Comparative Genomics of Rickettsia prowazekii Madrid E and Breinl Strains

Hong Ge,1,2 Yao-Yu Eric Chuang,3 Shuping Zhao,3 Min Tong,1,2 Mong-Hsun Tsai,3 Joseph J. Temenak,1,4 Allen L. Richards,1 and Wei-Mei Ching1,2*

Rickettsial Diseases Department, Infectious Diseases Directorate, Naval Medical Research Center, Silver Spring, Maryland 209101; Department of Preventive Medicine and Biometrics, Uniformed Services University of The Health Sciences, Bethesda, Maryland 20814-4797; Microarray Laboratory, Radiation Oncology Sciences Program, National Cancer Institute, National Institutes of Health, Gaithersburg, Maryland 20892-7712; and Division of Vaccines and Related Products Applications, Office of Vaccine Research and Review, Center for Biologics Evaluation and Research, Food and Drug Administration, Rockville, Maryland 20852*

Received 22 August 2003/Accepted 10 October 2003

Rickettsia prowazekii, the causative agent of epidemic typhus, has been responsible for millions of human deaths. Madrid E is an attenuated strain of R. prowazekii, while Breinl is a virulent strain. The genomic DNA sequence of Madrid E has recently been published. To study the genomic variations between Madrid E (reference) and Breinl (test) DNAs, cohybridization experiments were performed on a DNA microarray containing all 834 protein-coding genes of Madrid E. Of the 834 genes assessed, 24 genes showed 1.5- to 2.0-fold increases in hybridization signals in Breinl DNA compared to Madrid E DNA, indicating the presence of genomic variations in ~3% of the total genes. Eighteen of these 24 genes are predicted to be involved in different functions. Southern blot analysis of five genes, virB4, ftsK, rfbE, lpxA, and rpoH, suggested the presence of an additional paralog(s) in Breinl, which might be related to the observed increase in hybridization signals. Studies by real-time reverse transcription-PCR revealed an increase in expression of the above-mentioned five genes and five other genes. In addition to the elevated hybridization signals of 24 genes observed in the Breinl strain, one gene (rp084) showed only 1/10 the hybridization signal of Madrid E. Further analysis of this gene by PCR and sequencing revealed a large deletion flanking the whole rp084 gene and part of the rp083 gene in the virulent Breinl strain. The results of this first rickettsial DNA microarray may provide some important information for the elucidation of pathogenic mechanisms of R. prowazekii.

Rickettsiae are obligate intracellular gram-negative bacteria that belong to the alpha subdivision of Proteobacteria (48). Rickettsial diseases are widely distributed throughout the world and are characterized by sudden onset of acute fever, skin eschars, headache, and, in cases of the fulminant disease, disseminated intravascular coagulation that leads to failure of the cardiovascular system. Diagnosis of rickettsial diseases is difficult due to the lack of specific signs and symptoms and the inability to grow the bacterium on axenic microbiological media for confirmation (32). There are no licensed vaccines available for the prevention of rickettsial diseases in the United States. Moreover, three rickettsial agents, Rickettsia prowazekii, Rickettsia rickettsii, and Coxiella burnetii, have been listed as bioterror agents by the Centers for Disease Control and Prevention in the United States.

R. prowazekii is the etiologic agent of epidemic typhus in humans. Its presence is commonly associated with war, famine, and social disruption. Typhus killed >3 million people and infected in excess of 20 million people during World War I, and it killed several million more in World War II (20). As recently as 1997, there was an epidemic of typhus in Burundi that infected well over 100,000 people (31). Other outbreaks in Eastern Europe, Russia, Ethiopia, and South America further support the view that epidemic typhus is a reemerging infectious disease worldwide (20, 21). Present-day threats posed by typhus bacteria include natural exposure to antibiotic-resistant strains and the use of laboratory-manipulated strains as biological warfare agents (28). The World Health Organization has estimated that 104,000 casualties (19,000 dead and 85,000 incapacitated) would arise following a hypothetical air deployment of 50 kg of typhus agent (50).

