

## Modeling suggests frequency estimates are not informative for predicting the long-term effect of horizontal gene transfer in bacteria

Ann-Kathrin PETTERSEN<sup>1</sup>, Thomas BØHN<sup>2</sup>, Raul PRIMICERIO<sup>1,3</sup>, Paul R. SHORTEN<sup>4</sup>, Tanya K. SOBOLEVA<sup>4</sup> and Kaare M. NIELSEN<sup>1,2\*</sup>

<sup>1</sup> Department of Pharmacy, University of Tromsø, N9037 Tromsø, Norway

<sup>2</sup> Norwegian Institute of Gene Ecology, Science Park, N9294 Tromsø, Norway

<sup>3</sup> Department of Aquatic Biosciences, Norwegian College of Fishery Science, University of Tromsø, N9037 Tromsø, Norway

<sup>4</sup> AgResearch Limited, Ruakura Research Centre, Private Bag 3123, Hamilton, New Zealand

Horizontal gene transfer (HGT) is an important mechanism by which bacteria recombine and acquire novel genes and functions. Risk scenarios where novel plant transgenes transfer horizontally into bacteria have been addressed in numerical theoretical assessments and experimental studies. A key outcome of these studies has been that the frequencies of such inter-domain transfer are very low, if occurring at all, suggesting that such transfers would not occur at a level that is biologically significant. The relationship between transfer frequencies and the subsequent selection or genetic drift of transgene carrying bacteria often remains unresolved in these studies and assessments. Here we present a stochastic model to better understand the initial establishment and population dynamics of rare bacterial transformants carrying horizontally acquired (trans)genes. The following key parameters are considered: initial transformant numbers, strength of selection, bacterial population size and bacterial generations (time). We find that the initial number of transformants is important for the subsequent persistence of transformants only in the range of 1 to ~50 independent HGT events. Our simulations show that transformant populations under a wide range of HGT rates and selection coefficients undergo stochastic developments where they persist at low frequencies for up to several years (at frequencies that are below detection using available field sampling methodology), after which they eventually may go to fixation. Stochastic variability may thus play a crucial but disregarded role in the design of field monitoring strategies e.g. in biosafety assessments. We also estimate the time required for transformants to reach 0.0002% prevalence in a bacterial population, a threshold that allows experimental detection of transgene carrying bacteria through sampling of the larger bacterial populations.

**Keywords:** horizontal or lateral gene transfer / stochasticity / selection / population dynamics / survival / transgenes / biosafety

### INTRODUCTION

Bacteria acquire new genetic variability through random mutations and horizontal gene acquisitions (Lawrence and Hendrickson, 2003). Mutational processes have traditionally been considered to represent the main source of new genotypes in bacteria. However, horizontal gene transfer (HGT) is now shown to contribute substantially to the current composition and dynamics of bacterial chromosomes (Heinemann and Traavik, 2004; Nakamura et al., 2004; Ochman et al., 2000; Spratt et al., 2001). Some

studies suggest that HGT can be more important than mutations in producing new genotypes (Feil and Spratt, 2001; Townsend et al., 2003). The chance to gain a chromosomal nucleotide change by recombination compared to mutation can be as high as 20:1 (Feil and Spratt, 2001; Lawrence, 2002). The observation of HGT in bacterial genomes is, however, not indicative of the HGT frequency alone, but rather of the combined effect of donor DNA release or transfer, bacterial DNA uptake

\* Corresponding author: knielsen@farmasi.uit.no

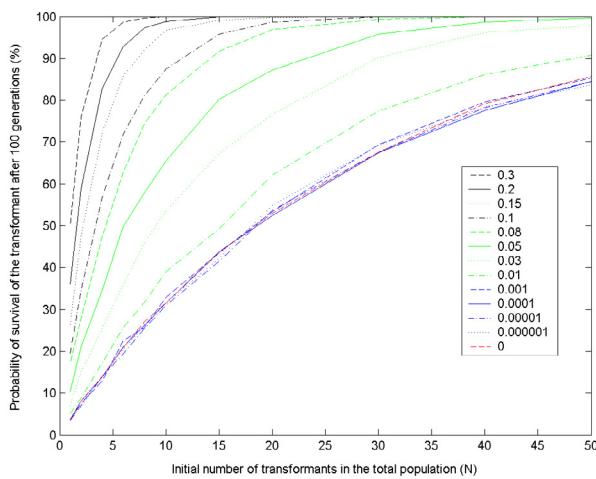
and integration, and genetic drift or positive selection of the transformed cells (Nielsen and Townsend, 2001, 2004). The acquisition of a genetic trait conferring a selective advantage in a bacterium can take place within the time span of a few bacterial generations, nevertheless, it may subsequently take many thousands of generations and many years before the trait will become widespread after clonal division and directional selection in the total bacterial population. Whereas the acquisition of shorter DNA fragments leading to minor nucleotide changes in bacteria *e.g.* introduction of single nucleotide polymorphisms (SNPs) in housekeeping genes may be effectively neutral and persist over time due to random genetic drift (Maynard Smith et al., 2000; Spratt et al., 2001), most horizontally acquired chromosomal genes in bacteria depend on directional selection to determine their persistence and transmission success over bacterial generations. Horizontally transferred DNA that results in a positively selected transformant has the potential to rapidly disseminate in a larger bacterial population. The rate depends mainly on the strength of selection (for non-mobile DNA). However, most non-homologous DNA inserted into the bacterial chromosome will undergo negative selection resulting in loss of the acquired genes from the population (Elena et al., 1998). Neutrally selected acquired genes are only affected by genetic drift and have the same dynamics to fixation as neutral mutations. Protein-coding genes under weak or near-neutral selection are, however, unlikely to persist in a bacterial population over time without acquiring mutations including small and large deletions, hence, they will ultimately lose their function and will be lost from the population. Pseudogenes are only infrequently found in bacterial genomes (Lawrence et al., 2001).

