Identification of *Haemophilus ducreyi* genes expressed during human infection

Margaret E. Bauer,† Kate R. Fortney,‡ Alistair Harrison, Diane M. Janowicz, Robert S. Munson, Jr3,4 and Stanley M. Spinola1,2,5,6

1Department of Microbiology and Immunology, Indiana University School of Medicine, 635 Barnhill Drive, Room MS420, Indianapolis, IN 46202, USA
2Department of Medicine, Indiana University School of Medicine, 545 Barnhill Drive, Room EH 435, Indianapolis, IN 46202, USA
3The Center for Microbial Pathogenesis in The Research Institute at Nationwide Children’s Hospital, 700 Children’s Drive, Columbus, OH 43205-2696, USA
4The Center for Microbial Interface Biology and Department of Pediatrics, The Ohio State University, 700 Children’s Drive, Columbus, OH 43205-2696, USA
5Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN 46202, USA
6Center for Immunobiology, Indiana University School of Medicine, Indianapolis, IN 46202, USA

To identify *Haemophilus ducreyi* transcripts that are expressed during human infection, we used selective capture of transcribed sequences (SCOTS) with RNA isolated from pustules obtained from three volunteers infected with *H. ducreyi*, and with RNA isolated from broth-grown bacteria used to infect volunteers. With SCOTS, competitive hybridization of tissue-derived and broth-derived sequences identifies genes that may be preferentially expressed *in vivo*. Among the three tissue specimens, we identified 531 genes expressed *in vivo*. Southern blot analysis of 60 genes from each tissue showed that 87% of the identified genes hybridized better with cDNA derived from tissue specimens than with cDNA derived from broth-grown bacteria. RT-PCR on nine additional pustules confirmed *in vivo* expression of 10 of 11 selected genes in other volunteers. Of the 531 genes, 139 were identified in at least two volunteers. These 139 genes fell into several functional categories, including biosynthesis and metabolism, regulation, and cellular processes, such as transcription, translation, cell division, DNA replication and repair, and transport. Detection of genes involved in anaerobic and aerobic respiration indicated that *H. ducreyi* likely encounters both microenvironments within the pustule. Other genes detected suggest an increase in DNA damage and stress *in vivo*. Genes involved in virulence in other bacterial pathogens and 32 genes encoding hypothetical proteins were identified, and may represent novel virulence factors. We identified three genes, *IspA1*, *IspA2* and *tadA*, known to be required for virulence in humans. This is the first study to broadly define transcripts expressed by *H. ducreyi* in humans.

INTRODUCTION

*Haemophilus ducreyi* is the causative agent of chancroid, a genital ulcer disease. *H. ducreyi* facilitates both the acquisition and transmission of human immunodeficiency virus (HIV)-1 and contributes to the HIV-1 pandemic in certain regions of Africa and Asia (Steen, 2001).

To study the pathogenesis of *H. ducreyi* infection, we developed an experimental model of infection in which healthy adult volunteers are infected on the upper arm with 101–102 c.f.u. *H. ducreyi*. The clinical course of disease in the model accurately mimics natural disease through the pustular stage (Al-Tawfiq et al., 1998; Palmer et al., 1998; Spinola et al., 2002). The histology of experimental pustules and naturally acquired ulcers are nearly identical, and *H. ducreyi* maintains the same general relationship with host cells during experimental and natural infection.

†These authors contributed equally to this work.

Abbreviation: SCOTS, selective capture of transcribed sequences.
expression in Listeria monocytogenes in response to temperature changes (Liu et al., 2002) and to compare gene expression between related bacterial species or strains (Dozois et al., 2003; Morrow et al., 1999).

In the present study, we employed the SCOTS procedure to capture \textit{H. ducreyi} transcripts that are expressed during experimental human infection. This study has led to the identification of several genes that are required for infection in humans.

