another host allows to constantly "probe" for a potentially more reproductive partnership. This combination of "serving" in the current host but "probing" for a potentially better host, ensures maximal ICE fitness for any given environment and provides an explanation for their evolutionary success.

With the above understood, the main question that remained after the surprising results of Chapter 4 was: how does a general strategy of an ICE to not compromise host cell fitness comply with the observed ICE-caused reproductive inhibition and cell lysis? At least theoretically, two mutually non-exclusive explanations can be imagined. Firstly, the discovered reproductive inhibition/cell death system may, in fact, represent a beneficial rather than a negative trait to the host. For example might fractional cell death be part of an extreme form of division of labour such as self-destructive cooperation (2). In this case cell lysis would somehow increase the average fitness of individual cells. For example, it has been shown in pathogens previously, that subpopulations are sacrificed thereby facilitating successful infection by the

remaining subpopulation (2). A self-destructive cooperation strategy regarding ICE*clc*-mediated cell death may be envisioned in context of biofilm development, where cell lysis may contribute favourably towards the structural integrity and heterogeneity of biofilms. For example, one could imagine that randomly dispersed lysis events throughout a biofilm, either of single tc cells or aggregates of tc cells (as a result of TCM formation), may influence biofilm architecture through release of signalling molecules and structural components (DNA), as well as the creation of spaces and micro-channels influencing nutrient and signal circulation (6).

The second possibility why we see cell death may be because it represents a prerequisite of achieving efficient horizontal gene transfer, but whose negative impact upon host fitness is kept low enough so as to not be selected against. The important role of cell death in this case might be due to a variety of reasons. For example could the transfer mechanism itself be causing damage; for example through membrane disruption of the host cell, or indirectly, through the high energetic costs required for the expression of the conjugation machinery. Alternatively, growth inhibition and death may represent adaptations allowing for transfer rates at the required level. For example could nutrients released from donor cell lysis be available for recipients cells and thereby contribute to a successful integration of ICE*clc* into the new host genome. In addition, the few divisions that tc cells still undergo before lysis to form TCM, which we termed "mating bodies", may serve to increase the chance of contact of a suitable recipient on a surface. Indeed, we have shown that in mutants that had lost the ability to form mating bodies (*shi*, *parA*^{ICEclc}) transfer efficiency was reduced up to five-fold. However, independent of the possible role of cell death in horizontal gene transfer, and given the complete dependence of ICE*clc* maintenance on the host, the question still remains, how at all can a disadvantage conferred by the ICE upon the host represent an evolutionary stable strategy? For instance, one could imagine that spontaneous mutants lacking the reproductive inhibition process could arise. Mathematically, such mutants would outgrow the wildtype after a few hundred cycles already, despite a rare occurrence of reproductively inhibited cells at non-growing phase (SI Explanation). An explanation why no such mutants have been seen yet must be due to the important role of genetic drift as an evolutionary mechanism (SI Explanation, Chapter 4). Genetic drift can be simulated by modelling simple serial batch growth, where after each cycle (reaching stationary phase), a random sub-sample is taken and diluted with fresh medium. As a consequence a bottle-neck in genetic diversity occurs after each growth cycle (the genetic drift) due to the filtering out of mutants that have failed to become abundant enough to be represented in the sub-sample. Indeed, modelling suggests that at a sub-sampling rate of 0.001, proportions of below 20% of transfer competent

cells, and at appearance of mutants at rates of 10^{-8} per generation, mutants are unlikely to establish (Chapter 4, SI Explanation). The reason for this may be twofold: firstly and more importantly, since ICE*clc*-related growth inhibition is only expressed in stationary phase (i.e., non-growing cells) there is no fitness loss for the host in terms of exponential growth rate of the population. Secondly and less importantly, reproductive inhibition in tc cells is confined to a small proportion (3%) of all cells in the population. In this respect, it is interesting to note that in the model (Chapter 4, SI Explanation), as long as cell damage is confined to the non-growing phase and genetic drift is kept at a constant level, a much higher fraction of reproductively inhibited cells would theoretically be allowed (up to 20%) than currently observed in the wildtype (3-5%) before mutants would start to replace wiltype. This leads to the rather difficult-to-answer questions how determinants lead to the tc cell/non-tc cell ratios currently observed, and why different proportions of tc cells are seen under different growth conditions (Chapter 2). Some answers may lie hidden in the evolutionary dynamics of host-parasite interactions (22). For example Lipsitch predicts that increasing efficiency in horizontal gene transfer in parasites that can also vertically transfer, will result in reduced virulence of the parasite. The surprising finding of the presence of two functional origins of transfer in ICE*clc* augmenting horizontal transfer efficiency (27) might, according to this theory, have been selected for because such increase in transfer efficiency would allow for a low proportion of cells to engage in horizontal gene transfer, thus reducing virulence on the host as defined by reproductive inhibition in tc cells.

However, a clearer picture of the pathways of growth inhibition and transfer competence in tc cells will certainly help to define also clearer evolutionary questions related to the tc cell growth inhibition phenomenon.

Further, it should also be interesting to see if the ICE*clc*-inherent cell death system might be exploited for the design of novel antibiotic drugs. For example one could imagine that factors that prevent the activation of the inherent cell death mechanism in a higher proportion of cells than naturally seen might represent interesting novel targets for bacteriostatic or bacteriocidal drugs.

Chapter 5. Previously, we showed that the proportion of cells expressing P_{inR} and P_{int} depends on RpoS levels, with cells having higher than population-average RpoS levels showing a higher likelihood of activating the tc-cascade (26). These results thus suggested that the decision to activate or not P_{inR} and P_{int} in an individual cell is a stochastic process. However, apparent stochasticity (e.g., the distribution of RpoS among cells) may have underlying causes, such as, some cells being in a slightly different growth phase or having experienced more

previous biochemical damage, and thus producing more RpoS. The specific hypotheses we thus wanted to test in the work of Chapter 5 were whether ICE*clc* activation in individual cells was dependent on cell pole age, whether it was confined to specific cellular lineages, or occurred as a consequence of pre-existing detectable biochemical damage. We also tested whether tc cell formation occurs spatially at random within a microcolony. For this purpose, most of the experiments in this chapter were based on time-lapse imaging of individual cells in growing microcolonies (see Chapter 2). This had the advantage of tracing the behaviour and life history of all cells throughout exponential and stationary phases, when the colony stops growing and ICE*clc* tc-formation was initiated.