Selection is defined as “nonrandom differential survival or reproduction of classes of phenotypically different entities” (Futuyma, 1998). This entity may be a gene, a bacterial cell, a group, or a bacterial population (Day and Otto, 2001). In haploids such as bacteria, selection is realized as differential population growth (Hartl and Clark, 1997). The Malthusian fitness parameter ( $m$ ) is often used in fitness calculations since bacterial populations have overlapping generations and short generation time (Hartl and Clark, 1997). The rate horizontally transferred DNA is spread within a bacterial population is thus a function of its effect on bacterial fitness (Burke and Riesenbergs, 2003). However, due to stochasticity at low initial numbers, even positively selected bacterial transformants are highly vulnerable to extinction. To understand and eventually predict the outcome of rare HGT events, two consecutive steps must

be understood; i) the stochastic (*i.e.* random) processes underlying transformant extinction or establishment from low numbers, and ii) the deterministic development under a given selection criterion after establishment (Nielsen and Townsend, 2001, 2004; Rouzine et al., 2001). Fisher (1930) and Wright (1931) investigated random drift in small populations using a binomial sampling model to derive the distribution of gene frequencies in populations (Hartl and Clark, 1997). Random genetic drift causes stochastic changes in gene frequencies and can be modeled as a Markov Chain (Maruyama and Kimura, 1975). To calculate the time to and probability of fixation, Kimura (1954, 1962) developed a diffusion model based on the Wright-Fisher model. The diffusion approximation has been widely used to understand the probability of fixation and average time to loss or fixation of (mainly disadvantageous) mutants in diploid populations (Kimura and King, 1979; Kimura and Ohta, 1973; Lande, 1994, 1998; Maruyama and Kimura, 1975; Nei and Roychoudhury, 1973; van Herwaarden and van der Waal, 2002). Kimura and Ohta (1969), and Barker and Butcher (1966) performed Monte Carlo experiments to test the results obtained with theoretical predictions and the results corresponded well with the theoretical work. The majority of this early work has focused on the fate of deleterious mutations in diploid organisms.

Based on the approaches mentioned above, we have developed a model for a haploid population where different levels of selective advantages (relative fitness) are assigned to the transformant population relative to the wild-type. We use the model to better understand how selection and stochasticity influence the outcome of rare HGT events (*e.g.* of transgenes) occurring in large bacterial populations. We use fixed selection coefficients, and focus on the initial stochastic phase for positively selected genes and the bacterial generations following rare HGT events. This initial stochastic phase may last a considerable number of generations, depending on selection and rate of HGT, expressed through the initial number of transformants. Parameters varied in our model are the size of the initial transformant population (or initial number of HGT events) as a subset of the larger total population, duration of the growth period (time in bacterial generations) and different advantages of the transferred DNA expressed as selection coefficients.

The survival of transformants is largely stochastic until they have established at a relatively large number. Each run of our model simulates one realization of this stochastic growth process. From the simulations, we calculate the probability of survival for the transformant



**Figure 1.** Effects of increasing the selection coefficients and the initial number of transformants on their probability of survival. Selection coefficients ( $s$ ) range from neutral ( $s = 0$ ) to strong positive selection ( $s = 0.3$ ). The lower dotted line shows neutral selection where visible population changes are due to genetic drift. The total bacterial population size is  $10^9$ .

from the number of simulated transformant populations alive after a certain number of generations.

In the recent work of Nielsen and Townsend (2004), the probability of detecting rare transformants in large bacteria populations is discussed. We use their parameters and our model to estimate the probability of transformants reaching detectable proportions within different numbers of generations. We also calculate the probability that the transformants are alive after a certain number of generations but remain under the detectable level (defined here as the minimum proportion of transformants in the sampled population that allow detection).

Drawing on population genetic theory developed for diploid populations, we show how selection determines the outcome of rare HGT events, allowing for a more precise hypothesis to be tested in HGT studies, an improved design of field and experimental monitoring studies of HGT, and a better-informed biological risk assessment of genetically modified organisms and their potential long-term impact on the environment.

## RESULTS

### The role of the initial number of transformants and selective advantage on the probability of transformant survival

The probability of transformant survival after 100 generations as a function of the initial number of

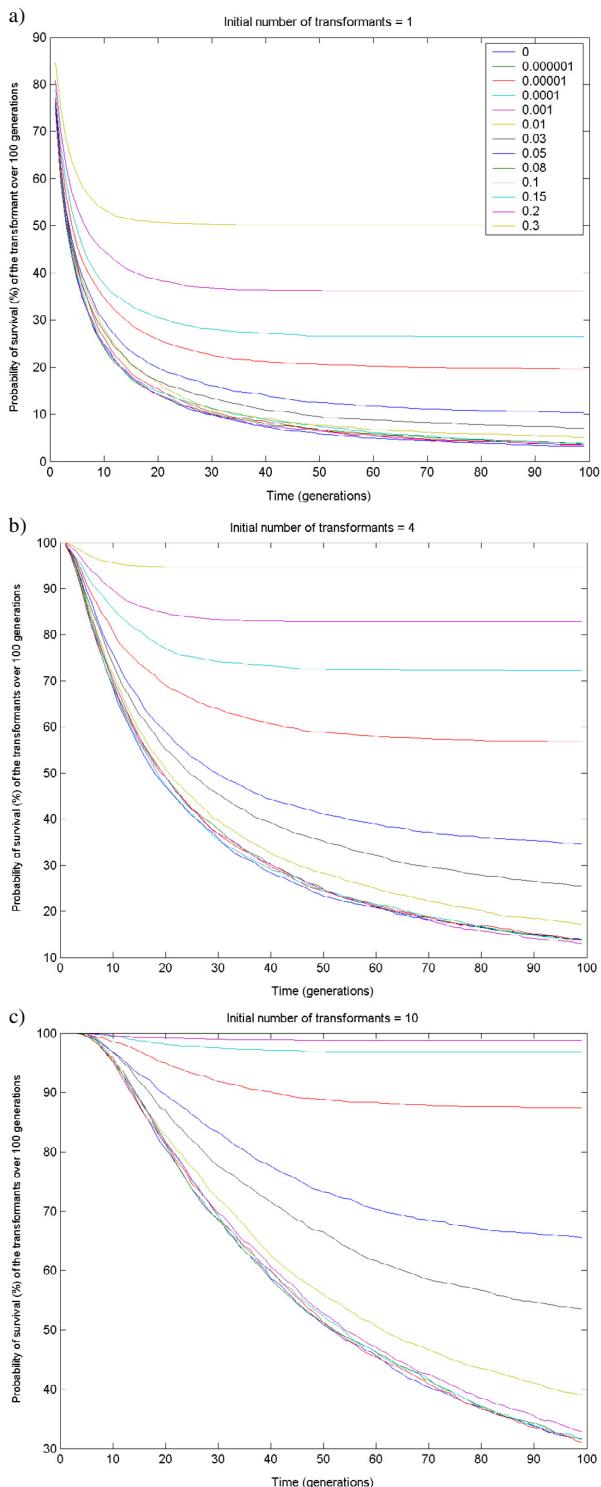
transformants and the selective advantage of the transformant is shown in Figure 1. The probability of transformant survival increases with increasing selective advantage and initial transformant numbers. Selective advantage does not make a substantial contribution to the probability of transformant survival unless the selection is strong ( $s \geq 0.01$ ). Therefore, transformants that exhibit weak positive selection require a larger initial number of transformants in order to survive. For example, 20 initial transformants with selection of  $s \leq 0.001$ , have approximately the same probability of surviving after 100 generations as one initial transformant with a selection coefficient  $s = 0.3$ .

