Faculteit Farmaceutische, Biomedische en Diergeneeskundige Wetenschappen
Departement Biomedische Wetenschappen

Contemporary syphilis epidemics: efforts to improve syphilis diagnostics

---

Hedendaagse syfilis epidemieën: inspanningen om de diagnostiek van syfilis te verbeteren

Proefschrift voorgelegd tot het behalen van de graad van doctor in de biomedische wetenschappen aan de Universiteit Antwerpen
te verdedigen door

Kara K. Osbak

Promoteren:
Prof. dr. Chris R. Kenyon
Prof. dr. Xaveer van Ostade

Antwerpen, 2017

INSTITUTE OF TROPICAL MEDICINE ANTWERP

Universiteit Antwerpen

COLOFON

Contemporary syphilis epidemics: efforts to improve syphilis diagnostics. Academic Thesis, University of Antwerp, Antwerp, Belgium.

The research described in this thesis was performed at the Institute of Tropical Medicine Antwerp and at the laboratory for Protein Science, Proteomics and Epigenetic Signaling (PPES), University of Antwerp. Kara Osbak was funded by a SOFI-B research grant (Project ID: 757003) from the Flemish Government-Department of Economy, Science & Innovation granted to Chris Kenyon.

Front cover photograph: An electron photomicrograph of two spiral shaped *Treponema pallidum*, magnified 36,000x. (Photo by BSIP/UIG via Getty Images)

Doctoral Committee

Promoters:

Prof. dr. Chris R. Kenyon	Institute of Tropical Medicine Antwerp University of Cape Town
Prof. dr. Xaveer van Ostade	University of Antwerp

Chairman:

Prof. dr. Stuart Maudsley	University of Antwerp

Members:

Dr. Alje P. van Dam	Public Health Laboratory (GGD) Amsterdam
Prof. dr. Dieter Deforce	Ghent University
Prof. dr. Patrick D'Haese	University of Antwerp
Prof. dr. Erlangga Yusuf	University of Antwerp

Table of Contents

List of Abbreviations

ACN	Acetonitrile
AIDS	Acquired immune deficiency syndrome
ART	Antiretroviral therapy
bp	Base pair
BPG	Benzathine penicillin G
CDC	Centers for Disease Control and Prevention
CI	Confidence interval
CID	Collision induced dissociation
CSF	Cerebrospinal fluid
CT	*Chlamydia trachomatis*
DDA	Data dependent acquisition
DIA	Data independent acquisition
DNA	Deoxyribonucleic acid
DRT	Drug resistance testing
DTT	Dithiothreitol
ECDC	European Centre for Disease Prevention and Control
EDTA	Ethylenediamine tetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionisation
FA	Formic acid
FDR	False discovery rate
HIV	Human immunodeficiency virus
HPLC	High performance liquid chromatography
hPTPs	Isotopically labelled proteotypic surrogate peptides
Ig	Immunoglobulin
IHME	Institute for Health Metrics and Evaluation
IL	Interleukin
IQR	Interquartile range
IT	Ion trap
ITM	Institute of Tropical Medicine Antwerp
iTRAQ	Isobaric tags for relative and absolute quantification
IVDU	Intravenous drug user
LC	Liquid chromatography
LC-MS/MS	Liquid chromatography–tandem mass spectrometry
LOD	Limit of detection
LTIA	Latex turbidimetric immunoassay
m/z	Mass-to-charge-ratio

MALDI	Matrix assisted laser desorption/ionization
MRM	Multiple reaction monitoring
mRNA	Messenger RNA
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MTCT	Mother to child transmission
MudPIT	Multidimensional protein identification technology
ND	Not determined
NG	*Neisseria gonorrhoeae*
NSAF	Normalized spectral abundance factor
NTT	Nontreponemal test
OR	Odds ratio
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PMLR	Penalized Multinomial Logistic Regression
POC	Point of care
PrEP	Pre-exposure prophylaxis
PTM	Post-translational modification
PTP	Proteotypic peptide
Q	Quadrupole
RP	Reversed phase
RPR	Rapid plasma reagin test
RPR-C	Rapid plasma reagin card test
RPR-S	Sekure rapid plasma reagin test
RU	RPR units
SCX	Strong cation exchange
SRM	Selected reaction monitoring
STD	Sexually transmitted disease
STI	Sexually transmitted infection
TaSP	Treatment as prevention
TOF	Time of flight
TPPA	*Treponema pallidum* particle agglutination test
TRUST	Toludine Red Unheated Serum Test
TT	Treponemal test
VDRL	Venereal disease research laboratory test
WHO	World Health Organization
WIV-ISP	Scientific Institute of Public Health Belgium

Summary

Syphilis, a multistage curable chronic disease caused by the spirochete *Treponema pallidum* ssp. *pallidum* (*T. pallidum*), has re-emerged during the last 15 years as a major global public health problem, with an estimated 6 million new infections each year worldwide. Men who have sex with men (MSM) populations have been particularly affected, accounting for more than 90% of incident syphilis infections in Belgium, whereby half of these are reinfections. If not treated promptly, syphilis can cause serious morbidity and adverse pregnancy events such as stillbirth and congenital defects. Epidemiological studies investigating the relationship between syphilis and HIV on different levels can yield insights into possible underlying determinants, such as sexual network connectivity, that influence prevalence and incidence trends. Molecular phylogenetic approaches applied to sexually transmitted infection (STI) studies can also deepen our understanding of transmission dynamics.

Work presented in **Section I** of this thesis aimed to answer two important questions related to the relationship between syphilis and HIV. First, we assessed if there was a country-level association between antenatal syphilis prevalence from early in the HIV epidemics (1990-1999) and peak HIV prevalence. The purpose of this study was to help better elucidate the underlying determinants of variations in HIV spread. If evidence of a positive association would be found then this could be interpreted as evidence that the same underlying determinant(s) played a role in the spread of both HIV and syphilis and/or that syphilis was an important cofactor in the spread of HIV. Linear regression analyses of data from 76 countries revealed a strong association- syphilis prevalence in the 1990s predicted approximately 53% of the variation in peak HIV prevalence **(chapter 2)**. Second, we assessed if we could use the HIV phylogenetic tree constructed from HIV-1 sequences of 1169 patients in follow-up at a single-centre cohort in Antwerp, Belgium to see if there was evidence for clustering of syphilis infections. We found no evidence of clustering, however, analyses revealed potential cases of sexual identity misclassification of MSM as heterosexuals. We discuss the role these individuals may play as a high-risk bridge population **(chapter 3)**.

Diagnosis of *T. pallidum* (re)-infection and post-treatment follow-up to determine pathogen eradication remains onerous for clinicians due to inadequate assays based on century-old techniques. Therefore, the development of a diagnostic test that could directly detect *T. pallidum* antigens in human biofluids such as blood and urine would represent considerable progress in improving individual care and prevention efforts. Mass spectrometry (MS)-based protein biomarker discovery methods may hold the key to providing novel robust diagnostic targets.

Section II of this thesis details studies related to improving syphilis diagnostics by employing MS-based protein biomarker discovery techniques for antigen test development and a laboratory evaluation of an existent commercial nontreponemal syphilis assay. Since *T. pallidum* cannot be continuously cultured *in vitro*, bacteria were isolated from infected rabbits and purified to study the *T. pallidum* proteome during infection. Extracted proteins were trypsinized and subjected to multidimensional peptide separation and protein identification via matrix-assisted laser desorption ionization-time of flight (MALDI-TOF/TOF) and electrospray ionization (ESI-LTQ-Orbitrap) tandem mass spectrometry. These efforts resulted in the most extensive proteome investigation of *T. pallidum* to date; we were able to detect, characterize and semi-quantify 54% of the whole *T. pallidum* proteome (577 proteins), yielding novel insights into *T. pallidum* biology and potential biomarkers for diagnostic applications **(chapter 4)**. Eleven candidate biomarker proteins were then shortlisted from this analysis based on predicted high abundance, physiochemical properties and *T. pallidum* specificity. Isotopically labelled surrogate proteotypic peptides (N=30) corresponding to these proteins were synthesized and incorporated into a targeted multiple reaction monitoring assay conducted on a triple quadrupole mass spectrometer. Unfortunately, no endogenous *T. pallidum* peptide signals were detected in undepleted protein extracts from urine and plasma from individuals with syphilis. Further *T. pallidum* polyclonal antibody magnetic bead enrichment studies showed that the limit of detection of MS instruments we used was likely too high to uncover biomarker proteins in very low (femtomoles/liter) predicted concentrations without alternative sample preparation strategies such as antibody enrichment to increase sensitivity **(chapter 5).** Nevertheless, findings from this pioneering study could be useful to other researchers considering employing similarly challenging techniques. Lastly, in collaboration with an industry partner, an automated immunoturbidimetric rapid plasma reagin assay was evaluated for clinical appropriateness for syphilis diagnosis and post-treatment follow-up using serum samples collected during a two-year observational cohort study of syphilis patients (N=120) and controls (N=30) conducted at the Institute of Tropical Medicine in Antwerp, Belgium **(chapter 6).** Test performance deficiencies were highlighted, which in clinical practice could lead to missed syphilis diagnoses and suboptimal clinical follow-up. Independent and comprehensive assay evaluations using well-characterized clinical samples are essential to improving diagnostic methods.

In conclusion, (molecular) epidemiological studies of STIs can yield insights that could help inform prevention efforts. The application of proteomics approaches to study syphilis has broadened our understanding of the enigmatic spirochete *T. pallidum.* These findings could be useful for the development of a syphilis antigen test.

Samenvatting

Syfilis, een chronische aandoening die via verschillende stadia verloopt en wordt veroorzaakt door de spirocheet *Treponema pallidum* ssp. *pallidum* (*T. pallidum*), is de voorbije 15 jaar opnieuw opgekomen en blijft met 6 miljoen nieuwe infecties per jaar een wereldwijd gezondheidsprobleem. Populaties van mannen die seks hebben met mannen ("men who have sex with men" of MSM) werden deels getroffen en betreffen meer dan 90% van de huidige syfilis infecties in België, waarvan de helft re-infecties zijn. Indien syfillis niet onmiddellijk behandeld wordt kan het leiden tot hoge ziektecijfers en nadelige effecten op de zwangerschap leidend tot doodgeboorte en aangeboren afwijkingen. Epidemiologische studies die de relatie tussen syfilis en HIV op verschillende niveaus onderzoeken kunnen inzicht verschaffen in mogelijke onderliggende factoren zoals seksuele netwerken, die van invloed kunnen zijn op trends in prevalentie en incidentie. Eveneens kunnen moleculaire fylogenetische studies toegepast worden in seksueel overdraagbare infectie (SOA)-studies waardoor ze meer inzicht verschaffen in de transmissiedynamiek van deze ziekten.

Het werk voorgesteld in **Sectie I** van deze thesis heeft als doel twee belangrijke vragen in verband met de relatie tussen syfilis en HIV te beantwoorden. In de eerste plaats hebben we onderzocht of er een associatie op het niveau van de landen bestond tussen de prevalentie van antenatale syfilis vanaf het begin van de HIV-epidemie (1990-1999) en de piekprevalentie van HIV. Het doel van deze studie was om de onderliggende factoren die variaties in HIV-verspreiding veroorzaken, te bestuderen. Indien er een bewijs van een positieve associatie zou worden gevonden, kan dit geïnterpreteerd worden als bewijs dat dezelfde onderliggende determinant(en) een rol hebben gespeeld in de verspreiding van zowel HIV als syfilis en/of dat syfilis een belangrijke cofactor was in de verspreiding van HIV. Lineaire regressieanalyses van gegevens uit 76 landen lieten een sterke associatie zien: syfilisprevalentie voorspelde in de jaren negentig ongeveer 53% van de variatie in piek-HIV-prevalentie **(hoofdstuk 2)**. Daarna werd onderzocht of we de HIV-fylogenetische stamboom, opgebouwd uit HIV-1-sequenties van 1169 follow-up patiënten uit een single-center cohort in Antwerpen, België, konden gebruiken om na te gaan of er aanwijzing was voor clustering van syfilisinfecties. Hoewel voor een dergelijk clustering geen bewijs werd gevonden, toonden analyses aan dat mogelijke gevallen van misclassificatie voorkwamen van MSM als heteroseksuelen. De rol die deze individuen kunnen spelen als brugpopulatie met hoog risico **(hoofdstuk 3**) wordt tevens besproken.

De diagnose van *T. pallidum* (her)infectie en follow-up na de behandeling om de uitroeiing van pathogenen te bepalen, blijft nog steeds erg lastig voor clinici omwille van ontoereikende assays die gebaseerd zijn op eeuwenoude technieken.

Daarom zou de ontwikkeling van een diagnostische test die *T. pallidum*-antigenen direct in menselijke biovochten zoals bloed en urine kan detecteren, een aanzienlijke vooruitgang betekenen in het verbeteren van de individuele zorg en preventie-inspanningen. Massaspectrometrie (MS) -gebaseerde methoden voor eiwit biomerker identificatie kunnen daarbij aan de grondslag liggen voor het ontwikkelen van dergelijke nieuwe en robuuste diagnostische methoden.

In **sectie II** van dit proefschrift worden studies beschreven die betrekking hebben op het verbeteren van syfilisdiagnostiek door gebruik te maken van MS technieken voor ontwikkeling van antigeentesten en door een bestaande commerciële nontreponemal syfilis assay te evalueren. Omdat *T. pallidum in vitro* niet gekweekt kan worden, werden bacteriën geïsoleerd uit geïnfecteerde konijnen en vervolgens gezuiverd om het *T. pallidum* proteoom tijdens infectie te bestuderen. De geëxtraheerde eiwitten werden getrypsiniseerd en gescheiden door multidimensionale chromatografie waarna eiwitidentificatie door tandem massaspectrometrie plaatsvond via matrix-geassisteerde laser desorptie ionisatie (MALDI-TOF/TOF) en electrospray ionisatie (ESI-LTQ-Orbitrap). Dit werk resulteerde in het meest uitgebreide proteoomonderzoek van *T. pallidum* tot nu toe; we konden 54% van het gehele T. *pallidum* proteoom (577 eiwitten) detecteren, karakteriseren en semi-kwantificeren, hetgeen nieuwe inzichten opleverde in *T. pallidum* biologie en potentiële biomerkers voor diagnostische toepassingen **(hoofdstuk 4)**. Elf kandidaat biomerker eiwitten werden vervolgens uit deze analyse geselecteerd op basis van abundantie, fysicochemische eigenschappen en *T. pallidum*-specificiteit. Isotopisch gelabelde proteotypische peptiden (N=30) die representatief zijn voor deze eiwitten werden gesynthetiseerd en geanalyseerd in een gerichte *multiple reaction monitoring* (MRM) test welke werd uitgevoerd op een *triple quadrupole* massaspectrometer. Helaas werden er geen endogene *T. pallidum* peptidesignalen gedetecteerd in eiwitextracten uit urine en plasma van individuen met syfilis. Verdere aanrijking van *T. pallidum* eiwitten door middel van polyklonale antilichamen, gekoppeld aan magnetische beads toonden aan dat de detectielimiet van de MS instrumenten waarschijnlijk te hoog was om onze biomerkereiwitten te detecteren waarvan voorspeld werd dat ze in zeer lage concentraties (femtomol/liter) voorkwamen. Alternatieve aanrijkingstechnieken welke de gevoeligheid verhogen, zijn dus noodzakelijk **(hoofdstuk 5)**. Desalniettemin kunnen de bevindingen van deze studie nuttig zijn voor andere onderzoekers die overwegen om gelijkaardige technieken toe te passen. Ten slotte werd, in samenwerking met een industriële partner, een geautomatiseerde snelle immunoturbidimetrische *rapid plasma reagin* assay geëvalueerd op zijn klinische geschiktheid voor syfilisdiagnose en follow-up na de behandeling. Dit gebeurde met behulp van serummonsters verzameld gedurende een twee jaar durende observatie van een cohort van syfilispatiënten (N= 120) en met

serummonsters van controles (N= 30), uitgevoerd aan het Instituut voor Tropische Geneeskunde in Antwerpen, België **(hoofdstuk 6)**. Tekortkomingen in de prestaties van de test werden vermeld, welke in de klinische praktijk zouden kunnen leiden tot gemiste syfilisdiagnoses en suboptimale klinische follow-up. Onafhankelijke en uitgebreide testevaluaties met behulp van goed gekarakteriseerde klinische monsters zijn essentieel voor het verbeteren van diagnostische methoden.

In conclusie kunnen we stellen dat (moleculaire) epidemiologische studies van SOAs inzichten opleveren die kunnen helpen bij het informeren van preventie initiatieven. Verder heeft de toepassing van proteoomanalyse in het bestuderen van syfilis ons begrip van de enigmatische spirocheet *T. pallidum* verbreed. Deze bevindingen kunnen nuttig zijn voor de ontwikkeling van een syfilis-antigeentest.

Chapter 1
General Introduction

Few human pathogens have made such a detrimental impact on individual and public health during the last 600 years than *Treponema pallidum* ssp. *pallidum*, henceforth referred to as *T. pallidum*, the etiological agent of syphilis. This multistage chronic disease manifests with a myriad of clinical symptoms, earning itself the title 'the great pretender' by the father of modern medicine, Sir William Osler. Typically transmitted through unprotected sexual contact, *T. pallidum* can also pass through the placenta during pregnancy to cause serious adverse birth events. Despite continued advances in the biological understanding of this enigmatic organism and implementation of effective antibiotic treatment strategies, the medical community seems to lag one step behind, with marked resurgences in syphilis incidence during with the last 15 years among men who have sex with men (MSM) populations, in addition to persistent endemic presence in middle and low income countries that continues to cause considerable morbidity. Suboptimal diagnostic and prevention strategies play important roles in perpetuating these trends.

In this introduction, the epidemiology of syphilis will be explored with an emphasis on the HIV co-epidemic, along with a summary of syphilis clinical presentations, diagnostic, treatment and prevention strategies. Various important aspects of *T. pallidum* biology and pathogenesis will also be introduced, together with a short instruction to proteomics-based biomarker discovery approaches.

Syphilis epidemiology

In 2012, globally there were an estimated 18 million prevalent cases of syphilis and 5.6 million incident cases in men and women aged 15 to 49 years [1]. Syphilis during pregnancy also continues to cause substantial perinatal morbidity and mortality; in 2008, approximately 1.36 million (range 1.16-1.56 million) pregnant women were estimated to have a probable active syphilis infection [2], causing an estimated 520 905 cases of adverse outcomes, including an estimated 212 327 stillbirths and 91 764 neonatal deaths [2]. *Fig 1* represents recent global syphilis prevalence estimates in antenatal care attendees [3].

Syphilis prevalence has substantially varied during the last century [4]. During the acquired immunodeficiency syndrome (AIDS) epidemic in the late 1990s, syphilis prevalence plummeted in a number of hyperendemic countries, namely in Sub-Saharan Africa, likely due to the effect of AIDS mortality breaking up sexual networks [5–7] and the widespread application of sexually transmitted infection (STI)

Fig 1 *Percentage of antenatal care attendees positive for syphilis. Taken from reference* [3]

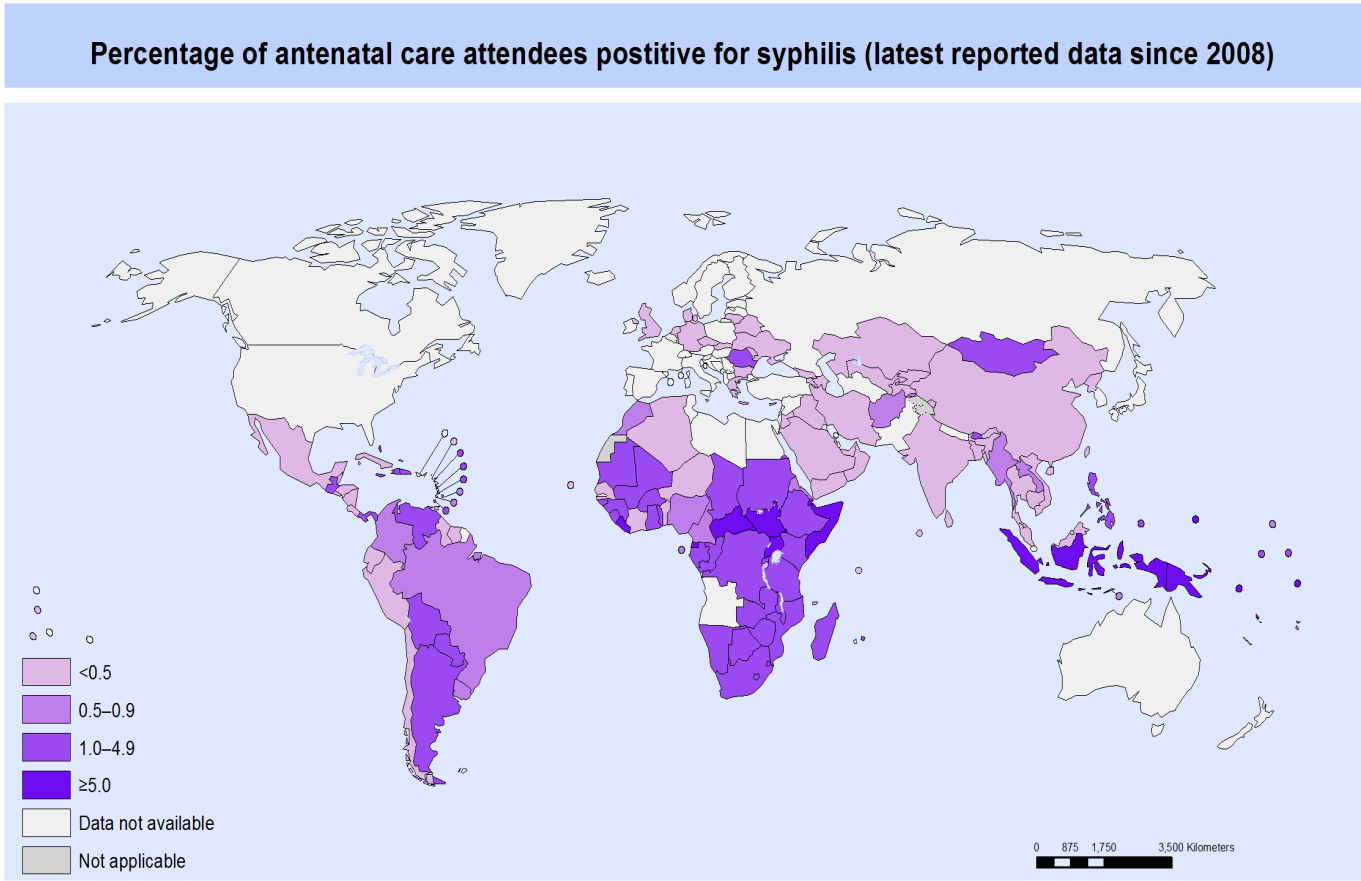
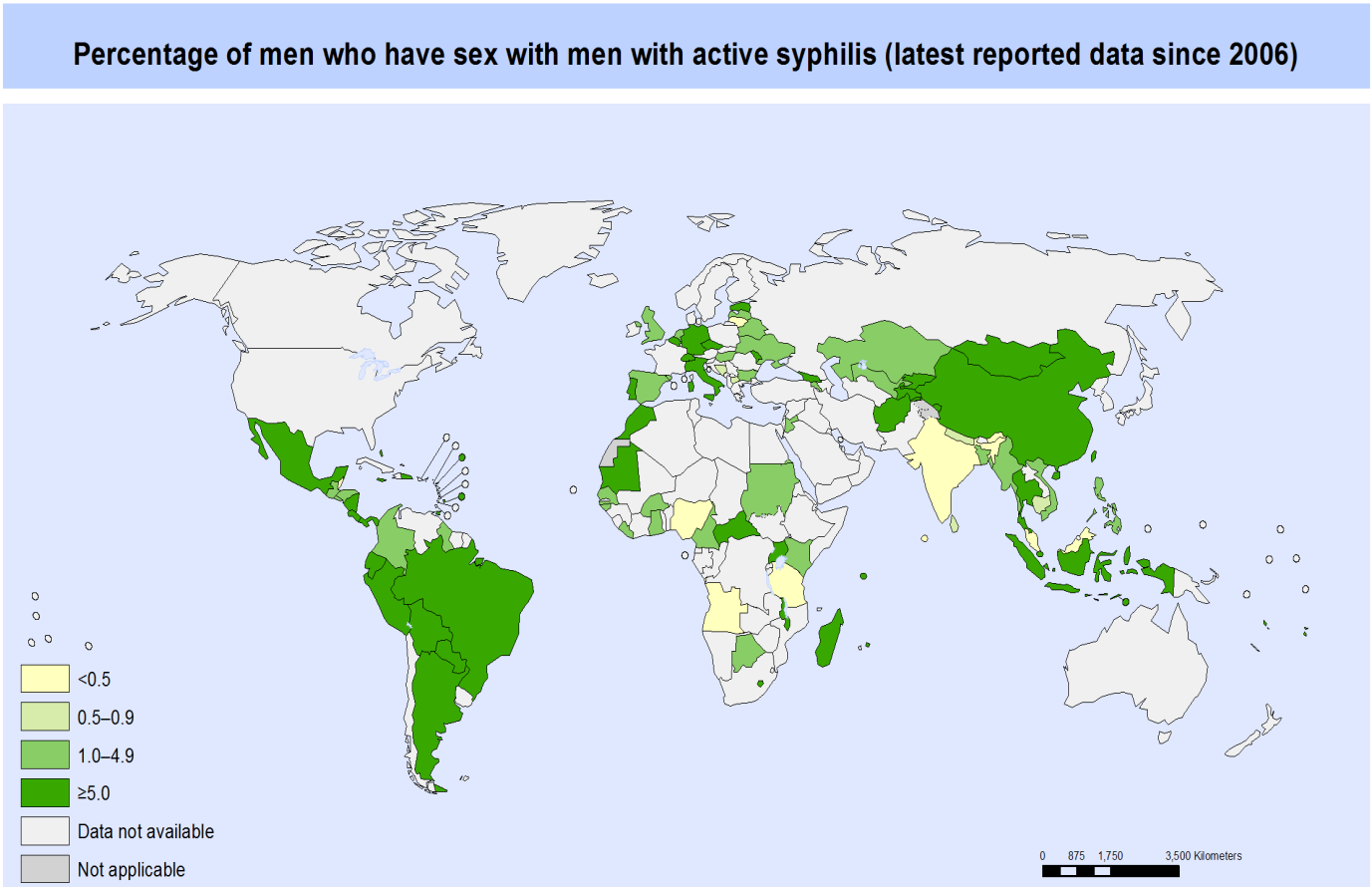
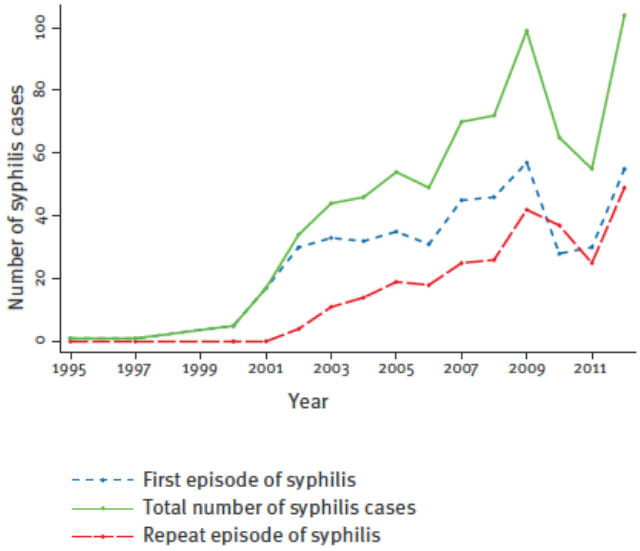


Fig 2 *Percentage of men who have sex with men (MSM) with active syphilis. Taken from* [8]



syndromic management [9]. Syphilis has re-emerged during the last 15 years in epidemic proportions, especially in North America [10], Western Europe [11] and China [12], with disproportionate disease burdens among males, particularly MSM. *Fig 2* represents a global overview of prevalence estimates of syphilis in MSM. A review of syphilis in 31 high income countries between 2000-2013 showed an increased male-to-female ratio in all geographical regions, from 4.1 in 2000 to 7.9 in 2013 [13]. In 2015, more than 80 % of male primary and secondary syphilis cases were diagnosed in MSM in the United States (US) [10]. China has also been experiencing widespread syphilis outbreaks; data from the national surveillance system in 2011 estimated around 3 million prevalent cases of syphilis [12], accounting for more than 15 % of the global estimate. Belgium has also witnessed sharp increases in syphilis incidence since the mid-1990s. In 2010, an estimated 7.7 % of MSM had active syphilis [14]. A medium-sized STI/HIV clinic of the Institute of Tropical Medicine (ITM) in Antwerp, Belgium, reported steady increases in syphilis cases since the mid 1990's, whereby most (>90 %) cases of syphilis presented in MSM with a high occurrence of reinfection [15,16] (*Fig 3*).

Fig 3 *Cases of initial and repeat syphilis diagnosed per year at the Institute of Tropical Medicine, Antwerp, Belgium between 1992*-2012 (n=729). Taken from reference* [15]

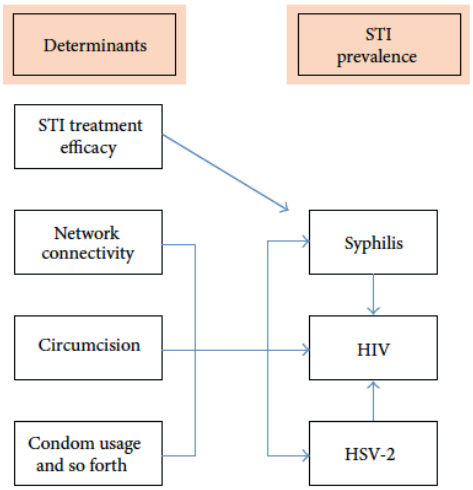


Syphilis persistence and recent outbreaks have been associated with a number of individual and population factors that perpetuate disease spread, such as the decreasing use of condoms [17,18], the prominent role of unprotected oral sex in *T. pallidum* transmission [19] and partner concurrency, defined as multiple sexual relations with different persons during the same time period [18,20].

Syphilis and Herpes Simplex Virus-2

Since Herpes Simplex Virus-2 (HSV-2) infection is incurable, sexually transmitted infection screening and treatment efficacy (STI-STE) would not have influenced HSV-2 prevalence [21], unlike syphilis where STI-STE is thought to have played a large role in the decline of syphilis prevalences (*Fig 4*) [22]. Moreover, there is no evidence for HSV-2 and syphilis enhancing the transmission of one another [23]. For these two reasons, HSV-2 prevalence is thought to be a useful measure to study non-STI-STE related risk factors in a population [24]; these studies could shed light on the determinants of variations in STI prevalence. A recent study I was a co-author of found a positive country-level association between HSV-2 and syphilis prevalence during two time periods (1990-1999 and 2008) using linear regressions analyses [25]. This hints that shared risk factors, such as low condom usage and connectivity of local sexual networks, were more likely to be responsible for variations in syphilis and HSV-2 prevalence, rather than STI-STE.

Fig 4 *Conceptual framework illustrating the relationship between syphilis, HIV and HSV-2 prevalence and four underlying risk factors. Syphilis prevalence is directly influenced by screening and treatment efficacy (STI-STE), HIV is indirectly influenced by this. Being non-curable, HSV-2 is not influenced by STI-STE. Reproduced from reference* [25]



HIV and syphilis

There are two distinct types of the human immunodeficiency virus (HIV), HIV-1 and HIV-2, both are RNA viruses that have been classified as members of the *Retroviridae* family and the *Lentivirus* genus [26]. HIV-infection is life-long and if left untreated

individuals develop fatal opportunistic infections within a decade as a result of HIV-induced deficiencies in the immune system, termed AIDS. This syndrome was first described in 1981 [27] and since then has become a worldwide epidemic affecting populations and geographic regions in varying magnitudes. A 2016 UNAIDS report estimated that 36.7 million (range 30.8-42.9 million) people were infected with HIV-1 [28] worldwide. Into the 21st century, major achievements in the development and roll-out of antiretroviral therapy (ART) for chronic HIV suppression and prevention campaigns have helped curb the worldwide epidemic.

Syphilis and HIV co-infection is an increasingly common occurrence [11,29]. According to a recent literature review [30], there is an estimated 9.5 % prevalence of syphilis among HIV-infected adults worldwide. In 2015, it was reported that half of primary and secondary syphilitic infections occurred in HIV-coinfected MSM in the US [10]. In China a substantial proportion of men with syphilis are also co-infected with HIV [29,31]. Trends are also similar in the province of Antwerp- at the ITM during 2000-2015, of 896 syphilis diagnoses, 642 (71.7 %) were diagnosed in HIV-positive individuals (personal communication V. Maes WIV-ISP).

These trends are not coincidental, as multiple studies have observed significant associations between syphilis and increased risk of HIV acquisition with hazard ratios ranging from 2.3 to 8.6 [32,33], particularity in MSM populations where diagnosis of syphilis carries a high risk of subsequent HIV seroconversion [34,35]. While behavioural factors such as low condom use [18] may explain this relationship, immunological and bacteriological factors also likely play a role HIV-spread. On a biological level, syphilitic lesions (chancres) enhance HIV-transmission by facilitating viral access to lymphocytes present in chancres and through the close proximity of HIV-infected cells to transmit virus to a sexual partner [36–38]. *T. pallidum* also increases HIV viral load [39], while treponemal lipoproteins are thought to increase the expression of the CCR5 chemokine cell receptor on monocytes that acts as co-receptors for HIV entry into $CD4^+$ T-cells [36,37].

Clinical manifestations of syphilis in HIV co-infected individuals is often more complex and atypical [40], with a higher risk of neurosyphilis [41] and poorer serological post-treatment outcomes [42]. Significantly reduced *T. pallidum*-specific opsonic activity of macrophages [43] and elevated levels of interleukin (IL)-10 expression in HIV-infected individuals [44] have also been reported, indicating differences in immunological response.

Chapters 2 and **3** of this thesis explore two aspects of the relationship between HIV and syphilis. In **chapter 2** we assess if there was a country-level association between antenatal syphilis prevalence during the early phase of the HIV epidemics (1990-1999) and peak HIV prevalence. Studying this association can help elucidate underlying determinants that may account for variations in HIV spread, especially in light of the recent HSV-2/syphilis study findings [25] which showed that factors such as network connectivity and not only STI-STE was responsible for drops in syphilis prevalence. A positive association could be interpreted as evidence that similar determinant(s) played a role in the spread of HIV and syphilis and/or that syphilis contributed to HIV spread. Data from 76 countries were subjected to linear regression analyses, revealing a strong association. We concluded that syphilis prevalence in the 1990s predicted approximately half of the variation in peak HIV prevalence.

Molecular epidemiological methods to enhance STI surveillance and prevention

Phylogenetics, the study of the evolutionary history and relationships among individuals or groups of organisms, can capture the underlying structures of transmission networks based on the genetic interrelatedness of viral sequences and cluster networks that could otherwise not be identified using conventional epidemiological methods. Phylogenetic inference usually involves computational approaches that implement maximum likelihood, parsimony and Markov chain Monte Carlo based Bayesian inference methods. An overview of these principles and practice is provided by Yang *et al.* [45]. Phylogenetic analyses of HIV sequences passively collected in the context of drug resistance testing has provided a plethora of information on HIV origins and transmission dynamics on many population and geographical levels [46–48]. These data are helpful to identify proximate and determinant traits associated with transmission, yielding information that can be used to inform prevention efforts. For example, a recent HIV molecular epidemiological study of a Belgian cohort elucidated characteristics of individuals present in inferred HIV transmission clusters, showing that individuals with a history of syphilis and chlamydia had a higher chance of clustering with one another, indicating that some sexual networks could be more high-risk for STIs than others [49]. It is expected that molecular bioinformatic methods will be increasingly applied to inform public health efforts [46].