The pathogenic mechanism of rickettsiae is unclear but may involve potential virulence factors, such as phospholipases (37, 46, 49), antigenic surface proteins (10, 19, 24, 46), hemolysins (25), type IV secretion systems (2, 25, 26), and the recently described nudix hydrolase (18). Significant biological differences exist between the virulent Breinl strain of R. prowazekii and the attenuated avirulent Madrid E strain (42). The Breinl strain grows well in macrophage-like cells, while the Madrid E strain grows poorly. The Breinl strain has a greater cytotoxic effect on macrophage-like cells treated with alpha/beta interferon (IFN-α/β) or IFN-γ than the Madrid E strain, and the Breinl strain is less sensitive to IFN-γ treatment than the Madrid E strain in L929 cells (42). However, the genes responsible for the variation in virulence between the two strains of R. prowazekii have not been well characterized. The published genomic sequence of the Madrid E strain (2) provides a foundation for comparative genomic studies of R. prowazekii. The circular genome consists of 1.1 Mb containing 834 protein-coding genes, 12 pseudogenes, and 38 RNA genes. About 24% of the genome belongs to the noncoding region, which repre-
Comparative Genomics of Rickettsia Prowazekii Madrid E and Breinl Strains

1. REPORT DATE
2004

2. REPORT TYPE
N/A

3. DATES COVERED
-

4. TITLE AND SUBTITLE
Comparative Genomics of Rickettsia Prowazekii Madrid E and Breinl Strains

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Submarine Medical Research Laboratory
Naval Submarine Base New London
Box 900 Bldg 148, Trout Avenue
Groton, CT 06349-5900

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:
   a. REPORT
      unclassified
   b. ABSTRACT
      unclassified
   c. THIS PAGE
      unclassified

17. LIMITATION OF ABSTRACT
   SAR

18. NUMBER OF PAGES
   10

19a. NAME OF RESPONSIBLE PERSON

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18
SENTS THE HIGHEST PROPORTION OF NONCODING DNA DETECTED IN ANY MICROBIAL GENOME SEQUENCED SO FAR (2). THE COMPARISONS OF FUNCTIONAL PROFILES OF MADRID E GENES WITH THOSE OF MITOCHONDRIAL GENES REVEALED MANY SIMILARITIES IN ATP PRODUCTION AND TRANSPORT SYSTEMS (2).


MATERIALS AND METHODS

Rickettsial strains and purifications. R. prowazekii strain Breinl was originally isolated from a patient in Warsaw, Poland, in 1922 and was brought to the United States in 1928. This strain was passaged multiple times and used as a representative strain of epidemic typhus. The strain used in this study was in its c1967th passage. Madrid E was the egg passage-attenuated strain of the Madrid strain, which was originally isolated from a typhus patient in Madrid (2). The Madrid E strain used here was in its c1967st passage. The exact passage histories of the two strains are unknown because of the lack of data collected early in the cultivation of the strains.

Six-day pathogen-free chicken embryos were inoculated with yolk sac suspensions from ampoules and incubated at 35°C. The infected yolk sacs were pooled, shaken with glass beads and further macerated by pipetting, diluted with brain heart infusion, and centrifuged at 10,000 × g for 30 min. The pellets were washed twice with buffer K36 (0.1 M KCl, 0.015 M NaCl, 0.05 M potassium phosphate, pH 7.0) by resuspension and sedimentation at 17,300 × g. The pellet was suspended in one-half volume of the original brain heart infusion, and 5 ml of this suspension was layered over 20 ml of 10% bovine albumin in buffer K36. After centrifugation at 480 rpm (SW27 rotor) for 1 h. After centrifugation, the areas of the gradient above and below the rickettsial band were removed with a Pasteur pipette. Then, the rickettsial band was drawn, and the final pellet was suspended in K36. Live rickettsiae purified by this method were essentially free of host material.

Construction of microarray slides. (i) DNA extraction and amplification. Renografin-purified R. prowazekii Madrid E was resuspended in ATL lysis buffer (Qiagen, Valencia, Calif.). After proteinase K digestion and RNase A treatment according to the instructions of the manufacturer, DNA was extracted three times with phenol-chloroform-isooamy alcohol and precipitated with ethanol-sodium acetate. PCR primers for all 834 genes covering the entire open reading frames (ORF) of each gene from R. prowazekii Madrid E were purchased from Research Genetics (Huntsville, Ala.), and they were 100 nucleotide sequences using a 96-well PE-9700 thermocycler (Applied Biosystems, Foster City, Calif.). The reaction mixture contained 15 ng of DNA, 0.1 mM deoxynucleoside triphosphate, 100 ng of each primer, 1× PCR buffer, 1.5 mM MgCl2, and 2.5 U of Taq polymerase (Invitrogen-Life Technology, San Diego, Calif.). The reaction conditions were as follows: initial denaturation at 94°C for 30 s and 32 cycles of denaturation (94°C for 1 s), annealing (53°C for 45 s), and extension (72°C for 45 s). The final extension was performed for 10 min after the residual NaOH was completely removed by vigorously washing the slides for 20 min in double-distilled H2O. The slides were immediately spun at 550 rpm for 3 min and then stored in desiccators at room temperature until spotting was performed.