A single transformant experiencing weak positive selection ( $s = 0.000001$  to  $0.001$ ) has approximately a 5% probability of surviving after 100 generations. Strong selective advantage sharply reduces the number of initial transformants required for their survival over 100 generations. Nevertheless, a single initial transformant with strong positive selective advantage ( $s \geq 0.08$ ) has approximately a 50 to 80% probability of extinction. Raising the initial number of transformants rapidly reduces the probability of extinction. From high initial transformant numbers, the transformant population may persist at low prevalence over extended periods of time, and also temporarily decrease before eventually becoming fixed. For example, with 10 initial transformants and very strong selection ( $s \geq 0.15$ ), the probability of transformant extinction after 100 generations is less than 5%.

### Time of transformant extinction after introduction

The probability of transformant survival for up to 100 generations for different selection coefficients and initial transformant population sizes ( $n_0 = 1, 4, 10$ ) is shown in Figure 2. There is a high probability of transformant extinction in the generations immediately following the HGT event(s). If the transformant population goes to extinction, then there is at least a 95% chance that extinction occurs within the first 65 generations when the initial number of transformants is one and the selection coefficients are low (see Fig. 2a).

With generation times of 30 min., 8 hours or 3 weeks, weakly positively selected transformants must survive approx. 30 generations (15 hours, 10 days or 90 weeks, respectively), and strongly positively selected transformants 10–20 generations (5–10 hours, 3–6 days or 30–60 weeks, respectively) to establish at numbers that prevent eventual extinction to occur. With strong positive selection coefficients, a transformant population may,



**Figure 2.** Probability of transformant survival during the first 100 generations with different selection coefficients. The initial number of transformants is 1 (a), 4 (b), or 10 (c). The total bacterial population size is  $10^9$ .

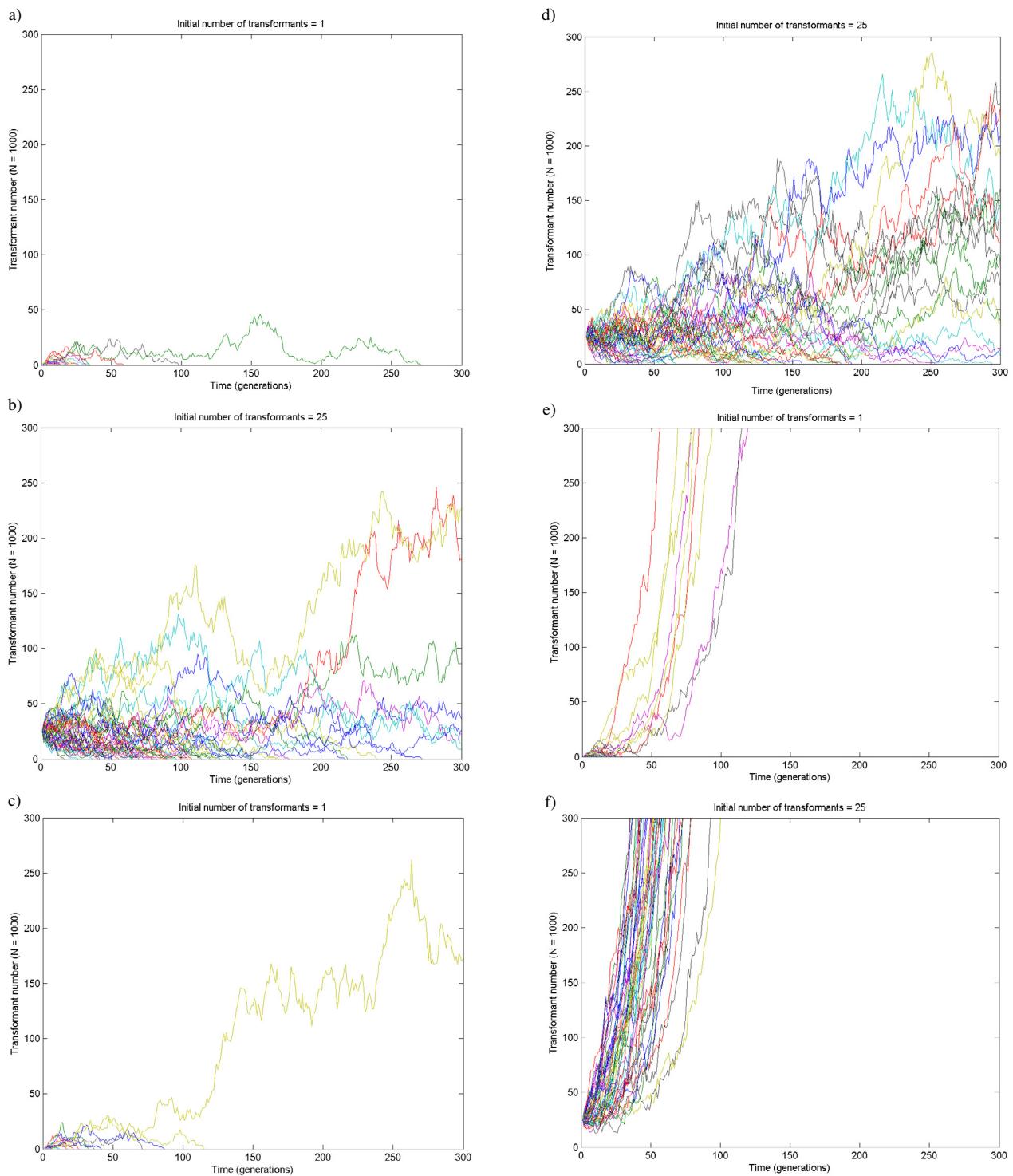
at high probability, amplify and go to fixation even when the initial number of transformants is as low as one individual. Such high selection coefficients are rarely documented in the literature. Low selection coefficients and a low number of initial transformants in the population leads to population extinction in about 95% of the cases. This is a high extinction rate, but it is important to note that in 5% of the independently starting populations, the transformants will be still alive. Repeated HGT events in time or space will add to the probability of long term survival and impact, but are not considered further here, as we focus on rare transformants arising after single exposure to a (trans)gene sources for a limited time.