Chapter 3 describes the application of phylogenetic methods to describe STI transmission patterns in a single centre-HIV cohort in Antwerp, Belgium. We assessed if a HIV phylogenetic tree constructed from HIV-1 sequences from 1169 individuals in follow-up at a HIV cohort in Antwerp, Belgium could be used to assess bacterial STI clustering. We found no evidence for STI clustering, however, analyses revealed six probable cases of sexual identity misclassification of MSM as heterosexuals. These individuals could represent a high-risk bridge population.

"The great pretender": clinical characteristics of syphilis

Clinical manifestations of syphilis, propensity for *T. pallidum* transmissibility and recommended treatment regimes vary over the course of infection [50]; these qualities extensively reviewed by Hook [51,52]. Clinical studies conducted during the antibiotic era involving individuals from low socioeconomic settings who were withheld syphilis treatment, in particular the Tuskegee and Guatemala trials, have been subject of considerable ethical controversies [53]. Recent excellent clinical studies (reviewed by Tipple [54]) have focused on syphilis presentation in HIV-infected individuals.

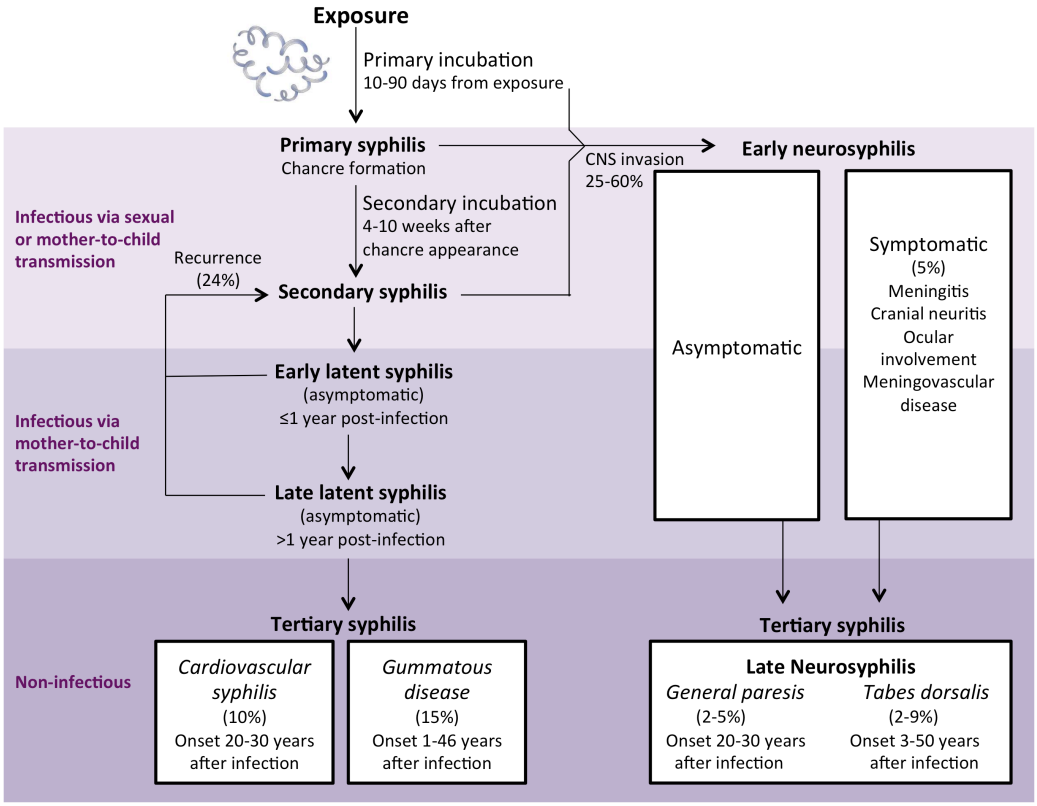
Disease manifestations

Natural progression of syphilis (*Fig 5*) is defined by two main stages: a highly variable symptomatic stage during the first 1-3 years of *T. pallidum* infection, followed by an asymptomatic prolonged latent phase, which in roughly 25 % of cases can evolve into a tertiary stage infection decades after initial infection [50,51]. Not all infected individuals experience all disease stages since syphilitic lesions often appear in inconspicuous anatomical sites, such as the anorectal area and vagina.

Primary stage syphilis is characterized by a local infection at the site of *T. pallidum* inoculation in the form of a painless, indurate, usually solitary ulcerative lesion (chancre) that appears 2-3 weeks (average 21 days and up to 90 days) after exposure. The incubation period is dependent on the size of inoculum; a 1936 study on prison volunteers (that would not pass contemporary ethical clearance) demonstrated that intradermal inoculation of individuals with 10 treponemes resulted in chancre development a mean 28.7 days later, whereas an inoculation with 104 organisms resulted in lesion development within a mean of 18.6 days [55]. Most chancres are localized on the distal penis in men, although they can also occur in atypical and/or inconspicuous anatomical sites such as in and near the rectum, vagina, cervix, fingers and mouth. Unprotected oral sex is thought to play an important role in syphilis transmission, as demonstrated by the high-prevalence of treponemes in the oral cavity

of persons with early syphilis [19]. Genital herpes (cause by herpes simplex virus) can mimic the clinical presentation of chancres, as well as chancroid (caused by infection with *Haemophilus ducreyi*), trauma and other dermatological processes. (Non)-tender regional lymphadenopathy can also be present during this stage. After suitable antibiotic treatment the lesions typically resolve within a few days, without treatment usually within 3-6 weeks.

Fig 5 *Natural history of untreated syphilis in immunocompetent individuals. Percentages of individuals developing to specific stages as well as time intervals are based on information in references* [56–58]. *Figure adapted from reference* [59]



Secondary stage syphilis is caused by systemic hematogenous dissemination of *T. pallidum*. Symptoms include a painless, erythematous, macular rash of 1-2 cm, typically located on the palms of hands or soles of feet. Rash manifestations can be highly variable- widespread or localized, pustular, papular and/or scaly in appearance and can occur simultaneously with primary stage lesions. Mucous membranes can also be affected during this stage, these lesions contain a highly infectious bacterial load in so-called 'mucous patches' or condylomata lata located in moist areas such as the genitals and perineum, often resembling the appearance of warts. Aspecific symptoms

such as malaise, low-grade fever, sore throat, headache are also common and hepatosplenomegaly, nephrotic syndrome and hepatitis can also be present during this stage. Without treatment the rash will typically resolve within weeks to several months without scarring.

After resolution of the symptomatic early stages, untreated syphilis progresses to the asymptomatic latent stage that can only be detected by serological testing. This stage is divided into early and late stages; 'early' is defined as within a year of initial infection and 'late' longer than one year. During the latent stage *T. pallidum* replicates more slowly and sequesters itself in privileged anatomical niches to evade immune detection [38,60]. After 2-3 years of untreated infection sexual transmission is rare, although approximately 25 % of individuals will present with secondary stage manifestations with a year or two after initial infection and in these period they are potentially infectious to sexual partners [57]. Clinical studies from the pre-antibiotic era have shown that approximately one third of infected individuals will progress to the much feared tertiary phase characterized by high morbidity and even mortality. This stage includes late neurosyphilis, presenting as general paresis or tabes dorsalis, gummateous and/or cardiovascular syphilis, most often presenting as aneurysm formation of the ascending aorta, coronary heart diseases or aortic valve insufficiency. Nowadays, with increased screening practices and frequent antibiotic tertiary syphilis is exceedingly rare.

Neurosyphilis, invasion of *T. pallidum* into the central nervous system, is a serious complication that can occur during any stage of syphilis [51]. Studies have suggested that mild meningeal signs and symptoms might be present in more than 40 % of individuals with secondary syphilis and these can resolve with or without therapy [61]. Clinical neurosyphilis can present in a number of ways, many of which roughly correlate with the duration of infection, with the exception of ocular involvement (eg. cranial nerve palsies, uveitis, etc.), which can occur at any stage. Rates of ocular involvement have been recently increasingly in HIV-infected patients [62] and higher levels of neurocognitive impairment in HIV-infected individuals have been reported in those with a previous syphilis history [63]. In later stages of infection, typically 5-10 years, or earlier in patients with HIV, inflammation of small and medium CNS arteries might result in stroke or premonitory stroke-like manifestations of meningulovascular syphilis, with the most common presenting symptoms being hemiplegia, aphasia or seizures [56]. General paresis, a cause of progressive dementia, seizures and a wide variety of psychiatric syndromes and in rare cases, tabes dorsalis [51,56] can also

appear in late stages of neurosyphilis, secondary to the involvement of nerves of the posterior columns presenting as lightning pains that occur as abrupt, unprovoked and severe radicular pain in combination with symptoms of ataxia, resulting from loss of proprioception. Both tabes dorsalis and general paresis are uncommon in the antibiotic era.

Syphilis during pregnancy

Besides the aforementioned clinical stages, *T. pallidum* also possesses the ability to damage foetuses. Hematogenously disseminating treponemes can pass through the placenta, resulting in 50 % of foetuses being aborted or stillborn and most exhibit diverse sequelae such as mucocutaneous lesions, hepatosplenomegaly and osteochondritis in the case of early congenital syphilis (before two years age). Late congenital syphilis children usually have clinical signs presenting later than two years of age, such as interstitial keratitis and blindness, cardiovascular lesions and bone deformations [64].

Syphilis Diagnostics

The reliable diagnosis of active syphilis is challenging. Owing to the difficulty culturing *T. pallidum in vitro* and lack of sensitive direct detection methods, syphilis diagnostics remains largely based on immunological detection [65].

Direct detection methods

T. pallidum can be visualized microscopically in fresh exudate using a dark-field condenser, however, considerable training experience and expertise is required for the proper use of this method, thus dark-field examination is either not done or not performed correctly in diagnostic settings [66]. Much effort have been invested in developing direct detection polymerase chain reaction (PCR) assays, although limited sensitivity has been achieved thus far. A 2013 systematic review and meta-analysis [67] showed that PCR analysis of swabs from primary genital or anal chancres have the highest sensitivity (78.4 %; 95 % Confidence interval (CI): 68.2-86.0), whereas latent stage infection (blood specimens) was related to low specificities, on average 31.2 % (CI: 22.4-41.5). The development of robust direct testing methods would revolutionize syphilis diagnostics, a topic that is addressed in **chapters 4** and **5** of this thesis.

Syphilis serological tests

Syphilis serological tests represented an important milestone in the history of diagnostic test development. The Wasserman Reaction developed in 1906 by Wasserman and colleagues [68] formed the basis for the modern nontreponemal tests (NTTs), including the rapid plasma reagin (RPR), Venereal Disease Laboratory (VDRL) and Toluidine Red Unheated Serum Test (TRUST) tests. These are based on bovine heart extract containing antigens that react with serum immunoglobulin (Ig)-G and IgM directed against lipoidal material (cardiolipin, cholesterol and lecithin) originating from damaged human cells and *T. pallidum* cellular components [69,70]. NTT titres roughly correlate with disease activity and typically become positive 10-15 days after chancre presentation. Without antibiotic treatment titres peak between 1-2 years following infection and remains positive with low titres in very late disease [71]. NTTs have evolved little during the last century and most are performed manually and are thus subject to interindividual interpretation and reagent batch variations [72,73]. Recently, automated quantitative latex turbidimetric immunoassays (LTIA) based on NTT principles have been widely implemented in Asian laboratories [74], however, large-scale comparative studies in high STI prevalence settings have been lacking. Automation of NTTs could contribute in various ways to improve syphilis diagnosis, for example, by increasing test accuracy with the use of an internal standardized calibrant, higher throughput capability and reporting of more accurate continuous values instead of titre values. **Chapter 6** of this thesis details a laboratory evaluation of a commercial RPR LTIA assay (Sekisui Chemical Co. Ltd., Osaka, Japan).

NTTs are still the mainstay tests for post-treatment follow-up to determine *T. pallidum* eradication and for the diagnosis of reinfection [38], despite many shortcomings. False negative results may occur during late latent stages or as a result of prozone phenomenon occasionally (0.3 % of cases) seen during early stages (primary/secondary) of infection when a higher than optimal amount of antibody in the tested sera prevents the formation of the antibody antigen lattice network needed to visualize a positive flocculation reaction [75–78]. Serum dilution is necessary to rectify prozone (false negative) or hook effects (lower than expected titre values). Moreover, early detection of syphilis can also be hampered by lagging antibody production during very early infection resulting in negative or non-responsive RPR tests [71]. False positive biological NTT reactions are also not uncommon and can be secondary to (auto) immune reactions, co-infections such as malaria and HIV and physiological factors such as older age [79–81] and pregnancy [82]. A recent study of 63 765 RPR tested blood samples revealed that 206 (0.32 %) had a biological false positive NTT [80].

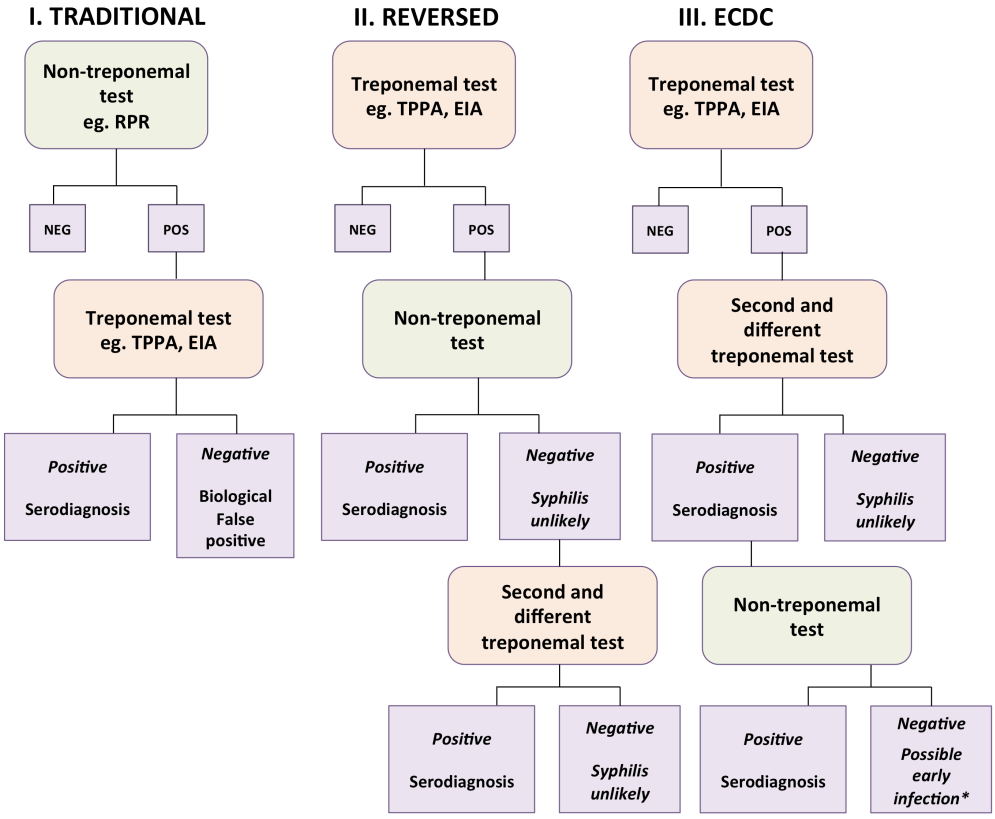
Treponemal antibody (TT) tests, such as the *Treponema pallidum* particle agglutination assay (TPPA) and Fluorescent treponemal antibody absorption (FTA-ABS) assay, measure IgG and IgM directed specifically against *T. pallidum* proteins. These tests generally perform well (*Table 1*) [83], however, they are not useful for the diagnosis of reinfections since they remain positive many years after primary exposure. In the 'traditional algorithm' [84] TTs are used to confirm reactive NTT results, however, with the increased availability of automated, low-cost, high-throughput *T. pallidum* chemiluminescent/luminescent immunoassays many laboratories in low to medium STI prevalence settings have switched to the 'reverse sequence algorithm' in which a TT is conducted first, and if reactive, a confirmatory NTT is performed [85]. The European Centre for Disease Prevention and Control (ECDC) recommends an extended version of this algorithm: a reactive TT is followed by a second, different TT that is used as a confirmatory test which is then followed by a quantitative NTT assay when the second TT is positive [86]. *Fig 6* is a schematic overview of the three main syphilis diagnostic algorithms. Generally, the reversed/ECDC algorithm has been found to perform better than the traditional [87,88]. A recent HIV-cohort study by Chen *et al.* [89] compared to the performance of these two algorithms; they found a missed serodiagnosis rate of 42.8 % using the traditional algorithm using a TRUST test the first screening assay. It should be noted that in high syphilis prevalence settings the probability of syphilis reinfections is high [15], therefore combined screening with NTT/TT is a pragmatic screening approach for high-risk individuals. Enzyme immunoassays (EIAs) for the detection of IgM can be performed to detect very early infection when NTTs may still be negative, yet, the sensitivity of these tests may not be optimal [90], especially for those with repeat infection. A comprehensive overview of the recent developments syphilis serological testing can be found in a review by Seña *et al*. [85].

Table 1 *Performance of serological tests for syphilis. Reproduced from* [71]

	Sensitivity (%) by stage of untreated syphilis				
Test	**Primary**	**Secondary**	**Early Latent**	**Late Latent**	**Specificity (%)**
VDRL	78 (74–87)	100	96 (88–100)	71 (34–94)	98 (96–99)
RPR	86 (77–99)	100	98 (95–100)	73	98 (93–99)
TRUST	85 (77–86)	100	98 (95–100)		99 (98–99)
FTA-ABS	84 (70–100)	100	100	96	97 (84–100)
TPPA	88 (86–100)	100	100		96 (95–100)

Legend: FTA-ABS- fluorescent treponemal antibody absorption; RPR- rapid plasma reagin; TPPA- T. pallidum particle agglutination; TRUST- Toluidine red unheated serum test; VDRL- Venereal Disease Research Laboratory

Fig 6 *Syphilis testing algorithms. Legend: EIA- enzyme immunoassay; ECDC- European Centre for Disease Prevention and Control; *- in cases where confirmatory NTT is negative consider performing an IgM EIA but in all cases treat regardless of result*



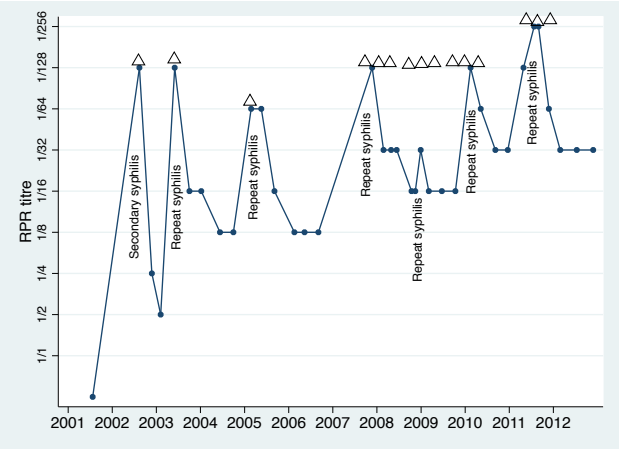
Post-treatment serological follow-up of syphilis & the 'serofast' conundrum

In addition to the cessation of clinical manifestations, serological post-treatment follow-up is also required to ascertain cure and detect reinfection or relapse. European guidelines recommend NTT testing 1, 3, 6 and 12 months after initial treatment [86], whereby a requisite fourfold (or two dilution; e.g., 1/64 to 1/16) decline of NTT titres after treatment is necessary as a "proof of cure". This stipulation is largely based on a study by Brown *et al.* [91] that evaluated the seroresponse of a cohort of primary and secondary stage patients whose symptoms resolved after treatment. The titre response curves generated from VDRL test data demonstrated a fourfold decline in titres at three months following therapy. However, in clinical practice not all individuals follow this ideal serological evolution; this stagnation of NTT titres after adequate therapy is referred to in the literature as "serological failure", "serological nonresponse", "seroresistance", "reagin-fast", or the "serofast state" and can be indicative of a persistent infection (treatment failure), reinfection, benign immune response unassociated with *T. pallidum*-infection or a residual immune response still ongoing in the absence of viable *T. pallidum* organisms [42]. Antibiotic therapy is typically readministered in these cases [86], although this usually does little to promote further titre [92]. A literary review of serological outcomes after syphilis treatment reported a median proportion of 'serofast' (this term will henceforth be used) individuals of 20.5 % at six months, which decreased to 11.2 % at ≥12 months for all stages of syphilis [42]. The serofast state occurs more frequency in patients with latent late latent syphilis and lower baseline test titres [93,94]. Particular *T. pallidum* strain subtypes and higher serum concentrations of the cytokine chemerin [95] may also increase susceptibility. For HIV-positive individuals, the relationship between serological cure and HIV status was found to be inconsistent in a systematic review by Seña *et al*. [42], whereby $CD4^+$ T-cell count and HIV viral load were not associated with serological cure.

Both treatment failure and reinfection are typically defined as a ≥ fourfold rise in NTT titres after treatment; these two conditions can only be distinguished by a history of having had unprotected sex with a potentially infected person. The diagnosis of reinfections poses particular challenges to clinicians since reinfections often present asymptomatically, especially in individuals who have a history of multiple infections. For example, a retrospective study at the ITM reported 729 episodes of syphilis diagnosed in 454 persons between 1992-2012. A majority (445/729; 61 %) of these individuals had more than one episode and these individuals frequently presented

without signs or symptoms of syphilis [15]. This phenomenon has also been observed in other populations [55,96,97] and is likely associated with an altered immune response, as reflected by a recent study that showed higher baseline RPR titres and lower levels of the cytokine IL-10 in the plasma of persons with repeat *T. pallidum* [16]. *Fig 7* is an example of a patient presenting with multiple reinfection episodes.

Fig 7 *Illustration of difficulties associated with depending on NTTs for the diagnosis of syphilis reinfections. Over an 11-year period this individual was diagnosed with an initial episode of symptomatic secondary syphilis, followed by six subsequent diagnoses and treatments (each hollow triangle represents benzathine penicillin G intramuscular treatment) for asymptomatic syphilis. Each syphilis diagnosis was based solely on RPR-titre fluctuations and his RPR titre never dropped below 1/8. We do not know if these diagnoses of repeat syphilis are accurate, or represent over- or under-diagnoses (personal communication C. Kenyon)*



Treatment

Recently, the European Centers for Disease Control and Prevention (ECDC) [86], Centers for Disease Control and Prevention (CDC) [98] and World Health Organization (WHO) [99] guidelines for syphilis management were updated in an effort to provide standardized evidence-based recommendations. It should be noted, however, that most evidence is based on dated clinical series and there were few randomized control trial series performed to effectively evaluate treatment efficacy. Recommended treatment regimes vary over the course of infection and are usually differentiated between early (≤1 year) and late (>1 year) syphilis and the presence of neurological invasion. In rare cases individuals can experience shakes, chills, a rise in temperature and intensification of skin rashes related to the Jarisch-Herxheimer reaction after treatment with symptoms

usually resolving within a few hours (reviewed by Butler [100]). A summary of the European treatment recommendations is detailed in *Table 2*.

Table 2 *Treatment of syphilis in Europe. Reproduced from* [86]

Early syphilis (Primary, Secondary and Early latent, i.e. acquired ≤1 year previously)
First line therapy option Benzathine penicillin G 2.4 million units intramuscularly (IM) (one injection of 2.4 million units or 1.2 million units in each buttock) on day 1
Penicillin allergy or parenteral treatment refused Doxycycline 200 mg daily (either 100 mg twice daily or as a single 200 mg dose) orally for 14 days or azithromycin 2 g orally single dose
Late latent (i.e. acquired >1 year previously or of unknown duration), cardiovascular and gummatous syphilis
First line therapy option Benzathine penicillin G 2.4 million units IM (one injection 2.4 million units single dose or 1.2 million units in each buttock) weekly on day 1, 8 & 15
Penicillin allergy or parenteral treatment refused Desensitization to penicillin or doxycycline 200 mg daily orally during 21–28 days
Neurosyphilis, ocular and auricular syphilis
First line therapy option Benzyl penicillin 18–24 million units IV daily, as 3–4 million units every 4 h during 10–14 days
Second line therapy option (if hospitalization and IV benzyl penicillin is impossible) Ceftriaxone 1–2 g IV daily during 10–14 days Procaine penicillin 1.2–2.4 million units IM daily AND probenecid 500 mg four times daily, both during 10–14 days
Penicillin allergy Desensitization to penicillin followed by the first line regimen
Syphilis in pregnancy
Pregnant women should be treated with the first line therapy option appropriate for the stage of syphilis and if allergic to penicillin should be desensitized.
Syphilis in HIV
Treatment should be gives as for non-HIV infected patients, although there are very few data on the use of second line options

Parenteral benzathine penicillin G (BPG) revolutionized modern medicine in the 1940's and still remains the best treponemicidal therapeutic drug [101] and first-line treatment for syphilis. Unlike many other human pathogens [102], such as enterococci and *Staphylococcal aureus, T. pallidum* has not (yet) developed resistance against penicillin. BPG treatment is thought to be almost always effective since no cases of treatment failure have been verified microbiologically by demonstration of viable *T. pallidum* and

exclusion of reinfection [86]. However, in some studies [60,103] treponemes have been shown to persist even after successful treatment, the clinical significance of this is unknown. One theory regarding this continued sensitivity is the contribution of *T. pallidum* protein Tp47, a penicillin binding protein with high beta-lactamase activity that is inhibited by its own catalytic products [104]. There is also a lack of recombination or uptake of genetic information through horizontal gene transfer [105], a mechanism oftentimes present in other (pathogenic) bacteria [106], rendering the probability of *T. pallidum* developing penicillin resistance unlikely.

The efficacy of alternative therapies such as doxycycline/tetracycline and ceftriaxone was evaluated in a recent meta-analysis by Liu *et al*. [107], wherein ceftriaxone was reported to have a similar efficacy to BPG in treating early syphilis. Moreover, in the same study when doxycycline/tetracycline was compared with ceftriaxone, ceftriaxone appeared to be more effective [107]. Macrolide antibiotics were historically recommended in penicillin-allergic patients presenting with primary, secondary and latent syphilis [108], however, clinical failures following azithromycin treatment were observed in several syphilis patients in San Francisco, United States from 2002-2003 that upon laboratory investigation showed the presence of A2058G mutation in the *T. pallidum* 23S rRNA gene [109], the same mutation that was previously identified in a SS14 strain isolate from a patient who failed intensive erythromycin treatment and exhibited cross-resistance to azithromycin [110]. In 2009, Matejkovà *et al.* [111] identified a new mutation 23s rRNA gene mutation A2059G in syphilis patients in the Czech Republic that also confers macrolide resistance. The presence of macrolide-resistant *T. pallidum* strains have been drastically increasing over time [101], with varied geographical prevalence, for example, 0.7 % in Taiwan [112] to 83 % in Australia [113] and nearly 100 % in China [114]. L. Stamm comprehensively reviewed the recent antibiotic resistance patterning of *T. pallidum* [101,115]. Furthermore, blood samples collected between 2014-2016 from syphilitic patients attending the ITM's STI clinic where subjected to treponeme typing in collaboration the University of Masaryk, Czech Republic. *T. pallidum* DNA was detected via PCR analysis in 29/72 (40.3 %) of whole blood samples, whereby all samples with amplifiable 23S rRNA (n=19; 65.5 %) possessed the A2058G mutation, suggesting 100 % prevalence of macrolide-resistant *T. pallidum* strains in the Antwerp region [116].

In lieu of reports of HIV-infected individuals having an increased predilection for the development of early symptomatic neurosyphilis [117,118] and evidence that single dose of 2.4 MU BPG may not be sufficient to clear the CSF of *T. pallidum* in HIV-infected

individuals with early syphilis [61], enhanced treatment regimes for co-infected persons has been the subject of intense debate, as reviewed by Lawrence *et al*. [119]. Results of studies are divided: two groups have reported little changes in treatment efficacy when comparing intensified treatments such as three injections of BPG *versus* one dose [120,121], conversely, a prospective observational study conducted in China showed increases in serological failure with only one BPG dose [122]. Furthermore, oral doxycycline has been reported to be similarly effective as BPG for the treatment of HIV-infected individuals with early syphilis [123].

Prevention

Disease prevention is a multifaceted complex process- a comprehensive explanation thereof would go beyond the scope of this thesis introduction. An important aspect of effective prevention is correctly assessing risk factors and risk groups associated with a particular disease. Baral *et al.* [124] proposed a modified social ecological model to help visualize multi-level domains of HIV-infection risks to help guide epidemiological study design. In many ways, these approaches can be generalized to syphilis prevention by considering the various levels of interventions and risk in order to attain an accurate estimation of disease burden and possible gaps in care.

Prompt treatment of pregnant women with a single dose of long acting BPG before 24 weeks of pregnancy can prevent syphilis transmission to the foetus. Therefore, syphilis testing of all pregnant women is part of the recommended basic antenatal care package recommended by the WHO, a strategy successful in reducing the incidence of perinatal death and stillbirth attributable to syphilis by approximately 50 % [125]. According to a recent report, only 45 of 89 reporting countries were successful in treating over 95 % of pregnant women who tested positive for syphilis [2], indicating that implementation of syphilis treatment policies may be limited in many settings, especially those with highest disease burden. At least 18 countries, including the US, Canada, Portugal, France, Belgium and Brazil, have faced recent shortages of BPG according to the WHO [126] and CDC [127]. Estimates indicate high volume demands of BPG needed to treat pregnant women [128], hopefully the few global manufacturers left will step up to meet these demands.

Syphilis stigmatization has played a part in hampering prevention efforts such as partner notification and population screening throughout the years [129]. Especially 'ping-pong' syphilis, a phenomenon occurring when one partner is treated but not the other, with the inherent possibility of later reinfecting oneself via the untreated partner,

contributes to sustaining syphilis prevalence. This usually occurs in cases where patients are reluctant to disclose their infection status or unable to contact previous (anonymous) sexual partners. Modern communication technologies are expected to increasingly facilitate prevention through the promotion of screening and (anonymous) partner notification [130]. Evidence of prevention success was detailed in recent Australian study of MSM that reported decreases in secondary syphilis and increases in the detection of asymptomatic early latent stage syphilis, likely owing to the increases in screening and possible interruption in syphilis progression [97]. Moreover, molecular epidemiologic analyses of may also have a prominent role in informing public health strategies in the future, as discussed by German *et al.* [46] and in **chapter 3** of this thesis.

Progress related to syphilis vaccine development was expertly reviewed by Lithgow and Cameron [131]. Anno 2017, a syphilis vaccine still evades us, although recent advancements using a *T. pallidum* protein Tp0751 (pallilysin) as a vaccine candidate yielded promising results in rabbit infectivity studies [132]. A mathematical modelling study demonstrated the potential of an efficacious syphilis vaccine to sharply reduce syphilis prevalence under a wide range of scenarios [133].

Biology of *T. pallidum* ssp. *pallidum*: a unique stealth pathogen

The microaerophilic fastidious spirochete *T. pallidum* ssp. *pallidum* is a helically coiled shaped bacterium aptly built to burrow into tissue (see *Fig 8/9* from [134]). The body of *T. pallidum*, 6-15 μm long by 0.15 μm wide, is surrounded by a cytoplasmic membrane enclosed by a loosely bound fragile outer membrane, with a peptidoglycan layer located between these two membranes to provide structural support. Periplasmic endoflagella encased in an outer sheath and wound around the helical protoplasmic cylinder facilitate the rapid rotation around the long axis causing flexation of the cell and locomotion along a helical path, providing its characteristic corkscrew motility [135]. *T. pallidum* is an organism notorious for its ability to avoid immune detection during infection, earning the name 'stealth pathogen'. A detailed description of the fascinating studies related to *T. pallidum's* stealth properties has been reviewed by Radolf *et al.* [136].

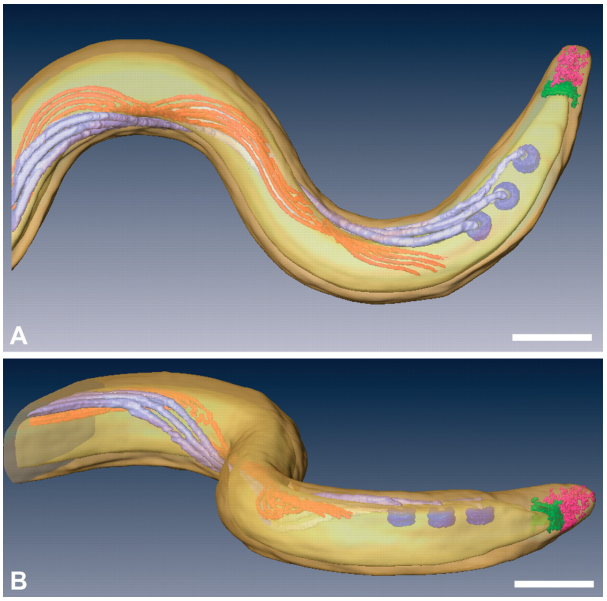


Fig 8 *Top (A) and side (B) views of a surface-rendered model of* T. pallidum*, showing the outer and cytoplasmic membranes (transparent yellow), basal bodies (dark lavender), flagellar filaments (light lavender), cytoplasmic filaments (orange), cap (green), and cone (pink). The peptidoglycan layer was not rendered. Bars: 200 nm. Reproduced from* [134]

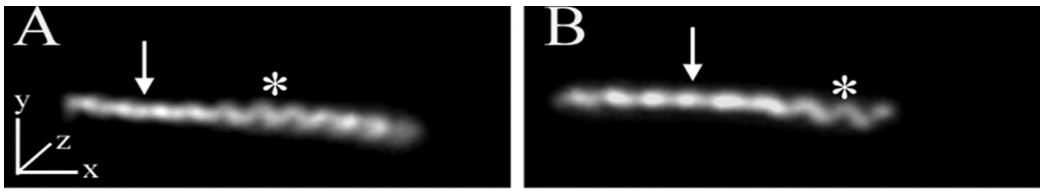
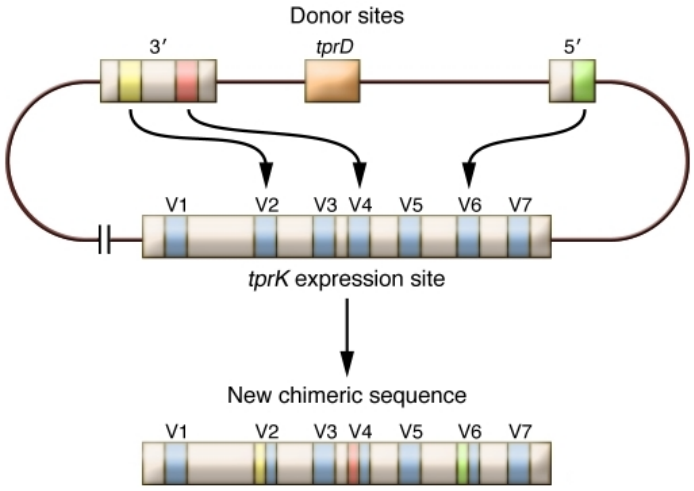


Fig 9 *Representative micrographs show the flat-wave morphology of* T. pallidum *as revealed by dark-field microscopy (A and B). The arrows and asterisks indicate regions of T. pallidum that are parallel to the z axis or in the x-y plane, respectively. Reproduced from* [134]

Briefly, unlike other gram negative bacteria, *T. pallidum* lacks a lipopolysacchide layer [69,105] that is usually highly immunogenic. Freeze-fracture microscopic [137,138] and cryo-electron tomographic studies [134] elucidated the flat-wave morphology of *T. pallidum*, revealing that its outer membrane contains 100-fold less membrane-spanning proteins than the outer membranes of typical gram-negative bacteria such as *Escherichia coli*. This paucity of potentially immunogenic surface exposed proteins is one of main determinants of *T. pallidum's* ability to avoid immune detection. Characterization of these proteins in relation to the chronicity of syphilitic infection has been the subject of intense debate [38,136]. Antigenic variation through gene conversion also plays a role in *T. pallidum*'s propensity to evade acquired immune system responses (reviewed in [38,136]), a mechanism also found in other persistent pathogenic spirochetes such as *Borrelia burgedorferi* (Lyme disease) [139] and *Borrelia hermsii* (relapsing fever) [140]. Members of the paralogous family of 12 *T. pallidum* repeat (Tpr) proteins have been identified as integral outer membrane protein candidates [141,142], among these

protein TprK is the most studied. TprK sequences have been shown to vary substantially within and between individual *T. pallidum* strains [143–146], with sequence diversity increasing during the course of infection [145,147]. This process is thought to facilitate the development of secondary stage syphilis [146]. Sequence diversity is located in seven discrete variable TprK protein regions that are predicted to be surface exposed. New variants of TprK arise by segmental gene conversion, with new sequence segments originating from donor sites located elsewhere on the chromosome [148,149]. *Fig 10* describes this process.

