Curr Microbiol (2009) 58:258–263
DOI 10.1007/s00284-008-9318-2

Identification of Small Noncoding RNAs in *Helicobacter pylori* by a Bioinformatics-Based Approach

Bin Xiao · Wei Li · Gang Guo · BoSheng Li · Zhen Liu · KeRan Jia · Ying Guo · XuHu Mao · QuanMing Zou

Received: 30 July 2008 / Accepted: 15 October 2008 / Published online: 3 January 2009

Abstract Small noncoding RNAs (sRNAs) are a group of regulatory RNA molecules normally without a protein-coding function. In recent years, the importance of sRNAs as mediators of gene expression in bacteria has begun to be recognized. More than 70 sRNAs have been known in *Escherichia coli*. However, little is known about sRNAs in *Helicobacter pylori*, a human pathogen associated with gastric diseases. Here, we systematically identified sRNAs in the *H. pylori* genome by a computational approach based on gene location, sequence conservation, promoter and terminator search, and secondary structure. Among a total of six candidate sRNAs initially predicted, two novel sRNAs (IG-443 and IG-524) were confirmed by Northern blot and reverse transcription–polymerase chain reaction (RT-PCR). Virtually, they were a class of natural antisense transcripts, which were complementary to partial sequences of the following genes: flagellar motor switch gene (*fliM*) and fumarase (*fumC*). Taken together, the results indicate that there exist novel sRNAs in *H. pylori* and these RNAs might play a potential role in regulating gene expression.

B. Xiao · W. Li · G. Guo · B. Li · Z. Liu · K. Jia · Y. Guo · X. Mao (✉) · Q. Zou (✉)
Department of Clinical Microbiology and Immunology, Faculty of Medical Laboratory Science, Third Military Medical University, Chongqing, People's Republic of China
e-mail: mxh95xy@mail.tmmu.com.cn

Q. Zou
e-mail: qmzou@mail.tmmu.com.cn

Introduction

Recently, small noncoding RNAs have attracted great interest as key regulators in both eukaryotic and prokaryotic life. In bacteria, these regulatory RNAs are also termed "small RNAs" (sRNAs). These are a group of functional RNA molecules normally without a protein-coding function, and they range from 50 to 500 nucleotides in size. The majority of sRNAs can control gene expression at the posttranscriptional level via base-pairing with complementary sequences in target transcripts [4]. It is well established that sRNAs control plasmid replication, transposition of transposable elements, bacterial virulence, quorum sensing, and bacterial expression regulation [6, 7, 14, 15]. The accumulating data suggest that bacteria use sRNAs to fine-tune their physiology and adapt to rapidly changing environments. As a result of recent systematic searches, over 70 sRNAs are now known in *Escherichia coli*, and more and more sRNAs are found in other pathogens (e.g., *Salmonella*, *Yersinia*, *Vibrio harveyi*, etc.) [5]. However, the biological functions and mechanisms of action of bacterial sRNAs remain to be elucidated.

Helicobacter pylori is a Gram-negative, spiral, microaerophilic human pathogen associated with gastric diseases like chronic active gastritis, peptic ulceration, gastric carcinoma, and mucosa-associated lymphoid-tissue lymphoma. A remarkable feature of *H. pylori* is its ability to survive in the extremely acidic environment of the stomach. Despite the limited number of transcriptional regulators, evidence is accumulating for the existence of new and complex circuits regulating gene transcription and virulence. The recent increased recognition of important regulatory roles for sRNAs has expanded our interest in how to find such regulatory RNAs in *H. pylori*. However,

sRNAs were hard to detect using traditional screening strategy due to their lack of defined sequence features, small size, and poor expression level. In a previous study, we have successfully identified several regulatory sRNAs in *H. pylori* by RNomics strategy (data not shown). To circumvent some of the limitations of the previous screens and identify additional sRNAs in *H. pylori*, we systematically screened small RNA genes in *H. pylori* genome by a computational approach, aiming to enrich the knowledge of regulatory network of *H. pylori*.

Materials and Methods

Bacterial Strains and Culture Conditions

Helicobacter pylori strain 26695 was obtained from ATCC. The bacteria were grown on CDC anaerobe blood agar plates (BD). Incubation was performed at 37°C for 2 days in an anaerobic jar containing a gas mixture of 5% O_2, 10% CO_2, and 85% N_2 (DU Scientific).