**METHODS**

**Tissues.** Pustules for SCOTS were obtained by biopsy from three women who had participated in human-challenge trials (see Supplementary Table S1). Nine additional pustules were obtained from seven men and two women for RT-PCR (Supplementary Table S1). All pustules were obtained 6–9 days after inoculation when the subjects reported pain. Informed consent for participation and for HIV serology was obtained from the volunteers in accordance with the human experimentation guidelines of the US Department of Health and Human Services and the Institutional Review Board of Indiana University-Purdue University at Indianapolis.

**RNA preparation.** Whole biopsies were immediately placed in RNalater (Qiagen). After 30 min, tissues were homogenized in buffer RLT provided in the RNeasy Fibrous Tissue kit (Qiagen) using either a tissue homogenizer or a bead beater (Biospec Products) with 2.4 mm Zirconia beads (Biospec). RNA was extracted with the RNeasy Fibrous Tissue kit following the manufacturer’s directions, except that lysozyme (400 µg ml\(^{-1}\)) was added to the protease K digestion step to lyse the bacteria. \textit{In vitro}-derived RNA was obtained from a culture of \textit{H. ducreyi} 35000 HP (HP, human passage) (Al-Tawfiq et al., 1998) grown in broth and used to inoculate human volunteers, as described previously (Al-Tawfiq et al., 1998; Spinola et al., 1994, 1996). At mid-exponential phase (OD\(_{660}\)=0.2), 4 ml bacteria were collected by centrifugation and suspended in 1 ml Ultraspec RNA (Biotec Laboratories). Total RNA was extracted following the manufacturer’s directions. After isolation, all RNA samples were treated with DNase (Ambion) following the manufacturer’s directions. Integrity of the RNA samples was determined with an Agilent Bioanalyzer (Agilent Technologies).

**cDNA synthesis.** RNA was converted to double-stranded cDNA by random priming with the Advantage RT-for-PCR kit (Clontech Laboratories), following the manufacturer’s instructions, and second-strand synthesis as described elsewhere (Froussard, 1992; Hou et al., 2002). The primers had defined 5’ sequences followed by nine random 3’ bases. Primers with different 5’ sequences were used to differentially tag the 5’ ends of cDNA derived from tissue (5’-GACACTCTCGAGACAGGCGGTACC-3’) and from broth-grown bacteria (5’-CCTCTGAAAGTTCCTGAGATCGATAG-3’). After synthesis, each cDNA pool was amplified by PCR using primers consisting of the appropriate 5’ tag without the random 3’ sequences. The amplimers were precipitated and suspended in 10 mM HEPES–1 mM EDTA to a final concentration of 375 ng ml\(^{-1}\).

**rDNA plasmid construction.** A 5.8 kb locus containing 5S, 16S and 23S rRNA genes was amplified by PCR from 35000 HP using the following primers: (forward) 5’TTCGCAAGAGGAAAAAGATATCCTATAATGG-3’ and (reverse) 5’TATGACCAATTACCTTATTGTGTGTGGTGGG-3’. The amplicon was cloned into the TA cloning vector pCR-XL-TOPO (Invitrogen) to make pSCOTS1 and propagated in \textit{Escherichia coli} TOP10F’ (Invitrogen).
SCOTS. The SCOTS procedure (illustrated in Supplementary Fig. S1) (Daigle et al., 2002; Faucher et al., 2006; Graham & Clark-Curtiss, 1999; Graham et al., 2002; Liu et al., 2002; Morrow et al., 1999) was followed as described by Hou et al. (2002). Graham & Clark-Curtiss (1999) showed that three rounds of SCOTS was sufficient to effectively eliminate eukaryotic sequences and normalize the relative levels of transcripts in each cDNA pool. Because the M. tuberculosis chromosome contains one set of rRNA genes, while the H. ducreyi chromosome contains six rRNA gene clusters, we increased the molar ratio of blocking rDNA 10-fold from the original protocol, to ensure an excess of blocking rDNA (Graham & Clark-Curtiss, 1999). With this modification, a pilot study confirmed that three rounds of SCOTS eliminated detectable rRNA-derived sequences from our cDNA pools (data not shown).

