Reports

A new technique for obtaining whole pathogen transcriptomes from infected host tissues

Tatyana L. Azhikina¹, Timofey A. Skvortsov¹, Tatyana V. Radaeva², Andrey V. Mardanov³, Nikolay V. Ravin³, Alexander S. Apt², and Eugene D. Sverdlov¹

¹*Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia,* ²*Central Institute for Tuberculosis, Moscow, Russia,* and ³*Centre "Bioengineering," Russian Academy of Sciences, Moscow, Russia*

BioTechniques 48:139-144 (February 2010) doi 10.2144/000113350

Keywords: intracellular pathogen transcriptome; *Mycobacterium tuberculosis*; coincidence cloning

Supplementary material for this article is available at www.BioTechniques.com/article/113350.

We propose a novel experimental approach based on coincidence cloning for analyzing sequences of bacterial intracellular pathogens specifically transcribed in affected tissues. Co-denaturation and co-renaturation of excess bacterial genomic DNA with the cDNA prepared on total RNA of the infected tissue allows one to select the bacterial fraction of the cDNA sample. We used this technique for preparing and characterizing the *Mycobacterium tuberculosis* cDNA pool, representing the transcriptome of infected mouse lungs in the chronic phase of infection. A cDNA pool enriched in fragments of mycobacterial cDNA was analyzed by the high-throughput 454 sequencing procedure. We demonstrated that its composition corresponded to what can be expected in the chronic phase of infection and, after the adaptation of *M. tuberculosis* to the host immune system, was characterized by an active lipid metabolism and switched from aerobic to anaerobic respiration. The technique is universal and requires no prior knowledge of the pathogen genome sequence. Pools of transcribed sequences obtained by this technique retain the main characteristics of the genome-wide gene transcription pattern within infected tissue, and can be used for in vivo analysis of gene expression of a wide spectrum of infection agents, such as viruses, bacteria, and protista.

Introduction

It is now well acknowledged that during pathogen penetration, the host organism triggers the defense system by recognizing and responding to microbial exposure. For its part, the pathogenic microorganism also responds to the signals generated by its interaction with the defense system by changing the gene expression pattern to neutralize (or at least decrease) the destroying potential of the host defense. The understanding of host-pathogen interactions not only requires an investigation of host innate and adaptive immune responses to the infection, but also knowledge of parasite factors that provide adaptation of the pathogen to the host environment. At the moment, this understanding is limited, and new experimental approaches to analyze these complex interactions are needed. The sequencing of hundreds of microbial genomes stimulated development of novel approaches for functional

studies at the genome-wide level. Modern technologies, such as serial analysis of gene expression (1), subtractive hybridization (2), and DNA microarray analysis (3,4) enable the assessment of bulk gene expression profiles in a single experiment, and provide a deeper insight into host-pathogen interactions (for review, see References 5 and 6). However, the efficacy of these technologies critically depends upon the availability of bacterial RNA samples that precisely reflect the real ratios of individual bacterial mRNAs in the infected host tissues. This is very challenging, given the paucity of bacterial mRNA compared with the amounts of mammalian RNA in samples. To resolve this problem, several experimental approaches were proposed, including differential lysis for bacterial RNA extraction (5,7), enrichment of bacterial RNA by cDNA-RNA subtractive hybridization (8), the DECAL method (9), and hybridization-based positive cDNA selection (selective capture of transcribed

sequences, or SCOTS) (10). At present, the most popular ways of preparing cDNA probes to be further used for hybridization with microarrays are (i) selective amplification of the corresponding cDNA with primers against ORFs of a given bacterial genome on a template of previously enriched bacterial RNA isolated from the infected tissue (11,12), and (ii) polyadenylation of the bacterial RNA followed by further amplification using T7 oligo(dT) primer and T7 RNA polymerase (13). However, these approaches are not universal, and since they are based on a known strain's genome sequence, they do not take into account genomic variability of bacteria due to insertion-deletion polymorphisms, which is known to be rather high (14,15). Here we propose an easy-to-implement hybridization method based upon the coincidence cloning (CC) approach (16) that allows isolation of representative bacterial cDNA pools from infected organs. Co-denaturation and co-renatur-

