

Multiple Pathways of Duplication Formation with and Without Recombination (RecA) in *Salmonella enterica*

Andrew B. Reams, Eric Kofoid, Elisabeth Kugelberg, and John R. Roth¹
Department of Microbiology, University of California, Davis, California 95616

ABSTRACT Duplications are often attributed to “unequal recombination” between separated, directly repeated sequence elements (>100 bp), events that leave a recombinant element at the duplication junction. However, in the bacterial chromosome, duplications form at high rates (10^{-3} – 10^{-5} /cell/division) even without recombination (RecA). Here we describe 1800 spontaneous *lac* duplications trapped nonselectively on the low-copy *F'*₁₂₈ plasmid, where *lac* is flanked by direct repeats of the transposable element IS3 (1258 bp) and by numerous quasipalindromic REP elements (30 bp). Duplications form at a high rate (10^{-4} /cell/division) that is reduced only about 11-fold in the absence of RecA. With and without RecA, most duplications arise by recombination between IS3 elements (97%). Formation of these duplications is stimulated by IS3 transposase (Tnp) and plasmid transfer functions (TraI). Three duplication pathways are proposed. First, plasmid dimers form at a high rate stimulated by RecA and are then modified by deletions between IS3 elements (resolution) that leave a monomeric plasmid with an IS3-flanked *lac* duplication. Second, without RecA, duplications occur by single-strand annealing of DNA ends generated in different sister chromosomes after transposase nicks DNA near participating IS3 elements. The absence of RecA may stimulate annealing by allowing chromosome breaks to persist. Third, a minority of *lac* duplications (3%) have short (0–36 bp) junction sequences (SJ), some of which are located within REP elements. These duplication types form without RecA, Tnp, or Tra by a pathway in which the palindromic junctions of a tandem inversion duplication (TID) may stimulate deletions that leave the final duplication.

GENE duplications are among the first mutations for which a physical basis was known. The Bar-Eye mutation of *Drosophila* was discovered in 1914 (Tice 1914) and shown genetically and cytologically in 1936 to be a duplication (Bridges 1936; Muller 1936)—well before the Watson–Crick model for DNA. The attached-X mutation was discovered even earlier and is essentially an inversion duplication (Morgan 1925). Recently, duplications have been shown to be exceedingly common polymorphisms in human populations (Conrad *et al.* 2010) and somatic amplifications play an important role in origins of malignancy and resistance to cancer therapies (Lu *et al.* 2008; Zhang *et al.* 2009; Conrad *et al.* 2010). Gene duplications also play a major role in the evolution of new genes (Ohno 1970; Bergthorsson

et al. 2007). Despite the high frequency and biological importance of duplications, the process by which they form remains uncertain.

Distinct models have been suggested for duplication formation in various genetic systems. In *Drosophila*, duplications often have transposable elements between copies (Green 1985; Tsubota *et al.* 1989). In humans, duplications most commonly arise between extensive sequence repeats (Redon *et al.* 2006) and are sometimes promoted by palindromic sequence elements (Conrad *et al.* 2010). A major question is how often duplications form by unequal-crossing-over model and how often by one of the multiple alternative pathways.

In bacteria, several properties of duplications are apparent:

1. Duplications of any particular gene form at extremely high rates (10^{-3} – 10^{-5} /cell/division) (Anderson and Roth 1981; Reams *et al.* 2010).
2. In unselected populations, the frequency of duplications and higher amplifications rapidly approach a steady state due to a balance between their high formation rate and their frequent loss and fitness cost (Reams *et al.* 2010).

Copyright © 2012 by the Genetics Society of America
doi: 10.1534/genetics.112.142570

Manuscript received June 2, 2012; accepted for publication July 24, 2012
Available freely online through the author-supported open access option.

¹Corresponding author: Department of Microbiology, University of California, Davis, 314 Briggs Hall, 1 Shields Ave., Davis, CA 95616. E-mail: jrroth@ucdavis.edu

- Duplications form most often between large direct sequence repeats >200 bp (Anderson and Roth 1981; Flores *et al.* 2000; Goldfless *et al.* 2006).
- Duplications often arise by recombination between directly oriented preexisting copies of transposable elements (Morris and Rownd 1974; Clugston and Jessop 1991; Haack and Roth 1995; Ogawa and Miyashita 1995; Reams and Neidle 2003; Reams and Neidle 2004b; Nicoloff *et al.* 2007).
- Even when substantial flanking repeats are involved, the rate of chromosomal duplication formation depends only weakly on RecA-mediated homologous recombination (Reams *et al.* 2010), which seems contrary to the idea of unequal crossing over.
- Short adjacent repeats in high-copy plasmids can recombine to form deletions and duplications without RecA (Morag *et al.* 1999; Bzymek and Lovett 2001; Reams and Neidle 2004a; Goldfless *et al.* 2006; Gore *et al.* 2006). These rare exchanges form by reciprocal exchanges at stalled replication forks leading to formation of a plasmid dimer with a deletion in one copy and a corresponding duplication in the other (Dianov *et al.* 1991; Mazin *et al.* 1991; Lovett *et al.* 1993).
- Following prolonged growth under selection for increased gene copy number, amplifications (more than two copies) whose structure differs from the duplication types that predominate in unselected cultures are recovered (Kugelberg *et al.* 2006; Kugelberg *et al.* 2010). Selected amplifications have tiny junction sequences (0–10 bp) that seem unlikely to form by recombination (Reams and Neidle 2004a; Kugelberg *et al.* 2006, 2010).
- Some selected amplifications (20%) have repeated tandem inversion duplications (TID), in which a parental sequence element (*e.g.*, ABCD) is repeated three times in alternating orientation ABCD–D' C' B' A'–ABCD and amplified by exchanges between flanking direct repeats. The parent sequence ABCD can be >10 kb and the initial TID junctions are flanked by extensive perfectly complementary sequences that can be rendered asymmetric (ABCD–C' B' A'–BCD) by junction deletions (Kugelberg *et al.* 2010; Lin *et al.* 2011).

In the experiments presented here, duplications of the *lacZ* gene were isolated nonselectively on a low-copy bacterial plasmid (F'₁₂₈) with several features that contribute to duplication formation. These include insertion sequence (IS) elements, a conjugative replication origin, and multiple palindromic repetitive extragenic palindromic (REP) elements. Pathways of duplication formation are inferred from the frequency of duplication types that appear in various mutant backgrounds. We first describe formal aspects of gene duplication that are relevant to these results.

Mechanisms of duplication formation

These mechanisms are intended as background to the experiments described here and will be referred to in describing this work.

Unequal crossing over: Most duplications are thought to form by a single “unequal crossover”—*i.e.*, homologous re-

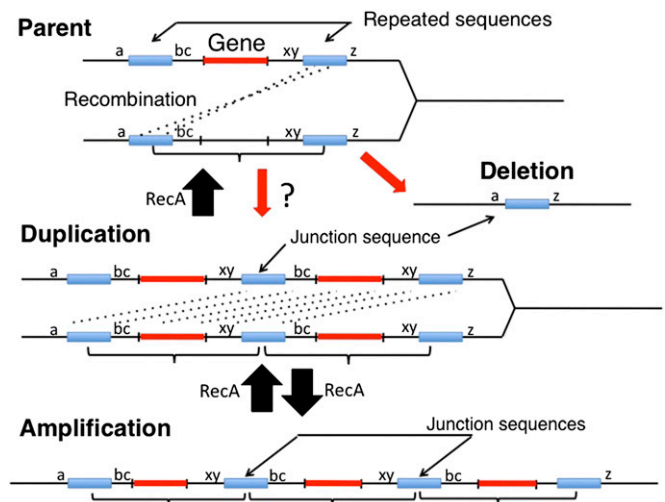


Figure 1 Duplication by unequal crossing over. This conventional view of duplication formation involves standard recombination events between separated sequence elements embedded in nonhomologous surrounding sequence (top). Recombination between the elements generates a duplication with a copy of the recombining element at the junction between duplicated regions and at the site of a corresponding deletion (middle). Regardless of how a duplication forms, the repeated segment can be lost or further amplified (bottom) by standard RecA-dependent recombination between nonallelic copies at any point of their extensive shared homology (see center).

combination between separated direct sequence repeats. Figure 1 diagrams this event as it might occur between bacterial sister chromosomes. Recombination creates a tandem head-to-tail duplication in one sister and, if the exchange is reciprocal, a corresponding deletion in the other. In either case, a copy of the recombining sequence is left at the rearrangement join point. The idea that duplications arise by unequal homologous recombination is called into question by the fact that bacterial duplications are frequent even in recombination deficient (*recA*) mutant strains and often have junction sequences that seem too short to support standard recombination.

Duplication formation without RecA was observed in the bacterial chromosome (Anderson and Roth 1978; Reams *et al.* 2010), but has been most extensively studied by selecting rare deletions between adjacent short repeats (20–700 bp) in small high-copy plasmids. These deletions were recovered in plasmid dimers—duplications of the plasmid whose second copy had a duplication or triplication of the parental repeats. This suggested that deletions and duplications form by unequal reciprocal exchanges between sister chromosomes (Dianov *et al.* 1991; Mazin *et al.* 1991; Lovett *et al.* 1993). The independence of RecA appears to reflect close spacing of short repeats that allows annealing between single DNA strands present in both sister chromosomes near a stalled replication fork (Lovett 2004). When repeats are farther apart, as in standard chromosomal duplication events, the process of duplication may be somewhat different, although the low dependence on RecA has been noted in the bacterial chromosome (Reams *et al.* 2010). The duplicated regions described

here are 20–131 kb in size, most of which arise between repeats >1 kb.

Unequal crossing over is also believed to cause loss or higher amplification of a duplicated segment. These secondary events depend heavily on RecA (Lin *et al.* 1984; Goldberg and Mekalanos 1986; Petes and Hill 1988; Galitski and Roth 1997; Poteete 2009). See Figure 1.

Role of fork collapse in gene duplication: In considering Figure 1, a mechanistic problem is the source of DNA nicks or breaks that might initiate an exchange between short separated sequences. Most recombination events in unperturbed cells are thought to occur during repair of spontaneously collapsed replication forks (Kuzminov 1995). Replication of a nicked DNA duplex can lead to fork collapse leaving a double-strand break in one sister chromosome. Fork stalling can also leave single-strand ends that are prone to annealing. Mechanisms have been proposed for repair of forks using the strand-exchange protein RecA or by pairing of single-strand regions as described later (Kuzminov 1995). The problem is how to obtain discontinuities near the specific sequence repeats.

Duplication by transposition: The frequent involvement of mobile elements in duplication formation suggests a role for transposition. Duplications formed directly by replicative transposition have a copy of the transposable element at the join point, but are unique in having a novel sequence junction between one end of the element and the target sequence (indicated by a circle in Figure 2). A full replicative transposition would simultaneously generate a corresponding deletion allele with another novel junction.

Replicative transpositions with one novel sequence junction (Figure 2) underlie some gene amplifications in *Acinetobacter* (Reams and Neidle 2003; Reams and Neidle 2004a) and on the *E. coli* F' plasmid (Kugelberg *et al.* 2006; Kugelberg *et al.* 2010), but they are rare and have been detected only after long-term selection for increased copy number. Extremely rare amplified duplications have been seen with novel sequence junctions at both ends of a transposable element at the duplication junction (Reams and Neidle 2003; E. Kofoid, unpublished results). It is not clear how these duplications form, but their structure suggests that two transposition events were required. Conservative transposition events can in principle generate a duplication by providing separated sequences that can then serve as homology for standard unequal recombination (described in Figure 1), but we have never seen duplications that formed by this concerted transposition/recombination process. Formation of the duplication types described here is stimulated by a transposase, but involves no act of transposition (generates no novel sequence junctions).

Single-strand annealing: Formation of a duplication by single-strand annealing requires two simultaneous breaks in different chromosomes near repeated sequences that flank the region to be duplicated (Figure 3). Resection can reveal complementary single-strand sequences that pair to form

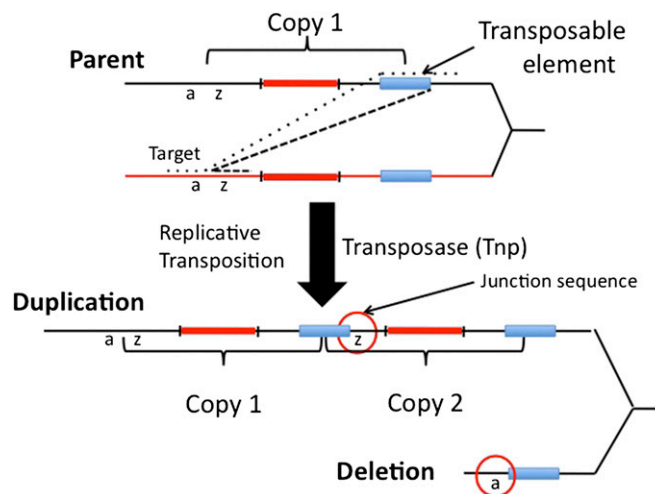


Figure 2 Duplication by replicative transposition. Replicative transposition can directly create a duplication with a copy of the transposable element at the rearrangement join point. The transposition event generates a novel sequence (circled above) at the juxtaposition of one end of the element and the transpositional target site. This event simultaneously generates a deletion that will appear in a different cell and have a novel sequence at the other end of the junction element. Duplications formed in this way have been observed in this system, but do not explain the involvement of transposase in the duplications described here. Conservative transposition can insert an element copy at a novel site for use by recombinational duplication, but we have not observed such events.

the duplication. This general annealing process has been demonstrated *in vivo* and *in vitro* and is independent of RecA (Mortensen *et al.* 1996; Bzymek and Lovett 2001; Liu *et al.* 2011). Duplication by annealing seems unlikely because two breaks must occur in different chromosomes, each near a recombining sequence repeat. The likelihood of annealing would increase if some feature of the repeated sequence increased the frequency of adjacent nicks that can be converted to double-strand breaks. This could happen at heavily transcribed genes (*e.g.*, *rrn*) if the nontemplate strand is prone to breakage. It could also happen at transposable elements, whose ends can be nicked by transposase as discussed below.