Fig 10 *Antigenic variation through gene conversion of the* T. pallidum *TprK protein. Variant DNA segments located adjacent to the* trpD *gene recombine non-reciprocally with the variable region (V1-V7) of the* tprK *gene in the expression site to generate new* TprK *mosaic proteins. Reproduced from* [59]



T. pallidum is slow-growing, with a replication time of 30-33 hours [150], owing to its limited metabolic capacity secondary to a lack of tricarboxylic acid cycle enzymes and an electron transport chain [105]. In the absence of biosynthetic pathways, it is thought that *T. pallidum* acquires most essential macromolecules from the host milieu, whilst relying on interconversion pathways to generate others. *T. pallidum* also poorly tolerates temperature changes and elevated oxygen tensions [151,152], rendering it extremely fragile. These factors contribute to its inability to survive outside of the mammalian host; persistent efforts to successfully propagate *T. pallidum in vitro* beyond several replications in tissue culture have failed [150,153,154]. Consequently, most studies on *T. pallidum* are conducted using the well-established methods of intratesticular inoculation of rabbits that yields viable organisms within 10 days [155,156] during peak orchitis [157,158]. Crude testicular bacterial extracts are typically subjected to density gradient centrifugation shortly after extraction to yield between

$1x10^{8-10}$ treponemes per extraction, which are then quantified using dark-field microscopy [159]. These techniques were applied in **chapters 4** and **5** of this thesis.

T. pallidum genome and proteome

T. pallidum ssp. *pallidum* shares close DNA sequence homology (>99 %) with other pathogenic spirochetes, including *T. pallidum* ssp. *pertenue*, the causative agent of yaws, *T. carateum*, the etiological agent of pinta and *T. pallidum* ssp. *endemicum* which causes endemic syphilis, also called bejel [160,161]. There is also a high degree of similarity to the rabbit pathogen *T. paraluiscuniculi*, a treponeme not infectious to humans [162]. A summary of characteristics related to the four human treponemal diseases is provided in *Table 3*. The pathogenic human treponemes vary in terms of virulence, with *T. pallidum* being the most invasive, *T. pallidum* ssp. *pertenue* moderately invasive and *T. carateum* being non-invasive; the implications of genetic diversity within this species on pathogenicity, evolution and diagnostics is reviewed by Šmajs [163].

Table 3 *Characteristics of the four treponemal diseases. Reproduced from* [164]

	***T. p. pallidum* (Syphilis)**	***T. p. pertenue* (Yaws)**	***T. p. endemicum* (Bejel)**	***T. carateum* (Pinta)**
Epidemiology				
Geographical distribution (climate)	Worldwide	Tropics (hot, human areas)	Deserts of Africa and Saudi Arabia (hot, dry areas)	Central and South America (hot, humid areas)
Age group (peak incidence)	Adults (18-30)	Children (2-10)	Children (2-10)	Adults (15-50)
Transmission	Sexual and congenital	Skin-to-skin contact	Mouth to mouth or utensils	Skin-to-skin contact
Clinical Characteristics				
Initial lesion (location)	Common (genitals)	Common (lower extremities)	Rare (oral mucosa)	Common (extremities)
Dissemination	Widespread and systemic	Widespread to skin and bone	Limited to intertriginous areas and facial bone	Limited only to skin
Late complications without treatment (%)	Gummas (10 %), neurological (10 %), cardiovascular (10-15 %)	Destructive lesions of skin and bones (10 %)	Destruction of nose/palate	Local skin hypo-achromia

The *T. pallidum* Nichols strain isolated in 1912 from the CSF of a patient with secondary syphilis (and since then has been propagated in rabbits) is the most ubiquitous and well-studied *T. pallidum* strain. The Nichols genome was first published in 1998 [105], followed by resequencing in 2013 [165], revealing a circular chromosome containing 1 139 633 base pairs (bp) containing 1039 predicted coding sequences (open reading frames; ORFs). In the meantime, the genomes of other *T. pallidum* strains such as DAL-1 [166], Chicago [167], SS14 [168] and Amoy [169] have been published, revealing small genetic differences in some loci, the clinical significance of these variations remains unclear. Compared to most bacteria, *T. pallidum* has one of the smallest bacterial genomes due to evolutionary reduction [163]. Evidence of this was indirectly demonstrated by a microarray transcriptome study that demonstrated the expression of almost all genes during experimental rabbit infection [170].

The predicted proteome of the *T. pallidum* reference strain (NCBI ID# NZ_CP003679.1; based on strain Sea 81-4 [171]) codes for 977 proteins. A combined proteome and serological analysis of *T. pallidum* using rabbit expression and gel-separation and MS analyses of bacterial protein extracts detected 88 polypeptides [172], many of which were shown to be reactive with human and rabbit *T. pallidum*-infected sera. **Chapter 4** of this thesis describes the characterization of the *T. pallidum* proteome during rabbit infection using liquid chromatographic (LC) peptide separation combined with high-resolution mass spectrometric analyses.

T. pallidum infection tactics

Shortly after contact with infectious exudate from another *T. pallidum*-infected individual, the transmitted treponemes burrows through the cutaneous layer to the vascular endothelium where it quickly disseminates through the vascular system to a variety of anatomical sites. Rabbit studies have demonstrated rapid systemic dissemination, whereby spirochetes were detectible in plasma 24 hours after inoculation, with the largest number of bacteria found in the spleen [173]. Recent studies have implicated *T. pallidum* proteins Tp0750 and Tp0751, in tissue invasion and dissemination processes [174–176]. Peak pathogen load during secondary syphilis has been estimated to be 10^4 to 3x 10^4 *T. pallidum* per millilitre of whole blood [177,178], although this number is expected to be considerably lower during latent infection when *T. pallidum* bacterial load drops. Therefore, latent stage, especially late, is expected to be the most challenging hurdle for *T. pallidum* antigen detection diagnostics.

Mass-spectrometry based proteomics

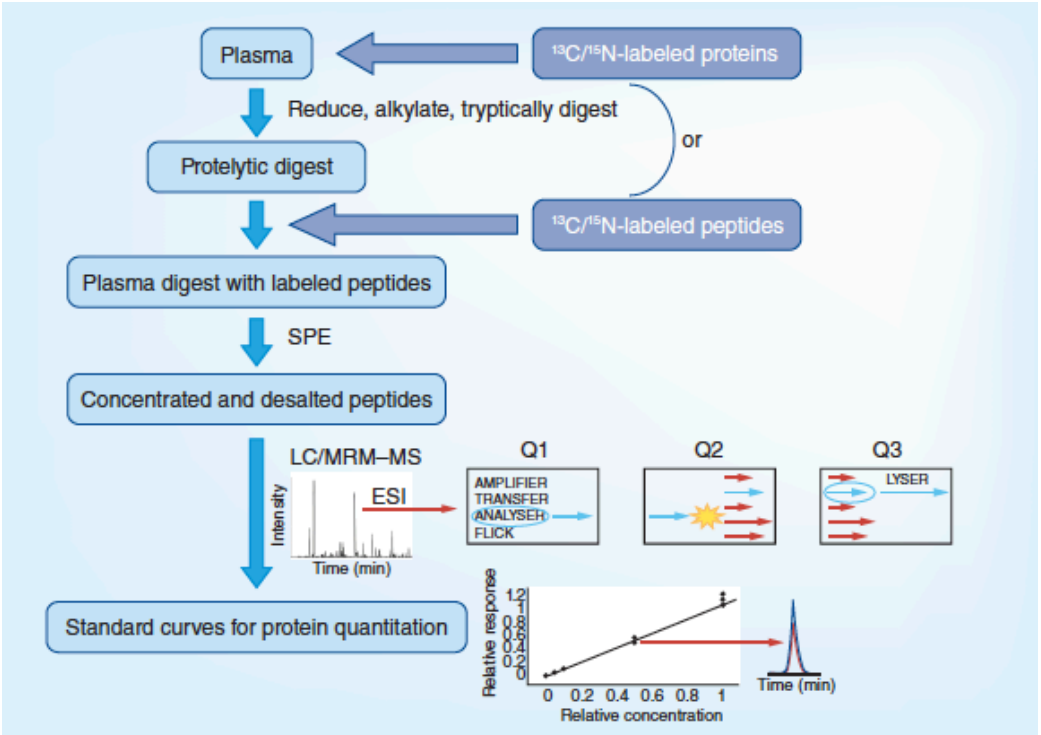
Mass spectrometry (MS) has emerged as a core tool for large-scale protein analyses, a technology that is increasingly being applied in clinical diagnostics [179–181] owing to its speed, sensitivity, specificity, throughput and ability to quantitatively analyze a proteome with high accuracy. During MS analysis, compounds of interest (e.g. proteins, peptides, lipids, metabolites) are typically ionized and introduced into the gas phase where they are analyzed according to their mass-to-charge (m/z) ratio in various MS modes. The past decade has seen rapid advancements in mass accuracy, sensitivity, scan rate and resolution of mass spectrometers used to analyse proteins, especially with the introduction of hybrid mass analysers, for example the linear ion trap-Orbitrap series [182].

The characterization of proteins by the analysis of peptides produced after proteolysis, performed with an enzyme such as trypsin, is termed a "bottom-up" proteomics. When this technique is applied to a mixture of proteins, it is referred to as "shotgun proteomics" [183,184], a name coined by the Yates group that draws an analogy to shotgun genomic sequencing [185]. An excellent review by Zhang *et al.* provides a full survey of shotgun proteomics, including fundamental techniques and applications [186].

Named the "Method of the Year" by Nature Methods in 2012 [187], selected reaction monitoring (SRM), also called multiple reaction monitoring (MRM), has emerged as an ideal method for finding and quantitating proteins in a variety of (biological) mediums [188]. MRM is a tandem MS technique usually performed on a triple-quadrupole mass spectrometer. The first analytical steps involve the selection of a precursor ion in the first quadrupole, a peptide that acts as a surrogate for the protein of interest. The precursor ion (protonated intact peptide) is then fragmented in the second quadrupole and following fragment selection in the third quadrupole the fragments hit the detector; the quantitation of the protein is based on the signal that reaches the detector. The precursor/product ion pair is referred to as the "transition". Since the fragmentation occurs on a millisecond time scale, different transitions can be selected as a function of retention time in an LC-MRM experiment, a type of analysis called "scheduled MRM". The addition of stable-isotope ($^{13}C/^{15}N$)-labelled standard (SIS) peptide analogues of the endogenous peptide targets into the MRM work-flow has made this the "gold-standard" MS-based quantitation method [189], allowing for the quantitation of hundreds of peptides (which can be extrapolated to the parent protein

concentrations). The reproducibility of MRM assay performances is similar to that of Enzyme-linked immunosorbent assays (ELISA), with coefficients of variation often below 10 % [190,191]. A schematic representation of these methods is shown in *Fig 11*. The speed and wide detection range of MRM was demonstrated in a study involving the quantitation of 142 high-to-moderate abundance proteins (from 31 mg/mL to 44 ng/mL) corresponding to 312 peptides in undepleted and non-enriched human plasma in a single run of 45 min, the equivalent of 9 seconds per assay [192]. Due to the inherent complexity of biofluids, low-abundance protein detection remains a considerable challenge, however, techniques such as front-end immunoaffinity depletion, fractionation and selective enrichment of target proteins/peptides have helped increase the sensitivity [193].

Fig 11 *Schematic diagram of the general LC/MRM-MS workflow with isotopically labelled standards (protein or peptide) for absolute protein quantitation (reproduced from reference* [194]*). The insets show the targeted MS detection of a tryptic peptide column effluent (upper panel) and a peptide standard curve with an extracted ion chromatogram of the heavy labelled (blue) and unlabelled (red) peptide (lower panel). Legend: SPE- solid phase extraction; LC- liquid chromatography; MS- mass spectrometry; ESI- electron spray ionization*

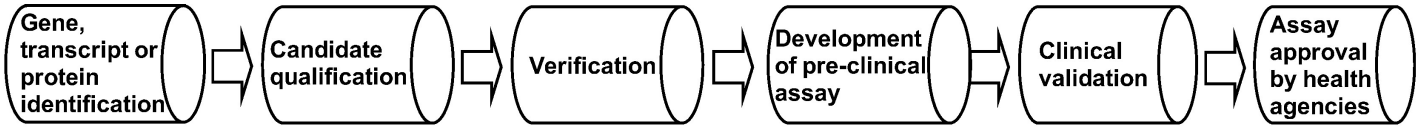


Biomarker discovery

For complex conditions such as cardiovascular disease and cancer, a 'one-biomarker (protein)-one-disease' approach is often not sufficient for adequate disease classification, prediction of therapeutic response, post-treatment therapeutic response monitoring and importantly, for early disease detection [195]. Multiple biomarkers for direct *T. pallidum* detection would also be necessary, owing to the complexity of this multi-stage disease whereby protein expression is expected to vary during the course of infection. Thus far few (published) studies [196] have applied targeted proteomics approaches to infectious disease applications, likely because of challenges related to infection variability and very low predicted pathogen protein concentrations in biofluids such as blood and urine.

The pipeline for proteomics-based biomarker development usually proceeds through three main phases —discovery, verification and analytical validation [197–201]. *Fig 12* outlines the biomarker workflow and inverse relationship between the number of samples analysed and number of analytes as the pipeline progresses. Various MS methods are applied to different phases of the biomarker discovery pipeline: in the initial discovery phase a non-targeted (shotgun) approach is used on a small number of 'ideal' samples, often combined with relative quantification of a large number of proteins using spectral counting methods [202,203]. Proteins of interest that are either differentially expressed between healthy and diseased samples, or in the case of infectious diseases expressed by the infectious agent, are then 'filtered' by performing studies on additional samples from different conditions and/or time points by using a higher-specificity mass spectrometry in a "qualification" step, which can include targeted proteomics (eg. MRM). Thereafter, a smaller number of biomarkers are "validated" in a larger (100-500) set of samples, followed by a clinical validation of final biomarkers on 500-1000s of samples. The last validation step is traditionally performed with an immunoassay platform, such as ELISA. These assays require antibodies targeted against each protein or peptide of interest, a process which is expensive and time-consuming. Further, ELISAs also have limited multiplexing capabilities [204] and can exhibit cross-reactivity, thus are not well-suited to quantify large numbers of candidate biomarker proteins. The current lack of a techniques to verify the multitude of potential biomarkers is termed the "bottleneck" in the biomarker pipeline [200]. During the last years MRM has emerged as a promising alternative to immunoassay-based validation [188].

Fig 12 *Integrated protein biomarker development pipeline, reproduced from reference* [205]. *Proteomics in combination with other '-omics', approaches provides biomarker candidates, which are assessed in the verification phase with an independent cohort of samples. Upon verification and development of pre-clinical assays, the most promising biomarker candidates are validated in thousands of clinical samples. Finally, the diagnostic assay is subjected to a rigorous evaluation and approval by health regulatory agencies. As biomarker candidates proceed through the pipeline, the number of clinical samples increases and low-throughput mass spectrometry-based methods are replaced with high-throughput immunoassays, however, recent applications of MRM assays have promising potential to complement or replace immunoassay validation*



	Gene, transcript or protein identification	Candidate qualification	Verification	Development of pre-clinical assay	Clinical validation	Assay approval by health agencies
Samples:	<10		100s		>1,000	
Analytes:	>1,000	100s	10-100	<10	1-5	
Timeline:	< 6 months	1-3 months	< 6 months	~12 months	< 6 months	
Approach:	• Whole genome sequencing • RNA sequencing • DNA or protein arrays • LC-MS/MS • Thorough sample fractionation • Low throughput	• Selection of candidates based on quantitative or semi-quantitative data • Statistical analysis • Selection of secreted, membrane or tissue - specific proteins • Availability of assays for verification	• LC-SRM or ELISA • Low sample fractionation • Moderate throughput	• Protein expression and purification • Development of monoclonal antibodies • Immunoassay validation	• ELISA • No fractionation • High throughput	

Plasma and serum have long been favoured in protein biomarker analyses since they contain the broadest concentration range of analytes (10 orders of magnitude, ranging from low mg/mL to low pg/mL [195]) that are released, secreted or leaked from neighbouring cells, tissues or microrganisms. Serum and plasma are both derived from whole blood, with plasma collection taking place in the presence of an anti-coagulant (e.g. ethylenediaminetetraacetic acid (EDTA)). Serum is the fluid that remains after coagulation and centrifugation, this process can contribute to variability, including the concomitant removal of non-specific analytes and cell lysis, therefore plasma has been recommended for use in MS-based investigations [206]. Biomarker discovery in urine has also been gaining ground, owing to its (non-invasive) ease of collection and function as a concentrated 'mirror' of systemic processes. Mermelekas *et al.* [207] recently reviewed these methods and advances in this area.

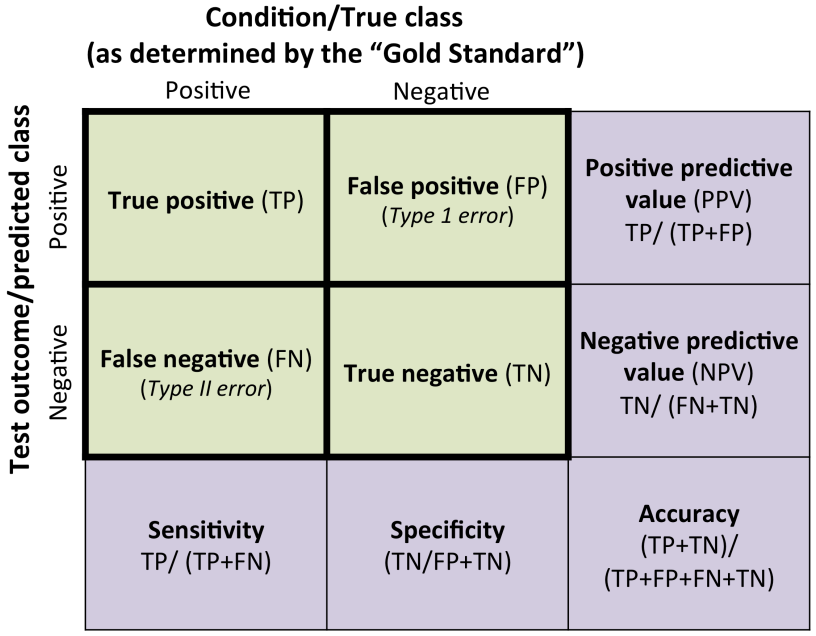
Work described in this thesis related to syphilis biomarker discovery (**chapters 4** & **5**) follows the first steps of the discovery pipeline.

Evaluation of diagnostic tests and biomarker performance

Diagnostic methods must be accurate, simple and affordable for the population in which they are intended, in addition to providing results in a timely manner to facilitate the administration of effective control measures, especially treatment. There are many factors that can influence test performance, including differences in the infectious agent or population (e.g. genetic variability, disease prevalence) and different testing methodologies (automated versus manual; reagent types). Therefore, test evaluations should be performed under a range of conditions that are likely to be encountered in practice and ideally through multicenter trials. A comprehensive report published by the TDR (Special Programme for Research and Training in Tropical Diseases) Diagnostics Evaluation Expert Panel [208] outlines the general principles of the design and conduct of trials to evaluate diagnostic tests for infectious diseases.

The potential of a biomarker (or diagnostic test) is measured primarily in accuracy, defined as the percentage of correct results obtained by a test compared to the results of a reference 'gold-standard' test. Various analytical concepts related test evaluation, such as sensitivity and specificity, are illustrated in *Fig 13.*

Fig 13 *Overview of frequently used concepts used to evaluate the accuracy of a diagnostic test or biomarker*



The SeTPAT study

Clinical observations reported and samples used in **Section II** of this thesis originated from a two-year prospective observational cohort study titled "*Treponema pallidum*-specific proteomic changes in patients with incident syphilis infection: an observational study in Belgium" that was conducted from February 2014 to September 2017 at the ITM in Antwerp, Belgium. This cohort study was part of the "Search for a *Treponema pallidum* Antigen Test (SeTPAT)" project conceptualized by Prof. Chris Kenyon, a study designed to investigate the proteomic, immunological, serological and clinical changes associated with pre- and post-treatment syphilis infection in a way that will ultimately lead to the development of a new diagnostic test of *T. pallidum* antigenaemia. The general aim of this study was to quantify a set of target proteins with the highest diagnostic potential for the diagnosis of initial *T. pallidum* infection and *T. pallidum* persistence.

In total, 120 patients with a new diagnosis of syphilis (active infection with serological test confirmation RPR/TPPA positivity) were prospectively recruited at the ITM's STI/HIV clinic. These individuals were treated according to European syphilis treatment guidelines [86] and followed up by a study doctor (myself) at 3, 6, 9, 12, 18

and 24 months after treatment. Thirty HIV-positive and syphilis negative control patients were also included. A majority of the syphilis positive study patients were HIV positive (N=103, 86 %). All except eight male patients have described themselves as men who have sex with men (MSM) and one woman was included in the study. Approximately 70 % of the study patients had a previous history of a STI infection (besides HIV), of which more than half (72/120; 60 %) presented as syphilis re-infections, in line with previous study findings [15]. Forty percent reported having more than 10 sexual partners during the last year, indicating this was generally a high-risk population. Blood and urine samples were collected during each study visit and were subjected to a variety of serological (e.g. RPR/TPA/TPPA/EIA) and PCR tests. These samples were evaluated by the MRM assay designed to asses candidate *T. pallidum* protein biomarkers (**chapter 5)**, as well as being used for the evaluation of commercial serological assays [90,209] (**chapter 6**), *T. pallidum* genotyping studies [116] and immunological studies designed to evaluate chemokine and cytokine responses during syphilis infection [16,44].

Aims of this thesis and study design

In **Section I** of this thesis, recent epidemiological patterns of syphilis in relation to HIV-infection were studied on two different levels in order to elucidate disease patterns that could be of help to inform prevention efforts. A study comparing antenatal syphilis seroprevelance data from 76 countries in the pre-peak HIV period of the 1990s to peak HIV prevalence revealed that countries with previously high antenatal syphilis prevalences went on to have generalized HIV epidemics (**chapter 2**). Further, user-friendly and streamlined phylogenetic analyses of HIV sequences passively collected in the context of drug resistance testing can enhance epidemiologic analyses of cohorts by highlighting trends that might not otherwise be uncovered by conventional methods. **Chapter 3** describes a study wherein a partial HIV-transmission network was constructed from phylogenetic analyses of 1169 HIV-1 sequences from individuals in follow-up at a single-center HIV cohort in Antwerp, Belgium to study clustering of incident STIs (syphilis/chlamydia/gonorrhea). The resultant phylogram was superimposed with STI-incidence data collected over 17-year period, leading to the identification of possible cases of sexual identity misclassification of MSM as heterosexual males, representing a possible high-risk bridge population.

Diagnosis of *T. pallidum* (re)-infection and post-treatment follow-up to determine pathogen eradication remains onerous for clinicians due to inadequate assays. Thus, **Section II** of this thesis is dedicated to preliminary work related to the identification of candidate protein biomarkers that could be used for antigen test development. **Chapter 4** details the comprehensive characterization of *T. pallidum* proteins expressed during rabbit infection through the application of complementary LC-MS/MS techniques. Insights from this proteome study were subsequently used to shortlist eleven candidate protein biomarkers that were incorporated into a scheduled MRM assay (**chapter 5**). The results from this study were negative, as no *T. pallidum* proteins were detected in undepleted plasma and urine protein extracts from individuals with syphilis. This was likely owing to the very low predicted concentrations of *T. pallidum* proteins in human biofluids, thus future studies should incorporate (antibody) enrichment steps to improve assay sensitivity. Independent evaluation of diagnostic assays with well-characterized clinical samples is key to ensuring test quality. A commercial automated LTIA RPR test was evaluated for clinical appropriateness using sera collected from SeTPAT study participants **(chapter 6)**; this study yielded insights into test performance deficiencies that can be used to improve future test versions. Finally, **chapter 7** brings together the concepts discussed in this thesis, putting them in a broader context with future perspectives.

In summary, the specific aims of this thesis are:

- to assess if the country-level prevalence of antenatal syphilis in the 1990-1999 period was associated with peak HIV prevalence
- to assess if the superimposition of incident syphilis cases on the HIV phylogram of the ITM HIV cohort would reveal clustering of syphilis cases
- to characterize and semi-quantify the *T. pallidum* proteome during rabbit infection to improve our understanding of the biology of *T. pallidum* and as a first step in identifying candidate protein biomarkers for antigen test development
- to investigate the performance of eleven candidate biomarker proteins shortlisted from the *T. pallidum* proteome investigation by designing a targeted proteomics MRM assay to analyze protein extracts from plasma and urine samples from individuals with syphilis
- to evaluate existing diagnostic technologies, namely an automated RPR assay, for clinical appropriateness

References

1. Newman L, Rowley J, Vander Hoorn S, Wijesooriya NS, Unemo M, Low N, et al. Global Estimates of the Prevalence and Incidence of Four Curable Sexually Transmitted Infections in 2012 Based on Systematic Review and Global Reporting. PLoS One. 2015;10(12):e0143304.
2. Newman L, Kamb M, Hawkes S, Gomez G, Say L, Seuc A, et al. Global estimates of syphilis in pregnancy and associated adverse outcomes: analysis of multinational antenatal surveillance data. [Internet]. Vol. 10, PLoS medicine. Public Library of Science; 2013. p. e1001396.
3. World Health Organization. Percentage of antenatal care attendees postitive for syphilis (latest reported data since 2008) [Internet]. 2017.
4. Kenyon CR, Osbak K, Tsoumanis A. The Global Epidemiology of Syphilis in the Past Century – A Systematic Review Based on Antenatal Syphilis Prevalence. Small PLC, editor. PLoS Negl Trop Dis. 2016;10(5):e0004711.
5. Chesson HW, Dee TS, Aral SO. Aids mortality may have contributed to the decline in syphilis rates in the United States in the 1990s. Sex Transm Dis. 30(5):419–24.
6. Kenyon CR, Osbak K. How many MSM in Europe could benefit from PrEP--a 9 billion Euro question? Int J STD AIDS. 2015;26(13):988–90.
7. Kenyon CR, Osbak K, Chico RM. What underpins the decline in syphilis in Southern and Eastern Africa? An exploratory ecological analysis. Int J Infect Dis. 2014;29:54–61.
8. World Health Organization. Global Health Observatory Map Gallery [Internet].
9. Pettifor A, Walsh J, Wilkins V, Raghunathan P. How effective is syndromic management of Stds?: A review of current studies. Sex Transm Dis. 27(7):371–85.
10. Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2015. Atlanta; 2016.
11. Abara WE, Hess KL, Neblett Fanfair R, Bernstein KT, Paz-Bailey G. Syphilis Trends among Men Who Have Sex with Men in the United States and Western Europe: A Systematic Review of Trend Studies Published between 2004 and 2015. Lima VD, editor. PLoS One. 2016;11(7):e0159309.
12. Jiang N, Gong X, Yue X. [Estimation of syphilis epidemic through application of workbook method among populations aged from 15 to 49 years old in China in 2011]. Zhonghua Yu Fang Yi Xue Za Zhi. 2014;48(8):693–8.
13. Read P, Fairley CK, Chow EPF. Increasing trends of syphilis among men who have sex with men in high income countries. Sex Heal. 2015;12(2):155–63.
14. World Health Organization. Men who have sex with men (MSM) with active syphilis data by country [Internet]. 2017.
15. Kenyon C, Lynen L, Florence E, Caluwaerts S, Vandenbruaene M, Apers L, et al. Syphilis reinfections pose problems for syphilis diagnosis in Antwerp, Belgium - 1992 to 2012. Euro Surveill. 2014;19(45):20958.
16. Kenyon, CR; Tsoumanis, A; Osbak, K; van Esbroeck, M; Florence, E; Crucitti, T; Kestens L. Repeat syphilis has a different immune response compared to initial syphilis. Sex Transm Infect. 2017.
17. Abara WE, Oraka E, Jeffries WL, Chavez P, Nasrullah M, DiNenno E. Prevalence and correlates of condom use among sexually active men who have sex with men in the United States: findings from the National Survey of Family Growth, 2002, 2006-10 and 2011-13. Sex Health. 2017;
18. Hess KL, Crepaz N, Rose C, Purcell D, Paz-Bailey G. Trends in Sexual Behavior Among Men Who have Sex with Men (MSM) in High-Income Countries, 1990–2013: A Systematic Review. AIDS Behav. 2017;
19. Yang CJ, Chang SY, Wu BR, Yang SP, Liu WC, Wu PY, et al. Unexpectedly high prevalence of Treponema pallidum infection in the oral cavity of human immunodeficiency virus-infected patients with early syphilis who had engaged in unprotected sex practices. Clin Microbiol Infect. 2015;21(8):787.e1-787.e7.
20. Kenyon CR, Osbak K. The Prevalence of Syphilis Is Associated with the Prevalence of Male Point-Concurrency: An Ecological Analysis. World J AIDS. 2015;5(2):131–9.
21. Gray RH, Wawer MJ. Reassessing the hypothesis on STI control for HIV prevention. Lancet. 2008;371(9630):2064–5.

22. Idsoe O, Guthe T. The rise and fall of the treponematoses. I. Ecological aspects and international trends. in venereal syphilis. Br J Vener Dis. 1967;43(4):227–43.
23. Ward H, Rönn M. Contribution of sexually transmitted infections to the sexual transmission of HIV. Curr Opin HIV AIDS. 2010;5(4):305–10.
24. Kenyon C, Colebunders R, Hens N. Determinants of generalized herpes simplex virus-2 epidemics: the role of sexual partner concurrency. Int J STD AIDS. 2013;24(5):375–82.
25. Kenyon CR, Tsoumanis A, Osbak K. Strong Country Level Correlation between Syphilis and HSV-2 Prevalence. J Sex Transm Dis. 2016;2016:1–6.
26. Weiss RA. How does HIV cause AIDS? Science. 1993;260(5112):1273–9.
27. Masur H, Michelis MA, Greene JB, Onorato I, Stouwe RA, Holzman RS, et al. An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. N Engl J Med. 1981;305(24):1431–8.
28. Joint United Nations Program on HIV/AIDS. UNAIDS data 2017. 2017.
29. Chow EPF, Wilson DP, Zhang L, Wang L, Guo H. HIV and Syphilis Co-Infection Increasing among Men Who Have Sex with Men in China: A Systematic Review and Meta-Analysis. Thorne C, editor. PLoS One. 2011;6(8):e22768.
30. Kalichman SC, Pellowski J, Turner C. Prevalence of sexually transmitted co-infections in people living with HIV/AIDS: systematic review with implications for using HIV treatments for prevention. Sex Transm Infect. 2011;87(3):183–90.
31. Wu Z, Xu J, Liu E, Mao Y, Xiao Y, Sun X, et al. HIV and Syphilis Prevalence Among Men Who Have Sex With Men: A Cross-Sectional Survey of 61 Cities in China. Clin Infect Dis. 2013;57(2):298–309.
32. Fleming DT, Wasserheit JN. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. Sex Transm Infect. 1999;75(1):3–17.
33. Darrow WW, Echenberg DF, Jaffe HW, O'Malley PM, Byers RH, Getchell JP, et al. Risk factors for human immunodeficiency virus (HIV) infections in homosexual men. Am J Public Health. 1987;77(4):479–83.
34. Girometti N, Gutierrez A, Nwokolo N, McOwan A, Whitlock G. High HIV incidence in men who have sex with men following an early syphilis diagnosis: is there room for pre-exposure prophylaxis as a prevention strategy? Sex Transm Infect. 2017;93(5):320–2.
35. Solomon MM, Mayer KH, Glidden D V., Liu AY, McMahan VM, Guanira J V., et al. Syphilis Predicts HIV Incidence Among Men and Transgender Women Who Have Sex With Men in a Preexposure Prophylaxis Trial. Clin Infect Dis. 2014;59(7):1020–6.
36. Sheffield JS, Wendel GD, McIntire DD, Norgard M V. Effect of genital ulcer disease on HIV-1 coreceptor expression in the female genital tract. J Infect Dis. 2007;196(10):1509–16.
37. Salazar JC, Cruz AR, Pope CD, Valderrama L, Trujillo R, Saravia NG, et al. Treponema pallidum elicits innate and adaptive cellular immune responses in skin and blood during secondary syphilis: a flow-cytometric analysis. J Infect Dis. 2007;195(6):879–87.
38. Lafond RE, Lukehart SA. Biological Basis for Syphilis. Clin Microbiol Rev. 2006;19(1):29–49.
39. Jarzebowski W, Caumes E, Dupin N, Farhi D, Lascaux A-S, Piketty C, et al. Effect of Early Syphilis Infection on Plasma Viral Load and CD4 Cell Count in Human Immunodeficiency Virus–Infected Men. Arch Intern Med. 2012;172(16):1237.
40. Rompalo AM, Lawlor J, Seaman P, Quinn TC, Zenilman JM, Hook EW. Modification of syphilitic genital ulcer manifestations by coexistent HIV infection. Sex Transm Dis. 2001;28(8):448–54.
41. Centers for Disease Control and Prevention (CDC). Symptomatic early neurosyphilis among HIV-positive men who have sex with men--four cities, United States, January 2002-June 2004. MMWR Morb Mortal Wkly Rep. 2007;56(25):625–8.
42. Seña AC, Zhang X-H, Li T, Zheng H-P, Yang B, Yang LG, et al. A systematic review of syphilis serological treatment outcomes in HIV-infected and HIV-uninfected persons: rethinking the significance of serological non-responsiveness and the serofast state after therapy. BMC Infect Dis. 2015;15(1):479.
43. Marra CM, Tantalo LC, Sahi SK, Dunaway SB, Lukehart SA. Reduced *Treponema pallidum*-Specific Opsonic Antibody Activity in HIV-Infected Patients With Syphilis. J Infect Dis. 2016;213(8):1348–54.
44. Kenyon C, Osbak KK, Crucitti T, Kestens L. The immunological response to syphilis differs by HIV

