Identification and localization of differences between *Escherichia coli* and *Salmonella typhimurium* genomes by suppressive subtractive hybridization

Received: 20 May 1999 / Accepted: 27 July 1999

**Abstract** The availability of bacterial genome sequences raises an important new problem – how can one move from completely sequenced microorganisms as a reference to the hundreds and thousands of other strains or isolates of the same or related species that will not be sequenced in the near future? An efficient way to approach this task is the comparison of genomes by subtractive hybridization. Recently we developed a sensitive and reproducible subtraction procedure for comparison of bacterial genomes, based on the method of suppression subtractive hybridization (SSH). In this work we demonstrate the applicability of subtractive hybridization to the comparison of the related but markedly divergent bacterial species *Escherichia coli* and *Salmonella typhimurium*. Clone libraries representing sequence differences were obtained and, in the case of completely sequenced *E. coli* genome, the differences were directly placed in the genome map. About 60% of the differential clones identified by SSH were present in one of the genomes under comparison and absent from the other. Additional differences in most cases represent sequences that have diverged considerably in the course of evolution. Such an approach to comparative bacterial genomics can be applied both to studies of interspecies evolution – to elucidate the “strategies” that enable different genomes to fit their ecological niches – and to development of diagnostic probes for the rapid identification of pathogenic bacterial species.

**Key words** *Escherichia coli* · *Salmonella typhimurium* · Subtractive hybridization

**Introduction**

Great progress in bacterial genomics made it possible to determine the structures of whole genomes of many bacteria, including the 4.6-Mb genome of the widely used research organism – and common pathogen – *E. coli* (Blattner et al. 1997), and has opened up the prospect of obtaining the genome sequences of most important bacterial strains in the near future. The availability of bacterial genomic sequences forms the basis for comparative bacterial genomics aimed at revealing the similarities and differences between bacterial organisms. It creates new opportunities for analysis of the organization of a genome’s regulatory machinery, the functions of particular genes, the evolutionary relationships between different bacterial genera, and for the identification of the minimal set of genes necessary for all living entities (Mushegian and Koonin 1996), as well as the specialized functions characteristic of a given strain, including entirely novel genes and gene families. Apart from their importance for these fundamental problems in the life sciences, the genome sequences form a solid base for the identification of genes responsible for the pathogenicity of infectious bacteria, allowing one to search for new antimicrobial medicines and vaccines as well as diagnostic tools.

However, it is unrealistic to expect that many of the great multitude of microbial strains and variants (several hundreds for each species) will be sequenced in the near future. Therefore the current problem is how to use the completely sequenced microorganisms as a reference to reveal structural features of other strains or isolates of the same species or even of different though related
species. This becomes increasingly important as more new infections emerge and old infections return to afflict human society. One possible way to solve this problem involves the parallel comparison of genomes by hybridization to DNA or oligonucleotide microarrays representing the whole genome of a sequenced microorganism. However, this approach, although generally universal and very powerful, will not be applicable to the identification of genes present in a strain under study but absent from the standard. An alternative and efficient approach is subtractive hybridization, which allows one to isolate genomic differences between related but phenotypically distinct bacterial strains (Tas et al. 1994; Strathdee and Johnson 1995; Brown and Curtiss 1996; Mahairas et al. 1996; Tinsley and Nassif 1996). In this case the differences can be isolated providing they are sufficiently divergent to prevent the annealing of different sequences under the conditions of subtraction. One area in which comparative integral genome analysis is highly desirable is in the comparison of the genomes of *E. coli* and its close relative *S. typhimurium*. The enteric bacteria are the principal etiological agents of gastroenteritis and enteric fever in animals and humans. Species of enteric bacteria differ both in virulence characteristics and in metabolic and biochemical properties. A rather detailed comparative analysis of *E. coli* and *S. typhimurium* has already been carried out at various levels. The chromosomes of the two species are of similar size and the known genes show a rather high degree of identity. However, comparison of the linkage maps shows considerable differences, including large inversions and loop structures (Riley and Krawiec 1987). Screening of a genomic library constructed from *S. typhimurium* DNA revealed a number of clones unique for Salmonellae. Most importantly, some of the clones contained sequences common to the majority of *Salmonella* strains but absent from *E. coli* strains. Such differences could be useful for designing diagnostic tests that can differentiate between enteric pathogens. A detailed molecular, functional, and evolutionary analysis of several sequences specific for *Salmonella* has been undertaken, and their nucleotide sequences and positions on the *S. typhimurium* genomic map (Riley and Sanderson 1990) were determined. However, a more comprehensive analysis is necessary to identify all or at least many of the sequences found in *Salmonella* and absent from *E. coli* and vice versa, in order to understand distinctions in the structure-function properties and pathogenic potential of the two genera and to develop reliable and efficient diagnostic tools.