Although it has been shown previously in *Caulobacter crescentus* (1) that cell age is correlated to slow growth and an increased probability of lysis, in our case we could exclude cell pole age to be the reason for growth inhibition and lysis phenomena in tc cells. We further showed that initiation of transfer competence is not the result of the physiological state of ancestor cells, or of a close relative (sister cell). Further, although tc-cells showed higher levels of reactive oxygen species and membrane damage than non-tc cells (as indicated by higher ROS and PI staining, respectively), we could not discern whether tc-initiation occurs because of damage, or whether damage occurs as a consequence of tc-initiation. Thus, in essence, none of our hypotheses was supported by the experimental observations, from which we concluded that ICE*clc* activation is a random process depending on stochastic variation among individual cells. Indeed, random ICE*clc* activation would be in accordance with our observation that ICE*clc* activation occurs spatially randomly in a microcolony. One could imagine that a spacially random arrangement of tc cells in a mixed culture biofilm might increase efficiency of horizontal gene transfer by maximising the chances to contact potential recipients. Similarly, one could imagine that a spatially random occurrence of lysis might represent an important feature in shaping biofilm structure and landscape, perhaps through the release of signalling and structural molecules, or the increase of biofilm porosity influencing medium flow and nutrient circulation (6).

For future studies, being able to differentiate between pre-existent or induced damage might be an interesting avenue to pursue. Specifically interesting tools in this regard would allow identification over time of the tc-state of cells as well as of reactive oxygen species presence. One possibility here would be reporter gene fusions to promoters in *P. knackmussii* or *P. putida* reactive to oxygen radicals, which, however, still need to be identified in these organisms.

Summary. In summary, the research of this thesis has provided novel analytical and observational solutions to single cell research. It has also led to proof for existing hypotheses of ICE*clc* activation and transfer competence. But perhaps most interestingly, this thesis has led to the unexpected discovery and characterisation of a novel phenotype corresponding to a severe fitness cost in transfer competent cells. This discovery undoubtedly poses many new exciting questions concerning the role of mobile DNA in manipulating host cell behaviour and differentiation in order to optimize horizontal transmission.

Finally, I am grateful and hopeful to have been witnessing the unfolding of an exciting story in microbial research.

REFERENCES

1. **Ackermann, M., A. Schauerte, S. C. Stearns, and U. Jenal.** 2007. Experimental evolution of aging in a bacterium. BMC Evol Biol **7:**126.

2. **Ackermann, M., B. Stecher, N. E. Freed, P. Songhet, W. D. Hardt, and M. Doebeli.** 2008. Self-destructive cooperation mediated by phenotypic noise. Nature **454:**987-990.

3. **Balaban, N. Q., J. Merrin, R. Chait, L. Kowalik, and S. Leibler.** 2004. Bacterial persistence as a phenotypic switch. Science **305:**1622-1625.

4. **Bates, D., J. Epstein, E. Boye, K. Fahrner, H. Berg, and N. Kleckner.** 2005. The *Escherichia coli* baby cell column: a novel cell synchronization method provides new insight into the bacterial cell cycle. Mol Microbiol **57:**380-391.

5. **Bates, D., and N. Kleckner.** 2005. Chromosome and Replisome Dynamics in *E. coli*: Loss of Sister Cohesion Triggers Global Chromosome Movement and Mediates Chromosome Segregation. Cell **121:**899-911.

6. **Battin, T. J., W. T. Sloan, S. Kjelleberg, H. Daims, I. M. Head, T. P. Curtis, and L. Eberl.** 2007. Microbial landscapes: new paths to biofilm research. Nat Rev Microbiol **5:**76-81.

7. **Beaber, J. W., B. Hochhut, and M. K. Waldor.** 2004. SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature **427:**72-74.

8. **Choonpradub, C., and D. McNeil.** 2005. Can the Box Plot be Improved? Songklanakarin Journal of Science and Technology **27:**649-657.

9. **Cleveland, W. S.** 1994. The Elements of Graphing Data, revised ed. Murray Hill, New Jersey.

10. **de Jong, I. G., J. W. Veening, and O. P. Kuipers.** 2010. Heterochronic phosphorelay gene expression as a source of heterogeneity in *Bacillus subtilis* spore formation. J Bacteriol **192:**2053-2067.

11. **Dhar, N., and J. D. McKinney.** 2010. *Mycobacterium tuberculosis* persistence mutants identified by screening in isoniazid-treated mice. Proc Natl Acad Sci U S A **107:**12275-12280.

12. **Dobrindt, U., B. Hochhut, U. Hentschel, and J. Hacker.** 2004. Genomic islands in pathogenic and environmental microorganisms. Nature Rev Microbiol **2:**414-424.

13. **Easter, J. J., and J. W. Gober.** 2002. ParB-stimulated nucleotide exchange regulates a switch in functionally distinct ParA activities. Mol Cell **10:**427-434.

14. **Eberhardt, A., L. J. Wu, J. Errington, W. Vollmer, and J. W. Veening.** 2009. Cellular localization of choline-utilization proteins in Streptococcus pneumoniae using novel fluorescent reporter systems. Mol Microbiol **74:**395-408.

15. **Engelberg-Kulka, H., S. Amitai, I. Kolodkin-Gal, and R. Hazan.** 2006. Bacterial programmed cell death and multicellular behavior in bacteria. PLoS Genet **2:**e135.

16. **Fozo, E. M., K. S. Makarova, S. A. Shabalina, N. Yutin, E. V. Koonin, and G. Storz.** 2010. Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. Nucleic Acids Res **38:**3743-3759.

17. **Gaillard, M., N. Pernet, C. Vogne, O. Hagenbüchle, and J. R. van der Meer.** 2008. Host and invader impact of transfer of the *clc* genomic island into *Pseudomonas aeruginosa* PAO1. Proc Natl Acad Sci U S A **105:**7058-7063.

18. **Gaillard, M., N. Pradervand, M. Minoia, V. Sentchilo, D. R. Johnson, and J. R. van der Meer.** 2010. Transcriptome analysis of the mobile genome ICE*clc* in *Pseudomonas knackmussii* B13. BMC Microbiol **10:**153.

19. **Gerdes, K., S. K. Christensen, and A. Lobner-Olesen.** 2005. Prokaryotic toxin-antitoxin stress response loci. Nat Rev Microbiol **3:**371-382.

20. **Gogarten, J. P., and J. P. Townsend.** 2005. Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol **3:**679-682.