If a positively selected transformant population survives the initial critical time period, it generally establishes and gradually out-competes the wild-type population. However, for low initial transformant numbers and low selection coefficients, the initial stochastic phase where both populations coexist may last from 100 up to at least 10000 generations, which may represent several years with a prevalence that is under the limit of detection (Nielsen and Townsend, 2004).

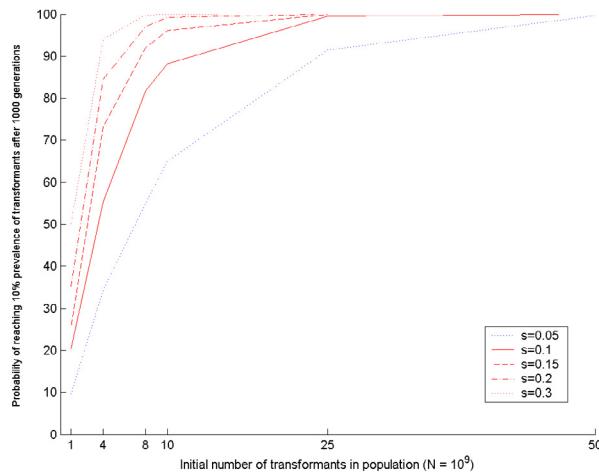
#### Variability in the population dynamics of transformants subject to neutral, weak or strong positive selection

Figure 3 illustrates the large random variation in the population dynamics of transformants within a bacterial population ( $N = 1000$ ), for 50 independent starting populations (number of permutations). As seen, the fate of each of the 50 populations, experiencing either neutral (Fig. 3a,b), weakly positive (Fig. 3c,d) or strongly positive (Fig. 3e,f) selection, varies considerably between independent starting populations. As seen in Figure 3a, starting with one transformant, all of the transformant populations go extinct over 300 generations examined under neutral selection. However, one population reaches a considerable relative population size before going extinct. A similar trend is seen for weakly positively selected transformant populations although they persist longer before extinction occurs. Under neutral selection, with a high initial number of transformants (25) (Fig. 3b), some of the transformant populations rise to high proportions over the 300 generations examined. The transformants have a probability of approximately 10% to survive and reach fixation in the small population ( $N = 1000$ ) within the first ~100 generations. Thus, although the underlying population genetic parameters are identical, the fate of individual HGT events varies considerably among independent realizations of the growth process.

## Modeling the fate of horizontally transferred genes



**Figure 3.** The dynamics of 50 independently starting transformant populations over 300 generations. Each line shows the trajectories of a single population. The initial population size is 1 (a) or 25 (b) with no selection, 1 (c) or 25 (d) with weak positive selection ( $s = 0.0002$ ), or 1 (e) or 25 (f) with strong positive selection ( $s = 0.1$ ). The (constant) total population size is 1000.



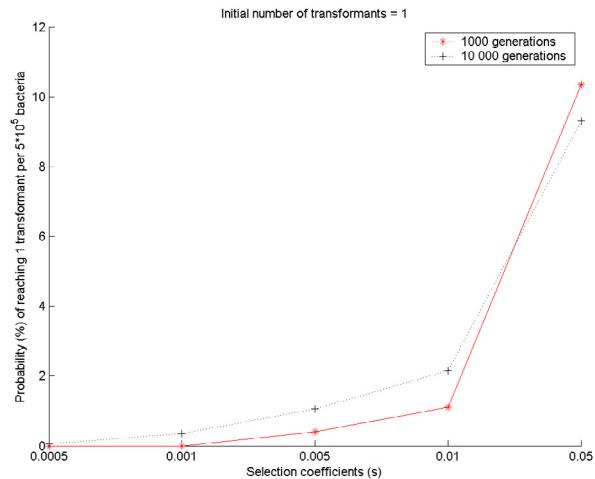
**Figure 4.** The probability to reach a 10% prevalence of transformants after 1000 generations as dependent on the initial number of transformants and different selection coefficients. The total population size is  $10^9$ . The probability of reaching a 10% prevalence level of transformants is close to zero for selection coefficients lower than 0.05.

#### Time to reach 10% prevalence of transformants in a larger bacterial population

For the HGT event to have a long-term impact, it must reach proportions that are biologically relevant within the larger bacterial population. However, the presence of the HGT event in all of the members of the transformed species/strain (fixation) is not necessary for an observable community effect to appear. Therefore, it is not possible to directly relate the time required for a transformant population to reach fixation with transformant proportions that are biologically meaningful. The proportions (and hence time needed) of transformants necessary in a bacterial community to create a biological effect are likely to vary with the specific HGT event, the selectable trait (transgene) transferred and the response in a particular ecological community. Figure 4 shows the probability to reach a prevalence of 10% transformants within a larger ( $N = 10^9$ ) wild-type bacterial population after 1000 generations. For the 1 to 50 initial transformants with selection coefficients substantially below 0.05 and for neutral selection, the probability of reaching a 10% prevalence in 1000 generations is close to zero, and hence not visible in the graph.