Recently we have developed a simple and reproducible subtraction procedure for comparison of bacterial genomes (Akopyants et al. 1998), based on the suppression subtractive hybridization (SSH) method that was originally employed for generation of subtracted cDNA libraries (Diatchenko et al. 1996; Gurskaya et al. 1996). In this report we describe the application of this technique to the development of clone libraries representing differences between *E. coli* and *S. typhimurium*, and the assignment of some *E. coli*-specific clones to the *E. coli* genomic sequence. We also evaluate the potential of the subtractive technique for the comparison of the genomes of related bacterial genera.

### Materials and methods

#### Bacterial strains

*Escherichia coli* K12 WZ-11 (F-, *met B*, Str*) (Silverstone et al. 1972) and *Salmonella typhymurium* Li2 were cultured aerobically in L-broth at 37°C overnight.

#### Oligonucleotides

Adapters for subtractive hybridization and PCR primers were synthesized using an ASM-102U synthesizer (Biosan). Sequences of oligonucleotides are presented in Table 1.

#### Genomic DNA isolation

DNAs from *E. coli* and *S. typhimurium* were isolated using the cetlytrimethylammonium bromide procedure described by Murray and Thompson (1980).

#### Suppression subtractive hybridization (SSH)

Two micrograms each of tester and driver DNA (*E. coli* and *S. typhimurium*, respectively) was digested to completion with *Hae*III restriction endonuclease. After digestion, DNA was extracted with phenol, precipitated with ethanol, and resuspended in 10 mM TRIS-HCl pH 7.5, at a final concentration of 100 ng/μl. Two aliquots of *E. coli* (tester) DNA (100 ng each) were ligated to the different adapter sets 1 and 2 (2 μM final concentration) using 400 U/μl of T4 DNA ligase (Clontech) and buffer supplied by the manufacturer, in a final volume of 10 μl. After the ligation 1 μl of 0.2 M EDTA was added and the samples were heated for 5 min at 70°C to inactivate the ligase.

Subtractive hybridization was carried out essentially as described by Akopyants et al. (1998). Two rounds of hybridization were done at 63°C, then nested PCR was performed. The first PCR using PCR primer 1 (see above) at 0.4 μM was carried out in a final volume of 25 μl. The reaction mixture was incubated for 2 min at 72°C, and then subjected to 20 cycles of 94°C for 10 s, 66°C for

### Table 1 Sequences of SSH adapters and PCR primers

| Adapter set 1 | 5′-GTAATACGACTCATATAGGGC-TCGAGGGCGCGCAGGTGAGT-3′/3′-GGCGGCTTCCAACGAGG-5′ |
| Adapter set 2 | 5′-GTAATACGACTCATATAGGG-CAGCGTGGTCGCGGCCGAGGT-3′/3′-GCCGGCTTCCAACGAGG-5′ |
| PCR Primer 1 | 5′-GTAATACGACTCATATAGGGGAGC-TCGAGGGCGCGCAGGTGAGT-3′ |
| Nested primer 1R | 5′-GCTTAACTTTCCTTCCACCAT-3′ |
| Nested primer 2R | 5′-GACGTGCGGGCGCCAGGTGAGT-3′ |
| Primer SL11F | 5′-GGCCCGGCGCGCCAGGTGAGT-3′ |
| Primer SL11R | 5′-CATGGTGTGTCTGAGG-3′ |
| Primer SL35F | 5′-TAGCGTGGTCGCGGCCGAGGT-3′ |
| Primer SL35R | 5′-GATACTTCTTCCAGC-3′ |
| Primer SL41F | 5′-TTTAAACTTTCCTTCCAGC-3′ |
| Primer SL41R | 5′-ACGCTACGTTCTTGGG-3′ |
| Primer SL47F | 5′-CAGCGACTCTTCTTGGGT-3′ |
| Primer SL47R | 5′-GATCATAGGTTGGGATCAG-3′ |