

GENOMIC PLASTICITY AND CHROMSOMAL COPY NUMBER OF

ACINETOBACTER BAYLYI ADP1

by

LAURA ELIZABETH CUFF

(Under the Direction of Ellen L. Neidle)

ABSTRACT

Acinetobacter baylyi ADP1, a ubiquitous soil bacterium, is an excellent model organism for systematic studies of gene duplication and amplification. This dissertation describes the role of a mobile insertion sequence element, IS1236, in chromosomal duplication and amplification events. Genetic rearrangements were further explored by characterizing the potential for recombination between copies of its single circular chromosome. Genomic copy number was evaluated at various growth stages and rates.

Gene amplification events are difficult to study due to their transient nature; however, by exploiting the pathway for benzoate consumption in *A. baylyi*, chromosomal amplification can be selected and maintained. *A. baylyi* strains lacking two transcriptional activators acquire the ability to grow on benzoate via amplification of a chromosomal region encompassing the *catA* gene and the *catBCIJFD* operon. Growth on benzoate selects amplification mutants, allowing detailed study of the recombination events generating the initial duplication. To expand this experimental system, the *cat*-gene DNA was relocated for the selection of amplification mutants in different chromosomal regions. Analysis of these amplification mutants revealed that proximity to one of six

chromosomal copies of IS1236 increases the likelihood that the insertion sequence is involved in the initial duplication event. Surprisingly, the mechanism underlying duplication formation differs from the well-characterized method of transposition and intra-molecular recombination between duplicated IS copies.

To determine whether duplication events might involve transposition and inter-molecular recombination between IS copies, the ploidy level of *A. baylyi* was studied. Quantitative PCR analysis of DNA content in cells was carried out on bacterial cultures at various growth stages and rates. This analysis showed *A. baylyi* contains a single copy of the chromosome per cell in both exponential and early-stationary phase cultures. Additionally, a single copy of the chromosome per cell was found in cultures with varying doubling times. These results make it unlikely that frequent recombination occurs between IS elements on different chromosomal copies within the same cell. The studies described herein clarify the ploidy level in *A. baylyi* and suggest a novel method of duplication involves DNA cleavage by the transposase encoded by IS1236.

INDEX WORDS: *Acinetobacter*, Gene Regulation, Gene Amplification, Genome Plasticity, Ploidy

GENOMIC PLASTICITY AND CHROMSOMAL COPY NUMBER OF
ACINETOBACTER BAYLYI ADP1

by

LAURA ELIZABETH CUFF
B.A., Grinnell College, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

© 2013

Laura Elizabeth Cuff

All Rights Reserved

GENOMIC PLASTICITY AND CHROMSOMAL COPY NUMBER OF

ACINETOBACTER BAYLYI ADP1

by

LAURA ELIZABETH CUFF

Major Professor: Ellen Neidle

Committee: Claiborne Glover
Anna Karls
Lawrence Shimkets

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2013

ACKNOWLEDGEMENTS

I gratefully acknowledge the members of my committee: Anna Karls, Larry Shimkets, and Claiborne Glover. I appreciate the guidance and knowledge each provided throughout the course of my research. I would like to thank my advisor, Ellen Neidle, in particular. Without her support and encouragement, I would not be the scientist I am today. I am grateful for the former and current members of the Neidle lab: Sarah Seaton, K.T. Elliott, Melissa Tumen-Velasquez, Nicole Laniohan, Stephanie Thurmond, and Maliha Ishaq. Thank you for everything I have learned from you each and every day. I appreciate the opportunities you have given me to grow as a scientist and a person.

I am forever indebted to my family for their unending support, love, and patience. I never would have achieved this goal without each of you providing your love in your own ways. I am amazed every day by your constant ability to help me see the bigger picture and to take pleasure from life's smallest details. To my dear friends near and far who have been physically and emotionally present for me, through the good and bad times, the fun, not-so-fun, and just downright tedious days, thank you from the bottom of my heart. I would not be where I am now without all of your influences. You each deserve a blue ribbon. Here it goes!

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	ix
CHAPTER	
1 INTRODUCTION AND LITERATURE REVIEW	1
Genetic elements affect genomic plasticity	2
Phenotypic results from changes in gene dosage.....	4
A model for the formation of gene duplication and amplification events ...	5
Emerging technologies for studying genomic plasticity.....	7
Gene amplification as a tool for the production of industrially and pharmaceutically relevant substances	13
Using <i>A. baylyi</i> ADP1 to study gene amplification events systematically	17
Final Remarks	20
References.....	23
2 ANALYSIS OF IS1236-MEDIATED GENE AMPLIFICATION EVENTS IN <i>ACINETOBACTER BAYLYI</i> ADP1	35
Abstract.....	36
Introduction.....	37
Materials and Methods.....	39

Results and Discussion	42
References.....	57
3 GENOMIC COPY NUMBER OF <i>ACINETOBACTER BAYLYI</i> ADP1	85
Abstract.....	86
Introduction.....	87
Materials and Methods.....	88
Results.....	91
Discussion.....	92
References.....	95
4 CONCLUSIONS	105
References.....	110

APPENDICES

A GENOME WIDE SELECTION FOR INCREASED COPY NUMBER IN <i>ACINETOBACTER BAYLYI</i> ADP1: LOCUS AND CONTEXT-DEPENDENT VARIATION IN GENE AMPLIFICATION.....	113
Summary.....	114
Introduction.....	115
Results.....	117
Discussion.....	125
Concluding Remarks.....	132
Experimental Procedures	133
Acknowledgements.....	141
References.....	142

LIST OF TABLES

	Page
Table 2.1: Strains and plasmids used in this study	74
Table 2.2: Predicted fragment sizes for PCR analysis.....	76
Table 2.3: Predicted fragment sizes for Southern hybridization.....	77
Table 2.S1: Primers used in this study.....	82
Table 2.S2: Amplicon analysis of Ben ⁺ mutants	84
Table 3.1: Primers used in this study	98
Table 3.2: Origin and termini copy numbers in pyruvate-grown early stationary phase <i>Acinetobacter baylyi</i> ADP1	99
Table 3.3: Origin and termini copy numbers in pyruvate-grown exponential phase <i>Acinetobacter baylyi</i> ADP1	100
Table 3.4: Origin and termini copy numbers in succinate-grown exponential phase <i>Acinetobacter baylyi</i> ADP1	101
Table 3.5: Origin and termini copy numbers in pyruvate grown <i>Escherichia coli</i> B.....	102
Table A.1: Strains and plasmids used in this study	157
Table A.S1: Characteristics of gene amplification mutants in this study.....	165
Table A.S2: Pearson's correlation coefficient analysis of the relationship between amplicon size and copy number.....	167
Table A.S3: Mann-Whitney analysis of pair-wise differences in <i>catA</i> copy number (p- values shown).....	168

Table A.S4: Mann-Whitney analysis of pair-wise differences in amplicon size (p-values shown).....	169
---	-----

LIST OF FIGURES

	Page
Figure 1.1: Experimental system for the study of GDA in <i>Acinetobacter baylyi</i> ADP1 ...	34
Figure 2.1: Selection and analysis of Ben ⁺ amplification mutants	64
Figure 2.2: Position of IS elements on circular map of the ADP1 chromosome.....	66
Figure 2.3: Roles for IS1236 in gene amplification.....	67
Figure 2.4: Amplicon size analysis by PFGE	69
Figure 2.5: IS1236 sequences identified in duplication junctions	70
Figure 2.6: PCR-based analysis of chromosomal configurations.....	71
Figure 2.7: Analysis of chromosomal configuration by Southern hybridization.....	72
Figure 2.S1: Sequence alignment of IS1236_1 and IS1236_6	78
Figure 2.S2: Transposition and 3-bp flanking repeats: different models of duplication predict the same chromosomal configuration in ACN1039-derived mutants	80
Figure 2.S3: Transposition and 3-bp flanking repeats: different models of duplication predict the same chromosomal configuration in ACN1050-derived mutants	81
Figure 3.1: Growth curves of <i>A. baylyi</i> ADP1 cultures for chromosomal copy number analysis.....	103
Figure 3.2: Growth curve of <i>E. coli</i> B cultures for chromosomal copy number analysis	104
Figure A.1: Selection for gene amplification by growth on benzoate	149
Figure A.2: Genomic analysis of amplification mutants	150

Figure A.3: DNA sequences involved in recombination to generate tandem duplication of the native <i>ben-cat</i> region.....	152
Figure A.4: Relationship between <i>catA</i> copy number and amplicon size	153
Figure A.5: Genome analysis of a Ben ⁺ duplication mutant (ACN1135)	154
Figure A.6: Transformation assay to determine spontaneous duplication frequency.....	155
Figure A.7: Relative <i>catA</i> duplication frequency at different loci.....	156
Figure A.S1: Genetic engineering allows relocation of <i>cat</i> gene region to loci throughout the ADP1 genome	162
Figure A.S2: Duplication junction sequences in ACN293-derived mutants selected on anthranilate.....	163
Figure A.S3: Genome analysis of a duplication mutant	164

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Genomic plasticity encompasses changes in the presence, direction, and copy number of DNA sequences. Changes at this level may be stochastic and transient, yet they facilitate adaptation to rapidly changing environments. This dissertation details studies of the plasticity of the genome of a soil bacterium, *Acinetobacter baylyi* ADP1. Topics include genetic changes as a result of gene amplification and the use of gene amplification as a tool to study chromosomal evolution.

The versatility of *A. baylyi* ADP1 affords opportunities for novel metabolic and genetic approaches. In addition, the natural transformation capabilities of this organism lead to tractability for extensive engineering of the chromosome [reviewed in (1-3)]. The ease with which the genome can be manipulated invites a deeper study of genetic rearrangements inherent in adaptation to environmental changes. Such rearrangements can be selected and maintained in response to nutrient variations and availability, the presence of toxins, or other environmental cues. Changes in protein levels result from regulatory changes at the transcriptional and translational level, but can also stem from variation in the chromosomal copy number of specific genes. The plasticity of the genome can be difficult to study in depth, due to the transiency of changes in genomic content. However, techniques to study these changes, such as high throughput sequencing, are becoming more accessible and less costly.

Gene amplification, an increase in the copy number of a segment of DNA relative to flanking regions, can confer a selective advantage (4-9). Amplification can act to regulate transcript levels temporarily via changes in gene dosage. This genetic malleability has medical and evolutionary significance, as gene amplification has been observed in all organisms, including bacteria, plants, insects, viruses, and humans (6, 10-13). Duplicated genes allow mutations to accrue in some gene copies, while the original function is maintained in others, permitting cells to adapt to environmental conditions (4, 7, 14, 15). In some cases, advantageous mutations are selected that allow new functions to evolve (16).

Genetic elements affect genomic plasticity

Mobile elements impact chromosomal content

Insertion sequence (IS) elements impact the plasticity and content of genomes in all classes of organisms. Duplications, inversions, and deletions that cause genome reorganization can occur through transposition of IS elements (17). IS movement affects gene presence, expression, and content. Large chromosomal inversions are one potential result of IS elements; IS6100 has been shown to generate chromosomal inversions in *Pseudomonas* affecting gene expression levels (18). Additionally, drug resistance in bacteria can be mediated by mobile DNA elements. For example, in *Pseudomonas aeruginosa*, acquisition of the transposon Tn6061 confers resistance to multiple antibiotics (18). Movement of these elements can cause significant changes in both genetic content and expression levels.

Transposable elements are frequently flanked by inverted repeating DNA sequences, called terminal inverted repeats. Recombination between these repeats leads to inversions of the intervening DNA sequences and may also generate deletions on the chromosome when multiple repeats are in tandem array (19). Complete deletions of insertion sequences were thought to be rare because of the paucity of DNA end joining mechanisms in bacteria. However, a recent study shows that an insertion sequence-excision enhancer (IEE) exists in *Escherichia coli* to excise IS629, an IS3 family member frequently found in O157 strains (20). Phylogenetic studies of the IEE suggest its presence in a wide range of bacteria and that the proteins that mediate transposon excision coevolved with the IS elements (20). The deletion of IS elements leads to variation in the chromosomal content and, depending on the genes that are removed from the chromosome, can have a strong impact on the resulting environment in which the bacterium can survive.

Transposable elements in Acinetobacter species

IS elements are common in the *Acinetobacter* genus and have been shown to affect levels of antibiotic resistance in the pathogenic members of the *Acinetobacter baumannii* complex (21). These pathogens are most frequently isolated from nosocomial infections and are often resistant to multiple antibiotics. Transmission of mobile elements within this species leads to difficulty in treating infections because of rapidly spreading multiple drug resistances (22). One example of a mobile element affecting antibiotic resistance is the insertion of IS*Aba825* in the promoter of *bla*_{OXA-58}, a gene that modulates levels of carbapenem resistance. The insertion generated a hybrid promoter for *bla*_{OXA-58}, leading to increased gene expression and carbapenem resistance (23). This

antibiotic is considered a drug of last resort, as it is one to which bacteria containing the common multiple drug resistance cassettes are still susceptible. Carbapenem resistance in *A. baumannii* can also be increased by another IS element, *ISAbal10*, which confers an additional promoter sequence to *bla_{OXA-23}* in some clinical isolates (24). With the additional promoter, higher expression of *bla_{OXA-23}* leads to increased carbapenem resistance.

In addition to affecting levels of antibiotic resistance, IS elements play a role in gene duplication and amplification. A third IS element in *A. baumannii*, *ISAbal25*, appears to have duplicated the *bla_{OXA-58}* gene which resulted in reduced susceptibility to carbapenem (25). Additionally, *ISAbal25* has been shown to act as a composite transposon, where two copies of the IS transpose acting as a single unit relocate any genes contained between the two elements, including those for carbapenem resistance (26). The mobility of genetic elements containing genes conferring resistance to antibiotics of last resort is troubling and merits further study. Rearrangement of the genome by IS elements can be transient or permanent, but the recombination that occurs has far-reaching effects.

Phenotypic results from changes in gene dosage

Gene amplification leading to overexpression of gene products has been implicated in increased virulence and the acquisition of antibiotic resistance in various bacteria. In *Haemophilus influenzae*, a stable tandem duplication of genes involved in capsule formation can lead to further amplification and overexpression of these capsular genes, thereby increasing invasiveness of the pathogen (27). Other examples of gene

amplification leading to increased virulence can be found in the *ctx* (cholera toxin) genes of *Vibrio cholerae* and the *alg* genes (alginate biosynthesis) in *Pseudomonas aeruginosa* (6, 28). Additional gene amplification events leading to antibiotic resistance are found in *Streptococcus faecalis*, which has an increased resistance to tetracycline resulting from amplified plasmid genes, and *E. coli*, in which amplified chromosomal regions lead to increased levels of ampicillin resistance (6). A review detailing these medically relevant bacterial gene duplications and amplifications has been published (11).

Gene amplification has evolutionary significance in addition to its medical relevance. It can serve as a means for adaptation to environmental stressors, as multiple copies of genes may allow a selective advantage under conditions in which growth is limited (5, 29-33). Amplification is associated with resistance to heavy metals, growth on exotic carbon sources, and growth at non-optimal temperatures (6, 34). Although amplification occurs frequently, tandem arrays are unstable due to the large regions of homology that can be deleted by homologous (RecA-mediated) recombination. This instability is one of the difficulties of studying gene duplication and subsequent amplification [reviewed in (35)]. However, the *Acinetobacter* model system described in Chapter 2 of this dissertation allows systematic characterization of prokaryotic gene amplification events.

A model for the formation of gene duplication and amplification events

A canonical model of gene duplication and amplification (GDA) stems from seminal studies in *E. coli* and *Salmonella typhimurium* (6, 10, 12, 13, 34). Briefly, a recombination event occurs via a RecA-dependent mechanism between two regions of

DNA with sequence identity. This initial recombination leads to a duplication of intervening DNA sequences. Notably, such tandem duplications can affect nearly any locus; however, various loci are affected at different frequencies (29, 34, 36). Following the initial duplication, RecA-mediated homologous recombination occurs between the large regions of sequence identity, allowing further amplification or deletion of these homologous stretches of DNA. The large size and sequence identity of the duplicated regions can increase the recombination frequency by as much as 150-fold compared to the wild type (6, 37). To summarize the model, gene amplification is a sequential process in which the rate of amplification is limited by the formation of the initial gene duplication.

While the model indicates the initial duplication is RecA-dependent, there are instances where the duplication does not depend on sequence identity and therefore is likely illegitimate (RecA-independent) (33, 36, 38, 39). Various mechanisms for duplication and amplification that are RecA-independent have been proposed. Mechanisms for illegitimate recombination include ligation of double strand breaks by DNA gyrase (40), strand exchange between sister chromosomes (41), and strand slippage during DNA replication or repair (42). In a study in *Salmonella enterica*, duplications formed at high rates (10^{-4} /cell), and RecA-independent recombination events formed with approximately an 11-fold lower frequency (43). Several of these recombination events resulted from activity by an IS3 transposase. Additionally, RecA was shown to be necessary for the loss of duplications, but not for their formation (44).

Duplication events are typically difficult to study systematically and detect experimentally. The size, location, and content of GDA events all factor into the relative

ease with which they can be studied. Methods for studying such events using 454 pyrosequencing were recently developed and are described below (45, 46). Maintenance of GDA on the chromosome also influences the ability to observe and characterize such events. As technology advances, more techniques with the capacity to study GDA events are becoming readily accessible.

Emerging technologies for studying genomic plasticity

Observing GDA events can be difficult due to the transiency of gene amplification. Next-generation sequencing (NGS) technologies (Solexa, 454 pyrosequencing, etc.) have recently improved the study of gene amplification events by providing detailed information and large-scale genomic coverage. Initially, most microbial NGS studies focused on determining diversity in soil or marine environments [reviewed in (47) and (48)], but these NGS technologies can also accommodate the study of microbial communities with clinical importance [reviewed in (49)]. Most studies observing copy number variation (CNV) have been undertaken in human subjects, often relating to cancer; however, there is no reason these techniques cannot also be applied to studying changes in gene copy number in microbes.

Pyrosequencing

Pyrosequencing can be used to determine changes in copy number at any locus, whether by duplication and/or amplification or through deletion. Copy number changes can be detected by inserting single nucleotide variations into reference and query probes in a method known as reference/query pyrosequencing (RQPS) (46). This method depends on quantitative pyrosequencing and probes designed to contain a single

nucleotide differing from the wild-type gene sequence. The reference probe recognizes a gene in known copy number, while the query probe is designed to recognize the gene of interest. If these reference and query probes are ligated to produce a single RQPS probe, then the molar ratio of each single nucleotide variation is fixed at 1:1. The RQPS probe can then be mixed with genomic DNA in varying ratios to determine copy number. Following quantitative pyrosequencing, the ratio between the introduced single nucleotide variation and the wild-type sequence found in the DNA mixture corresponds to the copy number of the query locus.

Liu *et al.* employed this RQPS method to determine the gender of animals by querying an *X*-linked gene, where females have two copies and males have a single copy of the gene. The gender of all animals was properly identified in this blind study (46). Additionally, array-based comparative genomic hybridization results measuring gene deletion or amplification at a specific locus were confirmed using RQPS, reiterating the effectiveness of using RQPS to detect changes in gene copy number. These methods can be readily applied to microorganisms for observation of copy number change at specific loci.

Real-time quantitative PCR

Another method for studying gene copy number changes at specific loci is real-time quantitative PCR (qPCR). Briefly, qPCR measures the amount of DNA under specific conditions. Two common methods of measuring DNA are TaqMan-based fluorescently labeled probes, which bind a specific sequence, and SYBR Green, which binds to double-stranded DNA. When levels of fluorescence begin to increase exponentially, the sample crosses the cycle threshold (C_T). Comparison between the C_T

values for the unknown DNA samples and the standard curve allow the determination of the precise amount of DNA in the starting reaction.

In one study, *rrn* copy numbers from *E. coli* were quantified using real-time PCR with SYBR Green I detection (50). The absolute copy number of *rrn* was determined from the C_T values in comparison to the standard curve and found to be 6.9. Relative copy number analysis of *rrn* in comparison to *dxs*, a gene known to be in single copy, returned the same result (50). The relative copy number was determined using a calibrator sample, which contained plasmid DNA with *dxs* and *rrn* both in single copy number. These values are very close to the copy number of 7 for *rrn* found using genome sequencing data (51).

Real-time qPCR was also used to determine bacterial gene copy number in other studies. Seaton *et al.* and Cuff *et al.* observed the copy number of genes necessary for aromatic carbon compound metabolism in *A. baylyi* (36, 38). Spontaneous gene amplification mutants were isolated, and the copy number of the amplified DNA was characterized using TaqMan-based qPCR. In these studies, the copy numbers of the amplified regions of the chromosome were compared to the copy number of a region of the chromosome assumed to be in single copy. The resulting ratios indicated the number of copies of the amplified region maintained on the chromosome.

When designing experiments to observe gene dosage using qPCR, it is important to take several things into consideration. Primer design and validation are crucial, as differing amplification efficiencies can skew results. qPCR has the ability to detect copy number variation at targeted loci rapidly and in a cost-effective fashion, but experiments must be judiciously designed and adhere to published standards for quality (52-54).

Microfluidics

In addition to conventional qPCR, microfluidic chips are now capable of high-throughput determinations of the dosage of many genes. A microfluidic chip requires less pipetting and therefore reduces pipetting error when compared to standard qPCR. These chips have chambers into which both template and primers are pushed via pressure and can quantify DNA with very little template. This technique increases the ability to test multiple samples and primer sets to be tested concurrently. Microfluidic arrays can be used to analyze gene expression or copy number levels.

The Fluidigm 48.48 dynamic chip is a commercially available single use chip that allows analysis of 48 samples with 48 different assays. This 48.48 chip was used in Spurgeon *et al.* to provide amplification curve data from the 18 different types of tissue samples and 44 different genes available in the GeneNote database over a 6-fold range (55). When these data were compared to qPCR-derived data, it was found to be of comparable quality. However, the microfluidic array can simultaneously analyze a greater number of genes than qPCR. When compared to publicly available microarray data for the same genes, the microfluidic array was found to result in more precise data. Additionally, the reproducibility within and between microfluidic chips was very high ($r > 0.99$).

In a study of the quantitative resolution of qPCR, Weaver *et al.* were able to distinguish between small copy number changes of chromosomal loci. An error model predicts that a 1.1-fold difference in gene copy number (the difference between 10 and 11 copies) can be distinguished by qPCR with sufficient replicates (56). Since large numbers

of replicates can be unwieldy using conventional 96-well plates, high throughput microfluidic platforms allow the highest level of quantitative resolution.

Digital PCR

Digital PCR (dPCR), an emerging method for DNA quantification, has been shown to be more sensitive for detecting small variations in copy number when compared to qPCR (57, 58). dPCR allows the absolute quantification of target genes by limiting dilutions without requiring standards. A limiting dilution of the DNA – dilution to the extent that most chambers will contain negative amplification results – allows quantification of DNA that is independent of the amplification efficiency. Using microfluidic devices, the need for serial dilution of the template DNA is abolished. On these microfluidic chips, DNA is partitioned into chambers where the reaction occurs. Chambers containing amplification are scored as positive events, whereas chambers with no amplification are scored as negative events. Counting the number of positive events, the target gene can be quantified regardless of the efficiency of amplification. Quantification of DNA prior to sequencing reactions consumes billions of DNA molecules; however, dPCR can greatly decrease the need for large amounts of DNA for NGS analysis. dPCR is an absolute, calibration-free method for determining the concentration of sequencing library molecules.

In a study of *HER2* copy number in genomic DNA from different breast cancer cells lines, dPCR was able to detect a smaller fold change in copy number than qPCR (1.17 compared to 1.27) (58). The authors also argue that dPCR is less variable between laboratories than qPCR, providing more accurate results. When the coefficient for variation (CV) for libraries assayed in replicate was determined, the mean CV for dPCR

was $11.8 \pm 1.5\%$. This CV was determined using a universal template approach, in which a probe-binding sequence is fused to one of the PCR primers (59). However, the CV for qPCR with this same approach was $21.2 \pm 2.6\%$. As there are neither internal nor external standards for the dPCR, this value corresponds to the real-world accuracy of the dPCR method. dPCR can be used to quantify DNA to allow high quality sequencing data from nanogram amounts of DNA.

dPCR was also used to observe the inhibitory effects of various compounds on PCR (60). Humic acids inhibit dPCR; however, once amplification occurs following the delay in reaction initiation, accurate copy numbers can be determined. dPCR was less affected by inhibitory compounds than qPCR (60). Although dilution gets rid of the inhibitory compounds, the DNA template is also diluted. If there is enough starting template, dilution is a viable option to remove inhibitory substances. dPCR estimates the number of molecules of DNA, not gene copies. To determine gene copies using dPCR, fragmentation of the template DNA may be an option. Increased fragmentation levels would make the occurrence of multiple gene copies within a single chamber less likely.

CNV-Seq

A technique termed CNV-Seq analyzes copy number variation between individuals and can be used to observe gene dosage in bacterial populations. CNV-Seq is based on a statistical model that uses a known sequence as a template and aligns two (or more) sets of shotgun sequencing reads to this template – each shotgun read comes from a different individual or clonal population (61). The developers of this technique simulated genomes with random CNVs and also compared CNV between two human individuals (Dr. James Watson and Dr. J. Craig Venter). The CNVs that were found with

the statistical model correlated well with those found in the genome sequences. The results show that the number of reads sequenced and not the read length is most important in determining the sensitivity and specificity of CNV evaluation.

The study of gene amplification across the chromosome is becoming easier with the increased availability of full genome sequencing, dPCR, and CNV-Seq. The technologies that are described here can reveal various levels of amplification, including changes in copy number that occur both globally and at specific loci. Minute changes in copy number can be differentiated with proper techniques and replicates. Using the methods described above, gene copy number can be precisely determined. With this knowledge, microbial production of pharmaceutically or industrially important compounds importance can be tightly regulated and optimized.

Gene amplification as a tool for the production of industrially and pharmaceutically relevant substances

Chemically induced gene amplification

As GDA becomes a more widely acknowledged phenomenon, its potential as a tool for microbial engineering is being developed. Gene amplification can be used to engineer microbial strains to produce products of industrial importance while minimizing toxic by-products (62). Historically, pathway construction and maintenance of expression levels were achieved using plasmid-based expression methods. However, plasmids are not stably maintained in cells for two reasons: segregational instability, which leads to plasmid-free daughter cells, and allele segregation, which results in non-productive selection-resistant plasmids. Chromosomal gene amplification can circumvent

this instability if selection is used to maintain multiple copies of the biosynthetic genes. However, in the absence of selection, the amplifications are generally lost through homologous recombination events that lead to a single copy of the gene(s) on the chromosome. Therefore, another means of stabilizing amplified DNA is to prevent homologous recombination after optimal gene dosage of a desired region is reached.

To achieve stable maintenance of amplification on a bacterial chromosome, Tyo *et al.* developed a system in *E. coli*, chemically inducible chromosomal evolution (CIChE), to engineer biosynthetic pathways that will be inherited by all daughter cells rather than being subjected to segregation (62). CIChE provides high gene copy number to facilitate expression using tandem head-to-tail repeats, which are then stabilized on the chromosome. The first step is to generate a DNA cassette containing the gene of interest and a selectable antibiotic marker flanked by two copies of a noncoding 1-kb region. These noncoding regions contain little homology to the *E. coli* chromosome, allowing specific amplification of the DNA cassette. Following integration of the cassette onto the chromosome, the large noncoding regions are candidates for homologous recombination events that generate extra copies of the integrated cassette. In the presence of increasing levels of the appropriate antibiotic (chloramphenicol), the cassette is amplified to higher gene copy numbers by RecA-dependent homologous recombination. When the cassette reaches the optimal copy number for compound production, *recA* is deleted from the chromosome. Once *recA* is deleted, the copy number of the cassette is stabilized, as deletions via homologous recombination cannot occur.

CIChE shows improved yields over plasmid-based yields for two well-studied pathways for production of poly-3-hydroxybutyrate (PHB), a biopolymer, and lycopene,

a polyisoprenoid (62). In the presence of antibiotic selection to maintain the plasmid, accumulated PHB was equivalent to CICH_E-derived levels; however, in the absence of selection, 30% less PHB was produced by the plasmid-carrying cells. The plasmid-based system lost productivity after 5 doublings without antibiotic selection and after 40 generations when the plasmid was maintained in the cells. This productivity loss was not found in the CICH_E-based strain, which maintained more than 90% of its productivity for 70 generations. CICH_E-produced levels of lycopene were also compared to plasmid-based systems. Yields for lycopene derived from CICH_E constructs were 60% greater than the plasmid-based yields, even under antibiotic selection for plasmid maintenance (62). CICH_E-based methods of pathway engineering take advantage of gene amplification to improve production levels of compounds with industrial or clinical importance, while minimizing extraneous or deleterious by-products.

Increased promoter copy number

Another method of increasing gene expression can occur through an increase in tandem promoter sequences upstream of a gene of interest. Li *et al.* (63) generated a system to observe increased gene expression under incrementally increasing tandemly repeated promoters. Using the core-*tac*-promoter, the expression of green fluorescence protein (GFP) was observed. The study found that the measured fluorescence of GFP was greatest with five tandem promoter sequences.

To test the system further, the *phaCAB* operon, which encodes PHB, was engineered to be under control of the five tandem core-*tac*-promoters (63). *phaCAB* was fused downstream of the promoter cluster and then the whole construct was integrated onto the *E. coli* chromosome. PHB accumulation was found to be 5.6-fold higher than in

a strain with only a single copy of the core-*tac*-promoter controlling gene expression (63). RT-qPCR analysis of transcriptional activity of these genes showed approximately a six-fold increase in comparison to the single copy promoter.

Gene amplification for production of actinorhodin

Gene amplification can also be used to produce pharmaceutically relevant metabolites. Stabilized gene amplification in microbial strains that produce such metabolites can lead to higher yields of these compounds. Actinomycetes, such as *Streptomyces* species, are frequently used to overproduce these metabolites. Tandem head-to-tail amplification of gene clusters necessary for the biosynthesis of antibiotics can lead to higher yields in such strains. Murakami *et al.* have detailed a system for the amplification of a biosynthetic pathway leading to higher production of actinorhodin that is dependent on two *Streptomyces kanamyceticus* specific recombination sites (RsA and RsB) and the *zouA* gene, which encodes a TraA-like protein that is homologous to proteins that mediate plasmid conjugation (64, 65).

A *Streptomyces coelicolor* strain was constructed that contains the three elements necessary for amplification (RsA, RsB, *zouA*) flanking the *act* gene cluster along with an apramycin resistance gene (64). The *act* gene cluster encodes the biosynthetic gene cluster for actinorhodins, a polyketide with antibiotic properties. Following the construction of this *S. coelicolor* strain, serial passage in medium containing progressively increasing concentrations of apramycin produced cultures that made large amounts of actinorhodin. PCR analysis to observe tandem amplification of the *act* gene cluster was performed and showed that *zouA*-dependent amplification occurred. This increase in copy number of the amplified region also correlated with increased levels of

apramycin resistance. Actinorhodin production increased more than 20-fold in response to a four- to 12-fold increase in amplification of the 35-kb region containing the *act* gene cluster. The RsA and RsB sites were necessary for specific amplification of the biosynthetic gene cluster (64). Using site-specific recombination methods in Actinomycetes to amplify secondary metabolic biosynthetic gene clusters can lead to increased production of pharmaceutically relevant compounds. These recombination methods can be broadly applied to the production of various antibiotics in a cost-effective manner.

The three studies detailed in the sections above use gene amplification as a tool to improve compound production through microbial biotransformation. Each uses a different method of gene amplification to achieve the goal of increased gene expression: stabilized chromosomal amplification, multiple promoter sequences upstream of the gene of interest, and site-specific recombination to increase gene dosage. While each technique has benefits and drawbacks, the use of gene amplification to engineer strains for optimal chemical production is becoming more prominent.

Using *A. baylyi* ADP1 to study gene amplification events systematically

As indicated above, it is important to study the effects of GDA. A systematic study of GDA events will lead to greater understanding of the impact of chromosomal context on amplification. *A. baylyi* ADP1, with its tractable genetic system and diverse metabolic capacity, is an ideal organism for such a comprehensive investigation. The *A. baylyi* system for studying gene amplification builds on the ability of ADP1 to degrade benzoate as a sole carbon source with enzymes encoded by the *ben* and *cat* genes (Fig.

1.1A) (66, 67). Growth on benzoate typically requires transcription from three promoters that are normally activated by two regulators, BenM and CatM. In *A. baylyi*, GDA mutants are isolated from a strain that cannot grow on benzoate (Ben⁻) because it lacks BenM and CatM and is unable to express sufficiently high levels of the *cat* genes to degrade the benzoate. A promoter mutation in the strain lacking the regulators (* in Fig. 1.1B) enables high-level transcription of the *benABCDE* operon. Derivatives of the Ben⁻ strain that grow on benzoate (Ben⁺) are readily selected. These Ben⁺ mutants have an amplified DNA region (amplicon) that includes all the *cat* genes. The point in the DNA where the recombination occurred to generate the original duplication is termed the duplication junction. The duplication junction defines the endpoints and size of the amplicon. Homologous recombination between amplicons allows rapid expansion to increase amplicon copy number on the chromosome. In the absence of BenM and CatM, the basal level of *cat* gene expression is low, but multiple copies of the genes contained on the amplicon augment transcription, thereby allowing growth on benzoate as the sole carbon source (Fig. 1.1B, C) (33, 36, 68, 69).

This original Ben⁻ strain contains a *benA* promoter mutation that allows high-level expression of the *benABCDE* operon without a transcriptional regulator. The promoter mutation restricts the upstream endpoint of the duplication junction to the region between the promoter mutation in *benA* and the start of the promoter of *catA*. This restriction appears to stem from toxic metabolite imbalance if there are multiple copies of the *benA* promoter mutation as well as multiple copies of the *catA* promoter on the same amplicon. Such amplification, which generates levels of *ben* gene-encoded enzymes that are much

higher than the level of the CatA enzyme, can cause toxic levels of catechol to accumulate in the presence of benzoate (70).

The initial studies of gene amplification focused on the chromosomal region that normally houses the *ben-cat* genes and characterized gene amplification in mutants derived from a strain with the *benA* promoter mutation. When independent amplification events were evaluated by sequence analysis of the duplication junctions, illegitimate recombination appeared to generate most of the initiating duplications (68). In a large percentage of the mutants, amplicons were greater than 250 kb in size, and an inverse relationship was observed between amplicon size and copy number (36, 68). Despite the large size of these amplicons, the *benA* promoter mutation seemingly prevented the study of recombination events upstream of this mutation, as amplification mutants with amplicon endpoints upstream of this mutation were not isolated.

In order to alleviate the restriction on upstream amplicon endpoints, a new parent strain was engineered for selection of additional amplification mutants. This parent strain has a wild-type *benA* promoter region and deletions that prevent the synthesis of BenM and CatM. Without these regulatory proteins, cells cannot grow on benzoate as a sole carbon source without amplification of the *ben* and *cat* genes. To grow on benzoate, amplification mutants derived from the new Ben⁻ strain require a larger amplicon (*benA* through *catD*) than those originally isolated (*catA* through *catD*). This new strain allows isolation of gene amplification mutants with amplicon endpoints upstream of *benA* (36). The analysis of duplication junctions from Ben⁺ amplification mutants derived from the new parent strain showed generally smaller amplicon sizes with higher variability in copy

number (36). As with the previous Ben^r parent strain, there was little to no sequence identity at the point of duplication.

Following the analysis of GDA events surrounding the wild-type *cat* gene locus, we were interested in studying GDA events at other chromosomal loci. The construction of a relocation plasmid to move the *cat* genes to new positions on the chromosome affords the ability to study GDA events at locations beyond the native locus (36). Chromosomal context appears to influence both the size and copy number of the amplicons. When the *cat* genes were relocated to loci within 32 kb of a transposable element, this insertion sequence mediated the duplication events through an atypical transposition mechanism (38). The amplicon sizes determined from mutants isolated from these loci were relatively stable (42-95 kb), but the copy number varied to a greater extent (2-35 copies) (36, 38). However, at a locus near the origin of replication, the amplicon copy number was always less than six, but the size varied significantly (47-471 kb) (36). Moving these genes to novel loci as a selection for GDA events allows an in-depth study of the effects of chromosomal context on amplicon size and copy number as well as the frequency with which duplications occur in the *A. baylyi* ADP1 chromosome.