Identification of Putative Small RNA Genes

The sequences of *H. pylori* 26695 were extracted from the EMBL database. Intergenic regions (IGRs) are defined here as sequences between two neighboring ORFs on either strand. A file of all *H. pylori* IGRs sequences was created by a Perl package based on the gene annotations. This file was used as an input file to the profile search program. We then used a profile-based search program, pftools2.2, to identify putative promoters in the IGRs. The program pftools2.2 was obtained from the Swiss Bioinformatics Institute's ftp server (http://www.isrec.isb-sib.ch/ftp-server/pftools/). Then the prediction of terminators was performed by RNAMotif algorithm as described elsewhere [3]. An intrinsic terminator is composed of a G-C-rich RNA hairpin loop and a U-rich tail. Predicted promoters and terminators were combined to generate putative small RNA genes if the following criteria were met: (1) Promoter and terminator pairs were on the same strand and (2) the distance between the two was less than 350 nucleotides but greater than 45 nucleotides.

After the prediction of promoters and terminators, all of the putative small RNA genes were used in subsequent BLAST and secondary structure analyses. We compared the putative small RNA genes against NCBI Microbial genomes databases using the BLAST program to select conserved putative sRNAs. The secondary structure of putative sRNAs was predicted by the Vienna package (http://rna.tbi.univie.ac.at/).

Detection of the Expression of sRNAs by RT-PCR

Total RNA from overnight cultures of *H. pylori* was prepared by the Trizol method. To remove potential DNA contamination, total RNA was treated with RNase-free DNase I for 30 min at 37°C and recovered by phenol/chloroform extraction and ethanol precipitation. Total RNA (1 μg) was incubated with 1 μl specific RT primer and 1 μl dNTP mix (10 mM each) at 65°C for 5 min to denature the RNA secondary structure. The reactions were chilled on ice for at least 1 min and the remaining reagents [5× buffer, dithiothreitol (DTT), RNase inhibitor, SuperScript III] were added as specified in the SuperScript III protocol and the reaction proceeded for 60 min at 50°C. Finally, the reverse transcriptase was inactivated by a 15-min incubation at 70°C. The cDNA amplification was carried out for 35 cycles at a final annealing temperature of 50°C using the following primers: IG-443 (5′-GAAAAGACATTAAAACTTG-3′; 5′-GATCAAATGACTTATGGCG-3′) and IG-524 (5′-ATGTGCATTGCGGTAGGGA-3′; 5′-ATCCACCCTAACGATGACG-3′). The polymerase chain reaction (PCR) product was separated on 3% agarose gel electrophoresis. Then the gel slices containing DNA with a size about 60 bp were excised and eluted according to the E.Z.N.A.TM Gel Extraction Kit (omega). The DNA fragment was directly subcloned into pGEM-T vector (Promega) and sequenced (Shanghai Sangon Biotechnology).

Northern Blot

Total RNA (30 μg) was separated on a denaturing 15% polyacrylamide gel, and 5S rRNA was visualized using ethidium bromide staining to ensure the quality of the RNA. Then total RNA was transferred to Hybond N+ membranes (Amersham) using a Trans-Blot SD semidry transfer (Bio-Rad). The RNAs were immobilization to the membrane by ultraviolet (UV) crosslinking (120 mJ/cm^2; UVP) and baking at 80°C for 1 h. The sequences of the DNA oligonucleotides used to generate complementary probes were as follows: IG-443 (5′-CCTAGCCCTACAAGTTTTAATG-3′); IG-524 (5′-CGACACTTTCCCTACCGCAATGC-3′). The DNA oligonucleotide probes were labeled with [γ-^{32}P] ATP (Yahui) using T4 polynucleotide kinase (NEB). The membrane was prehybridized in ExpressHyb (CloneTech) solution at 42°C for at least 1 h and hybridized in fresh ExpressHyb solution at 42°C for 12 h. The membranes were washed twice with 2× SSC (1× SSC buffer is 0.15 M NaCl plus 0.015 M sodium citrate)/0.1% sodium dodecyl sulfate (SDS) and 0.5× SSC/0.1% SDS at room temperature. Finally, the membranes were exposed to X-ray film at −70°C for 48 h.