#### Probability of reaching a detectable proportion of transformants with time

Nielsen and Townsend (2004) reviewed field HGT monitoring studies and concluded that there must be at



**Figure 5.** Probability of reaching the detection limit (1 transformant per  $5 \times 10^5$  bacteria, Nielsen and Townsend, 2004) after 1000 and 10000 generation with 1 initial transformant in a population of  $10^9$  bacteria using different selection coefficients. After 100 generations, the probability is zero over all selection coefficients (not shown).

least one transformant per  $5 \times 10^5$  bacteria in the sample analyzed in order for current methodology to be able to detect the transformant. This equals a proportion of 0.0002% transformants. The initial numbers of transformants, and hence proportions in our examples are 1 to 50 per  $10^9$  cells, or  $1 \times 10^{-9}$  to  $5 \times 10^{-8}$ . This is a small number of transformants and comparable to those produced in experimental gene transfer studies in soil (Nielsen et al., 1997, 2000). Figure 5 shows the probabilities that the descendants of the rare initial transformants will reach detectable proportions after 1000 to 10000 generations (see Tab. 1 for the relationship between bacterial generations and time).

After 100 generations, the probability of reaching detectable proportions of transformants is close to zero for all selection coefficients examined ( $s = 0.0005, 0.001, 0.005, 0.01, 0.05$ ). Thus, sampling only a few generations after the HGT event(s) has happened almost certainly will not detect any transformant bacteria (with  $s < 0.05$ ) even though they are there, and under positive selection. For higher selection coefficients, the outcome will be that the transformants are on their way to rapid fixation with little effect of stochasticity (Nielsen and Townsend, 2001). Our stochastic model is focused on understanding the effects of smaller selection coefficients where chance is important for the outcome (survival or death).

After 1000 generations, the probability of reaching detectable proportions is very low for selection coefficients  $< 0.001$  and only about 0.5% for a selection

**Table 1.** Time required for a bacterial population to reach 100, 1000 or 10000 generations.

Generation time	Generations		
	100	1000	10000
0.5 hour	2.1 d	21 d	210 d
8 hours	33 d	333 d	9 y
1 day	100 d	≈ 3 y	28 y
14 days	≈ 4 y	≈ 40 y	385 y
21 days	≈ 6 y	58 y	577 y

d = days, y = years.

coefficient of  $s = 0.005$ , 2% for  $s = 0.01$  and around 10% for  $s = 0.05$ . After 10000 generations, the probability to reach detectable proportions is around 0.2% for  $s = 0.0005$  and  $s = 0.001$ , 1%,  $s = 0.005$ , 2% for  $s = 0.01$  and around 10% for  $s = 0.05$ . The similar probability after 1000 and 10000 generations suggests that most populations have died out and the survivors have established at high numbers in the population.

In practical sampling of bacterial communities, it is important to consider the limitations of the sampling methodology and the probability that transformants are present below the limit of detection. Our model has been used to calculate the prevalence of transformants in a population of  $10^9$  after 100, 1000 and 10000 generations with two different selection coefficients (0.001 and 0.01 respectively), as a result of a single event of HGT. With low  $s$  ( $s = 0.001$ ), the probability that the number of transformants exceeds the detectable proportion (sample size 50000) after 100 and 1000 generations is close to zero. Yet, it is a real probability that some transformants remain alive. After 10000 generations, the probability of the transformants reaching a detectable proportion is 0.18%. With higher  $s$  ( $s = 0.01$ ), the probability to reach detectable proportions remains close to zero after 100 generations. However, after 1000 generations the probability of reaching detectable proportions is 1.7%, and after 10000 generations, the probability is about 2%.

## DISCUSSION

Horizontal gene transfer (HGT) has, through comparisons of sequenced bacterial genomes, been shown to be an important mechanism by which bacterial communities share, recombine and acquire novel functions (Eisen 2000; Nakamura et al., 2004; Ochman et al., 2000). The identification of patterns of horizontally transferred genes and mobile elements in sequenced bacterial genomes rarely provides information on the underlying processes

such as: the specific DNA sources, the vectors, the exact time and location where the HGT event(s) occurred, the HGT rate to which the sequenced genome was exposed to the same DNA fragment prior to the successful integration event, the frequency to which other organisms are or have been exposed to the same DNA fragment, or the number of bacteria carrying the same HGT event that escapes detection or has gone extinct during the course of evolution. It is therefore important that complementary approaches to understand HGT processes are developed. These include the development of: theoretical approaches to understand the population dynamic aspects of HGT events, field methodology to monitor HGT events occurring in natural bacterial communities, and experimental laboratory approaches that can identify and characterize the molecular, cellular and environmental conditions and factors governing the HGT process.

Progress in the understanding of HGT processes is not only important to improve our basic knowledge of one of the main causes of evolutionary change in bacterial communities, but is also for experimentally promoting advantageous HGT events in directed evolution in metabolic engineering, and in actively preventing unwanted HGT events (e.g. of antimicrobial resistance genes and pathogenicity traits in disease-causing bacteria). Moreover, concern over potential HGT of transgenes from genetically modified organisms to bacteria has called for detailed knowledge and risk evaluations of the factors preventing or promoting HGT events in natural bacterial communities (EFSA, 2004; Nielsen, 2003; Nielsen et al., 1997, 2005).

With the exception of the spread of mobile genetic elements, most HGTs occurring between distantly related or unrelated bacterial species occur infrequently when measured in an experimental time perspective. Several barriers to the HGT process have been described (Majewski and Cohan, 1998; Majewski et al., 2000; Vulic et al., 1997). Thus, successful HGT into the bacterial chromosome is likely to be rare and HGT events persisting and spreading in bacterial populations are likely to do so due to selective advantage, eventually in combination with a continuous supply of new transformants.