Final Remarks

Genetic elements can affect genomic plasticity; IS elements lead to inversions, duplications, and deletions. Duplications change gene dosage, and depending on which genes are affected, these duplications and further amplifications can impact antibiotic resistance levels. These increased resistance levels are becoming a significant problem in the treatment of bacterial infections. However, new technologies are making the study of

GDA more accessible and cost effective. Additionally, not all consequences of GDA events are negative; as GDA becomes more thoroughly understood, it can be used to produce compounds with industrial or pharmaceutical relevance. Systematic studies to assess the importance of genomic context are crucial to a deeper understanding of the driving mechanisms behind GDA.

The capability of ADP1 to regulate the degradation of diverse aromatic compounds via gene amplification has been well documented (33, 36, 38, 68, 69). This regulatory mechanism, when combined with the suite of available genetic tools, makes this bacterium a model organism for studying GDA. The single gene deletion collection allows study of the import and effect of each gene (71). Using the single gene deletion collection a global metabolic model of *A. baylyi* ADP1 was constructed (72). Comparison of this model with the *Bacillus subtilis* (73, 74), *P. aeruginosa* PAO1 (75, 76), and *E. coli* (77, 78) collections and metabolic models gives a greater insight into the differences in metabolism between bacteria found in varying environments. Using a combination of these metabolic tools and the genetic malleability of *A. baylyi* ADP1, genomic plasticity can be studied in greater depth.

This dissertation contains an introductory chapter detailing the significance and importance of studying gene amplification in an in-depth and systematic manner. Chapter 2 is comprised of a manuscript published in *Journal of Bacteriology* describing the determination of DNA duplication junctions from gene amplification mutants isolated from two specific loci on the *A. baylyi* chromosome and atypical transposition events that generated the initial duplication. Chapter 3 contains data detailing studies of

chromosomal copy number per cell in *A. baylyi*. The final concluding chapter is a brief summary of the work and conclusions contained herein.

References

1. **Metzgar D, Bacher JM, Pezo V, Reader J, Döring V, Schimmel P, Marlière P, de Crécy-Lagard V.** 2004. *Acinetobacter* sp. ADP1: an ideal model organism for genetic analysis and genome engineering. *Nucl. Acids Res.* **32**:5780-5790.
2. **Young DM, Parke D, Ornston LN.** 2005. Opportunities for genetic investigation afforded by *Acinetobacter baylyi*, a nutritionally versatile bacterial species that is highly competent for natural transformation. *Annu. Rev. Microbiol.* **59**:519-551.
3. **Elliott KT, Neidle EL.** 2011. *Acinetobacter baylyi* ADP1: transforming the choice of model organism. *IUBMB Life* **63**:1075-1080.
4. **Myllykangas S, Knuutila S.** 2006. Manifestation, mechanisms and mysteries of gene amplifications. *Cancer Lett.* **232**:79-89.
5. **Pettersson ME, Andersson DI, Roth JR, Berg OG.** 2005. The amplification model for adaptive mutation: simulations and analysis. *Genetics* **169**:1105-1115.
6. **Romero D, Palacios R.** 1997. Gene amplification and genomic plasticity in prokaryotes. *Annu. Rev. Genet.* **31**:91-111.
7. **Roth JR, Andersson DI.** 2004. Amplification-mutagenesis--how growth under selection contributes to the origin of genetic diversity and explains the phenomenon of adaptive mutation. *Res. Microbiol.* **155**:342-351.
8. **Stark GR, Wahl GM.** 1984. Gene amplification. *Annu. Rev. Biochem.* **53**:447-491.
9. **Sandegren L, Andersson DI.** 2009. Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nat. Rev. Microbiol.* **7**:578-588.

10. **Anderson P, Roth J.** 1981. Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (*rrn*) cistrons. Proc. Natl. Acad. Sci. U. S. A. **78**:3113-3117.
11. **Craven SH, Neidle EL.** 2007. Double trouble: medical implications of genetic duplication and amplification in bacteria. Future Microbiol. **2**:309-321.
12. **Haack KR, Roth JR.** 1995. Recombination between chromosomal IS200 elements supports frequent duplication formation in *Salmonella typhimurium*. Genetics **141**:1245-1252.
13. **Whoriskey SK, Nghiem VH, Leong PM, Masson JM, Miller JH.** 1987. Genetic rearrangements and gene amplification in *Escherichia coli*: DNA sequences at the junctures of amplified gene fusions. Genes Dev. **1**:227-237.
14. **Kugelberg E, Kofoid E, Reams AB, Andersson DI, Roth JR.** 2006. Multiple pathways of selected gene amplification during adaptive mutation. Proc. Natl. Acad. Sci. U. S. A. **103**:17319-17324.
15. **Ohno S.** 1970. Evolution by gene duplication. Springer-Verlag, Berlin, New York.
16. **Bergthorsson U, Andersson DI, Roth JR.** 2007. Ohno's dilemma: evolution of new genes under continuous selection. Proc. Natl. Acad. Sci. U. S. A. **104**:17004-17009.
17. **Craig NL.** 2002. Mobile DNA II. ASM Press, Washington, D.C.
18. **Coyne S, Courvalin P, Galimand M.** 2010. Acquisition of multidrug resistance transposon Tn6061 and IS6100-mediated large chromosomal inversions in *Pseudomonas aeruginosa* clinical isolates. Microbiology **156**:1448-1458.

19. **Ling A, Cordaux R.** 2010. Insertion sequence inversions mediated by ectopic recombination between terminal inverted repeats. *PLoS One* **5**:e15654.
20. **Kusumoto M, Ooka T, Nishiya Y, Ogura Y, Saito T, Sekine Y, Iwata T, Akiba M, Hayashi T.** 2011. Insertion sequence-excision enhancer removes transposable elements from bacterial genomes and induces various genomic deletions. *Nat. Commun.* **2**:152.
21. **Karah N, Sundsfjord A, Towner K, Samuelsen O.** 2012. Insights into the global molecular epidemiology of carbapenem non-susceptible clones of *Acinetobacter baumannii*. *Drug Resist. Updat.* **15**:237-247.
22. **Howard A, O'Donoghue M, Feeney A, Sleator RD.** 2012. *Acinetobacter baumannii*: an emerging opportunistic pathogen. *Virulence* **3**:243-250.
23. **Ravasi P, Limansky AS, Rodriguez RE, Viale AM, Mussi MA.** 2011. IS*Aba825*, a functional insertion sequence modulating genomic plasticity and *bla*_{OXA-58} expression in *Acinetobacter baumannii*. *Antimicrob. Agents. Chemother.* **55**:917-920.
24. **Lee Y, Kim CK, Lee H, Jeong SH, Yong D, Lee K.** 2011. A novel insertion sequence, IS*Aba10*, inserted into IS*Aba1* adjacent to the *bla*_{OXA-23} gene and disrupting the outer membrane protein gene *carO* in *Acinetobacter baumannii*. *Antimicrob. Agents. Chemother.* **55**:361-363.
25. **Evans BA, Hamouda A, Towner KJ, Amyes SG.** 2010. Novel genetic context of multiple *bla*_{OXA-58} genes in *Acinetobacter* genospecies 3. *J. Antimicrob. Chemother.* **65**:1586-1588.

26. **Pfeifer Y, Wilharm G, Zander E, Wichelhaus TA, Gottig S, Hunfeld KP, Seifert H, Witte W, Higgins PG.** 2011. Molecular characterization of *bla*_{NDM-1} in an *Acinetobacter baumannii* strain isolated in Germany in 2007. *J. Antimicrob. Chemother.* **66**:1998-2001.
27. **Corn PG, Anders J, Takala AK, Kayhty H, Hoiseth SK.** 1993. Genes involved in *Haemophilus influenzae* type b capsule expression are frequently amplified. *J. Infect. Dis.* **167**:356-364.
28. **Mekalanos JJ.** 1983. Duplication and amplification of toxin genes in *Vibrio cholerae*. *Cell* **35**:253-263.
29. **Lawrence JG.** 2003. Gene organization: selection, selfishness, and serendipity. *Annu. Rev. Microbiol.* **57**:419-440.
30. **Reams AB, Neidle EL.** 2004. Selection for gene clustering by tandem duplication. *Annu. Rev. Microbiol.* **58**:119-142.
31. **Hastings PJ, Slack A, Petrosino JF, Rosenberg SM.** 2004. Adaptive amplification and point mutation are independent mechanisms: evidence for various stress-inducible mutation mechanisms. *PLoS Biol.* **2**:e399.
32. **Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV.** 2002. Selection in the evolution of gene duplications. *Genome Biol.* **3**:8.
33. **Reams AB, Neidle EL.** 2003. Genome plasticity in *Acinetobacter*: new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. *Mol. Microbiol.* **47**:1291-1304.
34. **Anderson RP, Roth JR.** 1977. Tandem genetic duplications in phage and bacteria. *Annu. Rev. Microbiol.* **31**:473-505.

35. **Andersson DI, Hughes D.** 2009. Gene amplification and adaptive evolution in bacteria. *Annu. Rev. Genet.* **43**:167-195.
36. **Seaton SC, Elliott KT, Cuff LE, Laniohan NS, Patel PR, Neidle EL.** 2012. Genome-wide selection for increased copy number in *Acinetobacter baylyi* ADP1: locus and context-dependent variation in gene amplification. *Mol. Microbiol.* **83**:520-535.
37. **Rocha EP, Cornet E, Michel B.** 2005. Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS Genet.* **1**:e15.
38. **Cuff LE, Elliott KT, Seaton SC, Ishaq MK, Laniohan NS, Karls AC, Neidle EL.** 2012. Analysis of IS1236-Mediated Gene Amplification Events in *Acinetobacter baylyi* ADP1. *J. Bacteriol.* **194**:4395-4405.
39. **Brochet M, Couve E, Zouine M, Poyart C, Glaser P.** 2008. A naturally occurring gene amplification leading to sulfonamide and trimethoprim resistance in *Streptococcus agalactiae*. *J. Bacteriol.* **190**:672-680.
40. **Ikeda H, Shiraishi K, Ogata Y.** 2004. Illegitimate recombination mediated by double-strand break and end-joining in *Escherichia coli*. *Adv. Biophys.* **38**:3-20.
41. **Lovett ST, Drapkin PT, Sutura VA, Jr., Gluckman-Peskind TJ.** 1993. A sister-strand exchange mechanism for *recA*-independent deletion of repeated DNA sequences in *Escherichia coli*. *Genetics* **135**:631-642.
42. **Trinh TQ, Sinden RR.** 1993. The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in *Escherichia coli*. *Genetics* **134**:409-422.

43. **Reams AB, Kofoid E, Kugelberg E, Roth JR.** 2012. Multiple pathways of duplication formation with and without recombination (RecA) in *Salmonella enterica*. *Genetics* **192**:397-415.
44. **Reams AB, Kofoid E, Savageau M, Roth JR.** 2010. Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. *Genetics* **184**:1077-1094.
45. **Sun S, Ke R, Hughes D, Nilsson M, Andersson DI.** 2012. Genome-wide detection of spontaneous chromosomal rearrangements in bacteria. *PLoS One* **7**:e42639.
46. **Liu Z, Obenauf AC, Speicher MR, Kopan R.** 2009. Rapid identification of homologous recombinants and determination of gene copy number with reference/query pyrosequencing (RQPS). *Genome Res.* **19**:2081-2089.
47. **Simon C, Daniel R.** 2011. Metagenomic analyses: past and future trends. *Appl. Environ. Microbiol.* **77**:1153-1161.
48. **Simon C, Daniel R.** 2009. Achievements and new knowledge unraveled by metagenomic approaches. *Appl. Microbiol. Biotechnol.* **85**:265-276.
49. **Dunne WM, Westblade LF, Ford B.** 2012. Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory. *Eur. J. Clin. Microbiol. Infect. Dis.* **31**:1719-1726.
50. **Lee C, Lee S, Shin SG, Hwang S.** 2008. Real-time PCR determination of rRNA gene copy number: absolute and relative quantification assays with *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **78**:371-376.

51. **Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y.** 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453-1462.
52. **Bustin SA.** 2010. Why the need for qPCR publication guidelines?--The case for MIQE. *Methods* **50**:217-226.
53. **Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT.** 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**:611-622.
54. **D'haene B, Vandesompele J, Hellemans J.** 2010. Accurate and objective copy number profiling using real-time quantitative PCR. *Methods* **50**:262-270.
55. **Spurgeon SL, Jones RC, Ramakrishnan R.** 2008. High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. *PLoS One* **3**:e1662.
56. **Weaver S, Dube S, Mir A, Qin J, Sun G, Ramakrishnan R, Jones RC, Livak KJ.** 2010. Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution. *Methods* **50**:271-276.
57. **Qin J, Jones RC, Ramakrishnan R.** 2008. Studying copy number variations using a nanofluidic platform. *Nucl. Acids Res.* **36**:e116.
58. **Whale AS, Huggett JF, Cowen S, Speirs V, Shaw J, Ellison S, Foy CA, Scott DJ.** 2012. Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. *Nucl. Acids Res.* **40**:e82.

59. **White RA, Blainey PC, Fan HC, Quake SR.** 2009. Digital PCR provides sensitive and absolute calibration for high throughput sequencing. *BMC Genomics* **10**:116.
60. **Hoshino T, Inagaki F.** 2012. Molecular quantification of environmental DNA using microfluidics and digital PCR. *Syst. Appl. Microbiol.* **35**:390-395.
61. **Xie C, Tammi MT.** 2009. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* **10**:80.
62. **Tyo KEJ, Ajikumar PK, Stephanopoulos G.** 2009. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat. Biotechnol.* **27**:760-765.
63. **Li M, Wang J, Geng Y, Li Y, Wang Q, Liang Q, Qi Q.** 2012. A strategy of gene overexpression based on tandem repetitive promoters in *Escherichia coli*. *Microb. Cell Fact.* **11**:19.
64. **Murakami T, Burian J, Yanai K, Bibb MJ, Thompson CJ.** 2011. A system for the targeted amplification of bacterial gene clusters multiplies antibiotic yield in *Streptomyces coelicolor*. *Proc. Natl. Acad. Sci. U. S. A.* **108**:16020-16025.
65. **Murakami T, Sumida N, Bibb M, Yanai K.** 2011. *ZouA*, a putative relaxase, is essential for DNA amplification in *Streptomyces kanamyceticus*. *J. Bacteriol.* **193**:1815-1822.
66. **Bundy BM, Collier LS, Hoover TR, Neidle EL.** 2002. Synergistic transcriptional activation by one regulatory protein in response to two metabolites. *Proc. Natl. Acad. Sci. U. S. A.* **99**:7693-7698.

67. **Ezezika OC, Collier-Hyams LS, Dale HA, Burk AC, Neidle EL.** 2006. CatM regulation of the *benABCDE* operon: functional divergence of two LysR-type paralogs in *Acinetobacter baylyi* ADP1. *Appl. Environ. Microbiol.* **72**:1749-1758.
68. **Reams A, Neidle E.** 2004. Gene Amplification Involves Site-specific Short Homology-independent Illegitimate Recombination in *Acinetobacter* sp. Strain ADP1. *J. Mol. Biol.* **338**:643-656.
69. **Craven SH.** 2009. Gene Expression and Dosage: Distinct Mechanisms Regulate Benzoate Degradation in *Acinetobacter baylyi* ADP1. Ph.D. University of Georgia, Athens, GA.
70. **Collier LS, Gaines GL, 3rd, Neidle EL.** 1998. Regulation of benzoate degradation in *Acinetobacter* sp. strain ADP1 by BenM, a LysR-type transcriptional activator. *J. Bacteriol.* **180**:2493-2501.
71. **de Berardinis V, Vallenet D, Castelli V, Besnard M, Pinet A, Cruaud C, Samair S, Lechaplais C, Gyapay G, Richez C, Durot M, Kreimeyer A, Le Fevre F, Schachter V, Pezo V, Doring V, Scarpelli C, Medigue C, Cohen GN, Marliere P, Salanoubat M, Weissenbach J.** 2008. A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* **4**:174.
72. **Durot M, Le Fevre F, de Berardinis V, Kreimeyer A, Vallenet D, Combe C, Smidtas S, Salanoubat M, Weissenbach J, Schachter V.** 2008. Iterative reconstruction of a global metabolic model of *Acinetobacter baylyi* ADP1 using high-throughput growth phenotype and gene essentiality data. *BMC Syst. Biol.* **2**:85.

73. **Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K, Chapuis J, Christiansen LC, Danchin A, Debarbouille M, Dervyn E, Deuerling E, Devine K, Devine SK, Dreesen O, Errington J, Fillinger S, Foster SJ, Fujita Y, Galizzi A, Gardan R, Eschevins C, Fukushima T, Haga K, Harwood CR, Hecker M, Hosoya D, Hullo MF, Kakeshita H, Karamata D, Kasahara Y, Kawamura F, Koga K, Koski P, Kuwana R, Imamura D, Ishimaru M, Ishikawa S, Ishio I, Le Coq D, Masson A, Mauel C, Meima R, Mellado RP, Moir A, Moriya S, Nagakawa E, Nanamiya H, Nakai S, Nygaard P, Ogura M, Ohanan T, O'Reilly M, O'Rourke M, Pragai Z, Pooley HM, Rapoport G, Rawlins JP, Rivas LA, Rivolta C, Sadaie A, Sadaie Y, Sarvas M, Sato T, Saxild HH, Scanlan E, Schumann W, Seegers JF, Sekiguchi J, Sekowska A, Seror SJ, Simon M, Stragier P, Studer R, Takamatsu H, Tanaka T, Takeuchi M, Thomaidis HB, Vagner V, van Dijl JM, Watabe K, Wipat A, Yamamoto H, Yamamoto M, Yamamoto Y, Yamane K, Yata K, Yoshida K, Yoshikawa H, Zuber U, Ogasawara N. 2003. Essential *Bacillus subtilis* genes. Proc. Natl. Acad. Sci. U. S. A. **100**:4678-4683.**
74. **Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R. 2007. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. J. Biol. Chem. **282**:28791-28799.**
75. **Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, Will O, Kaul R, Raymond C, Levy R, Chun-Rong L, Guenther D, Bovee D,**

- Olson MV, Manoil C.** 2003. Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. U. S. A. **100**:14339-14344.
76. **Oberhardt MA, Puchalka J, Fryer KE, Martins dos Santos VA, Papin JA.** 2008. Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. J. Bacteriol. **190**:2790-2803.
77. **Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H.** 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. **2**:2006 0008.
78. **Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO.** 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. **3**:121.

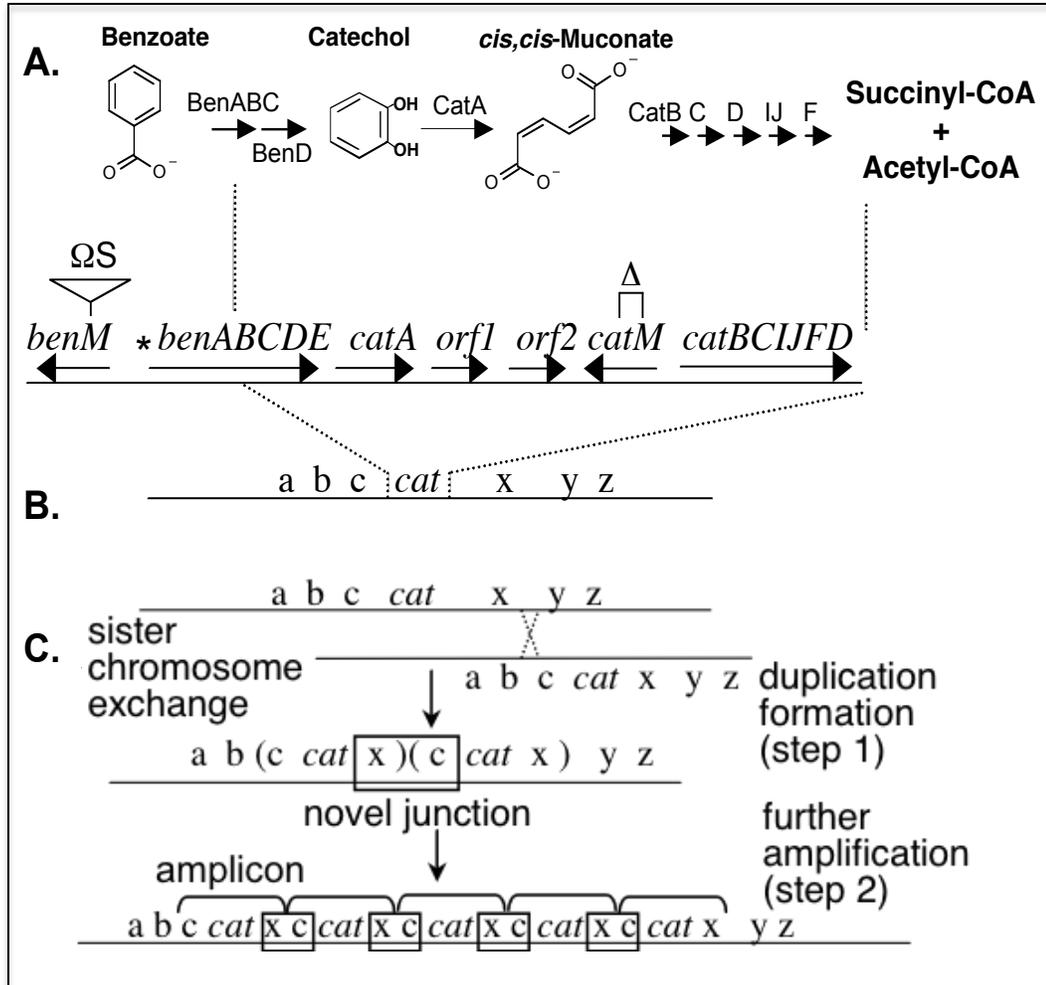


Figure 1.1. Experimental system for the study of GDA in *Acinetobacter baylyi* ADP1. A. Degradation pathway for benzoate. B. Chromosome of a mutant that does not grow on benzoate due to inactivation of *benM* (ΩS), and deletion of *catM* (Δ). * indicates *benA* promoter mutation conferring high-level transcription in the absence of a transcriptional regulator. Abbreviations: *cat* indicates the *catA* gene and the *catBCIJFD* operon; abc and xyz are genetic regions normally upstream and downstream, respectively, of the *cat* genes. C. Two-step model for amplification. Duplication formation does not require RecA (step 1). Further amplification depends on RecA (step 2).

CHAPTER 2
ANALYSIS OF IS1236-MEDIATED GENE AMPLIFICATION EVENTS IN
ACINETOBACTER BAYLYI ADP1¹

¹ Cuff, L.E., Elliott, K.T., Seaton, S.C., Ishaq, M.K., Laniohan, N.S., Karls, A.C., and Neidle, E.L. 2012. *Journal of Bacteriology* 194:4395-4405. Reprinted here with permission of publisher.

L.E.C., K.T.E., A.C.K., and E.L.N designed experiments, L.E.C., K.T.E., S.C.S., M.K.I., and N.S.L. performed experiments, L.E.C., K.T.E., and E.L.N. authored manuscript.

Abstract

Recombination between insertion sequence copies can cause genetic deletion, inversion, or duplication. However, it is difficult to assess the fraction of all genomic rearrangements that involve insertion sequences. In previous gene duplication and amplification studies in *Acinetobacter baylyi* ADP1, an insertion sequence was evident in approximately 2% of the characterized duplication sites. Gene amplification occurs frequently in all organisms and has a significant impact on evolution, adaptation, drug resistance, cancer, and varied disorders. To understand the molecular details of this important process, a previously developed system was used to analyze gene amplification in selected mutants. The current study focused on amplification events in two chromosomal regions that are near one of six copies of the only transposable element in ADP1, *IS1236* (an *IS3* family member). Twenty-one independent mutants were analyzed, and in contrast to previous studies of a different chromosomal region, *IS1236* was involved in 86% of these events. *IS1236*-mediated amplification could occur through homologous recombination between insertion sequences on both sides of a duplicated region. However, this mechanism presupposes that transposition generates an appropriately positioned additional copy of *IS1236*. To evaluate this possibility, PCR and Southern hybridization were used to determine the chromosomal configurations of amplification mutants involving *IS1236*. Surprisingly, the genomic patterns were inconsistent with the hypothesis that intramolecular homologous recombination occurred between insertion sequences following an initial transposition event. These results raise a novel possibility that the gene amplification events near the *IS1236* elements arise from illegitimate recombination involving transposase-mediated DNA cleavage.

Introduction

Acinetobacter baylyi ADP1 serves as a model bacterium for investigating gene duplication and amplification (GDA) (40). GDA plays a critical role in adaptation and evolution, and it also underlies a range of health issues such as drug resistance, cancer, and various human genetic disorders (1, 3, 4, 9, 13, 43). One mechanism for generating duplications involves homologous recombination between stretches of directly repeated DNA such as closely spaced rRNA operons or different copies of insertion sequences (IS) (2, 10, 23, 27, 30). Therefore, it was surprising that few examples of IS-mediated rearrangement were identified during previous studies of chromosomal GDA in *A. baylyi* (34, 35, 40).

Gene amplification mutants are readily selected from an *A. baylyi* parent strain that is unable to consume benzoate (Ben⁻) (12, 16, 34, 40). This parent lacks two regulators that normally activate transcription from multiple promoters in a clustered group of genes for benzoate catabolism. Spontaneous Ben⁺ derivatives that acquire the ability to grow on benzoate carry tandem head-to-tail arrays of a chromosomal segment (amplicon) encompassing genes needed to degrade benzoate (*cat* genes, Fig. 2.1A) (16, 34, 40). These arrays increase gene expression by providing multiple copies of genes with weak promoters and thereby compensate for the absence of the transcriptional regulators (34, 35, 40). The selection demands increased expression of genes controlled by at least two different promoters, and it yields gene amplification mutants almost exclusively. Furthermore, continued selection maintains the amplified DNA and permits the analysis of amplicon size and copy number. The precise site of duplication, termed

the duplication junction, can be identified using an assay that exploits the high efficiency of natural transformation in *A. baylyi* (Fig. 2.1B) (34).

The *A. baylyi* system was expanded for genome-wide studies using the 10 kbp *cat* gene region, which encompasses *catA* and the *catBCIJFD* operon, as a selectable cassette for relocation into any nonessential chromosomal locus (14, 40). The cassette is used to generate diverse Ben⁻ parent strains from which Ben⁺ amplification mutants are selected and analyzed. The frequency of spontaneous duplications that encompass *catA* can also be assessed regardless of the chromosomal location of this gene (40). The duplication frequency assay detects a small fraction of cells that retain a functional *catA* after one copy is inactivated by allelic replacement. When grown under conditions that do not select for duplications, 10⁻⁴ cells carry a duplication of *catA* in its native locus (40).

In *A. baylyi* ADP1, there is a single type of IS element, IS1236 (5, 18, 19, 25). In the chromosome, there are six copies of IS1236, five of which are identical. In one location (near locus *i* in Fig. 2.2), two adjacent copies (IS1236_2 and IS1236_3) can function as a composite transposon (41). IS1236 belongs to the large IS3 family, members of which transpose using a replicative excision and conservative integration mechanism, also known as the copy-and-paste method of transposition (15, 18). As a result, the element is retained in its original locus as well as being inserted at a new location (8, 15, 30, 37, 38). Transposition can insert IS1236 elements at novel chromosomal positions that might enable gene duplication via homologous recombination, as depicted in Fig. 2.3A. This type of scenario was inferred to account for the involvement of IS1236 in two of 91 GDA events in our initial studies of gene

amplification in the vicinity of the native *cat*-gene locus (34, 40). This locus is 190 kbp away from the nearest IS element (Fig. 2.2).

In our current study, we investigated whether proximity of the *cat*-gene cassette to an IS1236 element influences the percentage of selected GDA events that are IS-mediated. We analyzed gene amplification mutants derived from parent strains in which the relocated *catA* gene was within 24 or 32 kbp of the nearest IS1236 element (locus *i* or *ii*, respectively, in Fig. 2.2). As described here, duplication junction sequences were determined in order to identify IS involvement. When IS-mediated rearrangements were detected, the genomic configurations of the GDA mutants were further characterized.

Materials and methods

Strains and Growth Conditions

A. baylyi strains derived from the wild type, ADP1, were given ACN designations (Table 2.1). *E. coli* strains, DH5 α and Top10 (Invitrogen), were used as plasmid hosts and were grown at 37 °C in Luria-Bertani (LB) medium, also known as lysogeny broth [10 g of Bacto-tryptone, 5 g of yeast extract, and 10 g of NaCl per liter (39)]. *A. baylyi* strains were grown at 30 or 37 °C in LB or in minimal salts medium (42) with 10 mM succinate or 1 mM benzoate as the sole carbon source. Antibiotics were used as necessary at the following final concentrations: kanamycin, 25 μ g/mL; spectinomycin and streptomycin, 12.5 μ g/mL; ampicillin 150 μ g/mL.

DNA purification, plasmid construction, and the generation of plasmid libraries

Standard molecular techniques were used for restriction digests, ligations, *E. coli* transformations, and plasmid isolation (39). Genomic DNA was isolated from *A. baylyi*

strains with the Easy-DNA kit (Invitrogen). Based on previously described methods (35), libraries of DNA from amplification mutants were generated by ligation of EcoRI-cleaved genomic DNA from ACN1040, ACN1101, ACN1102, or ACN1127 into the plasmid vector pUC19. For all other strains, including the newly isolated mutants ACN1222 and ACN1223, a different plasmid vector was used, pZErO-2 (Invitrogen). Libraries were maintained in *E. coli* DH5 α for pUC19 constructs and *E. coli* Top10 (Invitrogen) for pZErO-2 constructs. Plasmids were screened for their ability to transform a Ben⁻ recipient to allow growth on benzoate as a sole carbon source (35). Plasmids that conferred the ability to consume benzoate were sequenced to determine the duplication junction site. DNA sequencing was performed at Genewiz, Inc.

Selection of Ben⁺ GDA mutants, chromosomal analysis by Pulsed Field Gel Electrophoresis (PFGE), and quantitative PCR (qPCR)

The Ben⁻ parent strain ACN1039 was grown on succinate medium, washed, concentrated, and plated on solid medium with benzoate as the sole carbon source, as described previously (35, 40). Two individual colonies that arose independently from different starting cultures after 7 and 10 days of incubation were streak purified, and respectively designated ACN1222 and ACN1223. They were further characterized by PFGE analysis (CHEF-DR II, Bio-Rad). Genomic DNA was prepared in agarose plugs, and the embedded DNA was digested with the restriction enzyme NotI as previously described (20, 35, 40). PFGE was used to separate the NotI-cleaved chromosomal fragments using a 24 hour protocol with a constant 6 V cm⁻² current and pulse switch times that increased from 1-150 seconds (ACN1039-derived mutants) or 10-200 seconds (ACN1050-derived mutants).

To evaluate the amplicon copy number in ACN1222 and ACN1223, the amount of *catA* DNA was compared to that of a control gene, *antA*, assumed to be in the genome in single copy. A real-time qPCR assay was used with TaqMan 5'-exonuclease (Applied Biosystems). The primers and method were previously described (40).

Analysis of the chromosomal configurations of GDA mutants by PCR and Southern hybridization

Strain specific oligonucleotides (Table 2.S1) were used for PCR analysis to determine the genomic rearrangements present in individual GDA mutants. LongAmp polymerase (New England Biolabs) and extension times of 2 min 30 sec were used to ensure the amplification of long products. Concentrations of reagents in the 25 μ L reaction were as follows: template DNA, \sim 10 ng; dNTPs, 300 μ M; oligonucleotides, 0.4 μ M each; 1X LongAmp Reaction Buffer; 2.5 units LongAmp Polymerase. PCRs were run on agarose gels at the appropriate concentrations for differentiating products of expected sizes (39) (Table 2.2).

Southern hybridizations were performed on select amplification mutants. A digoxigenin-labeled (DIG) DNA probe was generated by a random-primed labeling method following the Roche Molecular Biochemicals DIG Application Manual using the primers listed in Table 2.S1. Briefly, 10 ng ADP1 genomic DNA, 200 μ M DIG-dNTPs, 0.5 μ M each primer, and 2.6 units of Expand High Fidelity polymerase were used in a 30-cycle PCR. The labeled probe was stored at -20 $^{\circ}$ C in hybridization buffer (5X SSC [0.75 M NaCl, 0.075 M Na-citrate, pH 7.0], 50% formamide, 0.02% SDS, 0.1% N-laroyl sarcosine, 2% Bio-Rad blocking reagent). PCR products were generated as a size ladder for visualization of band sizes following hybridization with the DIG-labeled probe using

the primers in Table 2.S1. DNA from the amplification mutants (2 µg) and 4 µg of DNA from wild-type ADP1 and the parent strain were digested with PstI (ACN1050, ACN1056, ACN1057, ACN1058) or EcoRV (ACN1039, ACN1101). The digested DNA was separated on a standard (1.0 %) agarose gel and transferred to a nitrocellulose membrane using a TurboBlotter system (Schleicher and Schuell). Following transfer, DNA was cross-linked to the membrane by a total dose of UV light (254 nm) of 120 m J cm⁻². Prehybridization, hybridization for 16 hours with the DIG-labeled probe at 42 °C, and low and high stringency washes were carried out according to manufacturer's instructions (Roche). Following the washes, incubation of the membrane with an anti-DIG antibody (Roche) immediately followed by incubation with a chemiluminescent substrate of alkaline phosphatase (Roche) was used to visualize the band sizes on X-ray film according to manufacturer instructions and as previously described (20, 35).

Results and discussion

Selection and characterization of GDA mutants

In a previous study we described the construction of two Ben⁻ strains, ACN1039 and ACN1050, in which the *cat*-gene cassette was deleted from its native locus and relocated to the chromosomal positions indicated in Fig. 2.2 as locus *i* and locus *ii*, respectively (40). Here we chose 19 of the Ben⁺ amplification mutants that were derived from ACN1039 or ACN1050 for further analysis. In addition, we selected two new Ben⁺ mutants from ACN1039, designated ACN1222 and ACN1223 (Table 2.1). These mutants arose similarly to those that were previously isolated, each appearing spontaneously as a

colony after ACN1039 cells were spread on solid medium with benzoate as the sole carbon source.

Characterization of NotI-digested ACN1223 DNA by PFGE revealed one fragment in the mutant that is not in its parent (marked by an asterisk in Fig. 2.4A). When the wild-type chromosome is digested with NotI, its 3.6 Mb circular chromosome is cut into 6 fragments. As depicted in Fig. 2.4B, a NotI site in the *cat*-gene region was introduced when the cassette was placed in locus *i* of ACN1039. Introduction of this site results in cleavage of the wild-type fragment into two similarly sized fragments that are not resolved with PFGE (Fig. 2.4B). In GDA mutants, multiple copies of the *cat*-gene region result in a novel NotI-generated fragment whose size equals that of its amplicon (Fig. 2.4B). Thus, the banding pattern of ACN1223 indicates the presence in this strain of an amplicon of ~41 kbp. Similar analysis of ACN1222 indicated the presence of a ~44 kbp amplicon (data not shown).

ACN1039-derived mutants were found to have a relatively narrow range of amplicon size (39 to 61 kbp) with copy number of the amplicon ranging from 9 to 35 relative to the rest of the genome ((40) and Table 2.S2). Therefore, the amplicon sizes of ACN1222 and ACN1223 are consistent with previous observations. When amplicon copy number was evaluated (Table 2.S2), that of ACN1222 was determined to be 23, within the range previously observed for ACN1039-derived mutants. While ACN1223 was found to have only 2 copies of its amplicon, *cat* gene duplications in other genomic regions have previously been shown to be sufficient for slow Ben⁺ growth (40). The ACN1050-derived GDA mutants have amplicon sizes spanning from 49 to 95 kbp with copy numbers that range from 7 to 26 ((40) and Table 2.S2).

