The genome of *Treponema pallidum*: new light on the agent of syphilis

George M. Weinstock a,c,*, John M. Hardham b, Michael P. McLeod a, Erica J. Sodergren a, Steven J. Norris b,c

a Department of Microbiology and Molecular Genetics, University of Texas Medical School, 6431 Fannin Street, Houston, TX 77030, USA
b Department of Pathology and Laboratory Medicine, University of Texas Medical School, Houston, TX 77030, USA
c Center for the Study of Emerging and Re-emerging Pathogens, University of Texas Medical School, Houston, TX 77030, USA

Received 3 June 1998; received in revised form 25 August 1998; accepted 30 August 1998

Abstract

*Treponema pallidum* subsp. *pallidum*, the causative agent of the sexually transmitted disease syphilis, is a fastidious, microaerophilic obligate parasite of humans. This bacterium is one of the few prominent infectious agents that has not been cultured continuously in vitro and consequently relatively little is known about its virulence mechanisms at the molecular level. *T. pallidum* therefore represented an attractive candidate for genomic sequencing. The complete genome sequence of *T. pallidum* has now been completed and comprises 1 138 006 base pairs containing 1041 predicted protein coding sequences. An important goal of this project is to identify possible virulence factors. Analysis of the genome indicates a number of potential virulence factors including a family of 12 proteins related to the Msp protein of *Treponema denticola*, a number of putative hemolysins, as well as several other classes of proteins of interest. The results of this analysis are reviewed in this article and indicate the value of whole genome sequences for rapidly advancing knowledge of infectious agents.

Keywords: Molecular pathogenesis; Virulence factor; *Treponema pallidum*; Syphilis; Genome analysis

Contents

1. Introduction ........................................................................................................ 324
2. The DNA sequence of the *T. pallidum* genome .............................................. 325
   2.1. Overall characteristics of the sequence ..................................................... 325
   2.2. Methods of analysis ............................................................................... 325
3. Genes that might contribute to infection ............................................................ 326
   3.1. Virulence factors .................................................................................. 326
   3.2. *tpr* Genes: a treponeme-specific gene family ....................................... 326
   3.3. *tpr*-Associated open reading frames are also often treponeme-specific .................................................. 327
   3.4. Hemolysins ......................................................................................... 328

* Corresponding author. Tel.: +1 (713) 500-6083; Fax: +1 (713) 500-5499; E-mail: georgew@utmmg.med.uth.tmc.edu

0168-6445/98/$19.00 © 1998 Federation of European Microbiological Societies. Published by Elsevier Science B.V.
PII: S0168-6445(98)00017-5
1. Introduction

Syphilis was first recognized as a disease entity when it rapidly spread through Europe in the late fifteenth century, coinciding with the return of Columbus and his sailors from the New World. It was a classic example of an emerging infectious disease and became one of the most prevalent and devastating human infections in the world [1–3]. Early theories suggested that syphilis was one of the few epidemics to have spread from the New World to the Old, but this is now less certain. The disease quickly reached epidemic proportions in Europe and spread across the world during the 16th century with the age of exploration. Syphilis was ubiquitous by the 19th century, and has been called the AIDS of that era [4].

As the malady marched rapidly across Europe and around the world, it was called the French disease, Spanish disease, German disease, Polish disease, Portuguese disease, as well as other names, depending on one's point of view. As is often true for emerging infectious diseases, the initial version of syphilis that appeared in Europe was highly virulent and was often fatal in the early stages of infection. However, after a few decades the modern version of syphilis appeared, after acquiring the properties of a more chronic infection. Over the centuries, bizarre therapies, including oral administration of mercurial compounds and intentional inoculation of the patient with the malaria parasite to induce fever (a Nobel prize-winning therapy), were developed and used widely. The causative agent of syphilis, Treponema pallidum subsp. pallidum, was first identified by Schaudinn and Hoffmann in 1905 [5,6], and many of the leading scientists of that landmark era, including Elie Metchnikoff, Karl Landsteiner, and Paul Ehrlich, contributed to the increased understanding of this unusual, spiral-shaped bacterium. The term ‘magic bullet’, coined by Ehrlich for the compound arsphenamine, provided the first reasonably effective therapy for syphilis. But the disease persisted and peaked at over half a million reported new cases in the United States in 1943 before its rapid decline after the widespread availability of penicillin. Despite this decrease, syphilis and the related treponemal diseases yaws, endemic syphilis, and pinta still represent major world health problems. For example, over 134,000 cases of syphilis were reported in the U.S. in 1990 during a recent epidemic, including 2867 cases of the most devastating form, congenital syphilis [7].