Each round of SCOTS used 1.2 μg biotinylated chromosomal DNA, 66 μg rRNA and 3 μg of the appropriate cDNA pool. In order to obtain sufficient material and to minimize bias introduced by individual PCRs, the first round of SCOTS for each cDNA pool consisted of 10 independent reactions that were pooled after PCR amplification. Two reactions were carried out and pooled in each of the subsequent rounds.

The tissue-derived and broth-derived cDNAs were each subjected separately to three rounds of SCOTS to generate four cDNA pools. The three tissue-derived cDNA pools were individually subjected to three rounds of competitive SCOTS in which the chromosomal DNA was prehybridized with the broth-derived cDNA pool to block capture of genes expressed both in vitro and in vivo. The final pool of selectively captured cDNAs was cloned, and the clones were screened for inserts on Luria–Bertani (LB) agar plates supplemented with ampicillin (50 μg ml⁻¹), X-Gal (40 μg ml⁻¹) and IPTG (20 mM). Individual clones containing inserts were picked into LB broth in 96-well plates. Following overnight culture at 37 °C, the plates were subjected to plasmid isolation and automated sequencing in a 96-well format, using universal primers (M13 forward and reverse), exactly as described previously (Manson et al., 2004). Three or four 96-well plates of clones were sequenced from each of the three cDNA libraries. Sequences of cloned inserts were compared with the H. ducreyi genome using the BLASTN algorithm.

Southern analysis. Inserts from cDNA clones were amplified by PCR using M13 primers, separated electrophoretically and transferred to nitrocellulose by capillary action. Normalized, tissue-derived and broth-derived cDNA pools were labelled using the PCR DIG Probe Synthesis kit (Roche) for use as probes. Blots were hybridized with the labelled probes and developed using the DIG Easy Hyb system (Roche).

RT-PCR. RNA was extracted from pustules as described above or as described previously (Throm & Spinola, 2001). RT-PCR was performed to amplify H. ducreyi genes shown previously to be expressed in vivo (Throm & Spinola, 2001) or on the genes listed in Table 1. Each PCR reaction was performed alongside controls that received the same amount of RNA template but no reverse transcriptase; all such controls were negative. Each reaction was performed once on each tissue indicated in Table 1. Primers are listed in Supplementary Table S2 available with the online version of this paper.

RESULTS AND DISCUSSION

RNA isolation

RNA was obtained from pustules of three volunteers (subjects 238, 240 and 250) who were experimentally infected with H. ducreyi (Supplementary Table S1). Pustules contain necrotic cells that may release RNases; RNA later was required to stabilize the specimens. The integrity of RNA in the samples was determined by analysis with an Agilent Bioanalyser, which provides a spectrophotometric tracing and a virtual gel image of the RNA preparation (Fig. 1). The eukaryotic RNA was intact (Fig. 1), but the prokaryotic RNA levels were too low to visualize in these samples. Amplification of mRNA from several H. ducreyi genes, including cdhB, losB, dsrA and lspA1, by RT-PCR was successful (data not shown),