Strand slippage: Strand slippage was suggested to explain addition or removal of a base or two in formation of +1 and –1 frameshift mutations (Streisinger *et al.* 1966). In principle, the mechanism could also explain deletions and duplications between any pair of directly repeated sequences (Farabaugh *et al.* 1978). Slippage could underlie the tendency of large palindromic sequences to be removed during replication (Trinh and Sinden 1991; Leach 1994; Lovett 2004). If large single-strand regions are generated during replication, any included palindromic region might form a hairpin structure in either a nascent or template strand and favor slipped pairing of nearby regions to stimulate duplication or deletion of the palindromic region. Mechanisms such as this are probably responsible for the frequent duplication and deletion of small regions with palindromes (Sinden *et al.* 1991; Leach 1994; Bzymek and Lovett 2001; Lovett 2004;

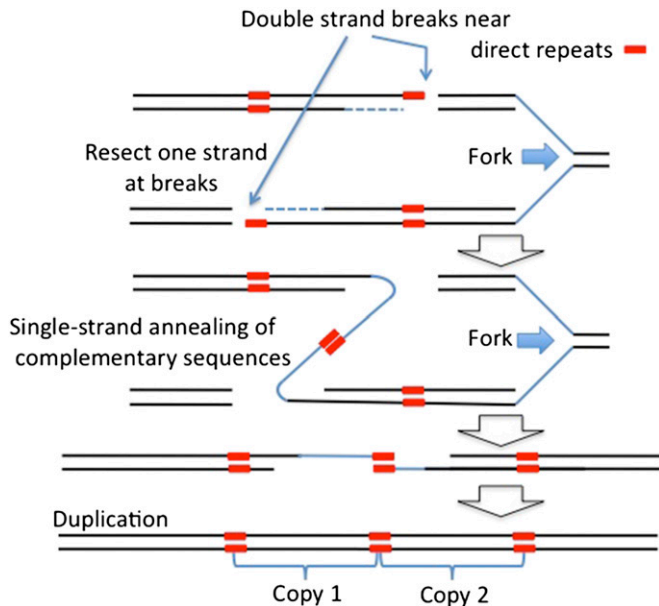


Figure 3 Duplication by single-strand annealing. In this diagram, breaks form at separated sites in different sister chromosomes behind a single replication fork. Breaks may form when a replication fork encounters a nick in its template or by conversion of a simple nick to a double-strand break. In the diagram, these breaks form near sequence repeats. This nicking could be caused by transposase or by fork stalling. The 5' strands at these breaks are resected, leaving 3'-single-strand extensions that could permit pairing of the repeats. Repair of the paired structure by nucleases and synthesis can lead to a duplication. Simultaneous single strands may coexist at a single collapsed replication fork and allow duplication between nearby repeats as diagramed previously (Lovett *et al.* 1993) or may form adjacent to transposable elements that are subject to transposase nicking.

Gore *et al.* 2006; Quiñones-Soto and Roth 2011), but seem unlikely to explain rearrangements of extensive chromosomal regions, as seen in the duplications described here.

Tandem inversion duplication: This duplication type was identified in Lac⁺ revertants isolated after prolonged selection of a leaky *lac* mutant for improved growth on lactose (Kugelberg *et al.* 2006; Kugelberg *et al.* 2010; Lin *et al.* 2011). Roughly 20% of the Lac⁺ revertants with an amplification of the region ABCD had multiple tandem copies of a basic structure that could be described as: ABCD–C'B'A'–BCD. This unit can be amplified by RecA-dependent unequal exchanges between the flanking tandem repeats (BCD). Note that the observed repeat junction sequences are asymmetric (nonpalindromic). It was proposed that the asymmetric structure (aTID) arises when deletions modify an initial symmetrical structure (sTID: ABCD–D'C'B'A'–ABCD) (see Figure 4A). Formation of this structure with perfectly paired adjacent sequences at each junction can form by snap-back repair from an initial short quasipalindromic sequence (Kugelberg *et al.* 2010). The palindrome stimulates deletions that generate the asymmetric (aTID) and tandem duplications with short junction sequences (see Figure 4B). Duplications of the initial symmetrical sTID type have been observed in yeast fol-

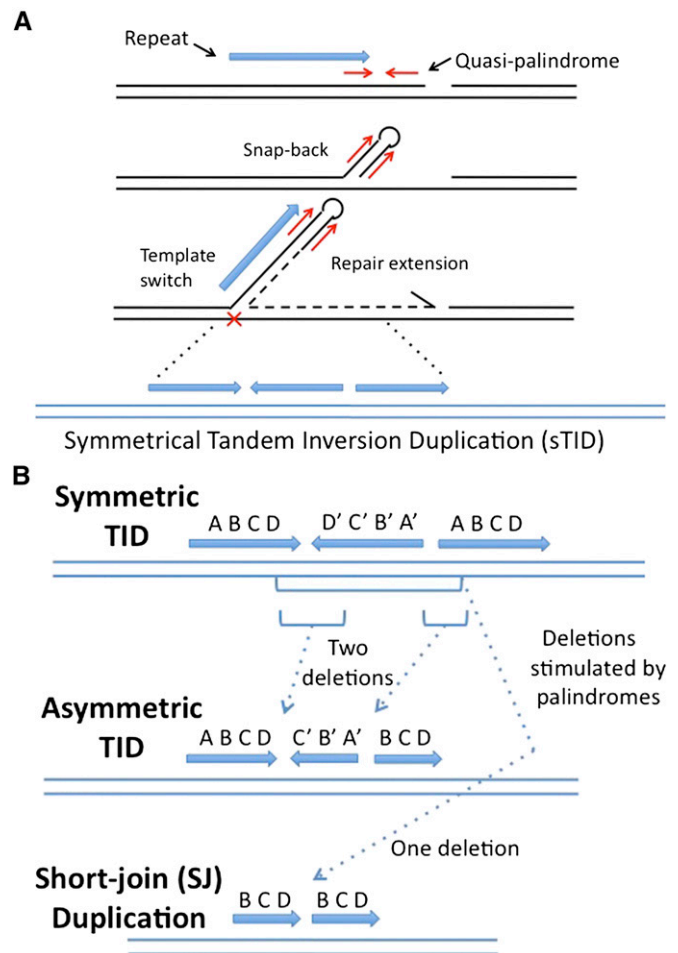


Figure 4 Tandem inversion duplications and their modification. Amplifications of this type were recovered in bacteria and yeast after prolonged growth under selection for additional gene copies (Araya *et al.* 2010; Kugelberg *et al.* 2010) and are also found in certain cancer cells, especially following exposure to cancer chemotherapy. Their formation is thought to be initiated by short quasi-palindromic structures. (A) One model for TID formation that relies on initiating repair replication at snap-back structures and switching templates to produce symmetrical triplications whose junctions have extended complementary palindromic sequences. (B) Deletions that modify a sTID to generate asymmetric junctions or simple tandem duplications.

lowing extensive growth under selection for copy-number increase (Araya *et al.* 2010). A TID amplification with one symmetrical and one modified junction has been reported in bacteria (Kugelberg *et al.* 2010).

A mechanism for forming the symmetrical precursor (sTID) is described in Figure 4A (Kugelberg *et al.* 2010). In this model, duplication formation is initiated by a quasipalindromic sequence near a strand end that snaps back to prime repair synthesis using itself as template. A second template switch generates a branched structure that resolves to leave the symmetrical TID (sTID). A correlation between short palindromes and duplications has been noted in several systems (Mehan *et al.* 2004; Ratray *et al.* 2005; Narayanan *et al.* 2006). Alternative models for TID formation have been suggested (Hastings *et al.* 2009; Brewer *et al.* 2011). In bacteria,

the asymmetric TID structures found after growth under selection have been attributed to secondary modifications of sTIDs that form during growth under selection (Kugelberg *et al.* 2010).

Additional mechanisms of duplication and amplification:

Several alternative duplication formation models have been proposed on the basis of systems in which amplifications arise under selection by a series of events. These include models in which palindromic sequences lead to snap-back replication to form dicentric chromosomes that break at cell division to produce a terminal inversion duplication (Gordenin *et al.* 1993; Narayanan *et al.* 2006). In addition, the “onion skin” model suggests amplification by nonhomologous end joining following local overreplication in chromosomes with multiple replication origins (Botchan *et al.* 1979; Spradling 1981). Rolling circle replication has been suggested for achieving sudden amplification (Petit *et al.* 1992; Roth *et al.* 1996; Hastings and Rosenberg 2002). These models seem unlikely to contribute to duplication formation in the system described here.

The F'_{128} *lac* plasmid

The duplications described here arise on the *E. coli* F'_{128} plasmid carried by *S. enterica*. Figure 5 diagrams the genome of this plasmid with some relevant sequence features (Kofoed *et al.* 2003). The *lac* operon is flanked by identical copies of the IS3 insertion sequences IS3A and IS3C, each 1258 bp separated by 131 kb. A third IS3 copy (IS3B) differs from the others by seven base substitutions (four amino acid substitutions in the transposase) that appear to eliminate transposase activity and prevent recombination between IS3B and the other IS3 copies. A plasmid operon encoding conjugational transfer functions (*tra*) encodes the TraI protein, which nicks plasmid DNA at the transfer origin (OriT) and displaces a 5'-ended single strand for transfer (Traxler and Minkley 1987; Frost *et al.* 1994). These nicks stimulate RecA-dependent recombination on the plasmid from 10- to 50-fold (Seifert and Porter 1984; Syvanen *et al.* 1986). The F'_{128} plasmid is rich in quasipalindromic REP elements (Stern *et al.* 1984; Gilson *et al.* 1990; Lupski and Weinstock 1992), which are found as clusters of paired inverse-order repeats (30 bp) as diagrammed in Figure 5. Duplication of the plasmid *lac* region are trapped by a method that detects any cell with two copies of *lacZ* regardless of where the duplications might end (see *Materials and Methods*).

Formation of duplications by multistep pathways on the F' plasmid

The idea of multistep duplication formation was first suggested by the *lac* amplifications isolated on F'_{128} following prolonged growth under selection for improved growth on lactose (Andersson *et al.* 1998). Selected amplifications were of two types. One type had tandem repeats of a small region (10–20 kb) with short (10 bp) junction sequences. The other carried multiple copies of a TID (described above)

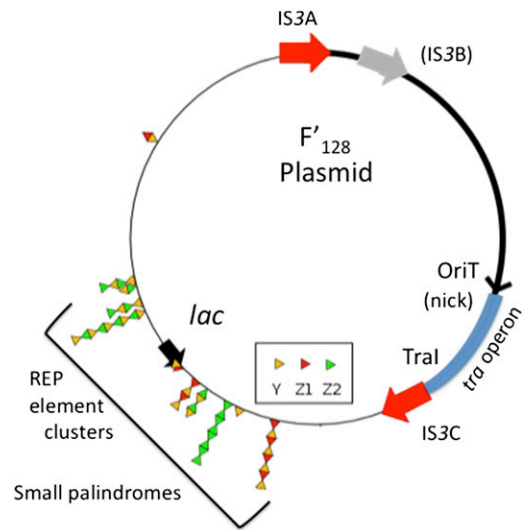


Figure 5 Features of the F'_{128} plasmid. The thick line represents the portion of the plasmid derived from the F-plasmid. The *lac*-containing segment between IS3A and IS3C is derived from the *E. coli* chromosome. Elements IS3A and IS3C are identical and differ from IS3B by seven base substitutions. REP elements are quasipalindromic repeated sequences that fall into three related families (Y, Z1, and Z2). These elements are typically found as clusters with multiple pairs of inversely oriented REP sequences.

with *lac* sequences in alternating orientation (Kugelberg *et al.* 2006). In contrast, the *lac* duplications trapped in unselected populations show duplicated regions that are generally larger than those amplified under selection, with most (97%) having a copy of IS3 at the junction. (See Appendix for a comparison.)

We describe a set of *lac* duplications that arose in several genetic backgrounds during nonselective growth. The formation rates of the several duplication types are interpreted in terms of several pathways for duplication formation, some of which have multiple steps, even though they arose with no deliberately applied selection. These pathways can be distinguished by their dependence on RecA. We propose that the RecA-dependent pathway involves formation of a plasmid dimer, which is remodeled by deletions stimulated by the transposase activity of IS3. The RecA-independent duplications may form either by annealing-mediated exchanges between sister copies of a monomeric plasmid or by remodeling an initial sTID duplication formed in a single plasmid.