status; a prospective observational cohort study. BMC Infect Dis. 2017;17(1):111.
45. Yang Z, Rannala B. Molecular phylogenetics: principles and practice. Nat Rev Genet. 2012;13(5):303–14.
46. German D, Grabowski MK, Beyrer C. Enhanced use of phylogenetic data to inform public health approaches to HIV among men who have sex with men. Sex Health. 2016;14(1):89.
47. Grabowski MK, Redd AD. Molecular tools for studying HIV transmission in sexual networks. Curr Opin HIV AIDS. 2014;9(2):126–33.
48. Brenner BG, Wainberg MA. Future of phylogeny in HIV prevention. J Acquir Immune Defic Syndr. 2013;63 Suppl 2:S248-54.
49. Chalmet K, Staelens D, Blot S, Dinakis S, Pelgrom J, Plum J, et al. Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. BMC Infect Dis. 2010;10(1):262.
50. Homes, KK; Sparling PF; Stamm W et al. Clinical Manifestations of Syphilis. In: Homes, KK; Sparling PF; Stamm W et al., editor. Sexually Transmitted Diseases. 4th ed. New York: McGraw Hill; 2008. p. 661–84.
51. Hook EW, Marra CM. Acquired Syphilis in Adults. N Engl J Med. 1992;326(16):1060–9.
52. Hook EW. Syphilis. Lancet. 2016;6736(16).
53. Cuerda-Galindo E, Sierra-Valenti X, González-López E, López-Muñoz F. La sífilis y la experimentación en humanos: perspectiva histórica y reflexiones éticas. De la Segunda Guerra Mundial a la actualidad. Actas Dermosifiliogr. 2014;105(9):847–53.
54. Tipple C. Impact of HIV-1 infection on the clinical presentation of syphilis in men who have sex with men. Sex Heal. 2015;12(2):110–8.
55. Magnuson HJ, Thomas EW, Olansky S, Kaplan BI, De Mello L, Cutler JC. Inoculation syphilis in human volunteers. Medicine (Baltimore). 1956;35(1):33–82.
56. Merritt, HH; Adams, RD; Solomon H. Neurosyphilis. New York, New York, USA: Oxford University Press; 1946.
57. Clark EG, Danbolt N. The Oslo study of the natural history of untreated syphilis. J Chronic Dis. 1955;2(3):311–44.
58. Stokes, JH; Beerman, H; Ingram N. Modern Clinical Syphilology. Philadelphia, Pennsylvania, USA: W.B. Sanders; 1944.
59. Ho EL, Lukehart SA. Syphilis: using modern approaches to understand an old disease. J Clin Invest. 2011;121(12):4584–92.
60. Pinto M, Antelo M, Ferreira R, Azevedo J, Santo I, Borrego MJ, et al. A retrospective cross-sectional quantitative molecular approach in biological samples from patients with syphilis. Microb Pathog. 2017;104:296–302.
61. Lukehart SA, Hook EW, Baker-Zander SA, Collier AC, Critchlow CW, Handsfield HH. Invasion of the Central Nervous System by Treponema pallidum: Implications for Diagnosis and Treatment. Ann Intern Med. 1988;109(11):855.
62. Oliver SE, Aubin M, Atwell L, Matthias J, Cope A, Mobley V, et al. Ocular Syphilis ? Eight Jurisdictions, United States, 2014?2015. MMWR Morb Mortal Wkly Rep. 2016;65(43):1185–8.
63. Marra CM, Deutsch R, Collier AC, Morgello S, Letendre S, Clifford D, et al. Neurocognitive impairment in HIV-infected individuals with previous syphilis. Int J STD AIDS. 2013;24(5):351–5.
64. Radolf JD. Treponema. In: Medical Microbiology. University of Texas Medical Branch at Galveston; 1996.
65. Tipple C, Taylor GP. Syphilis testing, typing, and treatment follow-up. Curr Opin Infect Dis. 2015;28(1):53–60.
66. Miller JN. Value and limitations of nontreponemal and treponemal tests in the laboratory diagnosis of syphilis. Clin Obstet Gynecol. 1975;18(1):191–203.
67. Gayet-Ageron A, Lautenschlager S, Ninet B, Perneger T V, Combescure C. Sensitivity, specificity and likelihood ratios of PCR in the diagnosis of syphilis: a systematic review and meta-analysis. Sex Transm Infect. 2013;89(3):251–6.
68. Wasserman, A; Neisser, A; Bruck C. Eine serodiagnostische reaction bei Syphilis. Dtsch Med Wochenschr. 1906;(32):745–76.
69. Belisle JT, Brandt ME, Radolf JD, Norgard M V. Fatty acids of Treponema pallidum and Borrelia

burgdorferi lipoproteins. J Bacteriol. 1994;176(8):2151–7.
70. Matthews HM, Yang TK, Jenkin HM. Unique lipid composition of Treponema pallidum (Nichols virulent strain). Infect Immun. 1979;24(3):713–9.
71. Ratnam S. The laboratory diagnosis of syphilis. Can J Infect Dis Med Microbiol. 2005;16(1):45–51.
72. Gupta SM, Bala M, Muralidhar S, Ray K. Evaluation of test results of microbiology laboratories of North India for standard tests for syphilis under an external quality assurance scheme. Eur J Clin Microbiol Infect Dis. 2009;28(5):461–8.
73. Muller I, Brade V, Hagedorn H-J, Straube E, Schorner C, Frosch M, et al. Is Serological Testing a Reliable Tool in Laboratory Diagnosis of Syphilis- Meta-Analysis of Eight External Quality Control Surveys Performed by the German Infection Serology Proficiency Testing Program. J Clin Microbiol. 2006;44(4):1335–41.
74. Huh HJ, Chung J-W, Park SY, Chae SL. Comparison of Automated Treponemal and Nontreponemal Test Algorithms as First-Line Syphilis Screening Assays. Ann Lab Med. 2016;36(1):23.
75. Eagle H. The Laboratory Diagnosis of Syphilis: The Theory, Technic, and Clinical Interpretation of the Wassermann and Flocculation Tests with Serum and Spinal Fluid. St. Louis: CV Mosby; 1937. 199 p.
76. Berkowitz K, Baxi L, Fox HE. False-negative syphilis screening: the prozone phenomenon, nonimmune hydrops, and diagnosis of syphilis during pregnancy. Am J Obstet Gynecol. 1990;163(3):975–7.
77. Jurado RL, Campbell J, Martin PD. Prozone Phenomenon in Secondary Syphilis. Arch Intern Med. 1993;153(21):2496.
78. Berkowitz KM, Stampf K, Baxi L, Fox HE. False negative screening tests for syphilis in pregnant women. N Engl J Med. 1990;322(4):270–1.
79. Hernández-Aguado I, Bolumar F, Moreno R, Pardo FJ, Torres N, Belda J, et al. False-positive tests for syphilis associated with human immunodeficiency virus and hepatitis B virus infection among intravenous drug abusers. Valencian Study Group on HIV Epidemiology. Eur J Clin Microbiol Infect Dis. 1998;17(11):784–7.
80. Liu F, Liu LLL, Guo XJJ, Xi Y, Lin LRR, Zhang HLL, et al. Characterization of the classical biological false-positive reaction in the serological test for syphilis in the modern era. Int Immunopharmacol. 2014;20(2):331–6.
81. Monath TP, Frey SE. Possible autoimmune reactions following smallpox vaccination: the biologic false positive test for syphilis. Vaccine. 2009;27(10):1645–50.
82. Lam TK, Lau HY, Lee YP, Fung SM, Leung WL, Kam KM. Use of the INNO-LIA syphilis score assay in the resolution of discordant positive screening enzyme immunoassay results for the serological diagnosis of syphilis. Int J STD AIDS. 2014;25(1):52–6.
83. Malm K, Andersson S, Fredlund H, Norrgren H, Biague A, Månsson F, et al. Analytical evaluation of nine serological assays for diagnosis of syphilis. J Eur Acad Dermatology Venereol. 2015;29(12):2369–76.
84. Report MW. Discordant results from reverse sequence syphilis screening--five laboratories, United States, 2006-2010. MMWR Morb Mortal Wkly Rep. 2011;60(5):133–7.
85. Seña ACC, White BLL, Sparling PFF. Novel Treponema pallidum serologic tests: a paradigm shift in syphilis screening for the 21st century. Clin Infect Dis. 2010;51(6):700–8.
86. Janier M, Hegyi V, Dupin N, Unemo M, Tiplica GS, Potočnik M, et al. 2014 European guideline on the management of syphilis. J Eur Acad Dermatology Venereol. 2014;28(12):1581–93.
87. Tong ML, Lin LR, Liu LLLL, Zhang HL, Huang SJ, Chen YY, et al. Analysis of 3 algorithms for syphilis serodiagnosis and implications for clinical management. Clin Infect Dis. 2014;58(8):1116–24.
88. Rourk AR, Nolte FS, Litwin CM. Performance Characteristics of the Reverse Syphilis Screening Algorithm in a Population With a Moderately High Prevalence of Syphilis. Am J Clin Pathol. 2016;146(5):572–7.
89. Chen B, Peng X, Xie T, Jin C, Liu F, Wu N. The tradition algorithm approach underestimates the prevalence of serodiagnosis of syphilis in HIV-infected individuals. Caimano MJ, editor. PLoS Negl Trop Dis. 2017;11(7):e0005758.
90. Osbak K, Abdellati S, Tsoumanis A, Esbroeck M Van, Crucitti T, Kenyon C. P1.46 Comparison of two enzyme immunoassays for the detection of igg and igm anti- *treponema pallidum antibodies*. In:

Systems Biology and Novel Technologies For Molecular Analysis and Diagnosis. BMJ Publishing Group Ltd; 2017. p. A61.2-A61.

91. Brown ST, Zaidi A, Larsen SA, Reynolds GH. Serological response to syphilis treatment. A new analysis of old data. JAMA. 1985;253(9):1296–9.
92. Seña AC, Wolff M, Behets F, Van Damme K, Martin DH, Leone P, et al. Response to therapy following retreatment of serofast early syphilis patients with benzathine penicillin. Clin Infect Dis. 2013;56(3):420–2.
93. Li J, Wang LN, Zheng HY. Predictors of serological cure and serofast state after treatment in HIV-negative patients with early syphilis in China. Sex Transm Infect. 2013;89(1):69.
94. Sena AC, Wolff M, Martin DH, Behets F, Van Damme K, Leone P, et al. Predictors of serological cure and Serofast State after treatment in HIV-negative persons with early syphilis. Clin Infect Dis. 2011;53(11):1092–9.
95. Zhang RL, Wang QQ, Zhang JP, Yang LJ. Molecular subtyping of Treponema pallidum and associated factors of serofast status in early syphilis patients: Identified novel genotype and cytokine marker. Stevenson B, editor. PLoS One. 2017;12(4):e0175477.
96. Courjon J, Hubiche T, Dupin N, Grange PA, Del Giudice P. Clinical aspects of syphilis reinfection in HIV-infected patients. Dermatology. 2015;230(4):302–7.
97. Chow EPF, Callander D, Fairley CK, Zhang L, Donovan B, Guy R, et al. Increased Syphilis Testing of Men Who Have Sex With Men: Greater Detection of Asymptomatic Early Syphilis and Relative Reduction in Secondary Syphilis. Clin Infect Dis. 2017;
98. Workowski KA, Bolan GA, Centers for Disease Control and Prevention. Sexually transmitted diseases treatment guidelines, 2015. MMWR Recomm reports Morb Mortal Wkly report Recomm reports. 2015;64(RR-03):1–137.
99. World Health Organization. WHO guidelines for the treatment of treponema pallidum (Syphilis). 51 p.
100. Butler T. The Jarisch–Herxheimer Reaction After Antibiotic Treatment of Spirochetal Infections: A Review of Recent Cases and Our Understanding of Pathogenesis. Am J Trop Med Hyg. 2017;96(1):46–52.
101. Stamm LV. Syphilis: antibiotic treatment and resistance. Epidemiol Infect. 2014;143(8):1567–74.
102. Lobanovska M, Pilla G. Penicillin's Discovery and Antibiotic Resistance: Lessons for the Future? Yale J Biol Med. 2017;90(1):135–45.
103. Dunlop EM. Survival of treponemes after treatment: comments, clinical conclusions, and recommendations. Genitourin Med. 1985;61(5):293–301.
104. Cha JY, Ishiwata A, Mobashery S. A novel beta-lactamase activity from a penicillin-binding protein of Treponema pallidum and why syphilis is still treatable with penicillin. J Biol Chem. 2004;279(15):14917–21.
105. Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, Dodson R, et al. Complete Genome Sequence of Treponema pallidum, the Syphilis Spirochete. Science (80-). 1998;281(5375):375–88.
106. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000;405(6784):299–304.
107. Liu H, Han Y, Chen X, Bai L, Guo S, Li L, et al. Comparison of efficacy of treatments for early syphilis: A systematic review and network meta-analysis of randomized controlled trials and observational studies. De Socio GV, editor. PLoS One. 2017;12(6):e0180001.
108. Centers for Disease Control (CDC). Sexually transmitted diseases treatment guidelines 1982. MMWR Morb Mortal Wkly Rep. 1982;31 Suppl 2:33S–60S.
109. Katz KA, Klausner JD. Azithromycin resistance in Treponema pallidum. Curr Opin Infect Dis. 2008;21(1):83–91.
110. Stamm LV, Bergen HL. A point mutation associated with bacterial macrolide resistance is present in both 23S rRNA genes of an erythromycin-resistant Treponema pallidum clinical isolate. Antimicrob Agents Chemother. 2000;44(3):806–7.
111. Matejková P, Flasarová M, Zákoucká H, Borek M, Kremenová S, Arenberger P, et al. Macrolide treatment failure in a case of secondary syphilis: a novel A2059G mutation in the 23S rRNA gene of Treponema pallidum subsp. pallidum. J Med Microbiol. 2009;58(Pt 6):832–6.
112. Wu BR, Yang CJ, Tsai MS, Lee KY, Lee NY, Huang WC, et al. Multicentre surveillance of prevalence

of the 23S rRNA A2058G and A2059G point mutations and molecular subtypes of Treponema pallidum in Taiwan, 2009–2013. Clin Microbiol Infect. 2014;20(8):802–7.
113. Read P, Tagg KA, Jeoffreys N, Guy RJ, Gilbert GL, Donovan B. Treponema pallidum Strain Types and Association with Macrolide Resistance in Sydney, Australia: New TP0548 Gene Types Identified. J Clin Microbiol. 2016;54(8):2172–4.
114. Chen XS, Yin YP, Wei WH, Wang HC, Peng RR, Zheng HP, et al. High prevalence of azithromycin resistance to Treponema pallidum in geographically different areas in China. Clin Microbiol Infect. 2013;19(10):975–9.
115. Stamm LV. Global challenge of antibiotic-resistant Treponema pallidum. Antimicrob Agents Chemother. 2010;54(2):583–9.
116. Mikalová L, Grillová L, Osbak K, Strouhal M, Kenyon C, Crucitti T, et al. Molecular Typing of Syphilis-Causing Strains Among Human Immunodeficiency Virus-Positive Patients in Antwerp, Belgium. Sex Transm Dis. 2017;44(6):376–9.
117. Ghanem KG, Moore RD, Rompalo AM, Erbelding EJ, Jonathan M, Gebo KA. Neurosyphilis in a clinical cohort of HIV-1-infected patients. 2008;22(10):1145–51.
118. Flood JM, Weinstock HS, Guroy ME, Bayne L, Simon RP, Bolan G. Neurosyphilis during the AIDS epidemic, San Francisco, 1985-1992. J Infect Dis. 1998;177(4):931–40.
119. Lawrence D, Cresswell F, Whetham J, Fisher M. Syphilis treatment in the presence of HIV. Curr Opin Infect Dis. 2015;28(1):1.
120. Costa-Silva M, Azevedo C, Azevedo F, Lisboa C. Early syphilis treatment in HIV-infected patients: single dose vs. three doses of benzathine penicillin G. J Eur Acad Dermatology Venereol. 2016;30(10):1805–9.
121. Ganesan A, Mesner O, Okulicz JF, O'Bryan T, Deiss RG, Lalani T, et al. A Single Dose of Benzathine Penicillin G Is as Effective as Multiple Doses of Benzathine Penicillin G for the Treatment of HIV-Infected Persons With Early Syphilis. Clin Infect Dis. 2015;60(4):653–60.
122. Yang CJ, Lee NY, Chen TC, Lin YH, Liang SH, Lu PL, et al. One Dose versus Three Weekly Doses of Benzathine Penicillin G for Patients Co-Infected with HIV and Early Syphilis: A Multicenter, Prospective Observational Study. Tucker JD, editor. PLoS One. 2014;9(10):e109667.
123. Tsai JC, Lin YH, Lu PL, Shen NJ, Yang CJ, Lee NY, et al. Comparison of serological response to doxycycline versus benzathine penicillin G in the treatment of early syphilis in HIV-infected patients: A multi-center observational study. PLoS One. 2014;9(10):1–7.
124. Baral S, Logie CH, Grosso A, Wirtz AL, Beyrer C. Modified social ecological model: a tool to guide the assessment of the risks and risk contexts of HIV epidemics. BMC Public Health. 2013;13(1):482.
125. Hawkes S, Matin N, Broutet N, Low N. Effectiveness of interventions to improve screening for syphilis in pregnancy: a systematic review and meta-analysis. Lancet Infect Dis. 2011;11(9):684–91.
126. World Health Organization. Technical consultation on preventing and managing global stock outs of medicines. In Geneva; 2015.
127. Centers for Disease Control and Prevention (CDC). Bicillin-LA (benzathine pencillin G) shortage. [Internet]. Sexually Transmitted Diseases (STDs). 2016.
128. Taylor MM, Nurse-Findlay S, Zhang X, Hedman L, Kamb ML, Broutet N, et al. Estimating Benzathine Penicillin Need for the Treatment of Pregnant Women Diagnosed with Syphilis during Antenatal Care in High-Morbidity Countries. Gantt S, editor. PLoS One. 2016;11(7):e0159483.
129. Quetel C. History of Syphilis. Cambridge, UK: Polity Press; 1990.
130. Down I, Wilson DP, McCann PD, Gray R, Hoare A, Bradley J, et al. Increasing gay men?s testing rates and enhancing partner notification can reduce the incidence of syphilis. Sex Health. 2012;9(5):472.
131. Lithgow KV., Cameron CE. Vaccine development for syphilis. Expert Rev Vaccines. 2017;16(1):37–44.
132. Lithgow KV., Hof R, Wetherell C, Phillips D, Houston S, Cameron CE. A defined syphilis vaccine candidate inhibits dissemination of Treponema pallidum subspecies pallidum. Nat Commun. 2017;8:14273.
133. Champredon D, Cameron CE, Smeija M, Dushoff J. Epidemiological impact of a syphilis vaccine: a simulation study. Epidemiol Infect. 2016;144(15):3244–52.
134. Izard J, Renken C, Hsieh C-E, Desrosiers DC, Dunham-Ems S, La Vake C, et al. Cryo-electron

tomography elucidates the molecular architecture of Treponema pallidum, the syphilis spirochete. J Bacteriol. 2009;191(24):7566–80.

135. Jepsen OB, Hougen KH, Birch-Andersen A. Electron microscopy of treponema pallidum Nichols. Acta Pathol Microbiol Scand. 1968;74(2):241–58.
136. Radolf JD, Deka RK, Anand A, Šmajs D, Norgard M V, Frank Yang X. Treponema pallidum, the syphilis spirochete: making a living as a stealth pathogen. Nat Publ Gr. 2016;14:744–59.
137. Walker EM, Zampighi GA, Blanco DR, Miller JN, Lovett MA. Demonstration of rare protein in the outer membrane of Treponema pallidum subsp. pallidum by freeze-fracture analysis. J Bacteriol. 1989;171(9):5005–11.
138. Radolf JD, Norgard M V, Schulz WW. Outer membrane ultrastructure explains the limited antigenicity of virulent Treponema pallidum. Proc Natl Acad Sci U S A. 1989;86(6):2051–5.
139. Zhang JR, Hardham JM, Barbour AG, Norris SJ. Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. Cell. 1997;89(2):275–85.
140. Barbour AG, Burman N, Carter CJ, Kitten T, Bergström S. Variable antigen genes of the relapsing fever agent Borrelia hermsii are activated by promoter addition. Mol Microbiol. 1991;5(2):489–93.
141. Cox DL, Luthra A, Dunham-Ems S, Desrosiers DC, Salazar JC, Caimano MJ, et al. Surface immunolabeling and consensus computational framework to identify candidate rare outer membrane proteins of Treponema pallidum. Infect Immun. 2010;78(12):5178–94.
142. Centurion-Lara A, Castro C, Barrett L, Cameron C, Mostowfi M, Van Voorhis WC, et al. Treponema pallidum major sheath protein homologue Tpr K is a target of opsonic antibody and the protective immune response. J Exp Med. 1999;189(4):647–56.
143. Centurion-Lara A, Godornes C, Castro C, Van Voorhis WC, Lukehart SA. The tprK gene is heterogeneous among Treponema pallidum strains and has multiple alleles. Infect Immun. 2000;68(2):824–31.
144. Stamm LV, Bergen HL. The sequence-variable, single-copy tprK gene of Treponema pallidum Nichols strain UNC and Street strain 14 encodes heterogeneous TprK proteins. Infect Immun. 2000;68(11):6482–6.
145. Giacani L, Molini BJ, Kim EY, Godornes BC, Leader BT, Tantalo LC, et al. Antigenic variation in Treponema pallidum: TprK sequence diversity accumulates in response to immune pressure during experimental syphilis. J Immunol. 2010;184(7):3822–9.
146. Reid TB, Molini BJ, Fernandez MC, Lukehart SA. Antigenic variation of TprK facilitates development of secondary syphilis. Infect Immun. 2014;82(12):4959–67.
147. LaFond RE, Centurion-Lara A, Godornes C, Van Voorhis WC, Lukehart SA. TprK sequence diversity accumulates during infection of rabbits with Treponema pallidum subsp. pallidum Nichols strain. Infect Immun. 2006;74(3):1896–906.
148. Centurion-Lara A, LaFond RE, Hevner K, Godornes C, Molini BJ, Van Voorhis WC, et al. Gene conversion: a mechanism for generation of heterogeneity in the tprK gene of Treponema pallidum during infection. Mol Microbiol. 2004;52(6):1579–96.
149. Giacani L, Brandt SL, Puray-Chavez M, Reid TB, Godornes C, Molini BJ, et al. Comparative investigation of the genomic regions involved in antigenic variation of the TprK antigen among treponemal species, subspecies, and strains. J Bacteriol. 2012;194(16):4208–25.
150. Fieldsteel AH, Cox DL, Moeckli RA. Cultivation of virulent Treponema pallidum in tissue culture. Infect Immun. 1981;32(2):908–15.
151. Radolf, JD; Hazlett, KRO; Lukehart S. Pathogenesis of Syphilis. In: Radolf, JD; Lukehart S, editor. Pathogenic Treponemes: Cellular and Molecular Biology. Norfolk, UK: Caister Academic Press; 2006. p. 197–236.
152. Stamm LV, Gherardini FC, Parrish EA, Moomaw CR. Heat shock response of spirochetes. Infect Immun. 1991;59(4):1572–5.
153. Fieldsteel AH, Cox DL, Moeckli RA. Further studies on replication of virulent Treponema pallidum in tissue cultures of Sf1Ep cells. Infect Immun. 1982;35(2):449–55.
154. Norris SJ. In vitro cultivation of Treponema pallidum: independent confirmation. Infect Immun. 1982;36(1):437–9.
155. Baseman JB, Nichols JC, Rumpp JW, Hayes NS. Purification of Treponema pallidum from infected rabbit tissue: resolution into two treponemal populations. Infect Immun. 1974;10(5):1062–7.

156. Lukehart SA, Marra CM. Isolation and laboratory maintenance of Treponema pallidum. Curr Protoc Microbiol. 2007;Chapter 12.
157. Chesney AM, Schipper GJ. The effect of the method of inoculation upon the course of experimental syphilis in the rabbit. Am J Syph Gonorrhea Vener Dis. 1950;34(1):18–24.
158. Baker-Zander S, Sell S. A histopathologic and immunologic study of the course of syphilis in the experimentally infected rabbit. Demonstration of long-lasting cellular immunity. Am J Pathol. 1980;101(2):387–414.
159. Hanff PA, Norris SJ, Lovett MA, Miller JN. Purification of Treponema pallidum, Nichols strain, by Percoll density gradient centrifugation. Sex Transm Dis. 11(4):275–86.
160. Giacani L, Lukehart SA. The endemic treponematoses. Clin Microbiol Rev. 2014;27(1):89–115.
161. Cejková D, Zobaníková M, Chen L, Pospíšilová P, Strouhal M, Qin X, et al. Whole genome sequences of three Treponema pallidum ssp. pertenue strains: yaws and syphilis treponemes differ in less than 0.2% of the genome sequence. PLoS Negl Trop Dis. 2012;6(1):e1471.
162. Strouhal M, Smajs D, Matejková P, Sodergren E, Amin AG, Howell JK, et al. Genome differences between Treponema pallidum subsp. pallidum strain Nichols and T. paraluiscuniculi strain Cuniculi A. Infect Immun. 2007;75(12):5859–66.
163. Smajs D, Norris SJ, Weinstock GM. Genetic diversity in Treponema pallidum: implications for pathogenesis, evolution and molecular diagnostics of syphilis and yaws. Infect Genet Evol. 2012;12(2):191–202.
164. Mitjà O, Šmajs D, Bassat Q. Advances in the diagnosis of endemic treponematoses: yaws, bejel, and pinta. PLoS Negl Trop Dis. 2013;7(10):e2283.
165. Pětrošová H, Pospíšilová P, Strouhal M, Čejková D, Zobaníková M, Mikalová L, et al. Resequencing of Treponema pallidum ssp. pallidum Strains Nichols and SS14: Correction of Sequencing Errors Resulted in Increased Separation of Syphilis Treponeme Subclusters. Dong Q, editor. PLoS One. 2013;8(9):e74319.
166. Zobaníková M, Mikolka P, Cejková D, Pospíšilová P, Chen L, Strouhal M, et al. Complete genome sequence of Treponema pallidum strain DAL-1. Stand Genomic Sci. 2012;7(1):12–21.
167. Giacani L, Jeffrey BM, Molini BJ, Le HT, Lukehart SA, Centurion-Lara A, et al. Complete genome sequence and annotation of the Treponema pallidum subsp. pallidum Chicago strain. J Bacteriol. 2010;192(10):2645–6.
168. Matejková P, Strouhal M, Smajs D, Norris SJ, Palzkill T, Petrosino JF, et al. Complete genome sequence of Treponema pallidum ssp. pallidum strain SS14 determined with oligonucleotide arrays. BMC Microbiol. 2008;8(1):76.
169. Tong ML, Zhao Q, Liu LL, Zhu XZ, Gao K, Zhang HL, et al. Whole genome sequence of the Treponema pallidum subsp. pallidum strain Amoy: An Asian isolate highly similar to SS14. Rockey D, editor. PLoS One. 2017;12(8):e0182768.
170. Smajs D, McKevitt M, Howell JK, Norris SJ, Cai W-W, Palzkill T, et al. Transcriptome of Treponema pallidum: gene expression profile during experimental rabbit infection. J Bacteriol. 2005;187(5):1866–74.
171. Giacani L, Iverson-Cabral SL, King JCK, Molini BJ, Lukehart SA, Centurion-Lara A. Complete Genome Sequence of the Treponema pallidum subsp. pallidum Sea81-4 Strain. Genome Announc. 2014;2(2):e00333-14-e00333-14.
172. McGill MA, Edmondson DG, Carroll JA, Cook RG, Orkiszewski RS, Norris SJ. Characterization and serologic analysis of the Treponema pallidum proteome. Infect Immun. 2010;78(6):2631–43.
173. Salazar JC, Rathi A, Michael NL, Radolf JD, Jagodzinski LL. Assessment of the kinetics of Treponema pallidum dissemination into blood and tissues in experimental syphilis by real-time quantitative PCR. Infect Immun. 2007;75(6):2954–8.
174. Houston S, Taylor JS, Denchev Y, Hof R, Zuerner RL, Cameron CE. Conservation of the host-interacting proteins Tp0750 and pallilysin among treponemes and restriction of proteolytic capacity to treponema pallidum. Infect Immun. 2015;83(11):4204–16.
175. Houston S, Russell S, Hof R, Roberts AK, Cullen P, Irvine K, et al. The multifunctional role of the pallilysin-associated Treponema pallidum protein, Tp0750, in promoting fibrinolysis and extracellular matrix component degradation. Mol Microbiol. 2014;91(3):618–34.
176. Kao W-CA, Pětrošová H, Ebady R, Lithgow K V, Rojas P, Zhang Y, et al. Identification of Tp0751

(Pallilysin) as a Treponema pallidum Vascular Adhesin by Heterologous Expression in the Lyme disease Spirochete. Sci Rep. 2017;7(1):1538.

177. Tipple C, Hanna MOF, Hill S, Daniel J, Goldmeier D, McClure MO, et al. Getting the measure of syphilis: qPCR to better understand early infection. Sex Transm Infect. 2011;87(6):479–85.
178. Gayet-Ageron A, Ninet B, Toutous-Trellu L, Lautenschlager S, Furrer H, Piguet V, et al. Assessment of a real-time PCR test to diagnose syphilis from diverse biological samples. Sex Transm Infect. 2009;85(4):264–9.
179. Sabbagh B, Mindt S, Neumaier M, Findeisen P. Clinical applications of MS-based protein quantification. Proteomics- Clin Appl. 2016;10(4):323–45.
180. Baker ES, Liu T, Petyuk VA, Burnum-Johnson KE, Ibrahim YM, Anderson GA, et al. Mass spectrometry for translational proteomics: progress and clinical implications. Genome Med. 2012;4(8):63.
181. Clark AE, Kaleta EJ, Arora A, Wolk DM. Matrix-Assisted laser desorption ionization-time of flight mass spectrometry: A fundamental shift in the routine practice of clinical microbiology. Clin Microbiol Rev. 2013;26(3):547–603.
182. Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, et al. Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. Anal Chem. 2006;78(7):2113–20.
183. Yates JR, Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, et al. Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol. 1999;17(7):676–82.
184. Wolters DA, Washburn MP, Yates JR. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem. 2001;73(23):5683–90.
185. Yates JR. Mass spectrometry and the age of the proteome. J Mass Spectrom. 1998;33(1):1–19.
186. Zhang Y, Fonslow BR, Shan B, Baek M-C, Yates JR, III. Protein analysis by shotgun/bottom-up proteomics. Chem Rev. 2013;113(4):2343–94.
187. Chief EC. Method of the Year 2012 - Targeted Proteomics. Nat Methods. 2012;10(1):2013.
188. Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods. 2012;9(6):555–66.
189. Parker CE, Pearson TW, Anderson NL, Borchers CH. Mass-spectrometry-based clinical proteomics--a review and prospective. Analyst. 2010;135(8):1830–8.
190. Domanski D, Percy AJ, Yang J, Chambers AG, Hill JS, Freue GVC, et al. MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. Proteomics. 2012;12(8):1222–43.
191. Percy AJ, Yang J, Hardie DB, Chambers AG, Tamura-Wells J, Borchers CH. Precise quantitation of 136 urinary proteins by LC/MRM-MS using stable isotope labeled peptides as internal standards for biomarker discovery and/or verification studies. Methods. 2015;81:24–33.
192. Percy AJ, Chambers AG, Yang J, Hardie DB, Borchers CH. Advances in multiplexed MRM-based protein biomarker quantitation toward clinical utility. Biochim Biophys Acta - Proteins Proteomics. 2014;1844(5):917–26.
193. Cheng K, Chui H, Domish L, Hernandez D, Wang G. Recent development of mass spectrometry and proteomics applications in identification and typing of bacteria. Proteomics- Clin Appl. 2016;10(4):346–57.
194. Percy AJ, Parker CE, Borchers CH. Pre-analytical and analytical variability in absolute quantitative MRM-based plasma proteomic studies. Bioanalysis. 2013;5(22):2837–56.
195. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics. 2002;1(11):845–67.
196. Kruh-Garcia NA, Wolfe LM, Chaisson LH, Worodria WO, Nahid P, Schorey JS, et al. Detection of Mycobacterium tuberculosis Peptides in the Exosomes of Patients with Active and Latent M. tuberculosis Infection Using MRM-MS. Koomen JM, editor. PLoS One. 2014;9(7):e103811.
197. Khalilpour A, Kilic T, Khalilpour S, Álvarez MM, Yazdi IK. Proteomic-based biomarker discovery for development of next generation diagnostics. Appl Microbiol Biotechnol. 2017;101(2):475–91.
198. Parker CE, Borchers CH. Mass spectrometry based biomarker discovery, verification, and validation - Quality assurance and control of protein biomarker assays. Mol Oncol. 2014;8(4):840–58.
199. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain

path to clinical utility. Nat Biotechnol. 2006;24(8):971–83.

200. Paulovich AG, Whiteaker JR, Hoofnagle AN, Wang P. The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. Proteomics Clin Appl. 2008;2(10–11):1386–402.
201. Surinova S, Schiess R, Hüttenhain R, Cerciello F, Wollscheid B, Aebersold R. On the Development of Plasma Protein Biomarkers. J Proteome Res. 2011;10(1):5–16.
202. Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol. 2010;2010.
203. Zegels G, Van Raemdonck GAA, Coen EP, Tjalma WAA, Van Ostade XWM. Comprehensive proteomic analysis of human cervical-vaginal fluid using colposcopy samples. Proteome Sci. 2009;7:17.
204. Krastins B, Prakash A, Sarracino DA, Nedelkov D, Niederkofler EE, Kiernan UA, et al. Rapid development of sensitive, high-throughput, quantitative and highly selective mass spectrometric targeted immunoassays for clinically important proteins in human plasma and serum. Clin Biochem. 2013;46(6):399–410.
205. Drabovich AP, Martínez-Morillo E, Diamandis EP. Toward an integrated pipeline for protein biomarker development. Biochim Biophys Acta - Proteins Proteomics. 2015;1854(6):677–86.
206. Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Yi J, Schuchard MD, et al. HUPO Plasma Proteome Project specimen collection and handling: Towards the standardization of parameters for plasma proteome samples. Proteomics. 2005;5(13):3262–77.
207. Mermelekas G, Vlahou A, Zoidakis J. SRM/MRM targeted proteomics as a tool for biomarker validation and absolute quantification in human urine. Expert Rev Mol Diagn. 2015;15(11):1441–54.
208. TDR Diagnostics Evaluation Expert Panel, Banoo S, Bell D, Bossuyt P, Herring A, Mabey D, et al. Evaluation of diagnostic tests for infectious diseases: general principles. Nat Rev Microbiol. 2010;8(12 Suppl):S17-29.
209. Osbak K, Abdellati S, Tsoumanis A, Van Esbroeck M, Kestens L, Crucitti T, et al. Evaluation of an automated quantitative latex immunoturbidimetric non-treponemal assay for diagnosis and follow-up of syphilis: a prospective cohort study. J Med Microbiol. 2017;66(8):1130–9.

Section I
Epidemiology

Chapter 2

The prevalence of syphilis from the early HIV period is correlated with peak HIV prevalence at a country level

Authors: Kara K. Osbak[1], Jane T. Rowley[2], Nicholas J. Kassebaum[3], Chris R. Kenyon[1,4]

Affiliations

1. HIV/STI Unit, Institute of Tropical Medicine, Antwerp, Belgium
2. London, United Kingdom
3. Institute for Health Metrics and Evaluation, Seattle Children's Hospital, University of Washington, Seattle, WA, United States
4. Division of Infectious Diseases and HIV Medicine, University of Cape Town, Cape Town, South Africa

I contributed to the data collection, data analysis, writing and finalization of the manuscript.

Published in Sexually Transmitted Disease Journal, April 2016
doi: 10.1097/OLQ.0000000000000422

Conflicts of Interest and Source of Funding: None declared.

Abstract

Background

Could we have predicted national peak HIV based on syphilis prevalence in the 1990s? Earlier studies have shown positive correlations between various STIs at different population levels. In this paper we test the hypothesis that there was a residual variation in the national prevalence rates of syphilis and that these rates could predict subsequent peak-HIV prevalence rates.