ation of the excess of bacterial genomic DNA with the cDNA transcribed from total RNA of the infected tissue enabled selective isolation of the bacterial cDNA fraction from the sample, and a single round of coincidence cloning resulted in >1000-fold enrichment of bacterial transcripts. As a model organism, we chose *Mycobacterium tuberculosis*, the intracellular pathogen responsible for tuberculosis in millions of people per year. We used this approach to selectively isolate *M. tuberculosis* cDNA from the lung tissue of A/Sn mice after aerosol challenge, the infection model developed in A.S.A.'s laboratory (17). To make an exhaustive description of the enriched bacterial cDNA sample, we performed high-throughput pyrosequencing of this sample, which is less biased and more reliable than other methods including microarrays. 454 pyrosequencing is an extremely powerful sequencing technique that allows for gathering sequence data of hundreds of millions of nucleotides in one experiment (18). A major advantage of the proposed technique over other methods of transcriptome analysis is that it does not require prior knowledge of the pathogen's genome structure, requires only the genomic DNA, and permits analysis of even small amounts of tissue from the infected organism.

Materials and methods

Basic protocols are given in the Supplementary Materials.

RNA isolation and cDNA synthesis

RNA was isolated from the lung tissue of mice at week 9 of infection with *M. tuberculosis* H37Rv (strain was cultivated at the Central Institute for Tuberculosis, Moscow, Russia) using the SV Total RNA Isolation System (Promega, Madison, WI, USA) according to the manufacturer's recommendations. RNA samples were treated with 1 U/ μ L DNaseI (MBI Fermentas, Vilnius, Lithuania) to remove residual DNA. The first cDNA strand was synthesized using BR and SMART primers (synthesized at the Institute of Bioorganic Chemistry, Moscow, Russia) (Supplementary Table S1). The primers (12 pM each) were annealed in 11 μ L of a mixture containing 2 μ g total RNA. The mixture was heated for 2 min at 70°C and then chilled on ice for 10 min. cDNA synthesis was performed according to the manufacturer's protocol with (RT+) or without (RT-) the addition of PowerScript II reverse transcriptase (Clontech, Mountain View, CA, USA). The RT+

and RT- reaction mixtures were then incubated at 37°C and 42°C for 10 and 120 min, respectively. Preparative synthesis of cDNA was performed with 5S primer for 30 cycles (95°C for 20 s, 64°C for 20 s, and 72°C for 2 min). cDNA was cleaned with a QIAquick PCR Purification kit (Qiagen, Valencia, CA, USA).

Coincidence cloning (CC)

Mycobacterial genomic DNA was exhaustively hydrolyzed with *Rsa*I (MBI Fermentas), and the suppression adapter I (Supplementary Table S1) was ligated to the resulting fragments (genomic samples). *Rsa*I-fragmented total cDNA was similarly ligated to the adapter II (cDNA sample). The mixture containing the genomic sample and cDNA sample (100 ng each) in 2 μ L hybridization buffer (HB; 50 mM HEPES pH 8.3, 0.5 M NaCl, 0.02 mM EDTA pH 8.0) was incubated for 5 min at 99°C (denaturation) and then for 18 h at 68°C (renaturation). Thereafter, 100 μ L pre-warmed HB (68°C) was added, and 1 μ L resulting solution was used as a template in the following PCR protocol.

The reaction was performed in a 25- μ L volume containing 10 pM external primer T7. After pre-incubation at 72°C for 5 s (filling-in of the sticky ends), the reaction mixture was PCR-amplified for 20 cycles (94°C for 30 s, 66°C for 30 s, and 72°C for 90 s). The second PCR was performed for 15 cycles (94°C for 30 s, 68°C for 30 s, and 72°C for 90 s) using 1 μ L 1:10-diluted initial PCR product as a template in a 25- μ L volume containing 10 pM of internal primers Not1Srf and Not1Rsa. The CC PCR product (5 μ g) was purified with a QIAquick PCR Purification kit (Qiagen) and used for 454 sequencing. An aliquot of the PCR product was ligated into a pGEM-T vector (Promega) and cloned into *Escherichia coli* (strain DH-5 α , provided by Paul Khil, NIDDK, National Institutes of Health, Bethesda, MD, USA). Randomly chosen clones were sequenced.

Results and discussion

Here we propose a technique for analyzing bacterial mRNA in infected host tissues using corresponding bacterial cDNA pools.