### Table 1. Verification of in vivo gene transcription

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Annotation</th>
<th>Tissues in which cDNA or mRNA was detected (not detected) by:</th>
<th>Overall results*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SCOTS</td>
<td>RT-PCR</td>
</tr>
<tr>
<td>HD0192</td>
<td>Hypothetical lipoprotein</td>
<td>238† (240, 250)</td>
<td>231, 276, 281</td>
</tr>
<tr>
<td>HD0286</td>
<td>Hypothetical protein</td>
<td>240 (238, 250)</td>
<td>160, 164, 276</td>
</tr>
<tr>
<td>HD0646</td>
<td>Conserved hypothetical protein</td>
<td>238, 240 (250)</td>
<td>142, 231, 276</td>
</tr>
<tr>
<td>HD0805</td>
<td>Conserved hypothetical protein</td>
<td>238 (240, 250)</td>
<td>160, 164, 276</td>
</tr>
<tr>
<td>HD1170</td>
<td>Outer-membrane protein P4</td>
<td>250 (238, 240)</td>
<td>231, 276, 281</td>
</tr>
<tr>
<td>HD1280</td>
<td>Possible serine protease homologue</td>
<td>238 (240, 250)</td>
<td>231 (164, 281)</td>
</tr>
<tr>
<td>HD1589</td>
<td>Lipoprotein NlpI homologue</td>
<td>238 (240, 250)</td>
<td>276, 281</td>
</tr>
<tr>
<td>HD1629</td>
<td>Outer-membrane lipoprotein LolB</td>
<td>250 (238, 240)</td>
<td>249, 252 (231, 276, 281)</td>
</tr>
<tr>
<td>HD1808</td>
<td>HIC protein</td>
<td>238 (240, 250)</td>
<td>281 (231)</td>
</tr>
<tr>
<td>HD1829</td>
<td>Probable outer-membrane protein</td>
<td>238 (240, 250)</td>
<td>231 (276, 281)</td>
</tr>
</tbody>
</table>

*Number of tissues in which the transcript was detected/total number of tissues examined.
†Subject number from which tissue sample was obtained.
suggesting that the prokaryotic RNA was also intact. Intact RNA was also derived from a broth culture of 35000 HP that was used to infect volunteers (Fig. 1). The four samples were independently subjected to SCOTS to produce four cDNA pools. The in vitro-derived cDNA pool was then used to block capture of each in vivo-derived cDNA pool in competitive SCOTS. Competitive SCOTS theoretically identifies transcripts that may be exclusively expressed or more abundant in one condition than in another. Increased transcript abundance may be due to higher rates of transcription, variations in mRNA degradation, and quality of the RNA samples. Thus, our results were not likely to be biased by differences in the quality of RNA among the samples.

**Sequence analysis of clones obtained after competitive SCOTS**

We sequenced inserts of 924 clones obtained after competitive SCOTS. All inserts contained *H. ducreyi* DNA. Only one insert was derived from *H. ducreyi* rRNA, indicating a blocking efficiency of >99%. DNA in eight inserts was derived from tRNAs. A few clones contained sequences that mapped to intergenic regions or pseudogenes; the remaining clones had inserts containing DNA homologous to known or putative ORFs in the *H. ducreyi* genome.

The 924 clones contained sequences from 531 genes, representing nearly one-third of the genome. We were somewhat surprised by the large number of captured transcripts, as *H. ducreyi* only infects humans and has no other known environmental niche, and presumably would not require the complex regulation observed in bacteria that infect humans and other species or occupy environmental niches. However, within the human host, *H. ducreyi* infects a variety of tissues, including stratified squamous or mucosal epithelium and lymph nodes, external sites such as the foreskin and labia, and internal sites such as the vagina and cervix. The cutaneous immune response to the organism is complex and consists of serum components, polymorphonuclear neutrophils, macrophages, T cells and dendritic cells. In addition, *H. ducreyi* infects men and women, who differ in their susceptibility to infection (Bong et al., 2002b; Spinola et al., 2003a). The *H. ducreyi* genome contains at least two intact systems that could sense extracytoplasmic stress; specifically, the Cpx two-component system and the sigma E (RpoE) system. Thus, *H. ducreyi* likely regulates its gene expression in response to different host pressures and microenvironments, as has been shown for other strict human pathogens such as *Bordetella pertussis*, which utilizes the bvg two-component system to regulate virulence genes.