21. **Juhas, M., J. R. van der Meer, M. Gaillard, R. M. Harding, D. W. Hood, and D. W. Crook.** 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev **33:**376-393.

22. **Lipsitch, M.** 1996. The evolution of virulence in pathogens with vertical and horizontal transmission. Evolution **50:**1729-1741.

23. **MacArthur, B., D., R. Tare, S., C. P. Please, P. Prescott, and R. Oreffo, O., C.** 2006. A non-invasive method for *in situ* quantification of subpopulation behaviour in mixed cell culture. J. R. Soc. Interface **3:**63-69.

24. **Minoia, M., M. Gaillard, F. Reinhard, M. Stojanov, V. Sentchilo, and J. R. van der Meer.** 2008. Stochasticity and bistability in horizontal transfer control of a genomic island in *Pseudomonas*. Proc. Natl. Acad. Sci. U S A **105:**20792-20797.

25. **Miyazaki, R., M. Minoia, N. Pradervand, V. Sentchilo, S. Sulser, F. Reinhard, and J. R. van der Meer.** 2011. The *clc* element and related genomic

islands in *Proteobacteria. In* A. P. Roberts and P. Mullany (ed.), Bacterial integrative mobile genetic elements. Landes Bioscience.

26. **Miyazaki, R., M. Minoia, N. Pradervand, S. Sulser, F. Reinhard, and J. R. van der Meer.** 2012. Cellular variability of RpoS expression underlies subpopulation activation of an integrative and conjugative element. PLoS Genet **8:**e1002818.

27. **Miyazaki, R., and J. R. van der Meer.** 2011. A dual functional origin of transfer in the ICE*clc* genomic island of *Pseudomonas knackmussii* B13. Mol Microbiol **79:**743-758.

28. **Ravatn, R., S. Studer, A. J. B. Zehnder, and J. R. van der Meer.** 1998. Int-B13, an unusual site-specific recombinase of the bacteriophage P4 integrase family, is responsible for chromosomal insertion of the 105-kilobase *clc* element of *Pseudomonas* sp. strain B13. J. Bacteriol. **180:**5505-5514.

29. **Reinhard, F., R. Miyazaki, N. Pradervand, and J. R. van der Meer.** 2013. Cell differentiation to "mating bodies" induced by an integrating and conjugative element in free-living bacteria. Curr Biol **23:**255-259.

30. **Sentchilo, V., K. Czechowska, N. Pradervand, M. Minoia, R. Miyazaki, and J. R. van der Meer.** 2009. Intracellular excision and reintegration dynamics of the ICE*clc* genomic island of *Pseudomonas knackmussii* sp. strain B13. Mol Microbiol **72:**1293-306.

31. **Sentchilo, V., R. Ravatn, C. Werlen, A. J. Zehnder, and J. R. van der Meer.** 2003. Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. Strain B13. Journal of Bacteriology **185:**4530-4538.

32. **Seoane, J., T. Yankelevich, A. Dechesne, B. Merkey, C. Sternberg, and B. F. Smets.** 2010. An individual-based approach to explain plasmid invasion in bacterial populations. FEMS Microbiol Ecol **75:**17-27.

33. **Springael, D., K. Peys, A. Ryngaert, S. Van Roy, L. Hooyberghs, R. Ravatn, M. Heyndrickx, J. R. van der Meer, C. Vandecasteele, M. Mergeay, and L. Diels.** 2002. Community shifts in a seeded 3-chlorobenzoate degrading membrane biofilm reactor: indications for involvement of in situ horizontal transfer of the clc-element from inoculum to contaminant bacteria. Environ Microbiol **4:**70-80.

34. **Stewart, E. J., R. Madden, G. Paul, and F. Taddei.** 2005. Aging and death in an organism that reproduces by morphologically symmetric division. PLoS Biol **3:**e45.

35. **Tukey, J. W.** 1977. Exploratory data analysis. McGraw-Hill, Reading, MA.

36.   **Veening, J. W., E. J. Stewart, T. W. Berngruber, F. Taddei, O. P. Kuipers, and L. W. Hamoen.** 2008. Bet-hedging and epigenetic inheritance in bacterial cell development. Proc Natl Acad Sci U S A **105:**4393-4398.

37.   **Wang, P., L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun.** 2010. Robust growth of *Escherichia coli*. Curr Biol **20:**1099-1103.

38.   **White, R., S. Chiba, T. Pang, J. S. Dewey, C. G. Savva, A. Holzenburg, K. Pogliano, and R. Young.** 2011. Holin triggering in real time. Proc Natl Acad Sci U S A **108:**798-803.

39.   **Wozniak, R. A., and M. K. Waldor.** 2009. A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. PLoS Genet **5:**e1000439.

ANNEX: SUPPORTING INFORMATION

SI Table 1. Varying sub-population sizes in different strains of *Pseudomonas putida* and *Pseudomonas knackmussii* B13.

Category	Strain	% Sub-population[1]	Mean[2]	Strain description	Significantly different category[3]
A	*P.knackmussii* B13- 3201[4]	16.2	18.3 ± 4.0	P_{int}-*egfp*, P_{int}-*eCherry*, P_{rpoS}-*rpoS*	B*, C*, D*, E*, F*
	P.knackmussii B13- 3202[4]	15.9			
	P.knackmussii B13- 3206[4]	22.9			
B	*P.putida* UWC-2507[5]	8.2	4.6 ± 3.2	ICE*clc*, P_{int}-*egfp*	A*
	P.putida UWC-2508[5]	2.3			
	P.putida UWC-2510[5]	3.2			
C	*P.putida* UWC-3414[5]	2.1	5.4 ± 5.4	ICE*clc*, P_{int}-*eCherry*	A*
	P.putida UWC-3415[5]	2.5			
	P.putida UWC-3416[5]	11.7			
D	*P.putida* UWC-3403[5]	2.1	2.6 ± 0.6	ICE*clc*, P_{int}-*eCherry*, Δ*attL*	A*
	P.putida UWC-3404[5]	2.5			
	P.putida UWC-3405[5]	3.2			
E	*P.putida* UWC-3406[5]	7.3	6.0 ± 1.8	ICE*clc*, P_{int}-*eCherry*, Δ*intB13*	A*
	P.putida UWC-3407[5]	3.9			
	P.putida UWC-3408[5]	6.7			

SI Table 1 *Continued*

Category	Strain	% Sub-population[1]	Mean[2]	Strain description	Significantly different category[3]
F	*P.putida* UWC-3409[5]	0.1	0.1 ± 0.1	ICE*clc*, P$_{int}$-*eCherry*, Δ*orf18502*	A*, D*, E*
	P.putida UWC-3410[5]	0			
	P.putida UWC-3411[5]	0.1			

1) ICE*clc* transfer competent sub-population of cells (percent of total) expressing *egfp* or *echerry* from P$_{int}$. Sampled 15 - 20 h after onset of stationary phase. Determined via R command *find.sub.pop()* in *Default* mode.