Determination of DNA Junction Sequences

The 21 GDA mutants chosen for this study resulted from amplification events in which the *cat* genes were closer to an IS element than they are in their native locus. The *cat* gene cassette in ACN1039 (locus *i*) is 24 kbp from IS1236_3, and the cassette in ACN1050 (locus *ii*) is 32 kbp from IS1236_6 (Fig. 2.2). We sought to evaluate whether proximity to a transposable element affects the likelihood of selecting mutants with IS-mediated rearrangements. Since DNA from the precise duplication site provides critical information about the underlying recombination event, such duplication junction sequences were analyzed in the ACN1039- and ACN1050-derived GDA mutants.

To identify the duplication junction sequences, we used a previously developed assay that exploits the high efficiency of natural transformation and homologous recombination in *A. baylyi* (35). The first step in this method involves construction of a recombinant plasmid library in *E. coli* carrying random chromosomal DNA fragments from an amplification mutant. The presence of a plasmid-borne duplication junction sequence can be identified by its ability to enable a Ben⁻ parent strain to grow on benzoate after transformation and homologous recombination. Individual plasmid-containing *E. coli* colonies are patched onto a lawn of the Ben⁻ parent strain on succinate medium, on which the *E. coli* strain cannot grow. Plasmid DNA that is released by the lysis of *E. coli* is taken up by the neighboring Ben⁻ parent strain. If the plasmid DNA includes a duplication junction, homologous recombination between this DNA and the chromosome can generate a duplication identical to the one found in the original amplification mutant from which the plasmid library was constructed (Fig. 2.1B). When transferred to benzoate medium, Ben⁺ transformants arise from cells in which the

chromosomal duplication has undergone further amplification. Plasmids are then isolated from the *E. coli* donors that are able to generate the Ben⁺ phenotype in *A. baylyi*, and the duplication junction is identified by DNA sequence analysis.

In this fashion, the duplication junctions of 21 amplification mutants were isolated and sequenced (Fig. 2.5). They reveal the precise point of recombination between DNA that had been downstream of the *cat*-gene cassette and DNA that was previously upstream of it. Inverted black triangles mark these sites in Fig. 2.1A and Fig. 2.3A. As detailed below, the majority of these duplication junctions involved the same sequence upstream of the *cat* genes. The DNA sequence of the duplication junction defines the endpoints of an amplicon and therefore also the amplicon size. In all cases, the sizes inferred from junction sequence analyses (Fig. 2.5) are in good agreement with those determined by PFGE (Fig. 2.4A and data not shown).

Predominance of IS-mediated rearrangement leading to amplification of loci i and ii

Sequence analysis showed that 11 out of 11 duplication events from locus *i* and 7 out of 10 duplication events at locus *ii* involved a copy of IS1236 at the duplication junction (Fig. 2.5). The duplication junction in each of these 18 independent GDA mutants occurred exactly at the end of the IS1236 terminal inverted repeat. This observation could be readily explained if homologous recombination between two IS elements occurred as shown in Fig. 2.3A. In this case, the sequence identified as the duplication junction represents the target site into which the IS element had initially inserted (the leftward pointing triangle in Fig. 2.3A). According to this scenario, the precise point of recombination (the duplication junction) really occurs somewhere in the 1.2 kbp region of sequence identity between two IS elements.

IS1236 sequences were evident in 86% of the duplication sites. These results differed from previous investigations where only a small fraction (2%) of 91 characterized duplication junctions involved IS1236 (34, 40). The previous studies identified junction sequences resulting from duplication events in a different chromosomal region, the native *cat*-gene locus, which suggests that chromosomal position influences the type of gene amplification mutants that are selected. Chromosomal position was also concluded to be a factor in the correlation between size and copy number of amplicons in different GDA mutants (40).

The effect of genomic location was also evaluated for spontaneous *catA* duplication frequency. This frequency was previously assessed in five chromosomal regions, the wild-type locus and the positions to which *catA* had been relocated, loci *i-iv* in Fig. 2.2. Interestingly, these values were similar regardless of genomic position (40). In the absence of selection, the duplication frequency of *catA* ranged from 10^{-4} to 10^{-5} . The value for locus *i* was the same as for the wild-type locus (10^{-4}), whereas that for locus *ii* was 10-fold lower. Therefore, the increase in the proportion of duplication sites involving IS1236 cannot be explained by an overall increase in the frequency of spontaneous duplications in locus *i* or *ii*. The relative distance of the *cat* genes to an IS1236 element is one difference between the wild-type locus and the two regions examined in our current study. However, further work is needed to establish if, and how, such proximity affects mechanisms underlying GDA.

IS1236 was first discovered in studies of mutations that prevent the degradation of *p*-hydroxybenzoate (18, 19). The spontaneous insertion of IS1236 plays a significant role in the inactivation of genes for aromatic compound catabolism in two different

chromosomal regions of ADP1 (18, 19, 26, 41). One of these regions corresponds to locus *i*, where the *cat*-gene cassette is inserted in *vanK*. This chromosomal area, which encompasses several *van* genes involved in the metabolism of ferulate and vanillate, undergoes frequent rearrangements including large spontaneous deletions and the transposition of a composite element that combines IS1236_2 and IS1236_3 (20, 41). Our analysis indicates that IS1236_3 alone, not the composite IS element, was located at the duplication junctions of GDA mutants involving locus *i* in the current study (data not shown).

The second region that is characterized by numerous IS1236 insertions resides approximately 300 kbp away, in a clockwise direction, from the native *cat* genes (*wt* locus, Fig. 2.2). This region makes up an approximately 20 kbp supraoperonic gene cluster involved in aromatic compound catabolism. IS1236 transposition was shown to occur preferentially in certain genes within the cluster, although the basis of such preference remains unclear (18, 26).

GDA events near locus ii, ACN1050-derived mutants

While all of the duplication junctions from locus *i* involved IS1236, three from locus *ii* did not (Fig. 2.5). These junctions were isolated from the GDA mutants ACN1162, ACN1164, and ACN1165. There were no obvious features to distinguish these amplicons from any others in terms of size or copy number (Table 2.S2). We examined DNA at the point of recombination in these mutants to identify short stretches of nucleotides that are identical between the sequences that recombine to form duplication junctions (bold nucleotides in Fig. 2.5). The lack of significant identity between the sequences generating the duplication junctions indicates that these duplications were formed by illegitimate

recombination. At the precise point of recombination, there was only one nucleotide of identity between the sequences comprising the duplication junction of two mutants (ACN1164 and ACN1165), and there was no identity at the duplication junction in the other (ACN1162). This type of pattern is characteristic of most of the duplication junctions isolated in previous studies of GDA at the native *cat*-gene locus (34, 40).

The seven remaining ACN1050-derived GDA mutants displayed IS involvement. Such involvement is intriguing in this chromosomal region because the nearest IS element to locus *ii* is *IS1236_6* (Fig. 2.2). This *IS1236* copy shares only 82% sequence identity with the other copies, which are all identical to one another. Moreover, the sequence variation suggests that *IS1236_6* does not encode a functional transposase (Fig. 2.S1). Specifically, the canonical catalytic triad (DDE motif) is disrupted by a frameshift that occurs prior to the codon for the glutamate residue (8, 31, 37).

Potential role for transposition in gene amplification

Amplification mutants could be generated by transposition of an *IS1236* element followed by homologous recombination between the two IS copies, as shown in Fig. 2.3A. The duplication junction should identify the precise target site into which *IS1236* transposed. This target site, marked by a diamond in Fig. 2.3A, is expected to be different in each mutant. Insertion of a copy of *IS1236* into this site would create two characteristic joints where wild-type sequence is interrupted by the IS element (represented in Fig. 2.3A by two black triangles of a split diamond). Following duplication, homologous recombination could result in further amplification. An amplification mutant would be expected to carry multiple copies of the duplication junction, as well as a single copy of one joint that marks the downstream portion of the transposition target site (labeled as the

transposition join point, Fig. 2.3A). This single join point highlights that the process of transposition followed by homologous recombination results in a hallmark copy of the insertion sequence downstream of the final copy of the amplicon. As described below, we failed to identify this “terminal IS element,” a result that led us to consider different possible duplication mechanisms. These possibilities are illustrated in Fig. 2.3B and C and are discussed later.

PCR analysis of chromosomal configurations in GDA mutants

We used PCR methods to detect the presence or absence of an IS element downstream of the *cat* genes in each of the ACN1039- and ACN1050-derived GDA mutants. Primers were designed to amplify the target DNA where *IS1236* might have transposed. The upstream portion of the target site (the leftward pointing triangle of the split diamond, Fig. 2.3A) was identified experimentally by the junction sequences (Fig. 2.5). PCR primer pairs were designed to anneal within and immediately downstream of the final amplicon, flanking the position where a terminal insertion sequence would be located following transposition of *IS1236* (Table 2.S1). These primers, when used with the chromosomal DNA of an amplification mutant as template, should distinguish between the presence of an IS element inserted in the target DNA at the transposition join point (as shown in Fig. 2.3A) and the wild-type configuration in this region (as shown in Fig. 2.3D).

The expected differences between PCR products from different chromosomal templates are shown in Fig. 2.6A-C and Table 2.2. Fragment size predictions were calculated for each of the 18 amplification mutants in which an IS element was detected at the duplication junction. When no terminal IS element is located downstream of the

final amplicon of a GDA mutant, its PCR product size will correspond to that of the wild type (ADP1) and the Ben⁻ parent, x bp (Table 2.2, Fig. 2.6A, B). If there were a terminal IS element present, the PCR product size of the wild-type configuration (x bp) would be increased by the size of IS1236, 1237 bp, yielding a total product size for the GDA mutant of $x + 1237$ bp (Table 2.2, Fig. 2.6C).

In ACN1039- and ACN1050-derived GDA mutants, these primer sets (Table 2.S1) generated a PCR product of a size corresponding to the wild-type configuration (Fig. 2.6D and data not shown). In no case was a larger product detected that would have indicated a terminal IS element, despite sufficient extension time and use of a polymerase mix competent for the formation of the longer product (Materials and Methods). These results contradict the model displayed in Fig. 2.3A of how the gene amplification mutants could be generated by transposition of IS1236 followed by homologous recombination between identical sequences to yield duplication and further amplification.

Confirmation of results by Southern hybridization

Southern hybridization was additionally used to confirm the absence of an IS element downstream of the final amplicon. In a similar fashion to the PCR method, the size of fragments detected by hybridization can distinguish the two potential configurations of an amplification mutant (Fig. 2.7B and C). A probe (depicted as a thick black line in Fig. 2.7) was designed to hybridize to DNA adjacent to the IS1236 sequence in the duplication junction. When chromosomal DNA from an amplification mutant is digested with a restriction enzyme that does not cut within IS1236, this probe should hybridize to a fragment corresponding to the junction fragment labeled z in Fig. 2.7B and C. Using information from the duplication junction sequence (Fig. 2.5), the size of this

fragment can be predicted (Table 2.3). Because this fragment contains the duplication junction sequence, a fragment of this size would be expected only in the Ben⁺ GDA mutants and not in comparable samples of DNA from the wild type or Ben⁻ parent strain. As shown in Fig. 2.7D, junction fragments of the expected sizes were detected in three mutants, ACN1056, ACN1057, and ACN1058 (labeled z). As expected, these fragments were not detected in the ADP1 wild-type strain or the ACN1050 parent strain (Fig. 2.7D).

In the region downstream of the final amplicon, this same probe should distinguish between two possible chromosomal configurations. In a GDA mutant with no terminal IS element downstream of the final amplicon (Fig. 2.7B), the probe would hybridize to a fragment of wild-type size, indicated as y in Fig. 2.7A and B, whose value is shown in Table 2.3. In a GDA mutant with the configuration shown in Fig. 2.7C, the presence of a terminal IS element would increase the size of the fragment detected by the probe. Relative to fragment y (Fig. 2.7B), the probe should hybridize to a fragment whose size increases by the number of nucleotides in the IS element, $y + 1237$ bp (Fig. 2.7C, Table 2.3). Southern hybridization analysis of four GDA mutants (ACN1056, ACN1057, ACN1058, and ACN1101) demonstrated the presence of a wild-type sized fragment in each strain (Fig. 2.7D and data not shown). These results are consistent with the PCR-based conclusion that there is no terminal IS element downstream of the final amplicon.

Contradiction of the predicted model for transposition-based gene amplification

Our initial model proposed that an IS-mediated duplication was generated by a transposition event that places IS1236 downstream of the *cat* genes. Following transposition, duplication and further amplification were proposed to occur by homologous recombination (Fig. 2.3A). In this scheme, recombination could occur via

intra-molecular interactions on a single chromosome or between identical copies of the chromosome within a cell. This model is based on standard assumptions for the well-documented behavior of transposable elements and homologous recombination . Moreover, this combination of transposition and homologous recombination has often been reported to underlie genetic duplication (21, 23, 28-30). Nevertheless, the absence of a detectable terminal IS element in our mutants indicates that this model for IS involvement is incorrect for the GDA events investigated here.

Alternative models of IS-mediated gene amplification

Two different models, shown in Fig. 2.3B and 2.3C, might account for the observed chromosomal configuration in the GDA mutants. As illustrated in Fig. 2.3B, the initiating duplication event could involve some type of illegitimate recombination between wild-type sequence downstream of the *cat*-gene cassette and the *IS1236* element that normally resides upstream of *catA*. The recombining DNA sequences that form each duplication junction are shown in Fig. 2.5. If these duplication junctions form by illegitimate recombination, then the precision of these events is clearly significant. Each duplication junction occurs at the exact basepair of the *IS1236* terminal inverted repeat where the transposase would nick to yield a free 3' hydroxyl that initiates the first strand transfer reaction in copy-and-paste transposition (15).

Illegitimate recombination has previously been considered to play a role in duplications involving an IS element. Specifically, in an experimental system involving Tn9, illegitimate recombination was proposed to initiate tandem amplification (7). In this example, high concentrations of chloramphenicol can select multiple copies of the Tn9-borne drug-resistance determinant. When this process was assessed in a *recA* mutant, the

initial step of the transposon duplication was observed to occur at a low frequency in a *recA*-independent fashion (7).

Because some illegitimate recombination mechanisms are mediated by short stretches of identity, we evaluated the level of identity between the DNA sequences that recombined. At the exact duplication site, all GDA mutants contain six or fewer identical nucleotides (highlighted by bold text in Fig. 2.5). In ten mutants there are no identical nucleotides at the duplication junction. Thus, the duplication mechanism does not appear to depend on sequence identity. However, the precision of the recombination site suggests that an *IS1236* transposase is involved even though the model for illegitimate recombination shown in Fig. 2.3B does not invoke a typical transposition mechanism.

A transposase-mediated duplication process for the ACN1050-derived mutants is especially interesting. The nearest IS element to locus *ii* in this parent strain is *IS1236_6*, and, as noted earlier, *IS1236_6* appears not to encode a functional transposase (Fig. 2.S1). At this locus, a transposase encoded by a different element may act in *trans*. However, *trans*-acting activity is presumed to occur infrequently with the IS3 family elements (11, 30).

Another explanation for the chromosomal configuration of the GDA mutants is illustrated in Fig. 2.3C. In this scenario, a typical transposition event places a novel *IS1236* copy downstream of the *cat*-gene cassette on one copy of the chromosome. Chromosomal duplication could then result from homologous recombination between this element and an *IS1236* element upstream of the *cat*-gene cassette on a different chromosomal copy that has not undergone transposition. Having different chromosomal copies in a bacterial cell may not be as rare as previously thought, since recent studies

indicate that prokaryotic polyploidy is common (6, 22, 32, 44, 45). Unless the original copy of the chromosome that underwent the novel transposition event is replicated before homologous recombination with a sister chromosome, our Southern hybridization and PCR methods would not detect it.

To determine whether a mechanism involving transposition (Fig. 2.3C) could be distinguished from one without transposition (Fig. 2.3B), we investigated target sequence duplication. A key characteristic of IS3-family transposition is that it generates a small duplication of the target such that the IS element is typically flanked by 3-bp direct repeats (30). Unfortunately, the analysis of potential target sequence duplication is unable to distinguish between the alternative duplication mechanisms shown in Fig. 2.3B and C (Fig. 2.S2 and 2.S3).

Copy-and-paste transposition

The model of illegitimate recombination shown in Fig. 2.3B was proposed because our results could not be explained by the documented behavior of other IS elements. In the well-studied copy-and-paste transposition mechanism of the IS3-family element *IS911*, the transposase binds and assembles terminal inverted repeat sequences (IRL and IRR) in a synaptic complex (37). After a single end of the element is cleaved, it is transferred to the same DNA strand 3 bp from the opposite end. This intermediate is resolved by host-mediated replication, which restores the donor DNA and also yields a circular transposition molecule. This circular element can synapse with a site adjacent to an *IS911* IR in the target, thereby allowing the transposase to cleave and transfer one IR strand of the mobile element to the target DNA. Host recombination enzymes complete the *IS911* insertion by resolving the four-way branched DNA intermediate of this IR-

targeted process (46). If the transposase brings together one *IS911* IR sequence with that of another *IS911* element (47) or with a chromosomal surrogate sequence (33), the resulting circular transposition intermediate includes the intervening sequence. By this method, the intervening sequence can be duplicated.

For an analogous mechanism to account for *cat*-gene duplication, there must be a second *IS1236* or a surrogate IR site downstream of the *cat* genes, and host replication would need to generate a >40 kbp circular transposition intermediate. However, the presence of a second appropriately placed *IS1236* element was ruled out by our experiments described earlier. Moreover, sites that might serve as IR surrogates were not detected when the target sequences surrounding duplication junctions (Fig. 2.5) were compared to the terminal 30 bp for IRL and IRR of *IS1236* (data not shown). Therefore, while the *IS1236*-associated duplication products are not consistent with an IR-targeted mechanism, the *IS911* studies highlight the importance of assessing target sequences and proteins that might contribute to processing alternative transposition products.

Concluding remarks

In previous studies of gene amplification in *A. baylyi*, *IS1236* was not found to be a significant factor in GDA events and was identified in only two of 91 characterized duplication junctions (12, 34, 40). In contrast, 86% of the 21 GDA mutants that were characterized in our current study display IS-involved rearrangements to form the duplication junctions. One difference between the past and present studies lies in the chromosomal regions from which amplification mutants were selected. Here, the selection cassette resided closer to the nearest IS element than in previous studies, within

32 kbp rather than 190 kbp (Fig. 2.2). This proximity may be responsible for the higher proportion of IS-mediated rearrangements.

The paucity of systematic GDA studies prevents the direct comparison of these results with those from other bacteria in which genetic rearrangements have been studied, such as *E. coli* and *Salmonella enterica*. Typically, gene amplification is difficult to study because of its transient and variable nature. The selection system developed to capture and characterize chromosomal amplicons in *A. baylyi* is unique and continues to reveal novel features of genomic rearrangement.

Surprisingly, PCR and Southern hybridization methods demonstrated the inability of a common model, which assumes transposition is followed by homologous recombination (Fig. 2.3A), to account for the formation of *A. baylyi* GDA mutants. The actual events responsible for generating these GDA mutants remain unknown. They may involve transposase-mediated DNA cleavage and illegitimate recombination (Fig. 2.3B). Alternatively, different chromosomal copies may undergo unequal homologous recombination (Fig. 2.3C). Aspects of the schemes shown in Fig. 2.3B and 2.3C differ from commonly accepted models of insertion sequence-mediated GDA. Nevertheless, the isolation of 18 independent mutants that display identical characteristics of IS-involved rearrangement suggests that a common mechanism is responsible for their formation. A greater understanding of the role of insertion sequences in gene duplication should lead to a deeper appreciation of recombination as a driving force in genomic variation.

References

1. **Albertson, D. G.** 2006. Gene amplification in cancer. *Trends Genet.* **22**:447-455.
2. **Anderson, P., and J. Roth.** 1981. Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (*rrn*) cistrons. *Proc. Natl. Acad. Sci. U. S. A.* **78**:3113-3117.
3. **Andersson, D. I.** 2011. Evolving promiscuously. *Proc. Natl. Acad. Sci. U. S. A.* **108**:1199-1200.
4. **Andersson, D. I., and D. Hughes.** 2009. Gene amplification and adaptive evolution in bacteria. *Annu. Rev. Genet.* **43**:167-195.
5. **Barbe, V., D. Vallenet, N. Fonknechten, A. Kreimeyer, S. Oztas, L. Labarre, S. Cruveiller, C. Robert, S. Duprat, P. Wincker, L. N. Ornston, J. Weissenbach, P. Marliere, G. N. Cohen, and C. Medigue.** 2004. Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. *Nucleic. Acids. Res.* **32**:5766-5779.
6. **Breuert, S., T. Allers, G. Spohn, and J. Soppa.** 2006. Regulated polyploidy in halophilic archaea. *PLoS One* **1**:e92.
7. **Chandler, M., E. B. de la Tour, D. Willems, and L. Caro.** 1979. Some properties of the chloramphenicol resistance transposon Tn9. *Mol. Gen. Genet.* **176**:221-231.
8. **Chandler, M., and J. Mahillon.** 2002. Insertion Sequences Revisited, p. 306-366. *In* N. L. Craig, R. Craigie, M. Gellert, and A. Lambowitz (ed.), *Mobile DNA II*. ASM Press, Washington, DC.

9. **Conrad, B., and S. E. Antonarakis.** 2007. Gene duplication: a drive for phenotypic diversity and cause of human disease. *Annu. Rev. Genomics Hum. Genet.* **8**:17-35.
10. **Coyne, S., P. Courvalin, and M. Galimand.** 2010. Acquisition of multidrug resistance transposon Tn6061 and IS6100-mediated large chromosomal inversions in *Pseudomonas aeruginosa* clinical isolates. *Microbiology* **156**:1448-1458.
11. **Craig, N. L.** 1996. Transposition, p. 2339-2362. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd ed., vol. 2. ASM Press, Washington, DC.
12. **Craven, S. H.** 2009. Gene expression and dosage: distinct mechanisms regulate benzoate degradation in *Acinetobacter baylyi* ADP1. Ph.D. University of Georgia, Athens, GA.
13. **Craven, S. H., and E. L. Neidle.** 2007. Double trouble: medical implications of genetic duplication and amplification in bacteria. *Future Microbiol.* **2**:309-321.
14. **de Berardinis, V., D. Vallenet, V. Castelli, M. Besnard, A. Pinet, C. Cruaud, S. Samair, C. Lechaplais, G. Gyapay, C. Richez, M. Durot, A. Kreimeyer, F. Le Fevre, V. Schachter, V. Pezo, V. Doring, C. Scarpelli, C. Medigue, G. N. Cohen, P. Marliere, M. Salanoubat, and J. Weissenbach.** 2008. A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* **4**:174.

15. **Duval-Valentin, G., B. Marty-Cointin, and M. Chandler.** 2004. Requirement of *IS911* replication before integration defines a new bacterial transposition pathway. *EMBO J.* **23**:3897-3906.
16. **Elliott, K. T., and E. L. Neidle.** 2011. *Acinetobacter baylyi* ADP1: transforming the choice of model organism. *IUBMB Life* **63**:1075-1080.
17. **Eraso, J. M., and S. Kaplan.** 1994. *prpA*, a putative response regulator involved in oxygen regulation of photosynthesis gene expression in *Rhodobacter sphaeroides*. *J. Bacteriol.* **176**:32-43.
18. **Gerischer, U., D. A. D'Argenio, and L. N. Ornston.** 1996. *IS1236*, a newly discovered member of the IS3 family, exhibits varied patterns of insertion into the *Acinetobacter calcoaceticus* chromosome. *Microbiology* **142**:1825-1831.
19. **Gerischer, U., and L. N. Ornston.** 1995. Spontaneous mutations in *pcaH* and *-G*, structural genes for protocatechuate 3,4-dioxygenase in *Acinetobacter calcoaceticus*. *J. Bacteriol.* **177**:1336-1347.
20. **Gralton, E. M., A. L. Campbell, and E. L. Neidle.** 1997. Directed introduction of DNA cleavage sites to produce a high-resolution genetic and physical map of the *Acinetobacter* sp. strain ADP1 (BD413UE) chromosome. *Microbiology* **143**:1345-1357.
21. **Gray, Y. H.** 2000. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends Genet.* **16**:461-468.
22. **Griese, M., C. Lange, and J. Soppa.** 2011. Ploidy in cyanobacteria. *FEMS Microbiol. Lett.* **323**:124-131.

23. **Haack, K. R., and J. R. Roth.** 1995. Recombination between chromosomal IS200 elements supports frequent duplication formation in *Salmonella typhimurium*. *Genetics* **141**:1245-1252.
24. **Juni, E., and A. Janik.** 1969. Transformation of *Acinetobacter calco-aceticus* (*Bacterium anitratum*). *J. Bacteriol.* **98**:281-288.
25. **Kichenaradja, P., P. Siguier, J. Pérochon, and M. Chandler.** 2010. ISbrowser: an extension of ISfinder for visualizing insertion sequences in prokaryotic genomes. *Nucl. Acids Res.* **38**:D62-68.
26. **Kok, R. G., D. A. D'Argenio, and L. N. Ornston.** 1997. Combining localized PCR mutagenesis and natural transformation in direct genetic analysis of a transcriptional regulator gene, *pobR*. *J. Bacteriol.* **179**:4270-4276.
27. **Kugelberg, E., E. Kofoid, A. B. Reams, D. I. Andersson, and J. R. Roth.** 2006. Multiple pathways of selected gene amplification during adaptive mutation. *Proc. Natl. Acad. Sci. U. S. A.* **103**:17319-17324.
28. **Kusumoto, M., T. Ooka, Y. Nishiya, Y. Ogura, T. Saito, Y. Sekine, T. Iwata, M. Akiba, and T. Hayashi.** 2011. Insertion sequence-excision enhancer removes transposable elements from bacterial genomes and induces various genomic deletions. *Nat. Commun.* **2**:152.
29. **Ling, A., and R. Cordaux.** 2010. Insertion sequence inversions mediated by ectopic recombination between terminal inverted repeats. *PLoS ONE* **5**:e15654.
30. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**:725-774.

31. **Montano, S. P., and P. A. Rice.** 2011. Moving DNA around: DNA transposition and retroviral integration. *Curr. Opin. Struct. Biol.* **21**:370-378.
32. **Pecoraro, V., K. Zerulla, C. Lange, and J. Soppa.** 2011. Quantification of ploidy in proteobacteria revealed the existence of monoploid, (mero-)oligoploid and polyploid species. *PLoS One* **6**:e16392.
33. **Polard, P., L. Seroude, O. Fayet, M. F. Prère, and M. Chandler.** 1994. One-ended insertion of *IS911*. *J. Bacteriol.* **176**:1192-1196.
34. **Reams, A. B., and E. L. Neidle.** 2004. Gene amplification involves site-specific short homology-independent illegitimate recombination in *Acinetobacter* sp. strain ADP1. *J. Mol. Biol.* **338**:643-656.
35. **Reams, A. B., and E. L. Neidle.** 2003. Genome plasticity in *Acinetobacter*: new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. *Mol. Microbiol.* **47**:1291-1304.
36. **Roth, J. R., N. Benson, T. Galitski, K. R. Haack, J. G. Lawrence, and L. Miesel.** 1996. Rearrangements of the Bacterial Chromosome: Formation and Applications, p. 2256-2276. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd ed., vol. 2. American Society for Microbiology, Washington, DC.
37. **Rousseau, P., C. Normand, C. Loot, C. Turlan, R. Alazard, G. Duval-Valentin, and M. Chandler.** 2002. Transposition of *IS911*, p. 367-383. *In* N. L.

- Craig, R. Craigie, M. Gellert, and A. Lambowitz (ed.), Mobile DNA II. ASM Press, Washington, DC.
38. **Rousseau, P., C. Tardin, N. Tolou, L. Salomé, and M. Chandler.** 2010. A model for the molecular organisation of the IS911 transpososome. *Mob. DNA* **1**:16.
 39. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1987. Molecular cloning : a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 40. **Seaton, S. C., K. T. Elliott, L. E. Cuff, N. S. Laniohan, P. R. Patel, and E. L. Neidle.** 2012. Genome-wide selection for increased copy number in *Acinetobacter baylyi* ADP1: locus and context-dependent variation in gene amplification. *Mol. Microbiol.* **83**:520-535.
 41. **Segura, A., P. V. Bünz, D. A. D'Argenio, and L. N. Ornston.** 1999. Genetic analysis of a chromosomal region containing *vanA* and *vanB*, genes required for conversion of either ferulate or vanillate to protocatechuate in *Acinetobacter*. *J. Bacteriol.* **181**:3494-3504.
 42. **Shanley, M. S., E. L. Neidle, R. E. Parales, and L. N. Ornston.** 1986. Cloning and expression of *Acinetobacter calcoaceticus* *catBCDE* genes in *Pseudomonas putida* and *Escherichia coli*. *J. Bacteriol.* **165**:557-563.
 43. **Stankiewicz, P., and J. R. Lupski.** 2010. Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* **61**:437-455.
 44. **Tobiason, D. M., and H. S. Seifert.** 2010. Genomic content of *Neisseria* species. *J. Bacteriol.* **192**:2160-2168.

45. **Tobiason, D. M., and H. S. Seifert.** 2006. The obligate human pathogen, *Neisseria gonorrhoeae*, is polyploid. PLoS Biol. **4**:e185.
46. **Turlan, C., C. Loot, and M. Chandler.** 2004. IS911 partial transposition products and their processing by the *Escherichia coli* RecG helicase. Mol. Microbiol. **53**:1021-1033.
47. **Turlan, C., B. Ton-Hoang, and M. Chandler.** 2000. The role of tandem IS dimers in IS911 transposition. Mol. Microbiol. **35**:1312-1325.

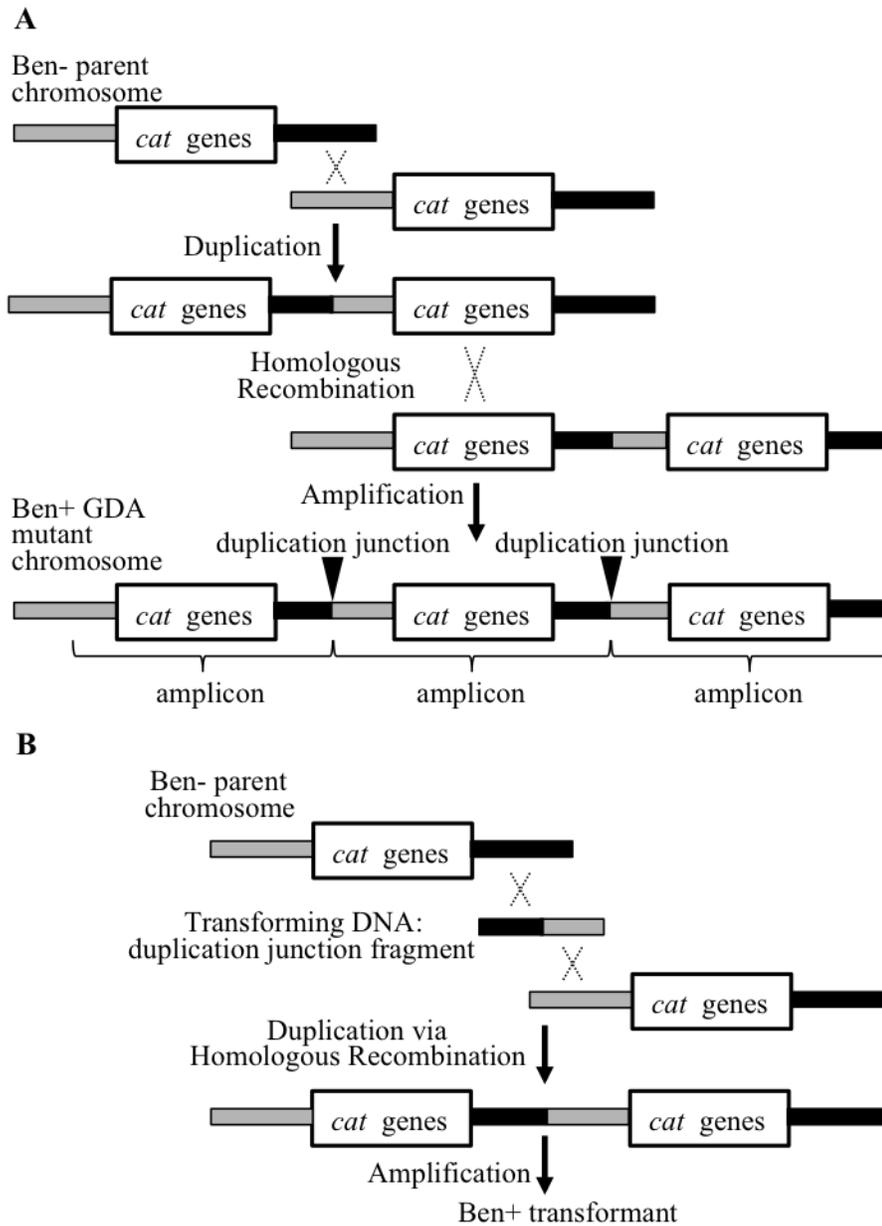


Figure 2.1. Selection and analysis of Ben⁺ amplification mutants. (A) Two-step model for the formation of Ben⁺ amplification mutants. In the first step, a chromosomal region is duplicated by recombination (X) between DNA that is normally downstream of the *cat* genes (black area) and DNA normally upstream of these genes (gray area) on the Ben⁻ parent chromosome. This step does not depend on homology and may be mediated by illegitimate recombination (34, 35, 40). The “*cat genes*” label denotes an approximately 10 kbp region including the *catA* gene and the *catBCIJFD* operon, which are transcribed from separate promoters and expressed weakly due to the absence of two transcriptional activators. In the second step of the amplification process, identity between the repeated

segments of the duplication enables further increases in copy number via homologous recombination. Multiple copies of the repeated segment, the amplicon, can increase gene expression and thereby generate a Ben⁺ mutant. The duplication junction defines the endpoints of the amplicon. Its sequence allows inferences about the recombination event that generated the duplication. (B) To identify the duplication junction, fragments of DNA from a Ben⁺ GDA mutant are used to transform a Ben⁻ parent strain. A Ben⁺ phenotype generated by homologous recombination signals the presence of the duplication junction DNA fragment.

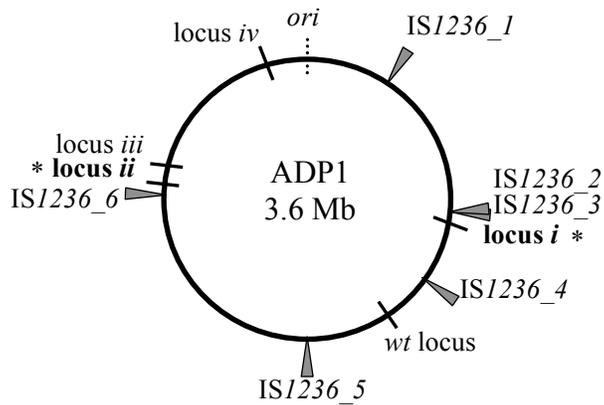


Figure 2.2. Position of IS elements on circular map of the ADP1 chromosome. Six genomic copies of *IS1236* reside in the locations shown relative to the origin of replication (*ori*). *IS1236_1* is oriented with the coding sequences of the transposases in a clockwise direction, and all the other copies are oppositely oriented. The native position of the *cat* genes (*wt* locus) is shown relative to where these genes were repositioned (loci *i* to *iv*) in engineered parent strains, as described in the text. Mutants described in this study arose with the *cat* genes located in the regions marked by bold text (*): locus *i* (ACIAD0982), which is 24 kbp from *IS1236_3*, and locus *ii* (ACIAD2822), which is 32 kbp from *IS1236_6*. The *wt* locus is 190 kbp from *IS1236_4*.