**Confirmation that the genes identified by SCOTS are expressed in vivo**

To confirm that the identified genes were expressed *in vivo*, 60 clones from each tissue were analysed by Southern blotting, which has been used as a confirmatory test in most other SCOTS-based studies (Baltes & Gerlach, 2004; Daigle et al., 2001; Graham & Clark-Curtiss, 1999; Haydel et al., 2002; Hou et al., 2002; Morrow et al., 1999).

Duplicate blots of cloned inserts were probed with normalized cDNA pools from tissues or broth cultures after three rounds of SCOTS, and the signals from the two probes were compared. Representative results are shown in Fig. 2. Of 180 probed clones, 87% hybridized with the normalized tissue-derived *H. ducreyi* cDNA but either did not hybridize with the normalized broth-derived *H. ducreyi* cDNA (Fig. 2, lanes 1, 2 and 4–6) or showed a much stronger signal with the tissue-derived cDNA than with the broth-derived cDNA (Fig. 2, lane 7). These data show that the competitive SCOTS procedure preferentially captured sequences that were present in the *in vivo*-derived cDNA pool. Of the clones that did not show a stronger signal with...
the tissue probe, only one (Fig. 2, lane 8) showed a similar signal with both probes, while the remainder showed no signal with either probe and were therefore inconclusive (Fig. 2, lane 3).

To further confirm the SCOTS data, we also performed RT-PCR on RNA isolated from *H. ducreyi*-infected pustules obtained from nine additional volunteers (Supplementary Table S1). Ten of 11 (91%) of the genes captured by SCOTS were amplified from one or more additional tissues (Table 1). The one gene not detected by RT-PCR was identified in two of the subjects who provided tissue for SCOTS. The data suggest that genes identified by SCOTS are transcribed in multiple volunteers.

We considered but could not use quantitative real-time RT-PCR (qRT-PCR) to confirm our findings. qRT-PCR requires the ability to normalize results from each sample to an objective reference reflecting total bacterial genomes or RNA levels, or to a gene constitutively expressed under the conditions being compared (Shelburne & Musser, 2004). Our in vivo RNA is almost entirely eukaryotic, while our in vitro RNA is entirely prokaryotic RNA. RNAlater was required to stabilize the tissue samples, which were processed entirely for RNA. Thus, we have no information about the number of bacteria or bacterial genomes in the samples. Unfortunately, multiple attempts to process biopsies simultaneously for DNA and RNA so that we could quantify bacterial genomes yielded degraded RNA. The identity of genes constitutively expressed by *H. ducreyi* in vitro and in vivo is unknown. In the absence of this information, 16S rRNA is commonly used to normalize bacterial RNA samples. However, 16S rRNA levels are highly dependent on growth conditions and nutrient availability (Gourse et al., 1996). The doubling time of 35 000 HP in broth is approximately 90 min, while the estimated doubling time of 35 000 HP in lesions is 16.5 ± 3.8 h (Throm & Spinola, 2001). Thus, we cannot assume that the 16S rRNA content of the bacteria grown in vitro is similar to that of the bacteria in a pustule. Although the chief limitation of SCOTS is that it is qualitative and we did not prove that the captured transcripts were upregulated in vivo, the captured genes must be expressed in vivo.

Redundancy of gene expression among volunteers

Fig. 3 shows the distribution of the 531 identified genes among the three volunteers. Although the majority of transcripts were identified in cDNA from only one volunteer, 139 transcripts were identified in at least two volunteers, and 28 transcripts were identified as being expressed in all three volunteers (Fig. 3). Analysis of the functional categories of the identified genes demonstrated similar patterns of gene expression in all three volunteers (Table 2).

With the exception of microarray analysis, none of the published in vivo expression procedures, including SCOTS, as well as promoter-trap and mutagenic strategies, is comprehensive for the entire genome (Rediers et al., 2005; Shelburne & Musser, 2004). Data from SCOTS are limited to the number of clones sequenced. Thus, we do not know whether 139 genes represents the true level of overlap of gene expression from host to host, or whether analysis of additional clones would have yielded greater overlap. RT-PCR of additional tissues showed that many transcripts captured in only one tissue in SCOTS were present in additional volunteers, suggesting that we would have found more overlap had we sequenced more clones.