(ii) Preparation of glass slides for microarray. Microscope slides were selected manually in order to exclude those slides with dust or scratches. The slides were first washed for 2 h and then coated with the residual NaOH was completely removed by vigorously washing the slides for 20 min in double-distilled H2O. The slides were immediately spun at 550 rpm for 3 min and then stored in desiccators at room temperature until spotting was performed.

(iii) Spotting and posttreatment of slides. All PCR products were purified by ethanol-acetate precipitation and then spotted onto polysine-coated slides by the computerized OMNI Grid Arrayer (Gene Machines, San Carlos, Calif.). The eight-pin model was used for spotting the DNAs of 834 genes plus a 96-gene duplicate as the internal control. The spotted DNA was cross-linked to polysine by UV irradiation and treated with sucinic anhydride to minimize background staining.

Hybridization of genomic DNA. Genomic DNAs from the Madrid E and Breinl strains were used as templates for direct incorporation of fluorescent nucleotide analogs (Cy3 and Cy5 dCTPs; Amersham, Piscataway, N.J.) by a random-priming polymerization reaction (6). After denaturation of the DNA, the two 50-μl labeling reaction mixtures containing 2 μg of template DNA (one from Madrid E and another from Breinl), 1× random-primer reaction buffer (Invitrogen-Life Technology), 1.2 mM (each) dATP, dGTP, and dTTP; 0.6 μM Cy3 dCTP; 3.75 nmol of Cy5-dCTP or Cy5-sdCTP; and 40 U of Klenow fragment (Invitrogen-Life Technology) were incubated at 37°C for 2 h. By adding Cot-1 DNA, yeast tRNA, and Tris-EDTA buffer, the labeled DNA was then concentrated, denatured, and applied to microarray slides. The hybridization was carried out at 65°C overnight, and then the slides were washed for 5 min in 2× SSC buffer (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) with 0.3% sodium dodecyl sulfate, followed by 5 min in 1× SSC and then 1 to 2 min in 0.2× SSC buffer. In a subsequent experiment, each DNA was reciprocally labeled in order to eliminate artifacts due to differences in labeling intensity.

Imaging and data analysis. The hybridized arrays were scanned at 10-μm resolution on a GenePix 4000 scanner (Axon Instruments, Inc., Foster City, Calif.) at variable photomultiplier tube voltage settings to obtain maximal signal intensities with c1967% probe saturation. The Cy5- and Cy3-labeled DNA samples were scanned at 635 and 532 nm, respectively. The resulting TIFF images were analyzed by GenePix Pro version 3.0. For each spot, the background intensity was subtracted from the spot signal intensity to produce a channel-specific value. The Cy5/Cy3 ratios for all targets were determined after background subtraction. Gene expression analysis was performed using the National Cancer Institute-Center for Information Technology microArray database system. Hierarchical clustering was performed according to the method of Eisen et al. using an “uncentered” correlation algorithm (14). The cluster data were filtered for signal intensities and background of >1.5-fold in both channels. The raw data were converted into log2 data for the analysis of the genomic distribution versus log2 ratio.

Southern blot analysis. Southern blot hybridization was carried out by using standard protocols (35) with the following modifications. Probes were prepared using a PCR DNA biotinylation kit (Kirkegaard & Perry Laboratories, Gaithersburg, Md.) that yielded products between 0.5 and 1.0 kb in size. MacVector version 7.0 software (Genetics Computer Group, Madison, Wis.) was used to select restriction enzymes for the digestion of genomic DNA to generate proper length distributions and to predict signal sizes in Southern blot analysis. Rickettsial DNA was digested with EcoRI/XbaI and HindIII for the detection of fksA, rfbE, lpaA, and rpoH. HindIII was replaced by EcoRV/ClaI for the detection of rvB4. None of these restriction enzymes cut Madrid E DNA sequences that were designed for the generation of probes. The RNA segments were separated on 0.8% agarose gels by overnight electrophoresis and then transferred onto a nitrocellulose membrane. The hybridization was performed by using a Terminal deoxynucleotidyl transferase cDNA labeling kit (Pharmacia, Piscataway, N.J.). The hybridization was carried out at 65°C for 24 h. The slides were then washed at high stringency (SSC 0.4×, 0.1×, and 0.05× at 65°C) and then exposed to X-ray film. The band intensities were quantitated by a computerized image analyzer (Delta GenRad, Northridge, Calif.) using the National Cancer Institute software. The gene expression levels were expressed as copy numbers per genome and were compared with standard curves of known copy numbers for each gene on each slide.
TABLE 1. Primers and probes used in real-time PCR