Results

Identification of Putative Small RNA Genes

As a starting point for detecting novel sRNAs in *H. pylori*, we considered a number of common properties of the previously found sRNAs in *Escherichia coli* that might serve as a guide to identify novel sRNAs. We are defining sRNAs as relatively short RNAs without protein-encoding function that are encoded by freestanding genes in the empty IGSs. Therefore, to identify novel sRNAs, we first created a file containing 1262 IGR sequences (>120 nucleotides) from *H. pylori* by a Perl program. Second, for the bacterial small RNA genes, the promoters and terminators are not qualitatively different from those for other traditional genes. The signals of promoters and terminators might give us some clue to discover sRNAs. Additionally, a high conservation was found between the sRNAs in closely related organisms. In this study, after the prediction of promoters and terminators respectively by pftools2.2 and the RNAMotif algorithm, all of the putative sRNA genes were used in subsequent BLAST to select highly conserved sRNA genes. The whole process of identifying putative sRNA genes is schemed as shown in Fig. 1. As a result, the final six novel sRNA genes shown in Table 1 were found in our computational search.

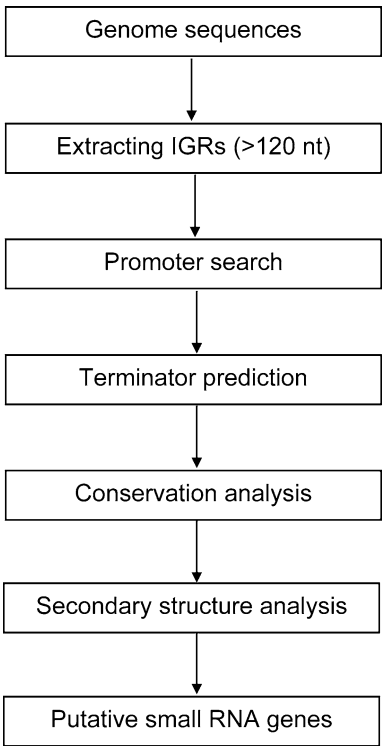

Fig. 1 Scheme of a genome-wide computational search for putative small RNA genes

Annotation of the Candidate sRNA Genes

To investigate the regulatory role of sRNAs in *H. pylori*, we first mapped the genomic location of candidate sRNAs. To our surprise, we found that most sRNAs had matches within a gene or open reading frame (ORF) but in antisense orientation, supporting previous experimental evidence for significant global antisense transcription. Virtually, these candidate sRNAs are a class of natural antisense transcripts (NATs) that exhibit partial or complete complementarity to other RNAs. They might be transcribed in cis from opposing DNA strands at the same genomic locus (*cis*-NATS); for example, IG-443 is encoded in the IGR between *fur* and *HP1033*, in the opposite orientation to the adjoining gene, flagellar motor switch gene (*fliM*) (Fig. 2a); IG-524 is encoded in the IGR between *HP1322* and *HP1326*, in the opposite orientation to the adjoining ORF fumarase (*fumC*) (Fig. 2b). The results suggest that IG-443 and IG-524 might respectively regulate the expression of the above genes in a sequence complementary manner. The results of secondary structure showed that putative sRNAs had typical stem-loop structure (Fig. 2c, d).

Detection of the Expression of sRNAs by RT-PCR and Northern Blot

To verify the expression of candidate sRNAs, we performed RT-PCR and Northern blot analysis. We performed PCR reaction with the template of DNase I-treated total RNA to confirm the quality of RNA without DNA contamination. The PCR product was analyzed on 3% agarose gel electrophoresis and an ~60-bp band was detected in IG-443 and IG-524, whereas there was no band in other candidate sRNAs and negative control (Fig. 3a). The results showed that the expression of IG-443 and IG-524 could be detected by a two-step RT-PCR. At the same time, the results of Northern blot were consistent with RT-PCR, and only the expression of IG-443 and IG-524 were confirmed (Fig. 3b).