Analysis of genes expressed in at least two volunteers

Of the 139 genes identified in at least two tissues (see Supplementary Table S3), 68 of 74 (92%) tested hybridized to a greater extent with tissue-derived than with broth-derived cDNA. Although the RT-PCR data suggested that there may be greater overlap among the three volunteers, observed host effects on *H. ducreyi* infection indicate that some of the detected genes could be expressed only in a subset of volunteers (Bong et al., 2002b; Spinola et al., 2003a). We therefore conservatively confined our examination of the potential functional significance of in vivo expressed genes to those from more than one volunteer.

The 139 genes identified in at least two volunteers fell into a variety of functional categories, based on homology with previously characterized bacterial genes (Table 2, final column). The largest group was the broad category of biosynthetic and metabolic genes. There were also genes involved in a number of cellular processes, including cell
division, DNA replication and repair, transcription, and translation. Some identified genes, including those encoding heat-shock proteins or transporters, can play important roles in responses to stress or nutrient deprivation. Additionally, two regulatory genes were expressed in vivo, including the sensor kinase of a two-component regulator. We also identified 32 genes encoding hypothetical proteins. Homologues of at least 21 of the genes (shown in bold type in Supplementary Table S3) have been identified by promoter-trap strategies as in vivo-expressed during infections caused by other pathogens (Rediers et al., 2005). Overall, these data are consistent with in vivo bacterial gene induction analyses, in which many categories of genes are affected in addition to traditional virulence factors (Rediers et al., 2005; Shelburne & Musser, 2004), and suggest that H. ducreyi responds to the host in a global fashion for in vivo survival.

Six identified electron transport genes are involved in anaerobic respiration in other bacterial systems, including nitrate reductases napA and nirA, 1-lactate dehydrogenase lldD, anaerobic glycerol-3-phosphate dehydrogenase subunits glpB and glpC, and torY, a c-type cytochrome involved in trimethylamine N-oxide reduction. Similarly, two metabolic genes, pflB encoding pyruvate formate-lyase and ansB encoding L-asparaginase II, are homologous to genes that are induced under anaerobic conditions in E. coli (Jennings & Beacham, 1990; Sawers & Bock, 1988). In contrast, we also identified atpA, an aerobic electron transfer chain subunit, suggesting that some aerobic respiration may also occur. Typically, the environment within pustules is anaerobic (Hays & Mandell, 1974). However, H. ducreyi cells are found both deep within the pustule and at the pustule surface, where the environment may be more aerobic (Bauer & Spinola, 2000; Bauer et al., 2001). Expression of both aerobic and anaerobic respiration genes could reflect differential responses within the bacterial population to microenvironments with different oxygenation levels in vivo.

Six genes identified in this study are involved in amino acid biosynthesis, including argA, argC and argE from the arginine biosynthetic pathway, and carB, which is involved in synthesis of arginine and pyrimidines. Arginine synthesis genes have been identified during in vivo expression studies of several other bacterial pathogens (Rediers et al., 2005). Notably, homologues of argA and carAB were identified during in vivo expression studies in Vibrio cholerae and Salmonella enterica serovar Typhimurium, respectively (Camilli & Mekalanos, 1995; Mahan et al., 1993). Mutational analysis in these pathogens has demonstrated that argA and carAB are required for virulence of V. cholerae and S. enterica, respectively, in mouse models of infection (Camilli & Mekalanos, 1995; Mahan et al., 1993). These data suggest that in vivo, H. ducreyi, among other pathogens, may require increased arginine, which is a semi-essential amino acid for mammals and thus likely to be in limited supply in tissues. In addition to its utility as an amino acid, arginine is a substrate in other synthesis pathways, such as polyamine production (Cunin et al., 1986). Possibly, one or more arginine-requiring biosynthetic pathway(s) is induced in vivo, leading to an increased arginine requirement.