Materials and Methods

Strains and media

All strains were derivatives of *Salmonella enterica* (*Typhimurium*) strain LT2 and are listed in Table 1. Rich medium was Luria broth (LB) with antibiotics as described below. The chromogenic β -galactosidase substrate, 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-gal), was obtained from Diagnostic Chemicals (Oxford, CT) and used in plates at 40 μ g/ml.

Table 1 Strains used in this study

Strain	Genotype	Source or reference
TR10000	Wild-type <i>S. enterica</i> (serovar Typhimurium) LT2	Lab collection
TT22240	<i>recA643::Tn10dTc</i>	Williams <i>et al.</i> (2006)
TT25467	<i>his-644(ΔOGDCBHAF) pro-621 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ /pSIM5(Cm^R)</i>	Reams <i>et al.</i> (2010)
TT25468	<i>his-644(ΔOGDCBHAF) pro-621 recA650::Gnt(sw) /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ /pSIM5(Cm^R)</i>	Reams <i>et al.</i> (2010)
TT25469	<i>his-644(ΔOGDCBHAF) pro-621 recA651::Rif(sw) /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ /pSIM5(Cm^R)</i>	Reams <i>et al.</i> (2010)
TT25470	<i>his-644(ΔOGDCBHAF) pro-621 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3C::Rif(sw) /pSIM5(Cm^R)</i>	This report
TT25471	<i>his-644(ΔOGDCBHAF) pro-621 recA643::Tn10dTc /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3C::Rif(sw) /pSIM5(Cm^R)</i>	This report
TT25474	<i>leuD21 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ /pSLT - /pSIM5(Cm^R)</i>	This report
TT25488	<i>leuD21 recA650::Gnt(sw) /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ /pSLT - /pSIM5(Cm^R)</i>	This report
TT25489	<i>leuD21 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3C::Rif(sw) /pSLT - /pSIM5(Cm^R)</i>	This report
TT25491	<i>leuD21 recA650::Gnt(sw) /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3C::Rif(sw) /pSLT - /pSIM5(Cm^R)</i>	This report
TT25562	<i>his-644(ΔOGDCBHAF) pro-621 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ tral::Rif(sw) /pSIM5(Cm^R)</i>	This report
TT25563	<i>leuD21 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ tral::Rif(sw) /pSLT - /pSIM5(Cm^R)</i>	This report
TT25564	<i>his-644(ΔOGDCBHAF) pro-621 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3A::Rif(sw) IS3C::Rif(sw) IS3B::Spec(sw) /pSIM5(Cm^R)</i>	This report
TT25565	<i>leuD21 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3A::Rif(sw) IS3C::Rif(sw) IS3B::Spec(sw) /pSLT - /pSIM5(Cm^R)</i>	This report
TT25586	<i>his-644(ΔOGDCBHAF) pro-621 recF521::Tn5d-rif /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ /pSIM5(Cm^R)</i>	This report
TT25601	<i>his-644(ΔOGDCBHAF) pro-621 recBC653::Tc /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ /pSIM5(Cm^R)</i>	This report
TT25603	<i>his-644(ΔOGDCBHAF) pro-621 recBC653::Tc recF521::Tn5d-rif /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ /pSIM5(Cm^R)</i>	This report
TT25604	<i>his-644(ΔOGDCBHAF) pro-621 recA650::Gnt(sw) /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ tral::Rif(sw) /pSIM5(Cm^R)</i>	This report
TT25681	<i>recA650::Gnt(sw) /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3A::Rif(sw) IS3B::Spec(sw) /pSIM5(Cm^R)</i>	This report
TT25682	<i>leuD21 recA650::Gnt(sw) /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3A::Rif(sw) IS3C::Rif(sw) IS3B::Spec(sw) /pSLT - /pSIM5(Cm^R)</i>	This report
TT25841	<i>F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ tral::Spec(sw) IS3C::Rif(sw) /pSIM5(Cm^R)</i>	This report
TT25868	<i>leuD21 proB1657::Tn10 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3C::mut2 IS3B::Spec(sw) /pSIM5(Cm^R)</i>	This report
TT25870	<i>leuD21 proB1657::Tn10 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ ΔIS3C IS3B::Spec(sw) /pSIM5(Cm^R)</i>	This report
TT25893	<i>leuD21 proB1657::Tn10 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3A::mut2 IS3B::Spec(sw) /pSIM5(Cm^R)</i>	This report
TT25894	<i>leuD21 proB1657::Tn10 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3A::mut2 IS3C::mut2 IS3B::Spec(sw) /pSIM5(Cm^R)</i>	This report
TT25900	<i>leuD21 proB1657::Tn10 recA651::Rif(sw) /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3A::mut2 IS3C::mut2 IS3B::Spec(sw) /pSIM5(Cm^R)</i>	This report
TT25902	<i>F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ tral::Rif(sw) IS3A::mut2 IS3C::mut2 IS3B::Spec(sw) /pSIM5(Cm^R)</i>	This report
TT25906	<i>recA643::Tn10dTc /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ tral::Rif(sw) IS3A::mut2 IS3C::mut2 IS3B::Spec(sw) /pSIM5(Cm^R)</i>	This report
TT25973	<i>leuD21 proB1657::Tn10 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3A::mut2 IS3C::mut2 IS3B::Spec(sw) /pDR5A /pSIM5(Cm^R)</i>	This report
TT25975	<i>leuD21 proB1657::Tn10 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3A::mut2 IS3B::Spec(sw) /pDR5A /pSIM5(Cm^R)</i>	This report
TT26107	<i>leuD21 proB1657::Tn10 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3C::mut2 IS3B::Spec(sw) /pDR5A /pSIM5(Cm^R)</i>	This report
TT26108	<i>leuD21 proB1657::Tn10 /F' 128 pro+ lacI4650::Ka(sw, Kn^S), lacZY+ IS3A::mut2 IS3C::mut2 IS3B::Spec(sw) /pDR6A(Amp^R) /pSIM5(Cm^R)</i>	This report

sw, alleles in which the normal sequence was replaced by a drug resistance determinant. *lacI4650::Ka(sw, Kn^S)*, a *lacI* gene replaced by a promoterless kanamycin resistance determinant. *pSIM5(Cm^R)*, plasmid carries the recombination genes (red) of phage lambda. *mut2*, an engineered point mutation, which changes the ATG initiation codon of IS3 transposase into an ATA (Spielmann-Ryser *et al.* 1991). *pDR5A(Amp^R)*, plasmid with the L-arabinose-inducible IS3 transposase gene. *pDR6A(Amp^R)*, plasmid with the L-arabinose-inducible mutant IS3 transposase gene carrying *mut2*.

Mutations that inactivated the IS3 transposase were constructed by linear transformation as described by Yu *et al.* (2003). To construct these strains, a cassette (Sac-Kan) encoding sucrose sensitivity (SacB) and kanamycin resistance (Npt1) was inserted into the initiation codon of the transposase gene, selecting kanamycin resistance (Lawes and Maloy 1995; Muro-Pastor and Maloy 1995). This cassette was then replaced with a mutant version of the initiation codon region with AUA in place of AUG by selection for resistance to sucrose. The Sac-Kan cassette was originally amplified from plasmid pPC217 with primers: TP2183 5' GGCTAAGTGAGTAAACTCTCAGTCAGAGGTGACTCACATATGGCGGCCGCTCTAGAAGTAG 3' and TP2184 5' GCTGTTTACGGGGTTTTTTTAC TGGTTGATACTGTTTTTGTGACGGTATCGATAAGCTTGA 3'. Strains carrying the Sac-Kan cassette within IS3 were transformed with single-stranded DNA that included the point mutation: 5' GTAAACTCTCAGTCAGAGGTGACTCACATAACAAAAACAGTATCAACAGTAAAAAACCC 3'. Sucrose-resistant

transformants that had lost kanamycin resistance were verified by sequencing of PCR fragments.

Trapping duplications formed in overnight cultures

Three methods for measuring duplication frequencies in unselected overnight cultures were described previously (Reams *et al.* 2010). These are: (1) Ka-Kan, (2) T-Recs, and (3) the standard transduction-based assays. The Ka-Kan assay was used here and its results were confirmed by both of the other two assays (Reams *et al.* 2010).

Duplication frequency measurements were confirmed by quantitative PCR to estimate the frequency of cells with the IS3A/C duplication joint point. One primer directed replication counterclockwise across IS3C and the other directed synthesis clockwise across IS3A as the plasmid is drawn in Figure 5. These primers were: TP1040 5' TGTGAATATGCTGACATGCC 3' and TP1041 5' CGGCGAATGGCTGGGATA 3', respectively. Since these primers direct divergent replication

in the parent plasmid, a PCR product will be amplified only in strains with an IS3A/C duplication junction. To estimate the frequency of cells with an IS3A/C duplication, multiple independent cultures of various tester strains were grown overnight in LB. Each culture was diluted serially and a PCR reaction was run on each dilution. All PCR products were separated on a 0.7% agarose gel using standard electrophoresis procedures and visualized after staining with ethidium bromide. The frequency of cells with an IS3A/C duplication was estimated by determining the greatest dilution that still generated a detectable junction PCR product.

In the PCR duplication assay, it was possible that the join point fragment could be artifactually generated by hybridization between single strands of the IS3A and C sequence produced by the PCR process. This possibility was tested using as template a mixture of DNA from two strains, each with only one of the IS3 elements, IS3A or IS3C, and therefore unable to form a duplication. This template mixture did not produce the junction point PCR fragment under the PCR conditions used for our tests ($<10^4$ cells per assay). The artifactual fragment was seen only when 100-fold more cells were tested.

Construction of a plasmid that overproduces IS3 transposase

The IS3-transposase gene was PCR amplified with an accurate polymerase (Pfx) and cloned into a pBAD-TOPO TA vector (Invitrogen, San Diego, CA) using the following primers: TP2425 5' TAACACGCGGCTAAGTGAGTAAA 3' and TP2426 5' GCCTAAGCGAGGTTCTTGTT 3'. This placed expression of the IS3-transposase gene under control of the arabinose-regulated pBAD promoter where it could be induced by 0.06% L-arabinose. As a negative control, the mutant IS3-transposase gene whose initiation codon is changed to ATA was inserted into pBAD-TOPO TA vector using the same primers. Overexpression of the normal transposase gene (but not the inactive mutant gene) showed an effect on duplication and deletion formation.

Determination of duplication junction sequences

Junction sequences were amplified using combinations of primer pools, as previously described (Kugelberg *et al.* 2006). All primers directed synthesis away from the *lac* operon. In some pools, all primers directed synthesis clockwise away from *lac* and, in other pools all primers directed synthesis counterclockwise away from *lac*. Combinations of a clockwise and a counterclockwise pool were used to test each duplication strain. No single pool (same direction) or pair of pools (divergent replication) is expected to prime fragment amplification from the parent template. However, cells with a *lac* duplication allowed divergent primer pairs to converge across a duplication join point. For each pool combination that produced a fragment, the responsible primer pair was identified, and its product was sequenced. All trapped duplications were initially screened for the most common duplication type, with an IS3A/C junction, using the

primer pools. Duplications yielding a PCR product with a size consistent with that of the IS3A/C junction fragment were scored as such. In a set of 100 independent duplications for which pools generated an appropriate-sized fragment, the IS3A/C junction was confirmed using two primers most proximal to the IS3A/C junction: TP1040 and TP1041. Thus, the fragment generated by the pools is reliable evidence of an IS3A/C duplication. Other duplication types generated amplified fragments of distinct sizes that were characterized by sequencing.

Determining rates of duplication formation (k_f) and loss (k_l)

Rates of duplication formation were estimated after 33 generations of growth as described previously (Reams *et al.* 2010). In determining duplication rate, observed increases in duplication frequency with time was corrected for both duplication loss and fitness costs using a spreadsheet simulation described previously (Reams *et al.* 2010).

Fitness cost measurements

The fitness cost of duplications was determined by comparing the growth rate of the parent haploid strain with that of a strain harboring a *lac* duplication as described previously (Reams *et al.* 2010). Similarly, the fitness advantage of cells with a deletion on the plasmid was measured by comparing the growth rates of the parent haploid strain with a strain harboring a *lac* deletion. To prevent duplication loss during the growth rate tests, all strains carried a *recA* mutation. This *recA* mutation reduced all growth rates slightly, but did not affect the relative fitness cost. Similarly, the fitness advantage of strains that lost a duplication was the same in both *recA*⁺ and *recA* strains. The small fitness changes due to *lac* duplications and deletions on plasmid F'₁₂₈ were determined by direct growth competition between the haploid parent strain and either a duplication or deletion strain. Relative fitness was calculated using the spreadsheet simulation method described previously (Reams *et al.* 2010).