<table>
<thead>
<tr>
<th>Target</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>virB4</td>
<td>CCT TTG CAT CTC TIT GAA GAT ACT AGT GAG A</td>
</tr>
<tr>
<td>Reverse</td>
<td>TAT CTT GCG CAG TTA AAC TTT CAC C</td>
</tr>
<tr>
<td>Probe</td>
<td>AAG TTA CAA GCA CGC GAA GCC ACT CTA AA</td>
</tr>
<tr>
<td>ftsK</td>
<td>GCA GCC TTA TAG CCG ATT CTA</td>
</tr>
<tr>
<td>Reverse</td>
<td>AAA AGC AGT ACA GAT AGT GCG TAA TG</td>
</tr>
<tr>
<td>Probe</td>
<td>AGC TAA ACG CTC TAG CTA TAT CAA AAG GT</td>
</tr>
<tr>
<td>rfbE</td>
<td>TCT TGT TAT GAG GGA AAA ATA GCT</td>
</tr>
<tr>
<td>Reverse</td>
<td>ATT CCA GCA ATC AGT TTT AGA AGT GA</td>
</tr>
<tr>
<td>Probe</td>
<td>TTT TAC CTG AAC CAT TAC TCC CAA TAA</td>
</tr>
<tr>
<td>lpxA</td>
<td>AAG CAC TGC GCG TTT ACT ACT TAC</td>
</tr>
<tr>
<td>Reverse</td>
<td>ATT CGA TAG GCG GAC TAT CA</td>
</tr>
<tr>
<td>Probe</td>
<td>TCC AAA CGG TAT TAC ATC TGC GGC AAC</td>
</tr>
<tr>
<td>rpoH</td>
<td>TCA GTG CAT GAA CAC GGT AAG TC</td>
</tr>
<tr>
<td>Reverse</td>
<td>GCT GCG CTC CAT CTT TAC G</td>
</tr>
<tr>
<td>Probe</td>
<td>TAC CAA TAT CTG GCT TAT TCT CCA AA</td>
</tr>
<tr>
<td>ampG2</td>
<td>GCG ATA AAC ATC CTG CCA AGA</td>
</tr>
<tr>
<td>Reverse</td>
<td>GCA TAA GTG AAG CGG CIT GT A</td>
</tr>
<tr>
<td>Probe</td>
<td>AGT ACC TCT TTC ATC ACA ATG CCG TTC TAC AAG AGT A</td>
</tr>
<tr>
<td>pal</td>
<td>AGC TGC TGC CCT TCT TCT G</td>
</tr>
<tr>
<td>Reverse</td>
<td>AGC TTG CTG GGT ATC AAA ACA TCC</td>
</tr>
<tr>
<td>Probe</td>
<td>AGT ACC TCT TTC ATC ACA ATG CCG TTC TAC AAG AGT A</td>
</tr>
<tr>
<td>iles</td>
<td>AGG CGA CGG AAC AGG AAT AGT A</td>
</tr>
<tr>
<td>Reverse</td>
<td>TCA ACA GGG CAA ACA AGT TCA A</td>
</tr>
<tr>
<td>Probe</td>
<td>TAT GGC TCC AGG CTT TGG TGA AGA TGA</td>
</tr>
<tr>
<td>uvrB</td>
<td>CAA CTC CAC CTC AAT GCT TGA TAT</td>
</tr>
<tr>
<td>Reverse</td>
<td>CAG GTG AGA AAC TTA CTC AGC TGG ATA</td>
</tr>
<tr>
<td>Probe</td>
<td>CCT TTC TCG TGG CAT AAC AAA GTG TGA ATT ACC</td>
</tr>
<tr>
<td>16S RNA</td>
<td>GGA GCA TGC GGT TTA ATT CG</td>
</tr>
<tr>
<td>Reverse</td>
<td>AAA GCA TCT CTG CAA TCC GTA AC</td>
</tr>
<tr>
<td>Probe</td>
<td>ACG CGA AAA ACC TTA CCA ACC CTT GAC AT</td>
</tr>
</tbody>
</table>

* Probes were sense or antisense sequences labeled with FAM at the 5’ end and with TAMRA at the 3’ end.

for 45 s); and a final extension step at 72°C for 10 min. Amplification of the 16S rRNA gene was done as an internal control under the same conditions as for virB4 with the following primers: 5’ GAA TTA ACC GCA TGC TCC AC 3’ (forward) and 5’ AAT TAA ACC GCA TGC TCC AC 3’ (reverse).