2) Mean ICE*clc* transfer competent sub-population of cells (percent of total) ± standard deviation calculated from strains within the same category.

3) * indicates a significant difference between the means of different categories at P<0.05 as determined by the Welch Two Sample t-test.

4) these strains have been constructed and analyzed previously in a paper by Miyazaki and collegues (3)

5) these strains have been published by Reinhard and collegues (4)

SI Explanation: Modelling of Development of Reproductive Inhibition-Independent Mutants in Populations of ICE*clc*-Containing Hosts

SUMMARY. A recent study by Reinhard et al. (4) described a phenomenon by which 3%-5% of cells in a population of the genus *Pseudomonas* lyse as a result of the induction of ICE*clc*, an integrating and conjugative element. Here we investigate how reproductive inhibition in a subpopulation may represent an evolutionary favourable strategy. By use of algebraic theoretical theory and computer simulations, we show that under conditions where genetic drift is absent a mutant showing no fractional reproductive inhibition of transfer-competent cells in populations of ICE*clc*-containing hosts would out-compete the wildtype. However, under conditions where genetic drift is prevailing, as shown in simulations of serial batch transfer growth experiments, out-competition does not occur.

AIM. The goal of this model was to assess if the selective advantage of a mutant lacking fractional reproductive inhibition as described by Reinhard et al. (4) would have been

sufficiently high to allow competitive replacement of the wildtype. If no, this would be one explanation of why we have not seen any such mutants yet. If yes, this indicates that fractional cell death might in fact provide a fitness advantage to the host population rather than a disadvantage and would therefore represent a selectable trait in analogy to destructive cooperation (1). To achieve this goal, we used two approaches, an algebraic theoretical analysis predicting the change of mutant-to-wildtype ratio as a function of growth rate, and a computer-based simulation, mimicking batch transfer events.

SI METHODS

Computer Simulations in R. For computer simulations the function *death.sim()* was created in *R* (R Development Core Team (2009). Version R 2.9.2. http://www.R-project.org). The function is based on a similar model as described by Reinhard et al. (2013) (4). Output of the previous model (limited to 21 growth cycles) and *death.sim()* was compared across different parameters and was found identical indicating the correct translation of the previous model into R language. *death.sim()* provides graphical as well as numerical output. Graphical output shows mutant-to-wildtype ratio as a function of growth cycles whilst including a legend of the simulation parameter settings used. Numerical output includes raw data used to construct the graph as well as the simulation parameter settings. After each simulation output is stored as R objects named *my.data* , *my.ratio* and *my.parameters*. *death.sim()* parameters are shown in Table SI 2. If no parameter is set in the function, the function calls default parameters as shown in Table SI 2.

R code

```
rm(list=ls(all=TRUE))

death.sim <- function (cycles = 300, probability = 3.00E-2, mutation_rate = 1.00E-5, Nm_start = 1.00E1, N_start = 5.00E4,
GenT_g = 2.00E0, GenT_ng = 1.00E0, death_rate=4.00E-1, Total_C = 1.00E-4, Per_cell = 2.00E-12, yield = 3.00E-1, Ks = 2.00E-
5, dilution = 1.00E-3, Ks.m = 2.00E-5, deltaT = 3.00E-1) {

my.list<-list()


N_start         <- N_start # N cells at start
probability     <- probability # Probability of a cell becoming transfer-competent
deltaT          <- deltaT # as described in Jan's simulation in Excel: ExcelSim
Nm_start            <- Nm_start # Number of mutants at start
GenT_ng             <- GenT_ng # Generation time non-active cells (h)
GenT_g          <- GenT_g # Generation time ICEclc transfer-competent cells (h)
death_rate      <- death_rate # Death rate of ICEclc transfer-competent cells (1/h)
Total_C             <- Total_C # Total amount of substrate per cycle (g /mL)
Per_cell        <- Per_cell # Amount of carbon per cell (g)
yield           <- yield # Biomass yield (g/g)
N3                  <- Per_cell/yield # as described in Jan's simulation in Excel: ExcelSim
N5                  <- Total_C/N3 # as described in Jan's simulation in Excel: ExcelSim
Total_cells     <- Total_C/Per_cell*yield # (g/ml)
Ks                  <- Ks # Ks (g/ml)
MuMax           <- log(2)/GenT_ng # µ(max) (1/h)
dilution        <- dilution # Population dilution factor per cycle
mutation_rate   <- mutation_rate # mutation rate (per generation)
Ks.m                <- Ks.m# g/ml, as described in Jan's simulation in Excel: ExcelSim
MuMax.m             <- MuMax # µ(max, mutant) (1/h)
cycle           <- numeric() # empty vector for later use
gen.cyc             <- numeric() # empty vector for later use
gen.tot         <- numeric() # empty vector for later use


# define ExcelSim's first cycle's first row
NN1                     <- numeric()
NN2                     <- numeric()
NN3                     <- numeric()
Carbon              <- Total_C
Mu                      <- MuMax*(Carbon/(Carbon+Ks))
Time                <- seq(0,14,0.5)
N1                      <- N_start
Netto_new           <- N1
N2                      <- if (Mu<0.1) {probability*N1} else {0}

Mu.m                <- MuMax.m*(Carbon/(Carbon+Ks))
N1.m                <- N1*mutation_rate
N2.m                <- if (N1.m>1) {floor(N1.m)} else {0}
Ntot2.m             <- Nm_start
```

```r
Ntot1.m                        <- N2.m + Ntot2.m
Netto_new.m            <- Ntot2.m



# define ExcelSim's first cycle's successive rows
for (i in 2:29) {
# i<-2
NN1[i]                 <- NA
NN2[i]                 <- NA
NN3[i]                 <- NA
N1[i]                  <- if(Mu[i-1]>0) {N1[i-1]*2^((Time[i]-Time[i-1])/(log(2)/Mu[i-1]))} else {N1[i-1]} # 4
Netto_new[i]           <- N1[i]-N1[i-1] # 5
N1.m[i]                <- N1[i]*mutation_rate   #8
N2.m[i]                <- if (N1.m[i]>1) {floor(N1.m[i])} else {0}#9
Ntot2.m[i]             <- if (Mu.m[i-1]>0) {Ntot1.m[i-1]*2^((Time[i]-Time[i-1])/(log(2)/Mu.m[i-1]))} else {Ntot2.m[i-1]}#10
Ntot1.m[i]             <- N2.m[i] + Ntot2.m[i]#11
Netto_new.m[i]  <- Ntot2.m[i] - Ntot2.m[i-1]#12
Carbon[i]                      <- if(Carbon[i-1]-((Netto_new[i]+Netto_new.m[i])*N3)<0) {0} else {Carbon[i-1]-
(Netto_new[i]+Netto_new.m[i])*N3}
Mu[i]                  <- if(Carbon[i]>0) {MuMax*(Carbon[i]/(Carbon[i]+Ks))} else {0}
N2[i]                  <- if (Mu[i]<0.1) {probability*N1[i]} else {0} #6
Mu.m[i]                <- MuMax.m*(Carbon[i]/(Carbon[i]+Ks))   #7
cycle[1]               <- 1#j
cycle[i]               <- 1#j
gen.cyc[1]             <- 1
gen.cyc[i]             <- i


}



my.list[[1]] <-
                data.frame(
                        Carbon,
                        Mu,
                        Time,
                        N1,
                        Netto_new,
                        N2,
                        Mu.m,
                        N1.m,
                        N2.m,
                        Ntot1.m,
                        Ntot2.m,
                        Netto_new.m,
                        NN1,
                        NN2,
                        NN3,
```