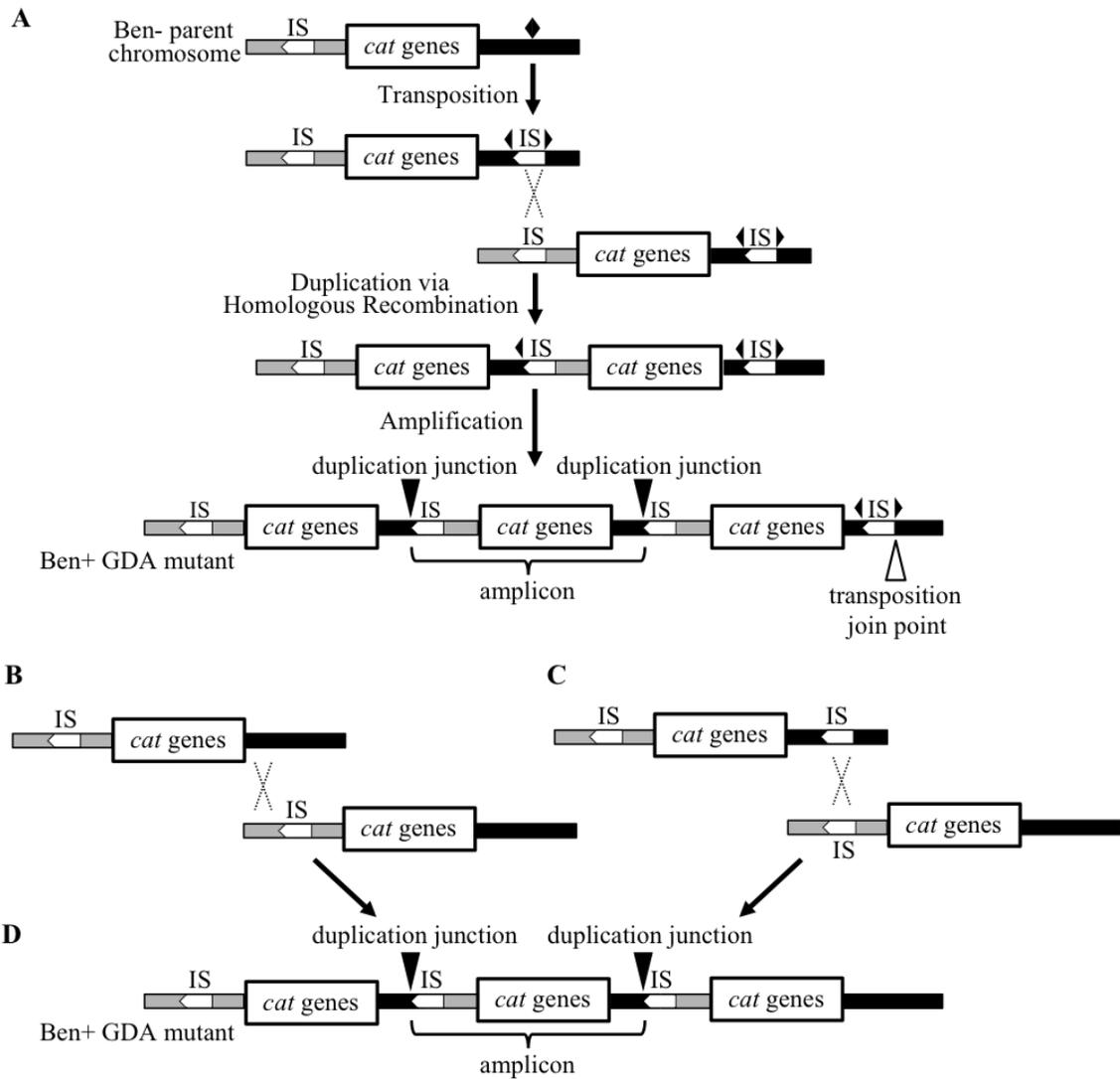


Figure 2.3. Roles for IS1236 in gene amplification. (A) Model for the formation of Ben⁺ GDA mutants following the transposition of IS1236 (IS) to a new target site (marked by a black diamond). After the IS element is inserted in this site (depicted by two outward facing triangles of a split diamond), homologous recombination between IS elements can generate a duplication. Further amplification can occur by homologous recombination. According to this model, there should be an IS element present downstream of the final amplicon in which the connection to downstream DNA creates a unique junction (the transposition join point). Other labels and representations are the same as for Fig. 2.1 (B) In an alternative model, illegitimate recombination could generate a duplication (not shown) that undergoes further amplification by homologous recombination. (C) In a different alternative model, there is transposition of an IS element to a position downstream of the *cat* genes on one copy of the chromosome (top). Homologous recombination could then occur between this transposed IS element and an IS element upstream of the *cat* genes on a different copy of the chromosome (bottom). (D)

Homologous recombination between duplications formed by the methods shown in panels B and C could generate a Ben⁺ GDA mutant with the depicted chromosomal configuration. This configuration lacks a terminal IS element downstream of the *cat* genes.

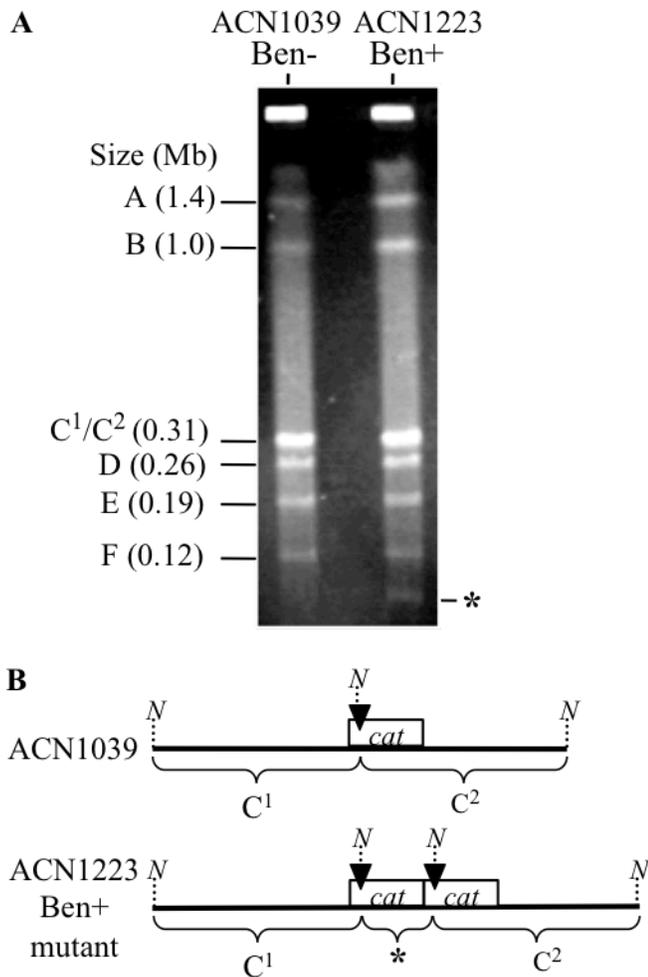


Figure 2.4. Amplicon size analysis by PFGE. (A) Separation of NotI-digested genomic DNA from a Ben⁺ GDA mutant (ACN1223) and its Ben⁻ parent strain (ACN1039). Labels A, B, D, E, and F correspond to wild-type NotI-generated chromosomal fragments. The wild type additionally has a fragment (C, ~0.62 Mb) that is cleaved in both ACN1039 and ACN1223 by a NotI site in the relocated *cat* genes. The resulting fragments (C¹ and C²) are nearly equal in size and cannot be separated by PFGE. One fragment, which is present in ACN1223 (*) but not in ACN1039, corresponds to an amplicon of ~41 kbp, as discussed in the text. (B) Illustration of restriction pattern differences between ACN1039 and ACN1223 in the vicinity of the relocated *cat*-gene cassette. A black arrowhead marks the position of a NotI restriction site (N) that generates two nearly equal sized fragments in ACN1039 (C¹ and C²). When there is gene amplification of the *cat* region, multiple copies of the marked NotI site generate a fragment whose size corresponds to that of its amplicon, as depicted for ACN1223 (*).

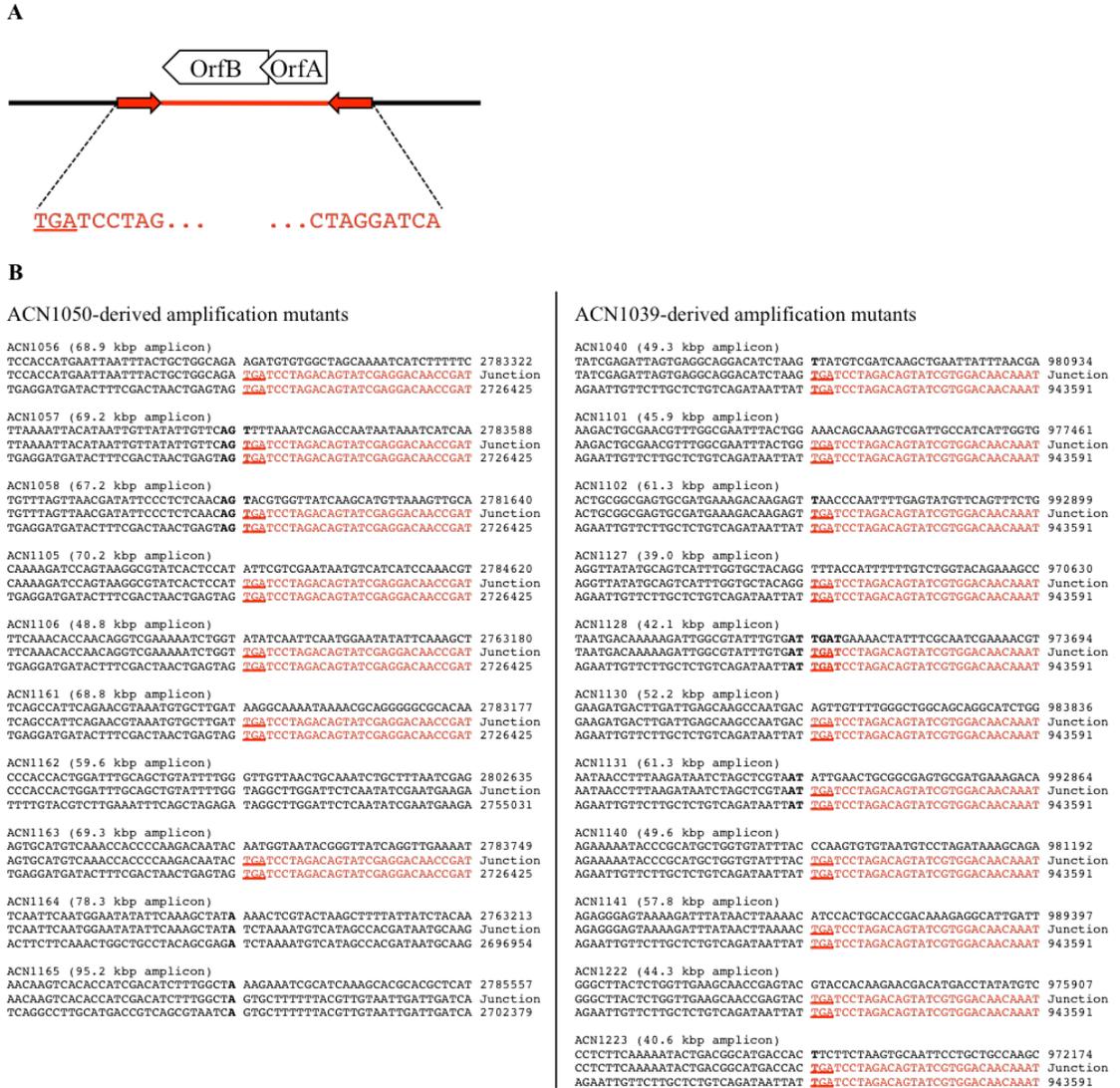


Figure 2.5. IS1236 sequences identified in duplication junctions. (A) Diagram of the IS1236 element with red arrows representing inverted repeats. Ten nucleotides are shown at each end of these repeats. OrfA, OrfAB, and OrfB are presumed to be transposase proteins, and the coding orientations of their genes (indicated by arrows) are opposite to those of the *cat* genes for the IS elements closest to the cassette in locus *i* or *ii*. (B) Duplication junction sequences for Ben⁺ amplification mutants. The middle line shows the DNA sequence of each experimentally determined duplication junction. The top and bottom lines indicate the known chromosomal sequence of DNA that normally occurs downstream or upstream, respectively, of the relocated *cat*-gene cassette starting with the indicated genomic coordinate (from GenBank entry CR543861). Nucleotide identity at the duplication junction is shown in bold typeface. IS1236 sequence is shown in red text with the first three basepairs of its inverted repeat underlined.

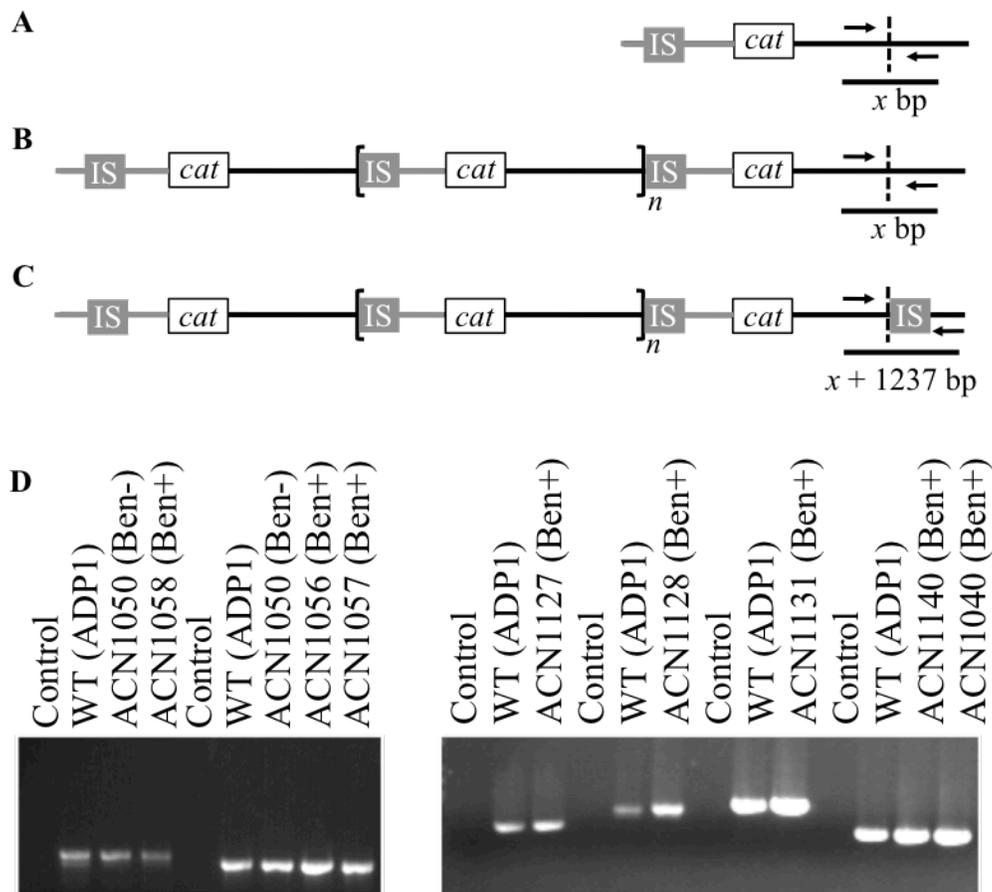


Figure 2.6. PCR-based analysis of chromosomal configurations. Panel A diagrams the expected PCR product in the Ben^- parent strain. The wild-type chromosome is identical to this configuration except the *cat*-gene cassette is in its native locus. The chromosomal shading (grey vs. black) is as described in Figure 2.1. Panels B and C depict two different chromosomal arrangements that could occur downstream of the terminal amplicon in a Ben^+ GDA mutant. The region within the bracket represents n copies of an amplicon. The vertical dashed line indicates the endpoint of the final amplicon or the equivalent nucleotide in the wild-type or parent chromosome. In panel B, the chromosomal configuration in the PCR-amplified region matches that of the wild-type strain (WT, ADP1) as well as its Ben^- parent. In the configuration shown in panel C, the presence of an IS element could result from the transposition and amplification process depicted in Fig. 2.3A. To differentiate between these two possibilities, PCR primers (Table 2.S1) were designed for specific GDA mutants based on DNA sequence from the experimentally determined duplication junction. PCR product sizes (Table 2.2) can be predicted from the primers used, either x bp for configurations A and B or x bp + 1237 (the number of nucleotides in the IS element) for configuration C. (D) The results of PCR with the wild type, Ben^- parent strain, Ben^+ GDA mutant or no DNA template are shown. DNA size standards (not shown) indicated in all cases that the wild-type fragments matched the expected sizes shown in Table 2.2. In all cases, the PCR-generated fragments from the amplification mutants were the same size as those from the wild type and Ben^- parent strain.

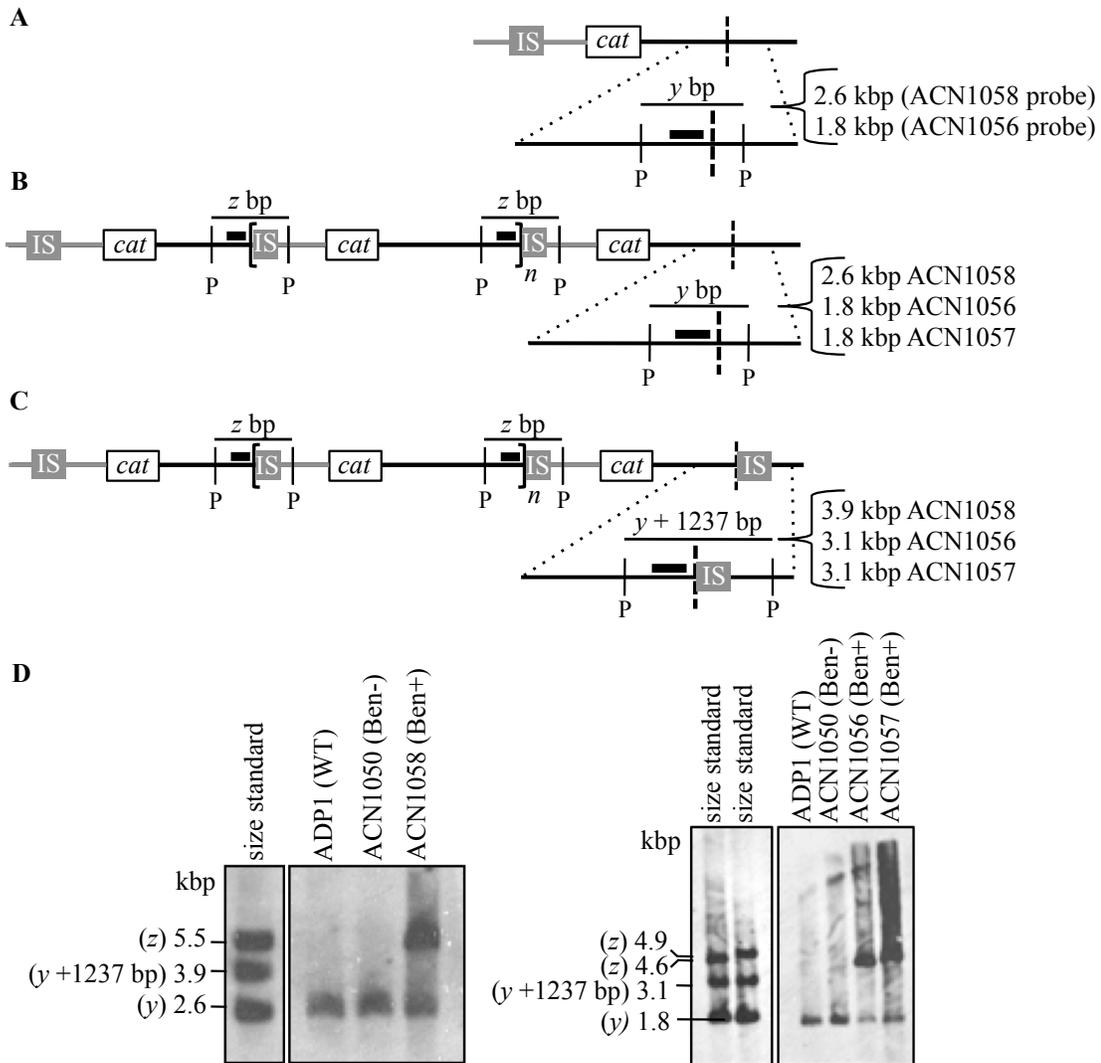


Figure 2.7. Analysis of chromosomal configurations by Southern hybridization. Panels A, B, and C depict the Ben⁻ parent and two different possible chromosomal arrangements downstream of the terminal amplicon in a Ben⁺ GDA mutant. The wild-type chromosome is identical to the Ben⁻ parent except the *cat*-gene cassette is in its native locus. These arrangements are the same as those described in Fig. 2.6A, B, and C. The chromosomal shading is as described in Fig. 2.1. Dotted lines show an enlarged view of a region that hybridizes to a labeled probe that abuts the duplication junction (depicted as a thick black bar). When chromosomal DNA is digested with the restriction enzyme PstI (P), sequence analysis allows the size prediction of the fragment that encompasses the duplication junction to which the probe will hybridize (z). This analysis also predicts the size of the wild-type PstI-generated fragment to which the same probe will hybridize (y). For wild-type and the Ben⁻ parent strain, the size of y depends on the probe used. For GDA mutants with the chromosomal configuration shown in B or C, the probe should detect a fragment of size y bp or y bp + 1237 (the size of the IS element), respectively. The predicted fragment sizes for different mutants vary according to the specific junction and probe sequences (Table 2.3). (D) The results of Southern hybridizations are shown for chromosomal DNA from the wild type (ADP1) and the mutants indicated. Standards

corresponding to the size of predicted fragments were generated by PCR (as described in Materials and Methods). In no case was a band of size $y + 1237$ bp detected.

Table 2.1. Strains and plasmids used in this study.

Strains used in this study		
<i>A. baylyi</i> Strain	Relevant characteristics^a	Source
ADP1	Wild type (BD413)	(24)
ACN1039	<i>benMΩS4036</i> , <i>benA5147</i> , Δ <i>cat5825</i> (<i>catA-catD</i>); <i>cat</i> region (1439430-1449722) ^b in locus <i>i</i> (ACIAD0982, <i>vanK</i>), <i>catMΩK5541</i> ; Ben ⁻ parent strain	(40)
ACN1040, ACN1101, ACN1102, ACN1127, ACN1128, ACN1130, ACN1131, ACN1140, ACN1141	Benzoate ⁺ mutant derived from ACN1039	(40)
ACN1050	<i>benMΩS4036</i> , <i>benA5147</i> , Δ <i>cat5825</i> ; <i>cat</i> region (1439430-1449722) ^b in locus <i>ii</i> (ACIAD2822), <i>catMΩK5541</i> ; Ben ⁻ parent strain	(40)
ACN1056-ACN1058, ACN1105, ACN1106, ACN1161-ACN1165	Benzoate ⁺ mutant derived from ACN1050	(40)
ACN1222-ACN1223	Benzoate ⁺ mutant derived from ACN1039	This study
Plasmids used in this study		
Plasmid	Relevant characteristics	Source
pZErO-2	Km ^R ; Zero background cloning vector	Invitrogen
pUC19	Ap ^R ; cloning vector	(40)
pBAC941	ACN1056 junction plasmid pZErO-2 with <i>EcoRI</i> insert (2779591-2783351; 2726455-2728826) ^b	This study
pBAC942	ACN1057 junction plasmid pZErO-2 with <i>EcoRI</i> insert (2779591-2783617; 2726455-2728826) ^b	This study
pBAC943	ACN1058 junction plasmid pZErO-2 with <i>EcoRI</i> insert (2779591-2781669; 2726455-2728826) ^b	This study
pBAC956	ACN1105 junction plasmid pZErO-2 with <i>EcoRI</i> insert (2779591-2784649; 2726455-2737877) ^b	This study
pBAC957	ACN1106 junction plasmid pZErO-2 with <i>EcoRI</i> insert (2762288-2763209; 2726455-2728826) ^b	This study
pBAC993	ACN1161 junction plasmid pZErO-2 with <i>EcoRI</i> insert (2779591-2783206; 2726455-2728826) ^b	This study
pBAC994	ACN1162 junction plasmid pZErO-2 with <i>EcoRI</i> insert (2801423-2802664; 2755061-2761112; 1449722-1444251) ^b	This study
pBAC995	ACN1163 junction plasmid pZErO-2 with <i>EcoRI</i> insert (2779591-2783778; 2726455-2728826) ^b	This study
pBAC996	ACN1164 junction plasmid pZErO-2 with <i>EcoRI</i> insert (2763053-2763242; 2696984-	This study

	2706500) ^b	
pBAC999	ACN1101 junction plasmid pUC19 with <i>EcoRI</i> insert (975200-977490; 943621-946613) ^b	This study
pBAC1000	ACN1102 junction plasmid pUC19 with <i>EcoRI</i> insert (983094-992928; 943621-946634) ^b	This study
pBAC1008	ACN1128 junction plasmid pZErO-2 with <i>EcoRI</i> insert (969600-973723; 943621-946634) ^b	This study
pBAC1009	ACN1140 junction plasmid pZErO-2 with <i>EcoRI</i> insert (975200-981221; 943621-946634) ^b	This study
pBAC1010	ACN1127 junction plasmid pUC19 with <i>EcoRI</i> insert (969600-970659; 943621-946634) ^b	This study
pBAC1011	ACN1040 junction plasmid pUC19 with <i>EcoRI</i> insert (975200-980963; 943621-946634) ^b	This study
pBAC1012	ACN1131 junction plasmid pZErO-2 with <i>EcoRI</i> insert (985699-992893; 943621-946634) ^b	This study
pBAC1013	ACN1141 junction plasmid pZErO-2 with <i>EcoRI</i> insert (985899-989426; 943621-946634) ^b	This study
pBAC1014	ACN1130 junction plasmid pZErO-2 with <i>EcoRI</i> insert (983092-983865; 943621-946634) ^b	This study
pBAC1015	ACN1165 junction plasmid pZErO-2 with <i>EcoRI</i> insert (2779591-2785586; 2702408-2706593) ^b	This study
pBAC1039	ACN1223 junction plasmid pZErO-2 with <i>EcoRI</i> insert (969000-972203; 943621-946634) ^b	This study
pBAC1063	ACN1222 junction plasmid pZErO-2 with <i>EcoRI</i> insert (975200-975936; 943621-946634) ^b	This study

^a Ap^R, ampicillin resistant; ΩK, omega cassette conferring Km^R (17); ΩS, omega cassette conferring resistance to streptomycin and spectinomycin.

^b Position in the ADP1 genome sequence according to GenBank entry CR543861.

Table 2.2. Predicted fragment sizes for PCR analysis^a

Strain	x bp fragment	$x + 1237$ bp fragment
ACN1056	818	2055
ACN1057	818	2055
ACN1058	859	2096
ACN1105	846	2083
ACN1106	866	2103
ACN1161	819	2056
ACN1163	819	2056
ACN1040	745	1982
ACN1101	742	1979
ACN1102	1018	2255
ACN1127	817	2054
ACN1128	968	2205
ACN1130	881	2118
ACN1131	1018	2255
ACN1140	745	1982
ACN1141	976	2033
ACN1222	883	2120
ACN1223	1113	2350

^a The size of fragment x is determined by the primers used (listed in Table 2.S1) to amplify the region depicted in Fig. 2.6, based on genomic DNA sequence (GenBank entry CR543861).

Table 2.3. Predicted fragment sizes for Southern hybridization^a

Strain	<i>y</i> fragment	<i>y</i> + 1237 bp fragment	<i>z</i> fragment (junction)
ACN1056	1826	3096	4586
ACN1057	1826	3096	4853
ACN1058	2598	3868	5503
ACN1101	1501	2738	10807

^a The sizes of fragments *y* and *z* are calculated based on the labeled probes used for hybridizations, which are generated with the PCR primers listed in Table 2.S1, and evaluation of the corresponding regions depicted in Fig. 2.7 in consideration of genomic DNA sequence (GenBank entry CR543861).

IS1236_6, mutations prevent production of full-length OrfA and OrfB proteins. Due to frameshift, the identical aligned GAA nucleotides corresponding to the glutamate (E) residue in the DDE motif do not encode this amino acid in IS1236_6.

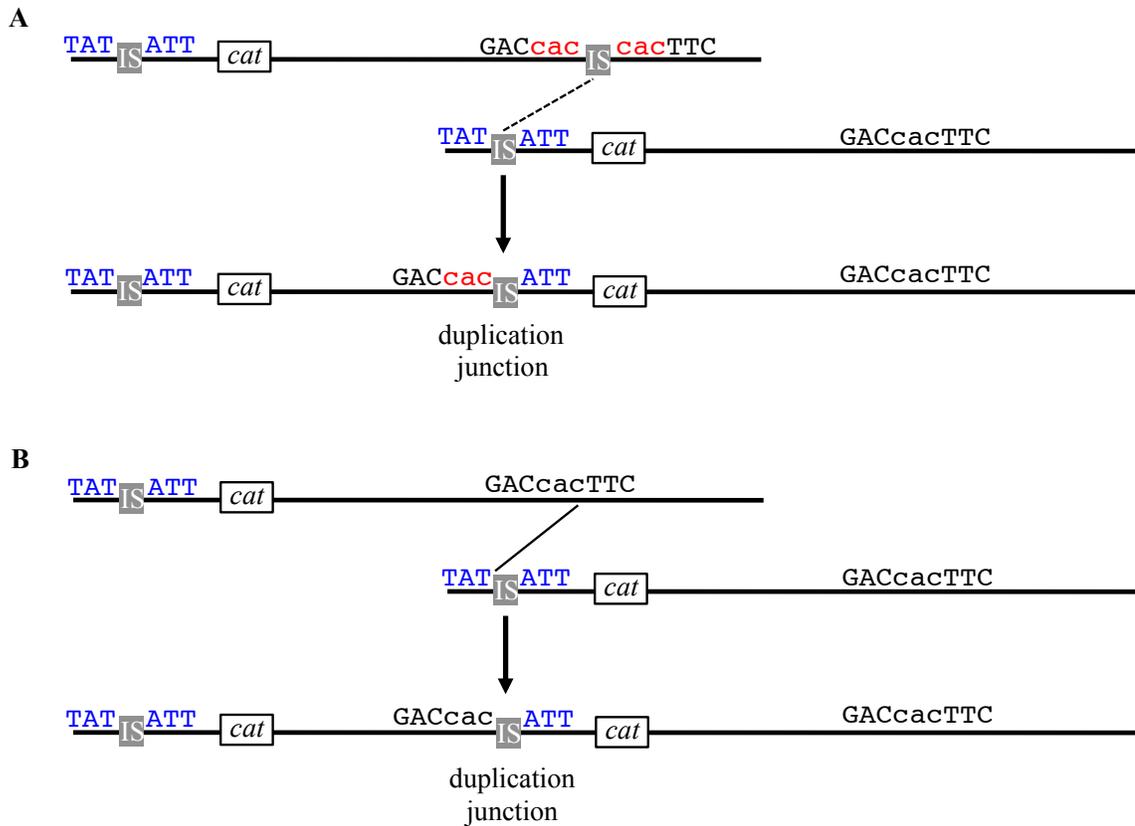


Figure 2.S2. Transposition and 3-bp flanking repeats: different models of duplication predict the same chromosomal configuration in ACN1039-derived mutants. The bottom line of each panel represents the chromosome of an ACN1039-derived Ben⁺ amplification mutant. Two models are illustrated for the formation of the mutant (corresponding to the models in Fig. 2.3B and 2.3C). In panel A (and Fig. 2.3C), an *IS1236* element transposes on one copy of the chromosome and then undergoes unequal homologous recombination with a sister chromosome. In panel B (and Fig. 2.3B), duplication involves illegitimate recombination without transposition. The lowercase typeface (*cac*) indicates the sequence that was identified immediately upstream of the IS element in the duplication junction of mutant ACN1223. In panel A, this sequence corresponds to the 3-bp transposition target that typically becomes duplicated during the transposition of *IS1236* (red text). The absence of direct repeats flanking *IS1236_3* (blue text) may reflect sequence divergence that occurred after an ancient insertion. In the model shown in panel B, this lowercase sequence corresponds to the sequence downstream of the *cat* genes that undergoes illegitimate recombination with the end of *IS1236_3*. As shown by the bottom line of both panels, the resulting chromosomal sequence of the mutant (ACN1223) is the same whether it is generated by homologous recombination between two IS elements (dashed line, panel A) or by illegitimate recombination (thin solid line, panel B). Thus, efforts to identify 3-bp flanking repeats are unable to distinguish between different models of duplication (A and B).

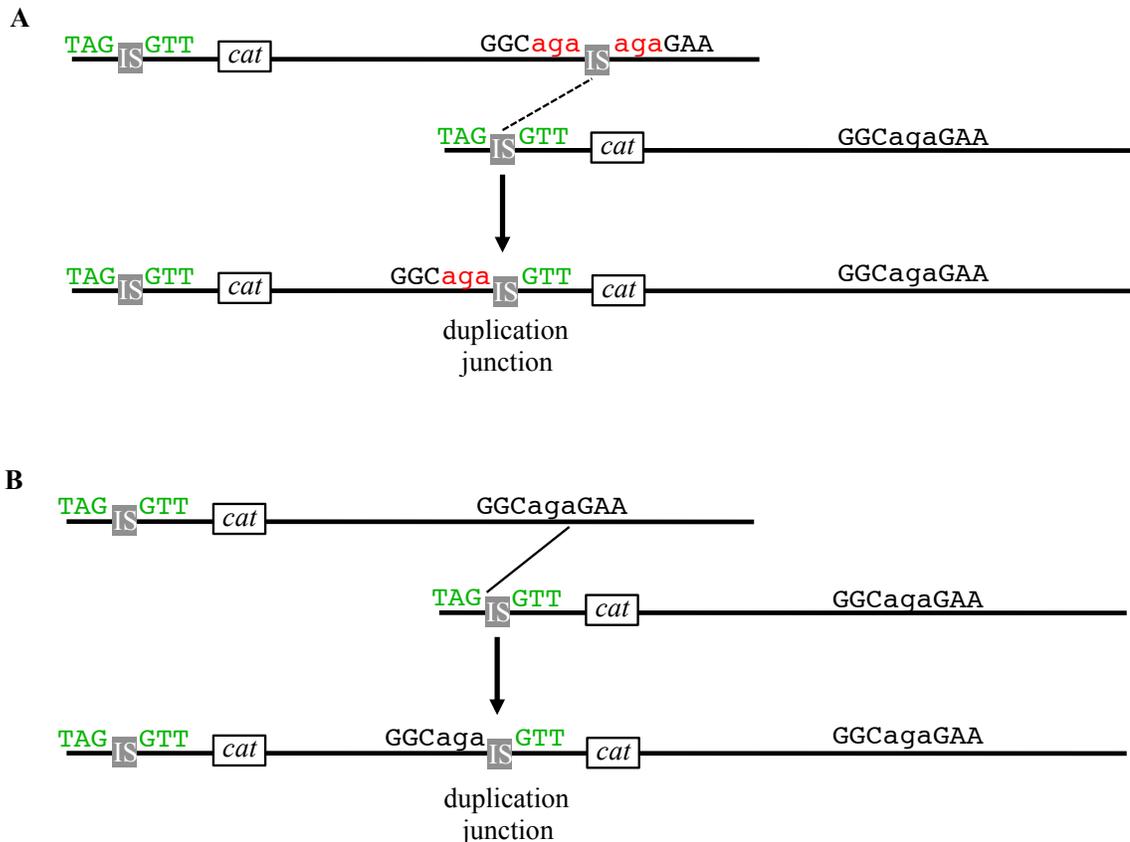


Figure 2.S3. Transposition and 3-bp flanking repeats: different models of duplication predict the same chromosomal configuration in ACN1050-derived mutants. The bottom line of each panel represents the chromosome of an ACN1050-derived Ben⁺ amplification mutant. Two models are illustrated for the formation of the mutant (corresponding to the models in Fig. 2.3B and 2.3C). In panel A (and Fig. 2.3C), an IS1236 element transposes on one copy of the chromosome and then undergoes unequal homologous recombination with a sister chromosome. In panel B (and Fig. 2.3B), duplication involves illegitimate recombination without transposition. The lowercase typeface (aga) indicates the sequence that was identified immediately upstream of the IS element in the duplication junction of mutant ACN1056. In panel A, this sequence corresponds to the 3-bp transposition target that typically becomes duplicated during the transposition of IS1236 (red text). The absence of direct repeats flanking IS1236_6 (green text) may reflect sequence divergence that occurred after an ancient insertion. In the model shown in panel B, this lowercase sequence corresponds to the sequence downstream of the *cat* genes that undergoes illegitimate recombination with the end of IS1236_6. As shown by the bottom line of both panels, the resulting chromosomal sequence of the mutant (ACN1056) is the same whether it is generated by homologous recombination between two IS elements (dashed line, panel A) or by illegitimate recombination (thin solid line, panel B). Thus, efforts to identify 3-bp flanking repeats are unable to distinguish between different models of duplication (A and B).