DNA sequencing. PCR primers that flanked the sequence of Madrid E from the end of npt to the beginning of npt were designed. A PCR product from Madrid E was calculated to be a 4.3-kb fragment. The PCR product from Breinl was ~2.5 kb and was directly sequenced with the Big-Dye terminator cycle-sequencing kit using the automated sequencing machine (Applied Biosystems). In addition, the full-length genes virB4, ftsK, lpxA, rfbE, and rpoH of Breinl and Madrid E were PCR amplified and cloned into the BlueScript vector (Stratagene, La Jolla, Calif.) and sequenced as described above. DNA sequence analyses and alignments were performed with MacVector version 7.0 and Blast software.

RNA extraction and RT. Total RNA was isolated from the Renografin-purified rickettsial strains Madrid E and Breinl with TRIzol (Invitrogen-Life Technology) according to the manufacturer's instructions. The RNA samples were treated with DNase I (Invitrogen-Life Technology) at 37°C for 1 h and then extracted with phenol-chloroform-isooamyl alcohol two or three times. The quality and quantity of RNA were confirmed by visualization of the agarose gel and spectrophotometer analysis. Two micrograms of total RNA was used for cDNA synthesis at 42°C for 50 min with SuperScript II (Invitrogen-Life Technology), followed by RNase H treatment at 37°C for 20 min. A reaction with no reverse
transcriptase was included as a control. All procedures were performed for both strains under identical conditions to allow comparison.

**Real-time RT-PCR.** The primer pairs and probes of selected genes for real-time PCR were designed using Primer Express software (ABI, Weiterstadt, Germany) as shown in Table 1. The probes were labeled with 5'-6-carboxyfluorescein and 5'-6-carboxytetramethyl-rhodamine at the 5' and 3' ends, respectively. PCR cycling was conducted using the SmartCycler (Cepheid, Sunnyvale, Calif.) in reaction mixtures with total volumes of 25 µl containing 1X TaqMan Universal PCR Master Mix (Applied Biosystems), 0.2 µM each primer, 0.1 µM probe, and 1 µl of diluted cDNA template. After incubation at 50°C for 2 min and denaturation at 95°C for 10 min, the reaction went to 45 to 60 cycles of denaturation (94°C for 15 s) and extension (60°C for 1 min). The 16S rRNA gene was included as an endogenous reference, and the comparative CT (threshold cycles) method was applied by using arithmetic formulas $(2^{-\Delta CT})$ (ABI system; AB, Piscataway, N.J.). In this method, the amount of target was normalized to that of the reference gene relative to the calibrator. The efficiency of the target amplification and the efficiency of the reference amplification were tested by assessing the variations of $\Delta C_T$ at different amounts of template from the reference and sample. No-cDNA templates in PCR and RT products without reverse transcriptase were used as controls. Real-time RT-PCR assays were performed in triplicate for each sample, and a mean value and standard deviation were calculated for the relative RNA expression levels.

**FIG. 1.** DNA microarray analysis of 834 *R. prowazekii* genes plus a 96-gene repeat as internal control. (A) Cohybridization of Cy3-labeled Madrid E DNA and Cy5-labeled Breinl DNA. (B) Reciprocal-labeling experiment. (C and D) Higher magnifications of representative images from panels A and B. (E) Locations of all 834 ORFs within the 1.1 Mb in Madrid E are mapped on the x axis. The log2(Cy5/Cy3) ratio for each gene is shown on the y axis. Cy5/Cy3 ratios of >1.5 or <0.667, equivalent to ± 0.58 in the log2 ratio, are indicated by the red bars.

**FIG. 2.** Cluster image of genes with fluorescence ratios of ≥1.5 or ≤0.667 in comparisons of Madrid E and Breinl in two separate experiments. In lane 1, Madrid E DNA was labeled by Cy3 (green) and Breinl DNA was labeled by Cy5 (red). Lane 2 shows a repeat experiment by reciprocal labeling. Ratio 1 and ratio 2 represent the values of 25 genes obtained from the two individual experiments.
RESULTS AND DISCUSSION

Comparison of genomic DNAs from Madrid E and Breinl by microarray. The cohybridization of Cy3-labeled Madrid E DNA (green fluorescence signal) and Cy5-labeled Breinl DNA (red fluorescence signal) was applied on the glass microarray slides. The resulting signal intensities from the two fluorophores were compared for each gene in order to detect the genomic differences, e.g., DNA duplications or deletions, in the two strains. It should be noted that no plasmid has been found in the two strains. It should be noted that no plasmid has been found in Madrid E, 5 of them (rfbE, lpxA, rpmH, and virB4) were cloned and sequenced. The results demonstrated that the sequences of all five genes were identical. No insertions, duplications, deletions, or mutations were seen.

DNA sequencing and Southern blot analysis. To rule out sequence variation among the 25 genes in Breinl and Madrid E, 5 of them (fisK, rfbE, lpxA, rpmH, and virB4) were cloned and sequenced. The results demonstrated that the sequences of all five genes were identical. No insertions, duplications, deletions, or mutations were seen.