Discussion

Helicobacter pylori is a spiral and microaerophilic human pathogen associated with gastric diseases. Infection with this pathogen is widespread and affects ~50% of the world's population. The 1.6-Mb genomes of five *H. pylori* strains (26695, J99, HPAG, Shi470, and G27) have been sequenced and annotated. A remarkable feature of *H. pylori* is its ability to survive in the extremely acidic environment of the stomach. It requires the regulation of bacterial gene expression to cope with the environmental fluctuations. However, *H. pylori* has the relative low abundance of

Table 1 Small RNA genes candidate in *H. pylori* genome

Candidate	Adjacent genes	Strand[a]	5′ End	Length (nucleotides)	Northern	RT-PCR
IG-39	*cpdB/metB*	← → ←	113232	53	−	−
IG-443	*fliY/fliA*	← → ←	1093352	60	+	+
IG-524	*HP132/HP1326*	← → ←	1385163	62	+	+
IG-81	*lpxD/ndk*	→ ← →	203721	92	−	−
IG-173	*dnaN/HP0502*	→ ← →	529205	62	−	−
IG-394	*HP1066/prmA*	→ ← →	1126629	131	−	−

[a] The middle arrow represents the small RNA genes and the flanking arrows indicate the orientation of the adjacent genes, respectively. Genes present on the strand given in the *H. pylori* genome database are indicated by →, and genes present on the complementary strand are indicated by ←

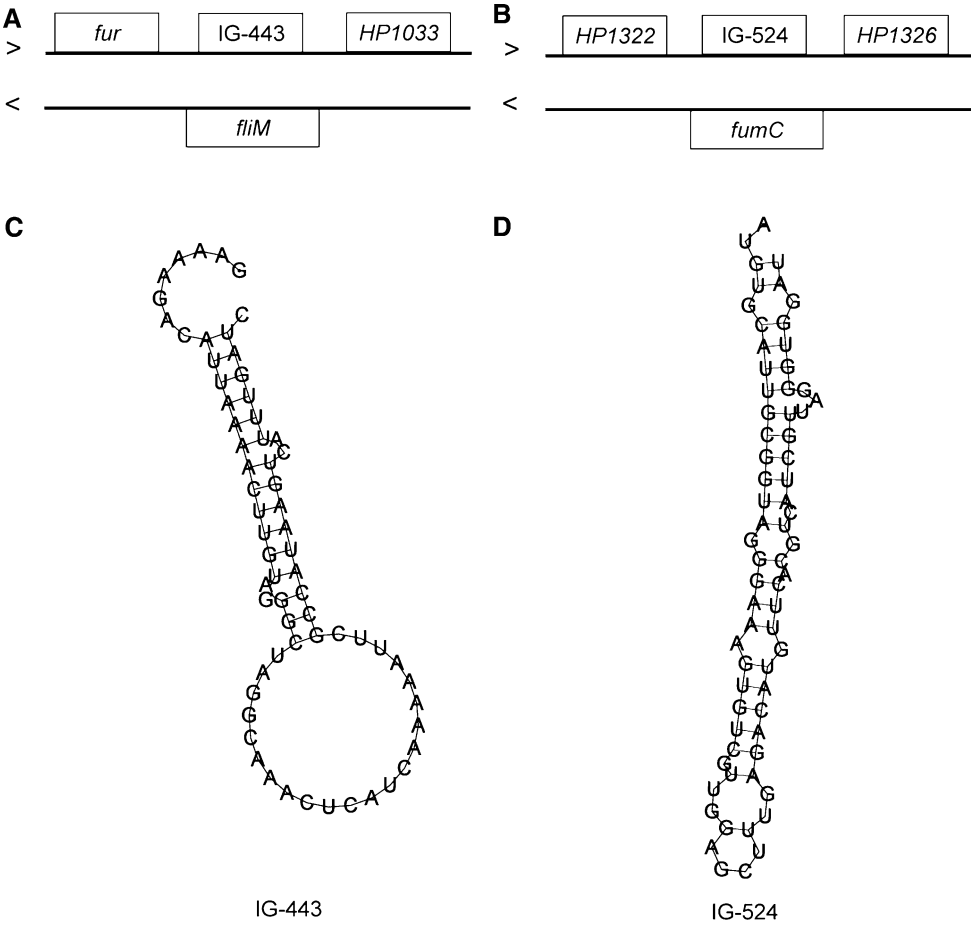

Fig. 2 The genomic location and secondary structure of sRNAs: **a** the location of IG-443 sRNA region and neighboring genes; **b** the location of IG-524 sRNA region and neighboring genes. Genes present on the strand given in the *H. pylori* genome database are indicated by the > and genes present on the complementary strand are indicated by the <; **c** the stem-loop secondary structure of IG-443 predicted by the Vienna package; **d** the stem-loop secondary structure of IG-524 predicted by the Vienna package