Methods

This analysis uses linear regression to evaluate the country-level relationship between antenatal syphilis prevalence (1990-1999) and peak HIV prevalence. Antenatal syphilis data were taken from an Institute for Health Metrics and Evaluation database on the prevalence of syphilis in low risk populations. Peak HIV prevalence was calculated based on data taken from the Global Health Observatory Data Repository of the World Health Organization.

Results

A moderately strong association is found for the 76 countries with data available ($R^2 = 0.53$, $P < 0.001$). The association was weakened but remained significantly positive when we adjusted for the type of syphilis testing used.

Conclusion

Syphilis prevalence in the 1990s predicted approximately 53% of the variation in peak HIV prevalence. Populations with generalized HIV epidemics had a higher prevalence of syphilis in the pre-HIV period. This finding provides additional rationale to carefully monitor sexual behavior, sexual networks and STI incidence in these populations.

Introduction

Previous studies have found positive correlations between the prevalence of various STIs at different levels of aggregation. At the level of ethnic/racial groups, studies have found a relationship between the prevalence of HIV and syphilis [1,2], HSV-2 [2,3], chlamydia [1] and gonorrhea [1]. At a national level peak HIV prevalence has been found to be correlated with HSV-2 prevalence in women [3]. At the world-regional level HIV prevalence was correlated with the prevalence of HSV-2, gonorrhea, chlamydia, syphilis and trichomonas but this relationship was only statistically significant in the case of HSV-2 and gonorrhea [4].

In this paper we extend these analyses to assess if the prevalence of syphilis from the early HIV period is correlated with peak HIV prevalence at a country level. We chose syphilis as the widespread testing for syphilis in routine antenatal care means that there are more data available from population-based samples for syphilis in the early HIV period than there is for other STIs [5].

Methods

Antenatal syphilis data for 1980 to 2010 were taken from an Institute for Health Metrics and Evaluation (IHME) database on the prevalence of syphilis in low risk populations compiled for GBD 2010 [6]. The IHME data sources included UNAIDS epidemiologic fact sheets, UNGASS country progress reports, reports from country specific surveillance systems, WHO reports on syphilis epidemiology, and unpublished data from correspondence with GBD 2010 collaborators. These were supplemented by a systematic literature review of syphilis seroprevalence (most recent PubMed search was October 2011). To be included, a study needed to provide data on the prevalence of syphilis in populations considered representative of the general population and have a sample size of at least 100. Data from high-risk populations (e.g. MSM, sex workers and STI clinic attendees) were excluded [6-9].

For each country we extracted the studies conducted in antenatal populations between 1990 and 1999 and used the median of these studies as a proxy for the national prevalence of antenatal syphilis (termed the *syphilis prevalence* variable).

For the *adjusted syphilis prevalence* we generated new syphilis prevalence estimates adjusted according to the syphilis-testing algorithm used since different types of testing algorithms influence the estimates of prevalence. Studies that use both a treponemal and a non-treponemal test to diagnose infection are regarded as offering the most accurate measure of active syphilis infection [10]. We applied correction

factors to different syphilis testing algorithms used based on a systematic review and meta-analysis that estimated the proportion of pregnancies with "probable active syphilis" according to the testing methodology used in the study [5]. Studies using only a treponemal test will falsely label persons with old or treated syphilis as having active syphilis, therefore a correction factor of 0.536 was applied to these prevalence estimates. Studies using only non-treponemal testing will falsely diagnose active syphilis in persons with other inflammatory conditions that lead to biological false positive reactions in the non-treponemal test. A correction factor of 0.522 was applied to these studies. In addition, the study recommended the use of a correction factor of 0.686 for studies that did not report the type of testing used [5].

National peak HIV prevalence was obtained for 149 countries for which HIV prevalence estimates were available from 1990 to 2009 in the Global Health Observatory Data Repository of the World Health Organization (http://apps.who.int/gho/data/node.main.622) [11]. These estimates are based on population-based testing, antenatal clinical surveillance and epidemic models [12]. From this data we calculated the *peak HIV prevalence* variable as the highest HIV prevalence (in 15 to 49 year-olds) that each country obtained between the years 1990-2009 (median year 1998, interquartile range 1996-2005) [11,13]. We chose to use peak HIV prevalence rather than HIV prevalence at a defined period after the syphilis prevalence was measured per country, as this has been shown to avoid the HIV introduction time bias which is caused by the widely differing times that national HIV epidemics began [14]. Peak HIV prevalence is a composite measure of all the factors which enhance and prevent the spread of HIV in a particular population. Since all national populations have now been exposed to HIV for more than a decade the peak prevalence of HIV that they attained represents a useful summary measure of how permissive different sexual networks are to the spread of HIV [14]. It may also be a useful measure of sexual network susceptibility to other STIs such as syphilis. Because HIV epidemics have been shown to reduce syphilis prevalence both through AIDS mortality [15,16] and improved STI treatment [17] it is important to test the association between peak HIV and syphilis prevalence using syphilis prevalence data preceding the peak HIV prevalence. The period 1990-1999 was chosen for syphilis prevalence data as this was the first decade for which sufficient data was available in the dataset and the beginning of this period predates the year that most countries attained their peak HIV prevalence. Only 16 countries had data points for the 1980s.

Simple linear regression was used to evaluate the relationship between syphilis prevalence and peak HIV prevalence – unadjusted and adjusted for laboratory testing used. Because the severity and heterogeneity of the HIV epidemics have been greater in sub-Saharan Africa than elsewhere the analyses were repeated limited to the countries of sub Saharan Africa. All analyses were performed in STATA 12.0 (StataCorp LP, College Station, TX, USA).

Results

We found a moderately strong association between syphilis prevalence and peak HIV prevalence for the 76 countries with data available ($R^2 = 0.53$, $P < 0.001$; Figure 1). The association was weakened but remained significantly positive when we adjusted for the type of syphilis testing used ($R^2 = 0.34$, $P < 0.001$). A similar positive association was found when we restricted the analyses to 28 countries of sub-Saharan Africa ($R^2 = 0.69$, $P < 0.001$, for unadjusted syphilis prevalence versus peak HIV prevalence).

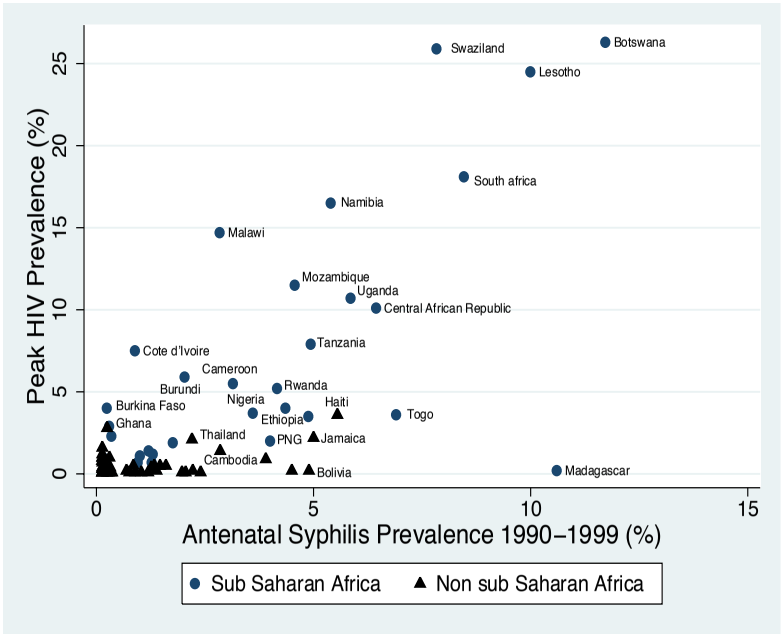
Discussion

This study found a positive association between syphilis prevalence during the period 1990-1999 and peak-HIV prevalence at a country level. Syphilis prevalence in the 1990s predicted approximately 53 % of the variation in peak HIV prevalence. This same relationship has been found at the levels of individuals [18,19], ethnic/racial groups [1,2] and world regions [4]. For example, a study that assessed this relationship at the ethnic group level in the United States of America and South Africa found a strong positive association between syphilis prevalence (predating peak HIV prevalence) and HIV prevalence [2]. However, not all studies have found this positive association [19]. The fact that the same positive association has been found between HIV and other STIs at the levels of individuals, ethnic/racial groups and countries [2,4] reduces the chance that this relationship is explained by confounding, or is due to an ecological or other form of bias. Rather, this association suggests that syphilis played a significant role in the spread of HIV and/or that both are determined by the same or similar factors.

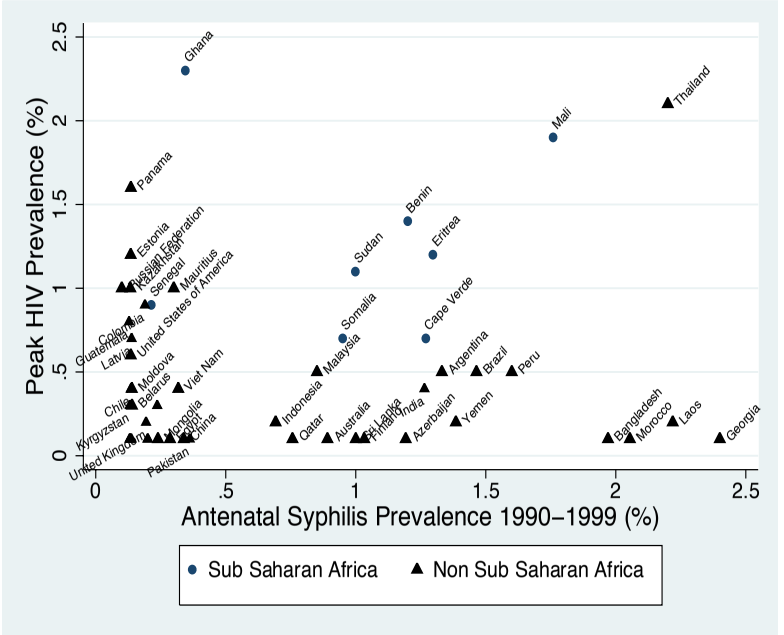
There are a number of limitations with this type of retrospective study; chief amongst these is the quality of the syphilis prevalence data. Peak HIV data was only reported for 149 countries and we only had combined HIV and syphilis data for 76 countries. Of the countries with data, prevalence estimates were based on relatively small or somewhat selected samples that may not be representative of the general population. Specifically, many of the studies are from urban areas and are studies of women presenting for antenatal care. It would also would have been optimal to have had an

Fig 1a *Association between peak HIV prevalence and the median prevalence of syphilis (expressed as a percentage) in all studies performed between the years 1990-1999 in antenatal women in 76 countries unadjusted for testing strategy ($R^2 = 0.53$, $P < 0.001$, $n = 76$).* **1b** *represents a magnified view of the same scatterplot but limited to countries with peak HIV and syphilis prevalences below 2.5%.*

a)



b)



earlier range syphilis prevalence estimates since some countries started attaining their peak HIV prevalences in the early/mid 1990s, such as Uganda where HIV peaked in 1991. We cannot exclude the possibility that the association between HIV and syphilis is due to unmeasured confounders. In certain countries intravenous drug use has been the major mode of HIV transmission and in these countries one would expect to find HIV prevalences to be disproportionately elevated vis-à-vis syphilis. We have also not considered the role that the endemic trepanematoses may have played in giving false positive syphilis results. The effect of each of these data and methodological limitations would dilute the strength of an association between syphilis and peak HIV prevalence.

A syphilis testing regime correction factor was also included in the analysis and when this factor was applied the association between syphilis and HIV was weakened. This could be explained by the fact that syphilis prevalences were overestimated in the studies that only use antibody testing which cannot distinguish reinfections from previous infections. The correction factors have not been validated in separate studies and hence they may lead to inaccurate prevalence estimates, which is an alternative explanation for the observed attenuation in HIV/syphilis association. A number of countries experienced considerable decreases or increases in the prevalence of syphilis in the 1990s [7,20]. Our measure of syphilis prevalence was the median value for the 1990s and our prevalence estimates may not have captured these changes.

There is increasing evidence that AIDS mortality played a role in the marked decline of syphilis prevalence in general populations in Southern and Eastern Africa [17,20,21] and MSM populations in North America [15]. In high income countries, the widespread availability of antiretroviral therapy has led to a decline in AIDS mortality and partly as a result, MSM sexual networks are returning towards a high-risk pre-AIDS structure with a concomitant increase in syphilis and other STIs [22,23]. It is important to appreciate that populations with generalized HIV epidemics had a higher prevalence of other STIs such as syphilis in the pre-HIV period [2] as this provides an additional rationale to carefully monitor sexual behavior, sexual networks and STI incidence in these populations to allow for the early detection of any return towards to a pre HIV risk configuration [21].

References

1. Morris M, Kurth AE, Hamilton DT, Moody J, Wakefield S. Concurrent partnerships and HIV prevalence disparities by race: linking science and public health practice. Am J Public Health. 2009;99(6):1023-31.
2. Kenyon C, Osbak K, Buyze J. The prevalence of HIV by ethnic group is correlated with HSV-2 and syphilis prevalence in Kenya, South Africa, the United Kingdom and the United States. Interdisciplinary Perspectives on Infectious Diseases. 2014:Article ID 284317.
3. Kenyon C, Colebunders R, Hens N. Determinants of generalized herpes simplex virus-2 epidemics: the role of sexual partner concurrency. Int J STD AIDS. 2013;24(5):375-82.
4. Kenyon C, Colebunders R, Buyze J. Classification of incidence and prevalence of certain sexually transmitted infections by world regions. Int J of Infec Dis. 2014;18:73-80.
5. Ham DC, Lin C, Newman L, et al. Improving global estimates of syphilis in pregnancy by diagnostic test type: A systematic review and meta-analysis. Int J Gynaecol Obstet. 2015;130 Suppl 1:S10-4.
6. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2095-128.
7. World Health Organization. Prevalence and incidence of selected sexually transmitted infections, Chlamydia trachomatis, Neisseria gonorrhoeae, syphilis, and Trichomonas vaginalis: methods and results used by the WHO to generate 2005 estimates. Geneva World Health Organization; 2011.
8. Rowley J, Toskin I, Ndowa F, Organization WH, Organization WH. Global incidence and prevalence of selected curable sexually transmitted infections, 2008: World Health Organization; 2012.
9. Meredith S, Hawkes S, Schmid G, Broutet N. The global elimination of congenital syphilis: rationale and strategy for action. 2007.
10. Tramont E. Syphilis In: Mandell GL, Bennett JE and Dolin R. Principles and Practice of Infectious Diseases 7th ed, Churchill Livingstone Inc. 2010.
11. Kenyon C, Buyze J. Should the threshold for generalized HIV epidemics be 1% or 5%? Int J STDs & AIDS. 2014;25:898.
12. UNAIDS. Methodology – Understanding the HIV estimates. Geneva: UNAIDS, 2014.
13. Kenyon CR, Buyze J. No association between gender inequality and peak HIV prevalence in developing countries - an ecological study. AIDS care. 2015;27(2):150-9.
14. Kenyon C, Colebunders R, Voeten H, Lurie M. Peak HIV prevalence: a useful outcome variable for ecological studies. Int J Infect Dis. 2013;17(5):e286-8.
15. Chesson HW, Dee TS, Aral SO. AIDS mortality may have contributed to the decline in syphilis rates in the United States in the 1990s. Sexually Transmitted Diseases. 2003;30(5):419-24.
16. Kenyon CR, Osbak K, Chico RM. What underpins the decline in syphilis in Southern and Eastern Africa? An exploratory ecological analysis. Int J Infect Dis. 2014;29:54-61.
17. Johnson LF, Dorrington RE, Bradshaw D, Coetzee DJ. The effect of syndromic management interventions on the prevalence of sexually transmitted infections in South Africa. Sex Reprod Healthc. 2011;2(1):13-20.
18. Nelson KE, Eiumtrakul S, Celentano D, Maclean I, Ronald A, Suprasert S, et al. The association of herpes simplex virus type 2 (HSV-2), Haemophilus ducreyi, and syphilis with HIV infection in young men in northern Thailand. Journal of acquired immune deficiency syndromes and human retrovirology : official publication of the International Retrovirology Association. 1997;16(4):293-300.
19. Buve A, Weiss HA, Laga M, Van Dyck E, Musonda R, Zekeng L, et al. The epidemiology of gonorrhoea, chlamydial infection and syphilis in four African cities. AIDS. 2001;15 Suppl 4:S79-88.
20. Kenyon CR, Osbak K, Buyze J, Chico RM. The changing relationship between bacterial STIs and HIV prevalence in South Africa - an ecological study. Int J STD AIDS. 2015;26(8):556-64.
21. Kenyon CR, Osbak K, Chico RM. What underpins the decline in syphilis in Southern and Eastern Africa? An exploratory ecological analysis. Int J Infect Dis. 2014;29:54-61.
22. Chesson HW, Gift TL. Decreases in AIDS mortality and increases in primary and secondary syphilis in men who have sex with men in the United States. Jaids-J Acq Imm Def. 2008;47(2):263-4.
23. Fenton KA, Imrie J. Increasing rates of sexually transmitted diseases in homosexual men in Western europe and the United States: why? Infectious disease clinics of North America. 2005;19(2):311-31.

Chapter 3

Superimposing incident STIs on HIV phylogram reveals possible misclassification of MSM as heterosexuals in a single-centre cohort in Antwerp, Belgium between 1997-2014

Authors: Kara K. Osbak[1]*, Conor J. Meehan[2]* Sergio G. Ribas[1], Leo Heyndrickx[1,2], Kevin K. Ariën[2], Achilleas Tsoumanis[1], Eric Florence[1], Marjan van Esbroeck[1], Katrien Fransen[1], Chris R. Kenyon[1,3]

**these authors contributed equally to this work*

Affiliations

1. Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium
2. Department of Biomedical Sciences, Institute of Tropical Medicine, Antwerp, Belgium
3. Division of Infectious Diseases and HIV Medicine, University of Cape Town, Cape Town, South Africa

I contributed to the data analysis, interpretation and manuscript preparation.

Conflicts of Interest and Sources of Funding

For all authors no conflicts of interest were declared. This work was part of Project ID: 757003 funded by the Flemish Government- Department of Economy, Science & Innovation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

In this single-centre study, we assessed if the superimposition of incident sexually transmitted infections (STIs) on HIV phylogenetic analyses could deepen our understanding of the transmission dynamics of STIs in our HIV-infected population. HIV-1 sequences collected between 1997-2014 from 1169 individuals attending a HIV clinic in Antwerp, Belgium were analysed to infer a partial HIV transmission network using Cluster Picker software. In total, 438 (37.5 %) individuals were identified in 136 clusters, including 76 transmission pairs and 60 clusters consisting of three or more individuals. Individual demographic, clinical and STI laboratory data collected during routine HIV follow-up were used to compare clustered and non-clustered individuals using logistic regression analyses. Individuals in a cluster were more likely to have a history of syphilis, chlamydia and/or gonorrhoea ($P < 0.05$), however, when analyses were stratified by HIV transmission risk groups (heterosexual and men who have sex with men (MSM)), this association only remained significant for heterosexuals with syphilis ($P = 0.001$). Under closer scrutiny, this association in heterosexuals was driven by six heterosexual men who were located in six almost exclusively MSM clusters. A parsimonious conclusion is that these six individuals were potentially misclassified as heterosexual. Improving the accuracy of sexual orientation reporting could improve care.

Introduction

Phylogenetic analysis of HIV sequences passively accumulated through routine drug resistance testing (DRT) can provide useful insights into local HIV-transmission network characteristics when coupled with demographic, clinical and sexually transmitted infection (STI) data. These approaches have proven useful in highlighting HIV transmission networks at various geographical levels [1–3], thereby allowing the identification of hotspots within (sub)populations such as men who have sex with men (MSM) [4]. Few studies [1] have specifically investigated the relationship between bacterial STIs and HIV phylogenetic clustering; a recent Belgian study by Chalmet *et al.* [1] demonstrated a strong correlation between syphilis and clustering. With the advent of increasingly user-friendly and robust bioinformatic tools [5], it is becoming easier for clinicians and epidemiologists to apply these approaches to various cohorts. It is expected that in the future these methods will be employed to inform and improve public health efforts [6].

HIV-infected individuals play a disproportionately large role in the current STI epidemics in Europe and elsewhere [7,8]. HIV-infected MSM in particular contribute up to 49.8 % of those diagnosed with certain STIs such as syphilis [9]. A resurgence of several STI epidemics including syphilis, hepatitis C and lymphogranuloma venereum has been observed in our HIV/STI clinics during the last two decades [10–12; Personal communication V. Maes WIV-ISP] and we have noted that persons with multiple infections are playing an increasing role in STI transmission [13]. For example, during the syphilis epidemic observed at our clinic between 1992-2012, over half of infections occurred in patients with a previous diagnosis of syphilis and almost all were HIV-infected MSM [13].

In this study, we aimed to ascertain if mapping STI diagnoses onto our local HIV phylogenetic tree could help us to better understand the sexual networks underpinning local STI transmission.

Methods

Ethics

The Institutional Review Board of the Institute of Tropical Medicine (ITM) approved this study (901/13).

Study participants

A retrospective review of clinical data from HIV-1 infected individuals attending the ITM's HIV clinic who had at least one plasma sample submitted for HIV-sequencing in the context of DRT from December 31, 1997 to February 17, 2014 was performed. The ITM has a medium-sized HIV-outpatient clinic with 3371 HIV-infected individuals in regular care during the study period. From 2008 onwards, DRT in the form of reverse transcriptase (RT) and protease (PR) gene sequencing was incorporated into the HIV diagnostic work-up; prior to 2008 HIV sequencing was performed in the case of treatment failure.

Demographic and clinical data were extracted from our clinical database and pseudonymised. Variables of interest included: geographical birth region, biological gender, age at HIV diagnosis/blood sampling for HIV sequencing and antiretroviral therapy (ART) intake at the time of sequencing sampling. Self-reported HIV-transmission risk was recorded by the HIV-care physician at the time of diagnosis. Individuals were grouped into categories: 'heterosexual', 'MSM', 'homo/bisexual (females only)' and 'other', which included intravenous drug use (IVDU), mother to child transmission (MTCT), blood transfusion and occupational risk transmissions. Sexual preference classification was extrapolated from the transmission risk categories, with the exception of the 'other' classification for which no sexual identity was assigned. Hence, a male reporting heterosexual HIV transmission would be classified as heterosexual. MSM were defined as males ever reporting sexual contact with someone of the same sex. All episodes of gonorrhoea, chlamydia and syphilis that were diagnosed at the ITM laboratory from August 1993 to June 2014 were extracted from the ITM laboratory database for each individual. Nadir and CD4+ T-cell count at diagnosis values were reported.

Sexually transmitted infection laboratory testing

STI testing is typically performed at the ITM every 3-12 months for patients in routine HIV follow-up care depending on their risk profile [14]. Treponema pallidum ssp. pallidum infection was dual tested with a non-treponemal Macro-Vue Rapid Plasma Reagin (RPR) card test (Becton, Dickinson and Company, Sparks, MD, United States (US)) and treponemal antibody detection was performed via the Cellognost Syphilis H Combipack *T. pallidum* haemagglutination assay (TPHA) (Dade Behring, Frankfurt, Germany) up until November 2002, thereafter testing was performed with the SERODIA *T. pallidum* particle agglutination (TPPA) assay (Fujirebio, Inc., Tokyo, Japan). Syphilis episodes were defined as previously described [13]. *Neisseria*

gonorrhoeae (NG) and *Chlamydia trachomatis* (CT) presence in first void urine and/or (anal/pharyngeal/vaginal) swabs were assessed with a nucleic acid amplification test from 1999 onwards for CT and 2010 onwards for NG following manufacturer's instructions. Before 2010 NG presence was tested by culture.

HIV RNA extraction, amplification and sequencing

Viral RNA was extracted from plasma using the QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions. The resistance genotyping was performed with the Trugene HIV-1 genotyping kit (Siemens Healthcare Diagnostics, Tarrytown, NY, US). The PCR primers covered a 1.3 Kb fragment of the pol gene. Full length protease (PR) and partial reverse transcriptase (RT) sequences (codons 39-241) were sequenced as previously described [15]. Sequencing products were analyzed using the OpenGene DNA Sequencing System (Visible Genetics, Inc., Toronto, Canada) following manufacturer's instructions. Sequences were stored using the Integrated Database Network System (IDNS) software package (Smartgene, Zug, Switzerland).

Sequence analyses

HIV subtyping was performed using REGA version 3.0 [16] based on the RT gene. Sequence alignments were performed for both the PR and RT genes separately. HIV-1 pol sequence data is uniquely suited to describe transmission events and networks since there is little (<1 %) intra-individual divergence between baseline and longitudinal sequences [17] demonstrated for subtype B. Sequences were placed in-frame using Mesquite version 3.10 [18] and then aligned using the codon model of MUSCLE version 3.7 [19] implemented in MEGA7 [20]. Trailing sequence ends were trimmed using MEGA7 and then the two gene alignments were concatenated using Mesquite. A maximum likelihood phylogenetic tree of all samples was created using a GTR+Γ model of evolution with each gene as a separate partition and 1000 bootstrap replicates as implemented RAxML version 8.2.9 [21]. Patients for whom their multiple samples did not cluster on the phylogenetic tree (distance > 0.1) were identified using a custom python script implementing DendroPy [22]. Such large distances are likely the result of either multiple HIV-1 infections or a sample switch error. The above alignment and tree building steps were redone using only the first sample (as defined by the date of sequencing) for all individuals. Trees were visualised using FigTree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). Cluster Picker [23] version 1.2.3 was used to identify clusters of all individuals. Clusters were defined as those with >90 % bootstrap support for the clade and <4.5 % genetic difference between all patient samples. Mixed transmission clusters were defined as clusters consisting of two or

more individuals, whereby at least one individual was a heterosexual male who clustered with at least one MSM individual and regardless of the presence of female(s) in the cluster. Potentially misclassified HIV sexual transmission cases were defined as heterosexual males with a history of syphilis who were present in a mixed transmission cluster.

Statistical Analyses

Continuous variables are expressed in median and interquartile range (IQR). Clustered and non-clustered individuals were compared using chi-squared ($\chi 2$) test or Fishers exact test for categorical variables. Univariate logistic regression analyses using maximum likelihood ratio for the whole cohort (Stage 1 analysis) were performed. When stratifying STI (syphilis/chlamydia/gonorrhoea) diagnosis analyses by HIV transmission risk groups (heterosexual versus MSM) (Stage 2 analysis) Penalized Multinomial Logistic Regression (PMLR) [24,25] analyses were performed using the STATA module 'firthlogit' [26]. A P-value of ≤0.05 was regarded as significant. All analyses were performed in STATA version 13.0 (StataCorp LP, College Station, TX, US).

Results

General description of study cohort

In total, 1173 HIV-infected individuals out of 3371 patients in follow-up over the study period that submitted at least one plasma sample for HIV-sequencing were included in the study. Of the 183 individuals that submitted two samples for DRT, four were excluded from the study because the two samples did not cluster on the phylogram, which could be indicative of a new (mixed) infection or an erroneous sample switch. This could possibly lead to mismatching of the HIV sequence and demographic data, hence these four individuals were omitted from the study; a total of 1169 individuals were retained in the study.

General cohort characteristics are presented in *Table 1*. Briefly, most participants were male (914/1169; 78.2 %) and of these 76.5 % (699/914) were MSM. Almost all females reported heterosexual HIV transmission risk (248/255; 97.3 %) with the exception of four who reported homo/bisexual HIV transmission risk. Fourteen (1.2 %) individuals reported 'other' probable HIV transmission modes, including IVDU (n=8), MTCT (n=3), occupational (n=1) and blood transfusion risk (n=2) categorized individuals. Approximately two thirds (788/1169; 67.4 %) of the study population originated from the Europe, North America or Australia region and of these 88.1 % (694/788) were born

in Belgium. The second largest population was sub-Saharan African (286/1169; 24.5 %), a majority of whom were female (176/286; 61.5 %). Other world regions accounted for 95/1169 (8.2 %) of the study population.

The distribution of HIV diagnosis year and sampling for sequencing year is depicted in *Figure 1*. Sampling was performed at the time of diagnosis for most individuals (830/1169; 71.0 %). The median time interval from diagnosis to sampling for those not sampled at diagnosis was 3.7 years (IQR 1.9-6.6). Most started ART shortly after diagnosis (639/830; 77.0 %). At the time of sampling, 17.9 % (208/1169) were ART naïve, 79 % (924/1169) were taking ART and 3.1 % (36/1169) had a history of ART but were not taking treatment. The median age at diagnosis was 34.3 years (IQR 28.3-41.8) and at sampling 35.9 years (IQR 29.4-44.2).

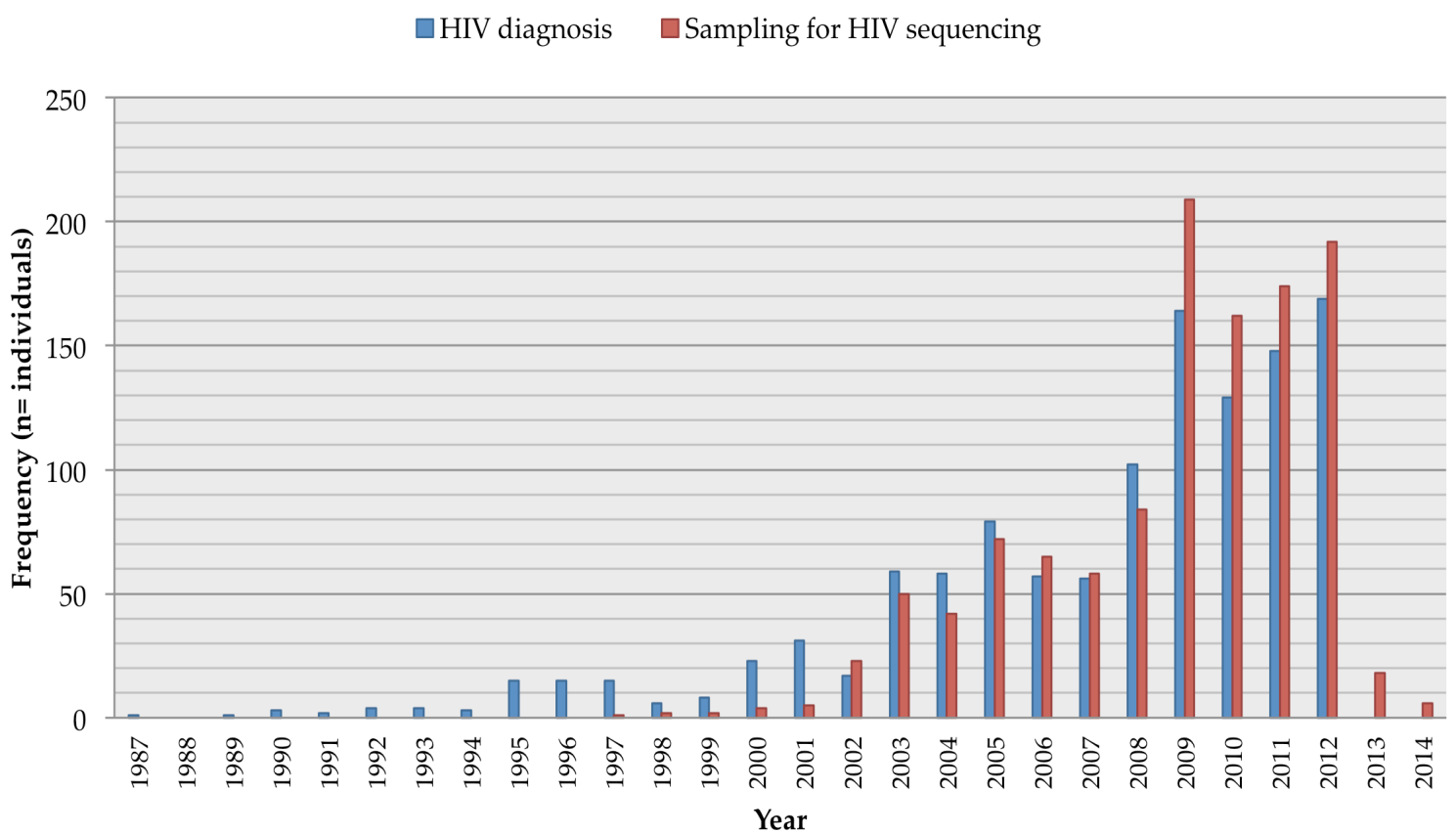
Approximately half of the participants were classified as HIV subtype B (676/1169, 57.8 %); most were MSM (584/676; 86.4 %). A majority of non-B subtype individuals were heterosexual (366/493; 74.2 %) and from the Sub-Saharan African region (275/493; 55.8 %) (*Table 1*).

Table 1 *Characteristics of study cohort HIV-infected individuals (n= 1169) who submitted plasma samples for drug resistance testing from 1997 to 2014 at a HIV clinic in Antwerp, Belgium*

Characteristic	**n= (%)**
Male	914 (78.2)
Female	255 (21.8)
HIV sampling for sequencing performed at time of HIV diagnosis	830 (71.0)
On ART at time of sequencing	924 (79.4)
Birth Region	
Europe, North America or Australia	788 (67.4)
Belgium	694 (59.4)
Asia	48 (4.1)
Sub-Saharan Africa	286 (24.5)
North-Africa	15 (1.3)
Central and South America	32 (2.7)
HIV subtype	
Subtype B	676 (57.8)
Non-B	493 (42.2)
Self-reported HIV transmission risk category	
Heterosexual Total	452 (38.7)
Male	204 (17.5)
Female	248 (21.2)
Homo/bisexual Total	703 (60.1)
Male (MSM)	699 (59.8)
Female (Homo/bisexual)	4 (0.3)
Other [a]	14 (1.2)
Male	11 (0.01)
Female	3 (0.003)
Other	
Characteristic	**Median (Interquartile Range)**
Age at time of sample for HIV sequencing (years)	35.9 (29.4-44.2)
Age at time of HIV diagnosis (years)	34.3 (28.3-41.8)
$CD4^+$ T-cell count nadir	295 (186-414)
$CD4^+$ T-cell count at HIV diagnosis	419 (266-610)

Legend: ART: antiretroviral therapy; a) 'other' transmission risk category includes: n=8 intravenous drug users, n=3 mother to child transmission, n=1 occupational & n=2 blood transfusion HIV transmissions

Fig 1 *Bar graph depicting per year frequency of study individuals diagnosed with HIV (blue) and submitting first sample for HIV PR/RT sequencing (red) (n= 1169 participants)*



High incidence of sexually transmitted infections among HIV-infected MSM

Syphilis was the most common STI, with a total of 298 episodes reported in 198 individuals. Most (133/198; 67.2 %) only had a diagnosis of one episode and 65/198 (33.9 %) individuals had two or more episodes. Almost all syphilis episodes were reported in MSM (290/298; 97.3 %), with the exception of seven heterosexual males (7/204; 3.4 %) and one homo/bisexual female. A diagnosis of syphilis was made in 27.2 % (190/699) of MSM. Chlamydia was the second most common STI with a total of 173 episodes in 131 individuals, mostly occurring in MSM (116/131; 88.5 %), in addition to three heterosexual females, 10 heterosexual males and two homo/bisexual females. In 34/131 (26.0 %) chlamydia was diagnosed on two or more occasions. With regards to gonorrhoea, there were 173 episodes in 104 individuals; most were MSM (98/104; 94.2 %), with the exception of three heterosexual males, one heterosexual female and two homo/bisexual females also reporting gonorrhoea episodes. A majority only had one episode (74/104; 71.2 %) and 30/104 (28.8 %) had two or more episodes. When considering all study individuals, 324/1169 (29.3 %) had at least one STI episode diagnosed (chlamydia/gonorrhoea/syphilis) during their HIV follow-up; this increased to 42.8 % (299/699) when only considering the MSM population.