T. pallidum, the causative agent of syphilis, is a spirochete, a phylogenetically ancient and distinct bacterial group. The bacterium has a helical or sinusoidal shape with outer and cytoplasmic membranes, a thin peptidoglycan layer, and flagella that lie in the periplasmic space and extend from both ends toward the middle of the organism. Multiple clinical stages separated by long periods of latent, asymptomatic infection characterize syphilitic T. pallidum infection. The primary infection is localized, but organisms rapidly disseminate and cause manifestations throughout the body, including the cardiovascular and nervous systems [8,9]. If untreated, infection can persist for decades, despite an active host immune response. T. pallidum would appear to exemplify an extreme in the range of invasive vs. toxigenic bacterial pathogens. T. pallidum is also unusual in terms of its degree of dependence on the host. It is an obligate parasite of humans and is one of the few medically important bacteria that has not been cultured continuously in vitro [10,11]. Limited multiplication can be obtained in a tissue culture system [12], but the standard means of propagating T. pallidum is through the intratesticular infection of rabbits. The inability to culture and hence clone the organism precludes most standard genetic approaches, includ-
ing mutagenesis and genetic transfer techniques. The fastidious nature of *T. pallidum* is most likely related to severe limitations in its metabolic capability [13].

Despite its importance as an infectious agent, relatively little is known about *T. pallidum* as compared to other bacterial pathogens [14]. Mechanisms of *T. pallidum* pathogenesis are poorly understood. No known virulence factors have been identified, and the outer membrane is mostly lipid with a paucity of proteins [15–17]. Consequently, existing diagnostic tests for syphilis are suboptimal and no vaccine against *T. pallidum* is available. Studies of this organism have clearly been held back because of its inability to be cultured continuously in vitro.

For these reasons, *T. pallidum* was an excellent candidate for genomic sequencing. Recently, the whole genome sequence was completed and the initial analysis of the sequence was presented [18]. In this review, we focus on analysis of the genomic sequence for virulence factors that are likely to lead to insights into infection. It is not the intent of this article to present rigorous scientific proof that these sequences are involved in infection. Rather, we present a speculative catalog of genes that are possibly important for infection. Much of this analysis is based on sequence similarity to known virulence functions. However, about one-third of the total predicted coding sequences bear no significant similarity to known genes, and thus what is described below is certainly an incomplete picture. Nevertheless, one of the goals of whole genome sequence projects is to identify important research possibilities, and this article aims to chronicle many of these.

2. The DNA sequence of the *T. pallidum* genome

2.1. Overall characteristics of the sequence

The genomic DNA sequence of *T. pallidum* subsp. *pallidum* (Nichols), as determined by the whole genome random sequencing method [19–24], comprises a circular chromosome of 1,138,006 bp with a G+C base composition of 52.8%. There are a total of 1,041 predicted ORFs, with an average size of 1,023 bp. The average size of these predicted proteins is 37,771 Da, ranging from 3,235 to 172,869 Da. The mean isoelectric point for the predicted proteins is 8.1, ranging from 3.9 to 12.3. These parameters are similar to those observed in other bacteria. These proteins are encoded by 92.9% of the genomic DNA. Biological roles have been suggested for 577 ORFs (55%) by the classification scheme of Riley [25], while 177 ORFs (17%) match hypothetical proteins from other species, and 287 ORFs (28%) have no database match and may be novel genes. When compared to another spirochete, *Borrelia burgdorferi*, whose genome has also been sequenced [24], 90 *T. pallidum* ORFs of unknown function match chromosome-encoded proteins in *B. burgdorferi*, but no *T. pallidum* ORFs match *B. burgdorferi* plasmid-encoded proteins, suggesting that the plasmid proteins are unique to *Borrelia*. The *T. pallidum* sequence and annotation information can be found at the web site for The Institute for Genomic Research at http://www.tigr.org/db/dmdb/tpdb/tpdb.html or the *Treponema pallidum* Molecular Genetics Server site at the University of Texas Medical School at Houston at http://utmmg.med.uth.tmc.edu/treponema/tpall.html.

All 61 triplet codons are used in *T. pallidum* genes, with a bias for G or C in the third codon position. This contrasts with the A or T bias in this position in *B. burgdorferi*. This observation is related to the higher G+C base composition in the *T. pallidum* genome, being almost twice that than in *B. burgdorferi*. The disparate G+C composition between the spirochete genomes is also related to a bias in overall codon usage, and a concomitant difference in amino acid composition in the predicted coding sequences.