Eight genes identified in vivo are involved in DNA repair, including mutL, mutS, radA, recB, recD, recG, ung and uvrA, suggesting that H. ducreyi encounters DNA-damaging stresses during human infection. Six of the identified

---

**Table 2. Functional categories of in vivo expressed genes**

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of genes identified in:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tissue 238</td>
</tr>
<tr>
<td>Adherence-related genes</td>
<td>2</td>
</tr>
<tr>
<td>Bacteriophage genes</td>
<td>8</td>
</tr>
<tr>
<td>Biosynthesis/metabolism</td>
<td>76</td>
</tr>
<tr>
<td>Cell division</td>
<td>5</td>
</tr>
<tr>
<td>DNA replication/repair</td>
<td>10</td>
</tr>
<tr>
<td>Electron transport</td>
<td>7</td>
</tr>
<tr>
<td>Hypothetical proteins</td>
<td>57</td>
</tr>
<tr>
<td>Outer-membrane proteins</td>
<td>3</td>
</tr>
<tr>
<td>Regulation</td>
<td>5</td>
</tr>
<tr>
<td>Ribosomal proteins</td>
<td>8</td>
</tr>
<tr>
<td>Secreted proteins</td>
<td>1</td>
</tr>
<tr>
<td>Stress response</td>
<td>5</td>
</tr>
<tr>
<td>Toxins</td>
<td>0</td>
</tr>
<tr>
<td>Transcription</td>
<td>4</td>
</tr>
<tr>
<td>Translation</td>
<td>11</td>
</tr>
<tr>
<td>Transport/uptake</td>
<td>15</td>
</tr>
<tr>
<td>Total number of genes</td>
<td>217</td>
</tr>
</tbody>
</table>
DNA repair genes were also identified during in vivo expression studies with other pathogens (Rediers et al., 2005). Additionally, the recBCD-encoded exonuclease V is required for repair of double-stranded DNA breaks and has been shown to be necessary in S. enterica for both survival in macrophages and virulence in mice (Cano et al., 2002). Thus, DNA repair could be an important mechanism for H. ducreyi survival in vivo.

In addition to genes involved in various cellular processes or regulation, we identified several genes that encode bacterial components that have been characterized by our laboratory and others as potential virulence determinants, including the large supernatant proteins LspA1 and LspA2, the cytotoxid distending toxin subunit CdtB, tight adherence protein TadA, and outer-membrane proteins, including MOMP, predicted outer-membrane protein HD1655, predicted outer-membrane lipoprotein HD1094, and a homologue of Haemophilus influenzae outer-membrane protein OMP P1 (Table 2 and Supplementary Table S3). The LspA proteins LspA1 and LspA2 share 86% sequence identity and confer antiphagocytic activity to the organism (Vakevainen et al., 2003). We identified clones that contained inserts corresponding to identical regions shared by LspA1 and LspA2. Although expression of MOMP and CdtC is not required for pustule formation in humans (Throm et al., 2000; Young et al., 2001), expression of LspA1 and LspA2 and the TadA-containing operon is required for full virulence in the human model of H. ducreyi infection (Janowicz et al., 2004; Spinola et al., 2003b). Prompted by the findings from this study, we constructed mutants in HD1844, which encodes WecA, the first enzyme in the enterobacterial common antigen pathway, and in HD0192 (Table 1), which encodes a hypothetical lipoprotein. Both mutants were partially attenuated for pustule formation in humans (K. E. Banks and others, unpublished observations; D. M. Janowicz and others, unpublished observations). Thus, several captured transcripts were relevant to pathogenesis. Identification of momp, lspa1 or lspa2, cdta, and tadA also showed that SCOTS detected some genes expressed by H. ducreyi in vitro (Nika et al., 2002; Stevens et al., 1999; Throm & Spinola, 2001; Ward et al., 1998). Southern blot analysis confirmed that momp, lspa2 and tadA were more abundant in the in vivo-derived cDNA pool (data not shown), suggesting that the expression of these genes is upregulated in vivo.