Measuring deletion formation rates

The rate of recombination between separated IS3 repeats in a single plasmid can be estimated as the rate at which the parent haploid plasmid loses the segment between IS3A and IS3C that includes the *lac* region. Deletion rates were estimated during nonselective growth (LB) of a *lac*⁺ strain with its IS3B locus replaced by a spectinomycin-resistance determinant to assure that only deletions between IS3A and C are being scored. Parallel cultures were initiated from single blue colonies on a LB plate containing spectinomycin and X-gal and grown by serial passage of saturated cultures by 1:10 dilutions in 3 ml LB. The increasing frequency of Lac⁻ deletion mutants was determined at three time points (*e.g.*, 66, 99, and 132 generations) by counting the number of white (Lac⁻) colonies on LB plates containing spectinomycin and X-gal. Spectinomycin was included to assure the strains maintained F' and did not form a white colony by simply

losing the plasmid. The number of generations was calculated by measuring viable cells at each time point and the total dilution factor. Deletion formation rates were calculated from the change in frequency of Lac⁺ cells using the spreadsheet simulation method (Reams *et al.* 2010) to correct for fitness differences between the parent and the deletion mutant. In *recA*⁺ strains, >99% (100/100) of Lac⁻ cells carried a deletion between IS3A and IS3C; in *recA* mutant strains 71% of Lac⁻ cells carried the IS3A–C deletion, which was used to correct results in estimating deletion rates.

Results

Detection of duplications

All *lac* duplications were isolated in nonselectively grown populations using the Ka-Kan trapping method described previously (Reams *et al.* 2010). Parent cells have a promoterless kanamycin resistance determinant (Ka) replacing (in inverse order) the *lacI* gene of an otherwise normal *lac* operon. These cells are phenotypically Kan^S and Lac⁺. Duplications of the *lac* region are detected by introducing a single-strand fragment that provides a promoter for the Ka gene and simultaneously deletes the *lacZ* gene. Cells without a duplication inherit this fragment to become kanamycin resistant (Kan^R Lac⁻) and form white colonies on rich medium with kanamycin and X-gal. Cells with a *lac* duplication become Kan^R but remain Lac⁺ and form blue colonies because of their extra copy of the *lac* region. That is, one copy retains the parental Ka, *lac*⁺ allele and the other acquires Kan^R and a *lacZ* deletion. The ratio of blue (duplication) to white (haploid) colonies gives the frequency of duplications in the recipient culture. This method detects *lac* duplications in RecA⁺ or RecA⁻ cultures because the transformation is mediated by recombination functions of phage lambda (Red) that are induced immediately before plating on selective medium. Recombination mediated by the Red enzyme does not require RecA and does not contribute to duplication formation (Reams *et al.* 2010).

As described previously, the frequency of duplications in an unselected population comes to a steady state dictated by the formation rate, loss rate, and fitness cost of duplications (Reams *et al.* 2010). After 33 generations in rich medium, the *lac* duplication frequency reaches 3×10^{-3} ($\pm 0.5 \times 10^{-3}$ SD), about two-thirds the steady-state level. The duplication formation rates were calculated on the basis of duplication frequency at 33 generations, corrected for fitness cost and loss rate (Reams *et al.* 2010). The formation rate of *lac* duplications is 3×10^{-4} /cell/generation ($\pm 0.5 \times 10^{-4}$ SD). Through the initial part of this article, duplication frequency at 33 generations (rather than rate) is used to compare formation of several duplication types. These duplication frequencies closely parallel formation rates because the sampling time is still within the period of initial linear duplication accumulation.

Characterizing trapped duplications

The junction sequences of trapped duplications were characterized using pools of PCR primers that directed transcription

away from *lac* on the F'₁₂₈ plasmid either clockwise or counterclockwise (Kugelberg *et al.* 2006). These primer pools detect the junction of virtually any tandem *lac* duplication on the F'₁₂₈ plasmid. Over 1800 *lac* duplications were trapped from the parent or several derivative strains. Duplications fell into three classes that differ by the nature of their junction sequence. These types are the IS3, REP, and SJ (short junction) duplications described in Figure 6 and Table 2. Most duplications (97%) trapped in wild-type populations are IS3 mediated in the sense that their join point has a hybrid IS3A/C element (top duplication in Figure 6). These duplications form by an exchange between the identical 1258-bp IS3A and IS3C elements, separated by 131 kb and the largest sequence repeats flanking *lac* in F'₁₂₈. A less common IS3-mediated duplication type (4%) arose between IS3C and the third IS3 element, IS3B, which differs from the other two IS3 elements by seven base substitutions, four of which cause amino acid substitutions within the transposase (Tnp) gene. The sequence changes in IS3B are presumably sufficient to reduce its ability to recombine with the other IS3 copies and may impair function of its transposase.

All of the other duplications have shorter junction sequences (< 37 bp). Most of these duplications are the REP type (2.7% of the total duplications) whose joint point sequence is derived from two flanking REP elements. REP elements are “repeated extragenic palindromic” sequences roughly 40 bp in size that are present in several hundred copies in the genomes of *S. enterica* and *Escherichia coli* (Lupski and Weinstock 1992; Bachellier *et al.* 1997, 1999). Clusters of REP sequences flank the *lacZ* operon of F'₁₂₈ (Kofoid *et al.* 2003). The REP duplication junctions have 4- to 36-bp sequences shared with non-allelic REP elements that flank *lac*. Thus, the duplications appear to have formed by exchanges between two (usually nonidentical) REP elements located in direct order on opposite sides of *lac*. The various REP elements found at duplication endpoints are summarized in the Appendix. The small size of REP junction sequences makes it unlikely that these duplications formed by RecA-dependent homologous recombination. The regions duplicated by REP duplications had an average size of 26 kb, with a range of 17–70 kb.

The rest of the non-IS3 duplications (0.3% of total duplications) have short junctions (0–9 bp) not derived from REP elements and an average repeat size of 65 kb, (range of 21–108 kb). Surprisingly, many SJ duplications (10/28) had one end within a particular small region between the closely spaced REP elements R27 and R28 and the other end within 90 bp of REP element, R31. The 10 duplications at this hot spot had 0- to 2-bp junction sequences and included a region 17.2–17.5 kb. While 8 of these 10 duplications were unique, two independently isolated duplications were identical, carrying the same 2-bp junction sequence.

Homologous recombination (RecA) contributes to formation of IS3 duplications but not REP or SJ types

As seen on the left side of Figure 7, elimination of RecA reduced total duplication frequency 7.8-fold ($0.037\% \pm 0.009\%$

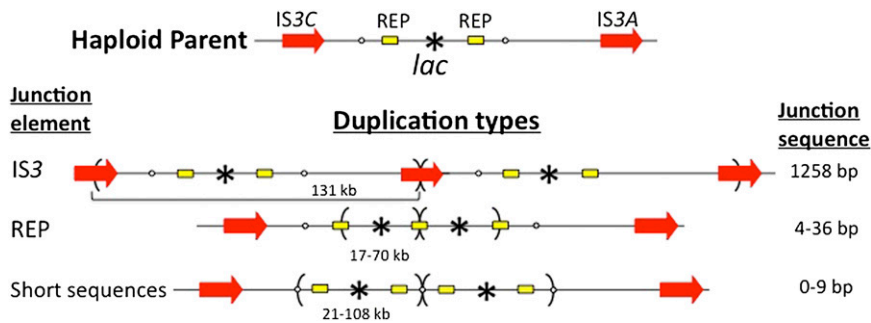


Figure 6 Structure of the duplication types. The haploid parent (top) has a *lac* operon flanked by copies of IS3 and several clusters of palindromic REP elements (see Figure 5). Duplications to form between separated short sequences and leave one copy of the sequence at the join point indicated by parentheses. While these junctions appear to have formed by unequal recombination, all duplication types can form in strains lacking RecA.

SD), but did not affect formation of REP and SJ types (see red bars).

The duplication rate in *recA* strains was still quite high (3×10^{-5} /cell/division). Most of the duplications formed without RecA have a copy of IS3 at their duplication junction and thus depend in some way on extensive IS3 repeats. The short junction sequences of REP and SJ duplications make it difficult to imagine their formation by homologous recombination. Consistent with this idea, the frequency of these duplication types was unaffected by the absence of RecA ($0.01\% \pm 0.006\%$ SD).

Removal of IS3 elements reduces *lac* duplication frequency

Formation of most *lac* duplications (with or without RecA) involves some kind of exchange between IS3A and IS3C, the only extensive sequences that flank *lac*. Deleting IS3C prevents both the IS3A/C and IS3B/C duplication types and reduces the overall frequency of *lac* duplications by 32-fold to 0.009% ($\pm 0.003\%$ SD). The *lac* duplications that form in the IS3C deletion strain were all REP or SJ types. Formation of these duplication types was unaffected by removal of the IS3C element and did not depend on RecA (see right side of Figure 7). This is consistent with the idea that REP- and SJ-mediated duplications form by a mechanism that involves neither IS3 homology nor recombination. In contrast, formation of the IS3C/A duplications depends on exchanges between IS3 elements that can be achieved either with or without RecA. The putatively defective IS3B element had very little effect on total duplication frequency. This supports the idea that the sequence changes that distinguish IS3B

from A and C reduce its ability to recombine and produce active transposase (see below).

IS3 transposase stimulates IS3A/C-mediated duplications

The involvement of IS3 in duplication formation without RecA suggested that IS3 elements might contribute more than simple sequence repeats. To test the role of transposase, the IS3A and C elements were replaced by directly oriented copies of the *rifR* gene (rifampicin resistance). This gene is nearly the same size as IS3 (1017 bp compared to 1258 bp), but is not derived from any transposable element and does not encode transposase. As shown in Figure 8, replacing only the IS3C element with *rifR* left no pair of sequences flanking *lac* and reduced duplication frequency 32-fold. Replacing both IS3A and C elements with *rifR* sequences left the *lac* region flanked by directly oriented copies of the *rifR* gene, but eliminated transposase. In the *rifR-rifR* strain, the *lac* duplication frequency was 10.7-fold lower than that in the parent IS3-IS3 strains ($0.288-0.027\% \pm 0.007\%$ SD) suggesting that sequence homology is insufficient. Of the *lac* duplications recovered in the *rifR-rifR* strain, 80% had a *rifR* gene at the join point and formed by exchanges that occurred between repeated *rifR* sequences without benefit of transposase nicking.

A *recA* mutation reduces the frequency of *rifR-rifR* duplications 2.5-fold (from 0.027 to $0.011\% \pm 0.007\%$ SD). In this *recA* mutant, 70% of the duplications had *rifR* junctions and the rest were the short junction types. Together, these results suggest that direct-order sequence repeats recombine at low frequency and can contribute to duplication formation even without RecA, but some property of the IS3 repeats stimulates

Table 2 Examples of join point sequences in *lac* duplications

Recombination region ^a	Join point ^b (bp)	Duplicated unit (kb)	Strain
Perfect match over 1258 bp of IS3 sequence	IS3C/A (1258)	131	TT26058
C-GATGCCTGATGCGACGCTGACGCGTCTTATCA-GC	REP26/REP32 (32)	22.7	TT26481
C-GATGCCTGATGCG-CG	REP25/REP19 (13)	31.6	TT26482
TGCCGGATGCGGCGTGAACGCCTTATCCGGCCTACG-A	REP27/REP31 (36)	17.4	TT26483
C-TATA-TG-T	SJ (4)	92.5	TT26484
GTGGA-AA	SJ (5)	108	TT26485

^a Exchange point within region shown in italics. Hyphens indicate mismatched bases between the exchange point and supporting homologous sequence.

^b IS3, IS3-mediated join point. REP, REP sequences at the join point. SJ, short homology join point.

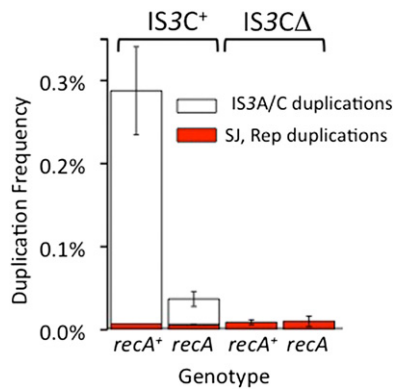


Figure 7 Effect of RecA and IS3 on formation of duplications. The frequency of several duplication types was determined for populations grown 33 generations without selection. RecA affects formation of duplications with IS3A/C junctions but not duplications with short junction sequences. Strains lacking RecA still produce an appreciable number of IS3A/C duplications.

the process significantly. In all of these tests, the formation rate of REP- and SJ-duplication types was not affected by the presence or absence of flanking IS3 or *rifR* repeats, or by a *recA* mutation, suggesting that these types form by an alternative mechanism.

To test contribution of IS3 transposase, expression of the IS3-transposase gene was eliminated by changing its ATG start codon to ATA. This point mutation was shown previously to eliminate expression of the IS3-transposase gene and reduce the IS3 co-integration rate by approximately 100-fold (Spielmann-Ryser *et al.* 1991). Thus, a single base change can eliminate transposase without reducing the size of the recombining sequence. Strains were constructed with this point mutation in one or both of the IS3 elements, IS3A and IS3C. In testing the effect of these mutations on duplication, the entire putatively defective IS3B element was removed to eliminate any potential contribution. As shown in Figure 8, removing IS3B had very little effect on total duplication frequency in an otherwise normal plasmid (second column).

Surprisingly, elimination of transposase expression from either one of the remaining IS3 elements reduced duplication formation 3.5- and 3.6-fold, respectively (to $0.080 \pm 0.015\%$ SD and $0.076\% \pm 0.022\%$ SD). Silencing both transposase genes caused very little additional drop in duplication rate (to $0.067\% \pm 0.019\%$ SD). This suggests that the transposases produced by both elements work together, possibly in *cis* on the element that encoded them, to stimulate duplication formation. This suggests that duplication may often involve simultaneous events at both elements. The double transposase defects reduced duplication frequency in both *rec+* (4.3-fold) and *recA* (2.6-fold) strains. This suggests the transposase contributes to duplication both with and without RecA.