Analysis of the predicted protein sequences indicates 129 of the ORFs (12%) can be assigned to 42 paralogous gene families. Among these, 15 families contain 44 genes that have no assigned biological role. The largest family, with 14 members, consists of ATP binding cassette proteins in ABC transport systems, while 30 families have only 2 members. Among 13 gene families are 16 clusters of adjacent genes that may represent duplications in the *T. pallidum* genome.

2.2. Methods of analysis

Following completion of the DNA sequence, coding regions were identified using GLIMMER [26]
and searched against a non-redundant database using the methods developed at TIGR. In addition, paralog families were analyzed using Pfam [27,28], membrane-spanning domains were predicted using TopPred [29], and signal peptides were predicted using Signal-P [30]. Although this procedure predicted the vast majority of ORFs, there may be a small number of genes that are not yet represented in the *T. pallidum* database, either because they are too small to have been considered or because they have unusual characteristics, for example different patterns of codon usage.

Subsequent to this analysis, a different search algorithm, PSI-BLAST [31], was used to search the database with each predicted ORF. In addition, searches of the BLOCKS [32] and ProDom [33] databases of protein domains as well as the COG database of orthologous groups of proteins [34] were performed. The results of these analyses of each putative ORF were used to make the predictions described in this review.

3. Genes that might contribute to infection

3.1. Virulence factors

Many genes are necessary for a microorganism to survive in a host. These include genes encoding intracellular proteins that are essential for cellular life in all situations, for example proteins needed for replication or gene expression, proteins necessary for the cell’s metabolism in the different environments in the host, regulatory proteins, as well as others. Besides these housekeeping functions are exported proteins required for metabolism (nutrient uptake for example) as well as interaction with the host. It is this latter group of molecules that we are concerned with in this review. These confer the pathogenic phenotype by allowing the microorganism to adhere to host tissue, invade new compartments, fight off or evade host responses, as well as other host-specific interactions.

A list of 67 genes that are candidates for this class of functions is given in Table 1 and their distribution around the chromosome is shown in Fig. 1. Note that there are three regions that appear to have a lower density of candidate genes. These regions are the locations for some of the larger gene clusters found on the chromosome. A ribosomal protein gene cluster and the two rRNA gene clusters are found in the 150–300 kb region, a cluster of genes involved in flagellum biosynthesis and another cluster involved in synthesis of a V-type ATPase is found in the 350–450 kb region, and another flagellar gene cluster is found in the 750–900 kb interval. This unequal distribution may reflect some aspect of chromosome evolution.

3.2. *tpr* Genes: a treponeme-specific gene family

Of great interest is the presence of a family of 12 related genes (paralogs) encoding predicted products with similarity to the major surface (or sheath) protein (Msp) of *Treponema denticola* [35]. In fact, this is the only entry in genomic databases that shows similarity to these 12 predicted products, including the genome of *B. burgdorferi*, and thus this seems to be a treponeme-specific gene family. These genes have been called *tpr* genes (*tprA*–*L*).

The *T. denticola* Msp is abundant, highly immunogenic [36–38], and forms a dense hexagonal array on the outer surface of the bacterium. Msp has been found to bind to fibronectin and laminin, and has porin-like activity [35–37]. Although a similar surface array has not been found on *T. pallidum*, it is tempting to speculate that the predicted Tpr proteins of *T. pallidum* are surface-localized and may represent some of the elusive outer membrane proteins of the organism. These putative membrane proteins may thus function as porins and adhesins. The fact that there are multiple versions of these genes in *T. pallidum* may reflect an antigen variation system, common to pathogenic borreliae, *Neisseria gonorrhoeae*, *Mycoplasma genitalium*, and many other pathogenic bacteria and protozoa. The extent of sequence similarity between various members ranges from complete identity throughout the whole gene to much more modest similarities. The similar regions do not always encompass the entire gene so that some regions are identical, but others can be highly variable.

The individual or coordinate expression and regulation of the *tpr* genes is under investigation. Preliminary findings indicate all genes are expressed and that there are upstream sequences that could be in-
involved in coordinate differential regulation (unpublished results). The tpr gene family in *T. pallidum* is reminiscent of a 32-member paralog family in *Helicobacter pylori* encoding outer membrane proteins (omp) [22]. The two gene families share features such as possible porin and adhesin functions. In addition, as in the *H. pylori* family, the *T. pallidum* tprA and tprF genes may contain frameshifts that could be corrected by slipped-strand mispairing during replication. Identification of the tpr family of putative outer membrane proteins is a major success of the genome project and may provide new targets for vaccine development.