Figure 1. Coincidence cloning procedure. (A) Flowchart of procedure. Total cDNA (murine cDNA and mycobacterial cDNA are designated by dotted and solid lines, respectively) and genomic mycobacterial DNA (solid circle) are hydrolyzed with *Rsa*I following ligation of adapters (white and black bars). The obtained cDNA and genomic samples are mixed, jointly denatured and slowly renatured. The target heteroduplexes are amplified in two PCR rounds, with primers corresponding to the outer (first round) and inner (second round) parts of suppression adapters. (B) Electrophoretogram of genomic DNA (gDNA) and total cDNA fragmented by *Rsa*I. (C) Second-round PCR amplification of mixture after coincidence cloning. The PCR product is amplified only by two primers, Not1Srf and Not1Rsa. Amplification by a single primer does not result in a visualized product.

Figure 2. Electrophoretogram of PCR products with primers against *Rv1664*. (A) Accumulation of PCR product using total cDNA as a template. (B) Accumulation of PCR product using the amplicon after coincidence cloning as a template. Numbers designate PCR cycles. +, positive control (*M. tuberculosis* genomic DNA); -, negative control (no template).

The technique is based on hybridization of total cDNA prepared from an infected tissue with a great molar excess of bacterial genomic DNA followed by selective PCR amplification. An excess of bacterial DNA is supposed to ensure unbiased quantitative analysis of the bacterial transcriptome within infected host cells. Pseudomonomolecular kinetics of bacterial cDNA reassociation with excess bacterial DNA provides practically invariable ratios between hybridized individual components of the cDNA pool if a sufficiently long hybridization time is used.

Total cDNA preparation

Total RNA was isolated from the mouse lung tissue using a standard method, and cDNA was synthesized using SMART technology (Clontech), with modifications. As a primer (BR), we used a statistical set of nonanucleotides (dN)₉ (synthesized at the Institute of Bioorganic Chemistry) attached to a 5'-constant 25-nucleotide fragment with a sequence therein identical to the SMART oligonucleotide. This constant part of the primer was supplemented with an *RsaI* restriction endonuclease site. The resulting reaction mixture contained single-stranded cDNA fragments of various lengths flanked by stretches with known nucleotide sequences. After termination of reverse transcription, cDNA was amplified using 5S primer complementary to the constant sequences at 5' and 3' ends of the first cDNA strands.

Coincidence cloning (CC)

To isolate bacterial pathogen cDNA from mixtures with mouse cDNA, we used the CC procedure, which allows one to selectively isolate only coincident fragments from two different sets of DNA fragments

(Figure 1A). Two sets of DNA fragments are preliminarily digested with a frequently cutting restriction enzyme (e.g., *RsaI*). The *RsaI* fragments in both sets are ligated to suppression adapters with identical external 5' ends [selective suppression of PCR (19)]. After this, the two sets are mixed together, denatured, and slowly renatured. The sequences common to both sets are able to form hybrid duplexes bearing different adapters at the 5' ends. In contrast, the DNA fragments unique to one of the sets can form only homoduplexes with identical adapters at their ends. The homoduplexes thereby cannot be PCR-amplified due to PCR-selective suppression, whereas the hybrid duplexes can.

Genomic DNA of *M. tuberculosis* H37Rv was digested with *RsaI*, and the resulting fragments were ligated with the adapter I. The mixture of mouse and mycobacterial cDNA was treated similarly with *RsaI* in order to remove sequences identical to the 5S PCR primer. The resulting fragments were ligated with the suppression adapter II. Restricted genomic and cDNA samples are illustrated in Figure 1B. Ligated genomic DNA and cDNA fragments were mixed in weight-equal proportions. Since the

Want **sensitive, specific,**
high-throughput and **fast**
qRT-PCR?
Be with MCLAB.

We are equipped with all the instruments necessary for performing your PCR project. Our services encompass the entire field of real-time PCR including gene expression analysis, customized design of qPCR assay, residual DNA analysis with qPCR, and plasmid quantification service using qPCR. Benefit from our experience with large pharmacies and biotech companies! Get better and faster results!!

mycobacterial cDNA fraction is estimated to represent as little as 0.04–0.2% of the total cDNA by weight (20), the molar ratio between mycobacterial genomic DNA and mycobacterial cDNA in the mixture is estimated to be $\geq 100:1$. Such an excess of genomic DNA prevents artificial changes in quantitative proportions between mycobacterial transcripts. The mixture obtained was denatured for 3 min at 99°C and slowly renatured for 16 h at 68°C. The resulting duplexes were then selectively amplified. The first PCR round was performed with T7 primer corresponding to the identical external parts of adapters I and II, while the second nested round was carried out with two primers identical to the 3'-terminal regions of adapters I (NotISrf) and II (NotIRsa), respectively. The use of nested PCR provides selective amplification of hybrid duplexes. PCR with each individual nested primer demonstrated the absence of any exponentially amplified product, thus confirming that the nested PCR resulted in selective amplification of heteroduplexes, but not homoduplexes (selective amplification of heteroduplexes is given in Figure 1C).