It should be noted that transposase contribution does not involve an act of transposition, since the scored duplications all showed an exchange between IS3 copies (as diagrammed in Figure 1) and did not generate any novel sequence junc-

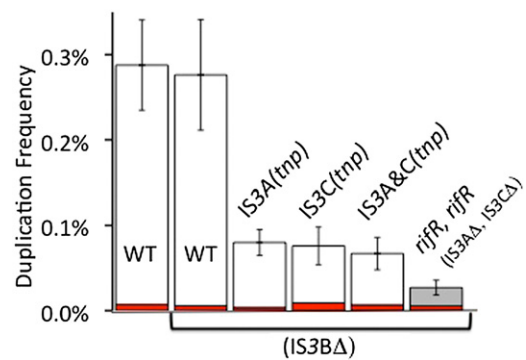


Figure 8 Effect of IS3 transposase on *lac* duplication formation. All strains except the one at the far left have a deletion of the defective IS3B element. For each strain, the two duplication types were determined: IS3 (open columns) and REP plus SJ (red columns). In the third, fourth, and fifth strain, IS3A and IS3C sequences are present but transposase (Tnp) production has been eliminated from one, the other, or both elements by a base substitution in the initiation codon of the *tpn* gene. In the strain at the far right (shaded column), both IS3A and IS3C were individually replaced by directly oriented copies of the *rifR* gene, leaving *lac* flanked by identical sequence repeats with no encoded transposase.

tion as would be expected for replicative transposition (see Figure 2). The absence of IS3 transposases had no effect on the overall frequency of REP or SJ duplication types.

In strains defective for transposase, duplication formation rates were partially restored by overexpressing IS3 transposase in *trans* from a plasmid. The plasmid is present at a copy number of about 10 per cell and expresses transposase from an arabinose-inducible promoter (pBAD). Inducing this transposase gene increased the duplication rate only about twofold in strains lacking either IS3A or IS3C transposase singly, and 1.8-fold in a strain defective for both transposases (data not shown). Given the strength of the arabinose promoter and the high plasmid copy number, these results suggest that IS3 transposase acts preferentially in *cis* on the sequence element that produced it, as shown previously for IS3 (Sekine *et al.* 1999) and many other transposable elements (Jain and Kleckner 1993; Weinreich *et al.* 1994; Derbyshire and Grindley 1996).

Plasmid conjugational transfer functions stimulate IS3-mediated duplications

Sequences on the F'_{128} plasmid are subject to intense recombination in *E. coli* caused by constitutive expression of the plasmid's conjugative transfer functions (*tra*) (Hopkins *et al.* 1980; Seifert and Porter 1984; Syvanen *et al.* 1986). The *Tra* functions encoded on F'_{128} accomplish cell-to-cell transfer by nicking F-plasmid DNA at the OriT site, displacing a single-strand 5'-end for transfer. In the *Salmonella* strains used here, *tra* expression is repressed by FinO, which is encoded by *Salmonella*'s resident pSLT plasmid (*i.e.*, *tra* expression is "low"). Transfer ability of F'_{128} is eliminated by deleting *traI*, which normally nicks at OriT (*i.e.*, *tra* activity is "none"). Conversely, transfer activity can be increased to high constitutive levels by removing the pSLT plasmid, thereby

eliminating the *tra* repressor (*i.e.*, *tra* expression is “high”). Strains with these three levels of *tra* expression (none, low, and high) were tested for their rate of duplication formation.

Expression of *tra* functions contributed substantially to formation of duplications in a *recA*⁺ strain, but its effect was reduced in absence of RecA (Figure 9A) and was eliminated in strains that lack both *recA* and IS3 transposase (Figure 9C). We conclude that nicking at OriT by TraI stimulates RecA-dependent exchanges on the plasmid that contribute to duplication formation. Transfer functions make no systematic contribution to the short junction duplication types (note the different scale in Figure 9B). These rates were determined in an IS3C deletion to eliminate any contribution from IS3 duplication types.

Role of RecB and RecF in duplication

In the model described below, RecA acts primarily to form plasmid dimers. These dimers arise by RecA-mediated exchanges between any homologous points on two plasmids and are stimulated by TraI nicking at OriT (Seifert and Porter 1984; Syvanen *et al.* 1986). Previous tests of recombination between extensive sequence repeats in the bacterial chromosome (segregation of a chromosomal duplication) suggested that the exchanges required RecA plus either the RecBC or the RecF recombination pathways (Galitski and Roth 1997). The rate of duplication formation on F'₁₂₈ is reduced in a *recA* mutant and in a *recB*, *recF* double mutant, but not in a *recB* or *recF* single mutant (Figure 10). This dependence is expected if recombination is initiated by single-strand nicks or double-strand breaks (Spies 2005) and blockage of either single pathway leaves the other functional. The model suggests that a major part of the RecA contribution to duplication is to promote exchanges between long homologies and make plasmid dimers that are precursors to duplication formation.

A model for duplication formation suggested by results

The results suggest that duplications with an IS3 join point form by both recombination-dependent and -independent pathways. The IS3 transposase contributes under both conditions without catalyzing transposition. Duplication formation is stimulated by plasmid conjugation functions that are known to stimulate homologous recombination between plasmids (Seifert and Porter 1984; Syvanen *et al.* 1986). Rare duplication types with short junction sequences (REP and SJ) form by a process that is independent of recombination, transposase, and conjugation functions.

Most duplications form in a plasmid dimer generated primarily by RecA-dependent exchanges between plasmid copies (Figure 11A, right) and stimulated by DNA nicks made at OriT by the conjugation function TraI. In this plasmid dimer, a deletion between nonallelic IS3 elements leaves a monomeric plasmid with a duplication. Both IS3 transposase and RecA contribute to the deletion event, which resembles the resolution step in IS3 replicative transposition. Such deletions can occur by homologous recombination or, less frequently, by single-strand annealing. We propose that IS3 transposase

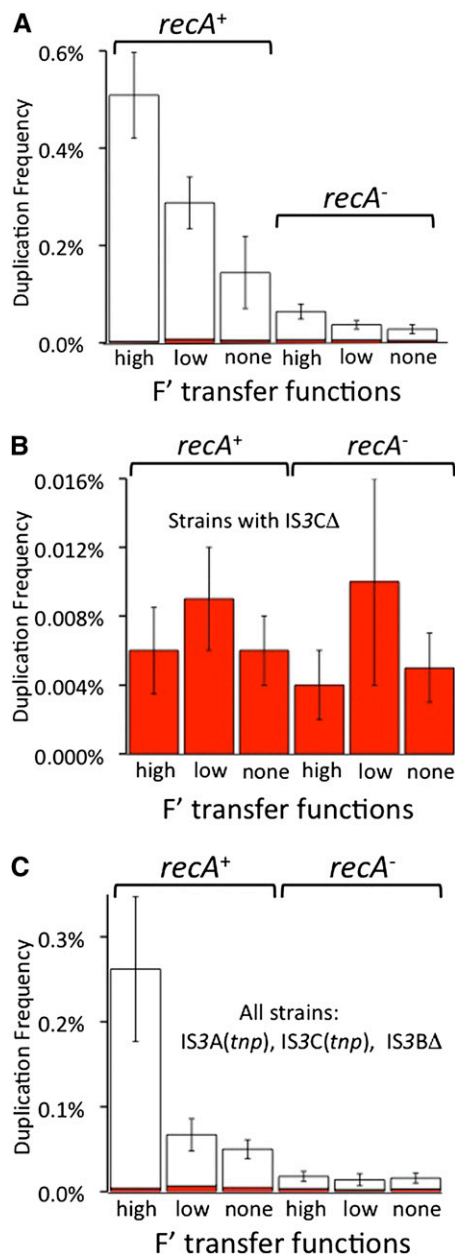


Figure 9 Effects of plasmid transfer functions. (A) The effect of *tra* expression—and therefore TraI nicking at OriT—on duplication frequency in *recA*⁺ and *recA*⁻ strains. (B) The effect of *tra* expression on the frequency of SJ plus REP duplication types in strains lacking IS3C and therefore incapable of making duplications with an IS3A/C junction. (C) The effect of *tra* on duplication frequency in strains with flanking IS3 elements but lacking transposase. Throughout the figure, the expression level “none” is a *traI* mutant with no nicking at OriT, “low” is a *traI*⁺ strain with *tra* operon repressed by pSLT, and “high” is a *traI*⁺ strain with derepressed *tra* operon (no pSLT).

nicks DNA at the ends of IS3, leading to breaks that initiate deletion by either recombination or single-strand annealing.

Duplications can also form without a dimer by standard unequal sister-strand exchanges between nonallelic IS3 elements (IS3A and IS3C) (diagrammed in Figure 1). In a *recA*⁺ strain, these events are mediated primarily by recombination,

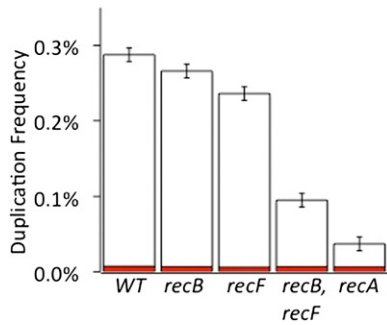


Figure 10 Effects of *recB* and *recF* mutations. For each strain, the two general duplication types were determined: IS3 (open columns) and REP plus SJ (red columns).

but also by single-strand annealing. Without RecA, dimers are less frequent, but duplications can arise by single-strand annealing of IS3 sequences (see Figure 3 and middle of Figure 11A). Annealing requires two coexisting breaks and is made more likely by frequent nicking at IS3 and by longer persistence of breaks in the absence of RecA.

Involvement of a plasmid dimer intermediate is one example of duplication by a multistep pathway. Another example is in Figure 4B above. The model shares general features with replicative transposition—a multistep pathway in which an IS element in one plasmid mediates co-integration with a different plasmid (Figure 11B). The co-integrate is an intermediate that breaks down by exchanges between IS copies to regenerate two separate plasmids, each now having a copy of the IS. In replicative transposition, the subsequent resolution step can be catalyzed by a dedicated resolvase or by a second activity of the transposase that catalyzes exchanges

between two IS elements. Resolution can also be achieved by recombination between the two elements. The deletions between IS3 elements described here can be considered a resolution activity of the IS3 transposase.

Plasmid dimers are known to form at high frequency in Rec⁺ strains, but are much less abundant in *recA* strains (Biek and Cohen 1986; Perals *et al.* 2000; Barre and Sherratt 2005). Some dimers clearly form at a low rate without RecA and can be detected by positive selection (Dianov *et al.* 1991; Mazin *et al.* 1991; Lovett *et al.* 1993). Our estimates suggest that dimers of F'₁₂₈ are present in about 25% of cells in an unselected RecA⁺ culture, while none (<0.1%) were found in *recA* strains (A. B. Reams, unpublished results). Most F plasmid dimers are resolved by recombination or by the ResD enzyme of F'₁₂₈, which catalyzes site-specific recombination at the *rtsF* resolution site (Lane *et al.* 1986).

The model proposes that the dimer is resolved by unequal exchanges between nonallelic IS3C and IS3A sequences (see Figure 11A left). If this exchange is reciprocal, it produces one plasmid with a tandem duplication and another with a corresponding deletion. These unequal exchanges can occur by simple recombination between any embedded sequence repeats, but are much more frequent when the recombining sequence is nicked, *e.g.*, when transposase nicks the ends of IS3.

Testing Predictions of the Model

In this model, duplications form by two steps. A plasmid dimer forms and is then resolved by a deletion between IS3 copies. The second step is tested below and provides evidence that deletions do form between IS3 copies and depend heavily on RecA and on IS3 transposase.