### 3.3. tpr-Associated open reading frames are also often treponeme-specific

Because of the prominence of the tpr genes as candidates for virulence factors, the genes neighboring the tpr loci are also of interest. Surprisingly, the

---

**Table 1: Possible virulence functions of *Treponema pallidum***

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
<th>Start coordinate</th>
<th>Stop coordinate</th>
<th>Number</th>
<th>Name</th>
<th>Start coordinate</th>
<th>Stop coordinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP0009</td>
<td>tprA</td>
<td>10 164</td>
<td>8 343</td>
<td>TP0006</td>
<td>Tp75</td>
<td>7 014</td>
<td>7 178</td>
</tr>
<tr>
<td>TP0011</td>
<td>tprB</td>
<td>10 396</td>
<td>12 375</td>
<td>TP0020</td>
<td>76K</td>
<td>22 046</td>
<td>24 166</td>
</tr>
<tr>
<td>TP0117</td>
<td>tprC</td>
<td>136 697</td>
<td>134 904</td>
<td>TP0034</td>
<td>adhB</td>
<td>42 739</td>
<td>41 792</td>
</tr>
<tr>
<td>TP0131</td>
<td>tprD</td>
<td>152 897</td>
<td>151 104</td>
<td>TP0163</td>
<td>troA</td>
<td>184 611</td>
<td>185 534</td>
</tr>
<tr>
<td>TP0313</td>
<td>tprE</td>
<td>327 985</td>
<td>330 270</td>
<td>TP0171</td>
<td>tpp15</td>
<td>190 994</td>
<td>191 419</td>
</tr>
<tr>
<td>TP0316</td>
<td>tprF</td>
<td>332 334</td>
<td>331 143</td>
<td>TP0225</td>
<td>lrr</td>
<td>229 177</td>
<td>229 914</td>
</tr>
<tr>
<td>TP0317</td>
<td>tprG</td>
<td>334 663</td>
<td>332 396</td>
<td>TP0292</td>
<td>ompA</td>
<td>305 554</td>
<td>306 804</td>
</tr>
<tr>
<td>TP0610</td>
<td>tprH</td>
<td>661 246</td>
<td>663 324</td>
<td>TP0298</td>
<td>TpN38</td>
<td>311 131</td>
<td>312 159</td>
</tr>
<tr>
<td>TP0620</td>
<td>tprI</td>
<td>672 887</td>
<td>671 061</td>
<td>TP0319</td>
<td>tmpC</td>
<td>334 823</td>
<td>335 881</td>
</tr>
<tr>
<td>TP0621</td>
<td>tprJ</td>
<td>675 221</td>
<td>672 948</td>
<td>TP0326</td>
<td>Omp</td>
<td>344 276</td>
<td>346 834</td>
</tr>
<tr>
<td>TP0897</td>
<td>tprK</td>
<td>975 828</td>
<td>974 314</td>
<td>TP0327</td>
<td>ompH</td>
<td>346 894</td>
<td>347 409</td>
</tr>
<tr>
<td>TP1031</td>
<td>tprL</td>
<td>1 124 349</td>
<td>1 125 890</td>
<td>TP0435</td>
<td>tpp17</td>
<td>462 495</td>
<td>462 028</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0470</td>
<td></td>
<td>498 263</td>
<td>497 157</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0486</td>
<td></td>
<td>p83/100h</td>
<td>518 980</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0567</td>
<td></td>
<td>22.5kh</td>
<td>616 337</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0571</td>
<td></td>
<td>lemA</td>
<td>621 056</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0574</td>
<td></td>
<td>lag</td>
<td>623 570</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0624</td>
<td></td>
<td>Omp</td>
<td>678 822</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0702</td>
<td></td>
<td>nlpD</td>
<td>767 875</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0729</td>
<td></td>
<td>tap1</td>
<td>795 588</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0768</td>
<td></td>
<td>tmpA</td>
<td>833 922</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0769</td>
<td></td>
<td>tmpB</td>
<td>834 956</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0796</td>
<td></td>
<td>Lp</td>
<td>863 166</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0819</td>
<td></td>
<td>Lp</td>
<td>887 996</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0821</td>
<td></td>
<td>tsp2</td>
<td>889 699</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0957</td>
<td></td>
<td>Tp33h</td>
<td>1 038 069</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0971</td>
<td></td>
<td>tpd</td>
<td>1 054 742</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0989</td>
<td></td>
<td>P26h</td>
<td>1 073 472</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0993</td>
<td></td>
<td>rlpA</td>
<td>1 078 255</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP1016</td>
<td></td>
<td>tpa39b</td>
<td>1 107 859</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP1038</td>
<td></td>
<td>tpf1</td>
<td>1 135 337</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0038</td>
<td>pfoS/R</td>
<td>46 706</td>
<td>45 657</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0454</td>
<td>regA</td>
<td>484 019</td>
<td>483 333</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0516</td>
<td>mviN</td>
<td>557 907</td>
<td>556 330</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0520</td>
<td>regB</td>
<td>560 541</td>
<td>559 168</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0520</td>
<td>senB</td>
<td>561 739</td>
<td>560 554</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0877</td>
<td>regC</td>
<td>953 655</td>
<td>954 749</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0980</td>
<td>regD</td>
<td>1 063 017</td>
<td>1 064 036</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0981</td>
<td>regD</td>
<td>1 064 099</td>
<td>1 065 259</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0077</td>
<td>cap</td>
<td>84 254</td>
<td>85 867</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0078</td>
<td>spsC</td>
<td>85 875</td>
<td>87 110</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0107</td>
<td>licC</td>
<td>121 921</td>
<td>120 347</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0283</td>
<td>kdtB</td>
<td>297 994</td>
<td>298 470</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0288</td>
<td>spsF</td>
<td>301 183</td>
<td>302 127</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0440</td>
<td>spsA</td>
<td>465 732</td>
<td>466 880</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP0562</td>
<td>spsE</td>
<td>609 512</td>
<td>610 645</td>
</tr>
</tbody>
</table>