To better understand the population dynamics of rare bacterial transformants carrying horizontally acquired genes, we use a model that simulates the stochastic growth dynamics of transformants within larger bacterial populations by considering initial transformant numbers, numbers of bacterial generations, and the strength of selection of the transformants. The number of transformants surviving to generation  $i$  is computed according to a binomial frequency distribution, where

each individual is independent of the others and has constant fitness over time. Density-dependence, population bottlenecks, or limited population growth is not explicitly considered in the model. This is an obvious simplification of any ecological system, as individuals always interact within and between species, and may have changing fitness over time. Evolutionary outcomes are indeed difficult to estimate (Day and Otto, 2001). Nevertheless, since our model is mainly concerned with the fate of rare bacterial transformants within their first generations, spatial considerations including density dependent interactions or bottlenecks are of less concern here and will be addressed in separate work. When it comes to fixation of a beneficial mutant, Heffernan and Wahl (2002) showed that the inclusion of population bottlenecks, in a similar model, increases the probability of fixation compared to a situation where the population size is constant (which we use). Thus, our approach is on the conservative side.

We focus on the probability that rare transformants present in a larger population will establish and multiply to numbers that permit detection and long-term survival. All populations are vulnerable to allee-effects *i.e.* extinction from low densities (Drake, 2004; Itoh et al., 2004; Pimm, 1991). This applies to most introduced species and genotypes during the process of establishment, even when they have a selective advantage and a positive growth rate. Extinction from a low number of individuals may be modeled as a stochastic process. In contrast, rare populations with positive growth rates cannot go extinct in deterministic models, thereby questioning the utility of these models to resolve the initial phase of transformant survival (Heffernan and Wahl, 2002; Nielsen and Townsend, 2004). Our model demonstrates that the probability of survival of rare transformants with weak positive selection is almost exclusively dependent on random survival and on the number of initial transformants. A decrease in transformant extinction is mostly a function of increasing the initial transformant population size up to approximately 50 individuals, or repeating the introduction of single HGT events to ~50 times, since the transformants behave independently in the model. Our results indicate that the specific HGT rate is marginally important to the long-term outcome as long as some transfers into the bacterial population occur over time. Similarly, Fisher (1930) stated that new mutations have a high chance of being eliminated, but since they repeatedly occur, some will go to fixation.

This prediction questions the focus of the majority of experimental HGT studies performed to date, where often the main objective has been to observe and quantify HGT

rates in various experimental and natural systems with little consideration of the population dynamics of rare bacterial transformants under different selective regimes. Our results emphasize that HGT rates are not predictive of the long-term impact of HGT events as long as they occur within a 10-fold range above zero repeatedly over time. In this perspective, we would like to point to the fact that the common classification of HGT rates as "high" and "low" should be understood as an artificial grading system used to describe the range of measurable HGT rates in field and laboratory samples. The grading of frequencies is thus an artifact of the range of frequencies that can be measured applying currently available methodology to minute sample sizes. The grading is not linked to biological relevance of the HGT process as often erroneously implied. Thus, it is not meaningful to indicate the biological impact of measurable HGT events alone without simultaneous consideration of selection and drift of the descendants of the primary transformants. The exception to this being HGT events occurring continuously at sufficient high frequencies to allow persistence in a bacterial population despite negative selection, or HGT events of mobile elements that act independently, or at the expense, of the bacterial chromosome.

Selection stands out as the major factor determining the fate of rare HGT events in larger bacterial populations. However, selection does not make a substantial contribution to initial survival of rare transformants unless it is strong ( $s > 0.01$ ). Most single transformant populations experiencing positive selection lower than  $s = 0.01$  have similar survival probabilities to neutrally selected transformants, demonstrating the largely stochastic survival probability in the first few generations following the HGT events. The selection coefficient expresses the probability of survival of the transformant as compared to the wild-type bacteria. It is used here as a theoretically given value that does not specifically relate to any gene. Since we are interested in the effect of natural selection on transformant survival probabilities, we have adopted selection coefficients ranging from neutral or near-neutral to strongly positive. Since rare HGT events carrying negatively selected traits are likely to go extinct (Nielsen and Townsend, 2001), we do not consider these further here. The range of selection coefficients used is within those reported by others. For instance, Imhof and Schlötterer (2001) measured the fitness parameter (selection coefficient) for beneficial mutations in *Escherichia coli*. The Malthusian fitness ( $m$ ) ranged from 0.0006 to 0.059 with an average of 0.02. Lenski et al. (1991) used selection coefficients for beneficial mutants

between 0.05 to 0.2. Due to environmental fluctuations and complex spatial population structures, it is difficult to quantify selection in natural environments. Selection coefficients in nature that are considerably smaller than the selection coefficients regarded here are close to selective neutrality, and will make little difference to the initial stochastic establishing phase of the transformants. Selection coefficients on selective agar-plates can be regarded as close to  $s = 1$ , *i.e.* all sensitive bacterial cells die immediately upon exposure to the antibiotic.

The fate of transformants within their first few generations is crucial for their long-term impact in the bacterial population. Most transformant extinction occurs during the first five generations following the HGT events. Depending on the bacterial phenotype and environment, bacterial generation time may vary between minutes to months. Thus, rare transformants will be lost from the larger bacterial population within hours or weeks, respectively. The time to fixation or loss is shortened by selection (Maruyama and Kimura, 1974). The large variability in survival and survival time of the independently starting populations is clearly seen in Figure 3, thus demonstrating the random outcomes within bacterial populations experiencing the same environmental parameters. Consideration of the large stochastic variability in observable transformant numbers is important when designing field monitoring strategies of HGT events and when drawing inferences from observed transformant population dynamics over discrete sampling time points.

Given immediate survival, a positively selected transformant population is likely to rise in proportion in the bacterial population until fixation. The time to fixation depends on the total population size. However, the relevant proportions of transformants in a larger bacterial population that will impact the system dynamics and interactions will vary with population structure and the horizontally transferred trait. Only considering a 100% fixation endpoint of transformants is not informative to understand the biology of dynamically evolving bacterial populations. Most genes are unlikely to be fixed within larger bacterial populations that have panmictic, or epidemic population structures (Maynard Smith et al., 2000). In larger bacterial populations, weakly selected HGT events may require several thousand years before fixation (Nielsen and Townsend, 2004). The average time to fixation of a neutral mutant is  $2N$  (Kimura and Ohta, 1969), and the probability of fixation of a neutral mutant is  $1/N$  for haploid populations (Kimura, 1957, 1962). The rate at which neutral mutations are fixed is equal to the neutral mutation rate (Kimura, 1983). Haldane (1927) showed that the probability of fixation of a mutant allele

with selective advantage  $s$  is approximately  $2s$ , and this agrees well with Kimura's result.