Table 2.S1. Primers used in this study.

Primer Name	Sequence	Strain
PCR analysis (Fig. 2.6 and Table 2.2)		
2783029_UP	CGCGCATCATCTACCGTACAGTCG	ACN1056, ACN1057, ACN1161, ACN1163
2783847_DN	GGCAAGTGTATAAGAAATCGTAGC	ACN1056, ACN1057, ACN1161, ACN1163
2781256_UP	GATCTGATCACCTCTCAAAGC	ACN1058
2782115_DN	GTAGAATGGGTCGAGTCAGCAGATG	ACN1058
980893_UP	GGGATTGCAGATTTGGTTCGGG	ACN1040, ACN1140
981637_DN	CAAAGCGACGTTTCCATTGAGCCG	ACN1040, ACN1140
976931_UP	GAAGAGTGGCAACCGCGTGTAGAAG	ACN1101
977672_DN	CTTTGTGCCTTGCTGCTCATG	ACN1101
992475_UP	GGATTAGGATTGATATTGCTTGTGG	ACN1102, ACN1131
993492_DN	CATACCTGCGACTCTCCTTGCGG	ACN1102, ACN1131
2784210_UP	CAGGCCATGTGGTCGCGCTG	ACN1105
2785055_DN	GTAGTAGGGTTTTGTTACTAGGATCTGTG	ACN1105
2762759_UP	GCATGCTGGTTTTGAAGTAAGGCC	ACN1106
2763624_DN	CTGAATTGTTGTACTTGAGTGTGG	ACN1106
970003_UP	GATGGCTGGATTCAATTTGATGGG	ACN1127
970819_DN	GGTCATAAAGCGTAACCCATTC	ACN1127
973132_UP	CCTTACGCAAATATCCGCAAG	ACN1128
974099_DN	GTCCTGCTCATAACCGTTGG	ACN1128
983351_UP	CACATATCGCAGCAGATAAATTG	ACN1130
984231_DN	CATGGGCTTCAAGTAAACGGGC	ACN1130
988993_UP	GGGAATATCTAATGCAAAGCGATG	ACN1141
989788_DN	GATGCCATCATTACCCAAGATACC	ACN1141
975336_UP	GACCATTACATACATTACTGCACGG	ACN1222
976218_DN	GCTTCACTGATATCGGAGGGTTG	ACN1222
971804_UP	TCCGCACCCACTACAACATCA	ACN1223
972916_DN	GGAACGCGCTGGCCTGCACATG	ACN1223
DIG-labeling for Southern hybridization (Fig. 2.7 and Table 2.3)		
2782954_UP	GCAGACAAACACTCGGTATGCGTTAC	ACN1056, ACN1057
2783350_DN	CTGCCAGCAGTAAATTAATTCATGG	ACN1056, ACN1057
2781264_UP	CACCTCTCAAAGCAAAGATCAATGG	ACN1058
2781669_DN	CTGTTGAGAGGGAATATCGTTAACTAAAC	ACN1058
977090_UP	CGGTGCAGGTCAGGCCGCGAG	ACN1101
977490_DN	CCAGTAAATTCGCCAAACGTTTCGC	ACN1101
Southern Size Standards (Fig. 2.7 and Table 2.3)		
2782954_UP	GCAGACAAACACTCGGTATGCGTTAC	ACN1056, ACN1057
2784779_DN	AGGCATTAATGAAGTCATTCCTAAATG	ACN1056, ACN1057
2786049_DN	GAAAAAAAAACAATAGTGAATTTTGCGAATATGC	ACN1056, ACN1057
2787539_DN	TCACTGTTTCTACACCAGATTGAATTTTG	ACN1056
2787602_DN	CACCTTCGCCTGTAAATACATAATCAGC	ACN1057
2782750_UP	GCATGGTTATCCTCATAAAGTTTCGC	ACN1057
2781188_UP	CTATTGGTTATGTCCCACTTACGAAG	ACN1058
2781264_UP	CACCTCTCAAAGCAAAGATCAATGG	ACN1058
2783861_DN	TGATTTACACGCGCGCAAGTGTATAAG	ACN1058

2786690_DN	AATAAATGTGAAGGGGAATCAGTTTTAATC	ACN1058
2785055_DN	GTAGTAGGGTTTTGTTACTAGGATCTGTG	ACN1058
966684_UP	TTCGGTAAAAATCTTGCCTTG	ACN1101
974759_UP	CAATATGCAATTGATAATGTCATG	ACN1101
975990_UP	GCAACATCCACAGGATTTAAACG	ACN1101
977490_DN	CCAGTAAATTCGCCAAACGTTTCGC	ACN1101

Table 2.S2. Amplicon analysis of Ben⁺ mutants.

Parent strain	GDA mutant	Amplicon size (kbp)	Amplicon Copy Number	Amount of amplified DNA (kbp) ^a	IS-mediated?
ACN1039	ACN1040	49	9	444	Yes
ACN1039	ACN1101	46	15	681	Yes
ACN1039	ACN1102	61	19	1138	Yes
ACN1039	ACN1127	39	31	1213	Yes
ACN1039	ACN1128	42	20	852	Yes
ACN1039	ACN1130	52	35	1826	Yes
ACN1039	ACN1131	61	27	1619	Yes
ACN1039	ACN1140	50	14	705	Yes
ACN1039	ACN1141	58	12	684	Yes
ACN1039	ACN1222	44	23	1012	Yes
ACN1039	ACN1223	41	2	82	Yes
ACN1050	ACN1056	69	7	507	Yes
ACN1050	ACN1057	69	9	595	Yes
ACN1050	ACN1058	67	8	545	Yes
ACN1050	ACN1105	70	6	402	Yes
ACN1050	ACN1106	49	10	473	Yes
ACN1050	ACN1161	69	26	1768	Yes
ACN1050	ACN1162	60	6	351	No
ACN1050	ACN1163	69	16	1105	Yes
ACN1050	ACN1164	78	16	1242	No
ACN1050	ACN1165	95	17	1606	No

^aAmount of amplified DNA is inferred from amplicon size and copy number. If the product of the listed amplicon size and its copy number differs from the listed amount of DNA, the discrepancy arises from rounding.

CHAPTER 3

GENOMIC COPY NUMBER OF *ACINETOBACTER BAYLYI* ADP1²

² Cuff, L.E. and E.L. Neidle. To be submitted to *Journal of Bacteriology*.

Abstract

Genome copy number per cell, also known as ploidy level, is becoming an area of interest. Well-studied bacteria, such as *Escherichia coli* and *Bacillus subtilis*, have a single chromosome per cell when growing slowly. In an effort to determine whether *Acinetobacter baylyi* was more similar to *E. coli* or to bacteria in the closely related *Pseudomonas* genera, which have greater than ten copies of the genome per cell, genome copy number per cell was determined. qPCR analysis of cultures in exponential growth and early stationary phase show that *A. baylyi* has a low genome copy number per cell, similar to *E. coli*.

Introduction

Bacteria have long been thought to contain a single copy of their chromosome in each cell. The most common model organisms, *Escherichia coli* and *Bacillus subtilis*, both contain one copy of their genome under slow growth conditions (1, 2). However, there are a few notable exceptions. The bacterium *Deinococcus radiodurans* has been shown to have as many as ten copies of its linear chromosome per cell. This extra genetic material allows the bacterium to maintain viability even after exposure to high levels of radiation (3). Additionally, *Borrelia hermsii* is known to contain multiple genomic copies within a single cell, averaging 14 copies of an expressed plasmid and 16 copies of the chromosome (4). Multiple copies of a chromosome can afford additional advantages to cells. Rudimentary gene regulation resulting from expression of multiple copies of genes and quicker transport of gene products to different regions of the cell in the case of a large bacterium (5) are two such advantages. Gene redundancy is also beneficial in situations where deleterious mutation accrual can lead to cell death.

Recently, additional bacteria have been reported to have multiple copies of the genome per cell in slow growing cultures or in stationary phase. These studies contradict the canonical idea of a single copy of the chromosome per bacterial cell. Growth rate and culture medium can affect the ploidy level of cells, as evidenced by the propensity of *E. coli* to initiate multiple rounds of replication prior to completion of the first when doubling rapidly (6). Genomic copy numbers have been experimentally assessed for some proteobacteria: *Azotobacter vinelandii* (7, 8), several *Neisseria* species (9, 10), *Buchnera* species (11), and two *Desulfovibrio* species (12). Copy number varied greatly, from an average of 3 copies for *Neisseria* species to 120 copies for the endosymbiont

Buchnera. Additionally, *Psuedomonas putida*, is polyploid with 20 origins and 14 termini per cell in stationary phase cultures (13).

The ploidy level of *A. baylyi* ADP1 recently came into question when a study of duplication events found recombination that involved copies of *IS1236* without prior transposition on the same DNA molecule (14). *IS1236* is a member of the IS3 family of insertion sequences, which move through a copy-and-paste mechanism of transposition. Molecular-based analyses showed that hypotheses about *IS1236* transposition generating duplications were incorrect. Transposition of *IS1236* on the same chromosome followed by homologous recombination between insertion sequence elements did not form gene duplication events. However, the possibility remained for gene duplication that involved a copy of *IS1236* at the point of recombination. If transposition followed by recombination between multiple chromosomes occurred within a single cell, duplication events would involve *IS1236* at the point of recombination (see Chapter 2 for details). These results, in conjunction with the variable ploidy levels of different bacteria motivated the investigation of genomic copy number in *A. baylyi* ADP1.

Materials and Methods

Cell culture

A. baylyi ADP1 cultures were grown in M9 minimal medium (15) supplemented with 20 mM pyruvate or 15 mM succinate as the carbon source. *E. coli* B (16) cultures were grown in M9 minimal medium supplemented with 20 mM pyruvate. Five mL cultures were grown in triplicate for 16 hours and then diluted 1:10,000 into 50 mL fresh medium in 250 mL side-arm flasks with shaking at 200 rpm. Cell density was assessed

using a Klett colorimeter. Cells were harvested in exponential phase and early stationary phase. Exponential phase was determined by calculating growth rate (Klett reading \approx 30-35). Early stationary phase was reached when Klett colorimeter readings were steady for 3 hours.

Cell lysis and enumeration

When cells had reached the desired density, 300 μ L were harvested by centrifugation at 5000 x g for 4 minutes. Following removal of the supernatant, cells were resuspended in 1 mL Cell Lysis Solution (from Qiagen Puregene Yeast/Bacteria Kit B) and incubated at 80°C for 5 minutes followed by a 10 minute incubation at room temperature. Following lysis, the lysate was observed using phase contrast on a Petroff-Hausser counting chamber under 1000X magnification to ensure >95% lysis prior to downstream use.

To quantitate cell density, cells were diluted 1:50 for stationary phase harvests and 1:20 for exponential phase harvests and counted in a Petroff-Hausser counting chamber. In addition, cultures were plated to determine CFUs, which is expected to give a slightly lower number than the direct cell counts because not every cell in the culture will be viable. However, cells that are not viable will lyse and release DNA in the Cell Lysis Solution. Cell lysis efficiency was calculated by subtracting the number of unlysed cells visible in the cell lysate from the total cell counts resulting in the number of total lysed cells and the total lysed cell number was then divided by the total cell count to determine lysis efficiency.

Quantitative PCR

qPCR was performed to determine the DNA content of lysed cells. Chromosomal copy number of regions near the origin and terminus were analyzed in both *A. baylyi* ADP1 and *E. coli* B. The input DNA for the standard curve was a ~1 kb product generated by the primers listed in Table 3.1 (std curve). DNA quantitation for both the standard curve and the lysate was performed with primers that amplified a shorter fragment nested within the ~1 kb fragment (Table 3.1, copy no.). The standard curve was generated in Cell Lysis Solution that had been diluted 10-fold in Millipore H₂O. The lysate was also diluted 10-fold in Millipore H₂O for analysis, and 2.5 µL was used as template in each reaction. Five-fold dilutions of PCR product in Cell Lysis Solution were used to generate the standard curve (6.25×10^{-2} to 2.0×10^{-5} ng) for the origin and terminus qPCR. qPCR was performed on an Applied Biosystems StepOnePlus instrument using Applied Biosystems SYBRGreen Master Mix. Any wells with a variation of more than 0.5 C_T were excluded from further analysis.

Genomic content calculations

DNA content per reaction (in molecules of DNA) was calculated taking into account the molecular weight of the fragment size of the ~1 kb fragment (Table 3.1) as determined from DNA sequence using the calculator available from EnCor Biotechnology Inc. (<http://www.encorbio.com/protocols/Nuc-MW.htm>). A basic formula for conversion of grams of DNA to genomic content for each qPCR reaction follows.

$$\frac{\text{grams DNA}}{\text{molecular weight of std curve PCR product}} \times 6.02 \times 10^{23} = \text{genomic content in reaction}$$

With the value for genomic content in molecules per reaction, genomic content per cell can be calculated using the number of cells lysed per reaction (resulting in molecules per cell).

Results

Ploidy levels for *A. baylyi* ADP1 were determined using the qPCR method developed by Breuert *et al.* (13). Genome copy numbers for *A. baylyi* ADP1 grown with 20 mM pyruvate as the carbon source in early stationary and exponential phase are reported in Tables 3.2 and 3.3, respectively. Cells were harvested at the time of the final reading during the growth curves (Figure 3.1). *A. baylyi* contains 2.2 origins and 1.2 termini per cell in early stationary phase. Additionally, the number of genomes per cell does not vary significantly between exponential and stationary phase (see average origin and terminus values in Tables 3.2 and 3.3). For comparison of chromosomal copy number in cells with a faster growth rate, the genome copy number in exponential phase for cells grown on succinate was also determined. There was no change in the chromosomal copy number for faster growing cells (Table 3.4).

To ensure technique efficacy, the ploidy level of *E. coli* B was also determined using previously published primers (17) in early stationary phase (Figure 3.2 and Table 3.5). The calculated ploidy level is slightly higher than reported in previous studies, 4.6 origins and 5.1 termini per cell compared to 6.8 origins and 1.7 termini per cell, potentially resulting from decreased primer efficiency in different laboratory settings (13, 18). Another source of variation may come from the different carbon sources used for growth in the different studies.

Discussion

The ploidy level of *A. baylyi* ADP1 (2.2 origins and 1.2 termini per cell in stationary phase cultures) is similar to that of *E. coli* but distinctly lower than that of *P. putida*. It has been shown that *E. coli* can have multiple origins per cell when replication initiation is occurring more quickly than cell division (19, 20). Replication of the *E. coli* genome (4.6 Mb) occurs in approximately 40 minutes; while the replication rate of *A. baylyi* ADP1 is unknown, it is assumed to be similar. A slightly shorter replication time may exist in *A. baylyi* ADP1 given the smaller genome size (3.6 Mb). As the most rapid doubling time achieved by *A. baylyi* ADP1 is near 40 minutes, it is unlikely that replication is initiated more than once per cell division cycle (21). This assumption holds true when the values for origins and termini per cell are compared between early stationary and exponential phase growth for *A. baylyi* ADP1. If replication is initiated more than once per cell division, the number of origins per cell would differ between the exponential and stationary phase cells.

The study of ploidy levels in *A. baylyi* ADP1 was initiated to clarify results from a previous study of IS1236-mediated duplication events. Atypical transposition events generated duplications allowing growth on benzoate in strains lacking the transcriptional regulators necessary to degrade benzoate as the sole carbon source. Multiple copies of the genome were a prerequisite for one explanation that could account for the duplication formation events. In this scenario, transposition of IS1236 could occur on one copy of the chromosome. The chromosomal copy containing the newly transposed element could recombine with a second copy of the chromosome in the cell that did not undergo a transposition event. Such recombination would generate the chromosomal arrangements

found (14). However, as multiple copies of the genome are not found in *A. baylyi* ADP1 cells, it is unlikely that transposition occurred prior to recombination on a different chromosome within the same cell.

Multiple genome copies per cell have been found in various Bacteria and Archaea. The presence of more than one copy of a chromosome per bacterial cell is intriguing. In the large bacterium *Epulopiscium* spp. which can be up to 600 μm in length, multiple genomic copies may be a way to circumvent transport times for proteins through the cytoplasm (5). When gene products are produced in close proximity to where they are needed inside the cell, an increased ability to respond quickly and appropriately to stimuli is achieved. Additionally, multiple gene copies can lead to new gene functions; beneficial mutations can accrue in the extra gene copies, leading to new gene products. Multiple gene copies also can protect cells from death due to the acquisition of a deleterious mutation in an essential gene.

The canonical model of a single copy of the chromosome per bacterial cell may not be as prevalent as previously thought. Multiple chromosomes in a cell can be a form of rudimentary regulation, as expression from more gene copies leads to higher gene product levels. Additional genome copies can lead to increased viability through recovery from deleterious mutations or through the evolution of new gene functions. More than one copy of the chromosome per cell may provide an evolutionary advantage for microorganisms.

An awareness of polyploidy in bacteria is vital to comparing similarities between model organisms and less tractable pathogens. *A. baylyi* and *P. putida* are used for comparative purposes for the study of the pathogens *Acinetobacter baumannii* and

Pseudomonas aeruginosa, respectively. Knowing the ploidy level in the model organisms can lead to better methods for combatting the virulence of these pathogens. Additionally, various ploidy levels may contribute to varying mechanisms for antibiotic resistance and virulence in these pathogens. A clearer understanding of the relationship between genomic copy number in model organisms and pathogens may have important medical implications.

References

1. **Skarstad K, Steen HB, Boye E.** 1983. Cell cycle parameters of slowly growing *Escherichia coli* B/r studied by flow cytometry. *J. Bacteriol.* **154**:656-662.
2. **Webb CD, Graumann PL, Kahana JA, Teleman AA, Silver PA, Losick R.** 1998. Use of time-lapse microscopy to visualize rapid movement of the replication origin region of the chromosome during the cell cycle in *Bacillus subtilis*. *Mol. Micro.* **28**:883-892.
3. **Hansen MT.** 1978. Multiplicity of genome equivalents in the radiation-resistant bacterium *Micrococcus radiodurans*. *J. Bacteriol.* **134**:71-75.
4. **Kitten T, Barbour AG.** 1992. The relapsing fever agent *Borrelia hermsii* has multiple copies of its chromosome and linear plasmids. *Genetics* **132**:311-324.
5. **Mendell JE, Clements KD, Choat JH, Angert ER.** 2008. Extreme polyploidy in a large bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **105**:6730-6734.
6. **Skarstad K, Boye E, Steen HB.** 1986. Timing of initiation of chromosome replication in individual *Escherichia coli* cells. *EMBO J.* **5**:1711-1717.
7. **Maldonado R, Jimenez J, Casadesus J.** 1994. Changes of ploidy during the *Azotobacter vinelandii* growth cycle. *J. Bacteriol.* **176**:3911-3919.
8. **Maldonado R, Garzon A, Dean DR, Casadesus J.** 1992. Gene dosage analysis in *Azotobacter vinelandii*. *Genetics* **132**:869-878.
9. **Tobiason DM, Seifert HS.** 2006. The obligate human pathogen, *Neisseria gonorrhoeae*, is polyploid. *PLoS Biol.* **4**:e185.
10. **Tobiason DM, Seifert HS.** 2010. Genomic content of *Neisseria* species. *J. Bacteriol.* **192**:2160-2168.

11. **Komaki K, Ishikawa H.** 2000. Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. *Insect Biochem. Mol. Biol.* **30**:253-258.
12. **Postgate JR, Kent HM, Robson RL, Chesshyre JA.** 1984. The genomes of *Desulfovibrio gigas* and *D. vulgaris*. *J. Gen. Microbiol.* **130**:1597-1601.
13. **Breuert S, Allers T, Spohn G, Soppa J.** 2006. Regulated polyploidy in halophilic archaea. *PLoS One* **1**:e92.
14. **Cuff LE, Elliott KT, Seaton SC, Ishaq MK, Laniohan NS, Karls AC, Neidle EL.** 2012. Analysis of IS1236-mediated gene amplification events in *Acinetobacter baylyi* ADP1. *J. Bacteriol.* **194**:4395-4405.
15. **Sambrook J, Fritsch EF, Maniatis T.** 1987. *Molecular cloning : a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Daegelen P, Studier FW, Lenski RE, Cure S, Kim JF.** 2009. Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3). *J. Mol. Biol.* **394**:634-643.
17. **Pecoraro V, Zerulla K, Lange C, Soppa J.** 2011. Quantification of ploidy in proteobacteria revealed the existence of monoploid, (mero-)oligoploid and polyploid species. *PLoS One* **6**:e16392.
18. **Bremer H, Dennis P.** 1996. Modulation of chemical composition and other parameters of the cell growth rate. *In* Neidhardt F (ed.), *Escherichia coli* and *Salmonella*. ASM Press, Washington DC.
19. **Cooper S, Helmstetter CE.** 1968. Chromosome replication and the division cycle of *Escherichia coli* B/r. *J. Mol. Biol.* **31**:519-540.

20. **Fossum S, Crooke E, Skarstad K.** 2007. Organization of sister origins and replisomes during multifork DNA replication in *Escherichia coli*. EMBO J. **26**:4514-4522.
21. **McCarthy D, Minner C, Bernstein H, Bernstein C.** 1976. DNA elongation rates and growing point distributions of wild-type phage T4 and a DNA-delay amber mutant. J. Mol. Biol. **106**:963-981.

Table 3.1. Primers used in this study.

Primer Name	Sequence	Purpose	Fragment Size (bp)	Molecular weight (Daltons)
50115_UP	GACTGTGATTTAGAGGCGAGCG	<i>A. baylyi</i> std curve	1293	399431
51407_DN	CCACTGCGTTCTATCGGCACT			
50710_UP	CCTTTCAGCGACACTCAGGAT	<i>A. baylyi</i> copy no.	61	N/A ^a
50770_DN	TCCAGATCGCCATCAACAATT			
1790148_UP	CTGAAGATGTTGAAGACTTGGC	<i>A. baylyi</i> std curve	1289	398802
1791436_DN	GACTTCATTGTCTAGGGGTGTG			
1790750_UP	CGACGTGAGGCAGATGTCATT	<i>A. baylyi</i> copy no.	65	N/A ^a
1790814_DN	TTGGTGGCGTGGTAATTGC			
1550252_UPEc	GACAACGTAGGCTTTGTTCATGCCGG	<i>E. coli</i> std curve	875	269673
1551126_DNEc	TTGATGCTGTCCGTGCAGCGGTCGT			
1550623_UPEc	GGCAAACCTCGCGCAGGCTGACATTAA	<i>E. coli</i> copy no.	258	N/A ^a
1550880_DNEc	GCCGCATGGTGATTGATTTCCGTCAC			
3922921_UPEc	GACTGACGCCAAATTGTTTCGCCAGT	<i>E. coli</i> std curve	1052	326191
3925972_DNEc	CAGTTGATGCTTCAGCGTGTCGGCAT			
3925392_UPEc	CACTGGCGAAGTGGAAACGTCAGAC	<i>E. coli</i> copy no.	227	N/A ^a
3925618_DNEc	GGTTGCTTTAATTCGCCAGATCG			

^aMolecular weight was not used in chromosomal copy number calculations

Table 3.2. Origin and termini copy numbers in pyruvate-grown early stationary phase *Acinetobacter baylyi* ADP1.

Culture number	Doubling time (min)	Cell density (cells/mL)	Lysis efficiency (%)	No. origins per cell	Average origins per cell \pm s.d.	No. termini per cell	Average termini per cell \pm s.d.
1	41	4.7×10^9	98.0	2.3		1.2	
2	52	3.8×10^9	97.5	2.0	2.2 ± 0.2	1.1	1.2 ± 0.1
3	46	3.6×10^9	95.8	2.4		1.4	

Table 3.3. Origin and termini copy numbers in pyruvate-grown exponential phase *Acinetobacter baylyi* ADP1.

Culture number	Doubling time (min)	Cell density (cells/mL)	Lysis efficiency (%)	No. origins per cell	Average origins per cell \pm s.d.	No. termini per cell	Average termini per cell \pm s.d.
1	67	7.9×10^8	98.4	1.9		1.0	
2	59	8.3×10^8	97.0	1.8	1.9 ± 0.1	1.0	1.0 ± 0.04
3	58	6.9×10^8	98.2	2.0		1.1	

Table 3.4. Origin and termini copy numbers in succinate-grown exponential phase *Acinetobacter baylyi* ADP1.

Culture number	Doubling time (min)	Cell density (cells/mL)	Lysis efficiency (%)	No. origins per cell	Average origins per cell \pm s.d.	No. termini per cell	Average termini per cell \pm s.d.
1	42	8.4×10^8	95.5	1.5		1.0	
2	33	6.9×10^8	100	1.9	1.7 ± 0.2	1.3	1.1 ± 0.2
3	34	8.4×10^8	98.5	1.6		1.1	

Table 3.5. Origin and termini copy numbers in pyruvate-grown *Escherichia coli* B.

Culture number	Doubling time (min)	Cell density (cells/mL)	Lysis efficiency (%)	No. origins per cell	Average origins per cell \pm s.d.	No. termini per cell	Average termini per cell \pm s.d.
1	121	2.4×10^9	98.7	4.8		5.2	
2	112	2.5×10^9	96.3	4.6	4.6 ± 0.2	5.1	5.1 ± 0.07
3	106	2.5×10^9	98.7	4.5		5.1	
Ref (17)	25	7.8×10^8	96.4		6.8 ± 1.6		1.7 ± 0.4

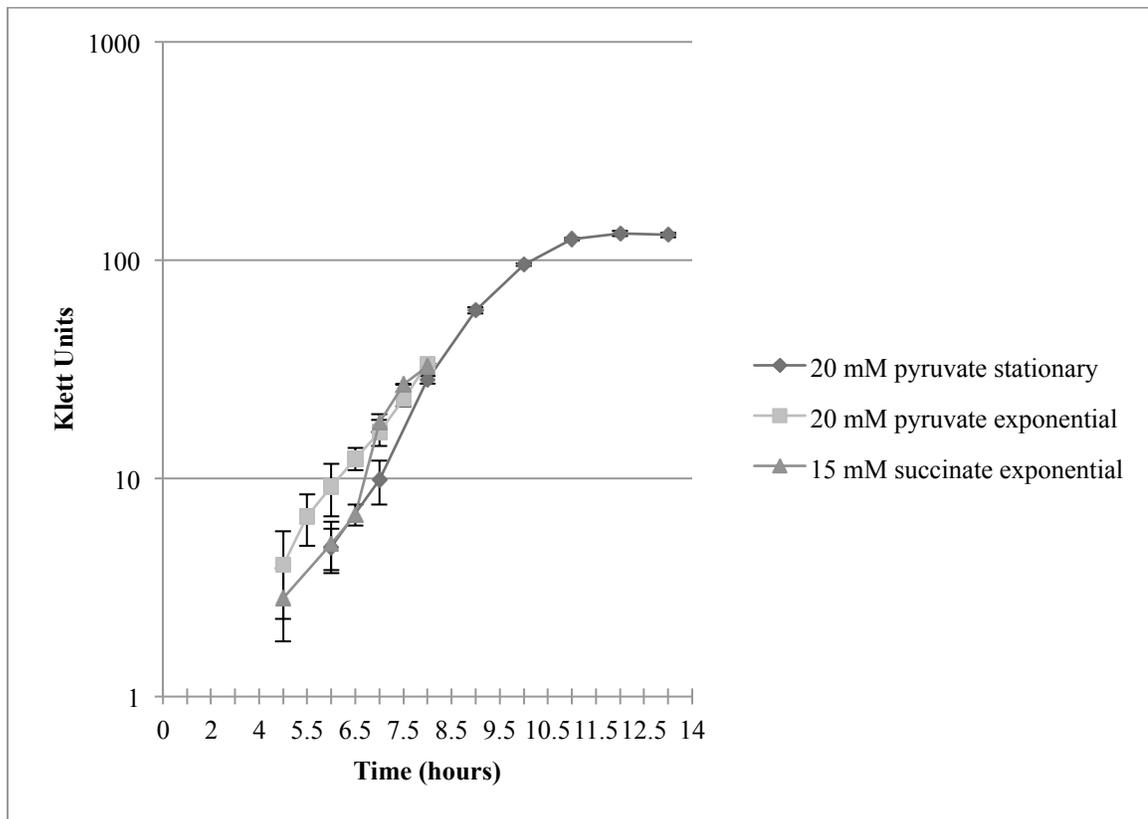


Figure 3.1. Growth curves of *A. baylyi* ADP1 cultures for chromosomal copy number analysis. Cells were grown in M9 medium with the carbon source indicated and harvested at the final time point. Doubling time was calculated from the three time points with the most rapid exponential growth. One Klett Unit is equivalent to approximately 8.5×10^5 cells.

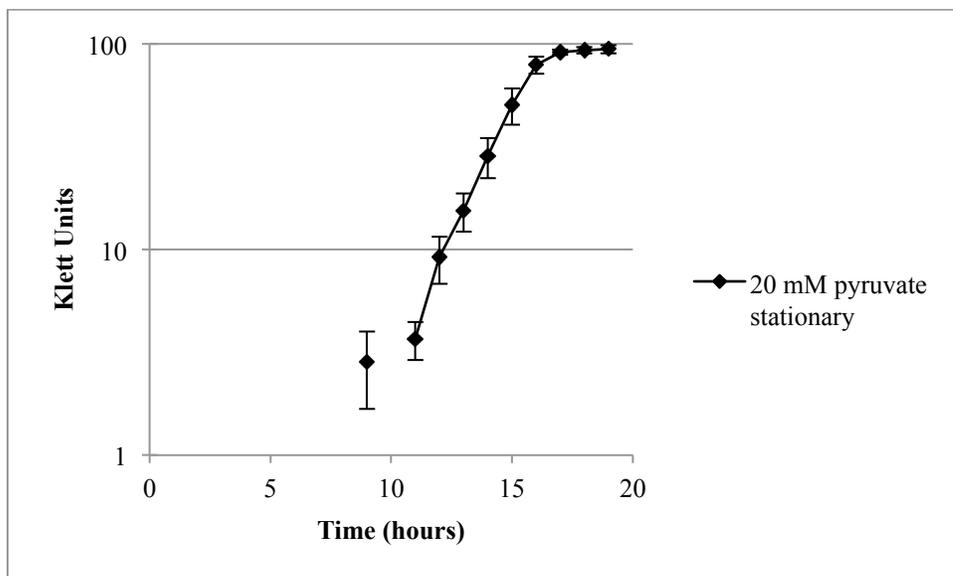


Figure 3.2. Growth curve of *E. coli* B cultures for chromosomal copy number analysis. Cells were grown in M9 medium with pyruvate and harvested at the final time point. Doubling time was calculated from the three time points with the most rapid exponential growth.

CHAPTER 4

CONCLUSIONS

This dissertation describes studies of chromosomal rearrangements and chromosomal copy number in *Acinetobacter baylyi* ADP1. The results have broad significance for prokaryotic molecular biology and physiology, medicine, and biotechnology.

Broad implications of gene amplification in bacteria

Gene duplication and amplification events occur in all organisms at surprisingly high frequencies (1-4). New studies show that roughly one out of every five bacterial cells in a population contains a duplication within a genomic region, despite the absence of selection (5). These duplication events occur at a rate that is several orders of magnitude higher than spontaneous point mutations (6-8). Despite the prevalence of this phenomenon, little is known about the specific mechanisms for spontaneous duplication events. A deeper understanding of contributing factors leading to gene amplification can inform better methods for combating the effects of gene amplification.

One well-documented result of gene amplification is increased levels of antibiotic resistance. Instances of direct modification include changes of drug availability via an increased number of transporters or through structural changes of β -lactamases (9, 10). Other amplifications generate increased levels of the antibiotic target or generate a bypass of cellular metabolic pathways (11, 12). We have recently reviewed the contribution of gene amplification to the acquisition of antibiotic resistance in bacteria (13). Gene

amplifications can have significant medical importance, as many bacteria are becoming resistant to the available and commonly used antibiotics.

Information about genomic plasticity, including gene duplication, is rapidly accumulating because of the increasing pervasiveness and decreasing cost of full genome sequencing. The sequencing of entire genomes of diverse bacteria will elucidate the propensity of specific genes or regions of the chromosome to be found in more than one copy. Determining the presence of multiple copies of a gene on the chromosome can lead to a better understanding of the importance of repeated DNA segments. In addition to full genome sequencing, several other methods to study chromosomal content and the copy number of specific genes are available and detailed in Chapter 1.

In Chapter 2, studies of genetic recombination in *A. baylyi* ADP1 revealed a novel, atypical transposition event mediating duplication formation. The duplications were found to originate at one end of the inverted repeat of IS1236, the locus where DNA nicking occurs to initiate transposition (14, 15). However, a full transposition event was not completed, as there was no terminal insertion sequence at the endpoint of the final amplicon (16). These are the first documented instances of IS3-family members generating duplications in this fashion.

The study of occurrences of gene duplication and amplification will shed light on the driving forces of adaptation. A recent model, termed the innovation-amplification-divergence model, enables the study of evolution in real time (17). This model posits that a mutation generates a new gene function with a low level of activity for an existing gene. Amplification of this new gene occurs and further adaptation enables enzymatic specialization. Selection for a beneficial mutation maintains the new gene functions

while less beneficial mutations are lost. In conjunction with the genetic tools available in *A. baylyi* ADP1, the innovation-amplification-divergence model can be used to study enzyme evolution in detail. The systematic analysis and thorough understanding of duplication formation events in *A. baylyi* ADP1 described here will enable a deeper study of microbial evolution. Understanding the precise genetic basis of chromosomal rearrangements and their evolutionary effects requires knowledge of the total genetic content within individual cells.

Impact of multiple chromosomal copies within a single cell

Bacteria and other microorganisms can have multiple genomic copies within one cell. This idea contradicts the long-held notion that bacteria only had one copy of the genome per cell, with a few exceptions. However, as technology allows for a more rapid and precise quantitation of DNA within a cell, greater numbers of microorganisms are found to have multiple copies of the chromosome per cell, even under slow growth conditions (18).

Chapter 3 details the study of the ploidy levels in *A. baylyi* ADP1. qPCR analysis revealed a single copy of the chromosome during both exponential and early stationary growth phases. This result indicates that *A. baylyi* ADP1 is more similar to *E. coli* than to the previously characterized *Pseudomonas* species with respect to chromosomal copy number.

Several potential mechanisms for duplication formation were suggested in Chapter 2. The characterization of the chromosomes of the amplification mutants ruled out the possibility that transposition occurred and was followed by intramolecular recombination between different IS copies. The ploidy level studies of *A. baylyi* ADP1

were instigated to evaluate the feasibility of other mechanisms involving intermolecular recombination. The finding that ADP1 has a single copy of the chromosome per cell suggests that interchromosomal recombination following a transposition event is unlikely because the opportunities for such interactions are limited to the period following replication before cell division. Nevertheless, the possibility of transposition and recombination between IS elements on different chromosomes cannot be eliminated. During replication, transposition of IS/236 could occur on one arm of the replicated chromosome. Homologous recombination between this newly transposed element and DNA on the opposite arm that had not undergone a transposition event would generate the results found in Chapter 2. This possibility invokes a mechanism of transposition and homologous recombination that only occurs between different chromosomal copies. Such exclusivity of interchromosomal recombination events might imply a replication-dependent mechanism.