The CC amplicon represents a pool of *M. tuberculosis* cDNA *RsaI* fragments selectively amplified from a mixture of mouse and mycobacterial cDNAs. To characterize the amplicon's properties, we cloned it into the pGEM-T cloning vector, transformed *E. coli* cells, and sequenced 75 recombinant clones containing inserts. Nucleotide sequences were compared with *M. tuberculosis* and mouse genome databases. Of 75 sequences, 22 (30%) made hits with the mouse genomic sequences, while the rest (70%) corresponded to the *M. tuberculosis* genome. As proved by the presence of two different adaptors, the inserts of *M. tuberculosis* origin were heteroduplexes (i.e., formed by identical fragments from genomic *M. tuberculosis* DNA and cDNA).

The CC-derived amplicon was quantitatively analyzed by PCR for some selected mycobacterial transcripts using corresponding gene-specific primers (Supplementary Table S2). We had to use semiquantitative PCR at 2-cycle intervals instead of real-time PCR due to very low mycobacterial content in the total cDNA, which allows non-specific priming. Figure 2 illustrates such semiquantitative analysis. We checked the presence of 15 randomly chosen transcribed genes in the CC-derived amplicon and total cDNA (Supplementary Table S5). Although all 15 genes were detected in the amplicon, only 10 of those were found in the total cDNA, likely due to a low content of *M. tuberculosis* RNA in the

total lung tissue RNA sample. In half of the cases, a cDNA product could be detected close to the 40th cycle. As the 16S RNA transcript was visualized in the total cDNA sample at the 32nd cycle, we divided the tested genes into three groups, according to their transcription level: high (35–36 cycles in the total cDNA sample; up to 22 cycles in the CC amplicon sample), middle (39–40 cycles in the total cDNA sample; 24–28 cycles in the CC amplicon sample) and low (undetected in the total cDNA sample; more than 28 cycles in the CC amplicon sample). We compared our data to the model of chronic infection presented by Talaat et al. (21) and found coincidence for several genes (e.g., the gene *Rv0592* is transcribed at the middle level, and genes *Rv0990*, *Rv1182*, *Rv1664*, *Rv0988*, and *Rv2030* belong to the low-transcription-level group). Our method allowed us to detect the transcription of genes that were not revealed by microarrays used in Talaat's experiments (e.g., *Rv0996*). Moreover, we found active transcription of *Rv1387*, *Rv1525*, and *Rv2941* that was easily detected both in total cDNA and the CC amplicon, even though these genes were assigned as untranscribed by Talaat et al. These differences in expression might be explained by the time of infection and/or host genetics (line of mice), as well as by the advantages or limitations of the different methods.

Overall, we were able to demonstrate that generally, the CC procedure does not qualitatively alter the transcription profiles, and that the levels of in vivo expression are reflected by the content of transcripts in CC products with appreciable accuracy. Remarkably, the CC method allows for enrichment of poorly represented transcripts that remain undetectable in vivo.

Whole-transcriptome description

We performed a massive parallel pyrosequencing of the CC amplicon obtained (GS FLX; Roche, Mannheim, Germany). Nucleotide sequences were read in 98,692 independent reactions; the average sequence length was ~200 bp. 454 sequence quality is known to be extremely high [$>99\%$ (22)], so we assigned the obtained sequences to the *M. tuberculosis* genome in cases of at least 90% nucleotide identity. 75,436 sequences (76.4%) were found to be homologous to the genome sequence of *M. tuberculosis*. Considering that the total RNA contains ~0.04% mycobacterial RNA (20), the content of bacterial cDNA was enriched at least 1000-fold. Using BLASTn, we next assessed the number of the cDNA sequences complementary to particular genes.