In summary, our simulations show that the selection coefficient and the initial size of the transformant population are crucial parameters for the establishment of horizontally transferred genes (transformants) in a larger bacterial population. The model predicts how the fate of various transformant populations is determined by selection, thereby allowing optimized design of experimental systems that seek to quantify various selection scenarios. This will be of particular value in biosafety assessments of GMOs where hypothesized ranges of selection can be analyzed to predict environmental outcomes if rare HGT events occur. Moreover, understanding the relevant range of selection can inform the design of appropriate GMO monitoring strategies (Nielsen and Townsend, 2004), and provide a basis for experimental verification of the model's outcome through long-term serial transfer experiments.

In the future, we aim to (i) develop spatial models to better understand the dynamics of HGT events occurring in structured environments, (ii) develop models of HGT that incorporate variable selection intensity and different types of density-dependent population regulation, and (iii) examine whether horizontally transferred genes can be categorized with regard to the different types of selection they induce.

## METHODS

Our stochastic model describes the population dynamics of rare bacterial transformants (*e.g.* carrying a transgene) in a larger clonally dividing bacterial (wildtype) population during the initial period where the transformants either establish or become extinct. A transformant's ability to become established depends on its selective advantage, the HGT rate and stochastic factors. Although we focus on HGT events that occur by natural transformation, the model is equally applicable to other HGT mechanisms resulting in transfer of genes co-replicating with the bacterial chromosome. In the model, an initially small number of transformants is introduced into the wildtype population, and no further HGT within the bacterial population is permitted. The introduced transformant bacteria are assumed to divide periodically. After division, each individual transformant survives independently with a fixed probability  $p = 1/(2-s)$ , where  $s$  is the selective advantage of the transformant. For neutral selection ( $s = 0$ ) the probability of transformant survival is 0.5 and it gradually increases with increasing selective advantage (higher relative fitness). We assume that the

**Table 2.** Values of the parameters used in the model.

Parameter	Value / range
$N$	$10^9$
$G$	50, 100, 300, 1000, 10000
$T$	2000 – 5000
$s$	0, 0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.03, 0.05, 0.08, 0.1, 0.15, 0.2, 0.3
$t_0$	Number of transformants <sup>a</sup> in the starting population 1 – 50

<sup>a</sup> Transformants are defined here as those bacteria that are identical to the remaining total bacterial population,  $N$ , except that they carry a horizontally transferred gene that change transformant selection according to  $s$ .

probability of individual transformant survival is density-independent, which is a reasonable assumption during the initial stages of transformant establishment. Furthermore, the total population size ( $N = 10^9$ ) is assumed to be constant and equal to the carrying capacity of the system. If the number of transformants at generation  $i$  is  $n_i$ , then it follows that the number of transformants surviving to generation  $i + 1$  ( $n_{i+1}$ ) ranges from 0 (all transformants die) to  $2n_i$  (all transformants survive) according to a binomial frequency distribution  $B(2n_i, p)$  (Hartl and Clark, 1997). The model is therefore

$$\begin{aligned} n_{i+1} &\sim B(2n_i, p) \\ y_{i+1} &= N - n_{i+1} \end{aligned} \quad (1)$$

where  $y_{i+1}$  denotes the wildtype population size at generation  $i + 1$  and  $n_{i+1}$  denotes a realization of the binomial distribution  $B(2n_i, p)$ . Equation (1) can be sequentially used to determine the change in the frequency distribution of transformants in the population over time. The model was implemented in MATLAB (www.mathworks.com) using the standard normal approximation to the binomial distribution for large numbers of transformants. Between 2000 to 5000 realizations of the stochastic growth process were used to generate the frequency distributions of the transformant population size. See Table 2 for a description of the various parameters used.

## ACKNOWLEDGEMENTS

A.-K. Pettersen, R. Primicerio, T. Bøhn and K.M. Nielsen acknowledge financial support from the Research Council of Norway. We thank J. Townsend, D. Daffonchio, O. Lazebny, P. Roughan and J. Eggert for comments on the manuscript, and J. Pope for suggestions on the model.

Received October 28, 2005; accepted March 15, 2006.

## REFERENCES

- Barker JSF, Butcher JC** (1966) A simulation study of quasi-fixation of genes due to random fluctuation of selection intensities. *Genetics* **53**: 261–268
- Burke JM, Riesenbergh LH** (2003) Fitness effects of transgenic disease resistance in sunflowers. *Science* **300**: 1250
- Day T, Otto SP** (2001) Fitness. In Encyclopedia of life sciences, Nature Publishing Group / www.els.net, pp 1–6
- Drake JM** (2004) Allee effects and the risk of biological invasion. *Risk Analysis* **24**: 795–802
- EFSA – European Food Safety Authority** (2004) Opinion of the scientific panel on genetically modified organisms on the use of antibiotic resistance genes as marker genes in genetically modified plants. *EFSA J.* **48**: 1–18
- Eisen J** (2000) Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. *Curr. Opin. Genet. Dev.* **10**: 606–611
- Elena SF, Ekunwe L, Hajela N, Oden SA, Lenski RE** (1998) Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. *Genetica* **102–103**: 349–358
- Feil EJ, Spratt BG** (2001) Recombination and the population structure of bacterial pathogens. *Annu. Rev. Microbiol.* **55**: 561–590
- Fisher RA** (1930) The genetical theory of natural selection, Clarendon Press, Oxford
- Futuyma DJ** (1998) Evolutionary Biology, 3rd edn, Sinauer Associates, Sunderland
- Haldane JBS** (1927) A mathematical theory of natural and artificial selection. Part V. Selection and mutation. *Proc. Camb. Phil. Soc.* **28**: 838–844
- Hartl DL, Clark AG** (1997) Principles of population genetics, 3rd edn, Sinauer Associates, Sunderland
- Heffernan J, Wahl LM** (2002) The effects of genetic drift in experimental evolution. *Theor. Popul. Biol.* **62**: 349–356
- Heinemann JA, Traavik T** (2004) Problems in monitoring horizontal gene transfer in field trials of transgenic plants. *Nat. Biotechnol.* **22**: 1105–1109