HD1895, a gene encoding a Hmwc-like protein, was identified in this screen. In H. influenzae, Hmwc is involved in glycosylation of the high-molecular-weight (HMW) adhesins, and a copy of this gene is present in both HMW gene clusters in H. influenzae (Grass et al., 2003; St. Gme & Grass, 1998). The N-terminal portions of LspA1 and LspA2 share homology with HMW1A and HMW2A (Ward et al., 1998), but there is no evidence of glycosylation of the LspA proteins, and the genes up- and downstream of HD1895 are not likely to encode surface-associated proteins. Thus, the substrate for the putative glycosyl transferase Hmwc is unknown.

Several genes encoding enzymes involved in lipooligosaccharide (LOS) biosynthesis were identified, including galE, kidK, rsaE and lpxC. Sialic acid is a component of the LOS of H. ducreyi. HD1669 and HD1670 were identified in the screen and have recently been shown to encode components of an ABC transporter that transports sialic acid (Post et al., 2005). HD1842 and HD1844 also encode proteins that are involved in complex carbohydrate biosynthesis. These genes are part of a gene cluster that has strong homology to genes involved in enterobacterial common antigen synthesis.

H. ducreyi has three Mu-like bacteriophage gene clusters. Genes in these regions have not been characterized, and no bacteriophage production or activity has been observed in H. ducreyi. However, seven genes from two of these clusters were identified in our study (Supplementary Table S3).

Our results were similar to those reported in a recent SCOTS-based study of in vivo gene expression by Actinobacillus pleuropneumoniae in a porcine model of infection (Baltes & Gerlach, 2004). A total of 46 A. pleuropneumoniae genes expressed in lung specimens pooled from four pigs were identified. The H. ducreyi genome contains homologues of 28 of the 46 A. pleuropneumoniae genes identified in the porcine study, and 12 of the 28 homologues were also expressed in our three human volunteers in vivo. These homologues included a gene encoding an outer-membrane protein, six biosynthetic genes, two regulatory genes, a stress response gene and genes encoding two transporters. While H. ducreyi and A. pleuropneumoniae have different host ranges and cause different diseases, they are members of the Pasteurellaceae, and both are extracellular pathogens that rely on similar strategies such as resistance to phagocytosis for in vivo survival.

In summary, our data suggest that H. ducreyi gene expression undergoes multiple changes within the human host compared with in vitro broth culture. These data also indicate that one or more regulatory mechanisms of H. ducreyi functions in vivo, and define a number of candidates for genes important to in vivo survival or virulence of the organism, including several hypothetical genes as well as genes required for virulence in other bacteria. Future studies will include defining regulatory mechanisms, identifying host factors to which the organism responds and defining the roles in virulence of genes expressed in vivo.

ACKNOWLEDGEMENTS

This work was supported by Public Health Service grants AI27863 and AI31494 from the National Institute of Allergy and Infectious Diseases (NIAAD). The human challenge trials were supported by Public Health Service grant MO1RR00750 to the General Clinical Research Center at Indiana University. D. M. J. was supported by NIH grants T32 AI007637 and K08 AI074657 from NIAID. M. E. B. was partially supported by the Indiana Genomics Initiative (INGEN) of Indiana University, which is supported in part by Lilly Endowment.
We thank Josephine Clark-Curtiss for training on the SCOTS procedure, many helpful discussions, and critical review of the manuscript. We thank Joan Hou for training and trouble-shooting on the SCOTS procedure, Carisa Townsend for technical assistance, Genshi Zhao, James Graham and France Daigle for helpful discussions, and Tricia Humphreys, Barbara Van Der Pol and Byron Batteiger for critical review of the manuscript.

REFERENCES


Edited by: M. P. Stevens