293

```
                cycle,

                gen.cyc

                )
```

define ExcelSim's successive cycle's first row

```
m<-cycles
for (j in 2:m) {
```

```
NN1                        <- floor(dilution*N2[i])
NN2                        <- NN1
NN3                        <- if (NN2>0) {NN2} else {0}
Carbon            <- Total_C
Mu                         <- MuMax*(Carbon[1]/(Carbon[1]+Ks))
Time              <- seq(0,14,0.5)
N1                         <- floor(dilution*(N1[i-1]-N2[i-1]))
Netto_new                  <- N1
N2                         <- if (Mu[1]<0.1) {probability*N1} else {0}


Mu.m              <- MuMax.m*(Carbon/(Carbon+Ks))
N1.m              <- N1*mutation_rate
N2.m              <- if (N1.m>1) {floor(N1.m)} else {0}
Ntot2.m                    <- floor(dilution*Ntot2.m[i])
Ntot1.m                  <- N2.m + Ntot2.m
Netto_new.m       <- Ntot2.m
```

define ExcelSim's successive cycle's successive rows

```
n<-29
for (i in 2:n) {
#i<-2
```

```
NN1[i]             <- if (NN1[i-1]-NN1[i-1]*(death_rate*Time[i])+(NN1[i-1]-(NN1[i-1]*death_rate*Time[i]))*2^((Time[i]-Time[i-
1])/GenT_g) > 0) {NN1[i-1]-NN1[i-1]*(death_rate*Time[i])+(NN1[i-1]-(NN1[i-1]*death_rate*Time[i]))*2^((Time[i]-Time[i-1])/GenT_g)} else
{0}
NN2[i]             <- NN1[i]-NN1[i-1]
NN3[i]             <- if (NN2[i]>0) {NN2[i]} else {0}
N1[i]              <- if(Mu[i-1]>0) {N1[i-1]*2^((Time[i]-Time[i-1])/(log(2)/Mu[i-1]))} else {N1[i-1]} # 4
Netto_new[i]       <- N1[i]-N1[i-1] # 5
N1.m[i]            <- N1[i]*mutation_rate    #8
N2.m[i]            <- if (N1.m[i]>1) {floor(N1.m[i])} else {0}#9
Ntot2.m[i]         <- if (Mu.m[i-1]>0) {Ntot1.m[i-1]*2^((Time[i]-Time[i-1])/(log(2)/Mu.m[i-1]))} else {Ntot2.m[i-1]}#10
Ntot1.m[i]         <- N2.m[i] + Ntot2.m[i]#11
Netto_new.m[i]  <- Ntot2.m[i] - Ntot2.m[i-1]#12
Carbon[i]                  <- if(Carbon[i-1]-((Netto_new[i]+Netto_new.m[i]+NN3[i])*N3)<0) {0} else {Carbon[i-1]-
(Netto_new[i]+Netto_new.m[i]+NN3[i])*N3}
Mu[i]              <- if(Carbon[i]>0) {MuMax*(Carbon[i]/(Carbon[i]+Ks))} else {0}
N2[i]              <- if (Mu[i]<0.1) {probability*N1[i]} else {0} #6
```

```r
Mu.m[i]               <- MuMax.m*(Carbon[i]/(Carbon[i]+Ks))   #7
cycle[1]              <- j
cycle[i]              <- j
gen.cyc[1]            <- 1
gen.cyc[i]            <- i
}


my.list[[j]] <-
              data.frame(
                      Carbon,
                      Mu,
                      Time,
                      N1,
                      Netto_new,
                      N2,
                      Mu.m,
                      N1.m,
                      N2.m,
                      Ntot1.m,
                      Ntot2.m,
                      Netto_new.m,
                      NN1,
                      NN2,
                      NN3,
                      cycle,
                      gen.cyc
                      )
}



# create a data.frame from the list
my.data <- rbind(my.list[[1]],my.list[[2]])


for (k in 3:m) {
              my.data <- rbind(my.data,my.list[[k]])
              }


# store data in general directory
my.data <<- my.data
my.ratio <<- my.data$Ntot2.m[my.data$gen.cyc==29]/my.data$N1[my.data$gen.cyc==29]


# find the cycle closest to below ratio = 1
#close.to.1<<-1-min(1-my.ratio[1-my.ratio>0])
#my.ratio.dat <- data.frame (my.ratio, 1:length(my.ratio))
#my.ratio.dat
#cycle.below.ratio.1<<-my.ratio.dat[,2][my.ratio.dat$my.ratio==close.to.1]
```

```
#plot ratio
#dev.off()
quartz()
plot(1:length(my.ratio[my.ratio!=Inf]), my.ratio[my.ratio!=Inf], ylim=c(0,1.2), xlab="Cycles", ylab="Ratio (m/WT)")
abline(h=1, lty=2)
#abline(v=cycle.below.ratio.1)
#text(x=cycle.below.ratio.1+10, y=close.to.1 - 0.03,paste(cycle.below.ratio.1), col="red")
#points(cycle.below.ratio.1,close.to.1, col="red")
cex<-0.7
mtext(side=3, line=3, adj=0, cex=cex, paste("cycles","=",cycles))
mtext(side=3, line=3, adj=0.2,cex=cex,paste("probability","=",probability))
mtext(side=3, line=3, adj=0.5,cex=cex,paste("mutation_rate","=",mutation_rate))
mtext(side=3, line=3, adj=0.8,cex=cex,paste("Nm_start","=",Nm_start))
mtext(side=3, line=3, adj=1.0,cex=cex,paste("N_start","=",N_start))

mtext(side=3, line=2, adj=0,cex=cex,paste("GenT_g","=",GenT_g))
mtext(side=3, line=2, adj=0.2,cex=cex,paste("GenT_ng","=",GenT_ng))
mtext(side=3, line=2, adj=0.5,cex=cex,paste("death_rate","=",death_rate))
mtext(side=3, line=2, adj=0.8,cex=cex,paste("Total_C","=",Total_C))
mtext(side=3, line=2, adj=1.0,cex=cex,paste("Per_cell","=",Per_cell))

mtext(side=3, line=1, adj=0,cex=cex,paste("yield","=",yield))
mtext(side=3, line=1, adj=0.2,cex=cex,paste("Ks","=",Ks))
mtext(side=3, line=1, adj=0.5,cex=cex,paste("dilution","=",dilution))
mtext(side=3, line=1, adj=0.8,cex=cex,paste("Ks.m","=",Ks.m))
mtext(side=3, line=1, adj=1.0,cex=cex,paste("deltaT","=",deltaT))

mtext(side=3, line=0.1, adj=0,cex=cex,paste("µ(max)","=",round(MuMax,3)))
mtext(side=3, line=0.1, adj=0.2,cex=cex,paste("µ(max, mutant)","=",round(MuMax.m,3)))

# print parameter settings and ratios
my.parameters <<-
data.frame(cycles,probability,mutation_rate,Nm_start,N_start,GenT_g,GenT_ng,death_rate,Total_C,Per_cell,yield,Ks,dilution,Ks.m,
deltaT,MuMax,MuMax.m)
my.list<-list (my.parameters, my.ratio)
names(my.list)<-c("Parameters","End ratio after cycle (mutant/WT)")
my.list
}
```