---

**G.M. Weinstock et al. / FEMS Microbiology Reviews 22 (1998) 323–332**

---

Downloaded from https://academic.oup.com/femsre/article-abstract/22/4/323/593502 by guest on 13 January 2018
tpr genes are generally surrounded by predicted ORFs that do not bear any similarity to genes in databases, thus providing little clues as to associated functions. Not only are the tpr genes treponeme-specific, but this distinction applies to the neighboring genes as well. Among the neighbors of tpr genes are several paralog gene families, however. This suggests that these genes encode functions that may be functionally important for the Tpr proteins.

3.4. Hemolysins

T. pallidum is not generally thought of as being toxigenic, and has not previously been found to pro-
duce either lipopolysaccharide or exotoxins. Cytotoxicity against neuroblasts and other cell types was observed at high concentrations of the bacterium [39–41]. Nevertheless, five genes encoding proteins similar to bacterial hemolysins were identified in the genome. One of these resembles hemolysin III of Bacillus cereus [42], and shares similarity with other members of this family from other bacteria. The other four genes, which are related to each other, show sequence similarity to the hlyC hemolysin from Serpulina hyodysenteriae [43], a spirochete that is an important pathogen of swine. In the case of the B. cereus hemolysin, the recombinant protein produced in Escherichia coli has been shown to have pore-forming hemolytic activity [44]. On the other hand, the hemolytic phenotype of the S. hyodysenteriae gene was also observed with a gene that was cloned and expressed in E. coli, but the activity of the protein product from this gene has not been demonstrated. Thus it is necessary to verify that the T. pallidum proteins are in fact cytolytic before this function can be assigned rigorously.

3.5. Regulatory systems may be scarce

There are virtually no previous studies on the regulation of T. pallidum gene expression due to the lack of genetic manipulation of this system. Inspection of the DNA sequence indicates the possibility of as many as five two-component regulatory systems, which would be a slightly lower density of such regulators than is found in larger genomes, such as E. coli or B. subtilis. The degree of similarity of some of these genes to regulatory or sensory proteins is not high, so it is likely that T. pallidum has relatively few of these regulatory systems. In addition, T. pallidum has very few predicted proteins that show similarity to classical repressor or activator protein families. Those that are found appear to be involved in regulating metabolic functions, such as a cyclic AMP binding protein or the troR repressor, controlling a transport operon. Thus there appear to be few proteins that could be involved in virulence gene regulation. Outside of the possible two-component systems, there is a homolog of the mviN virulence regulator for regulation of virulence genes [45]. Homologs to mviN have been found, often by genome projects, in Haemophilus influenzae, Vibrio cholerae, Salmonella typhimurium, Chlamydia trachomatis, B. burgdorferi, E. coli, H. pylori, and Mycobacterium tuberculosis. Thus this protein, which affects virulence in mouse models, is of general interest.