To evaluate the possibility of multiple copies or paralogs of the disparate genes and to explain why the hybridization signal ratios were between 1.5 and 2.0 with these genes, Southern hybridizations were performed for the above-mentioned five genes (Fig. 3). Total DNAs were isolated from the R. prowazekii Madrid E and Breinl strains, respectively, and then digested with restriction enzymes, which were selected by accounted for ~3% (25 of 834) of all genes detected on the array. Eighteen out of 25 genes fell within functional categories based upon sequence homology to other bacterial genes (2). The remaining seven genes are unmatched to any homologous function (Table 2) (2). Since the fluorescence ratios of 24 genes were >1.5, we evaluated the possibility of additional gene copies in Breinl compared to Madrid E in light of a report of different copy numbers of homologous genes. There are 54 paralogous gene families comprising 147 gene products reported in the genome of R. prowazekii Madrid E (2). This observation may indicate specific adaptations of bacteria to their environments. The presence of four copies of SpoT in R. prowazekii and Rickettsia conorii might be related to the adaptation of Rickettsiae to environmental starvation in their vectors (25). Other possible factors contributing to the increased fluorescence signals may be differences in the number of gene repeat sequences, gene insertions in Breinl, or deletion of genes in Madrid E.

### TABLE 2. Functional categories for 25 genes identified by microarray

<table>
<thead>
<tr>
<th>Category</th>
<th>Protein type</th>
<th>Gene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell envelope</td>
<td>Surface polysaccharides,</td>
<td>RP003</td>
<td>O-antigen ABC export system; ATP-binding protein</td>
</tr>
<tr>
<td></td>
<td>lipopolysaccharides, and antigens</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP007</td>
<td>lpxA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP771</td>
<td>pal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP451</td>
<td>sca3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP825</td>
<td>mraY2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Murin sacculus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transport and binding</td>
<td>Cell division</td>
<td>RP823</td>
<td>Cell division protein</td>
</tr>
<tr>
<td>protein</td>
<td>Chaperone and stress-induced</td>
<td>RP670</td>
<td>Cold shock protein</td>
</tr>
<tr>
<td>Cellular processes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP617</td>
<td>iles</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP100</td>
<td>rpmE</td>
</tr>
<tr>
<td>Translation</td>
<td>Aminocyl-tRNA synthetase</td>
<td>RP773</td>
<td>rpmF</td>
</tr>
<tr>
<td></td>
<td>Ribosomal proteins; synthesis</td>
<td>RP456</td>
<td>rpmJ</td>
</tr>
<tr>
<td></td>
<td>and modification</td>
<td>RP660</td>
<td>rpmJ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP141</td>
<td>rpmC</td>
</tr>
<tr>
<td>Transcription</td>
<td>RNA synthesis and modification</td>
<td>RP303</td>
<td>rpmH</td>
</tr>
<tr>
<td>Replication</td>
<td></td>
<td>RP203</td>
<td>avrB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP103</td>
<td>virB4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP668</td>
<td>ampG2</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td>RP084</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP105</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP158</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP401</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP725</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP870</td>
<td></td>
</tr>
</tbody>
</table>

* Functions compiled according to Andersson et al. (2).
MacVector version 7.0 software based on the published Madrid E sequence. The restriction enzymes were predicted to cut frequently within Madrid E DNA but not within the regions for probes. The Southern blots were hybridized with five DNA probes synthesized by PCR. The resulting blots displayed one strong band each for ftsK, rfbE, lpxA, and rpoH in all lanes of Madrid E and Breinl DNAs digested with EcoRI/XbaI and HindIII, which was consistent with the predicted single copies of these genes in the Madrid E genome (2). However, the blots of Breinl DNA digested by EcoRI/XbaI appeared to have an additional weak band for all four genes. These additional bands were not seen in the Breinl strain after HindIII digestions. The blots hybridized with the rp103 probe for the detection of virB4 displayed two major bands in Madrid E. One was located at the expected position of 3,596 bp containing rp103. The second band resulted from the hybridization of the probe to another virB4 (pp784) and was located at a predicted position of 1,426 bp. However, the blots from Breinl DNA showed a third band at 1.3 kb, in addition to the two similar bands seen in Madrid E for virB4. The genome sequence for Madrid E has confirmed two paralogs (rp103 and pp784) of virB4, which showed 30.9 and 20.3% DNA identities to Agrobacterium tumefactions virB4 (2). Further sequencing alignments for rp103 and pp784 showed only 49 and 25% identities in DNA and protein, respectively, dispersed over the whole ORF, but both contain a consensus VirB4 domain. These findings with virB4 demonstrated that probe rp103 could hybridize to pp784 in addition to itself. Correspondingly, the third band in Breinl may represent another hybridization between rp103 and a paralog of virB4. All these data support the notion that there may be an additional paralog(s) of virB4, ftsK, rfbE, lpxA, and rpoH in the Breinl strain that is not found in the Madrid E strain. Since paralog genes are not identical in their DNA sequences, despite belonging to the same functional family, it is conceivable that the signals due to paralogous hybridization may be not as strong as those due to homologous hybridization. We are not sure why there are no weak bands due to paralogs in Breinl DNA for ftsK, rfbE, lpxA, and rpoH after HindIII digestion, as well as for virB4 after EcoRV/ClaI digestion. It is possible that the paralogs have restriction cutting sites within the probe-complementary regions, or it may simply be that due to weak hybridization the signals were too faint to be seen. These data collectively suggested that an additional paralog(s) of the five genes might exist in the virulent Breinl strain, providing an explanation for the observation of gene hybridization ratios between 1.5 and 2.0 from microarray experiments.