R commands

```
#..Step 1. Clean work directory
rm(list=ls(all=TRUE))

#..Step 2. Create a workfolder
```

#..Step 3. Place into the workfolder the file death_sim_function.R

#..Step 4. Set your workfolder. Open R. Type below command
setwd("type you work folder address here") # example setwd("/Users/fedor/R_workfolder")

#..Step 5. import death.sim function
source("death_sim_function_code.R")

#..Step 6. Run a demonstration of death.sim() with default parameters
death.sim()

#..Step 7. Call output
my.parameters # function parameters
my.ratio # ratios at end of each cycle
my.data # raw data

#..Step 8. Set the parameters to your liking. All parameters you do not set will run with default parameters.
death.sim (cycles = 100, mutation_rate = 1E-3, dilution = 1)

#..Parameter	Default setting	Parameter description
#..cycles	300	Growth cycles
#..probability	0.03	Probability of a cell becoming transfer-competent
#..mutation_rate	10-5	mutation rate (per generation)
#..Nm_start	10	Number of mutants at start
#..N_start	50000	N cells at start
#..GenT_g	2.0	Generation time ICEclc transfer-competent cells (h)
#..GenT_ng	1.0	Generation time non-active cells (h)
#..death_rate	0.4	Death rate of ICEclc transfer-competent cells (1/h)
#..Total_C	1e-4	Total amount of substrate per cycle (g /mL)
#..Per_cell	2e-12	Amount of carbon per cell (g)
#..yield	0.3	Biomass yield (g/g)
#..Ks	2e-5	Ks (g/ml)
#..Ks.m	2e-5	Ks (g/ml, mutant)
#..deltaT	0.3	
#..dilution	0.001	Population dilution factor per cycle
#..MuMax	0.963	μ(max) (1/h)
#..MuMax.m	0.963	μ(max, mutant) (1/h)

Table SI 2. Simulation parameter default settings in the R function *death.sim()*

death.sim() parameter	Default setting	Parameter description
cycles	300	Growth cycles
probability	0.03	Probability of a cell becoming transfer-competent
mutation_rate	10^*	mutation rate (per generation)
Nm_start	10	Number of mutants at start
N_start	50000	N cells at start
GenT_g	2.0	Generation time ICEclc transfer-competent cells (h)
GenT_ng	1.0	Generation time non-active cells (h)
death_rate	0.4	Death rate of ICEclc transfer-competent cells (1/h)
Total_C	10^*	Total amount of substrate per cycle (g /mL)
Per_cell	$2*10^*$	Amount of carbon per cell (g)
yield	0.3	Biomass yield (g/g)
Ks	$2*10^*$	Ks (g/ml)
dilution	0.001	Population dilution factor per cycle
MuMax	0.963	μ(max) (1/h)
MuMax.m	0.963	μ(max, mutant) (1/h)

RESULTS & DISCUSSION

Algebraic analytical theory shows that an ICE*clc*-containing hosts showing fractional reproductive inhibition of transfer-competent cells would be outcompeted by mutants that shows no such reproductive inhibition under conditions with no genetic drift. It was suggested previously (4) that because ICE*clc*-mediated lysis in genus *Pseudomonas* takes place only in a small proportion of cells (3%-5%) under non-growing conditions, it does not represent a significant growth disadvantage to the host as to compromise population fitness. Here, we complement this finding by showing algebraic theory that suggests that this hypothesis does not hold for conditions without genetic drift (resampling with replacement or dilution). Consider bacterial growth expressed as

$$(1) \qquad y = 2^x,$$

where *y* is the number of cells, and *x* is the number of cell divisions. Assuming that after each division a certain percentage of cells would become reproductively inhibited, the growth formula changes to

$$(2) \qquad y = 2^x(1-a)^x,$$

with the added constant *a* representing the proportion of cells reproductively inhibited. However, we know that ICE*clc*-mediated reproductive inhibition is only ever activated under non-growing conditions and after an exponential growth phase (4) (2) (5). To suit such scenario the formula needs to be modified to:

$$(3) \qquad y = \left[2^x(1-a)\right]^b,$$

where *x* now becomes cell divisions per exponential growth phase and *b* is the number of exponential growth cycles after each of which fractional reproductive inhibition manifests. With this formula at hand, we can now calculate the number of growth cycles after which the ratio of mutant cell numbers to wildtype cell numbers reaches 1. For this, we equalize the equations for wildtype and mutant,

$$(4) \qquad \left[2^x(1-a)\right]^b = \left(2^x\right)^b c,$$

and resolve for the number of growth cycles *b*,

$$(5) \qquad b = \frac{\log(c)}{\log(1-a)},$$

where *c* is the mutation rate determining the mutant to wildtype ratio at the start of growth.