The sequencing characterized the transcription of 536 genes, which amounted to ~14% of the total number of genes (Supplementary Tables S3, S6). Details of the transcriptome properties that are characteristic for the chronic phase of infection are described in the Supplementary Materials. We compared the obtained transcription pattern with previously reported data on the *M. tuberculosis* transcriptome present in the lung of infected mice during the chronic phase (21) and found some common characteristics, such as an activation of lipid metabolism and a switch from aerobic to anaerobic respiration (Supplementary Table S4). The differences observed could be due to a different infection timetable, host properties (different mouse strains were used), or peculiarities of any of the methods used. For example, our data confirmed the absence (or extremely low level) of expression for *nuoA*–*nuoN* genes (which code for NADH dehydrogenases). Conversely, the expression of *narH*, *narJ*, *narI*, and *narK* genes (which code for nitrate reductases) was shown only in our model.

All existing methods of transcriptome examination can be divided into two categories: hybridization or sequence-based approaches. Hybridization-based microarray techniques are widely used and powerful, but have several limitations: high background levels owing to cross-hybridization, problems of rare transcript detection, and expression quantification. Furthermore, microarray analysis is always based on preliminary knowledge of the genome sequence. The limitations of microarrays in comparison to high-throughput sequencing have been extensively discussed in many research reports and reviews (23).

Sequence-based approaches directly determine the cDNA sequence, but until now, the capacity, cost, and time expenditures of Sanger sequencing made whole-transcriptome description unfeasible. The development of novel high-throughput DNA sequencing methods (reviewed in Reference 24) provides the opportunity to study transcriptomes of several organisms (25). However, this includes only one study of prokaryotes (26). To our knowledge, no one has used deep sequencing to study pathogen transcription in infected tissues. In cases of intracellular infection, enrichment of preliminary bacterial RNA is required. For this purpose, researchers usually divide the host and pathogen cells by differential lysis procedures (5). The differential lysis approach involves the physical separation of bacteria from host cells before

RNA preparation. To leave the bacteria intact while performing efficient lysis of the host cells or tissue, detergents and lysis conditions need to be selected individually for each pathogen. This may be the reason why so few intracellular pathogens have been enriched by differential lysis. This additional step could be critical in work with highly unstable bacterial RNA and also with small amounts of material (e.g., that provided from surgical samples).

We intended to create a new method that would be universal and independent of the knowledge of genome structure, would not require any preliminary enrichment steps (such as differential lysis), and would use all of the advantages of deep-sequencing technologies. For this purpose, we applied the coincidence cloning procedure, combined with the PCR suppression effect, to a model of *M. tuberculosis* infection in mouse lungs. The procedure we propose consists of co-denaturation and co-renaturation of the excess bacterial genomic DNA with the cDNA transcribed from total RNA of the infected tissue, followed by selective amplification of target heteroduplexes (Figure 1A). This process does not depend on knowledge of genome structure;

it requires only a sample of the genomic bacterial DNA at hand.

The most critical step is the formation of proper heteroduplexes. To avoid formation of imperfect duplexes, genomic DNA of the precise strain of bacteria under investigation must be used, otherwise the genome variability between strains could lead to improper pairing and false sequencing results. Despite the fact that the size of homologous fragments could be unequal between cDNA and genomic DNA (Figure 1B), our experience gives us reason to believe that different sizes of hybridizing fragments—and later PCR amplification (16,27)—does not affect the effectiveness of hybridization, due to the filling-in of the sticky ends of the duplexes.

To evaluate the enrichment degree obtained through CC, we cloned a CC aliquot and sequenced several clones. Both clone sequencing and massive parallel 454 pyrosequencing returned similar data: the CC amplicon contained ~70% mycobacterial and ~30% mouse cDNA, which represents a >1000-fold enrichment. Taking into account the absence of any preliminary enrichment steps, this result appears satisfactory. We have no data to

compare the enrichment achieved by our method to that of other methods, since no one sample prepared for microarray studies was sequenced to directly show the percentage of host material.

By analyzing the sequences flanking cDNA fragments, we found out that all *M. tuberculosis* cDNA fragments were flanked by two different adapter sequences. This directly indicates that they were formed by identical fragments from genomic *M. tuberculosis* DNA and cDNA. All host cDNA fragments lacked the correct sequence of adapter I, which means the absence of a mycobacterial genomic DNA strand in the duplex. We suppose that these mouse cDNA homoduplexes resulted from PCR amplification of an extremely complex mixture.