Real-time RT-PCR analysis. To estimate the expression profiles of these genes in yolk sac-grown R. prowazekii, we conducted a comparative quantification of the gene expression profiles for 10 of the genes with detected differences in the microarray. Specific primers and probes were designed for each gene in order to perform real-time PCR. The target sequences were amplified from both Madrid E and Breinl. Control experiments without reverse transcriptase generated weak signals with Ct values near 60 cycles. Other controls without cDNA templates were negative in all experiments. The collected data were analyzed using the standard comparative 2−ΔΔCt method. With this method, the expression profiles of the two strains were compared using the 16S rRNA gene as the control.
internal control. The results showed that all real-time PCRs were positive for the corresponding target sequences with $C_T$ values from 27 to 42. The expression levels of all 10 genes were higher in the virulent Breinl strain than in the avirulent Madrid E strain, ranging from 1.3 to 9.9 times higher. The genes with twofold-higher or greater expression in Breinl were virB4, rfbE, lpxA, ampG2, pal, iles, and uvrB. These results are summarized in Table 3. The real-time PCR and microarray data for these 10 genes are illustrated in Fig. 4. The expression ratios of lpxA, ampG2, pal, and uvrB were higher than those of other genes compared to the DNA ratios by microarray experiments, suggesting different transcriptional regulations in Breinl than in Madrid E.

**Analysis of rp084.** One of the intriguing observations in this study was that the rp084 gene exhibited only $\sim 1/10$ (0.121, averaging from 0.158 and 0.084) of the intensity in Breinl DNA as in Madrid E DNA (Fig. 1 and 2). To verify the microarray result, amplification of rp084 by PCR and RT-PCR were performed for the two strains. Neither the DNA nor the mRNA of this gene was detected in Breinl, but both were present in Madrid E (Fig. 5A and B). PCR amplification of the region flanking rp083, rp084, and nearby noncoding regions produced a 4.3-kb fragment in Madrid E (expected size) and a 2.5-kb product in Breinl. Sequencing analysis of the amplicon from Breinl revealed that there was a deletion of 1,812 bp, including part of rp083, all of rp084, the noncoding region between rp083 and rp084, and part of the noncoding region after rp084 (Fig. 5C). We are not sure of the significance of this deletion of genomic DNA for the virulent Breinl strain. The presence of rp084 was detected by PCR amplification in another virulent strain of *R. prowazekii*, MADII, but not in the Ananier strain (data not shown). Both rp083 and rp084 belong to the category of unknown function in Madrid E (2). There were no obvious similarities in either the primary DNA or protein sequence after an exhaustive database search. Following a BLAST database search with the predicted mass (18,232 Da) and pI (9.52) for rp084, three mitochondrial proteins, CTAG (cytochrome c oxidase assembly protein), CYCJ (cytochrome c-type biogenesis protein), and OSCP (ATP synthase oligomycin sensitivity conferral protein), with the same pI and very similar molecular masses of $\sim 18$ kDa were identified. They show similarities in the predicted secondary structures (data not shown), and putative biochemical characteristics, such as overall hydrophilicity, a high content of basic amino acids (higher pI value), and a predicted $\alpha$-helical secondary structure, all of which are indicative of mitochondrion-related proteins (17). CTAG
found in the rickettsia Cowdria ruminantium was shown to affect cytochrome c oxidase synthesis at a terminal stage (45). CYCJ might be required for the biogenesis of c-type cytochromes in Bradyrhizobium japonicum and Rhizobium etli (33, 41). In Drosophila melanogaster and Saccharomyces cerevisiae, OSCP was implicated in proton conduction during ATP synthesis (4, 9, 34). Cytochrome c is believed to be an important factor in the initiation of the apoptotic cascade. The cytochrome c released in response to stress could activate a caspase initiator, caspase 9, and trigger the subsequent activation of downstream executioner caspases in the mammalian host (22, 52). Further functional studies of rp084 are necessary to demonstrate the specific role of this rickettsial gene.