Note that *x* is cancelled out in equation (5) indicating that number of cell divisions per growth cycle do not affect the rate of change of wildtype-to-mutant ratio with respect to the number of growth cycles.

Assuming a 3% reproductive inhibition after each growth cycle, and mutant to wildtype ratios ranging from 10^{-4} to 10^{-9}, we see that mutants will start out-competing wildtype after between 300 and 700 growth cycles (Fig.1).
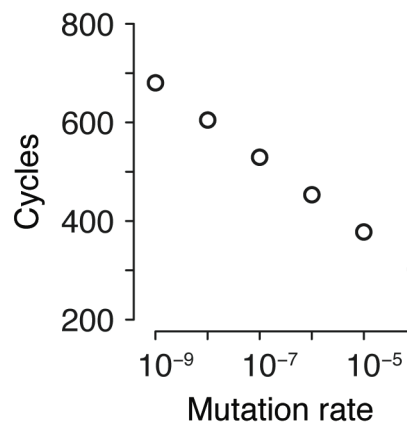

Fig.1. Growth cycles at mutant-to-wildtype ratio 1 as a function of mutation rate.

The function shows, mutants could still out-compete wildtype even though reproductive inhibition is restricted to stationary phase. It is also shown that the number of cell divisions in exponential phase (*x*, formulas 3 and 4), does not affect and is independent of the number of growth cycles (*b*, formulas 3 to 5) until catch-up.

Computer simulations show that genetic drift (dilution factor) is key to suppress mutant competitiveness. As a second approach to assess whether small proportion reproductive inhibition in non-growth phase represents a fitness cost we used computer simulation to predict mutant to wildtype ratios over growth cycles. Hereto, we translated the excel-based computer model as designed by (4) into R code creating the function *death.sim()* (SI R script). While previous results showed data for 21 growth cycles we used *death.sim()* to simulate results up to 1000 growth cycles. Like the previous model, *death.sim()* includes several settable parameters affecting the outcome. We found that the single most influential parameter determining whether mutant out-competes wildtype or not was the population dilution factor, which is a variable representing genetic drift.

When the population dilution factor was set sufficiently low (10^{-3}), both parameters mutation rate, and proportion of ICE*clc*-active cells, influenced the outcome such that when either was set to a low value, the mutant would fail to outgrow wildtype (as tested over 1000 cycles) (Fig.1.A). In contrast, at high proportion (0.25) or high mutation rate (10^{-5}) the simulation showed that mutant started to out-compete wildtype between after 20 and 200 cycles regardless of a low dilution factor (10^{-3}). Another series of simulations with the same parameter settings as above but this time at a high dilution factor setting (1) showed that mutants always outcompeted wildtype no matter if a low mutation rate or low proportion were set. Together, these results show that the dilution factor (genetic drift) is key to whether mutants will outgrow wildtype or not. Since genetic drift is always present, one can assume that transfer-competent mediated reproductive inhibition does not represent a high enough fitness disadvantage as to be selected against.
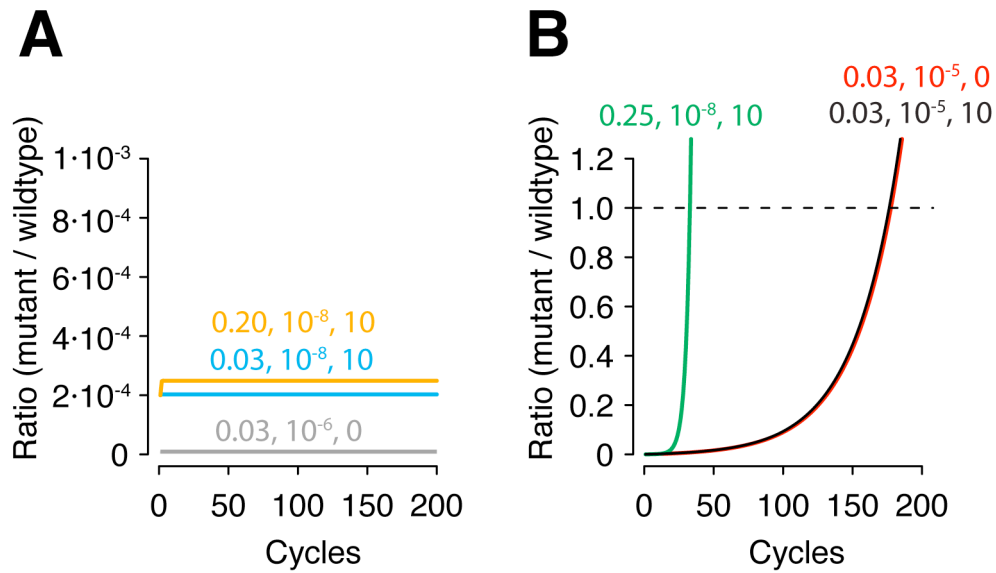
A