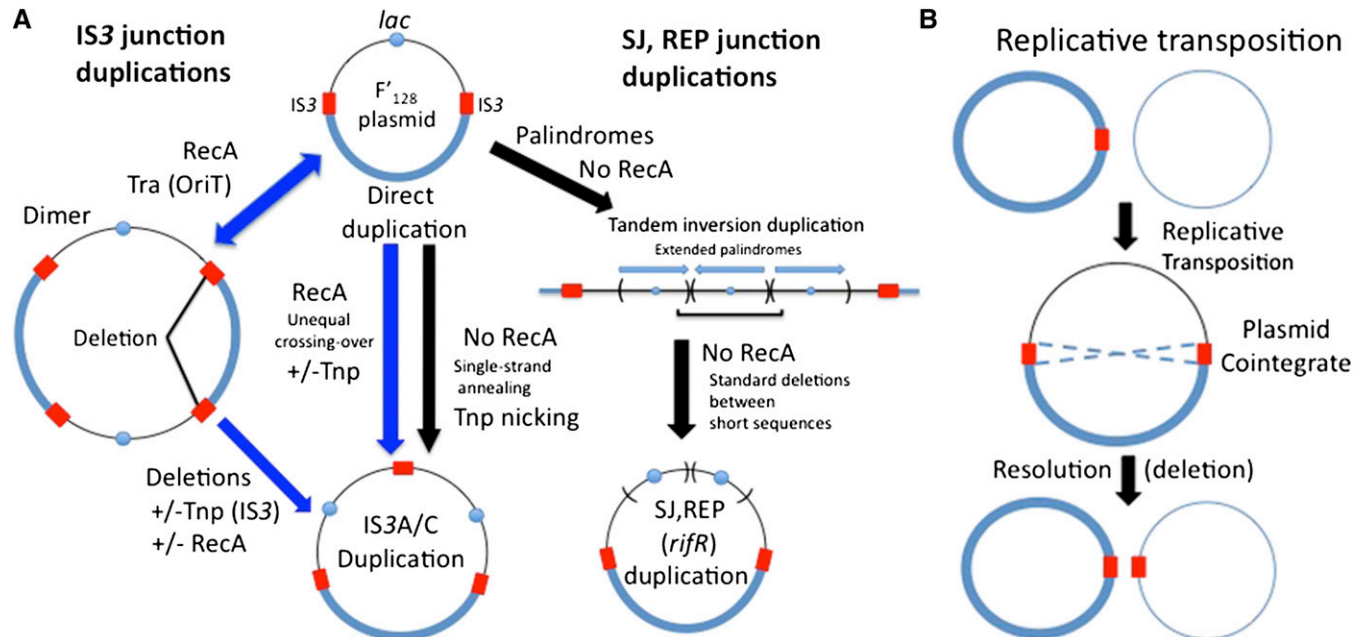


Figure 11 Pathways for duplication formation and replicative transposition. (A) Three different pathways that form duplications on F'₁₂₈. Each pathway can be distinguished on the basis of its dependence on RecA, IS3 transposase, and Tra functions. (B) The process of replicative transposition through a co-integrate intermediate. Note the similarity between the resolution event in B and the deletion affecting the plasmid dimer at left side of A.

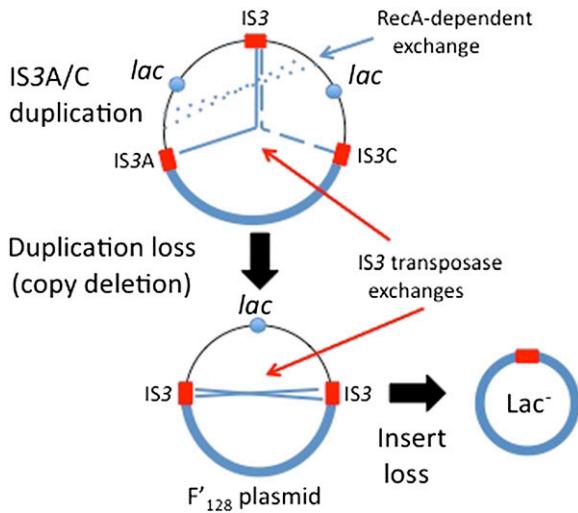


Figure 12 Assays of deletion formation on F'_{128} . The figure describes the two events assayed to assess formation of deletions between two IS3 elements. The first test, duplication loss, was performed in *recA* strains, where duplication loss occurs primarily by exchanges between two IS3 elements. The second test, insert loss rate, uses a monomeric F'_{128} plasmid and determines the rate at which the insert (*lac*) is lost by exchanges between nonallelic IS3 elements.

Deletions that cause loss of a duplication: The first test examined the recombination events diagrammed at the top of Figure 12. The parent IS3A/C duplication was trapped by the Ka-Kan method, which leaves one copy with a functional *lac* allele (Lac^+ , Kan^S) and the other with a *lac* deletion (Lac^- , Kan^R). This duplication strain is phenotypically Lac^+ Kan^R and produces haploid segregants that lose either their Lac^+ or Kan^R phenotype with equal probability. Rates of duplication loss were determined during nonselective growth by following a loss of the Lac^+ phenotype with time and correcting this loss rate for fitness effect as described in *Materials and Methods*. A duplication copy can be lost by exchanges between IS3 elements (dashed and solid lines) or between any other homologous sites in the two repeats (dotted lines). Loss by the second route is essentially eliminated by a *recA* null mutation, which decreases the formation rate of Lac^- segregants 50-fold and leaves a residual RecA-independent loss rate that reflects primarily events between IS3 copies. Rates of duplication loss were compared in *recA* strains with and without transposase point mutations in their IS3 copies. Figure 13 shows that in *recA* strains IS3 transposase contributes roughly two-thirds of the duplications. Overexpressing transposase in *trans* from a plasmid restored the deletion rate to strains with a *tnp* mutation. Thus, IS3 transposase can catalyze deletions between copies of IS3 repeats in the absence of RecA.

Deletions between IS3 copies that remove *lac* from a monomeric F'_{128} plasmid: The second test examined deletions that remove the *lac* segment (IS3A–*lac*–IS3C) from a plasmid monomer. Loss of the Lac^+ phenotype occurs primarily by deletions between nonallelic IS3 elements as diagrammed at

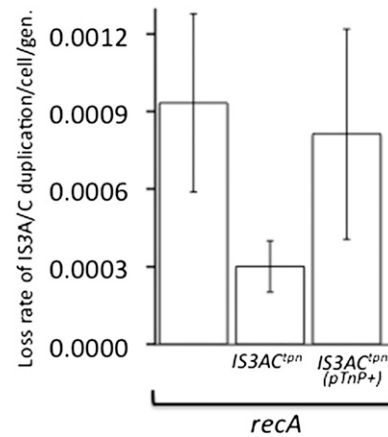


Figure 13 Duplication loss recombination-deficient strains depends on IS3 transposase. The duplication loss rate was measured in *recA* strains whose F'_{128} plasmid carries a *lac* duplication. Three such duplication strains were analyzed: (1) a strain with wild-type IS3 elements, (2) a strain in which all copies of IS3 lack transposase expression, and (3) a strain with impaired IS3 transposase expression and a plasmid that overexpresses IS3 transposase in *trans*. The frequency of cells without the duplication was measured at various time points during nonselective growth and the duplication loss rates were calculated after correcting for fitness effects using a spreadsheet simulation.

the bottom of Figure 12 and in the second step of the duplication model (Figure 11A, left). These deletions remove the entire 131-kb *lac* region and occur at a rate of 4×10^{-4} deletions/cell/generation (see row 1 of Table 3). PCR verified that >99% of Lac^- derivatives had an IS3A/C deletion join point. Removal of the IS3C element from the parent plasmid reduced Lac^- mutant formation over 1000-fold (row 5), demonstrating that the scored deletion events reflect exchanges between IS3A and IS3C, as suggested by the model.

The deletion rate dropped 300-fold in a *recA* mutant (Table 3, row 3). Of the reduced number of Lac^- cells arising without RecA, 71% had deletions between IS3A and C, showing that IS3 copies can recombine without RecA, albeit at severely reduced rates. Reduced expression of IS3 transposase lowered deletion formation about twofold in a RecA+ strain (rows 1 and 2) and about fourfold in a *recA* mutant (rows 3 and 4). These results suggest that IS3 transposase stimulates (perhaps by DNA nicking) formation of deletions both by RecA-dependent and RecA-independent mechanisms. Thus, the deletions proposed by the model can form and RecA and IS3 transposase are major and minor contributors.

Comparing rates of duplication and deletion: A surprising aspect of the duplication process was revealed by comparing the deletion rate (estimated by the insert loss) to the duplication rate (measured by the Ka-Kan assay). As seen in the top row of Table 3, the F'_{128} plasmid forms duplications and deletions at comparable rates in a *recA*⁺ strain, using IS3 elements as recombining sequences. A defect in transposase expression reduced duplication and deletion rates two- to fourfold in *recA*⁺ and *recA* strains. Unexpectedly,

Table 3 Duplication and deletion rates for *lac* on plasmid F'128

Genotype	Duplication rate (kDup) ($\times 10^7$)	Deletion rate (kDel) ($\times 10^7$)
<i>rec</i> ⁺	3000	4100
<i>rec</i> ⁺ , IS3A(<i>tnp</i>), C(<i>tnp</i>)	720	2100
<i>recA</i>	270	13
<i>recA</i> , IS3A(<i>tnp</i>), C(<i>tnp</i>)	120	3.5
<i>rec</i> ⁺ , IS3C Δ	62	3.8
<i>recA</i> , IS3C Δ	60	3.7

All measurements are made in strains with repressed (low) expression of plasmid transfer (Tra) function, and a deletion of the defective element IS3B. Units are expressed as events per cell per division.

a *recA* mutation reduced the deletion rate 300-fold and duplication rate only 10-fold (compare rows 1 and 3 in Table 3). The larger effect of *recA* on deletion rates was also seen in strains lacking IS3 sequences (rows 5 and 6). As a result, duplications are considerably more common than deletions in the absence of RecA. This was unexpected, because duplications and deletions are generally thought to form in the same way (unequal recombination).

To validate the larger effect of RecA on deletion formation, rates were assayed by direct PCR tests of whole populations. Cultures of the parent strain with a monomeric plasmid were used as template for PCR tests that amplified join points of IS3-mediated duplications and deletions, both which are absent in the parent strain. The frequency of cells with a duplication or deletion join point was estimated from the dilution endpoint at which junction fragments could no longer be detected. In this assay, a *recA* mutation reduced the IS3A/C deletion frequency about 500-fold and the IS3C/A duplication frequency only 10-fold. The larger effect of RecA on deletion than on duplications is consistent with the results using the two genetic assays. Reasons for this difference are discussed later.

Finally, Table 3 shows that strains lacking the entire IS3C element (row 5) form *lac* deletions at about the same rate as strains having both IS3 elements but lacking RecA and IS3 transposase (row 4). In both of these strains, the contribution of IS3 is eliminated and the rate of duplication is at least 20-fold higher than that of deletions, suggesting the existence of a low level pathway for duplication that does not require RecA, transposase, or direct order sequence repeats. We propose below that these duplications (the SJ and REP types) form by a pathway initiated by a TID (see Figure 4). The ability of this pathway to generate duplications but not deletions may explain the excess of duplications in strains lacking both recombination and IS3 sequences.

Discussion

Duplications of the *lac* operon are frequent (3×10^{-4} /per cell/division) and are of two general types. Roughly 97% arose by genetic exchanges between IS3 sequences and left a copy of IS3 at the duplication join point. Elimination of RecA, TraI endonuclease, or IS3 transposase reduced, but did not eliminate, formation of IS3-mediated duplications.

Much rarer *lac* duplication types (3%) have short junction sequences (REP or SJ) and end at a variety of sites other than IS3. The REP and SJ duplications appear to form by an independent mechanism.

Several aspects of these findings were unexpected:

1. While most duplications arise by exchanges between IS3A and C elements, the duplication rate is reduced only about 11-fold in the absence of RecA and rate of duplication in *recA* mutants is still quite high ($>10^{-5}$ /cell/division). See Figure 7 and Table 3.
2. The RecA dependence of duplication on F'128 is unexpected since RecA makes a smaller contribution to duplication at some sites in the chromosome (Reams *et al.* 2010) and on small plasmids (Dianov *et al.* 1991; Mazin *et al.* 1991).
3. While deletions and duplications between long repeats form at similar rates in RecA⁺ strains, deletion formation is more heavily RecA dependent than duplication. That is, in the absence of RecA, duplications are more common than deletions. This was seen in genetic assays and was verified by PCR amplification of deletion and duplication join points in unselected cultures. This is also seen in the chromosome between 5-kb *rrn* repeats, where a *recA* mutation reduces duplication formation only a few fold (Reams *et al.* 2010) but reduces duplication loss (40-kb deletion) several thousand fold (A. B. Reams, unpublished results).
4. IS3 transposase contributes to duplication formation both with and without RecA, yet no transposition event is involved.
5. The conjugative (transfer) replication origin of F contributes to duplication, probably because the TraI enzyme stimulates plasmid-plasmid recombination by nicking DNA at a site (OriT), far from the recombining IS3 sequences.
6. The functions that contribute to formation of IS3AC duplications and deletions (RecA, TraI, and Tnp) had no effect on formation of duplications with short junction sequences (REP and SJ).

Duplications form at a high rate in many genetic systems. The traditional idea that duplications form by unequal recombination between repeated sequences is attractive, but may be true only with long repeats subject to nicks that generate recombinogenic ends. Experimental genetics may have predisposed us to thinking of recombination as frequent, because the recombination we see in eukaryotic systems has been stimulated by the DNA breaks introduced during the meiotic process. Similarly, bacterial recombination assays provide linear fragments with broken ends. Separated sequence repeats embedded in a larger chromosome may recombine rather seldom, as seen here for IS3 sequences that lack their transposase or for the repeated *rifR* gene sequences. In light of these considerations, it is surprising that duplications of a typical chromosomal gene with no provided repeats form at a high rate (1×10^{-5} /cell/division). The high rate may result from a very large number of potential endpoints,

even though the rate of each individually type is very low. High duplication rates at particular sites may reflect catalyzed events between substantial sequence repeats.