Potential virulence genes identified in this study. The virB4, ftsK, lpxA, and rfbE genes are known to be virulence-related genes, and their overrepresentation in the virulent Breinl strain might indicate that they have roles in the virulence of R. prowazekii. Although there are limited reports of the study of these genes in Rickettsia, there have been investigations in Escherichia coli and other bacteria.

(i) **virB4.** The R. prowazekii Madrid E genome possesses two virB4 paralogs (rp103 and rp784) that are homologous to the virB4 genes encoding bacterial transport systems, such as T-DNA export in A. tumefaciens, with DNA sequence identities of 30.9 and 20.3%, respectively (2). VirB4 is believed to be an essential virulence protein in A. tumefaciens as a cytoplasmic membrane ATPase (11, 16, 23). VirB4 also functions in plasmid transport in E. coli, in P1 toxin export in Bordetella pertussis, in the Cag system in Helicobacter pylori, and in the icm/dot system in Legionella pneumophila (reference 12 and references therein). However, the role of virB4 in rickettsiae has not been elucidated. In addition, other virB family members have been reported in R. prowazekii Madrid E, such as virB8-11 and virD4 (2).

(ii) **ftsK.** FtsK is predicted to be involved in bacterial cell division. FtsK may possess two functions: the membrane domain (FtsKM) targets the protein to the closing septum (47), and the cytoplasmic portion (FtsKC) segregates chromosome dimers during cell division (39, 40, 51). A mutation in ftsK44 appeared to block the late phase in the cell division pathway (5). This information suggests that the FtsK protein plays a crucial role in cell division by triggering chromosome segregation. There has been no report of the study of ftsK in rickettsiae.

(iii) **lpxA and rfbE.** LpxA encodes UDP-N-acetylgalactosamine (UDP-GlcNAc) acetyltransferase, which is required for the first step of lipid A (endotoxin) biosynthesis (30). LpxA catalyzes the transfer of an R-3-hydroxyacyl chain from R-3-hydroxyacyl-acyl carrier protein to the glucosamine 3-OH of UDP-GlcNAc (1). The growth and virulence of most gram-negative bacteria require lipid A (27). It is also a potent activator of innate immunity in animal systems (30, 43). The rfbE gene encodes an enzyme necessary for the biosynthesis of the O-antigen side chain in E. coli O157:H7 (7). Rickettsial lipopolysaccharide appears to be an important immunogen, as well as a virulence factor (36, 38). Thus, rfbE has a potential role in the pathogenesis of R. prowazekii.

(iv) **rpoH.** The RNA polymerase sigma-32 factor (RpoH) has been characterized as a primary component of the heat shock response in the survival of E. coli (8). In response to a sudden heat shock or other stresses, the levels of RpoH rise transiently through enhanced synthesis and protein stabilization. In R. prowazekii, sigma-73 factor has been overexpressed in E. coli, and it can stimulate transcription by promoting RNA polymerase binding (3). Rickettsial heat shock proteins, including DnaK, GroEL, and GroES, have been detected by Western blot analysis (15), but the regulatory mechanism for heat shock or stress is unknown.
In conclusion, we have successfully assembled the first genomewide DNA microarray for *R. prowazekii* and used this array to compare genomic variation between the virulent Breinl and avirulent Madrid E strains. Differences in the number of paralogs may be present for a few genes, such as *virB4*, *ftsK*, *rfbE*, *lpxA*, and *rpoH*, in Breinl and Madrid E. The absence of the whole *rp084* and partial *rp083* genes in Breinl is of unknown relevance to rickettsial virulence. Our results in this report may provide new insights into the molecular pathogenesis of *R. prowazekii*, such as the differences in the numbers of paralog genes, that may be useful for the development of improved diagnostic assays and vaccines. Further investigations are planned using this microarray to perform differential expression profile analysis of rickettsial mRNAs from infected mammalian cells and arthropod vectors.

**ACKNOWLEDGMENTS**

This work was supported by the Naval Medical Research and Development Command, Work Unit 6000.ID.RD.A0239.

The opinions and assertions contained here are those of the authors and are not to be construed as of official.

The Renografin-purified rickettsiae were prepared under Gregory A. Dasch’s direction. We thank Edwin Oaks for his critical comments during the preparation of the manuscript.

**REFERENCES**