HIV transmission clustering analysis reveals high number of mixed transmission clusters

The HIV phylogram constructed from the sequence information and subsequent clustering analyses placed a total of 438/1169 (37.5 %) study participants into 136 transmission clusters, including 60 larger clusters defined as three or more individuals and 76 transmission pairs (also defined as clusters) (*Table 2; Figure 2*). Larger clusters contained a median of four individuals (IQR 3-6) with a maximum of 17 individuals (n= 1 cluster; MSM only). Clusters were defined based on the HIV-transmission risk category of the individuals, whereby 83 clusters contained 283 exclusively MSM, 27 clusters (containing 57 individuals) contained exclusively heterosexual men and females and 24 clusters (containing 94 individuals) were mixed transmission clusters. No homo/bisexual females clustered with others. Six heterosexual females were present in six different mixed transmission clusters; thus 18 clusters consisted of heterosexual males clustering exclusively with other MSM. There were two 'other' categorized clustered pairs: i) a male IVDU with a male heterosexual and ii) a male IVDU clustered with a heterosexual female.

Table 2 *HIV-transmission clusters identified by Cluster Picker and STI history of individuals in each cluster category*

	Total n= clusters	Total n= individuals	MSM only clusters	Heterosexual only clusters	MSM/ heterosexual mix clusters [a]	'Other' mix clusters
Clusters [a]	136	438	83 (283)	27 (57)	24 (94)	2 [c]
Cluster pairs	76	152	42 (84)	24 (48)	8 (16)	2 (4)
Clusters with ≥3 individuals	60	286	41 (199)	3 (9)	16 (78)	NA
STI [b]	**Total n= clusters**	**Total n= individuals**	**Total n= individuals**	**Total n= individuals**	**Total n= individuals**	**Total n= individuals**
Syphilis	NA	198	86	0	22	0
Chlamydia	NA	131	55	0	6	0
Gonorrhoea	NA	104	45	1	4	0

Legend: All data presented in n=clusters (n=individuals) unless otherwise noted; NA: not applicable; STI: sexually transmitted infection; a) Clusters defined as those with >90% bootstrap support for the clade and <4.5 % genetic difference between all individual samples as determined by Cluster Picker version 1.2.3; b) individuals ever diagnosed with an episode of chlamydia, gonorrhoea or syphilis at the ITM laboratory from 1993-2014; c) two transmission pairs: 1. male intravenous drug user (IVDU) + heterosexual female; 2. Male IVDU + heterosexual male

Fig 2 *HIV transmission phylogram of 1169 individuals in this study. Clusters (n= 136) are depicted in colour; blue represents clusters containing ≥3 individuals, green represents cluster pairs and red/yellow depicts mixed transmission clusters with possible misclassified HIV sexual transmission heterosexual males*

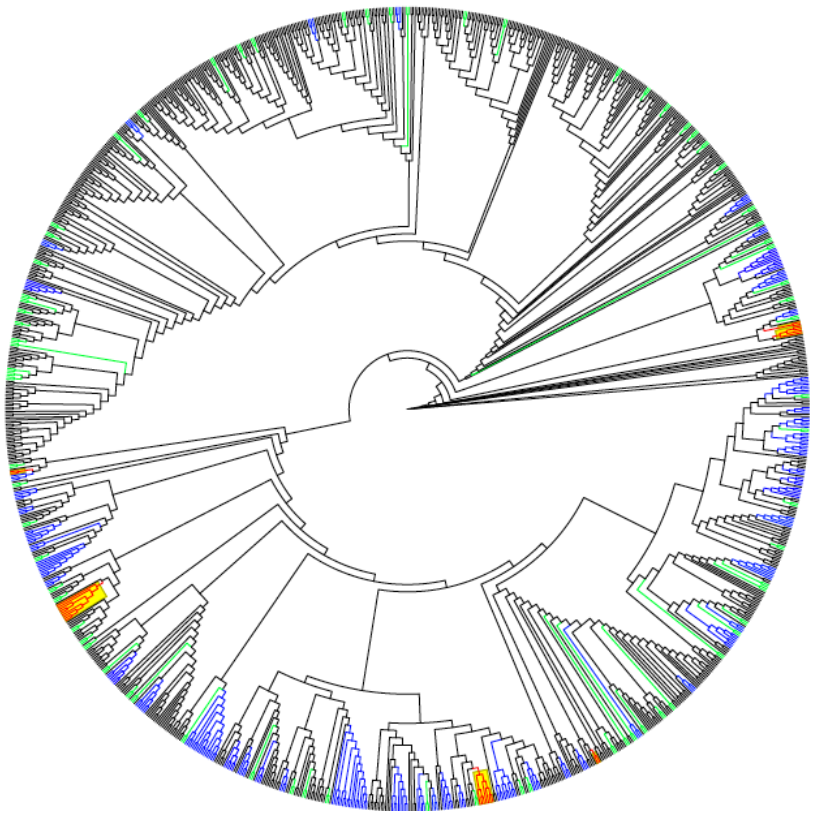


Table 3 outlines various univariate associations with HIV-transmission clustering. There was a strong positive association between male gender (OR 4.96 [CI 3.4-7.3]; P < 0.001), HIV B-subtype (OR 4.7 [CI 3.6-6.1]; P < 0.001), HIV sequencing performed at diagnosis (OR 2.3 [CI 1.7-3.0]; P < 0.001), MSM transmission (OR 3.9 [CI 3.0-5.2]; P < 0.001) and transmission clustering. Only 41/286 individuals of Sub-Saharan African origin clustered (OR 0.2 [CI 0.1-0.3]; P < 0.001) compared to 346/694 of Belgian origin (OR 4.1 [CI 3.2-5.4]; P < 0.001). There was a negative association between age ≥25 years and clustering (OR 0.5 [CI 0.4-0.8]; P = 0.008).

Table 3 *Characteristics of HIV-infected individuals who were identified in HIV transmission clusters compared to non-clustering individuals*

	Total n= (%)	In transmission cluster [a] n= (%)	Not in transmission cluster [a] n= (%)	Odds ratio (95 % CI) [e]
Whole study cohort	1169	438 (37.5)	731 (62.5)	NA
Gender				
Female	255 (21.8)	35 (13.7)	220 (86.3)	Ref
Male	914 (78.2)	403 (44.1)	511 (55.9)	4.7 (3.4-7.3)**
HIV Subtype				
Non-subtype B	493 (42.2)	91 (18.5)	402 (81.5)	Ref
Subtype B	676 (57.8)	347 (51.3)	329 (48.7)	4.7 (3.6-6.1)**
Belgian origin				
No	475 (40.6)	92 (19.4)	383 (80.6)	Ref
Yes	694 (59.4)	346 (49.9)	348 (50.1)	4.1 (3.2-5.4)**
Sub-Saharan African origin				
No	883 (75.5)	397 (45.0)	486 (55.0)	Ref
Yes	286 (24.5)	41 (14.3)	245 (85.7)	0.2 (0.1-0.3)**
Age at time of sampling for sequencing [b]				
< 25 years	112 (9.6)	55 (49.1)	57 (50.9)	Ref
≥ 25 years	1057 (90.4)	383 (36.2)	674 (63.8)	0.5 (0.4-0.8)*
HIV sampling for sequencing performed at time of diagnosis				
No	339 (29.0)	84 (24.8)	255 (75.2)	Ref
Yes	830 (71.0)	354 (42.7)	476 (57.3)	2.3 (1.7-3.0)**
Sexual orientation [b]				
Heterosexual	452 (39.2)	90 (19.9)	362 (80.1)	Ref
MSM	699 (77.4)	346 (49.5)	353 (50.5)	3.9 (3.0-5.2)**
Homo/bisexual female	4 (0.3)	0 (0)	4 (100)	NA [d]
Syphilis Overall association [b,c]				
No	971 (83.1)	330 (34.0)	641 (66.0)	Ref
Yes	198 (16.9)	108 (54.5)	90 (45.5)	2.3 (1.7-3.2)**
Syphilis MSM only [b,c]				
No	509 (72.8)	244 (47.9)	265 (52.1)	Ref
Yes	190 (27.2)	102 (53.7)	88 (46.3)	1.3 (0.9-1.8)
Syphilis Heterosexual only [b,c]				
No	455 (98.5)	84 (18.9)	361 (81.1)	Ref
Yes	7 (1.5)	6 (85.7)	1 (14.3)	18.5 (3.1-111.1)**

Table 3 (cont.) *Characteristics of HIV-infected individuals who were identified in HIV transmission clusters compared to non-clustering individuals*

	Total n= (%)	In transmission cluster [a] n= (%)	Not in transmission cluster [a] n= (%)	Odds ratio (95 % CI) [e]
Chlamydia Overall association [b,c]				
No	1038 (88.8)	337 (36.3)	661 (63.7)	Ref
Yes	131 (11.2)	61 (46.6)	70 (53.4)	1.6 (1.1-2.2)*
Chlamydia MSM only [b,c]				
No	583 (83.4)	286 (49.1)	297 (50.1)	Ref
Yes	116 (16.6)	60 (51.7)	56 (48.3)	1.1 (0.7-1.6)
Chlamydia Heterosexual only [b,c]				
No	439 (97.1)	89 (20.3)	350 (79.7)	Ref
Yes	13 (2.9)	1 (7.7)	12 (92.3)	0.5 (0.1-2.5)
Gonorrhoea Overall association [b,c]				
No	1065 (91.1)	388 (36.4)	677 (63.6)	Ref
Yes	104 (8.9)	50 (48.1)	54 (51.9)	1.7 (1.1-2.5)*
Gonorrhoea MSM only [b,c]				
No	601 (86.0)	298 (49.6)	303 (50.4)	Ref
Yes	98 (14.0)	48 (49.0)	50 (51.0)	1.0 (0.6-1.5)
Gonorrhoea Heterosexual only [b,c]				
No	452 (99.1)	88 (19.6)	360 (80.4)	
Yes	4 (0.09)	2 (50.0)	2 (50.0)	NA [d]

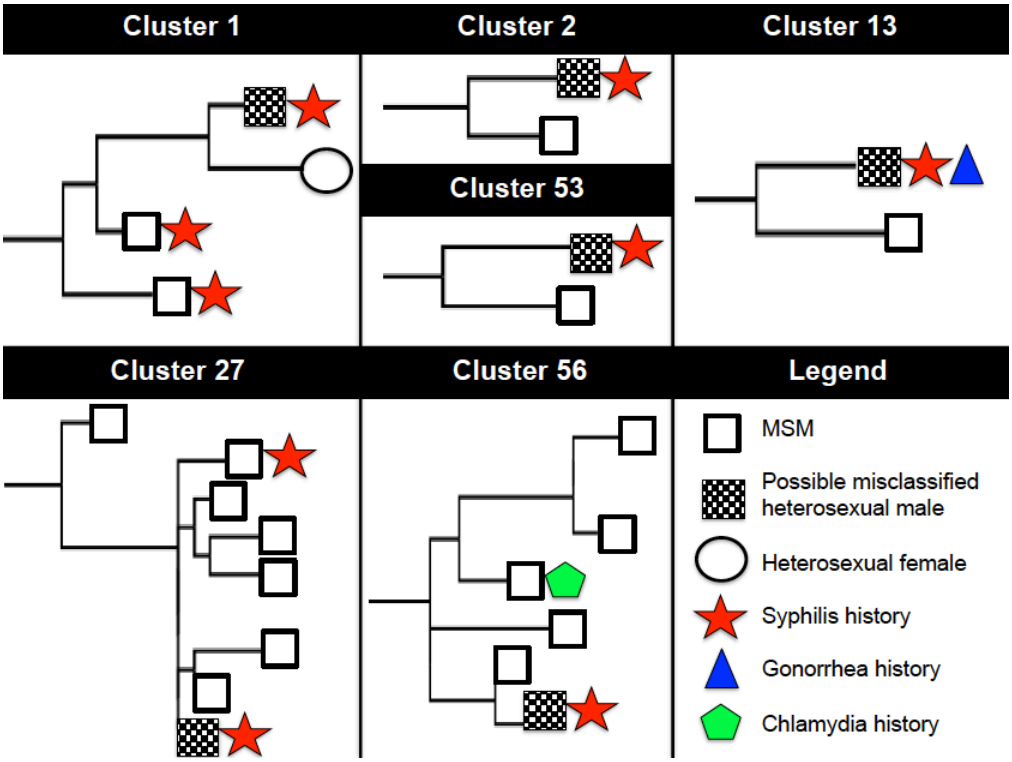
Legend: CI: confidence interval; NA: not applicable; P-value and Odds Ratio values are for comparison between clustered and non-clustered groups (clustered code= '1' in univariate analysis; non-clustered= '0'); Ref: reference comparison group for univariate analysis; Column 1 is presented as column percentages and column 2 & 3 are presented as row percentages; a) Clusters defined as those with >90 % bootstrap support for the clade and <4.5 % genetic difference between all individual samples as determined by Cluster Picker version 1.2.3; b) Calculated excluding the n=14 'other' (non-sexual) transmission category individuals and n=4 homo/bisexual females; c) individuals ever diagnosed at the ITM with an episode of syphilis, gonorrhoea or chlamydia during 1993-2014; d) not calculated due to low case number; e) P < 0.05, ** P < 0.005*

Those diagnosed with syphilis were more likely to be in a cluster (Overall association; OR 2.3 [CI 1.7-3.2]; P < 0.001) whereby 108/198 (54.5 %) of individuals with a diagnosis of syphilis clustered. Most of these individuals were located in MSM (86/108; 79.6 %) clusters while 22 (20.4 %) were located in mixed transmission clusters. When the analysis was stratified by sexual transmission categories ('MSM' and 'heterosexual'), clustering of syphilis only remained significant in the heterosexual group (OR 18.5 [CI 3.1-111.1]; P < 0.001). Individuals with a diagnosis of chlamydia and gonorrhoea were also more likely to cluster (Overall association: chlamydia: OR 1.6 [CI 1.1-2.2]; P = 0.02; gonorrhoea: 1.7 [CI 1.1-2.5]; P = 0.02), however, when stratified by transmission category, this relationship was no longer significant for both chlamydia and gonorrhoea (*Table 3*).

Six possible cases of HIV sexual transmission misclassification

Twenty-five heterosexual males were present in mixed transmission clusters, of which 6/25 had a diagnosis of syphilis and 5/6 clustered exclusively with other males. According to our definition, these males were categorized as possible HIV sexual transmission misclassifications. We therefore estimate the misclassification rate in our cohort to be 0.7 % of males (6/914). Two of these suspect misclassified individuals also had a history of chlamydia (n=1) and gonorrhoea (n=1) infections and 3/6 were born outside of Belgium, including sub-Saharan Africa (n=1), Eastern Europe (n=1) and Central and South America (n=1) regions. *Figure 3* is a graphical depiction of these six possible misclassified heterosexual males.

Fig 3 *Six HIV-transmission clusters containing possible misclassified HIV sexual transmission heterosexual males with a history of syphilis*



Discussion

This study exemplifies the utility of coupling phylogenetic analyses of HIV collected through DRT with (STI) clinical data to gain insights into local HIV transmission network characteristics. Simplified bioinformatic analyses revealed 136 transmission clusters encompassing 37.5 % of the study cohort. STIs were highly prevalent among MSM, with almost half (42.5 %) of MSM having a diagnosis of at least one STI episode (chlamydia, gonorrhoea and/or syphilis). We found that individuals with a STI diagnosis were more likely to cluster together. However, when analyses were stratified per risk transmission category, this relationship was attenuated and only remained significant for heterosexual males with a history of syphilis. This finding was unexpected since almost all syphilis-infections in our cohort occur in MSM [13]; the phylogenetic analyses, however, provided a possible explanation for this finding. The fact that 5 out of 6 of these men were in clusters that only contained MSM suggested that they might have been misclassified as non-MSM. In the sixth cluster a woman was present but her serological tests for syphilis remained negative until the

end of the study. The fact that 90.1 % (797/881) of all the men with a diagnosis of syphilis between 2000-2015 in our clinic reported being MSM (Personal communication V. Maes, ISP-WIV) further increases the probability that these men acquired their syphilis from other men.

A recent HIV phylogenetic study in the United Kingdom estimated that misclassification of homosexually acquired infection ranged between 1-11 % of classified male heterosexuals diagnosed with HIV [27]; this number increased to 1-21 % when considering only Black African populations. This estimation is higher than the 0.7 % of suspected misclassifications in our cohort. This discrepancy could be attributed to the larger national population study cohort that would likely cover sexual networks more extensively. Half of the possibly misclassified individuals in our study originated from outside Belgium. Although speculative, this finding could be due to a greater barrier to disclose same sex contacts in some populations [28]. Further sources of misclassification could be that doctors in our clinic are not taking accurate sexual histories, clinical databases are incorrectly filled in and/or not regularly updated if new patient information is obtained.

There are a number of caveats to this analysis. Only 34.6 % of HIV-infected persons in our cohort were included in this study. Since the introduction of routine DRT at the ITM from 2008 onwards, approximately half of the individuals in routine HIV follow-up clinical cohort have been sequenced. Nonetheless, 36.6 % of all individuals still clustered with other individuals, indicating there is moderate coverage of the local HIV transmission network. One must also take into account the high diversity of the cohort population. A quarter of the participants originated from sub-Saharan Africa; they were less likely to cluster ($P < 0.001$) with others, which could be indicative of transmission events outside the cohort or even the region. Mobility is also a factor, as demonstrated by a recent study investigating HIV transmission characteristics in the US which showed that of the American persons that clustered in the network, 457/13145 (3.5 %) were genetically linked to a potential transmission partner outside the US [29]. Moreover, the long study period spanning 17 years means that individuals diagnosed and sequenced early in the study period who then went on to transmit HIV years later within the same cohort may not be genetically linked due to the continued viral evolution that would slowly abolish phylogenetic linkage [30].

Furthermore, our phylogenetic tree is only based on HIV transmission. Since compared to other STIs, HIV is relatively inefficiently transmitted sexually (and especially so when concerning forms of oral sex) [31], our tree likely underestimates

the sexual network connectivity of individuals therein. Phylogenetic analyses are limited by the fact that they often cannot assign directionality, thus it cannot be ruled out that the one female clustered with the possible misclassified transmission individuals could have been the transmission link instead of a MSM in the same cluster. The number of individuals defined as possibly misclassified was very small. We conducted a PMLR analysis in an attempt to compensate for rare event biases [24,25], however, the sample size remains small. Lastly, we did not correct for known drug mutations since in other studies this did not substantially affect the clustering or tree structure [1,2].

Increased use of phylogenetic analysis brings with it important ethical, security and privacy considerations that need to be addressed [6]. With the further refinement of accessible bioinformatics tools, individual smaller centres will be able to use these methods to define and characterize their cohorts. Strategies using molecular epidemiology have already been developed with the potential to identify individuals most likely to transmit HIV infection, thereby allowing near real-time preventative strategy implementation [32,33]. Furthermore, patient care would benefit from standardized and optimized approaches to sexual orientation reporting. A recent study by Haider *et al.* [34] investigating patient-centered approaches to sexual orientation data collection found that patients were willing to be asked about their sexual orientation, were likely not to take offense at such questioning and both patients and clinicians prefer nonverbal data collection methods.

In our setting the superimposition of STIs on the HIV phylogram provided insights into how sexual orientation could have been misclassified for some individuals; this has important implications for patient care and understanding local STI transmission characteristics since these individuals could operate as a bridge population into otherwise low-risk groups. As far as patient care is concerned, it is crucial to know if individuals are MSM as they require different anal neoplasia and STI screening [14] than other populations.

Acknowledgements

The authors gratefully acknowledge the participation of the individuals in this study. Thanks to Virginie Maes from the Scientific Institute of Public Health (WIV-ISP) for providing a summary of the ITM's STI data. We appreciate Tania Crucitti's (ITM) constructive comments during the manuscript preparation and for supervision of the STI testing.

References

1. Chalmet K, Staelens D, Blot S, Dinakis S, Pelgrom J, Plum J, et al. Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. BMC Infect Dis. 2010;10(1):262.
2. Mehta SR, Wertheim JO, Brouwer KC, Wagner KD, Chaillon A, Strathdee S, et al. HIV Transmission Networks in the San Diego–Tijuana Border Region. EBioMedicine. 2015;2(10):1456–63.
3. Wertheim JO, Leigh Brown AJ, Hepler NL, Mehta SR, Richman DD, Smith DM, et al. The Global Transmission Network of HIV-1. J Infect Dis. 2014;209(2):304–13.
4. Bezemer D, Cori A, Ratmann O, van Sighem A, Hermanides HS, Dutilh BE, et al. Dispersion of the HIV-1 Epidemic in Men Who Have Sex with Men in the Netherlands: A Combined Mathematical Model and Phylogenetic Analysis. Celentano DD, editor. PLoS Med. 2015;12(11):e1001898; discussion e1001898.
5. Lam TT-Y, Hon C-C, Tang JW. Use of phylogenetics in the molecular epidemiology and evolutionary studies of viral infections. Crit Rev Clin Lab Sci. 2010;47(1):5–49.
6. German D, Grabowski MK, Beyrer C. Enhanced use of phylogenetic data to inform public health approaches to HIV among men who have sex with men. Sex Health. 2016;14(1):89.
7. Dougan S, Evans BG, Elford J. Sexually Transmitted Infections in Western Europe Among HIV-Positive Men Who Have Sex With Men. Sex Transm Dis. 2007;PAP(10):783–90.
8. Mayer KH, Bush T, Henry K, Overton ET, Hammer J, Richardson J, et al. Ongoing sexually transmitted disease acquisition and risk-taking behavior among US HIV-infected patients in primary care: implications for prevention interventions. Sex Transm Dis. 2012;39(1):1–7.
9. Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2015. Atlanta; 2016.
10. Bottieau E, Apers L, Van Esbroeck M, Vandenbruaene M, Florence E. Hepatitis C virus infection in HIV-infected men who have sex with men: sustained rising incidence in Antwerp, Belgium, 2001-2009. Euro Surveill. 2010;15(39):19673.
11. Apers L, Koole O, Bottieau E, Vandenbruaene M, Ophoff D, Van Esbroeck M, et al. Incidence of HCV and sexually transmitted diseases among hiv positive msm in antwerp, belgium, 2001-2011. Acta Clin Belg. 2013;68(6):421–6.
12. Vandenbruaene M, Ostyn B, Crucitti T, De Schrijver K, Sasse A, Sergeant M, et al. Lymphogranuloma venereum outbreak in men who have sex with men (MSM) in Belgium, January 2004 to July 2005. Euro Surveill. 2005;10(9):E050929.3.
13. Kenyon C, Lynen L, Florence E, Caluwaerts S, Vandenbruaene M, Apers L, et al. Syphilis reinfections pose problems for syphilis diagnosis in Antwerp, Belgium - 1992 to 2012. Euro Surveill. 2014;19(45):20958.
14. Workowski KA, Bolan GA, Centers for Disease Control and Prevention. Sexually transmitted diseases treatment guidelines, 2015. MMWR Recomm reports Morb Mortal Wkly report Recomm reports. 2015;64(RR-03):1–137.
15. Ribas SG, Heyndrickx L, Ondoa P, Fransen K. Performance evaluation of the two protease sequencing primers of the Trugene HIV-1 genotyping kit. J Virol Methods. 2006;135(2):137–42.
16. Pineda-Peña A-C, Faria NR, Imbrechts S, Libin P, Abecasis AB, Deforche K, et al. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. Infect Genet Evol. 2013;19:337–48.
17. Hightower GK, May SJ, Pérez-Santiago J, Pacold ME, Wagner GA, Little SJ, et al. HIV-1 clade B pol evolution following primary infection. PLoS One. 2013;8(6):e68188.
18. Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis. 2016.

19. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.

20. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870–4.

21. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.

22. Sukumaran J, Holder MT. DendroPy: a Python library for phylogenetic computing. Bioinformatics. 2010;26(12):1569–71.

23. Ragonnet-Cronin M, Hodcroft E, Hué S, Fearnhill E, Delpech V, Brown AJL, et al. Automated analysis of phylogenetic clusters. BMC Bioinformatics. 2013;14:317.

24. Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80(1):27–38.

25. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med. 2002;21(16):2409–19.

26. Coveney J. FIRTHLOGIT: Stata module to calculate bias reduction in logistic regression. 2015.

27. Hué S, Brown AE, Ragonnet-Cronin M, Lycett SJ, Dunn DT, Fearnhill E, et al. Phylogenetic analyses reveal HIV-1 infections between men misclassified as heterosexual transmissions. AIDS. 2014;28(13):1967–75.

28. Baral S, Logie CH, Grosso A, Wirtz AL, Beyrer C. Modified social ecological model: a tool to guide the assessment of the risks and risk contexts of HIV epidemics. BMC Public Health. 2013;13(1):482.

29. Wertheim JO, Oster AM, Hernandez AL, Saduvala N, Bañez Ocfemia MC, Hall HI. The International Dimension of the U.S. HIV Transmission Network and Onward Transmission of HIV Recently Imported into the United States. AIDS Res Hum Retroviruses. 32(10–11):1046–53.

30. Castro-Nallar E, Pérez-Losada M, Burton GF, Crandall KA. The evolution of HIV: Inferences using phylogenetics. Mol Phylogenet Evol. 2012;62(2):777–92.

31. Boily M-C, Baggaley RF, Wang L, Masse B, White RG, Hayes RJ, et al. Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. Lancet Infect Dis. 2009;9(2):118–29.

32. Little SJ, Kosakovsky Pond SL, Anderson CM, Young JA, Wertheim JO, Mehta SR, et al. Using HIV Networks to Inform Real Time Prevention Interventions. Harrigan PR, editor. PLoS One. 2014;9(6):e98443.

33. Wang X, Wu Y, Mao L, Xia W, Zhang W, Dai L, et al. Targeting HIV Prevention Based on Molecular Epidemiology Among Deeply Sampled Subnetworks of Men Who Have Sex With Men. Clin Infect Dis. 2015;61(9):1462–8.

34. Haider AH, Schneider EB, Kodadek LM, Adler RR, Ranjit A, Torain M, et al. Emergency Department Query for Patient-Centered Approaches to Sexual Orientation and Gender Identity. JAMA Intern Med. 2017;177(6):819.

Section II
Syphilis Diagnostics

Chapter 4

Characterizing the Syphilis-Causing *Treponema pallidum* ssp. *pallidum* Proteome Using Complementary Mass Spectrometry

Authors: Kara K. Osbak[1*]; Simon Houston[2]; Karen V. Lithgow[2]; Conor J. Meehan[3]; Michal Strouhal[4]; David Šmajs[4]; Caroline E. Cameron[2]; Xaveer Van Ostade[5]; Chris R. Kenyon[1,6]; Geert A. Van Raemdonck[1,5]

Affiliations

1. HIV/STI Unit, Institute of Tropical Medicine, Antwerp, Belgium
2. Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada
3. Unit of Mycobacteriology, Institute of Tropical Medicine, Antwerp, Belgium
4. Department of Biology, Faculty of Medicine, Masaryk University, Brno, Czech Republic
5. Laboratory for Protein Science, Proteomics and Epigenetic Signaling (PPES) and Centre for Proteomics (CFP), University of Antwerp, Wilrijk, Belgium

I contributed to the experimental work, data analysis, interpretation and manuscript writing.

Published in PLOS Neglected Tropical Diseases Journal, September 2016
doi: 10.1371/journal.pntd.0004988

Conflicts of Interest and Sources of Funding
This work was supported by the grants from the Flanders Research Foundation, SOFI B Grant to CK, a Public Health Service Grant from the National Institutes of Health to CEC, (grant # AI-051334) and a grant from the Grant Agency of the Czech Republic to DS and MS (P302/12/0574, GP14-29596P). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Background

The spirochete bacterium *Treponema pallidum* ssp. *pallidum* is the etiological agent of syphilis, a chronic multistage disease. Little is known about the global *T. pallidum* proteome, therefore mass spectrometry studies are needed to bring insights into pathogenicity and protein expression profiles during infection.

Methodology/Principal Findings

To better understand the *T. pallidum* proteome profile during infection, we studied *T. pallidum* ssp. *pallidum* DAL-1 strain bacteria isolated from rabbits using complementary mass spectrometry techniques including multidimensional peptide separation and protein identification via matrix-assisted laser desorption ionization-time of flight (MALDI-TOF/TOF) and electrospray ionization (ESI-LTQ-Orbitrap) tandem mass spectrometry. A total of 6033 peptides were detected, corresponding to 557 unique *T. pallidum* proteins at a high level of confidence, representing 54 % of the predicted proteome. A previous gel-based *T. pallidum* MS proteome study detected 58 of these proteins. One hundred and fourteen of the detected proteins were previously annotated as hypothetical or uncharacterized proteins; this is the first account of 106 of these proteins at the protein level. Detected proteins were characterized according to their predicted biological function and localization; half were allocated into a wide range of functional categories. Proteins annotated as potential membrane proteins and proteins with unclear functional annotations were subjected to an additional bioinformatics pipeline analysis to facilitate further characterization. A total of 116 potential membrane proteins were identified, of which 16 have evidence supporting outer membrane localization. We found 8/12 proteins related to the paralogous *tpr* gene family: TprB, TprC/D, TprE ,TprG, TprH, TprI and TprJ. Protein abundance was semi-quantified using label-free spectral counting methods. A low correlation (r=0.26) was found between previous microarray signal data and protein abundance.

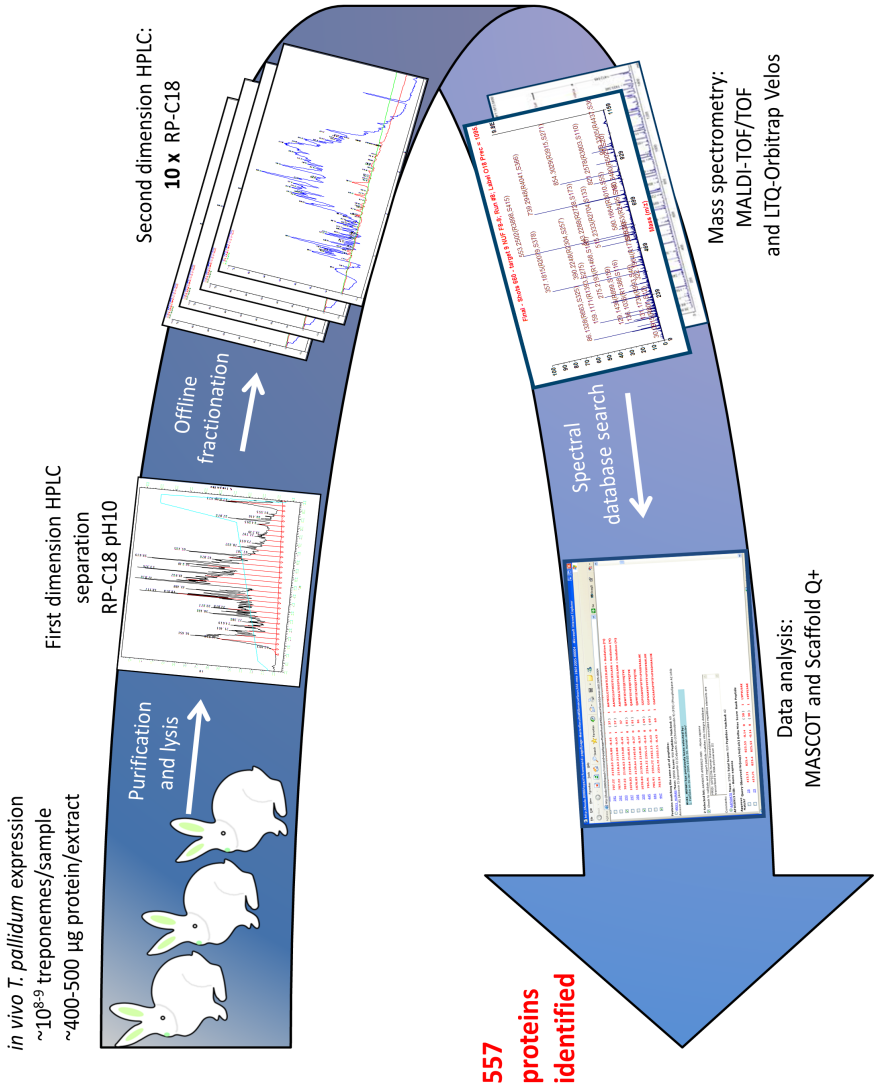
Conclusions

This is the most comprehensive description of the global *T. pallidum* proteome to date. These data provide valuable insights into *in vivo T. pallidum* protein expression, paving the way for improved understanding of the pathogenicity of this enigmatic organism.

Author Summary

Syphilis remains a major cause of morbidity and mortality worldwide. The bacterium causing syphilis, *Treponema pallidum* ssp. *pallidum*, has evolved into a highly distinctive organism that is only able survive (and be propagated) in mammals. In humans it can evade the immune system for decades with devastating consequences. Much remains to be learned about how it accomplishes this. Only a minority of its predicted proteins have been detected experimentally thus far. We aimed to more comprehensively characterize the proteins of this organism. Since it cannot be cultured *in vitro*, we cultured *T. pallidum* in rabbits and analyzed extracted proteins using different mass spectrometry methods, a manner of detecting proteins with high accuracy. In total, we detected more than half of the predicted number of proteins that could be expressed by this bacterium (N = 557). For approximately half of the proteins, we succeeded in characterizing their predicted cellular location using an array of bioinformatic tools and catalogued their function. This is the most comprehensive analysis of the *T. pallidum* proteome to date. This study lays the groundwork for other protein investigations of this unique organism.

Graphical Abstract



Introduction

Treponema pallium ssp. *pallidum*, henceforth referred to as *T. pallidum*, is the causative agent of syphilis, a multistage chronic disease with an estimated 8 million new cases per year [1]. Recent outbreaks of syphilis infection among certain populations such as men who have sex with men (MSM) [2], together with continuing substantial perinatal morbidity and mortality attributed to congenital syphilis infections [3], highlight the need for improved diagnostics and vaccine development.

T. pallidum is an obligate microaerophilic bacterial pathogen [4–6] that is aptly suited to invading mammalian tissue by the use of endoflagella that produce undulations in travelling planar waves [7], thereby driving its characteristic corkscrew-like movement [8]. The membrane of *T. pallidum* lacks lipopolysaccharide (LPS) and the loosely associated fragile outer membrane contains a low amount of proteins [8–11]. Many biomedical experimental approaches such as genetic manipulation have been hampered by its lack of *in vitro* cultivability [12]. Despite these limitations, numerous studies using *T. pallidum* harvested from the experimental rabbit model have increased our basic biological understanding of this unique organism, including the description of the genome [13,14], transcriptome [15] and proteome [16,17].

The *T. pallidum* Nichols strain genome was sequenced for the first time in the late 1990's [13], revealing only 1041 predicted open reading frames (ORFs) on a 1.14 Mb circular chromosome, making it one of the smallest human pathogen genomes. Resequencing of the Nichols strain [14] identified 102 errors that were predicted to affect protein-coding genes and reduced the number of ORFs to 1039, 968 of which are predicted to be protein coding. Similar to other obligate pathogens such as *Mycoplasma pneumoniae* [18], *T. pallidum* is predicted to have lost many non-essential genes though genome reduction. This theory is supported by extensive genome-wide transcriptional analyses [15], which revealed the uniform expression of almost all *T. pallidum* genes during experimental rabbit infection. Consequently, *T. pallidum* has severely limited metabolic and biosynthetic capabilities, rendering it highly dependent on the host milieu and nutrients [19].