Other potential mechanisms could explain the observed results. The data reveal recombination at the exact point in the DNA where the transposase cleaves the DNA. The generation of a single-stranded end in the DNA by the IS/236 transposase could lead to intramolecular recombination that does not involve a completed transposition event. With strand invasion elsewhere in the chromosome, resolution of the broken DNA could generate a duplication of the intervening chromosomal region. Further clarification of the mechanism(s) by which the nicked DNA could be repaired requires additional experimental investigation. DNA repair and recombination have not been well studied in *A. baylyi* ADP1, and genomic analysis suggests that there are some significant

differences between the set of known recombination genes in this strain and those from well-characterized bacteria such as *E. coli*.

Although studies of chromosomal copy number in ADP1 did not definitively reveal the mechanism of IS-mediated duplication, these investigations provide important data and a framework for evaluating recombination models. In general, characterizing the ploidy level of bacteria will lead to a deeper understanding of the advantages of multiple chromosomal copies and also will contribute to our expanding appreciation of genetic flexibility and variability in microorganisms. As illustrated by the research described here, the investigation of chromosomal plasticity with respect to both content and copy number is readily achievable in *A. baylyi* ADP1.

References

1. **Myllykangas S, Knuutila S.** 2006. Manifestation, mechanisms and mysteries of gene amplifications. *Cancer Lett.* **232**:79-89.
2. **Romero D, Palacios R.** 1997. Gene amplification and genomic plasticity in prokaryotes. *Annu. Rev. Genet.* **31**:91-111.
3. **Roth JR, Andersson DI.** 2004. Amplification-mutagenesis--how growth under selection contributes to the origin of genetic diversity and explains the phenomenon of adaptive mutation. *Res. Microbiol.* **155**:342-351.
4. **Craven SH, Neidle EL.** 2007. Double trouble: medical implications of genetic duplication and amplification in bacteria. *Future Microbiol.* **2**:309-321.
5. **Sun S, Ke R, Hughes D, Nilsson M, Andersson DI.** 2012. Genome-wide detection of spontaneous chromosomal rearrangements in bacteria. *PloS One* **7**:e42639.
6. **Andersson DI, Hughes D.** 2009. Gene amplification and adaptive evolution in bacteria. *Annu. Rev. Genet.* **43**:167-195.
7. **Anderson P, Roth J.** 1981. Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (*rrn*) cistrons. *Proc. Natl. Acad. Sci. U. S. A.* **78**:3113-3117.
8. **Seaton SC, Elliott KT, Cuff LE, Laniohan NS, Patel PR, Neidle EL.** 2012. Genome-wide selection for increased copy number in *Acinetobacter baylyi* ADP1: locus and context-dependent variation in gene amplification. *Mol. Microbiol.* **83**:520-535.

9. **Nicoloff H, Perreten V, McMurry LM, Levy SB.** 2006. Role for tandem duplication and lon protease in AcrAB-TolC-dependent multiple antibiotic resistance (Mar) in an *Escherichia coli* mutant without mutations in *marRAB* or *acrRAB*. *J. Bacteriol.* **188**:4413-4423.
10. **Sun S, Berg OG, Roth JR, Andersson DI.** 2009. Contribution of gene amplification to evolution of increased antibiotic resistance in *Salmonella typhimurium*. *Genetics* **182**:1183-1195.
11. **Nilsson AI, Zorzet A, Kanth A, Dahlstrom S, Berg OG, Andersson DI.** 2006. Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. *Proc. Natl. Acad. Sci. U. S. A.* **103**:6976-6981.
12. **Brochet M, Couve E, Zouine M, Poyart C, Glaser P.** 2008. A naturally occurring gene amplification leading to sulfonamide and trimethoprim resistance in *Streptococcus agalactiae*. *J. Bacteriol.* **190**:672-680.
13. **Elliott KT, Cuff LE, Neidle EL.** 2013. Copy number change: evolving views on gene amplification. *Future Microbiol.*
14. **Nagy Z, Chandler M.** 2004. Regulation of transposition in bacteria. *Res. Microbiol.* **155**:387-398.
15. **Loot C, Turlan C, Rousseau P, Ton-Hoang B, Chandler M.** 2002. A target specificity switch in *IS911* transposition: the role of the OrfA protein. *EMBO J.* **21**:4172-4182.
16. **Cuff LE, Elliott KT, Seaton SC, Ishaq MK, Laniohan NS, Karls AC, Neidle EL.** 2012. Analysis of *IS1236*-mediated gene amplification events in *Acinetobacter baylyi* ADP1. *J. Bacteriol.* **194**:4395-4405.

17. **Nasvall J, Sun L, Roth JR, Andersson DI.** 2012. Real-time evolution of new genes by innovation, amplification, and divergence. *Science* **338**:384-387.
18. **Pecoraro V, Zerulla K, Lange C, Soppa J.** 2011. Quantification of ploidy in proteobacteria revealed the existence of monoploid, (mero-)oligoploid and polyploid species. *PloS One* **6**:e16392.

APPENDIX A

GENOME-WIDE SELECTION FOR INCREASED COPY NUMBER IN
ACINETOBACTER BAYLYI ADP1: LOCUS AND CONTEXT-DEPENDENT
VARIATION IN GENE AMPLIFICATION³

³ Seaton, S.C., Elliott, K.T., Cuff, L.E., Laniohan, N.S., Patel, P.R., Neidle, E.L. 2012. *Molecular Microbiology*. 83(3): 520-535. Reprinted here with permission of publisher.

S.C.S, K.T.E., L.E.C., and E.L.N. designed experiments, S.C.S, K.T.E., L.E.C., N.S.L., and P.R.P. performed experiments, S.C.S., K.T.E., L.E.C., and E.L.N. authored manuscript.

Summary

Renewed interest in gene amplification stems from its importance in evolution and medical problems. However, amplified DNA segments (amplicons) are not fully characterized in any organism. Here we report a novel *Acinetobacter baylyi* system for genome-wide studies. Amplification mutants that consume aromatic compounds were selected under conditions requiring high-level expression from three promoters in a set of chromosomal genes. Tools were developed to relocate these catabolic genes to any non-essential chromosomal position, and 49 amplification mutants from five loci were characterized. Amplicon size (18-271 kbp) and copy number (2-105) indicated that 30% of mutants carried more than 1 Mbp of amplified DNA. Amplification features depended on genomic position. For example, amplicons isolated at one locus were similarly sized but displayed variable copy number, whereas those from another locus were differently sized but had comparable copy number. Additionally, the importance of sequence context was highlighted in one region where amplicons differed depending on the presence of a promoter mutation in the strain from which they were selected. DNA sequences at amplicon boundaries in 19 mutants reflected illegitimate recombination. Furthermore, steady-state duplication frequencies measured under non-selective conditions (10^{-4} to 10^{-5}) confirmed that spontaneous gene duplication is a major source of genetic variation.

Introduction

An increase in gene dosage through genetic duplication and amplification (GDA) is a common process that allows rapid adaptation to variable, limiting, or extreme conditions (Reams & Neidle, 2004b, Andersson & Hughes, 2009, Andersson, 2011). GDA not only plays a key role in evolution and chromosomal organization, but it also underlies medical issues ranging from chemotherapeutic resistance and cancer to developmental, cognitive, and autoimmune disorders (Albertson, 2006, Craven & Neidle, 2007, Stankiewicz & Lupski, 2010, Conrad & Antonarakis, 2007). In this report, we describe the development of *Acinetobacter baylyi* ADP1 as a bacterial model system for genome-wide studies of GDA.

Spontaneous duplications occur at rates that are orders of magnitude higher than those for point mutation, and thus represent a major source of genetic variability (Roth *et al.*, 1996). However, this type of genetic plasticity is difficult to study because GDA is readily reversible and often transitory. This dynamic nature motivates recent mathematical models of evolutionary processes to incorporate transient intermediate states of amplified DNA (Innan & Kondrashov, 2010, Pettersson *et al.*, 2009). In *Escherichia coli* and *Salmonella enterica* serovar Typhimurium, GDA has emerged as a central theme in efforts to assess mutation rates during bacterial adaptation to growth-limiting conditions (Andersson, 2011, Pranting & Andersson, 2011, Reams *et al.*, 2010, Roth, 2011). Such studies demonstrate how novel phenotypes emerge rapidly due to population heterogeneity since many cells (approximately 10%) typically carry a duplication of some region of the chromosome.

In *E. coli* and *S. enterica* studies of gene amplification, not all selected mutants carry amplified DNA. Some point mutants derive from ancestral cells with amplified DNA even though the amplified segment (amplicon) may not be retained or characterized (Roth *et al.*, 2006). Such is the case with methods to assess the adaptation of a partially defective *lac* mutation. During extensive investigations of *lac* reversion, many genome duplications have been characterized (e.g. (Slecht *et al.*, 2002). Yet, as detailed in a recent paper, idiosyncratic features of different methodology can complicate the interpretation of experimental results (Quinones-Soto & Roth, 2011). It is difficult to compare similar studies in *E. coli* and *S. enterica* due to variation in the specific allele that is investigated and whether it is located on the chromosome or on a conjugative plasmid (such as the F episome). Thus, it is important to study GDA systematically and to capture and characterize amplified DNA.

Most amplification studies focus on a few random events or a single locus that is amenable to characterization (Nicoloff *et al.*, 2006, Reams & Neidle, 2003, Paulander *et al.*, 2010). However, DNA sequence and local genetic architecture can influence processes that generate duplications, such as recombination, transposition, and replication slippage. Furthermore, the entire genetic content and context of an amplified segment can impact its fitness, which makes genetic neighborhood critical (Pettersson *et al.*, 2009, Reams *et al.*, 2010). Therefore, genome-wide studies are needed to evaluate context-dependent effects.

Toward this aim, we expanded a method to select spontaneous amplification mutants of the soil bacterium *A. baylyi* from a parent strain that lacks two transcriptional activators (Reams & Neidle, 2003, Reams & Neidle, 2004a). Consumption of benzoate

by this strain requires increased expression of *catA* and the *catBCIJFD* operon (Fig. A.1). Mutants selected by growth on benzoate (Ben⁺ phenotype) carry multiple copies of the weakly expressed *cat* genes in tandem head-to-tail chromosomal arrays (Reams & Neidle, 2003). There is a need for increased expression from multiple promoters and also a requirement to prevent toxic metabolites from accumulating. Because of these demands, all resulting mutants carry amplified DNA. According to our model, the first step of the amplification process involves recombination between DNA downstream and upstream of the *cat* genes to create a duplication (Fig. A.1). Two amplicon copies then serve as long direct repeats that may increase (or decrease) in copy number via homologous recombination.

As described here, we altered the selection conditions used previously to require that a larger chromosomal region be amplified in the region where the *cat* genes normally reside. In this case, recombination events occurred between DNA in regions not previously explored. Additionally, different chromosomal regions were studied after deletion of the *cat*-gene cluster from its native locus and relocation by genetic engineering. The *cat*-genes then served as a selection cassette to isolate duplication sites that reflect interactions between diverse flanking DNA in different genomic contexts. The size and copy number of independently isolated amplicons were characterized from five chromosomal positions, and in some cases the exact DNA sequences at duplication junctions were determined. In addition, we calculated the spontaneous duplication frequency at each locus.

Results

Relocation of the cat genes

To isolate new amplification mutants at different chromosomal loci, the entire *cat*-gene region was moved as a cassette. First, the *cat* genes were deleted from their native chromosomal position (wt locus, Fig. A.2) via allelic replacement, as described in Experimental procedures. The resultant strain, ACN825 (all strains and plasmids are shown in Table A.1), was unable to grow on catechol or other carbon sources that are degraded via catechol as an intermediate, such as benzoate or anthranilate.

To restore growth on benzoate, a functional *cat*-gene cassette was introduced at a new location. In ACN825, the conversion of benzoate to catechol is controlled by a *benA* promoter mutation that eliminates the need for an activator [Fig. A.1 and (Collier *et al.*, 1998)]. Further degradation of catechol requires expression of *catA* and the *catBCIJFD* operon. These *cat* genes were targeted to the ACN825 chromosome together with *catM*, encoding their transcriptional regulator. Integration of these genes at specific positions was accomplished by allelic replacement with a relocation plasmid developed for this study (Fig. A.S1). Strains in which the plasmid-borne *cat*-genes replaced targeted chromosomal DNA were selected by growth on benzoate. Their genotypes were confirmed by PCR, pulsed-field gel electrophoresis (PFGE) and/or hybridization methods. Using this approach, the *cat* genes were inserted in each of four new locations disrupting genes known to be dispensable for growth on defined medium [loci *i-iv*, Fig. A.2A; (de Berardinis *et al.*, 2008)]. The disrupted genes (ACIAD0982, ACIAD2822, ACIAD2969 and ACIAD3593, respectively) are predicted to encode membrane or

transport proteins (from Cluster of Orthogonal Groups 2116P, 2814G, 2271G, and 1113E) (Tatusov *et al.*, 2000).

Disruption of catM sets the stage to select new amplification mutants

After relocation, *catM* was inactivated to lower expression of the adjacent *cat* genes. Each of the four strains in which the *cat* genes had been relocated was transformed with a linearized version of plasmid pBAC6, which carries an insertionally-inactivated *catM*, *catM::ΩK5541* (Ezezika *et al.*, 2006). Transformants were selected using the drug resistance encoded by the mutated allele, and genotypic analysis confirmed allelic replacement. Inactivation of *catM* at locus *i*, *ii*, *iii*, and *iv* (Fig. A.2) in the resulting strains (ACN1039, ACN1050, ACN1061, and ACN1024; Table A.1) caused a Ben⁻ phenotype, defined by an inability to form visible colonies on solid benzoate medium during seven days of incubation at 30°C. Under these conditions, wild-type colonies form in 1-2 days.

New amplification mutants from different chromosomal regions

ACN1024, with the *cat* genes in locus *iv*, was used as a parent strain to select derivatives capable of growth on benzoate (Ben⁺). Small colonies arose on benzoate medium at a frequency of approximately one colony per 10⁸ cells within 20 days of incubation. To test for amplification, genomic DNA was digested with the restriction enzyme NotI, and analyzed by PFGE. The parent strain (ACN1024) has one extra genomic NotI site compared to the wild type due to its disrupted-*catM* allele (Fig. A.2). Because this site lies within each *cat*-gene copy, NotI-digested DNA of an amplification mutant should contain a fragment whose size is equal to that of its amplicon (Fig. A.2C). These fragments would not result from similar digestion of parent-strain DNA.

Consistent with the Ben⁺ derivatives carrying *cat*-gene amplicons, each mutant contained one NotI-generated fragment that is not found in the parent (marked by an asterisk in Fig. A.2B). Southern hybridization confirmed that *catA* resides on these fragments as well as on fragment A². Also consistent with these new fragments representing amplified DNA, they appear to be present in multiple copies compared to other genomic regions based on the relative band intensities (Fig. A.2B). To date, there are ten confirmed ACN1024-derived amplification mutants (Tables A.1 and A.S1; some shown in Fig. A.2B). Gene amplification mutants were similarly isolated and characterized from parent strains in which the *cat* genes were relocated to positions *i*, *ii*, and *iii* (Fig. A.2, Tables A.1 and A.S1).

New amplification mutants involving the wild-type cat locus

To complement the analysis of diverse chromosomal regions, we expanded our previous methods to assess novel GDA events involving the native *cat*-gene region. The parent strains described so far exploit a point mutation in *benA* to control the conversion of benzoate to catechol (Fig. A.1). Our initial studies indicated that this *benA* promoter mutation is not tolerated in multiple copy, and we presumed that the problem was due to the conversion of benzoate to toxic levels of catechol by high levels of the BenABCD enzymes. In studies of the native *cat* locus, the upstream boundary of all amplicons from Ben⁺ mutants resided within an approximately 5-kbp region between the *benA* and *catA* promoters (Reams & Neidle, 2004a, Reams & Neidle, 2003). Thus, the types of recombination events that were previously selected may have been influenced by this constraint. To assess new GDA events, we altered the carbon source selection and/or the genetic background of the parent strain used to isolate mutants.

As an alternative to benzoate, we selected amplification mutants by growth of the initial parent strain (ACN293) on anthranilate as the sole carbon source. This substrate is converted to catechol by the products of *antABC*, genes located distantly on the chromosome from the *cat* genes (Bundy *et al.*, 1998, Eby *et al.*, 2001). Further consumption of catechol requires high expression of the *cat* genes, which are poorly expressed in ACN293 without BenM and CatM. Thus, amplification of the *cat* genes should allow growth on anthranilate without demanding expression of the *benABCDE* operon. Derivatives of ACN293 that grew on anthranilate (Ant⁺) were readily selected at a frequency similar to that observed for benzoate. The Ant⁺ phenotype was associated with *cat* gene amplification as judged by PFGE (data not shown) and by copy-number evaluation of *catA* (described later).

Additionally, a new parent strain (ACN854) was constructed by deleting the *benM* and *catM* regulatory genes in an otherwise wild-type background. ACN854 lacks the point mutation that drives *benABCDE* gene expression in other parent strains. Therefore, to grow on benzoate, ACN854-derived mutants must harbor multiple copies of a chromosomal region encompassing not only the *cat* genes, but also the neighboring *benABCDE* operon. This selection increases the minimal amplicon size to approximately 15 kbp (Fig. A.1C). On solid medium, Ben⁺ derivatives arose with a frequency of approximately 1 in 10⁸ cells during a 14-day incubation time. Similarly, Ant⁺ mutants were selected from ACN854. Analysis of genomic DNA by PFGE was indicative of amplification (data not shown).

Sequences of recombination sites: DNA junctions

Twelve confirmed gene amplification mutants derived from ACN854 (Tables A.1 and A.S1) were characterized further by determining the sequence across the novel junction created by genetic duplication (depicted in Fig. A.1D). A previously described transformation assay was used to identify DNA spanning such junctions (Reams & Neidle, 2003), and the corresponding sequences from independently isolated Ben⁺ derivatives of ACN854 are shown in Fig. A.3. The precise junction represents the point at which DNA that is normally downstream of the *cat* genes joins DNA that is normally upstream of them, and junction DNA can be used to align these “downstream” and “upstream” sequences (dotted and solid lines, respectively, in Fig. A.1D). These alignments, shown in Fig. A.3B, can be used to infer the precise point of recombination and to evaluate the extent of sequence identity that might contribute to GDA events. The identical sequence was identified from two independently isolated mutants in the case of three different junctions. Therefore, the nine different sequences shown in Fig. A.3B reflect the junctions identified in twelve mutants. Junction sequences were similarly determined from the Ant⁺ amplification mutants derived from ACN293. These junction sequences are displayed in Fig. A.S2.

Amplicon size and copy number

The relationship between amplicon size and copy number was compared at different chromosomal positions (Fig. A.4). Amplicon size was inferred from the known junction sequence and/or was estimated from PFGE analysis of the genomic DNA from each mutant. Amplicon copy number was assessed using a real-time qPCR method. As described in Experimental procedures, copy number was determined with Taqman probes

specific for *catA* (in the amplified region) and for *antA* (assumed to be in single copy). Amplicon copy number was inferred to be equal to the *catA* to *antA* ratio.

In the set of amplification mutants derived from ACN293, the largest amplicons were maintained in the fewest copies (Fig. A.4A), and this inverse relationship was readily demonstrated with Pearson's correlation coefficient (p-value of 0.003, Table A.S2). In contrast, an inverse relationship was less evident for mutants derived from ACN854, which also harbors the *cat* genes in the native locus but lacks the *benA* promoter mutation that drives *ben* gene expression in ACN293 (Fig. A.4B). Analysis of the data for ACN854 indicated that an inverse relationship between amplicon size and copy number was not statistically significant for this data set (p-value of 0.43, Table A.S2).

When these attributes were examined from amplicons derived from the other parent strains, different patterns were observed at each locus (Fig. A.4C-E). To assess the statistical significance of these apparent differences between loci, we used the Mann-Whitney test (Tables A.S3 and A.S4). Although these data sets are relatively small, some significant differences were observed (see discussion). Among the notable differences, all of the amplicons in mutants derived from ACN1024 were maintained in relatively low copy number, with no more than six copies found in any mutant. In several cases, qPCR analysis suggested the presence of a simple duplication (copy number of two, Table A.S1).

Pulsed-Field Gel Analysis of Duplication Mutants

The isolation of duplication mutants was surprising since the Ben⁺ mutants from previous studies were found to have no fewer than three copies of the *cat*-gene region.

Therefore, PFGE with multiple restriction enzymes was used to evaluate the chromosomal configurations of the duplication mutants. In one example, the restriction enzyme I-CeuI was used to study an ACN1024-derived duplication mutant (ACN1135) that has an amplicon size of 0.14 Mbp. In comparison to its parent, the mutant had a larger version of the I-CeuI-generated fragment that carries the *cat*-genes (Fig. A.5). The increase in the size of this fragment (labeled III) is consistent with the presence of exactly two copies of the amplicon. These data (and additional analyses shown in Fig. A.S3) support the conclusion that ACN1135 is a duplication mutant. Similar analyses of ACN1136 and ACN1190 also confirmed a simple duplication in both of these ACN1024-derived mutants (data not shown). The only locus from which simple duplication mutants were isolated was locus *iv*.

Spontaneous duplication frequencies

Differences in the observed patterns of GDA mutants selected from different genomic regions led us to examine whether there were also differences in the frequency of spontaneously occurring duplications in the corresponding chromosomal locations. To address this question, we developed a transformation-based assay to evaluate the existence of cells that have at least two copies of *catA* in populations grown under conditions that do not select for high *cat*-gene expression. This assay, depicted in Fig. A.6, uses a cell-free preparation of linear donor DNA from a strain with an insertionally-inactivated *catA* allele (*catA*:: Ω K51025). When such donor DNA transforms recipient cells with a functional *catA* copy, drug resistance can be used to select recombinants in which allelic replacement has inactivated the chromosomal gene. If the *catA* gene were originally present in two (or more) copies within the recipient cell, the drug-resistant

transformant will remain Ben⁺ due to its retention of a functional copy of *catA*. Thus, the proportion of Ben⁻ to Ben⁺ cells within the set of Km^R transformants represents the *catA* duplication frequency in the original population.

This assay was conducted with the wild-type ADP1 as recipient and donor DNA from ACN1025, a Ben⁻ strain with the disrupted allele in its chromosome. The results revealed a duplication frequency of 10⁻⁴ at the native *cat*-gene locus. To evaluate the duplication frequency of the *catA* gene when it was in different chromosomal positions, donor strains that are Km^R and Ben⁻ were constructed in which *catA::ΩK51025* replaces the wild-type allele within the relocated cluster of *cat*-genes in locus *i* (ACN1028), *ii* (ACN1048), *iii* (ACN1080) or *iv* (ACN1037). Cell-free DNA lysates from these strains were then used to transform the corresponding strain that carried a functional set of relocated *cat*-genes in locus *i* (ACN1036), *ii* (ACN1046), *iii* (ACN1060), or *iv* (ACN1023). As shown in Fig. A.7, the duplication frequency of *catA* when located in these positions ranged from 10⁻⁴ to 10⁻⁵. Two loci relatively close to one another, with the *cat* genes in locus *ii* (ACIAD2822) or *iii* (ACIAD2969), displayed *catA* duplication frequencies that were 6-fold and 10-fold lower, respectively, than for *catA* in its native locus (Fig. A.7). These differences were statistically significant according to the Students t-test.

Discussion

Genome-wide studies of GDA in A. baylyi

An advantage of studying *A. baylyi* is the ease with which its high efficiency for natural transformation and recombination can be used to manipulate the genome (Elliott

& Neidle, 2011, Metzgar *et al.*, 2004, Young *et al.*, 2005). By exploiting these features, a system was developed for studying adaptation to growth limitation that is significantly different from those used in *E. coli* and *S. enterica*. Our system allows the capture of strains with large arrays of amplified chromosomal DNA and enables their junction sequences to be identified. In *A. baylyi*, GDA mutants can be isolated using the same core genes and identical selection conditions regardless of their genomic position. Thus, this approach prevents possible bias due to variability in selection methods. Furthermore, all selected mutants carry amplified DNA.

Method development relied on facile allelic replacement. The relocation plasmid carries a selectable cassette of catabolic genes flanked by cloning sites for the convenient insertion of PCR products that can be used to target the cassette to any nonessential locus in the chromosome (Fig. A.S1). Available information about the inactivation of individual genes in *A. baylyi* ADP1 helps guide strategic placement of the cassette and indicates genomic loci to avoid (de Berardinis *et al.*, 2008). Genetic relocation permits amplification mutants to be selected from key positions to test the contextual effects of features such as repetitive DNA, transposable elements, and sites for the initiation or termination of replication.

After genetic relocation, a functional *cat*-gene set can be used to determine duplication frequency with a transformation method based on a transduction assay developed in *S. enterica* (Anderson *et al.*, 1976). The recipient strains used in this transformation assay can be further modified by inactivation of *catM* in its new location to create the parents from which amplification mutants are selected. This process introduces a NotI site in *catM* that helps confirm the parent-strain genotype by PFGE

(Fig. A.2). Furthermore, this NotI site provides a marker for the PFGE-based detection of amplicons and their sizes (Fig. A.2). The *catA* copy number can then be precisely determined with a real-time qPCR assay that is sufficiently sensitive to detect a simple duplication (Figs. A.5 and A.S3).

A previously developed transformation method captured junction sequences that lie at the amplicon boundaries (Reams & Neidle, 2003). The determination of such sequences from diverse regions is arguably the greatest barrier to molecular studies of GDA. High-throughput hybridization and new sequencing methods have the potential for studying DNA junctions (Alkan *et al.*, 2011, Andersson & Hughes, 2009). However, until these methods are optimized and cost effective, our transformation assay and targeted sequencing approach is a practical and unique application for genome-wide GDA study.

Large amounts of amplified DNA

In the 49 newly isolated strains with *cat*-gene amplification, amplicon sizes ranged from 18 to 471 kbp, and copy number ranged from 2 to 105 (Table A.S1). The combination of both values can result in the carriage of large amounts of amplified DNA. Approximately 30% of the strains in this study appear to have amplified DNA in excess of 1 Mbp. In the most extreme example, 105 copies of a 27-kbp amplicon in ACN898 would correspond to 2.8 Mbp of amplified DNA. Considering that the wild-type genome is 3.6 Mbp, the flexibility of these bacteria to adapt to adverse conditions is remarkable. These results support models that amplification can provide ample material for mutation and functional diversification (Innan & Kondrashov, 2010, Bergthorsson *et al.*, 2007).

Sequence-based selective pressures: effects of a point mutation

In earlier studies of ACN293, nearly half of approximately 100 Ben⁺ derivatives had large amplicons (260-290 kbp), although selection requires only that a 10-kbp region be amplified (Fig. A.1)(Reams & Neidle, 2004a). These results may be affected by a demand to amplify the *cata* promoter without concomitant amplification of the *benA* promoter mutation. Amplification of this mutation was suspected to cause overproduction of the *ben*-encoded enzymes and hence the conversion of benzoate to a toxic level of catechol. This toxicity was not expected to occur with anthranilate rather than benzoate as a growth substrate even with a mutation-driven overproduction of Ben enzymes, since these enzymes do not accept anthranilate as a substrate. Despite this rationale, none of the ACN293-derived Ant⁺ mutants had amplicons that included the promoter mutation (Fig. A.S2). Moreover, as observed with selection on benzoate, a large fraction (3/7) of these Ant⁺ isolates had amplicons in the 260-290 kbp range (Table A.S1, Fig. A.4).

Under similar selection conditions, using a new parent strain without the promoter mutation, ACN854, none of the resulting mutants had a 260-290 kbp amplicon (Fig. A.4b). As expected, all amplicons selected on benzoate carried the wild-type *benA* promoter, since this is a requirement of growth. Interestingly, two of the three ACN854-derived Ant⁺ mutants carried the wild-type *benA* promoter on the amplicon, even though *ben*-gene expression is not needed for growth on anthranilate. The constraint that limits the location of the upstream endpoint of the amplicon in ACN293-derived mutants appears to have been lifted for ACN854 derivatives.

Contrary to our earlier assumptions, it appears that selection against amplification of the *benA* mutation is maintained during growth on anthranilate. Amplification of the

benA mutation may cause problems other than (or in addition to) the accumulation of catechol from benzoate. Overproduction of the *benABC*-encoded benzoate dioxygenase might cause oxidative stress without its normal substrate or in the presence of a substrate analog such as anthranilate. Normally, benzoate dioxygenase is only induced when needed (Collier *et al.*, 1998). Surprisingly, the constraint limiting the position of the upstream endpoint of the amplicon also affects the location of the downstream endpoint. Recombination involving DNA in the region between the *benA* and *catA* promoters appears to occur preferentially with a region nearly 300-kbp away.

DNA junction sequences

Alignments of the sequences involved in generating these large amplicons revealed little identity (Fig. A.S2). Duplication can result from illegitimate recombination between short direct repeats of only 15-30 base pairs or in the absence of discernible sequence identity (Reams & Neidle, 2004a, Ikeda *et al.*, 2004). Analysis of the smaller ACN854-derived amplicons similarly lacked obvious sequence identity (Fig. A.3B). In ACN854-derived mutants, there were three examples of the identical junction being found in two independent mutants. In one case, the mutants were isolated from different carbon sources, ACN951 (*Ant*⁺) and ACN1118 (*Ben*⁺). Copy number assessment for both strains in each of the three pairs was comparable (Table A.S1). For example, ACN869 and ACN906, had *catA* copy numbers of 13 and 15, respectively. Thus, strains with a given *cat*-amplicon appear to maintain it at a characteristic level under a specific selective condition.

Recurrent junction sequences were previously observed and inferred to reflect site-specific illegitimate recombination (Reams & Neidle, 2004a). To determine if the

newly identified recurrent junction sequences were present in unselected bacterial populations, PCR techniques were used (data not shown). Although these efforts were unsuccessful, such duplications might be present at low levels that fall below the detection limits of the methodology. Alternatively, these junctions might be generated by secondary rearrangements during the GDA process. Examples of such rearrangements have been reported, including the creation of a novel junction sequence via the deletion of a segment within a pre-existing duplication (Kugelberg *et al.*, 2006).

Duplication frequencies

In *S. enterica*, duplication frequencies at different chromosomal positions vary across several orders of magnitude, from 3×10^{-2} to 6×10^{-5} per bacterial cell (Pettersson *et al.*, 2009, Sonti & Roth, 1989, Anderson & Roth, 1981). In contrast, the *catA* duplication frequencies differed by no more than 10-fold regardless of genomic position (Fig. A.7). Based on segregation rates in *S. enterica* (Reams *et al.*, 2010), our results should correspond to steady-state duplication frequencies. Although the *A. baylyi* values fell within the lower end of the range for the other bacteria, this difference may not be significant. The higher frequencies typically correspond to loci flanked by regions of extensive DNA identity resulting from closely spaced copies of insertion sequences or ribosomal RNA operons (Anderson & Roth, 1981, Kroll *et al.*, 1991, Sun *et al.*, 2009, Haack & Roth, 1995). While our limited results may not yet reflect the full genomic range, it has been noted that the *A. baylyi* ADP1 genome has a relatively small fraction of repetitive DNA (Barbe *et al.*, 2004). Interestingly, *vanK* (locus *i*), previously reported to undergo frequent rearrangements due to two nearby copies of an insertion sequence, did

not have a significantly higher duplication frequency than the other loci examined (Segura *et al.*, 1999).

GDA variation based on position and context

There was greater variation observed in amplicon size and copy number than for duplication frequency. For example, a significant inverse relationship between size and copy number was only found for amplicons derived from ACN293 (Fig. A.4). Amplicons from ACN854, a parent with the *cat* genes in the same native locus as ACN293, had a different pattern that was associated with the presence or absence of a point mutation (Fig. A.4A & A.4B, discussed earlier). As noted in our previous studies, there was no correlation between the time it took for a mutant colony to appear initially on the selective medium and the features of its amplicon such as size or copy number. These observations are consistent with the amplified arrays arising from a stochastic process.

Some loci, notably *i* (*vanK*) and *ii* (ACIAD2822), gave rise to amplicons that were consistently small in size, but with large variations in copy number. These results may reflect proximity to nearby copies of an insertion sequence (Cuff *et al.*, 2012). In contrast, at locus *iv* (ACIAD3593), amplicons varied greatly in size (approximately 47 to 471 kbp) but all exhibited low copy numbers (≤ 6). Only this locus yielded mutants with simple duplications, providing the first evidence that two copies of the *cat* genes are sufficient for growth on benzoate (Fig. A.5 and A.S3). Results in this region could result from the presence of some nearby gene(s) that frequently co-amplify with the *cat* genes and that confer a high fitness cost.

Our earlier suggestion that larger amplicons in *A. baylyi* have a higher fitness cost and consequently lower copy number was biased by the extrapolation of results from

a single parent strain (ACN293) (Reams & Neidle, 2004a). Current results suggest that the fitness cost of a given amplicon, presumably reflected in its steady state copy number, is dictated by genetic content. This conclusion agrees with *Salmonella* studies that found no correlation between the fitness cost and the size of a duplicated region (from 72 to >1200 kbp) and indicated that such costs are variable and locus dependent (Pettersson *et al.*, 2009, Reams *et al.*, 2010).

Locus dependence was also observed recently in a different aspect of genomic plasticity that influences horizontal gene transfer (heterogamic transformation). In a study of *Acinetobacter* strains, genomic location was a key factor in interspecies transformation frequencies (Ray *et al.*, 2009). This study not only highlights the importance of examining multiple loci under identical conditions, but it also demonstrates the versatility of *A. baylyi* as a model system for genome-wide experiments.

Concluding remarks

A. baylyi ADP1 is a genetically malleable soil bacterium that displays differences in its genetic content, metabolic capabilities, and environmental niche from enteric bacteria such as *E. coli* and *S. enterica*. Because of these differences, *A. baylyi* can provide comparative, complementary, and novel information about GDA relative to other experimental studies. In this report, we demonstrate the ability to evaluate GDA features that have not previously been explored systematically in any organism. Results using this genome-wide experimental system highlight the complex contributions of genomic context to the nature of the selected gene duplication and amplification events.

Experimental procedures

Bacterial strains and growth conditions

A. baylyi strains with ACN designations are derivatives of the wild-type strain ADP1 (Table A.1). *E. coli* strains DH5 α and Top10 (Invitrogen) were used as plasmid hosts. *E. coli* was grown at 37°C in lysogeny broth (LB), also known as Luria-Bertani medium [10 g of Bacto-tryptone, 5 g of yeast extract and 10 g of NaCl per liter (Sambrook *et al.*, 1989)]. *A. baylyi* strains were cultured at 30°C or 37°C in minimal salts medium (MM) with succinate (10 mM), benzoate (1 mM), catechol (1 mM), anthranilate (1 mM), or *cis,cis*-muconate (2 mM) as the carbon source (Shanley *et al.*, 1986), or in LB. Antibiotics were added as needed at the following final concentrations: kanamycin (Km), 25 μ g/ml; spectinomycin and streptomycin 15 μ g/ml; ampicillin 150 μ g/ml. To select for loss of the *sacB* counter-selectable marker, cells were cultured on LB with 5% w/v sucrose, at 30°C (Jones & Williams, 2003).