We suggest here that the high rate of IS3-mediated duplications reflects frequent plasmid dimers that are remodeled by frequent deletion events to produce the detected tandem duplication (Figure 11A). Such multistep pathways are effective because each step is catalyzed and the intermediate cell type may grow or be subject to selection between steps. The first step, F plasmid dimerization, can result from a recombination event at any point in the plasmid (231 kb), events that are highly stimulated by the nicks introduced by TraI (Seifert and Porter 1984; Syvanen *et al.* 1986). We estimate that one cell in four has a dimer of its F'_{128} plasmid in *rec*⁺ strains (A. B. Reams, unpublished results). This agrees with frequencies of RecA-dependent dimers seen for other plasmids (Biek and Cohen 1986). The second step in the proposed duplication pathway is a deletion between nonallelic IS3 elements stimulated by transposase-introduced DNA nicks near substantial (1.2-kb) sequence repeats. The IS3 transposase recognizes and cleaves the 3'-ends of the inverted repeats flanking IS3 sequences (Sekine *et al.* 1999). We propose that IS3 normally uses these nicks (with or without RecA) to resolve the plasmid co-integrates it produces in the course of replicative transposition. In the F' plasmid, these nicks can generate deletions between nonallelic IS3 copies. With RecA, a break at either IS3 copy can be used to generate a duplication by homologous unequal recombination between sisters. Without RecA, double-strand ends near nonallelic IS3 elements on different sisters may be independently resected to expose single strands that can recombine by annealing. (See Figure 3.) Without catalyzed events, duplications and deletions arise at significantly lower rates.

Perhaps most surprising is the finding that RecA appears more important for deletion than for duplication. To account for these observations, we propose that the strand exchange activity of RecA contributes equally to the process of deletion and duplication. However, in the case of duplications, the reduced homologous recombination caused by lack of RecA may be compensated by an increase in single-strand annealing. This increase may occur because DNA breaks required for annealing persist longer in the absence RecA, making it more likely for two breaks to coexist. In addition, revealed single strands may anneal better when not involved in RecA protein filaments (A. B. Reams, unpublished results).

Duplications in the least frequent class (3%) have short junction sequences (SJ), some of which are within palindromic REP elements far from IS3A and C elements. The structure of these duplications is diagrammed in the Appendix. Formation rate of these duplications is about 30-fold lower than that of IS3-mediated duplications and is not affected by lack of RecA, Tra, or IS3 elements. The junctions of SJ and REP duplications resemble those of duplications found in the bacterial chromosome in regions far from *rrn* repeats (E. Kofoid, unpublished results). Like SJ and REP duplications, the RecA dependence of chromosomal duplica-

tion is very low. We suggest that in the absence of long repeats, duplications form by a pathway initiated when short a quasipalindromic sequence generates a symmetrical inversion duplication (sTID) as in Figure 4A (Kugelberg *et al.* 2010). This structure is later modified by deletions that leave the observed REP and SJ duplications. The proposed pathway is supported by the amplified TID structures found following selection for increased gene copy number in *E. coli*, *Salmonella*, and yeast (Ratray *et al.* 2005; Narayanan *et al.* 2006; Kugelberg *et al.* 2010; Brewer *et al.* 2011; Lin *et al.* 2011).

The results described here have implications for the origin of Lac⁺ revertants appearing during prolonged growth under selection (Cairns and Foster 1991; Andersson *et al.* 2011). In the Cairns system, all revertant colonies include some cells with a high-copy *lac* amplification, but none of these amplifications has the IS3 junction that is most common in the absence of selection. Selected amplifications have shorter repeats, mostly with SJ junctions. The revertants appearing under selection were first attributed to cells with an IS3-mediated *lac* duplication that grew slowly and acquired junction deletions that enhanced growth and left the observed short junction sequences (SJ) (Kugelberg *et al.* 2006). Results presented here argue against this idea by showing that the formation rate of SJ and REP duplications is not affected by removal of IS3. Furthermore, the yield of Lac⁺ revertants under long-term selection is reduced very little by removal of IS3 (S. Maisnier-Patin and J. P. Aboubechara, unpublished results). These results suggest that the common IS3 duplications are not precursors of the short-junction duplications described here or found in selected amplifications. It seems clear that the whole range of duplication types (IS3, SJ, REP) can form during nonselective growth. We propose that when a population including such cells is placed under selection for more *lac* copies, only short duplications have sufficiently low fitness cost to support growth and selective amplification (Reams *et al.* 2010). Consistent with this idea, amplifications recovered after prolonged selection (Kugelberg *et al.* 2006) have generally shorter repeats than the unselected SJ duplications described here. (See Appendix.)

Acknowledgments

We thank members of the lab for their advice and suggestions during the course of this work and during preparation of the manuscript, including John Aboubechara, Natalie Duleba, Manjot Grewel, Douglas Huseby, Sophie Maisnier-Patin, Semarhy-Quiñones-Soto, and Emiko Sano. This work was supported in part by a grant from the National Institutes of Health, GM27068.

Literature Cited

- Anderson, R. P., and J. R. Roth, 1978 Tandem chromosomal duplications in *Salmonella typhimurium*: fusion of histidine genes to novel promoters. *J. Mol. Biol.* 119: 147–166.
- 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. USA* 78: 3113–3117.
- Andersson, D. I., E. S. Slechta, and J. R. Roth, 1998 Evidence that gene amplification underlies adaptive mutability of the bacterial lac operon. *Science* 282: 1133–1135.
- Andersson, D. I., D. Hughes, and J. R. Roth (Editors), 2011 *The Origin of Mutants Under Selection: Interactions of Mutation, Growth, and Selection*. ASM Press, Washington, DC.
- Araya, C. L., C. Payen, M. J. Dunham, and S. Fields, 2010 Whole-genome sequencing of a laboratory-evolved yeast strain. *BMC Genomics* 11: 88.
- Bachelier, S., J. M. Clement, M. Hofnung, and E. Gilson, 1997 Bacterial interspersed mosaic elements (BIMEs) are a major source of sequence polymorphism in *Escherichia coli* intergenic regions including specific associations with a new insertion sequence. *Genetics* 145: 551–562.
- Bachelier, S., J. M. Clement, and M. Hofnung, 1999 Short palindromic repetitive DNA elements in enterobacteria: a survey. *Res. Microbiol.* 150: 627–639.
- Barre, F. X., and D. J. Sherratt, 2005 Chromosome dimer resolution, pp. 513–524 in *The Bacterial Chromosome*, edited by N. P. Higgins. ASM Press, Washington, DC.
- Bergthorsson, U., D. I. Andersson, and J. R. Roth, 2007 Ohno's dilemma: evolution of new genes under continuous selection. *Proc. Natl. Acad. Sci. USA* 104: 17004–17009.
- Biek, D. P., and S. N. Cohen, 1986 Identification and characterization of recD, a gene affecting plasmid maintenance and recombination in *Escherichia coli*. *J. Bacteriol.* 167: 594–603.
- Botchan, M., W. Topp, and J. Sambrook, 1979 Studies on simian virus 40 excision from cellular chromosomes. *Cold Spring Harb. Symp. Quant. Biol.* 43(2): 709–719.
- Brewer, B. J., C. Payen, M. K. Raghuraman, and M. J. Dunham, 2011 Origin-dependent inverted-repeat amplification: a replication-based model for generating palindromic amplicons. *PLoS Genet.* 7: e1002016.
- Bridges, C. B., 1936 The bar “gene”: a duplication. *Science* 83: 210–211.
- Bzymek, M., and S. T. Lovett, 2001 Evidence for two mechanisms of palindrome-stimulated deletion in *Escherichia coli*: single-strand annealing and replication slipped mispairing. *Genetics* 158: 527–540.
- Cairns, J., and P. L. Foster, 1991 Adaptive reversion of a frameshift mutation in *Escherichia coli*. *Genetics* 128: 695–701.
- Clugston, C. K., and A. P. Jessop, 1991 A bacterial position effect: when the F factor in *E. coli* K12 is integrated in cis to a chromosomal gene that is flanked by IS1 repeats the elements are activated so that amplification and other regulatory changes that affect the gene can occur. *Mutat. Res.* 248: 1–15.
- Conrad, D. F., D. Pinto, R. Redon, L. Feuk, O. Gokcumen *et al.*, 2010 Origins and functional impact of copy number variation in the human genome. *Nature* 464: 704–712.
- Cuff, L. E., K. T. Elliott, S. C. Seaton, M. K. Ishaq, N. S. Laniohan *et al.*, 2012 Analysis of IS1236-mediated gene amplification events in *Acinetobacter baylyi* ADP1. *J. Bacteriol.* 194: 4395–4405.
- Derbyshire, K. M., and N. D. Grindley, 1996 Cis preference of the IS903 transposase is mediated by a combination of transposase instability and inefficient translation. *Mol. Microbiol.* 21: 1261–1272.
- Dianov, G. L., A. V. Kuzminov, A. V. Mazin, and R. I. Salganik, 1991 Molecular mechanisms of deletion formation in *Escherichia coli* plasmids. I. Deletion formation mediated by long direct repeats. *Mol. Gen. Genet.* 228: 153–159.
- Farabaugh, P. J., U. Schmeissner, M. Hofer, and J. H. Miller, 1978 Genetic studies of the lac repressor. VII. On the molecular nature of spontaneous hotspots in the lacI gene of *Escherichia coli*. *J. Mol. Biol.* 126: 847–857.
- Flores, M., P. Mavingui, X. Perret, W. J. Broughton, D. Romero *et al.*, 2000 Prediction, identification, and artificial selection of DNA rearrangements in *Rhizobium*: toward a natural genomic design. *Proc. Natl. Acad. Sci. USA* 97: 9138–9143.
- Frost, L. S., K. Ippen-Ihler, and R. A. Skurray, 1994 Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* 58: 162–210.
- Galitski, T., and J. R. Roth, 1997 Pathways for homologous recombination between chromosomal direct repeats in *Salmonella typhimurium*. *Genetics* 146: 751–767.
- Gilson, E., S. Bachelier, S. Perrin, D. Perrin, P. A. Grimont *et al.*, 1990 Palindromic unit highly repetitive DNA sequences exhibit species specificity within Enterobacteriaceae. *Res. Microbiol.* 141: 1103–1116.
- Goldberg, I., and J. J. Mekalanos, 1986 Effect of a recA mutation on cholera toxin gene amplification and deletion events. *J. Bacteriol.* 165: 723–731.
- Goldfless, S. J., A. S. Morag, K. A. Belisle, V. A. Sutera Jr., and S. T. Lovett, 2006 DNA repeat rearrangements mediated by DnaK-dependent replication fork repair. *Mol. Cell* 21: 595–604.
- Gordenin, D. A., K. S. Lobachev, N. P. Degtyareva, A. L. Malkova, E. Perkins *et al.*, 1993 Inverted DNA repeats: a source of eukaryotic genomic instability. *Mol. Cell. Biol.* 13: 5315–5322.
- Gore, J. M., F. A. Ran, and L. N. Ornston, 2006 Deletion mutations caused by DNA strand slippage in *Acinetobacter baylyi*. *Appl. Environ. Microbiol.* 72: 5239–5245.
- Green, M. M., 1985 The role of mobile DNA elements in unequal and intrachromosomal crossing-over in *Drosophila melanogaster*, pp. 353–362 in *Aneuploidy*, edited by V. Dellarco, P. Voytek, and A. Hollaender. Plenum Press, New York.
- Haack, K. R., and J. R. Roth, 1995 Recombination between chromosomal IS200 elements supports frequent duplication formation in *Salmonella typhimurium*. *Genetics* 141: 1245–1252.
- Hastings, P. J., and S. M. Rosenberg, 2002 In pursuit of a molecular mechanism for adaptive gene amplification. *DNA Repair* 1: 111–123.
- Hastings, P. J., G. Ira, and J. R. Lupski, 2009 A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* 5: e1000327.
- Hopkins, J. D., M. B. Clements, T. Y. Liang, R. R. Isberg, and M. Syvanen, 1980 Recombination genes on the *Escherichia coli* sex factor specific for transposable elements. *Proc. Natl. Acad. Sci. USA* 77: 2814–2818.
- Jain, C., and N. Kleckner, 1993 Preferential cis action of IS10 transposase depends upon its mode of synthesis. *Mol. Microbiol.* 9: 249–260.
- Kofoed, E., U. Bergthorsson, E. S. Slechta, and J. R. Roth, 2003 Formation of an F' plasmid by recombination between imperfectly repeated chromosomal Rep sequences: a closer look at an old friend (F'(128) pro lac). *J. Bacteriol.* 185: 660–663.
- Kugelberg, E., E. Kofoed, A. B. Reams, D. I. Andersson, and J. R. Roth, 2006 Multiple pathways of selected gene amplification during adaptive mutation. *Proc. Natl. Acad. Sci. USA* 103: 17319–17324.
- Kugelberg, E., E. Kofoed, D. I. Andersson, Y. Lu, J. Mellor *et al.*, 2010 The tandem inversion duplication in *Salmonella enterica*: selection drives unstable precursors to final mutation types. *Genetics* 185: 65–80.
- Kuzminov, A., 1995 Collapse and repair of replication forks in *Escherichia coli*. *Mol. Microbiol.* 16: 373–384.
- Lane, D., R. de Feyter, M. Kennedy, S. H. Phua, and D. Semon, 1986 D protein of miniF plasmid acts as a repressor of transcription and as a site-specific resolvase. *Nucleic Acids Res.* 14: 9713–9728.
- Lawes, M., and S. Maloy, 1995 MudSacI, a transposon with strong selectable and counterselectable markers: use for rapid mapping