454 pyrosequencing gave us information on 20,000,000 nucleotides, 75% of which corresponded to mycobacterial transcripts. We were able to assign these sequences to 14% of *M. tuberculosis* genes. An apparently low number of transcribed genes can be explained by both insufficient sequencing depth (resulting in an underestimation of under-represented transcripts) and also by a general decrease in the

BioTechniques®

The International Journal of Life Science Methods

Update your subscription

Update your profile today to ensure that your copy of *BioTechniques* experiences no delivery disruptions.

Go to:

www.BioTechniques.com/subscribe

number of transcribed sequences followed by a drop in total levels of mycobacterial protein and RNA in the chronic phase of infection (21,28). Talaat et al. demonstrated that in the chronic phase of tuberculosis, 50% of genes are transcribed but only 8% are actively transcribed (cDNA/gDNA signal > 1.5). The remainder (42%) are of a low transcription level (~1–2 copies per genome). Genes with low expression could be detected only if sequencing data cover the genome size many times over. For example, Wang et al. (25) published their estimate of the sequencing volume necessary for a complete description of the *Saccharomyces cerevisiae* transcriptome (it has a genome length comparable to *M. tuberculosis*) and demonstrated that to detect 50% of ORFs, one would need to perform more than a million independent runs. Increasing the sequencing volume will certainly give information about rare transcripts.

In conclusion, we propose a new method for the evaluation of sequences of bacterial pathogens specifically transcribed in infected tissues, and demonstrate its potency by analyzing the *M. tuberculosis* transcriptome in chronically infected A/Sn mice (i.e., after adaptation of bacilli to the host immune system). The method is universal and does not require prior knowledge of the pathogen's genome sequence. Pools of transcribed sequences obtained by this method retain qualitative and quantitative characteristics of the initial sample and can be used in a wide spectrum of pathogen transcriptome studies. We believe that this method extends our approaches to a comprehensive examination of host-pathogen interactions.

Acknowledgments

The authors thank Boris Glotov and Darya Vanichkina for critically reading the manuscript. This work was supported by the Russian Foundation for Basic Research (RFBR; grant nos. 06-04-48976, 07-04-00988, and 08-04-01053), the President of the Russian Federation (grant no. 2006.2003.4), the Russian Academy of Sciences (a grant of the program "Physico-chemical biology. Structural, functional and evolutionary analysis of genomic cis-regulatory systems,") and the National Institutes of Health (NIH; grant no. AI078864, to A.A.). 454 sequencing was supported by the Federal Agency for Science and Innovation (Rosnauka) (grant no. GK 02.552.11.7045). The cloned DNA inserts were sequenced in the Interinstitute "Genom" Center (www.genome-centre.narod.ru) organized under the support of RFBR (grant no. 00-04-55000).

Competing interests

The authors declare no competing interests.

References

1. Velculescu, V.E., L. Zhang, B. Vogelstein, and K.W. Kinzler. 1995. Serial analysis of gene expression. *Science* 270:484-487.
2. Diatchenko, L., Y.F. Lau, A.P. Campbell, A. Chenchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov, et al. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA* 93:6025-6030.
3. Kato-Maeda, M., Q. Gao, and P.M. Small. 2001. Microarray analysis of pathogens and their interaction with hosts. *Cell. Microbiol.* 3:713-719.
4. Schoolnik, G.K. 2002. Functional and comparative genomics of pathogenic bacteria. *Curr. Opin. Microbiol.* 5:20-26.
5. Mangan, A., I. Monahan, and P. Butcher. 2002. Gene expression during host-pathogen interactions: approaches to bacterial mRNA extraction and labelling for microarray analysis, p. 137-151. In B.W. Wren and N. Dorrel (Eds.), *Methods in Microbiology*. Vol. 33. Academic Press, London.
6. Waddell, S.J., P.D. Butcher, and N.G. Stoker. 2007. RNA profiling in host-pathogen interactions. *Curr. Opin. Microbiol.* 10:297-302.
7. Schnappinger, D., S. Ehrt, M.I. Voskuil, Y. Liu, J.A. Mangan, I.M. Monahan, G. Dolganov, B. Efron, et al. 2003. Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. *J. Exp. Med.* 198:693-704.
8. Li, M.S., I.M. Monahan, S.J. Waddell, J.A. Mangan, S.L. Martin, M.J. Everett, and P.D. Butcher. 2001. cDNA-RNA subtractive hybridization reveals increased expression of mycocerosic acid synthase in intracellular *Mycobacterium bovis* BCG. *Microbiology* 147:2293-2305.
9. Alland, D., I. Kramnik, T.R. Weisbrod, L. Otsubo, R. Cerny, L.P. Miller, W.R. Jacobs, Jr., and B.R. Bloom. 1998. Identification of differentially expressed mRNA in prokaryotic organisms by customized amplification libraries (DECAL): the effect of isoniazid on gene expression in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* 95:13227-13232.
10. Graham, J.E., and J.E. Clark-Curtiss. 1999. Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proc. Natl. Acad. Sci. USA* 96:11554-11559.
11. Talaat, A.M., P. Hunter, and S.A. Johnston. 2000. Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. *Nat. Biotechnol.* 18:679-682.
12. Talaat, A.M., S.T. Howard, W. Hale 4th, R. Lyons, H. Garner, and S.A. Johnston. 2002. Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis*. *Nucleic Acids Res.* 30:e104.
13. Waddell, S.J., K. Laing, C. Senner, and P.D. Butcher. 2008. Microarray analysis of defined *Mycobacterium tuberculosis* popula-