The predicted proteins found within *T. pallidum* range in size from 3,235 to 172,869 Da with an average size of 37,771 Da [13,20]. Early studies on *T. pallidum* polypeptides, including pre-MS analysis gel-based techniques and the use of recombinant DNA technology has been extensively reviewed by Norris *et al*. [17] and Schouls [21]. A large scale *T. pallidum* recombinant protein study included the construction of a

bacterial artificial chromosome (BAC) library containing 901 of the 1039 *T. pallidum* predicted proteins for expression in *Escherichia coli* [22,23]; many of the expressed proteins were reactive with sera from syphilis-infected rabbits and/or humans [24,25] at different stages of infection as determined by serological reactivity studies. Subsequently, McGill *et al.* conducted a *T. pallidum* proteome investigation on *in vivo* expressed *T. pallidum* using gel-based approaches complemented with MALDI-TOF Mass Spectrometry (MS) and peptide mass fingerprinting [16]. A total of 88 polypeptides were identified and the immunoreactive potential of select proteins was characterized. Numerous bioinformatic approaches have been used to characterize *T. pallidum* proteins, including lipoprotein characterization [26], the determination of potential outer membrane proteins [27] and the re-annotation of *T. pallidum* strain SS14 hypothetical proteins [28]. However, despite rigorous analyses and major advances in genome sequencing, approximately 30 % of *T. pallidum* proteins still have no known orthologues and at present cannot be assigned a biological function [13]. This 'unknown' category of proteins may represent an arsenal of genes encoding virulence factors specific for *T. pallidum* [20].

Progress has been made on understanding virulence and persistence strategies of this unique pathogen. Genetic sequence diversity is primarily localized in six hot spots [29] in *T. pallidum* ssp. *pallidum* and *T. pallidum* ssp. *pertenue* (the causative agents of Yaws), including regions encoding members of the paralogous *tpr* gene family consisting of 12 genes categorized into subfamilies I (*tpr*C, D, F and I), II (*tpr*E, G and J) and III (*tpr*A, B, H, K and L). The Tpr proteins contribute to antigenic variation that aids in immune evasion [30]. Nonreciprocal gene conversion occurs between donor sites and several variable regions (V1-7) in *tprK* [31] and these variable regions in the encoded protein are targets of the host humoral response during infection [32–35]. Host immune pressure is capable of selecting against certain TprK sequence epitopes [36] and TprK sequence variability can help evade the host immune response [35] during infection. Recombinant protein studies have confirmed surface exposure, bipartite architecture and porin function related to the outer membrane proteins Tpr C/D [37] and TprI [38]. Moreover, *T. pallidum* lipoproteins, of which Tp47 is the most widely studied [39–42], play an important role in immune system activation and evasion as reviewed by Kelesidis *et al.* [43].

With the recent evolution of robust highly sensitive tandem MS instrumentation, the comprehensive description of bacterial proteomes, also referred to as shotgun proteomics (reviewed by Semanjski *et al*. [44]), is achievable. Many current state-of-the-art proteomic studies have approached 80 % coverage of the predicted expressed

proteome [45,46]. The study of pathogens expressed *in vivo* is of particular interest since this would be the closest approximation of human pathophysiological conditions. For example, previous studies on the *Mycoplasma tuberculosis* proteome from guinea pig infected lungs during early and chronic stages of disease [47] have provided valuable insights into pathogen protein expression. However, interference of host proteins present in large excess can hinder MS detection of low abundance pathogen proteins. Thus, several strategies have been used to overcome issues of sample complexity to enrich bacteria cells, such as the use of density gradient centrifugation [16].

Using highly sensitive non-gel based complementary proteomic techniques, we sought to further elucidate the global proteome of *T. pallidum* in order to gain insights into the fundamental physiological state of *T. pallidum* during rabbit infection. Three biological replicates of *in vivo* cultured *T. pallidum* were subjected to multidimensional chromatographic separation and tandem MS/MS analysis whereby 557 *T. pallidum* proteins were identified at a high level of confidence, representing 54 % of the predicted proteome. This is the first description of 499 *T. pallidum* proteins expressed *in vivo*, of which 106 were annotated as uncharacterized/hypothetical proteins. Detected proteins were comprehensively analysed to predict cellular localization and function. This unique 'snapshot' view of the *T. pallidum* proteome during infection extends our understanding of *T. pallidum* pathogenesis and forms the basis for further proteome investigations.

Methods

Rabbit inoculation and *T. pallidum* purification using Percoll density gradient centrifugation

Three biological samples, hereafter referred to as samples TPA-A, TPB-B and TPC-C, originated from three New Zealand White rabbits that were inoculated intra-testicularily with *T. pallidum* DAL-1 strain bacteria according to established methods [48]. Inoculations originated from two different bacterial stocks of DAL-1 strain bacteria, whereby sample TPB-B and TPC-C originated from the same stock. When peak orchitis was reached, on average 11-14 days post-inoculation, rabbits were sacrificed using T61 administration according to the manufacturer's instructions and the bacteria was extracted from the testes and purified using Percoll density gradient centrifugation as previously described [49]. Briefly, collected organisms were separated from host cellular gross debris by low-speed centrifugation at 34 800 g for 30 minutes followed by gradient separation via ultra-centrifugation at 100 000 g for 1

hour. Bacteria were quantified using darkfield microscopy and a counting chamber. For sample TPA-A, approximately 10^{8-9} treponemes were re-suspended and stored in 1 mL NaCl solution and frozen at -80 °C. For samples TPB-B and TPC-C, approximately 10^{8-9} treponemes were re-suspended in 1 mL phosphate buffered saline (PBS) (HiMedia Laboratories, Mumbai, India) and frozen at -80 °C. Two samples, TPA-A and TPB-B, were subjected to an extra thaw cycle before protein extraction due to inadvertent thawing during sample shipment. Each rabbit was tested serologically to rule out a naturally occurring infection with *T. paraluiscuniculi.*

Ethics statement

The treponemal DAL-1 strain was propagated in rabbits at the Veterinary Research Institute in Brno, Czech Republic. The handling of animals in the study was performed in accordance with the current Czech legislation (Animal Protection and Welfare Act No. 246/1992 Coll. of the Government of the Czech Republic). These specific experiments were approved by the Ethics Committee of the Veterinary Research Institute (Permit Number 20-2014).

T. pallidum protein sample preparation

Cell lysis of the purified *T. pallidum* extract was performed by conducting three consecutive freeze-thaw cycles, followed by ultrasonication on ice with an amplitude of 50 % and a pulser frequency of 2 seconds for 2 minutes (Sonics, Vibra cell; Newton USA). Protein concentration was determined by loading a small fraction of the lysed sample on a high performance liquid chromatographic (HPLC) reversed phase C4 system that was calibrated using a serial dilution of a protein standard mixture. Protein concentrations were determined based on the area under the curve (AUC) at 214 nm. Approximately 400-500 μg of protein was extracted from each biological replicate; a large proportion of this amount was host protein in the form of albumin. Samples were acetone precipitated by adding 6 volumes of LC-MS grade acetone (Biosolve, Valkenswaard, Netherlands) and incubated overnight at -20 °C. In all cases, lo-bind Eppendorf tubes (Eppendorf, Hamburg, Germany) were used to ensure high recovery rates of proteins/peptides.

Protein enzymatic digestion

Following protein precipitation, protein samples were re-suspended in 50 mM Tris-HCl/6 M urea/5 mM DTT/10 % beta-mercaptoethanol (25 μL/100 μg protein) at pH 8.7. For the denaturation and reduction process all samples were incubated at 65 °C during 1 hour. Subsequently, proteins in all fractions were diluted in 50 mM Tris-HCl/ 1 mM $CaCl_2$ (75 μL/100 μg protein) and alkylated by adding 200 mM iodoacetamide (10 μL/100 μg protein) during 1 hour at room temperature. Proteomics-grade

modified trypsin (Promega, Madison, Wisconsin, United States (US)) was added at a 30:1 protein-to-enzyme ratio. After incubation at 37 °C for 18 hour the digestion was stopped by freezing the samples.

Peptide separation by reversed phase C18 at high pH (1st dimension)

After tryptic digestion, peptides were separated in a first dimension based on hydrophobicity at high pH by using a reversed phase C18 column (X!Select, CSH, RP-C18, 2.1 x 150 mm, 3.5 µm, Waters) connected to a Waters Alliance e2695 HPLC bio-system and a Waters 996 PDA detector (Waters Corporation, Milford, MA, US). Solvent A contains 200 mM ammonium formate at pH 10, while solvent C contains 100% water and solvent D 100% acetonitrile (ACN) (LC-MS grade, Biosolve, Valkenswaard, Netherlands). During the chromatographic run, an ACN gradient was performed, while continuously 10% of solvent A was added to become an overall pH of 10 during the entire run. The following gradient was used at a constant flow rate of 200 µL/min: 5 % to 15 % D over the first 5 min, 15 % to 40 % D over 80 min, 40 % to 90 % D over 8 min, 5 min 90 % D, and 90 % to 5 % D over 2 min. In total, 30 fractions were collected starting from 10 to 100 min with an interval of 3 min/fraction. The peptide concentration of the different fractions was determined based on the AUC at 214 nm. Fractions were pooled in a concatenated way (e.g. fractions 1, 11 and 21) to obtain optimal orthogonality, yielding in total 10 fractions for further analysis. Collected fractions were lyophilized and re-suspended in RP mobile phase (97 % water, 3 % ACN, 0.1 % FA).

Peptide separation by micro-capillary reversed phase C18 (2nd dimension)

Peptide fractions were separated in a second dimension using an Agilent 1100 series micro-capillary HPLC system (Agilent Technologies, Waldbronn, Germany). For each fraction 15 µg of peptides was injected on a Zorbax 300SB-C18 guard column (0.3 mm x 5 mm; particle size 3.5 µm; Agilent Technologies) serially connected with a Zorbax 300SB-C18 analytical RP column (0.3 mm x 150 mm; particle size 3.5 µm; Agilent Technologies). Samples were online desalted by loading the peptides on the guard column before the ACN gradient was started. Solvent A contained 0.1 % formic acid (FA) in water while solvent B contained 0.1 % FA in 90 % ACN /10 % water. Following ACN gradient was performed using the capillary pump with a constant flow rate at 6 µL/min: 5 % to 60 % B in 56.7 min, ramp to 90 % B over 3.3 min persistent 90 % B for 5 min, 85 % B for 5 min and back to equilibrating conditions of 3 % B. Starting from minute 5 until minute 51.7 of the chromatographic run, 350 spots (800 nl/spot) for each fraction were spotted on an Opti-TOF MALDI-target (28 columns x 25 rows; 8 sec interval; 700 spots; 2 runs per target) (Applied Biosystems, Inc.). Afterwards, each

spot was covered with matrix (2 mg/ml α-cyano-4-hydroxycinnamic acid in 70 % ACN; internal calibrant: 93 pmol/ml human [Glu^1]-fibrinopeptide B) using an external syringe pump with a 4 second interval (800 nl matrix/spot) at a flow rate of 12 μL/min.

MALDI-TOF/TOF MS/MS analysis

Spotted fractions were offline analysed using a MALDI ABi4800 proteomics analyser (Applied Biosystems). MALDI-TOF MS-analysis (reflectron mode; laser intensity: 3400; 25 x 20 laser shots per spot; mass-range 800-3000 Da) was performed first, after which precursors were selected with a signal-to-noise (S/N) ratio above or equal to 100. [Glu^1]-fibrinopeptide B (m/z 1570.667) was used as internal standard to calibrate MS-spectra. MALDI-TOF/TOF MS/MS-analysis was performed on the selected MS precursors. A maximum of 50 unique precursors per spot were selected for fragmentation, starting from the precursors with the lowest S/N- ratio. These precursors were ionized (laser intensity: 4300; 25 x 20 laser shots per spot) and fragmented in a collision cell (CID, 1 kV collision energy).

MALDI-TOF/TOF MS/MS spectral data analysis

Spectra from each sample were extracted by Peak Explorer and screened against a *T. pallidum* database UniProt proteomes IDs UP000014259 and UP000000811 using the MASCOT search engine (Matrix Science; version 2.1.03) based on the digestion enzyme trypsin. We chose to screen against the Nichols strain database since the DAL-1 strain proteome is not well annotated and the genetic differences between the strains are minimal and described [50]. The latter database is generally used as the treponemal reference database while the former is a more recent version. Carbamidomethylation of cysteines was listed as fixed modification, while oxidation of methionine was set as a variable modification. A maximum of two missed cleavages of trypsin was tolerated. Mass tolerance was set to 200 ppm for the precursors and 0.20 Da for the fragment ions. The MudPIT scoring algorithm of MASCOT was used. Scaffold Q+ (version Scaffold 4.0.5, Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Because the *T. pallidum* proteome contains several small proteins with just one or a few detectable tryptic peptides, protein identifications based on one unique peptide were only allowed if they fulfilled certain stringent conditions; these criteria were comprised by the peptide prophet algorithm that was performed by Scaffold Q+. Protein identifications were accepted if they could be established at greater than 95.0 % probability according to the protein prophet algorithm.

Protein abundances were estimated based on the spectral counts (SC) of each identified protein by calculating the normalized spectral abundance factor (NSAF) as previously described [51]. In short, this approach includes a normalisation step based on (1) the observable peptides (OP) and (2) on the total number of identified peptides. The NSAF values reflecting an average of the biological and technical runs of each detected proteins are provided in *S3 Table*. Pearson's correlation test and Mann Whitney test were calculated to compare the cDNA/DNA signal data to the NSAF protein abundance data. A P-value of < 0.05 was considered statistically significant. All analyses were performed in Stata 12.0 (StataCorp LP, College Station, TX, USA).

In order to determine whether the identification methodology was stringent enough, the false discovery rate (FDR) was defined on protein level by using a concatenated database consisting of the target spectral database and a shuffled database. Calculation of FDR was performed as follows: 2x false positive identifications / (false positive identifications + true positive identifications) [52]. For all samples, the FDR on protein level had to be less than 5%. Spectra were also screened against the mammalian Swissprot database containing human (*Homo sapiens*) and rabbit (*Oryctolagus cuniculus*) proteomes for spectra verification to prevent assignment of peptides with a conserved amino acid sequence.

Orbitrap Velos LTQ MS/MS Analysis

Nano reverse phase liquid chromatography and mass spectrometry

The peptide mixtures were separated in the second dimension by reverse phase chromatography on an Eksigent nano-UPLC system using an Acclaim C18 PepMap100 nano-Trap column (200 μm x 20 mm, 5 μm particle size) connected to an Acclaim C18 analytical column (75 μm x 150 mm, 3 μm particle size) (Thermo Scientific, San Jose, CA). Peptide fractions were dissolved in mobile phase A, containing 2 % ACN and 0,1 % formic acid and spiked with 20 fmol [Glu^1]-fibrinopeptide B. A linear gradient of mobile phase B (0,1 % FA in 98 % ACN) in mobile phase A (0,1 % FA in 2 % ACN) from 2 to 45 % B in 35 min followed by a steep increase to 95 % mobile phase B in 2 min was used at a flow rate of 350 nl/min. The nano-LC was coupled online with the mass spectrometer using a PicoTip Emitter (New objective, Woburn, MA) coupled to a nanospray ion source. The LTQ Orbitrap Velos (Thermo Scientific, San Jose, CA) was set up in a data dependent MS/MS mode where a full scan spectrum (350–5000 m/z, resolution 60000) was followed by a maximum of ten CID tandem mass spectra (100 to 2000 m/z). Peptide ions were selected as the twenty most intense peaks of the MS scan. Collision induced dissociation (CID) scans were acquired in the LTQ ion trap part of the mass

spectrometer. The normalized collision energy used was 35 % in CID. A dynamic exclusion list of 45 sec for data dependent acquisition was applied.

Orbitrap Velos LTQ MS/MS spectral data analysis

Spectra from each sample were extracted by Proteome discoverer software (Thermo Scientific, San Jose, CA) and screened against a *T. pallidum* database (UniProt ID proteomes IDs UP000014259 and UP000000811) using the MASCOT search engine (Matrix Science; version 2.1.03) based on the digestion enzyme trypsin. Carbamidomethylation of cysteines was listed as fixed modification, while methionine oxidation was set as variable modification. A maximum of two missed cleavages of trypsin was tolerated. Mass tolerance was set to 10 ppm for the precursors and 0.8 Da for the fragment ions. The MudPIT scoring algorithm of MASCOT was used. Further protein identification, quantification and validation procedures were conducted as mentioned above for the MALDI-TOF/TOF analysis. All Orbitrap LTQ mass spectrometric data are available at PeptideAtlas [53]. The identifier is PASS00903.

Identification of known and predicted *T. pallidum* membrane proteins

Initially, all potential membrane proteins were identified from the 557 *T. pallidum* proteins detected by mass spectrometry by: (1) analyzing annotated functions (and subcellular localizations, if available) from all published *T. pallidum* ssp. *pallidum* genome sequences, (2) identification of lipoproteins based on previous bioinformatic analyses performed by Setubal *et al*. [26], (3) identification of rare outer membrane proteins based on previous bioinformatic analyses performed by Cox *et al*. [27], and (4) by additional review of experimental findings in the scientific literature. Next, all potential membrane proteins (and proteins annotated as 'uncharacterized', 'hypothetical' or 'conserved hypothetical') were analyzed using 5 bioinformatic prediction tools. The SignalP 4.1 server [54] and the LipoP 1.0 server [55] were used to predict the presence and location of potential signal peptide cleavage sites and lipoprotein signal peptides, respectively. PSORTb version 3.0.2 [56] was used to predict protein subcellular localization. TMHMM server version 2.0 [57] and PRED-TMBB [58] were used for predicting the presence and location of transmembrane alpha-helices and beta strands, respectively. Proteins with unclear subcellular localization predictions using the above bioinformatic pipeline were further analyzed using the following eight prediction tools. CELLO version 2.5 [59] was used to predict subcellular localization. Philius [60], Phobius [61], Octopus/Spoctopus [62], HMMTOP version 2.0 [63], and TMpred were used for the prediction of transmembrane alpha-helices BOMP [64] and TMBETADISC-RBF [65] were used for the prediction of beta-barrel outer membrane proteins.

Assignment of orthologous functional categories and cellular localization

The eggNOG version 4 database (retrieved 21/04/15) was used to assign COG and NOG categories to all genomes. First all proteins per sample were compared to the eggNOG database using USEARCH version 7.0.959 with an e-value of 1e-30 and a bit-score cut-off of 70% of the top hit to ensure only close matches were retrieved and reduce the likelihood of spurious annotations. An eggNOG membership is assigned to each protein if 70% of the UBLAST hits belong to the same eggNOG member. Distinctions are then made between proteins with no UBLAST hit to any eggNOG sequence (no_hit) and over 70% of hits to a member that is not assigned an eggNOG code (none). Annotations are also clustered at the 25 higher COG functional category levels as per the eggNOG assignments. Classification of proteins according to their cellular location was achieved using data extraction from online databases (Swissprot) and the methods as outlined for the membrane localized proteins.

Results and Discussion

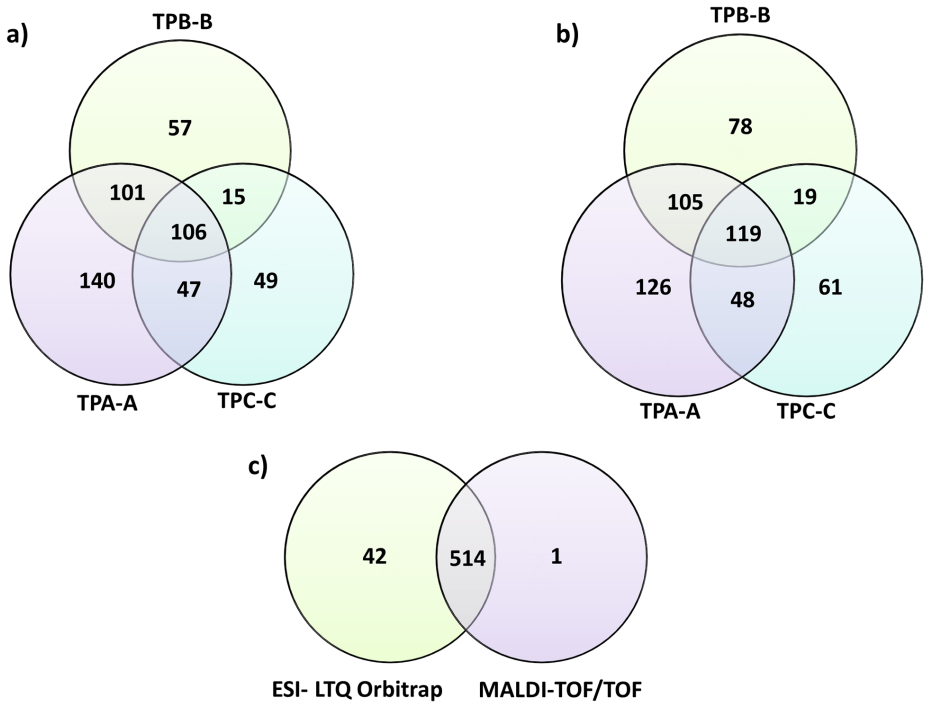
Mass spectrometry analysis

In short, from the three biological replicates, a total of 6033 *T. pallidum* peptides were detected corresponding to 557 proteins and 54 % of the total predicted proteome (*S1 Table*). Proteins ranged in size from 6-173 kDa with a pI range of 4.15 to 12.05. Acquired spectra were screened against two Nichols strain UniProt proteomes whereby three extra proteins (TP0248, TP0651 and TP0922) were uncovered compared to when solely screened against the Nichols reference UniProt proteome (ID: UP000000811) [13]. In the resequenced proteome (ID: UP000014259) [14] three of these proteins were below the 150bp annotation limit. We found 57/102 proteins containing previously reported sequencing errors [14] compared to the original genome analysis [13], including two genes with an authentic frameshift, 14 reannotated gene fusions and 5 novel ORFs reannotated in the new proteome (*S2 Table*).

Pertaining to the individual samples, 394/398 (TPA-A), 279/321 (TPB-B) and 217/247 (TPC-C) proteins were uniquely identified by MALDI-TOF/TOF and ESI- MS/MS analysis, respectively, of which 106 (MALDI-TOF/TOF) and 119 (ESI- MS/MS) proteins were present in all three biological samples (*Fig 1A/B*). Only 31 proteins were found with less than 2 peptide identifications in one biological and one technical run (*S3 Table*). For the individual MS analyses (MALDI- TOF/TOF *versus* ESI- LTQ Orbitrap MS/MS detection), 514 proteins were detected by both methods (*Fig 2C*). Only one and 42 additional proteins were exclusively identified by MALDI- TOF/TOF MS/MS analysis and ESI-MS/MS analysis, respectively (*Fig 2C*) indicating that we are

possibly approaching the upper limit of the detectable *T. pallidum* proteome and that the non-detected proteins are 1) not expressed, or 2) are expressed at a very low level. All *T. pallidum* designated spectra were rescreened against human and rabbit UniProt protein databases and no overlap was found.

Fig 1 *Venn Diagrams depicting the total number of unique* T. pallidum *proteins identified per rabbit biological replicate (N=3) analyzed by (A) Matrix-assisted laser desorption/ionization time of flight and (B) Electrospray Ionization LTQ- Orbitrap Velos MS/MS. All spectra were screened against the UniProt databases (ID: UP000000811 & UP000014259), with a peptide and protein identification confidence interval of 95%. There was considerable overlap between the complementary MS analytical methods whereby an additional 42 treponemal proteins were found in the Orbitrap analysis as depicted in diagram (C).*



A previous proteomics study of *in vivo* rabbit expressed *T. pallidum* Nichols strain bacteria [16] detected 88 proteins using MALDI-TOF MS with peptide mass fingerprinting. We detected 58 of these proteins, therefore, to date 58 % of the whole *T. pallidum* predicted proteome has been detected using MS methods. We failed to detect 30 of these previously identified proteins as outlined in *Table 1*. The protein detection differences between the studies could be attributed to different experimental methods, for example gel-based *versus* liquid chromatographic separation, which may favor the detection of proteins with certain physiochemical characteristics. Although the differences on the genomic level between the two strains are minimal [50], different duplication rates or other strain characteristics could contribute to different protein expression profiles found between these studies.

Table 1 T. pallidum *proteins found in previous proteome investigation (McGill* et al.*) [16] but not detected in this study*

TP number	UniProt Accession Number#	Protein Name(s)	Molecular Weight (kDa)	COG category code	cDNA/DNA signal ratio Smajs et al. [15]
TP0112	O83150	Aminopeptidase C (PepC)/Bleomycin hydrolase	51	none	1.0
TP0139	O83175	Trk family potassium (K+) transporter, NAD+ binding protein	25	P	0.6
TP0201	R9UV93	50S ribosomal protein L5	21	none	1.8
TP0216	R9UVA6	Chaperone protein DnaK/Heat shock 70 kDa protein	68	none	4.5
TP0239	R9UU25	50S ribosomal protein L10	20	none	1.9
TP0259	O83283	Peptidoglycan binding (LysM domain-bacterial cell wall degradation)	29	S	5.4
TP0298	R9UW99	ABC superfamily ATP binding cassette transporter, membrane protein/ Membrane lipoprotein TpN38	38	none	3.3
TP0349	R9UUA6	Peptidyl-prolyl cis-trans isomerase/ Metallochaperone SlyD	18	O	2.2
TP0356	O83375	Putative RNA-binding protein	12	none	10.7
TP0424	R9UWN5	Two-sector ATPase, V(1) subunit E	19	none	4.7

TP number	UniProt Accession Number#	Protein Name(s)	Molecular Weight (kDa)	COG category code	cDNA/DNA signal ratio Smajs *et al.* [15]
TP0435	R9UWP1	Copper resistance lipoprotein NlpE/17 kDa lipoprotein	17	none	5.1
TP0448	R9UUM3	Uracil phosphoribosyltransferase	41	F	0.3
TP0453	O67998	Outer membrane protein/30kLp	32	none	1.0
TP0476	R9UWS9	Acetate kinase/Acetokinase	49	none	1.3
TP0505	O83518	Hexokinase	48	none	1.8
TP0537	R9UTC2	Triosephosphate isomerase	27	G	1.9
TP0554	O83565	Phosphoglycolate phosphatase	25	R	1.1
TP0556	R9UWZ4	Aspartate--ammonia ligase/Asparagine synthetase A	37	E	1.8
TP0584	R9UTG3	Uncharacterized protein	54	none	0.8
TP0611	O83620	ABC superfamily ATP binding cassette transporter, ABC protein	28	none	1.2
TP0655	O83661	Spermidine/putrescine ABC superfamily ATP binding cassette transporter, binding protein	40	E	0.6
TP0734	R9UXE0	Purine-nucleoside phosphorylase	25	F	1.6
TP0769	R9UTV1	Treponemal membrane protein B/ Antigen TmpB	37	S	5.2
TP0821	R9UVI4	Lipoprotein/ Membrane lipoprotein TpN32; 29 kDa protein	29	P	1.5
TP0823	O83795	Desulfoferrodoxin (Rbo)	14	C	3.2
TP0862	O83834	Peptidyl-prolyl cis-trans isomerase	28	none	5.3
TP0925	R9UUA1	Nitrogenase (Flavodoxin)	16	C	9.4
TP0964	O83930	ABC transporter, ATP-binding protein	25	V	0.9
TP0971	R9UX80	Tp34 lipoprotein/ 34kDa membrane antigen	22	none	12.6
TP1013	R9UUH7	10 kDa chaperonin/GroES protein/Protein Cpn10	9	O	4.6

Legend: #: UniProt Proteome ID: UP000014259; Clusters of Orthologous Genes (COG)

Detection of possible *T. pallidum* heterogeneous sites at the protein level

All *T. pallidum* protein sequences were screened for possible heterogeneous sites by searching the spectral databases for amino acid sequences containing sites designated with 'X', meaning 'undetermined amino acid site'. Heterogeneous sites were defined as differing amino acids located at the same coordinate 'X' in the same protein sequence. A total of 25 *T. pallidum* proteins contained sites designated as 'X', of which four proteins were identified with heterogeneous peptide matches at site 'X' (*Table 2*). Amino acid sequence diversity was found within one sample for three proteins, TP0082 (TPC-C), TP0248 (TPC-C) and TP0922 (TPB-B). Protein TP0692 contained two peptides with heterogeneous sites within two samples (TPA-A/TPC-C). This is the first account of sequence heterogeneity at the protein level for these particular proteins. Although the amino acid sequence designation is of high confidence (95 %), cautious interpretation of these results is warranted as *de novo* peptide sequencing was not utilized so these analyses could represent falsely identified sites, therefore, further research is advised. *Treponema pallidum* intra-strain nucleotide sequence heterogeneity has been reported previously [14,66,67], including *tprK* [22,31,32,66,68,69] and heterogeneity in four DAL-1 strain genes related to chemotaxis and metabolism [66]. The functional relevance of this observed intra-strain variability in these proteins in currently unknown.

Table 2 *Potential heterogeneous sites identified within peptides of MS detected* T. pallidum *proteins*

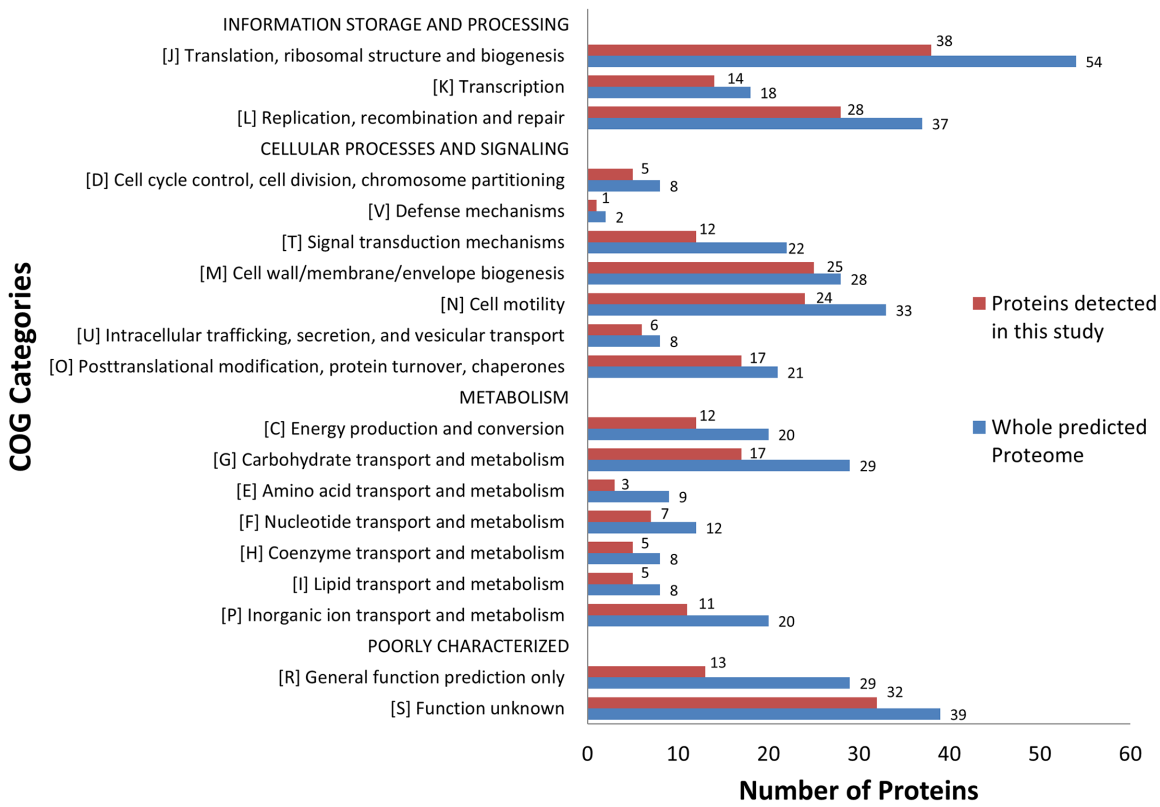
					Amino acid detected at site 'X' in peptide sequence					
					MALDI TOF/TOF			ESI-LTQ-Orbitrap		
					Sample			Sample		
UniProt Accession Number#	**TP number**	**Protein Name**	**Molecular Weight (kDa)**	**Identified Peptide Sequence**	**TPA-A**	**TPB-B**	**TPC-C**	**TPA-A**	**TPB-B**	**TPC-C**
O88098	TP0082	Formate hydrogenlyase transcriptional activator (FhlA)	66	(R)LYPIXNAR(K)			T			D
O83276	TP0248	Uncharacterized lipoprotein	15	(R)AYELXER(S)			T			H
O83690	TP0692	Protein RecA	44	(K)TLRRXASR(G)	Y		H	T		P
O83892	TP0922	Uncharacterized protein	33	(R)ALXGNDPSAAR(V)		V			A	

Legend: #: UniProt Proteome ID: UP000014259; Matrix-assisted laser desorption ionization-time of flight (MALDI-TOF/TOF); electrospray ionization LTQ-Orbitrap tandem mass spectrometry (ESI-LTQ-Orbitrap), X= undetermined amino acid site

Bioinformatic characterization of detected *T. pallidum* proteins

Bioinformatic analyses assigned 279 detected proteins to 19 higher Clusters of Orthologous Genes (COG) functional category levels according to their eggNOG assignments. Distributional description of these proteins and their categorical frequencies are depicted in *Fig 2* and extensive descriptions, including COG/NOG codes for all detected proteins, can be found in *S2 Table*. Of the proteins that were delegated into a clear functional category, the highest representative categories were 'J' (translation, ribosomal structure and biogenesis) (17 %) and 'L' (replication, recombination and repair) (12%). High category coverage was found for the categories 'M' (cell wall/membrane/envelope biogenesis) and 'O' (posttranslational modification, protein turnover and chaperones) with 25/28 and 17/21 proteins found, respectively. Forty-five proteins fell under category 'S' or 'R', indicating poor functional characterization. A total of 9 proteins had no UBLAST hit to any eggNOG sequence (category 'no_hit'), of which 5 proteins were ribosomal and 4 were uncharacterized. Many proteins (N=275) were at least 70 % homologous to a protein member not assigned an eggNOG code (category 'none') indicating that the *T. pallidum* proteome is very unique compared to other organisms. Six proteins were categorized under multiple COG categories. In almost all of the COG categories, more than half of the predicted proteins were detected, supporting the theory that *T. pallidum* has shed its unnecessary genes during its evolution [13].

Fig 2 *Bar Diagram depicting the distribution of the detected and undetected* T. pallidum *proteins distributed in 19 COG functional categories. Blue bars represent MS detected proteins in this study. Red bars represent all predicted proteins in the* T. pallidum *proteome. The functional category was automatically determined for genes that could be placed in Clusters of Orthologous Groups (COGs). For genes with more than one COG category both categories were used.*



The *T. pallidum* Nichols and SS14 strain genomes differ minimally [14], thus in the case of genetically congruent ORFs we extrapolated recent *T. pallidum* strain SS14 hypothetical protein function re-annotations [28] to 22 'uncharacterized/hypothetical' proteins detected in this analysis. In total, 114 proteins remained classified as 'uncharacterized proteins/ hypothetical proteins'. This category did not include 17 proteins with ambiguous "putative" membrane protein descriptions. A previous study [16] detected eight of these uncharacterized proteins, meaning that this is the first account of 106 'uncharacterized/ hypothetical' proteins at the protein level. This uncharacterized area of the *T. pallidum* proteome may contain novel proteins with important roles in pathogenesis and even represent novel biomarker, treatment or vaccination targets.

Predicted cellular localization of detected *T. pallidum* proteins

The global classification of detected proteins according to their cellular localization was achieved by screening online databases such as UniProt and by reviewing relevant literature. The cellular localization of the proteins was predicted for 292/557 proteins; these were largely localized in the cytoplasm (N=97, 17 %), membrane (N=116, 21 %), ribosome (N=33, 6 %) and flagella (N=19, 3 %). A schematic breakdown of the predicted cellular localizations for all detected proteins can be found in *Fig 3*, with comprehensive information for each protein provided in *S2 Table*.