DNA manipulation, strain and plasmid construction

Standard methods were used for molecular techniques (Sambrook *et al.*, 1989). DNA sequencing was done at the University of Georgia Integrated Biotech Laboratories core facility or Genewiz, Inc. Genomic DNA was isolated from *A. baylyi* strains using the Easy-DNA kit (Invitrogen) for genomic libraries and the illustra bacteria genomicPrep Mini Spin Kit (GE) for real-time quantitative PCR (qPCR). Linear chromosomal DNA from *A. baylyi* strains for use in transformations was prepared by a lysis procedure, as previously described (Juni, 1972, Neidle & Ornston, 1986). *A. baylyi* strains were transformed with linearized plasmid DNA or chromosomal DNA in cell lysates as previously described (Juni, 1972, Neidle & Ornston, 1986). Newly engineered strains

were tested for phenotype by screening for appropriate drug resistance and carbon source utilization. Genotypes were tested using PCR, sequencing, Southern hybridization, and/or PFGE. Newly constructed plasmids were sequenced as necessary to confirm correct construction. All cited genomic coordinates refer to GenBank entry CR543861.

PFGE and Southern hybridization methods

PFGE and Southern hybridization were performed as previously described (Reams & Neidle, 2003, Gralton *et al.*, 1997). PFGE band sizes were estimated using Gel-Pro Analyzer software, version 4.0 (Media Cybernetics, LP).

Allelic replacement of catM

To facilitate subsequent chromosomal manipulation, *catM* was disrupted with a counter-selectable *sacB*-Km^R cassette as follows. A deletion in the *catM* allele of pBAC6 (Ezezika *et al.*, 2006) was made by digestion with SalI followed by self-ligation to generate pBAC707. A SalI fragment with the desired cassette was excised from pRMJ1 (Jones & Williams, 2003) and ligated into the SalI site of pBAC707, creating pBAC708. This new allele (*catM::sacB*-Km^R5613) was used to replace *catM* of a recipient strain, ACN147, by transforming it with pBAC708 that had been linearized with XmnI. A Km^R-transformant with the correct genotype was designated ACN814.

Chromosomal cat-gene deletion

Selection for loss of ACN814's *sacB* marker facilitated acquisition of an engineered *cat*-gene deletion (allele Δ *cat5825* encompassing genomic coordinates 1439923-1449545). A plasmid-borne version of this allele was constructed via overlap-extension PCR to include surrounding DNA extending from genomic coordinates 1439140-1450109 (Horton *et al.*, 1990). The resulting plasmid, pBAC761, was linearized

with NdeI and used to transform ACN814. The resulting strain, ACN825, carries the engineered chromosomal deletion and is unable to grow on benzoate (Ben⁻).

A cat-gene relocation plasmid

Plasmid pBAC841 (Fig. A.S1) carries the wild-type *cat* genes flanked by unique restriction sites. It was constructed in multiple steps using DNA from several available plasmids. A fragment generated by EcoRI and ClaI digestion of pPAN4 (Shanley *et al.*, 1986) was ligated into similarly digested pIB3 (Neidle & Ornston, 1986) to generate pBAC184, which has DNA that spans from downstream of *catA* through *catD* with an internal ClaI deletion. An EcoRV to BglII fragment containing *catA* was cloned from pIB1361 (Neidle & Ornston, 1986) into pBAC184, previously cut with SmaI and BglII, to generate pBAC832. To add additional, unique restriction sites, an approximately 500-bp DraIII to EcoRI fragment from the multiple cloning site of pET21b (Invitrogen) was ligated into pBAC832 cut with EcoRI and NdeI after the DraIII and NdeI ends had been treated with T4 DNA Polymerase (New England Biolabs). The resulting plasmid, pBAC840, was linearized with ClaI and transformed into wild-type ADP1 to allow capture of the remaining *cat* region DNA via the gap-repair method (Gregg-Jolly & Ornston, 1990). This method generated the final plasmid, pBAC841, such that DNA fragments can be readily inserted on either side of the *cat* genes to target them to a specific chromosomal location via homologous recombination (Fig. A.S1).

Chromosomal relocation of functional cat genes

Four loci were chosen for the introduction of the *cat* genes into new chromosomal sites by allelic replacement: ACIAD0982 (*vanK*), ACIAD2822, ACIAD2969 (*aroP*), and ACIAD3593. The relocation method, described first for locus *i* (*vanK*), involves PCR

amplification of an approximately 1 kbp genomic region within the target region (*vanK*) using primers to introduce PstI and BamHI restriction sites at the 5' and 3' ends, respectively. These enzymes were then used to digest the PCR product and the relocation plasmid (pBAC841). Ligation of the *vanK* and plasmid fragments created an intermediate, pBAC873, with the plasmid-borne *cat* genes adjacent to one target for homologous recombination. A second *vanK* PCR product was generated with primers that introduced SacI and XhoI restriction sites at the 5' and 3' ends, respectively. The PCR product was digested with these enzymes and ligated into similarly digested pBAC873. The resulting plasmid, pBAC874, contains the *cat*-region flanked by *vanK* sequences (Fig. A.S1). Analogous methods were used to generate pBAC850, pBAC876, and pBAC872 in which the *cat*-genes are flanked by sequences of ACIAD3593, ACIAD2822 and ACIAD2969, respectively.

To introduce the *cat* genes into each new locus, the plasmid carrying DNA designed to replace corresponding wild-type sequence (pBAC874, pBAC850, pBAC876, or pBAC872) was linearized by digestion with PstI or XhoI. Individual samples of linearized DNA were used to transform the Ben⁻ strain lacking the *cat* genes, ACN825. Transformants in which the engineered configuration replaced the corresponding chromosomal region were selected by their ability to grow on benzoate (Ben⁺ phenotype). Ben⁺ strains with the correct genotypes containing the *cat* genes in the indicated locus (as shown in Fig. A.2) were designated ACN1036 (locus *i*), ACN1046 (locus *ii*), ACN1060 (locus *iii*), and ACN1023 (locus *iv*). In each strain, the *cat* genes were oriented on the chromosome in the same direction as in the wild-type strain, and the

transcriptional direction of *catA* and the *catBCIJFD* operon was the same as that of the disrupted gene.

Inactivation of relocated catM

Each Ben⁺ strain with the *cat*-genes in a new chromosomal location was transformed with linearized pBAC6 DNA (Ezezika *et al.*, 2006). This plasmid DNA, digested with XhoI, carried a disrupted *catM* allele encoding resistance to Km. Drug resistance was used to select transformants in which the disrupted *catM* replaced the wild-type allele. Strains with the correct genotype, carrying *catM::ΩK5541* in the indicated locus (Fig. A.2) became Ben⁻ and were designated ACN1039 (locus *i*), ACN1050 (locus *ii*), ACN1061 (locus *iii*), and ACN1024 (locus *iv*). Amplification mutants were selected from these parent strains.

Selection of amplification mutants

Ben⁺ mutants were selected on solid medium from concentrated lawns of Ben⁻ parent strains that had initially been grown in succinate medium, as described previously (Reams & Neidle, 2003). As in the past, no more than one mutant was selected from any individual succinate-grown culture to ensure independent selection. The same approach was used to select mutants on anthranilate (Ant⁺) or *cis,cis*-muconate (CCM⁺) as the sole carbon source.

Isolation of junction DNA

To determine the precise sequence at chromosomal duplication junctions, a transformation-based assay was employed as previously described (Reams & Neidle, 2003). This assay relies on the ability of the junction DNA to serve as a platform for homologous recombination such that it can regenerate a duplication and confer growth on

benzoate or anthranilate to an *A. baylyi* parent strain unable to grow on such medium. Genomic DNA fragments from amplification mutants were used to generate plasmid libraries in a cloning vector (pZErO-2 or pUC19). Recombinant plasmids, housed in *E. coli*, that gave a positive result in the transformation assay were sequenced and characterized as previously described (Reams & Neidle, 2003).

New parent strain lacking BenM and CatM (ACN854)

To generate a *benM* and *catM* mutant with a wild-type *benA* promoter, the *catM* gene of *benM*-deleted strain ACN389 was disrupted. A linearized version of pBAC708 was used for allelic replacement of *catM* in ACN389 to generate ACN843. This strain was selected by drug-resistance conferred by its acquired chromosomal *catM::sacB-Km^R5613* allele. Subsequently, this allele was replaced by selecting for the loss of *sacB* after introduction of the *catM*-deleted allele from linearized pIGG6 (446 bp *catM* deletion, *catM*Δ5293) (Reams & Neidle, 2004a). This replacement yielded parent strain ACN854.

Amplicon copy number assessment

A real-time qPCR assay, with TaqMan 5'-exonuclease (Applied Biosystems), was used to evaluate the amount of *catA* DNA relative to that of a control gene, *antA*, assumed to be in the genome in single copy. The following primer-probe sets were used (and designed with the PrimerExpress software, Applied Biosystems): *catA*_FOR: CATTACCTCGATATGCGTATGGAT, *catA*_REV: CGTGGTGTTCGCATTTTCAAT, *antA*_FOR: TGCCGTA CTTGGTGGAAAAGT, *antA*_REV: CGGTGCATGGCCCATT. The sequences of TaqMan probes to detect *catA* and *antA*, which were 5'-labeled with the fluorescent reporter 6-carboxyfluorescein (6FAM), follow: *catA*FAM: (6FAM)-

CCGAAGATGCCGCACT, antAFAM: (6FAM)-TTTCATATTCAAGAGCTGGTTTA.

To quantify *catA* and *antA* targets, a standard curve was generated using ADP1 genomic DNA as the template for PCR (12.5 ng to 0.02 ng). In ADP1, the *catA:antA* ratio is assumed to be 1. In an amplification mutant, the ratio of *catA:antA* is assumed to equal the amplicon copy number.

Duplication frequency assay

A natural transformation-based assay was developed to assess the duplication frequency of *catA*. For this purpose, an insertionally inactivated *catA* allele was constructed as follows. An internal BamHI site in *catA* was created by PCR amplifying this gene from the ADP1 chromosome in two fragments such that this site was created when the fragments were joined. The addition of an external EcoRI or HindIII site to the opposite end of these fragments allowed them to be cloned into appropriately digested pUC19. The resulting plasmid, pBAC763, was then digested with BamHI to cleave *catA* and allow insertion of a BamHI fragment with the Ω Km^R cassette derived from pUI1637 (Eraso & Kaplan, 1994). The resulting plasmid, pBAC852, was then used to inactivate *catA* in its native or relocated chromosomal position. After linearization with AatII, pBAC852 was used to replace *catA* with the new allele, *catA* Ω K51025. The following Ben⁻ and Km^R strains were generated with *catA* Ω K51025 in the indicated locus (Fig. A.2): ACN1025 (wt locus), ACN1038 (locus *i*), ACN1048 (locus *ii*), ACN1080 (locus *iii*), and ACN1037 (locus *iv*).

Lysates from these strains were used as donor DNA to transform recipient strains with the wild-type *catA* in the corresponding chromosomal locus. Transformations were conducted as follows: 10 μ l of 1 M succinate was added to a 5 mL overnight culture of

the recipient strain. After a 30-min incubation at 37°C, 100 µl of this culture was aliquoted into 900 µl of LB broth with 1.5 µg of the donor DNA or no DNA for the negative control. The recipient strain and DNA were incubated at 37°C with shaking for 5 h, then diluted and plated to determine cfu/mL on the following media: succinate MM, benzoate MM, succinate MM with Km, and benzoate MM with Km. Duplication frequency was calculated as (cfu/mL on benzoate MM with Km) ÷ (cfu/mL on succinate MM with Km). Without Km, plating efficiencies on benzoate MM were comparable to those on succinate MM. For each strain, this assay was performed in duplicate with two independent recipient strain cultures and two independent donor DNA preparations. The results from at least three independent experiments were combined and standard deviations are shown.

Statistical Analysis

The significance of differences in amplicon size or *catA* copy number between parent strains was assessed by a two-tailed Mann-Whitney test at <http://faculty.vassar.edu/lowry/utest.html>. Correlation between amplicon size and copy number was assessed with Pearson's correlation coefficient in Excel 2010, and significance was calculated at <http://faculty.vassar.edu/lowry/rsig.html>. The significance of differences in duplication frequency between loci was assessed by an unpaired, two-tailed Students t-test using Excel 2010. For all analyses, p-values < 0.05 were considered statistically significant.

Acknowledgements

We gratefully acknowledge Jan Mrazek for helpful discussions and assistance with statistical analysis. Sandra Haddad, Lauren Collier-Hyams, and Jennifer Morgan constructed plasmids used in the current study, and Jennifer Taylor assisted with selection of mutant *A. baylyi* strains. In addition, we acknowledge members of the Cold Spring Harbor Laboratory Advanced Bacterial Genetics course (Summer 2010) for assistance with copy number analysis, as well as Susan Lovett and Anca Segall for helpful discussions regarding this work. National Science Foundation grants (MCB 0920893 to ELN and DBI 0905829 to KTE) supported this work.

References

- Albertson, D.G., (2006) Gene amplification in cancer. *Trends Genet* **22**: 447-455.
- Alkan, C., B.P. Coe & E.E. Eichler, (2011) Genome structural variation discovery and genotyping. *Nat Rev Genet* **12**: 363-376.
- Anderson, P. & J. Roth, (1981) Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (*rrn*) cistrons. *Proc Natl Acad Sci USA* **78**: 3113-3117.
- Anderson, R.P., C.G. Miller & J.R. Roth, (1976) Tandem duplications of the histidine operon observed following generalized transduction in *Salmonella typhimurium*. *J Mol Biol* **105**: 201-218.
- Andersson, D.I., (2011) Evolving promiscuously. *Proc Natl Acad Sci USA* **108**: 1199-1200.
- Andersson, D.I. & D. Hughes, (2009) Gene amplification and adaptive evolution in bacteria. *Annu. Rev. Genet.* **43**: 167-195.
- Barbe, V., D. Vallenet, N. Fonknechten, A. Kreimeyer, S. Oztas, L. Labarre, S. Cruveiller, C. Robert, S. Duprat, P. Wincker, L.N. Ornston, J. Weissenbach, P. Marliere, G.N. Cohen & C. Medigue, (2004) Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. *Nucleic Acids Res* **32**: 5766-5779.
- Bergthorsson, U., D.I. Andersson & J.R. Roth, (2007) Ohno's dilemma: evolution of new genes under continuous selection. *Proc. Natl. Acad. Sci. U.S.A.* **104**: 17004-17009.

- Bundy, B.M., A.L. Campbell & E.L. Neidle, (1998) Similarities between the *antABC*-encoded anthranilate dioxygenase and the *benABC*-encoded benzoate dioxygenase of *Acinetobacter* sp. strain ADP1. *J Bacteriol* **180**: 4466-4474.
- Bundy, B.M., L.S. Collier, T.R. Hoover & E.L. Neidle, (2002) Synergistic transcriptional activation by one regulatory protein in response to two metabolites. *Proc Natl Acad Sci USA* **99**: 7693-7698.
- Collier, L.S., G.L. Gaines, 3rd & E.L. Neidle, (1998) Regulation of benzoate degradation in *Acinetobacter* sp. strain ADP1 by BenM, a LysR-type transcriptional activator. *J Bacteriol* **180**: 2493-2501.
- Conrad, B. & S.E. Antonarakis, (2007) Gene duplication: a drive for phenotypic diversity and cause of human disease. *Annu Rev Genomics Hum Genet* **8**: 17-35.
- Craven, S.H. & E.L. Neidle, (2007) Double trouble: medical implications of genetic duplication and amplification in bacteria. *Future Microbiol* **2**: 309-321.
- Cuff, L.E., K.T. Elliott, S.C. Seaton, M.K. Ishaq, N.S. Laniohan, A.C. Karls & E.L. Neidle, (2012) Analysis of *IS1236*-mediated gene amplification events in *Acinetobacter baylyi* ADP1. *J Bacteriol* **194**: 4395-4405.
- de Berardinis, V., D. Vallenet, V. Castelli, M. Besnard, A. Pinet, C. Cruaud, S. Samair, C. Lechaplais, G. Gyapay, C. Richez, M. Durot, A. Kreimeyer, F. Le Fevre, V. Schachter, V. Pezo, V. Doring, C. Scarpelli, C. Medigue, G.N. Cohen, P. Marliere, M. Salanoubat & J. Weissenbach, (2008) A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol Syst Biol* **4**: 174.

- Eby, D.M., Z.M. Beharry, E.D. Coulter, D.M. Kurtz, Jr. & E.L. Neidle, (2001) Characterization and evolution of anthranilate 1,2-dioxygenase from *Acinetobacter* sp. strain ADP1. *J Bacteriol* **183**: 109-118.
- Elliott, K.T. & E.L. Neidle, (2011) *Acinetobacter baylyi* ADP1: transforming the choice of model organism. *IUBMB Life* **63**: 1075-1080.
- Eraso, J.M. & S. Kaplan, (1994) *prpA*, a putative response regulator involved in oxygen regulation of photosynthesis gene expression in *Rhodobacter sphaeroides*. *J Bacteriol* **176**: 32-43.
- Ezezika, O.C., L.S. Collier-Hyams, H.A. Dale, A.C. Burk & E.L. Neidle, (2006) CatM regulation of the *benABCDE* operon: functional divergence of two LysR-type paralogs in *Acinetobacter baylyi* ADP1. *Appl Environ Microbiol* **72**: 1749-1758.
- Gralton, E.M., A.L. Campbell & E.L. Neidle, (1997) Directed introduction of DNA cleavage sites to produce a high-resolution genetic and physical map of the *Acinetobacter* sp. strain ADP1 (BD413UE) chromosome. *Microbiology* **143 (Pt 4)**: 1345-1357.
- Gregg-Jolly, L.A. & L.N. Ornston, (1990) Recovery of DNA from the *Acinetobacter calcoaceticus* chromosome by gap repair. *J Bacteriol* **172**: 6169-6172.
- Haack, K.R. & J.R. Roth, (1995) Recombination between chromosomal IS200 elements supports frequent duplication formation in *Salmonella typhimurium*. *Genetics* **141**: 1245-1252.
- Horton, R.M., Z.L. Cai, S.N. Ho & L.R. Pease, (1990) Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *Biotechniques* **8**: 528-535.

- Ikeda, H., K. Shiraishi & Y. Ogata, (2004) Illegitimate recombination mediated by double-strand break and end-joining in *Escherichia coli*. *Adv Biophys* **38**: 3-20.
- Innan, H. & F. Kondrashov, (2010) The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* **11**: 97-108.
- Jones, R.M. & P.A. Williams, (2003) Mutational analysis of the critical bases involved in activation of the AreR-regulated sigma54-dependent promoter in *Acinetobacter* sp. strain ADP1. *Appl Environ Microbiol* **69**: 5627-5635.
- Juni, E., (1972) Interspecies transformation of *Acinetobacter*: genetic evidence for a ubiquitous genus. *J Bacteriol* **112**: 917-931.
- Juni, E. & A. Janik, (1969) Transformation of *Acinetobacter calco-aceticus* (*Bacterium anitratum*). *J Bacteriol* **98**: 281-288.
- Kroll, J.S., B.M. Loynds & E.R. Moxon, (1991) The *Haemophilus influenzae* capsulation gene cluster: a compound transposon. *Mol Microbiol* **5**: 1549-1560.
- Kugelberg, E., E. Kofoid, A.B. Reams, D.I. Andersson & J.R. Roth, (2006) Multiple pathways of selected gene amplification during adaptive mutation. *Proc Natl Acad Sci USA* **103**: 17319-17324.
- Metzgar, D., J.M. Bacher, V. Pezo, J. Reader, V. Doring, P. Schimmel, P. Marliere & V. de Crecy-Lagard, (2004) *Acinetobacter* sp. ADP1: an ideal model organism for genetic analysis and genome engineering. *Nucleic Acids Res* **32**: 5780-5790.
- Neidle, E.L. & L.N. Ornston, (1986) Cloning and expression of *Acinetobacter calcoaceticus* catechol 1,2-dioxygenase structural gene *catA* in *Escherichia coli*. *J Bacteriol* **168**: 815-820.

- Nicoloff, H., V. Perreten, L.M. McMurry & S.B. Levy, (2006) Role for tandem duplication and *lon* protease in AcrAB-TolC- dependent multiple antibiotic resistance (Mar) in an *Escherichia coli* mutant without mutations in *marRAB* or *acrRAB*. *J Bacteriol* **188**: 4413-4423.
- Paulander, W., D.I. Andersson & S. Maisnier-Patin, (2010) Amplification of the gene for isoleucyl-tRNA synthetase facilitates adaptation to the fitness cost of mupirocin resistance in *Salmonella enterica*. *Genetics* **185**: 305-312.
- Pettersson, M.E., S. Sun, D.I. Andersson & O.G. Berg, (2009) Evolution of new gene functions: simulation and analysis of the amplification model. *Genetica* **135**: 309-324.
- Pranting, M. & D.I. Andersson, (2011) Escape from growth restriction in small colony variants of *Salmonella typhimurium* by gene amplification and mutation. *Mol. Microbiol.* **79**: 305-315.
- Quinones-Soto, S. & J.R. Roth, (2011) Effect of Growth Under Selection on Appearance of Chromosomal Mutations in *Salmonella enterica*. *Genetics*.
- Ray, J.L., K. Harms, O.G. Wikmark, I. Starikova, P.J. Johnsen & K.M. Nielsen, (2009) Sexual isolation in *Acinetobacter baylyi* is locus-specific and varies 10,000-fold over the genome. *Genetics* **182**: 1165-1181.
- Reams, A.B., E. Kofoid, M. Savageau & J.R. Roth, (2010) Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. *Genetics* **184**: 1077-1094.

- Reams, A.B. & E.L. Neidle, (2003) Genome plasticity in *Acinetobacter*: new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. *Mol Microbiol* **47**: 1291-1304.
- Reams, A.B. & E.L. Neidle, (2004a) Gene amplification involves site-specific short homology-independent illegitimate recombination in *Acinetobacter* sp. strain ADP1. *J Mol Biol* **338**: 643-656.
- Reams, A.B. & E.L. Neidle, (2004b) Selection for gene clustering by tandem duplication. *Annu Rev Microbiol* **58**: 119-142.
- Roth, J., N. Benson, T. Galitsky, K. Haak, J. Lawrence & L. Miesel, (1996) Rearrangements of the bacterial chromosome: formation and applications. In: *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd ed. R.C.I. FC Neidhardt, JL Ingraham, ECC Lin, KB Low, et al. (ed). Washington, DC: American Society for Microbiology, pp. 2256–2276.
- Roth, J.R., (2011) The joys and terrors of fast adaptation: new findings elucidate antibiotic resistance and natural selection. *Mol. Microbiol.* **79**: 279-282.
- Roth, J.R., E. Kugelberg, A.B. Reams, E. Kofoed & D.I. Andersson, (2006) Origin of mutations under selection: the adaptive mutation controversy. *Annu Rev Microbiol* **60**: 477-501.
- Sambrook, J., T. Maniatis & E. Fritsch, (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York.
- Segura, A., P.V. Bunz, D.A. D'Argenio & L.N. Ornston, (1999) Genetic analysis of a chromosomal region containing *vanA* and *vanB*, genes required for conversion of

- either ferulate or vanillate to protocatechuate in *Acinetobacter*. *J Bacteriol* **181**: 3494-3504.
- Shanley, M.S., E.L. Neidle, R.E. Parales & L.N. Ornston, (1986) Cloning and expression of *Acinetobacter calcoaceticus catBCDE* genes in *Pseudomonas putida* and *Escherichia coli*. *J Bacteriol* **165**: 557-563.
- Slechta, E.S., J. Harold, D.I. Andersson & J.R. Roth, (2002) The effect of genomic position on reversion of a *lac* frameshift mutation (*lacIZ33*) during non-lethal selection (adaptive mutation). *Mol Microbiol* **44**: 1017-1032.
- Sonti, R.V. & J.R. Roth, (1989) Role of gene duplications in the adaptation of *Salmonella typhimurium* to growth on limiting carbon sources. *Genetics* **123**: 19-28.
- Stankiewicz, P. & J.R. Lupski, (2010) Structural variation in the human genome and its role in disease. *Annu Rev Med* **61**: 437-455.
- Sun, S., O.G. Berg, J.R. Roth & D.I. Andersson, (2009) Contribution of gene amplification to evolution of increased antibiotic resistance in *Salmonella typhimurium*. *Genetics* **182**: 1183-1195.
- Tatusov, R.L., M.Y. Galperin, D.A. Natale & E.V. Koonin, (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**: 33-36.
- Young, D.M., D. Parke & L.N. Ornston, (2005) Opportunities for genetic investigation afforded by *Acinetobacter baylyi*, a nutritionally versatile bacterial species that is highly competent for natural transformation. *Annu Rev Microbiol* **59**: 519-551.

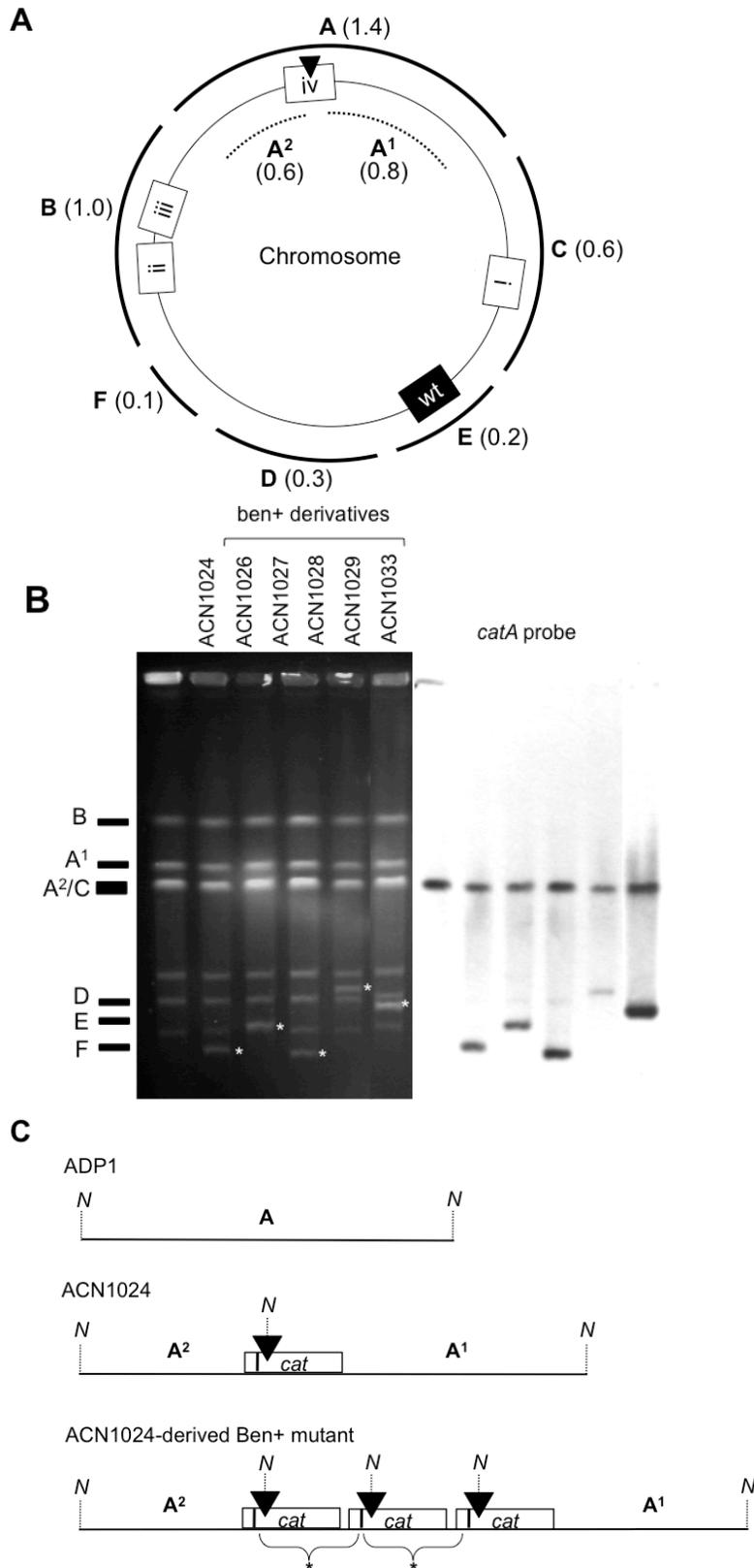


Figure A.2. Genomic analysis of amplification mutants. (A) Cleavage of the ADP1 chromosome with NotI yields six fragments, depicted as A-F (size in Mbp). The *cat*

genes, normally on fragment E (wt), were relocated in different strains to new positions *i*, *ii*, *iii*, and *iv* in genes ACIAD0982 (*vanK*), ACIAD2822, ACIAD2969, and ACIAD3593, respectively. In ACN1024 with the *cat* genes in locus *iv*, insertional inactivation of *catM* introduced an additional restriction site (black triangle) that causes NotI to cleave fragment A into two pieces (A1 and A2). (B) PFGE (left) of NotI-digested genomic DNA from ACN1024 and its Ben⁺ derivatives. Two similarly sized fragments appear to co-migrate. The band corresponding to fragments A2 and C is inferred to be a doublet because it is more intense than the band of the larger A1 fragment. With Southern hybridization (right panel), a *catA*-specific probe detected A2 and additional fragments in the mutants (*). (C) In Ben⁺ mutants derived from ACN1024 (bottom line), digestion with NotI generates a fragment that corresponds to the size of the amplicon (*) due to multiple copies of the disrupted *catM*. This amplicon-associated fragment is absent from the wild type (top line) and ACN1024 parent (middle line). The black rectangle in the *cat* gene region indicates the approximate location of the *catA* probe.

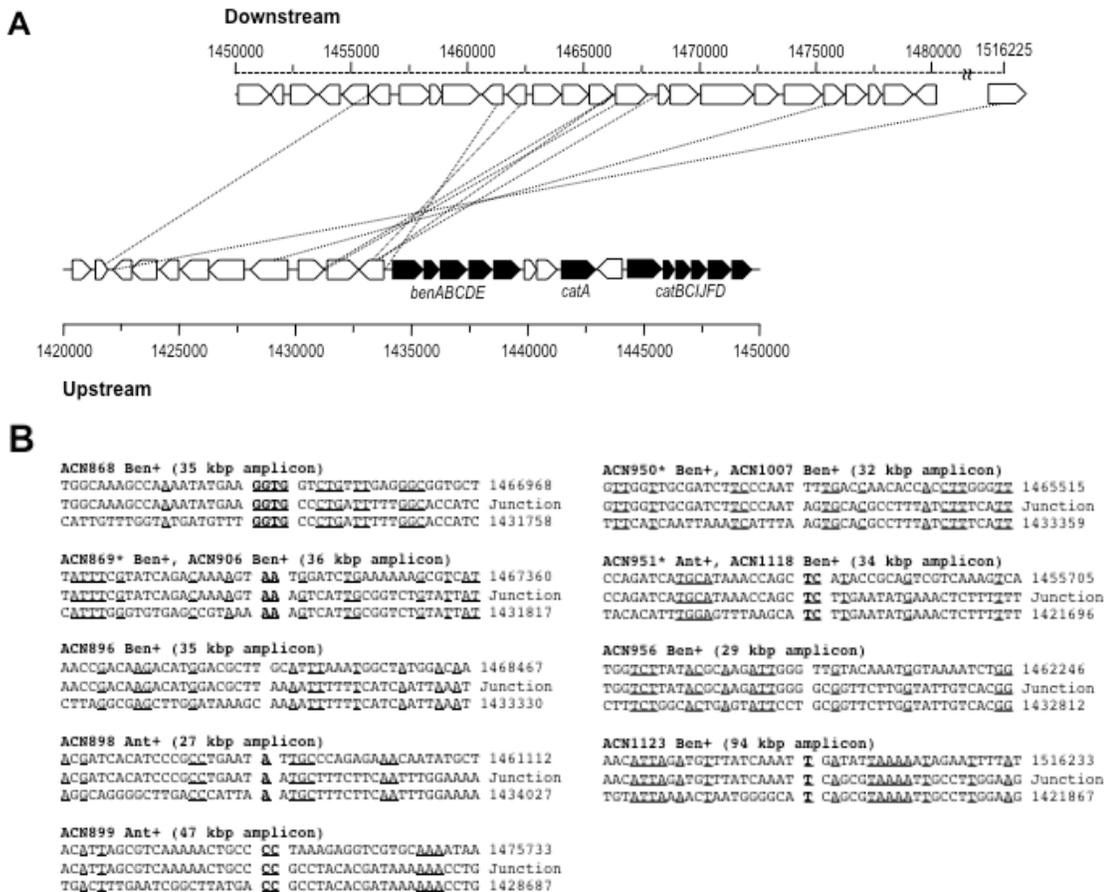


Figure A.3. DNA sequence involved in recombination to generate tandem duplication of the native *ben-cat* region. (A) The downstream (top, dotted line) and upstream (bottom, solid line) amplicon endpoints are shown relative to the *ben* and *cat* genes at the native locus. Each line represents an independent mutant, derived from ACN854, with a genomic duplication. Genomic coordinates correspond to the published genome sequence (GenBank entry CR543861). (B) Duplication junction sequences for ACN854-derived mutants, selected on benzoate (Ben⁺) or anthranilate (Ant⁺) are shown. The center line of text indicates the DNA junction from an amplification strain. The top line represents the parental downstream sequence, with the first displayed nucleotide identified by its position in the ADP1 genome sequence. The bottom line is the parental upstream sequence, with the position of the first nucleotide indicated. Identical nucleotides at the DNA junction are represented by bold and underlined text. Identity in the vicinity of the recombination junction is represented by underlined text. Note that three junctions were each identified in two independent amplification strains (indicated by *).

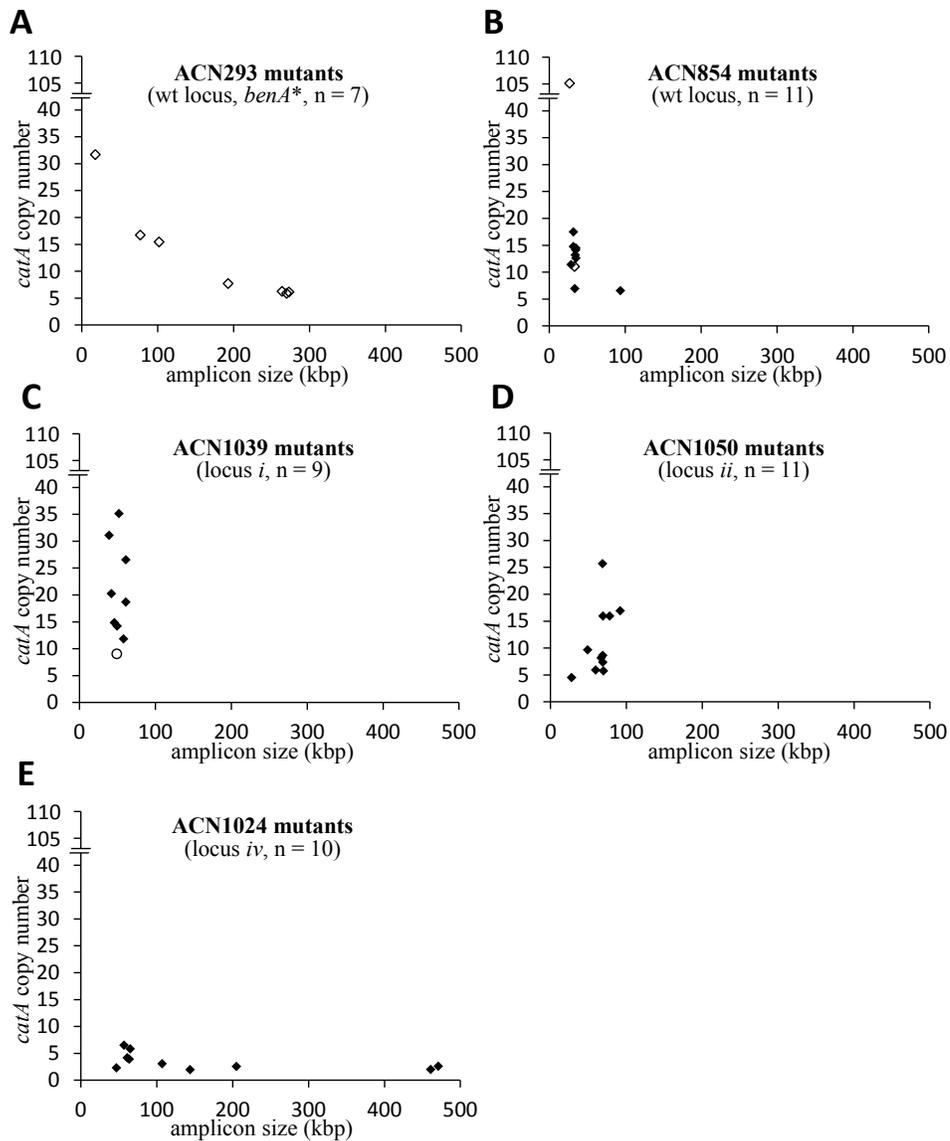


Figure A.4. Relationship between *catA* copy number and amplicon size. Each data point represents the analysis of an individual gene amplification mutant derived from the parent strain indicated. Closed diamond, selected on benzoate; open diamond, selected on anthranilate; open circle, selected on *cis,cis*-muconate. All mutants were isolated independently of one another. These traits are shown for mutants derived from (A) ACN293, (B) ACN854, (C) ACN1039, (D) ACN1050 and (E) ACN1024. The location of the *cat* genes in each parent strain is indicated parenthetically with reference to the locations shown in Fig. A.2.