tions using RNA amplification strategies. *BMC Genomics* 9:94.

14. Dobrindt, U. and J. Hacker. 2001. Whole genome plasticity in pathogenic bacteria. *Curr. Opin. Microbiol.* 4:550-557.
15. Mira, A., L. Klasson, and S.G. Andersson. 2002. Microbial genome evolution: sources of variability. *Curr. Opin. Microbiol.* 5:506-512.
16. Azhikina, T., I. Gainetdinov, Y. Skvortsova, A. Batrak, N. Dmitrieva, and E. Sverdlov. 2004. Non-methylated genomic sites coincidence cloning (NGSCC): an approach to large scale analysis of hypomethylated CpG patterns at predetermined genomic loci. *Mol. Genet. Genomics* 271:22-32.
17. Nikonenko, B.V., M.M. Averbakh, Jr., C. Lavebratt, E. Schurr, and A.S. Apt. 2000. Comparative analysis of mycobacterial infections in susceptible I/St and resistant A/Sn inbred mice. *Tuber. Lung Dis.* 80:15-25.
18. Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bembien, J. Berka, M.S. Braverman, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376-380.
19. Luk'ianov, S.A., N.G. Gurskaia, K.A. Luk'ianov, V.S. Tarabykin, and E.D. Sverdlov. 1994. Highly-effective subtractive hybridization of cDNA. *Bioorg. Khim.* 20:701-704.
20. Banaiee, N., W.R. Jacobs, Jr., and J.D. Ernst. 2006. Regulation of *Mycobacterium tuberculosis* whiB3 in the mouse lung and macrophages. *Infect. Immun.* 74:6449-6457.
21. Talaat, A.M., S.K. Ward, C.W. Wu, E. Rondon, C. Tavano, J.P. Bannantine, R. Lyons, and S.A. Johnston. 2007. Mycobacterial bacilli are metabolically active during chronic tuberculosis in murine lungs: insights from genome-wide transcriptional profiling. *J. Bacteriol.* 189:4265-4274.
22. Meyer, M., U. Stenzel, and M. Hofreiter. 2008. Parallel tagged sequencing on the 454 platform. *Nat. Protoc.* 3:267-278.
23. Shendure, J. 2008. The beginning of the end for microarrays? *Nat. Methods* 5:585-587.
24. Shendure, J. and H. Ji. 2008. Next-generation DNA sequencing. *Nat. Biotechnol.* 26:1135-1145.
25. Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10:57-63.
26. Yoder-Himes, D.R., P.S. Chain, Y. Zhu, O. Wurtzel, E.M. Rubin, J.M. Tiedje, and R. Sorek. 2009. Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc. Natl. Acad. Sci. USA* 106:3976-3981.
27. Buzdin, A., E. Kovalskaya-Alexandrova, E. Gogvadze, and E. Sverdlov. 2006. GREM, a technique for genome-wide isolation and quantitative analysis of promoter active repeats. *Nucleic Acids Res.* 34:e67.
28. Sohaskey, C.D. 2008. Nitrate enhances the survival of *Mycobacterium tuberculosis* during inhibition of respiration. *J. Bacteriol.* 190:2981-2986.

Received 30 June 2009; accepted 5 January 2010.

Address correspondence to Tatyana L. Azhikina, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Miklukho-Maklaya 16/10, Moscow 117997, Russia. e-mail: tatazhik@ibch.ru