- Imhof M, Schlötterer C** (2001) Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. *Proc. Natl. Acad. Sci. USA* **98**: 1113–1117
- Itoh Y, Tainaka K, Sakata T, Tao T, Nakagiri N** (2004) Spatial enhancement of population uncertainty near the extinction threshold. *Ecol. Model.* **174**: 191–201
- Kimura M** (1954) Process leading to quasi-fixation of genes in natural populations due to random fluctuation of selection intensities. *Genetics* **39**: 280–295
- Kimura M** (1957) Some problems of stochastic-processes in genetics. *Ann. Math. Stat.* **28**: 882–901
- Kimura M** (1962) On the probability of fixation of mutant genes in a population. *Genetics* **47**: 713–719
- Kimura M** (1983) The neutral theory of molecular evolution, Cambridge University Press, Cambridge
- Kimura M, King JL** (1979) Fixation of a deleterious allele at one of two “duplicate” loci by mutation pressure and random drift. *Proc. Natl. Acad. Sci. USA* **76**: 2858–2861
- Kimura M, Ohta T** (1969) Average number of generations until extinction of an individual mutant gene in a finite population. *Genetics* **61**: 763–771
- Kimura M, Ohta T** (1973) The age of a neutral mutant persisting in a finite population. *Genetics* **75**: 199–212
- Lande R** (1994) Risk of population extinction from fixation of new deleterious mutations. *Evolution* **48**: 1460–1469
- Lande R** (1998) Risk of population extinction from fixation of deleterious and reverse mutations. *Genetica* **102/103**: 21–27
- Lawrence JG, Hendrix RW, Casjens S** (2001) Where are the pseudogenes in bacterial genomes? *Trends Microbiol.* **9**: 535–540
- Lawrence JG** (2002) Gene transfer in bacteria: speciation without species? *Theor. Popul. Biol.* **61**: 449–460
- Lawrence JG, Hendrickson H** (2003) Lateral gene transfer: when will adolescence end? *Mol. Microbiol.* **50**: 739–749
- Lenski RE, Rose MR, Simpson SC, Tadler SC** (1991) Long-term experimental evolution in *Escherichia coli*. 1. Adaptation and divergence during 2000 generations. *Am. Nat.* **138**: 1315–1341
- Majewski J, Cohan FM** (1998) The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. *Genetics* **148**: 13–18
- Majewski J, Zawadzki P, Pickerill P, Cohan FM, Dowson CG** (2000) Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. *J. Bacteriol.* **182**: 1016–1023
- Maruyama T, Kimura M** (1974) A note on the speed of gene frequency changes in reverse directions in a finite population. *Evolution* **28**: 161–163
- Maruyama T, Kimura M** (1975) Moments for the sum of an arbitrary function of gene frequency along a stochastic path of gene frequency change. *Proc. Natl. Acad. Sci. USA* **72**: 1602–1604
- Maynard Smith J, Feil EJ, Smith NH** (2000) Population structure and evolutionary dynamics of pathogenic bacteria. *Bioessays* **22**: 1115–1122
- Nakamura Y, Itoh T, Matsuda H, Gojobori T** (2004) Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat. Genet.* **36**: 760–766
- Nei M, Roychoudhury AK** (1973) Probability of fixation and mean fixation time of an overdominant mutation. *Genetics* **74**: 371–380
- Nielsen KM** (2003) An assessment of factors affecting the likelihood of horizontal transfer of recombinant plant DNA to bacterial recipients in the soil and rhizosphere. *Collect. Biosafety Rev.* **1**: 96–149
- Nielsen KM, Townsend JP** (2001) Environmental exposure, horizontal transfer and selection of transgenes in bacterial populations. In Vurro M, Gressel J, Butt T, Harman D, Nuss D, Sands D, St. Leger R, eds, Enhancing biocontrol agents and handling risks, NATO Science Series, Vol. 339, IOS Press, Amsterdam, The Netherlands, pp 145–158
- Nielsen KM, Townsend JP** (2004) Monitoring and modeling horizontal gene transfer. *Nat. Biotechnol.* **22**: 1110–1114
- Nielsen KM, Bones AM, van Elsas JD** (1997) Induced natural transformation of *Acinetobacter calcoaceticus* in soil microcosms. *Appl. Environ. Microb.* **63**: 3972–3977
- Nielsen KM, van Elsas JD, Smalla K** (2000) Transformation of *Acinetobacter* sp. BD413(pFG4ΔnptII) with transgenic plant DNA in soil microcosms and effects of kanamycin on selection of transformants. *Appl. Environ. Microb.* **66**: 1237–1242
- Nielsen KM, Berdal KG, Kruse H, Sundsfjord A, Mikalsen A, Yazdankhah S, Nes I** (2005) An assessment of potential long-term health effects caused by antibiotic resistance marker genes in genetically modified organisms based on antibiotic usage and resistance patterns in Norway, VKM-Report, Norwegian Scientific Committee for Food Safety, Oslo, Norway
- Ochman H, Lawrence JG, Groisman EA** (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304
- Rouzine IM, Rodrigo A, Coffin JM** (2001) Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiol. Mol. Biol. R.* **65**: 151–185
- Townsend JP, Nielsen KM, Fisher D, Hartl DL** (2003) Horizontal acquisition of divergent chromosomal DNA in bacteria: effects of mutator phenotypes. *Genetics* **164**: 13–21
- Spratt BG, Hanage WP, Feil EJ** (2001) The relative contribution of recombination and point mutation to the diversification of bacterial clones. *Curr. Opin. Microbiol.* **4**: 602–606
- Pimm SL** (1991) The balance of nature? The University of Chicago Press, Chicago, pp 1–434
- van Herwaarden OA, van der Waal NJ** (2002) Extinction time and age of an allele in a large finite population. *Theor. Popul. Biol.* **61**: 311–318
- Vulic M, Dionisio F, Taddei F, Radman M** (1997) Molecular keys to speciation: DNA polymorphism and the control of gene exchange in enterobacteria. *Proc. Natl. Acad. Sci. USA* **94**: 9763–9767
- Wright S** (1931) Evolution in Mendelian populations. *Genetics* **16**: 97–159