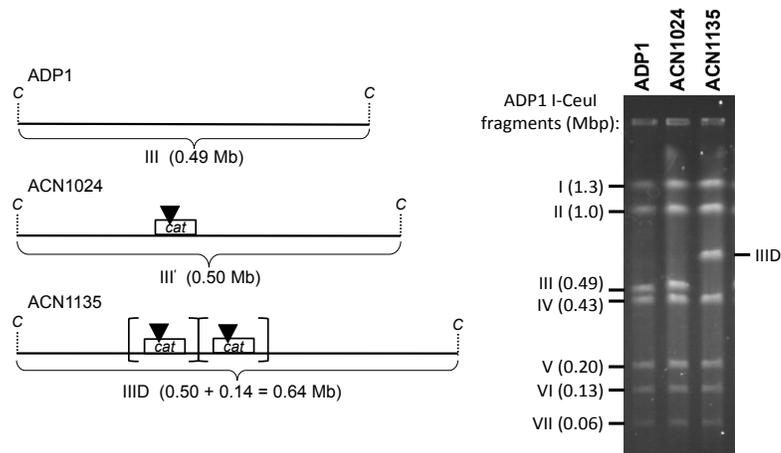


Figure A.5. Genome analysis of a Ben^+ duplication mutant (ACN1135). Left panel illustrates the size an I-CeuI fragment (III) from the wild-type (ADP1, top), the corresponding fragment (III') from the Ben^- parent strain (ACN1024, middle), and the corresponding fragment (IIID) from the ACN1024-derived Ben^+ duplication mutant (ACN1135). The amplicon size of ACN1135 was determined to be approximately 0.14 Mb (Fig. A.S3). Therefore, two copies of this *cat* gene amplicon (indicated by brackets) should increase fragment III' in size from 0.50 Mb to 0.64 Mb. A fragment of this size was observed in ACN1135 in the PFGE of I-CeuI-digested DNA (right, IIID). C indicates I-CeuI sites. Black triangles indicate the *catM*:: Ω K insertion. Sizes on the left of the gel image correspond to the I-CeuI fragment sizes in ADP1.

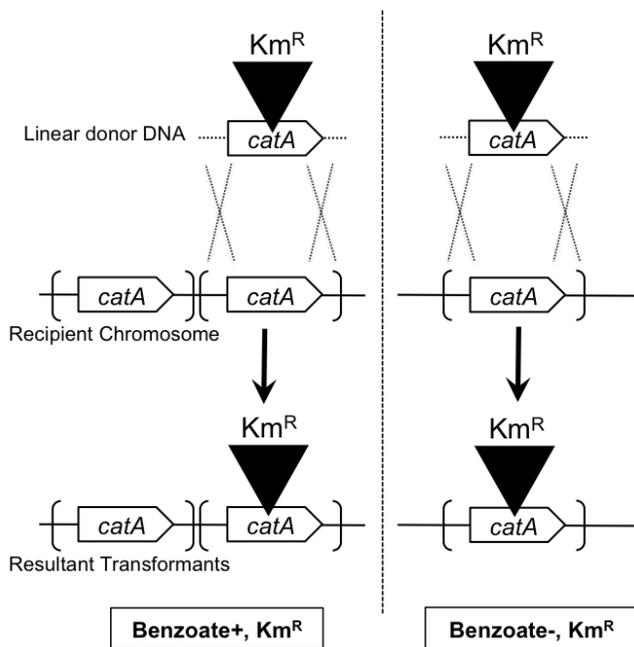


Figure A.6. Transformation assay to determine spontaneous duplication frequency. The source of donor DNA is a strain (Ben⁻ and Km^R) with a disrupted *catA* allele (top). Cell-free DNA from this strain is used to transform a population of cells grown under conditions that do not select for *catA* (middle). Most cells in this population have single copy of *catA* (right), although a small portion will have two (or more) copies due to GDA (left). Drug resistance (Km^R) is used to select transformants (bottom) in which the donor allele replaced *catA* on the recipient chromosome by homologous recombination (two crossover events, X). As described in the text, the ability (left) or inability (right) of these transformants to grow on benzoate as a sole carbon source can be used to assess the proportion of the starting population that carried a spontaneous duplication of *catA*.

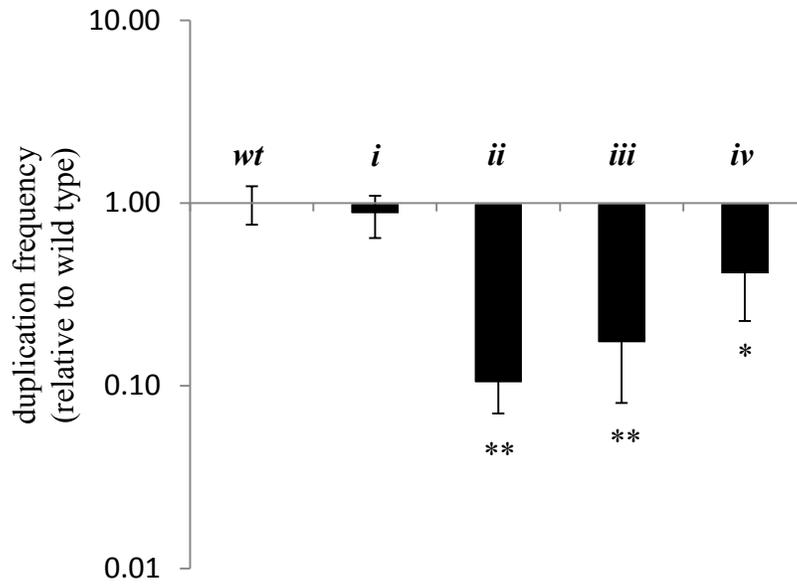


Figure A.7. Relative *catA* duplication frequencies at different loci. The duplication frequency of *catA* was measured at the native *catA* locus and nonnative loci *i-iv*. Results were normalized relative to the *catA* duplication frequency at the native locus (10^{-4}). Duplication frequencies significantly different from that of the native *catA* locus are indicated (* $p < 0.05$, ** $p < 0.01$).

Table A.1: Strains and plasmids used in this study.

<i>A. baylyi</i> strains used in this study		
Strain	Relevant characteristics ^a	Source
ADP1	Wild-type (BD413)	(Juni & Janik, 1969)
ACN147	<i>benM::ΩS4036, benA5147</i> ; promoter mutation allows <i>ben</i> -gene expression without BenM	(Collier <i>et al.</i> , 1998)
ACN293	<i>benM::ΩS4036, benA5147, ΔcatM5293</i> ; Ben ⁻ parent	(Reams & Neidle, 2003)
ACN389	<i>ΔbenM5389</i>	(Bundy <i>et al.</i> , 2002)
ACN814	<i>benM::ΩS4036, benA5147, catM::sacB-Km^R5613</i>	This study
ACN825	<i>benM::ΩS4036, benA5147, Δcat5825 (catA-catD)</i>	This study
ACN843	<i>ΔbenM5389, catM::sacB-Km^R5613</i>	This study
ACN854	<i>ΔbenM5389, ΔcatM5293</i> ; Ben ⁻ parent	This study
ACN868- ACN869	ACN854-derived, Ben ⁺	This study
ACN871- ACN875	ACN293-derived, Ant ⁺	This study
ACN877	ACN293-derived, Ant ⁺	This study
ACN883	ACN293-derived, Ant ⁺	This study
ACN896	ACN854-derived, Ben ⁺	This study
ACN898	ACN854-derived, Ant ⁺	This study
ACN899	ACN854-derived, Ben ⁺	This study
ACN906	ACN854-derived, Ben ⁺	This study
ACN950	ACN854-derived, Ben ⁺	This study
ACN951	ACN854-derived, Ant ⁺	This study
ACN956	ACN854-derived, Ben ⁺	This study
ACN1007	ACN854-derived, Ben ⁺	This study
ACN1023	<i>benM::ΩS4036, benA5147, Δcat5825 (catA-catD)</i> ; <i>cat</i> region (1439430-1449722) ^b in locus <i>iv</i> (ACIAD3593); constructed with pBAC850	This study
ACN1024	<i>benM::ΩS4036, benA5147, Δcat5825; cat</i> region (1439430-1449722) ^b in locus <i>iv</i> (ACIAD3593), <i>catM::ΩK5541</i> ; Ben ⁻ parent	This study
ACN1025	<i>catA::ΩK51025</i>	This study
ACN1026- ACN1029	ACN1024-derived, Ben ⁺	This study
ACN1032- ACN1033	ACN1024-derived, Ben ⁺	This study
ACN1036	<i>benM::ΩS4036, benA5147, Δcat5825 (catA-catD)</i> ; <i>cat</i> region (1439430-1449722) ^b in locus <i>i</i> (ACIAD0982, <i>vanK</i>); constructed with pBAC874	This study

ACN1037	<i>benM</i> :: Ω S4036, <i>benA5147</i> , Δ <i>cat5825</i> ; <i>cat</i> region (1439430-1449722) ^b in locus <i>iv</i> (ACIAD3593), <i>catA</i> :: Ω K51025	This study
ACN1038	<i>benM</i> :: Ω S4036, <i>benA5147</i> , Δ <i>cat5825</i> ; <i>cat</i> region (1439430-1449722) ^b in locus <i>i</i> (ACIAD0982, <i>vanK</i>), <i>catA</i> :: Ω K51025	This study
ACN1039	<i>benM</i> :: Ω S4036, <i>benA5147</i> , Δ <i>cat5825</i> (<i>catA-catD</i>); <i>cat</i> region (1439430- 1449722) ^b in locus <i>i</i> (ACIAD0982, <i>vanK</i>), <i>catM</i> :: Ω K5541; Ben ⁻ parent	This study
ACN1040	ACN1039-derived, CCM ⁺	This study
ACN1046	<i>benM</i> :: Ω S4036, <i>benA5147</i> , Δ <i>cat5825</i> (<i>catA-catD</i>); <i>cat</i> region (1439430- 1449722) ^b in locus <i>ii</i> (ACIAD2822); constructed with pBAC876	This study
ACN1048	<i>benM</i> :: Ω S4036, <i>benA5147</i> , Δ <i>cat5825</i> ; <i>cat</i> region (1439430-1449722) ^b in locus <i>ii</i> (ACIAD2822), <i>catA</i> :: Ω K51025	This study
ACN1050	<i>benM</i> :: Ω S4036, <i>benA5147</i> , Δ <i>cat5825</i> ; <i>cat</i> region (1439430-1449722) ^b in locus <i>ii</i> (ACIAD2822), <i>catM</i> :: Ω K5541; Ben ⁻ parent	This study
ACN1056- ACN1058	ACN1050-derived, Ben ⁺	This study
ACN1060	<i>benM</i> :: Ω S4036, <i>benA5147</i> , Δ <i>cat5825</i> (<i>catA-catD</i>); <i>cat</i> region (1439430- 1449722) ^b in locus <i>iii</i> (ACIAD2969); constructed with pBAC872	This study
ACN1061	<i>benM</i> :: Ω S4036, <i>benA5147</i> , Δ <i>cat5825</i> ; <i>cat</i> region (1439430-1449722) ^b in locus <i>iii</i> (ACIAD2969), <i>catM</i> :: Ω K5541; Ben ⁻ parent	This study
ACN1080	<i>benM</i> :: Ω S4036, <i>benA5147</i> , Δ <i>cat5825</i> ; <i>cat</i> region (1439430-1449722) ^b in locus <i>iii</i> (ACIAD2969), <i>catA</i> :: Ω K51025	This study
ACN1101- ACN1102	ACN1039-derived, Ben ⁺	This study
ACN1105- ACN1106	ACN1050-derived, Ben ⁺	This study
ACN1114	ACN1024-derived, Ben ⁺	This study
ACN1118	ACN854-derived, Ben ⁺	This study
ACN1123	ACN854-derived, Ben ⁺	This study
ACN1127- ACN1128	ACN1039-derived, Ben ⁺	This study
ACN1130-	ACN1039-derived, Ben ⁺	This study

ACN1131		
ACN1135- ACN1136	ACN1024-derived, Ben ⁺	This study
ACN1140- ACN1141	ACN1039-derived, Ben ⁺	This study
ACN1161- ACN1166	ACN1050-derived, Ben ⁺	This study
ACN1182	ACN1024-derived, Ben ⁺	This study
ACN1190	ACN1024-derived, Ben ⁺	This study
Plasmids used in this study		
Plasmid	Relevant characteristics	Source
pCR2.1- TOPO	Ap ^R , Km ^R ; PCR cloning vector	Invitrogen
pET-21b	Ap ^R ; cloning vector	Novagen
pZER0-2	Km ^R ; Zero background cloning vector	Invitrogen
pIB3	Ap ^R ; partial <i>catM</i> (1441380-1444461) ^b	(Neidle & Ornston, 1986)
pIB1361	Ap ^R ; partial <i>ben</i> and <i>cat</i> region (1437679-1444461) ^b	(Neidle & Ornston, 1986)
pIGG6	ApR; Δ <i>catM5293</i>	(Reams & Neidle, 2004a)
pPAN4	Ap ^R ; <i>catBCIJFD</i>	(Shanley <i>et al.</i> , 1986)
pRMJ1	Ap ^R , Km ^R ; <i>sacB</i> -Km ^R cassette	(Jones & Williams, 2003)
pUI1637	Ap ^R , Km ^R ; Ω Km ^R cassette	(Eraso & Kaplan, 1994)
pBAC6	Ap ^R , Km ^R ; <i>catM</i> :: Ω K5541	(Ezezika <i>et al.</i> , 2006)
pBAC184	Ap ^R ; partial <i>cat</i> region with an internal ClaI deletion (1441380-1442211; 1447467-1449722) ^b in pUC19	This study
pBAC707	Ap ^R ; <i>catM</i> region (1443514-1444461) ^b with internal multiple cloning site between 1444251 and 1444252 ^b in pUC19	This study
pBAC708	Ap ^R , Km ^R ; <i>catM</i> region (1443514-1444461) ^b with <i>sacB</i> -Km ^R cassette in Sall site within multiple cloning site between 1444251 and 1444252 ^b (<i>catM</i> :: <i>sacB</i> -Km ^R 5613) in pUC19	This study
pBAC761	Ap ^R ; Δ <i>cat5825</i> (1439140-1439923; 1449545-1450109) ^b in pUC19	This study
pBAC763	Ap ^R ; <i>catA</i> (1439463-1440751) ^b in pUC19 with an internal BamHI site engineered at 1440105 ^b	This study
pBAC832	Ap ^R ; <i>cat</i> region with an internal ClaI deletion (1439438-	This study

	1442211; 1447467-1449722) ^b in pUC19	
pBAC840	Ap ^R ; contains <i>cat</i> region with internal ClaI deletion (1439438-1442211; 1447467-1449722) ^b and multiple cloning region from pET21b in pUC19	This study
pBAC841	Universal <i>cat</i> region relocation plasmid; contains <i>cat</i> DNA (1439438-1449722) ^b	This study, Fig. A.S1
pBAC844	3509555-3510831 ^b (target-L) cloned into BamHI/PstI of pBAC841; intermediate plasmid for relocation to locus <i>iv</i> (ACIAD3593)	This study
pBAC845	3510884-3511786 ^b (target-R) cloned into SacI/XhoI of pBAC840; intermediate plasmid for relocation to locus <i>iv</i> (ACIAD3593)	This study
pBAC846	3509555-3510831 ^b (target-L) and 3510884-3511786 ^b (target-R) cloned into BamHI/PstI and SacI/XhoI of pBAC840; intermediate plasmid for relocation to locus <i>iv</i> (ACIAD3593)	This study
pBAC850	3509555-3510831 ^b (target-L) and 3510884-3511786 ^b (target-R) cloned into BamHI/PstI and SacI/XhoI of pBAC841; for relocation to locus <i>iv</i> (ACIAD3593)	This study, Fig. A.S1
pBAC852	Ap ^R , Km ^R , <i>catA</i> ::ΩK51025; ΩK in BamHI site of pBAC763	This study
pBAC857	ACN868 junction ^c ; EcoRI insert (1432131-1431778; 1466991-1464347) ^b	This study
pBAC858	ACN869 junction ^c ; EcoRI insert (1432131-1431838; 1467382-1464347) ^b	This study
pBAC859	ACN896 junction ^c ; EcoRI insert (1437229-1433352; 1468488-1464305) ^b	This study
pBAC860	ACN898 junction ^c ; HindIII insert (1437877-1434048; 1461133-1460612) ^b	This study
pBAC861	ACN899 junction ^c ; HindIII insert (1430928-1428708; 1475753-1475359) ^b	This study
pBAC862	ACN950 junction ^c ; EcoRI insert (1464347-1465535; 1433380-1437229) ^b	This study
pBAC863	ACN951 junction ^c ; EcoRI insert (1450761-1455727; 1421717-1437229) ^b	This study
pBAC864	ACN956 junction ^c ; EcoRI insert (1437229-1432834; 1462267-1450763) ^b	This study
pBAC865	ACN906 junction ^c ; EcoRI insert (1432131-1431838; 1467382-1464347) ^b	This study
pBAC866	ACN1007 junction ^c ; EcoRI insert (1437229-1433380; 1465535-1464347) ^b	This study
pBAC867	ACN1118 junction ^c ; EcoRI insert (1450761-1455727; 1421717-1437229) ^b	This study
pBAC868	ACN1123 junction ^c ; EcoRI insert (1511262-1516254; 1421888-1432129) ^b	This study
pBAC870	2898691-2899659 ^b (target-R) in SacI/NotI of pBAC841; intermediate plasmid for relocation to locus <i>iii</i> (ACIAD2969)	This study
pBAC871	2761112-2761860 ^b (target-R) in SacI/XhoI of pBAC841; intermediate plasmid for relocation to locus <i>ii</i> (ACIAD2822)	This study
pBAC872	2897617-2898690 ^b (target-L) and 2898691-2899659 ^b	This study,

	(target-R) in BamHI/Pst I and SacI/NotI of pBAC841; for relocation to locus <i>iii</i> (ACIAD2969)	Fig. A.S1
pBAC873	968309-969139 ^b (target-L) in BamHI/PstI of pBAC841; intermediate plasmid for relocation to locus <i>i</i> (ACIAD0982, <i>vanK</i>)	This study
pBAC874	968309-969139 ^b (target-L) and 969600-970784 ^b (target-R) in BamHI/PstI and SacI/XhoI of pBAC841; for relocation to locus <i>i</i> (ACIAD0982, <i>vanK</i>)	This study, Fig. A.S1
pBAC876	2759824-2761080 ^b (target-L) and 2761112-2761860 ^b (target-R) in BamHI/PstI and SacI/XhoI of pBAC841; for relocation to locus <i>ii</i> (ACIAD2822)	This study, Fig. A.S1

^aAp^R, ampicillin resistant; ΩK, omega cassette conferring Km^R (Eraso & Kaplan, 1994); ΩS, omega cassette conferring resistance to streptomycin and spectinomycin; *sacB*-Km^R, dual selection cassette containing a counterselectable marker and kanamycin resistance cassette (Jones & Williams, 2003), CCM⁺ indicates growth on *cis,cis*-muconate.

^bPosition in the ADP1 genome sequence according to GenBank entry CR543861.

^cJunction DNA was cloned in pZErO-2

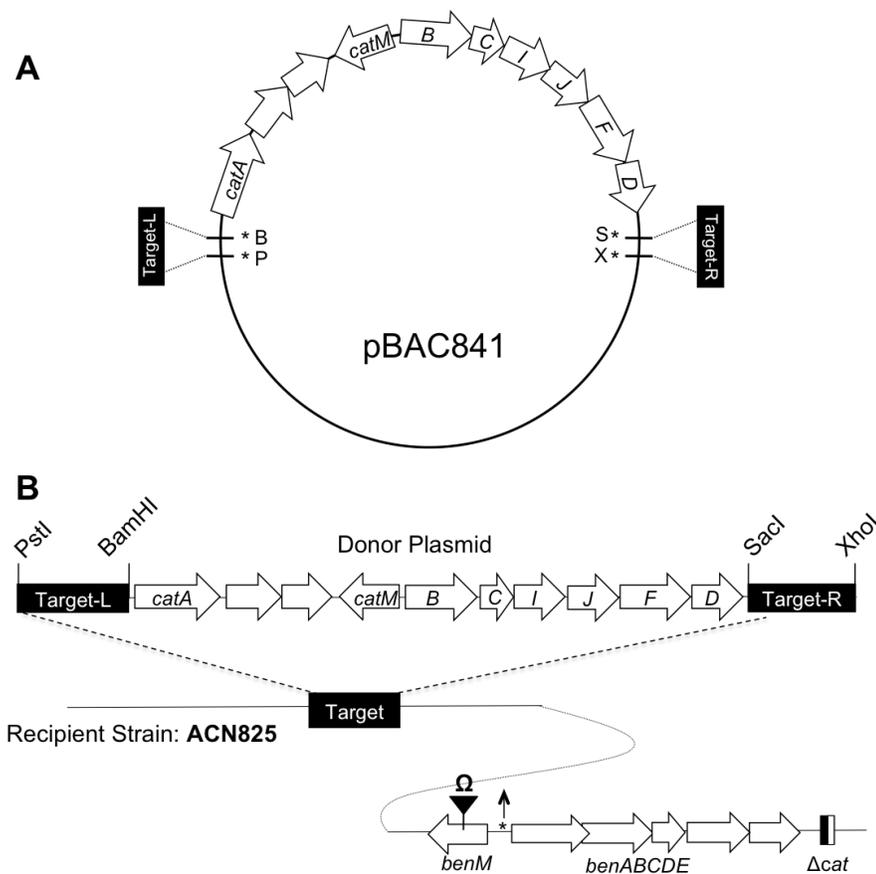


Figure A.S1. Genetic engineering allows relocation of *cat* gene region to loci throughout the ADP1 genome. (A) Plasmid, pBAC841, was engineered to serve as a universal relocation plasmid and encodes an entire functional *cat* gene region. DNA from the target locus is amplified by PCR and cloned into pBAC841 using unique restriction sites (P= PstI; B=BamHI; S=SacI; X=XhoI). (B) The *cat* gene cassette can be inserted into any *A. baylyi* target locus via allelic exchange. In the recipient strain, ACN825, *benM* is not expressed due to insertional inactivation with an antibiotic marker (Ω), the *benABCDE* operon is expressed at high levels without any transcriptional regulator due to a promoter mutation (*), and the native *cat* region has been deleted by overlap extension PCR.

ACN871 (270 kbp amplicon)
TTGTTGGTATAGGTGTATTTAGTCA AACGTTGTAAAGGATACTGATTGAG 1708344
TTGTTGGTATAGGTGTATTTAGTCA TTTGTTCCTCCAACATTAGTTTAAAT Junction
CACCCGGGACAGCATTGTTGGTTAC TTTGTTCCTCCAACATTAGTTTAAAT 1438290

ACN872 (193 kbp amplicon)
TTTTCAAATAATAAGTTGCAGCTTT TTTAAAGGCCCATCTGGAATATCG 1632114
TTTTCAAATAATAAGTTGCAGCTTT AATTTGGCGCTGCCGCTAATTTTGG Junction
AGAGTGGTCATGGTTTTTCAACCTTA AATTTGGCGCTGCCGCTAATTTTGG 1438678

ACN873 (18 kbp amplicon)
ATGCAGCTGCTGACAACACGCCCTG ATTGCACATGCTGGTGCCCAATAGT 1453820
ATGCAGCTGCTGACAACACGCCCTG TGCTGCGGAAAAAGAATTTGCATCG Junction
GGCTCAAAAGTATGAAGCAAGCGAT TGCTGCGGAAAAAGAATTTGCATCG 1435538

ACN874 (76 kbp amplicon)
GATAACTTAAGAGTTGCCAAGCAA **GA** TGCAACGGAGCAAGAGATGAAAGA 1511953
GATAACTTAAGAGTTGCCAAGCAA **GA** GGAGTTCCTGCCTGTCAGGCTGG Junction
GGAAATGGCGACGCCAGACGATCTT **GA** GGAGTTCCTGCCTGTCAGGCTGG 1435330

ACN875 (101 kbp amplicon)
TGTTCTTTCTAGCTGGCTATCCACA **T** CTTGAAGAGTCATACCGTCTTGGTG 1537635
TGTTCTTTCTAGCTGGCTATCCACA **T** TGGATGCCCCGCTGGGACGATAACG Junction
AATGTTATGCACCTCAAGCCTCATT **T** TGGATGCCCCGCTGGGACGATAACG 1435711

ACN877 (264 kbp amplicon)
GTTTATACTGGCGATGACCTGTGGT **A** TTTGTGCTGGATCGAATTATTATAA 1703635
GTTTATACTGGCGATGACCTGTGGT **A** GATGACATATCACTCAGCTTAATTT Junction
TCTAAGGCGAATTTTGTTTTTTAAG **A** GATGACATATCACTCAGCTTAATTT 1439285

ACN883 (273 kbp amplicon)
TACTCTGGAAGTGAATCAAAAAAT **TAT** AAAAAATGATCAGACAAGGGAAT 1709069
TACTCTGGAAGTGAATCAAAAAAT **TAT** TTTGGCATGTCACGCTATGTAAT Junction
TTCCGCTATAAAAAACAGTTACAGC **TAT** TTTGGCATGTCACGCTATGTAAT 1435971

Figure A.S2: Duplication junction sequences in ACN293 mutants selected on anthranilate. The center line of text represents the DNA junction from the amplification strain, with top and bottom lines of indicating the parental downstream and upstream sequences, respectively. The position of the first nucleotide of each line is given at right and corresponds to the published *A. baylyi* genome sequence (GenBank CR543861). Identical nucleotides at the DNA junction are shown in bold, underlined text. Homology in the vicinity of the junction is also shown (underlined).

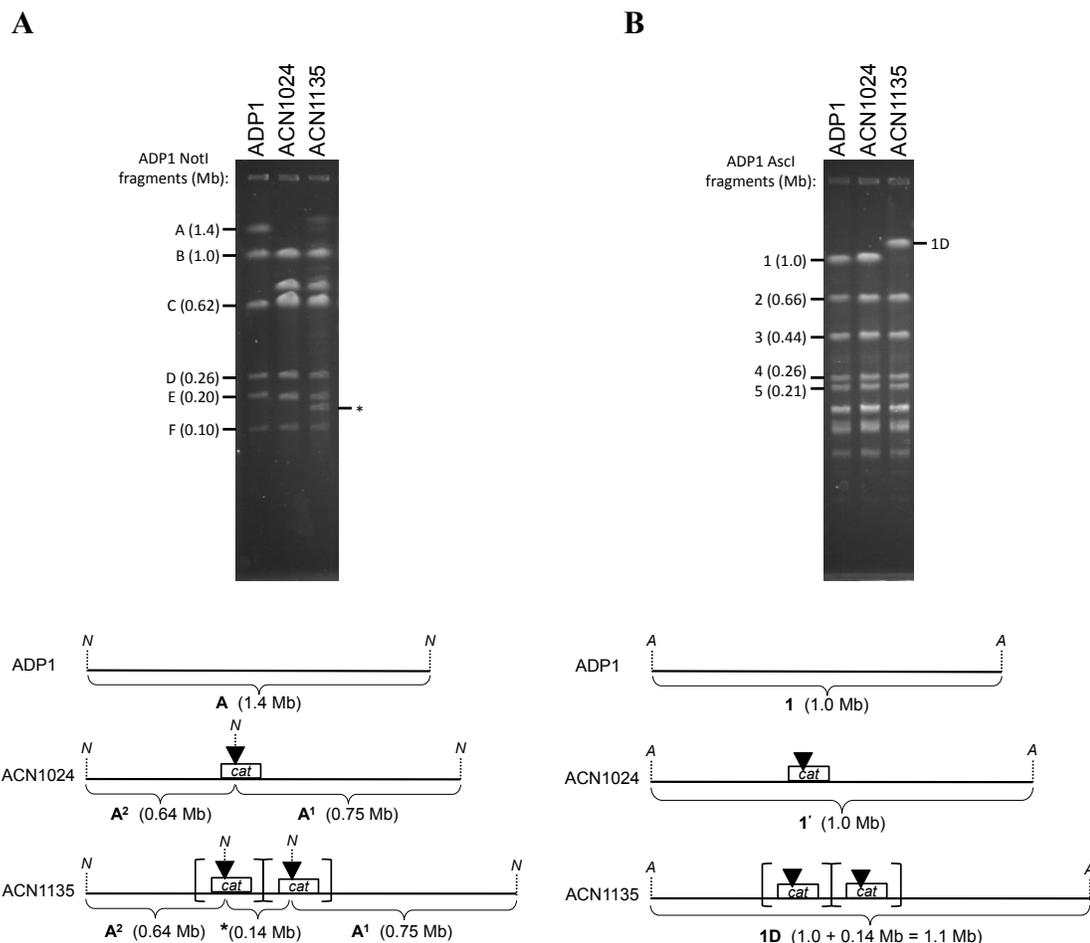


Figure A.S3. Genome analysis of a duplication mutant. **(A)** If ACN1135, derived from ACN1024, contains a duplication of the amplicon, there are two copies of the NotI site (*N*) that disrupts *catM*. Digestion of ACN1135 DNA with NotI generates a fragment that corresponds to the size of the amplicon (indicated by *). This amplicon-associated fragment is absent from the wild type or ACN1024 parent. In PFGE with NotI digested genomic DNA, this fragment is observed to be 0.14 Mb and is absent from ADP1 and ACN1024. **(B)** Based on the unique NotI fragment size in A, the ACN1135 AscI fragment 1D size in a duplication mutant is predicted. PFGE analysis of AscI digested genomic DNA is shown. The estimated size of fragment 1D in ACN1135 is consistent with the conclusion that this strain is a duplication mutant. *N* and *A* indicate NotI and AscI sites, respectively. Black triangles indicate the *catM*::ΩK insertion. Braces indicate the amplicon in ACN1135. For each PFGE, the sizes on the left of the gel image correspond to the fragment sizes in ADP1. In B, only the largest 5 of the 19 expected AscI fragments in ADP1 are indicated.

Table A.S1. Characteristics of gene amplification mutants in this study.

Strain	Parent Strain	Carbon Source ^a	Amplicon Size (kbp)	Amplicon Copy Number	Amount of amplified DNA(kbp) ^b
ACN868	ACN854	benzoate	34	14	486
ACN869	ACN854	benzoate	35	13	436
ACN871	ACN293	anthranilate	270	6	1585
ACN872	ACN293	anthranilate	193	8	1485
ACN873	ACN293	anthranilate	18	32	570
ACN874	ACN293	anthranilate	76	17	1270
ACN875	ACN293	anthranilate	101	15	1557
ACN877	ACN293	anthranilate	264	6	1649
ACN883	ACN293	anthranilate	273	6	1670
ACN896	ACN854	benzoate	34	13	448
ACN898	ACN854	anthranilate	27	105	2796
ACN899	ACN854	Benzoate	46	n/d ^c	n/d ^c
ACN906	ACN854	benzoate	35	14	502
ACN950	ACN854	benzoate	31	17	548
ACN951	ACN854	anthranilate	33	11	367
ACN956	ACN854	Benzoate	29	11	326
ACN1007	ACN854	benzoate	31	15	463
ACN1026	ACN1024	benzoate	64	4	251
ACN1027	ACN1024	benzoate	107	3	326
ACN1028	ACN1024	benzoate	61	4	253
ACN1032	ACN1024	benzoate	47	2	108
ACN1033	ACN1024	benzoate	205	3	518
ACN1040	ACN1039	<i>cis,cis</i> muconate	49	9	444
ACN1056	ACN1050	benzoate	69	7	507
ACN1057	ACN1050	benzoate	69	9	595
ACN1058	ACN1050	benzoate	67	8	545
ACN1101	ACN1039	benzoate	46	15	681
ACN1102	ACN1039	benzoate	61	19	1138
ACN1105	ACN1050	benzoate	70	6	402
ACN1106	ACN1050	benzoate	49	10	473
ACN1114	ACN1024	benzoate	65	6	379
ACN1118	ACN854	benzoate	33	7	230
ACN1123	ACN854	benzoate	94	7	611
ACN1127	ACN1039	benzoate	39	31	1213
ACN1128	ACN1039	benzoate	42	20	852
ACN1130	ACN1039	benzoate	52	35	1826
ACN1131	ACN1039	benzoate	61	27	1619
ACN1135	ACN1024	benzoate	144	2	282
ACN1136	ACN1024	benzoate	461	2	924
ACN1140	ACN1039	benzoate	50	14	705
ACN1141	ACN1039	benzoate	58	12	684
ACN1161	ACN1050	benzoate	69	26	1768
ACN1162	ACN1050	benzoate	60	6	351

ACN1163	ACN1050	benzoate	69	16	1105
ACN1164	ACN1050	benzoate	78	16	1242
ACN1165	ACN1050	benzoate	95	17	1606
ACN1166	ACN1050	benzoate	28	5	127
ACN1182	ACN1024	benzoate	57	6	370
ACN1190	ACN1024	benzoate	471	3	1211

^a Indicates the carbon source on which the gene amplification mutant was isolated.

^b Amount of amplified DNA is inferred from the amplicon size and copy number.

Differences between the amount indicated and the apparent product of amplicon size and copy number are due to rounding.

^c not determined.

Table A.S2. Pearson's correlation coefficient analysis of the relationship between amplicon size and copy number.

parent strain	r ^a	p-value ^b
ACN293	-0.92	0.0030
ACN854	-0.26	(0.43)
ACN1024	-0.52	(0.06)
ACN1050	0.50	(0.058)
ACN1039	-0.12	(0.38)

^ar-value is Pearson's correlation coefficient. Positive r-value indicates a direct correlation between amplicon size and copy number. Negative r-value indicates an inverse correlation.

^bbold indicates $p < 0.05$, parentheses indicate $p \geq 0.05$.

Table A.S3. Mann-Whitney analysis of pair-wise differences in *catA* copy number (p-values shown).^a

ACN293	ACN854	ACN1024	ACN1050	ACN1039	
*	(0.47)	0.0021	(1.0)	(0.11)	ACN293
	*	0.00010	(0.33)	(0.095)	ACN854
		*	0.00050	0.00030	ACN1024
			*	0.023	ACN1050
				*	ACN1039

^abold indicates $p < 0.05$, parentheses indicate $p \geq 0.05$.

Table A.S4. Mann-Whitney analysis of pair-wise differences in amplicon size (p-values shown).^a

ACN293	ACN854	ACN1024	ACN1050	ACN1039	
*	0.019	(0.59)	0.024	0.020	ACN293
	*	0.00050	0.0085	0.0024	ACN854
		*	(0.42)	0.0048	ACN1024
			*	0.014	ACN1050
				*	ACN1039

^abold indicates $p < 0.05$, parentheses indicate $p \geq 0.05$.