- of chromosomal mutations in *Salmonella typhimurium*. *J. Bacteriol.* 177: 1383–1387.
- Leach, D. R., 1994 Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *BioEssays* 16: 893–900.
- Lin, D., I. B. Gibson, J. M. Moore, P. C. Thornton, S. M. Leal *et al.*, 2011 Global chromosomal structural instability in a subpopulation of starving *Escherichia coli* cells. *PLoS Genet.* 7: e1002223.
- Lin, R. J., M. Capage, and C. W. Hill, 1984 A repetitive DNA sequence, rhs, responsible for duplications within the *Escherichia coli* K-12 chromosome. *J. Mol. Biol.* 177: 1–18.
- Liu, J., J. Sneeden, and W. D. Heyer, 2011 In vitro assays for DNA pairing and recombination-associated DNA synthesis. *Methods Mol. Biol.* 745: 363–383.
- Lovett, S. T., 2004 Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* 52: 1243–1253.
- Lovett, S. T., P. T. Drapkin, V. A. Suter Jr., and T. J. Gluckman-Peskind, 1993 A sister-strand exchange mechanism for recA-independent deletion of repeated DNA sequences in *Escherichia coli*. *Genetics* 135: 631–642.
- Lu, X. Y., M. T. Phung, C. A. Shaw, K. Pham, S. E. Neil *et al.*, 2008 Genomic imbalances in neonates with birth defects: high detection rates by using chromosomal microarray analysis. *Pediatrics* 122: 1310–1318.
- Lupski, J. R., and G. M. Weinstock, 1992 Short, interspersed repetitive DNA sequences in prokaryotic genomes. *J. Bacteriol.* 174: 4525–4529.
- Mazin, A. V., A. V. Kuzminov, G. L. Dianov, and R. I. Salganik, 1991 Mechanisms of deletion formation in *Escherichia coli* plasmids. II. Deletions mediated by short direct repeats. *Mol. Gen. Genet.* 228: 209–214.
- Mehan, M. R., N. B. Freimer, and R. A. Ophoff, 2004 A genome-wide survey of segmental duplications that mediate common human genetic variation of chromosomal architecture. *Hum. Genomics* 1: 335–344.
- Morag, A. S., C. J. Saveson, and S. T. Lovett, 1999 Expansion of DNA repeats in *Escherichia coli*: effects of recombination and replication functions. *J. Mol. Biol.* 289: 21–27.
- Morgan, L. V., 1925 Polyploidy in *Drosophila melanogaster* with two attached X Chromosomes. *Genetics* 10: 148–178.
- Morris, C. F., and R. Rownd, 1974 Transition of the R factor R12 in *Proteus mirabilis*. *J. Bacteriol.* 118: 867–879.
- Mortensen, U. H., C. Bendixen, I. Sunjevaric, and R. Rothstein, 1996 DNA strand annealing is promoted by the yeast Rad52 protein. *Proc. Natl. Acad. Sci. USA* 93: 10729–10734.
- Muller, H. J., 1936 Bar duplication. *Science* 83: 528–530.
- Muro-Pastor, A. M., and S. Maloy, 1995 Direct cloning of mutant alleles from the bacterial chromosome into plasmid vectors in vivo. *Biotechniques* 18: 386–388, 390.
- Narayanan, V., P. A. Mieczkowski, H. M. Kim, T. D. Petes, and K. S. Lobachev, 2006 The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. *Cell* 125: 1283–1296.
- Nicoloff, H., V. Perreten, and S. B. Levy, 2007 Increased genome instability in *Escherichia coli* lon mutants: relation to emergence of multiple-antibiotic-resistant (Mar) mutants caused by insertion sequence elements and large tandem genomic amplifications. *Antimicrob. Agents Chemother.* 51: 1293–1303.
- Ogawa, N., and K. Miyashita, 1995 Recombination of a 3-chlorobenzoate catabolic plasmid from *Alcaligenes eutrophus* NH9 mediated by direct repeat elements. *Appl. Environ. Microbiol.* 61: 3788–3795.
- Ohno, S., 1970 *Evolution by Gene Duplication*. Springer-Verlag, New York.
- Perals, K., F. Cornet, Y. Merlet, I. Delon, and J. M. Louarn, 2000 Functional polarization of the *Escherichia coli* chromosome terminus: the dif site acts in chromosome dimer resolution only when located between long stretches of opposite polarity. *Mol. Microbiol.* 36: 33–43.
- Petes, T. D., and C. W. Hill, 1988 Recombination between repeated genes in microorganisms. *Annu. Rev. Genet.* 22: 147–168.
- Petit, M. A., J. M. Mesas, P. Noirot, F. Morel-Deville, and S. D. Ehrlich, 1992 Induction of DNA amplification in the *Bacillus subtilis* chromosome. *EMBO J.* 11: 1317–1326.
- Poteete, A. R., 2009 Expansion of a chromosomal repeat in *Escherichia coli*: roles of replication, repair, and recombination functions. *BMC Mol. Biol.* 10: 14.
- Quiñones-Soto, S., and J. R. Roth, 2011 Effect of growth under selection on appearance of chromosomal mutations in *Salmonella enterica*. *Genetics* 189: 37–53.
- Ratray, A. J., B. K. Shafer, B. Neelam, and J. N. Strathern, 2005 A mechanism of palindromic gene amplification in *Saccharomyces cerevisiae*. *Genes Dev.* 19: 1390–1399.
- 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.
- Reams, A. B., and 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., and E. L. Neidle, 2004b Selection for gene clustering by tandem duplication. *Annu. Rev. Microbiol.* 58: 119–142.
- Reams, A. B., E. C. Kofoed, M. Savageau, and J. R. Roth, 2010 Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. *Genetics* 184: 1077–1094.
- Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry *et al.*, 2006 Global variation in copy number in the human genome. *Nature* 444: 444–454.
- Roth, J. R., N. Benson, T. Galitski, K. Haack, J. G. Lawrence *et al.*, 1996 Rearrangement of the bacterial chromosome: formation and applications, pp. 2256–2276 in *Escherichia coli and Salmonella*, edited by F. C. Neidhardt ASM Press, Washington, DC.
- Seaton, S. C., K. T. Elliott, L. E. Cuff, N. S. Laniohan, P. R. Patel *et al.*, 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.
- Seifert, H. S., and R. D. Porter, 1984 Enhanced recombination between lambda plac5 and F42lac: identification of cis- and trans-acting factors. *Proc. Natl. Acad. Sci. USA* 81: 7500–7504.
- Sekine, Y., K. Aihara, and E. Ohtsubo, 1999 Linearization and transposition of circular molecules of insertion sequence IS3. *J. Mol. Biol.* 294: 21–34.
- Sinden, R. R., G. X. Zheng, R. G. Brankamp, and K. N. Allen, 1991 On the deletion of inverted repeated DNA in *Escherichia coli*: effects of length, thermal stability, and cruciform formation in vivo. *Genetics* 129: 991–1005.
- Spielmann-Ryser, J., M. Moser, P. Kast, and H. Weber, 1991 Factors determining the frequency of plasmid cointegrate formation mediated by insertion sequence IS3 from *Escherichia coli*. *Mol. Gen. Genet.* 226: 441–448.
- Spies, M. S. K., 2005 Homologous recombination by the RecBCD and RecF Pathways, pp. 389–404 in *The Bacterial Chromosome*, edited by N. Higgins. ASM Press, Washington, DC.
- Spradling, A. C., 1981 The organization and amplification of two chromosomal domains containing *Drosophila* chorion genes. *Cell* 27: 193–201.
- Stern, M. J., G. F. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins, 1984 Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* 37: 1015–1026.
- Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita *et al.*, 1966 Frameshift mutations and the genetic code. *Cold Spring Harb. Symp. Quant. Biol.* 31: 77–84.

- Syvanen, M., J. D. Hopkins, T. J. t. Griffin, T. Y. Liang, K. Ippen-Ihler *et al.*, 1986 Stimulation of precise excision and recombination by conjugal proficient F⁺ plasmids. *Mol. Gen. Genet.* 203: 1–7.
- Tice, S. C., 1914 A new sex-linked character in *Drosophila*. *Biol. Bull.* 26: 221–230.
- Traxler, B. A., and E. G. Minkley Jr., 1987 Revised genetic map of the distal end of the F transfer operon: implications for DNA helicase I, nicking at oriT, and conjugal DNA transport. *J. Bacteriol.* 169: 3251–3259.
- Trinh, T. Q., and R. R. Sinden, 1991 Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. *Nature* 352: 544–547.
- Tsubota, S. I., D. Rosenberg, H. Szostak, D. Rubin, and P. Schedl, 1989 The cloning of the Bar region and the B breakpoint in *Drosophila melanogaster*: evidence for a transposon-induced rearrangement. *Genetics* 122: 881–890.
- Weinreich, M. D., A. Gasch, and W. S. Reznikoff, 1994 Evidence that the cis preference of the Tn5 transposase is caused by non-productive multimerization. *Genes Dev.* 8: 2363–2374.
- Williams, P. A., R. J. Ingebretsen, and R. J. Dawson, 2006 14.6 mT ELF magnetic field exposure yields no DNA breaks in model system *Salmonella*, but provides evidence of heat stress protection. *Bioelectromagnetics* 27: 445–450.
- Yu, D., J. A. Sawitzke, H. Ellis, and D. L. Court, 2003 Recombineering with overlapping single-stranded DNA oligonucleotides: testing a recombination intermediate. *Proc. Natl. Acad. Sci. USA* 100: 7207–7212.
- Zhang, F., C. M. Carvalho, and J. R. Lupski, 2009 Complex human chromosomal and genomic rearrangements. *Trends Genet.* 25: 298–307.

Communicating editor: S. Sandler

Appendix: The Size and Endpoint Distribution of Duplications with Junctions Other than IS3

The top line in Figure A1 is a genetic map of the affected region of the plasmid. Duplications with REP join points recurred repeatedly and the number of each type is presented. The REP-SJ hot spot is a small region near REPs that contains the endpoints of many SJ duplications. The junction types found in amplifications following prolonged growth under selection (Kugelberg *et al.* 2010) are presented at the bottom of the figure. One-third of selected

amplifications had REP junctions. The most common duplication (97%) found without selection (between IS3C and IS3A) was not found among selected amplifications. The size of REP duplications is dictated by the position of REP in the plasmid and is ~25 kb. The average size of SJ amplifications trapped nonselectively (65 kb) was larger than the average of the SJ duplications recovered after selection (34 kb). We suggest that all duplication types form prior to selection, but the lower fitness cost of shorter duplications allows them to be amplified successfully under selection.

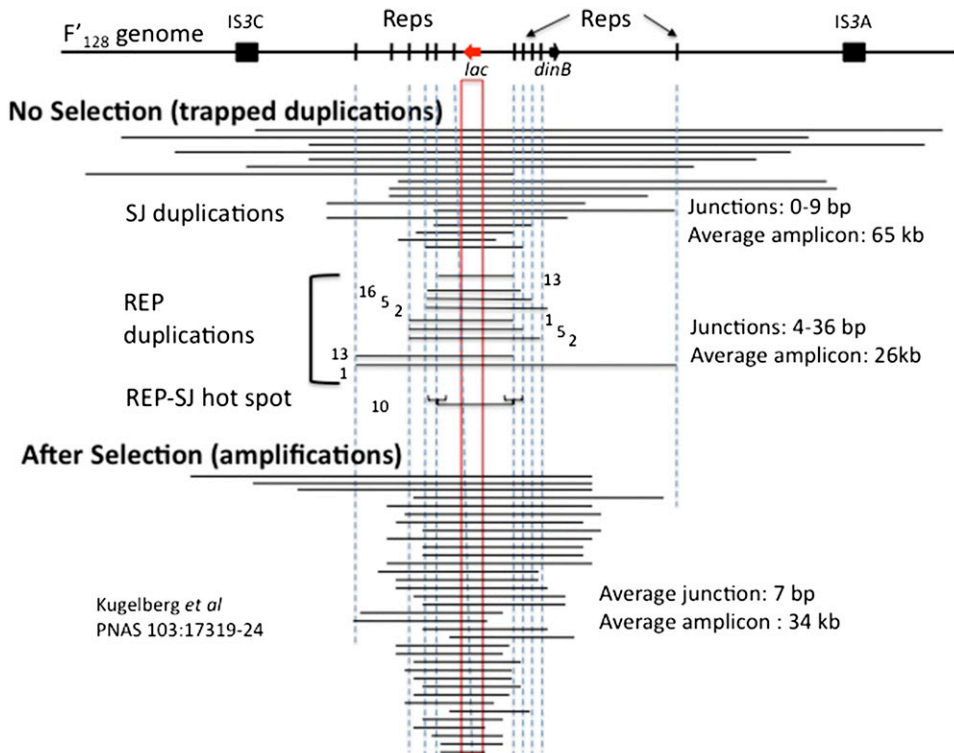


Figure A1 Duplications isolated with and without selection (SJ and REP types). Duplications other than the most frequent IS3C/IS3A type are described in the top half of the figure. The junctions of SJ duplications lie outside of any REP element. These duplications all include *lac* but their size is not otherwise constrained by the method used to trap them. The REP duplication all arise by exchanges between REP elements lying on either side of *lac*. The size of these duplications are constrained by the location of REP elements in the F'_{128} plasmid. The amplification-bearing strains described in the bottom half of the figure were all isolated as Lac⁺ colonies after prolonged growth under selection on lactose.