Surprisingly, T. pallidum encodes six genes that are homologous to sigma factors, which is a higher density than found in the larger E. coli genome. In addition there are a number of proteins that are similar to factors involved in controlling sigma factor activity or in transcription termination control. These general observations suggest control of virulence gene expression in T. pallidum may use different strategies than found in E. coli and its relatives.

3.6. Only a few possible genes for polysaccharide biosynthesis

The most important non-protein molecules for virulence are various types of polysaccharides. Lipopolysaccharide has many important properties, including activation of host defense systems. Capsules are made of exopolysaccharides that protect the cell from host response systems and can also play a role in other processes, such as adhesion. Often the genes for the synthesis of such polysaccharides are clustered in large units. However, no gene cluster with homology to polysaccharide biosynthesis functions was detected in the T. pallidum genome. A few scattered genes were identified by homology to functions in other organisms, principally spore coat polysaccharide biosynthesis in B. subtilis. The significance of this finding is not clear.

3.7. Few surface proteins

Considerable effort has been devoted over the years to the isolation of outer membrane and other surface proteins [14]. However, this has been a difficult task and T. pallidum has earned a reputation as a ‘stealth’ pathogen because of the apparent paucity of surface proteins. This has raised the possibility that the lack of surface antigens may be an important strategy in T. pallidum infection. Indeed, the outer membrane of T. pallidum shows relatively few membrane proteins in freeze fracture studies [15–17], suggesting this is a feature that helps the organism evade the immune response. Eighteen proteins that have been previously suggested as surface located (at
times the exact surface is controversial) are shown on the map. Inspection of the genomic sequence suggests another 13 possible surface localized proteins, not counting the 12 Tpr proteins, putative sensors of two-component regulators, and hemolysins. In addition, a number of other putative proteins (not shown on the map), that do not show similarities to database sequences, are predicted to contain membrane spanning regions and are likely surface localized. Thus, the number of surface proteins should more than double as a result of the genomic sequence.

3.8. Metabolic functions

There are many other functions that play a role in cell survival during infection, and some of these are involved in metabolic activities of the cell. Although these are not noted on the map, it is likely that some of these will be surface localized. These include both transport systems as well as enzymes, such as glycerophosphodiester phosphodiesterase, which is surface localized [46]. These proteins may provide good targets for vaccines.

3.9. Miscellaneous functions that might interact with the host

This group of proteins includes putative functions that have some characteristics that are suggestive of interaction with the host. For example, the gcp gene encodes a putative neutral metalloprotease that specifically cleaves O-sialoglycoproteins, such as glycophorin A. This sialoglycoprotease is similar in sequence to related proteins from many bacteria. In Pastewella haemolytica, where it has been best studied, the enzyme is secreted into the medium and thus appears targeted against host glycoproteins [47,48]. Somewhat more speculatively are the ankA and ankB genes, two paralogs that contain sequences similar to those found in mammalian ankyrin 3, a protein interacting with the cytoskeleton [49,50]. Finally there is the iev function, whose sequence suggests it is an integral membrane protein. It shows a region of sequence similarity to a viral protein that may play an immunoevasive role in the pathogenesis of Marek’s disease. It is a candidate for causing the early stage immunosuppression that occurs after MDHV infection.

4. Conclusions

*T. pallidum* has been a major pathogen of the civilized world for over 500 years. It has been one of the more refractory organisms to study and, in fact, was only identified in the early part of this century. However, as a result of the completion of the genomic sequence, there is now a wealth of leads to pursue to understand, diagnose, and treat syphilis. In this review, we have described a collection of 67 proteins that are of interest for future studies of *T. pallidum* virulence. Less than one-third of these had previously been noted and among the previously uncharacterized genes is the *tpr* gene family that is likely to play an important role in treponemal infections. Our future understanding of *T. pallidum*, as well as many other microorganisms, pathogenic or otherwise, is being profoundly altered by the availability of whole genome sequences.

Acknowledgments

We thank Dr. Claire Fraser and the staff at TIGR for their work on determining the *T. pallidum* genome sequence and providing an initial analysis, and Dr. Gerry Myers and Tom Brettin at Los Alamos National Laboratory for their efforts on subsequent annotation and analysis. This work was supported by NIH Grant AI31068 to G.M.W. MPM was supported in part by the graduate research associate program at LANL.

References