regulatory elements. Analysis of the *H. pylori* genome has identified only 32 genes with a possible regulatory function. This is approximately half the number of those reported for *H. influenzae*, which has a genome of comparable size to *H. pylori,* and less than a quarter of those predicted for *E. coli* [11]. Therefore, *H. pylori* seems to use complex and fascinating mechanisms to control gene expression. Recent studies have demonstrated that sRNAs have key roles in the bacterial response to stress. A small RNA in *E. coli*, RyhB, was found to downregulate a set of iron-storage and iron-using proteins when iron is limiting; it was negatively regulated by the ferric uptake repressor (Fur) protein [9]. In *H. pylori*, Fur has also been shown to play an intricate role in the adaptation of the bacterium to iron limitation and low pH. The above evidence suggests that *H. pylori* does seem to possess a number of sRNAs, and the regulation of gene expression by RNAs is more ubiquitous than we thought.

In the traditional view, there are three major classes of RNAs (mRNA, rRNA, and tRNA) in cells. Recent systematic screens for sRNAs with housekeeping or regulatory

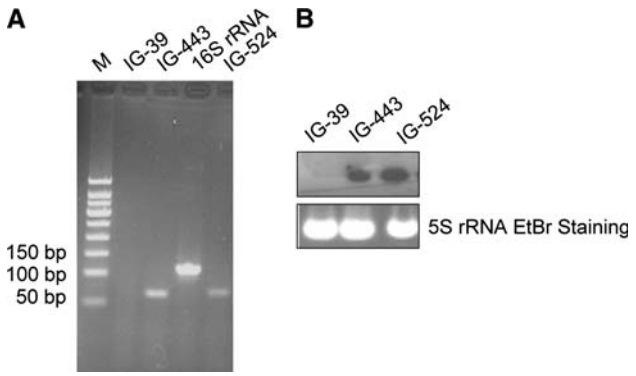

Fig. 3 Identification of sRNAs by RT-PCR and Northern blot: **a** detection of IG-443 and IG-524 by two-step RT-PCR; 16S rRNA was used as a positive control; **b** the expression of IG-443 and IG-524 sRNAs was identified by Northern blot. As a loading control, 5S rRNA was stained with ethidium bromide (EtBr)

functions have significantly changed our view about their prevalence. A large number of these molecules are now known to be encoded by bacterial, archaeal, and eukaryal genomes. In contrast to the numerous successful programs for identifying protein genes, the prediction of sRNA genes has not been a straightforward process. In 2001, 3 groups discovered 31 new sRNAs of *E. coli* by bioinformatics approaches. In these groups, the common denominator of sRNA genes prediction focused on the IGRs and sequence conservation [1, 10, 16]. In addition, a machine learning approach relying on neural networks was developed to extract common features among known sRNAs for the prediction of new candidates [2]. However, each approach has respective advantages and disadvantage. For instance, comparative analysis is complicated when sequence conservation is less important than structure conservation. Furthermore, it is very difficult for *H. pylori* to predict small RNA genes, because there were no known sRNAs in *H. pylori* to set up predictive criteria.

Here, we combined gene location, sequence conservation, promoter and terminator analysis, and secondary structure to build a strategy to search small RNA genes in the *H. pylori* genome. Using our method, IG-443 and IG-524 candidates were directly validated by Northern hybridization. In fact, the two sRNAs are a class of NATs; NATs are endogenous RNA molecules that exhibit partial or complete complementarity to other RNAs. They might be transcribed in cis from opposing DNA strands at the same genomic locus (*cis*-NATS) or in trans from separate loci (*trans*-NATs). NATs have already been found to function at several levels of gene regulation, including translational regulation, alternative splicing, and RNA stability [13].