B

Fig. 2. Simulations with population dilution factor set to 0.001 showing ratio of mutant versus wildtype as a function of growth cycles. Coloured curves indicate output of different simulations with different parameter settings. Corresponding parameter settings for each curve are shown in matching colours. The numbers correspond to the probability of a cell becoming transfer-competent, the mutation rate per generation, and the number of mutants at the start of the simulation, respectively. All other simulation parameters correspond to the default parameters as described in (Table SI 2). Dotted horizontal line indicates a ratio of 1. The population dilution factor of all simulations was set to 0.001. **A.** Curves indicating no tendency of mutants to out-compete wildtype. **B.** Curves with a tendency of mutants to out-compete wildtype.
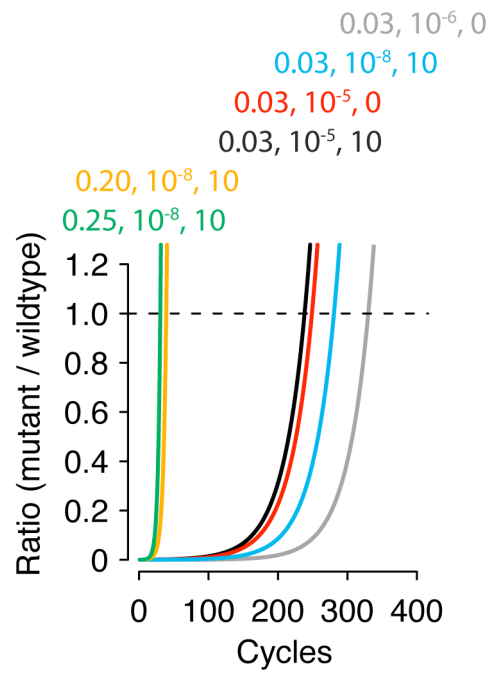
Fig. 3. Simulations with population dilution factor set to 1 showing ratio of mutant versus wildtype as a function of growth cycles. Same as Fig.2.A and B but with population dilution factor set to 1. Dotted horizontal line indicates a ratio of 1, beyond which mutants out-compete wildtype.

ACKNOWLEDGEMENTS

REFERENCES

1.   **Ackermann, M., B. Stecher, N. E. Freed, P. Songhet, W. D. Hardt, and M. Doebeli.** 2008. Self-destructive cooperation mediated by phenotypic noise. Nature **454:**987-990.

2.   **Gaillard, M., N. Pradervand, M. Minoia, V. Sentchilo, D. R. Johnson, and J. R. van der Meer.** 2010. Transcriptome analysis of the mobile genome ICE*clc* in *Pseudomonas knackmussii* B13. BMC Microbiol **10:**153.

3.   **Miyazaki, R., M. Minoia, N. Pradervand, S. Sulser, F. Reinhard, and J. R. van der Meer.** 2012. Cellular variability of RpoS expression underlies subpopulation activation of an integrative and conjugative element. PLoS Genet **8:**e1002818.

4.   **Reinhard, F., R. Miyazaki, N. Pradervand, and J. R. van der Meer.** 2013. Cell differentiation to "mating bodies" induced by an integrating and conjugative element in free-living bacteria. Curr Biol **23:**255-259.

5.   **Sentchilo, V., R. Ravatn, C. Werlen, A. J. Zehnder, and J. R. van der Meer.** 2003. Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. Strain B13. Journal of Bacteriology **185:**4530-4538.

CURRICULUM VITAE

Friedrich Fedor Reinhard

Born August 23 1979 in Munich, Germany

1999   Hochschulreifeprüfung at the Deutsche Höhere Privatschule in Windhoek, Namibia.

2003   **Bachelor of Science** at the University of Edinburgh, Scotland.
Bachelor of Science thesis under the supervision of Prof. Graeme Reid in the Department of Microbiology at the University of Edinburgh.
Title of thesis: **Dimethylsulfoxide reduction in *Shewanella oneidensis.***

2005   **Master of Research** at the University of Nottingham, England.
Master of Research thesis under the supervision of Dr. Martin Krüger in the Department of Microbiology at the Max Planck Institute for Marine Microbiology and Prof. Christine Dodd in the Division of Food Sciences at the University of Nottingham.
Title of thesis: **Sulfate reducing genes at sites with high activity of anaerobic oxidation of methane.**

2014   **PhD** at the University of Lausanne, Switzerland
PhD thesis under the supervision of Prof. J. R. van der Meer in the Department of Fundamental Microbiology.
Title of thesis: **A single cell approach to studying the ICE*clc* behaviour in *Pseudomonas*.**

Awards   Les Doctorials de UNIL 2011 Poster Award (Lausanne, Switzerland, 2011), David C. White Most Innovative Poster Award (Seattle, USA, 2010), ISME Poster Award (Seattle, USA, 2010), ISME Travel Award (Seattle, USA, 2010), D-day Poster Award (Lausanne, Switzerland, 2009)

LIST OF PUBLICATIONS

Reinhard, F., and J. R. van der Meer. 2014. Life history analysis of integrative and conjugative element activation in growing microcolonies of *Pseudomonas*. Journal of Bacteriology (accepted, doi: 10.1128/JB.01333-13).

Reinhard, F., and J. R. van der Meer. 2013. Improved statistical analysis of low abundance phenomena in bimodal bacterial populations. PLOS ONE 8:e78288

Reinhard, F., R. Miyazaki, N. Pradervand, and J. R. van der Meer. 2013. Cell differentiation to "mating bodies" induced by an integrating and conjugative element in free-living bacteria. Current Biology 23:255-9.

Miyazaki, R., M. Minoia, N. Pradervand, S. Sulser, F. Reinhard, and J. R. van der Meer. 2012. Cellular variability of RpoS expression underlies subpopulation activation of an integrative and conjugative element. PLoS Genetics 8:e1002818.

Miyazaki, R., M. Minoia, N. Pradervand, V. Sentchilo, S. Sulser, F. Reinhard, and J. R. van der Meer. 2011. The *clc* element and related genomic islands in *Proteobacteria*. In A. P. Roberts and P. Mullany (ed.), Bacterial integrative mobile genetic elements. Landes Bioscience.

Reinhard, F., and J. R. van der Meer. 2010. Microcolony growth assays p. 3562-3570. In K. N. Timmis, V. de Lorenzo, T. McGenity, and J. R. van der Meer (ed.), Handbook of Hydrocarbon and Lipid Microbiology, vol. 5. Springer Verlag.

Minoia, M., M. Gaillard, F. Reinhard, M. Stojanov, V. Sentchilo, and J. R. van der Meer. 2008. Stochasticity and bistability in horizontal transfer control of a genomic island in *Pseudomonas*. Proc. Natl. Acad. Sci. U S A 105:20792-20797.

Reinhard, F. 2005. Sulfate reducing genes at sites with high activity of anaerobic oxidation of methane. Max Planck Institute for Marine Microbiology, Bremen. (Master of Research thesis).

Reinhard, F. 2003. Dimethylsulfoxide reduction in *Shewanella oneidensis*. University of Edinburgh, Edinburgh. (BSc. Honours thesis).