The genomic location of sRNAs might give us some clue to investigate the target genes and functions of sRNAs. The results showed that IG-443 sRNA was complementary to flagellar motor switch gene (*fliM*). The coding product of *fliM* is required for motility of *H. pylori* and essential for the colonization and persistence in the human gastric mucosa [17]. Therefore, IG-443 might participate in the regulation of flagellar biosynthesis by pairing with target mRNA, *fliM*, and resulting in changes in the translation and stability of mRNA. In addition, fumarase is an important enzyme of the the tricarboxylic acid cycle, and IG-524 might participate in the regulation of fumarase catabolism of *H. pylori* in a sequence complementary manner.

Although generally successful in predicting sRNAs, our approach does have limitations. The main limitation is due to its reliance on IGRs. Although nearly all sRNAs identified to date are encoded in IGRs, indeed recent RNomics screens for *E. coli* sRNAs have identified a number of sRNAs that are partially encoded on the noncoding strands of ORFs [8, 12]. Another limitation of our approach is its reliance on Rho-independent terminators. However, Rho-dependent termination is also thought to occur in *H. pylori*.

In recent years, although bioinformatical strategies have been taken to identify a great number of novel sRNAs in bacteria, many sRNAs were identified by various experimental strategies [i.e., shotgun cloning of small-sized sRNAs (or RNomics), microarray analysis, genomic SELEX, and copurification with proteins]. In a previous study, we have successfully identified several regulatory sRNAs in *H. pylori* by RNomics strategy. In summary, what we know about sRNAs in bacteria is just the tip of the iceberg. Our understanding of the target and regulation functions of sRNAs is still limited. Therefore, the next challenge of our work will be to develop equally effective methods for finding the targets of those sRNAs in *H. pylori*. We believe that a better understanding of the rules of how sRNAs and their targets interact should help to unveil the full range of the function of sRNAs in *H. pylori*.

Acknowledgment This research was supported by a Chinese National Natural Science Foundation project (30770113).

References

1. Argaman L, Hershberg R, Vogel J et al (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. Curr Biol 11:941–950
2. Carter RJ, Dubchak I, Holbrook SR (2001) A computational approach to identify genes for functional RNAs in genomic sequences. Nucleic Acids Res 29:3928–3938
3. Chen S, Lesnik EA, Hall TA et al (2002) A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. Biosystems 65:157–177
4. Gottesman S (2004) The small RNA regulators of *Escherichia coli*: roles and mechanisms. Annu Rev Microbiol 58:303–328
5. Gottesman S (2005) Micros for microbes: non-coding regulatory RNAs in bacteria. Trends Genet 21:399–404

6. Guillier M, Gottesman S (2006) Remodelling of the *Escherichia coli* outer membrane by two small regulatory RNAs. Mol Microbiol 59:231–247
7. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. Cell 118:69–82
8. Livny J, Fogel MA, Davis BM, Waldor MK (2005) sRNAPredict: an integrative computational approach to identify sRNAs in bacterial genomes. Nucleic Acids Res 33:4096–4105
9. Masse E, Gottesman S (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. Proc Natl Acad Sci USA 99:4620–4625
10. Rivas E, Klein RJ, Jones TA, Eddy SR (2001) Computational identification of noncoding RNAs in *E. coli* by comparative genomics. Curr Biol 11:1369–1373
11. Scarlato V, Delany I, Spohn G, Beier D (2001) Regulation of transcription in *Helicobacter pylori*: simple systems or complex circuits? Int J Med Microbiol 291:107–117
12. Vogel J, Bartels V, Tang TH et al (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. Nucleic Acids Res 31:6435–6443
13. Wagner EG, Flardh K (2002) Antisense RNAs everywhere? Trends Genet 18:223–226
14. Wagner EG, Simons RW (1994) Antisense RNA control in bacteria, phages, and plasmids. Annu Rev Microbiol 48:713–742
15. Wassarman KM (2002) Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. Cell 109:141–144
16. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S (2001) Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev 15:1637–1651
17. Zhang ZW, Dorrell N, Wren BW, Farthingt MJ (2002) *Helicobacter pylori* adherence to gastric epithelial cells: a role for non-adhesin virulence genes. J Med Microbiol 51:495–502