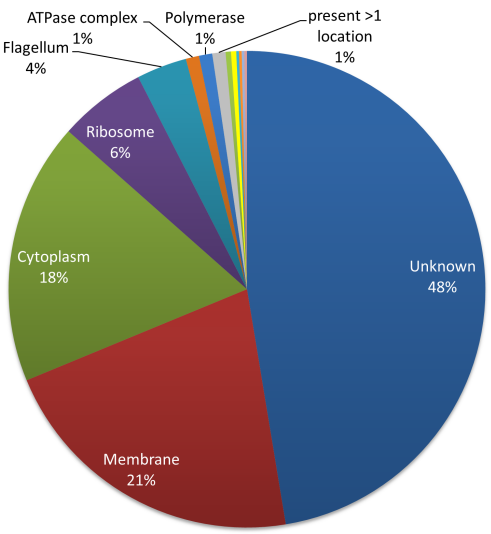
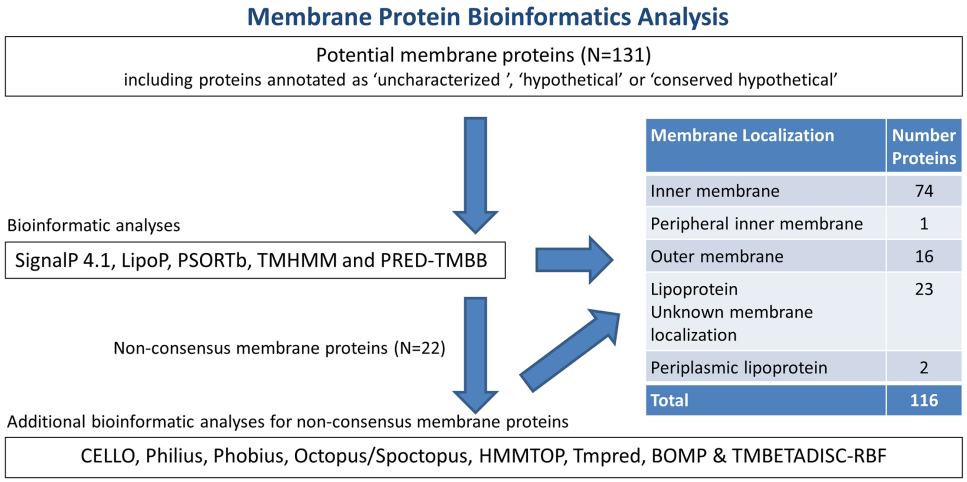


Fig 3 *Pie diagram depicting the subcellular localization of* T. pallidum *detected by MS analysis (N=557).* Almost half of the detected proteins (N=265; 48%) did not have an annotated cellular location. Of the known locations, membrane (N=116; 21%), cytoplasm (n=99; 18 %), ribosomal (N=33; 6 %) and flagella (N=19; 3 %) were the most represented cellular localizations.

Detected proteins were subjected to additional bioinformatic pipeline analyses in order to identify potential membrane proteins as detailed in the methods section. In short, all potential membrane proteins (N=131; including proteins annotated as 'hypothetical') were analyzed using five bioinformatic prediction tools: SignalP 4.1 [54], LipoP [55], PSORTb [56], TMHMM [57] and PRED-TMBB [58].

Proteins with unclear subcellular localization predictions using the above bioinformatics pipeline (N=25) were further analyzed using an additional eight prediction tools including CELLO [59], Philius [60], Phobius [61], Octopus/Spoctopus [62], HMMTOP [63], Tmpred, BOMP [64] and TMBETADISC-RBF [65]. A results summary of this analysis can be found in *Fig 4* and *S4 Table*.

Fig 4 *Bioinformatics pipeline analysis for potential* T. pallidum *membrane proteins*



In total, 116 proteins were designated as 'membrane' localized, with a majority (64 %; N=74) located within the inner membrane. Sixteen proteins (14 %) were predicted to be located in the outer membrane (OM) (*Table 3*). The OM localization of five detected proteins has been experimentally investigated: (TP0117 (TprC)/ TP0131 (TprD) [27,37,70], TprI (TP0620) [38], TP0126, an OmpW homologue [71] and TP0326, a BamA homologue [27,72–74].

The other 11 predicted OM proteins in this analysis were: TprB (TP0011), M23B subfamily peptidase (TP0155), TprE (TP0313), TprJ (TP0621), TP0421, TP0858, TP0324, TP0855, TP0865, TP0923 and TP0969.

Table 3 *Predicted* T. pallidum *outer membrane proteins detected by MS analysis*

TP number	UniProt Accession Number#	Protein Name	References
TP0011	O83055	Tpr protein B	Cox *et al*. 2010 [27]
TP0117*	O88138	Tpr protein C (TprC)/ Tpr protein D (TprD)	Anand *et al*. 2012 [37]; Cox *et al*. 2010 [27]; Centurion-Lara *et al*. 2013 [70]
TP0126	R9UV28	Uncharacterized protein	Giacani *et al*., 2015 [71]
TP0131*	O88138	Tpr protein C (TprC)/ Tpr protein D (TprD)	Anand *et al*. 2012 [37]; Cox *et al*. 2010 [27]; Centurion-Lara *et al*. 2013 [70]
TP0155	O83190	M23B subfamily peptidase	Cameron *et al*. 2004 [75]; Cox *et al*. 2010 [27]
TP0313	O83335	Tpr protein E (TprE)	Cox *et al*. 2010 [27]
TP0324	R9USU0	Putative outer membrane protein	Cox *et al*. 2010 [27]
TP0326	O83346	Outer membrane protein	Cameron *et al*. 2000 [74]; Cox *et al*. 2010 [27] Desrosiers *et al*. 2011 [72]
TP0421	O83436	Uncharacterized protein	Cox *et al*. 2010 [27]
TP0620	R9UV07	Tpr protein I (TprI)	Giacani *et al*., 2005 [76]; Centurion-Lara *et al*. 2013 [70]; Cox *et al*. 2010 [27]; Anand *et al*. 2015 [38]
TP0621	F7IWA5	Tpr protein J (TprJ)	Cox *et al*. 2010 [27]
TP0855	R9UWW7	Putative outer membrane protein	Cox *et al*. 2010 [27]
TP0858	R9UXR8	Uncharacterized protein	Cox *et al*. 2010 [27]
TP0865	R9UU78	Putative outer membrane protein	Cox *et al*. 2010 [27]
TP0923	R9UX30	Putative outer membrane protein	
TP0969	R9UUH1	Putative outer membrane protein	Cox *et al*. 2010 [27]

*Legend: #: UniProt Proteome ID: UP000014259, *tprC and tprD alleles are identical at the genomic level in DAL-1 strain* [50]

A previous *in silico* prediction analysis of the *T. pallidum* genome revealed 46 predicted lipoproteins [26]. Our analysis also detected 25 lipoproteins, 23 with unknown membrane locations and two located within the periplasm (TP0796;

TP0171), including the 15 kDa (Tpp15) lipoprotein (TP0171) and 47 kDa membrane antigen (TP0574). In spirochetes, lipoproteins are highly expressed molecules primarily localized in the periplasm anchored to the outer leaflet of the cytoplasmic membrane [9] where they are thought to modulate immune responses from both innate and adaptive immunity [43,77].

There were ambiguities regarding the subcellular localization of 13 proteins after analysis with the additional prediction tools (*S4 Table*), including TprG (TP0317), TprH (TP0610), ABC superfamily ATP binding cassette transporter (TP0786), flagellar hook length control protein FliK (TP0729) and two TolC-like proteins (TP0967 and TP0968). Of the 49/116 reported membrane proteins that could be designated to a COG category, two categories were most represented: 'P' (inorganic ion transport and metabolism) (N=9) and 'M' (cell wall/membrane/envelope biogenesis) (N=6). This agrees with the predicted biological functional location.

Important to note is the fact that most of the protein localization data is based on computational predictions. These types of predictions have an inherent risk of including false positives and also omitting real OM proteins. Further laboratory work is needed to experimentally confirm the cellular locations of these proteins.

Relative *T. pallidum* protein abundance as determined by spectral counting

We examined the relative abundance of the proteins detected by calculating the NSAF values [78] for the proteins detected in the biological and technical runs; all values are listed in *S3 Table* and the log distribution of all detected proteins can be found in *Fig 5A*. This approach is based on the number of observable peptides and normalizes technical variability between samples [78]. A value of '1' represents the mean protein level for all detected proteins. Proteins with an average NSAF value greater than 5.0 were regarded as 'highly abundant'. A summary of the top 50 highest abundant proteins according to the spectral counting averages is provided in *Table 4*. High abundant proteins (N=103) included two proteins related to redox balance, 22 proteins related to translation, two proteins related to chemotaxis and three ABC family transport proteins. Proteins related to motility were found to be high abundant, including flagellar filament proteins (TP0663; TP0792; TP0868; TP0870) and 3 proteins related to flagellar biosynthesis (TP0403; TP0658; TP0718). The fact that proteins related to motility, transport and chemotaxis are highly expressed can be indicative that these processes are essential and highly utilized for cell survival.

Fig 5 *(a) Histogram depicting the distribution of detected* T. pallidum *proteins sorted according to their protein abundance amounts (Log expression NSAF values) (b) Scatter Plot of Microarray cDNA/DNA ratio signal versus NSAF spectral count average for all MS detected proteins. Statistical analysis revealed no correlation between protein abundance and microarray data (Pearson's correlation coefficient r = 0.26). Proteins not detected did not have a significantly lower microarray signal as calculated by a two-sample Wilcoxon rank-sum test (P = 0.5). Red line designates cutoff for 'high abundant' proteins (NSAF value > 5.0)*

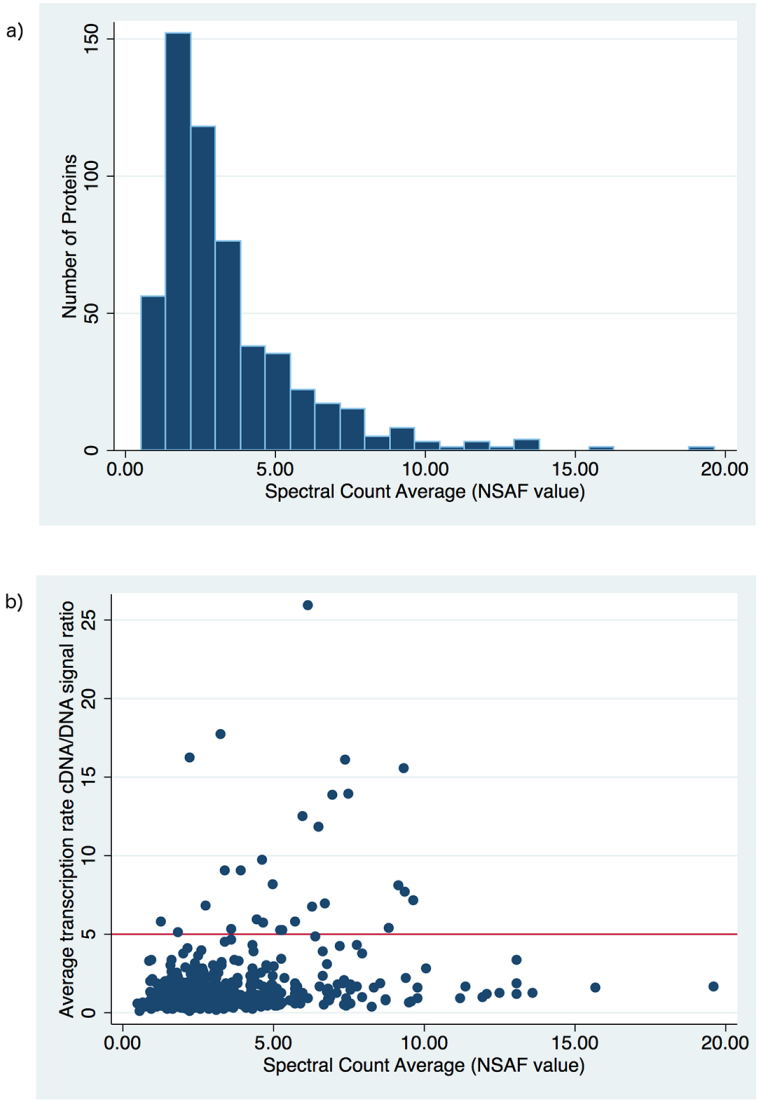


Table 4 *Top 50 most abundant* T. pallidum *proteins detected by MS analysis*

TP number	UniProt Accession Number[#]	Protein Name	Molecular Weight (kDa)	Average NSAF value	cDNA/ DNA signal ratio Smajs *et al.* [15]	GO Biological Process	Cellular location	COG category
TP0214	R9UVZ6	Uncharacterized protein	7	19.6	1.6		Unknown	no_hit
TP0708	R9UXB8	Uncharacterized protein	18	15.7	1.5		Inner membrane	none
TP0385	O83400	Uncharacterized protein	14	13.6	1.2		Unknown	none
TP0951	R9UVX4	50S ribosomal protein L34	6	13.1	1.8	Translation	Ribosome	no_hit
TP0126a[*]	R9UTR8	Uncharacterized protein	23	13.1	1.1		Unknown	none
TP0474	R9UUP5	Probable transcriptional regulatory protein	26	13.1	3.3	Regulation of transcription	Cytoplasm	K
TP0061	R9UTJ4	30S ribosomal protein S18	12	12.5	1.2	Translation	Ribosome	none
TP0048	O83087	Cell shape determination protein CcmA	15	12.1	1.1		Unknown	none
TP0657	R9UWD4	Carbon storage regulator homolog	8	12.0	0.9	mRNA catabolic process; regulation of carbohydrate metabolic process	Unknown	none
TP0409a	R9UT20	Preprotein translocase subunit YajC	14	11.4	1.6		Inner membrane	none

TP number	UniProt Accession Number#	Protein Name	Molecular Weight (kDa)	Average NSAF value	cDNA/ DNA signal ratio Smajs *et al.* [15]	GO Biological Process	Cellular location	COG category
TP0718	R9UXC7	IIISP family Type III (Virulence-related) secretory pathway protein/Flagellar biosynthesis protein FliP	31	11.2	0.8	Bacterial-type flagellum organization; protein secretion	Flagellum/Inner membrane	N
TP0808	R9UWS1	Acyl carrier protein	9	10.1	2.8	Fatty acid biosynthesis; Fatty acid metabolism	Cytoplasm	none
TP0558	O83569	Nickel/cobalt efflux system	32	9.8	0.9		Inner membrane	none
TP0198	R9USD7	30S ribosomal protein S17	10	9.8	1.5	Translation	Ribosome	none
TP0849	R9UU63	50S ribosomal protein L35	8	9.7	7.1	Translation	Ribosome	no_hit
TP0490	O83503	Uncharacterized protein	7	9.6	0.7		Unknown	no_hit
TP0649	O83655	Hemolysin (TlyC)	30	9.5	0.6		Inner membrane	P
TP0213	R9USF3	50S ribosomal protein L17	19	9.4	2.2	Translation	Ribosome	J
TP0072	O83111	Glutaredoxin-related protein	10	9.4	7.6	Cell redox homeostasis	Unknown	none
TP0509	O83522	Alkyl hydroperoxide reductase (AhpC)/Peroxiredoxin	21	9.3	15.5		Unknown	O

TP number	UniProt Accession Number#	Protein Name	Molecular Weight (kDa)	Average NSAF value	cDNA/ DNA signal ratio Smajs *et al.* [15]	GO Biological Process	Cellular location	COG category
TP0919	R9UVS2	Thioredoxin	11	9.2	8.1	Cell redox homeostasis; glycerol ether metabolic process	Unknown	none
TP0362	R9USW9	50S ribosomal protein L28	8	8.9	5.4	Translation	Ribosome	none
TP0368	R9UVP4	Uncharacterized protein	12	8.7	0.8		Inner membrane	none
TP0119	O83156	Amino acid ABC transporter, permease protein (YaeE)	23	8.7	0.7	Transport	Inner membrane	none
TP0195	R9USH3	30S ribosomal protein S3	28	8.5	1.9	Translation	Ribosome	J
TP0604	R9UUZ6	Ribosome-recycling factor	21	8.4	1.6	Protein biosynthesis; translation termination	Cytoplasm	J
TP0045	R9UUU2	Adenosine deaminase	35	8.3	0.4	Hydrolase activity; metal ion binding; deaminase activity; nucleotide metabolic process	Unknown	F
TP0700	R9UV74	Uncharacterized protein	14	8.0	3.7		Unknown	none
TP0315	R9USP4	Uncharacterized protein	23	8.0	0.9		Unknown	none
TP0922	O83892*	Uncharacterized protein	33	7.8	1.6		Unknown	none

TP number	UniProt Accession Number#	Protein Name	Molecular Weight (kDa)	Average NSAF value	cDNA/ DNA signal ratio Smajs *et al.* [15]	GO Biological Process	Cellular location	COG category
TP0887	R9UVP5	30S ribosomal protein S15	10	7.8	4.3	Translation	Ribosome	none
TP0707	R9UTN7	Putative membrane protein	17	7.6	0.6		Inner membrane	none
TP0202	R9UTZ5	30S ribosomal protein S14 type Z	7	7.6	1.8	Translation	Ribosome	no_hit
TP0433	R9UT05	Acidic repeat protein	66	7.6	1.4		Unknown	none
TP0663	R9UV45	Putative flagellar filament outer layer protein FlaA; Tromp-2	27	7.5	13.9	Bacterial-type flagellum-dependent cell motility	Flagellum	none
TP0076	R9UUX9	Sugar ABC superfamily ATP binding cassette transporter, membrane protein	30	7.4	0.4	Transport	Inner membrane	G
TP1029	R9UY97	Uncharacterized protein	25	7.4	0.9		Unknown	none
TP0249	R9UVD9	Flagellar filament outer layer protein (FlaA1)	39	7.4	16.1		Flagellum	N
TP1012	O83976	RNA polymerase sigma factor	35	7.4	0.5	DNA-templated transcription; initiation	Polymerase	K
TP0206a	R9UU00	50S ribosomal protein L30	7	7.4	2.0	Translation	Ribosome	no_hit
TP0194	R9UVW4	50S ribosomal protein L22	14	7.3	1.8	Translation	Ribosome	J
TP0234	R9UU20	50S ribosomal protein L33	7	7.2	4.2	Translation	Ribosome	no_hit

TP number	UniProt Accession Number#	Protein Name	Molecular Weight (kDa)	Average NSAF value	cDNA/ DNA signal ratio Smajs *et al.* [15]	GO Biological Process	Cellular location	COG category
TP0007	R9URZ2	Uncharacterized protein	35	7.1	1.8		Unknown	none
TP0412	O83427	Uncharacterized protein	12	7.1	1.2		Unknown	S
TP0792	R9UU06	Flagellin/ Flagellar filament 33 kDa core protein; Class B	31	7.0	13.8	Bacterial-type flagellum-dependent cell motility	Flagellum	N
TP0654	O83660	Spermidine/putrescine ABC superfamily ATP binding cassette transporter, membrane protein	30	6.9	1.0	Transport	Inner membrane	none
TP0563	O83574	Uncharacterized protein	16	6.9	0.8		Unknown	O
TP0248	O83276*	Uncharacterized lipoprotein TP_0248	15	6.8	1.5		Lipoprotein-unknown membrane localization	S

TP number	UniProt Accession Number[#]	Protein Name	Molecular Weight (kDa)	Average NSAF value	cDNA/ DNA signal ratio Smajs *et al.* [15]	GO Biological Process	Cellular location	COG category
TP0403	O83418	Flagellar biosynthesis protein FliJ	18	6.8	0.8	Bacterial-type flagellum-dependent cell motility; chemotaxis	Flagellum	none
TP0842	R9UU24	Methionine aminopeptidase	27	6.8	0.8	Protein initiator methionine removal	Unknown	J

*Legend: #: UniProt Proteome ID: UP000014259; NA: not available; NSAF normalized spectral abundance factor; no_hit: proteins with no UBLAST hit to any eggNOG sequence; none: over 70% of hits to a member that is not assigned an eggNOG code. *: the C-terminal 49 amino acids of TP0126a correspond to the N-terminal 49 amino acids of TP0126 due to a genome insertion* [29]

In terms of the cellular localizations of high abundant proteins, 18 were membrane localized. Four of these proteins were predicted lipoproteins (TP0248, TP0768, TP0895, TP0789) and two were predicted OM uncharacterized proteins (TP0858, TP0126). Surprisingly, approximately a third of the high-abundant proteins (N=37) were classified as uncharacterized/hypothetical and seven proteins did not have any significant (70 %) match with any other EggNOG sequences indicating these are highly specialized *T. pallidum* proteins that may play an important role in unique survival and virulence tactics. The most highly represented COG category of the highly expressed proteins was category 'J' (translation, ribosomal structure & biogenesis). A low correlation was found between previous gel-based studies [16,17] that determined protein abundance based on silver staining and protein abundance as determined in this study. For example, some highly abundant gel-detected proteins were not detected in our analysis, such as the uncharacterized protein TP0259 and the Tp34 lipoprotein (TP0971) [16,17]. We found a low correlation between the average transcriptional rate (cDNA/DNA signal) from a previous transcriptome study [15] and the average NSAF value for each detected protein found in this study (Pearson's correlation coefficient, r = 0.26; P = 0.000). The distribution of these data is depicted in *Fig 5B*. In general, flagellar proteins and proteins related to flagellar biosynthesis such as flagellar filament outer layer protein (TP0249), putative flagellar filament outer layer protein FlaA (TP0663), and flagellar biosynthetic protein FliP (TP0718) were highly expressed in both studies. There were some notable discordances between the data, such as the high gene expression level measured for lipoprotein antigen Tp47 (TP0574), galactose ABC superfamily ATP binding cassette transporter Tpp38 (TP0684) and the 60kDa chaperonin (TP0030), all of which were found in low abundance at the protein level in this study. Moreover, 27 proteins with high gene expression (cDNA/DNA signal ratios greater than 4.0) were not found in this analysis (*Table 1*). We theorized that the proteins we failed to detect in our analysis would have a lower mean transcription rate. There was however, no significant cDNA/DNA signal data difference between the detected and undetected proteins as determined by a Two-sample Wilcoxon rank-sum (Mann-Whitney) test (P = 0.5). Other studies have demonstrated low correlations between transcriptome and protein abundance data, as reviewed by Maier *et al.* [79]. Intermediary factors such as translation efficiency and protein half-life play a prominent role in accentuating the lack of a linear association between gene expression and protein abundance.

T. pallidum proteins confirmed or predicted to be related to virulence

Thirty-nine proteins implicated in *T. pallidum* virulence [20] were detected, including eight members of the *tpr* gene family and a protein related to a beta-barrel assembly machinery (BAM) complex. Brief descriptions of these proteins are detailed below.

Tpr proteins

Regarding the *tpr* gene family implicated in host-immune evasion [30], 8/12 of these proteins were detected in this analysis, including proteins TprB (TP0011), TprC/D (TP0117/TP0131), TprE (TP0313), TprG (TP0317), TprH (TP0610), TprI (TP0620) and TprJ (TP0621). Proteins TprA (TP0009), TprK (TP0897) and TprL (TP1031) were not detected. There was no unique TprF peptide sequence found in this analysis, although three peptides were uncovered that are homologous for TprC/D, F and I (*Table 5*). The ORF origin of these peptides cannot be determined. The *tprC* and *tprD* loci contain two identical coding sequences in the reference Nichols and DAL-1 strain genome [13,70], therefore we included the detection of both TprC and TprD since no distinction could be made between the coding ORF origin of these proteins. Even though *tprK* was previously shown to exhibit the highest level of transcription among tpr family genes [80], the fact that *tprK* displays high sequence variability [36] makes the likelihood of detecting this protein minimal due to rigid MS analytical criteria.

Table 5 *Peptide Sequences of proteins TprB, C/F/I, E/G/J and H, detected by MS analysis*

Protein	Peptide Sequence Identified
Tpr protein B	NELAAQMR
Tpr protein B	VKGKGTNSR
Tpr protein C	APMNALNIDALLRMQWK
Tpr protein C/F/I	GEARSGVWAQLQLK
Tpr protein C/F/I	SGVWAQLQLK
Tpr protein C/F/I	TALLWGVGGR
Tpr protein E	TNGTQVVNIDVTVPVNVRQSPVR
Tpr protein E	VEQAVQENIR
Tpr protein E/G/J	AGISASLIEK
Tpr protein E/G/J	DKLLWNVGGR
Tpr protein E/G/J	IPVQDYGWVKPSVTVHASTNRAHLNAPAAGGAVGATYLTK
Tpr protein E/G/J	TTNTVGVSFPLVMR
Tpr protein G	KKTDALDAGQQIR
Tpr protein G	KTDALDAGQQIR
Tpr protein G	TDALDAGQQIR
Tpr protein H	AGDAYTHLIDGLEAGMDVR
Tpr protein H	RVRSVGTWALLFMSSAAGLCAETR
Tpr protein H	LHTLASTPR
Tpr protein H	RTLLSPSAAVR
Tpr protein H	TKVTPGGPVAYAIAQR
Tpr protein I	FIQMALVK
Tpr protein I	VATDSGDR
Tpr protein J	MVGEALIKQQLSR
Tpr protein J	NNANMQAVGGSLGDTARMVGEALIK
Tpr protein J	NNNGNPLPSGGSSGHIGLPVVGK
Tpr protein J	QDLADLVPMMR

BAM-Complex

Outer membrane beta-barrel proteins (OMPs) are commonly involved in cellular process such as small molecule efflux (such as antibiotics) and nutrient acquisition [81,82] in bacteria. The beta-barrel assembly machinery (BAM) complex [83] is thought to facilitate OMP assembly, insertion and folding and in Gram-negative bacteria this complex is typically composed of five proteins: BamA, which is an integral membrane protein and four accessory lipoproteins, BamB-BamE [84]. The insertion and assembly of proteins into the outer membrane is controlled through interactions with periplasmic chaperones (SurA, Skp, and DegP) [85]. Studies [72,86]

have demonstrated the presence of a BAM complex in *T. pallidum* which is similar to that of *Escherichia coli* [72]. We detected the BamA orthologue (TP0326) [72,74,86,87]. Peptides identified corresponded to the POTRA 2 & 3 domains and a transmembrane domain/ extra-cellular Loop L3 [72,86] (*Table 6*).

Table 6 *Peptides related to BamA orthologue protein (TP0326) identified in MS analysis and corresponding (topological) domain locations*

Identified peptide sequence	Sequence location	(Topological) Domain Location*
MKVDQESLR	141-149	POTRA 2
VDQESLRR	143-150	POTRA 2
AFTESVLK	197-203	POTRA 3
KVLSTQEAR	205-213	POTRA 3
VEGVAKTVDK	246-255	POTRA 3
AGSYGNGLPHPYTSR	514-529	Trans-memebrane β-strand 5/ Extra-cellular Loop L3

**based on experimental predictions* [72,86]

Other detected proteins implicated in *T. pallidum* virulence

In our analyses we detected a selection of additional proteins that have been previously implicated in *T. pallidum* virulence and pathogenesis, as reported in *Table 7*.

Table 7 *Proteins identified in this study that have been previously implicated in* T. pallidum *virulence*

ORF Number(s)	Reported/predicted functional role	Supporting reference
TP0967, TP0968, TP0969	TolC-like proteins	[88,89]
TP0155	Fibronectin binding protein	[75]
TP1038	Bacterioferrin/ TpF1	[90]
TP0027, TP0649	Putative hemolysins	[91]

Exploring the undetected *T. pallidum* proteins

Of the predicted protein coding ORFs, 482/968 proteins were not detected in this study. Most of the undetected proteins are classified as 'uncharacterized proteins/hypothetical proteins' (N=197), 'conserved hypothetical integral membrane proteins (N=10), or 'conserved hypothetical protein' (N=1). The most plausible explanations for not detecting half of the proteome are i) very low protein abundance could evade MS detection, ii) lack of protein expression during *in vivo* expression during some or all stages of infection, iii) small proteins are less viable to detection since they contain fewer peptides and/or these protein sequences lack arginine or

lysine tryptic digestion sites or iv) the presence of (partial) sequence heterogeneity that would thwart peptide/database matching. Certain caveats of MS analyses will always preclude the detection of the whole proteome of organisms. A non-exhaustive list of other technical limitations include: i) hydrophobic peptides do not elute from LC columns during the applied gradient, ii) spectral masking of low abundant proteins by the presence of high abundant protein spectra, iii) co-elution and ion suppression that may prevent the ionization or detectability of some peptides by MS and iv) some peptides are unable to ionize sufficiently on the MS platform.

Variable *T. pallidum* genomic sequences as modulators of protein expression

To address the possibility that the presence of variable sequences may have affected proteome coverage, either by altered gene expression or by precluding MS detection, we searched for known and predicted heterogeneous sequences in the *T. pallidum* genome. Within this analysis we looked for sequences containing elements indicative of phase variation (homopolymeric tracts) or antigenic variation through gene conversion (tandem repeats, *tprK* donor sites and quadruplex forming G-rich sequences (G4FS)). Previous investigations have identified and characterized 19 genes with variable sequence elements, of which 9 proteins were detected in this analysis. Aside from the 12 aforementioned Tpr family proteins there are seven additional genes shown to contain variable sequence elements including: *tprK* donor sequences to promote gene conversion (*tp0130; tp0129; tp0128*), homopolymeric G-tracts (poly-G tracts) in promoter regions to alter transcription (TP0126), poly-G tracts in the ORF to induce phase variation (TP0127), or G4FS *cis-acting* DNA elements that form guanine quadruplexes to induce recombination and gene conversion (TP0104; TP0136) [32,36,67,70,92]. Notably, TP0136, a fibronectin binding protein implicated in *T. pallidum* virulence not detected in this analysis, harbors two G4FS sequences localized within tandem repeats in the ORF [93]. Surprisingly, the paralogues of TP0136: TP0133, TP0134, TP0462 and TP0463 were also not detected. Among these seven additional variable sequences only an OmpW homologue (TP0126) was detected in our analyses.

We also searched for predicted variable sequences in the *T. pallidum* genome. A previous *T. pallidum* genomic study predicted the presence of G4FS which may be involved in generation of *tprK* variants in pathogenic treponemes [32]. Similar G4 DNA structures have been implicated in the host immune evasion tactics of *Neisseria gonorrhoeae* where they function as recombination activation elements to regulate gene conversion and the expression of cell surface pilin proteins (PilE) [94]. Giacani *et al.* (2012) identified 46 putative G4FS sequences located in 33 different genes and eight

unique intergenic regions (IGRs) of *T. pallidum*; 21 of the 33 predicted G4FS-containing ORFs were detected in this analysis. Among the eight putative G4FS residing within unique IGRs, only two of the downstream proteins were detected in this study (TP0104; TP0549). Additionally, we searched for the presence of tandem repeats [95,96] in ORFs and IGRs of genes for which peptides had previously been detected using MS [16] or exhibited high transcript abundance [15]. The presence of highly mutable sites, or contingency loci, such as tandem repeats have been suggested to represent a mechanism for rapid environmental adaptation and virulence within a host [97]. This has been explored in a recent study involving serial *in vivo* passage of *Campylobacter jejuni* in mice that resulted in increased phases in the contingency loci and virulence [98]. This analysis identified three additional genes harboring tandem repeats (*tp0470; tp0424; tp0769*), providing a possible rationale for why these proteins remained undetected in this study. Our study detected 30 proteins out of a total of 60 proteins with known and predicted variable sequences. Remarkably, four proteins discovered in this analysis were only annotated in the original *T. pallidum* genome [13], mostly due to the fact that sequences below 150 base pairs were not annotated as protein coding in the resequenced genome [14]. Perhaps there is a need for deeper mining of the *T. pallidum* genome and re-evaluation of the definition of protein coding sequences, especially in light of the recent attention brought to classes of endogenous polypeptides called 'SEPs' (sORF-encoded polypeptides). These polypeptides are encoded by short open reading frames (small ORFs or smORFs) (generally <150 amino acids in length) in bacteria and eukaryotic organisms and are thought to play an important function in biological functions [99] such as cell survival under conditions of glucose toxicity as studied in *E. coli* [100]. Interestingly, in *M. pneumonia*, 53% of all smORFs are deemed essential to cell survival whilst another 11% affect the fitness of the organism [101], indicating that these may also play a large (unknown) role in *T. pallidum* biological function.

In general, proteins in small genomes are more likely to function as proficient "multitaskers" and have been shown to interact with other proteins from a wider range of functions in comparison to their orthologues in larger genomes [102]. It is possible that many *T. pallidum* proteins perform multiple biological functions, especially under different environmental conditions. A growing area in proteomics is the concept of 'protein moonlighting', defined as a single protein that displays multiple functions that are not related to multiple RNA splice variants, multiple proteolytic fragments or gene fusions [103]. Many bacterial species employ protein moonlighting and the role of this phenomenon in bacteria virulence has been excellently reviewed by Henderson *et al.* [104,105]. Some bioinformatic approaches

have been suggested to approach genome wide annotation of potential moonlighting proteins [106,107]; these may be useful for future *T. pallidum* proteome studies.

One of the many intriguing aspects of *T. pallidum* is the fact so many proteins lack homology with proteins from other bacteria. This is exemplified by the fact that only 59 % (581/968) of the *T. pallidum* protein coding genes were allocated into COG or NOG categories. With the demonstrated expression of 114 uncharacterized/ hypothetical *T. pallidum* proteins in this study, some even at high abundance, further experimental analysis is needed to elucidate the functions of these proteins such as looking at protein binding partners. Periodic re-evaluations of 'uncharacterized' *T. pallidum* proteins are warranted, especially with the rapid sophistication of bioinformatics tools and the growing repertoire of proteins with known predicted functions from other organisms.

We are confident in the quality and extent of the protein coverage of this analysis. For example, we performed analysis on three biological replicates, employed multidimensional peptide separation techniques together with complementary MS analyses in order to improve the dynamic range and coverage of the analyses. Nevertheless, there are a number of limitations related to this study.

Limitations

Even though our experimental approaches aimed to closely mimic the physiological conditions of human infection, a distinct advantage over the artificial conditions of *in vitro* studies, we cannot exclude the effects of inter-rabbit variability. Different rabbits may exert unique immune pressures, which in turn may influence gene expression. The fact that infected rabbits typically do not transition into the secondary stage of syphilis [108] and there is no tertiary stage in rabbits [109] may suggest that the infectious dynamics of rabbit syphilis may differ from that of humans. Moreover, there may be differential gene expression depending upon the tissue environment [15], therefore the analysis of intradermal rather than intratesticular infections of rabbits, or sampling of human syphilitic lesions (pending ethical consent) could provide interesting insights into differing protein expression profiles. Lastly, technical handling after testicular extraction and treponemal purification may 'stress' the bacteria into a non-characteristic infectious expression state and some proteins may degrade quickly after extraction since individual protein half-life ranges can vary from several seconds to tens of hours [110]. Gentle and prompt sampling and handling of treponemal extract samples may help to alleviate these potential interferences.

Despite purification efforts through Percoll density gradient centrifugation, the high abundant rabbit albumin may have masked the spectra of some lower abundant *T. pallidum* proteins. Additional purification or pre-fractionalization steps could be added to facilitate the detection of low abundant proteins, however, there is a risk of inadvertently depleting treponemal proteins through methods such as albumin depletion. Possible experimental method improvements include altering the LC-MS/MS settings to be focused on either small or large proteins and/or using alternative protease and/or multi protease protein digestion [111]. Regarding the use of spectral counting, this method remains a semi-quantitative estimation of protein abundance since proteins are not measured compared to a reference. More absolute and precise protein quantification methods could be used in the future such as isobaric tags for relative and absolute quantification (iTRAQ) or selected reaction monitoring (SRM) as reviewed by Maaß and Becher [112].

Conclusions

This study makes a number of contributions to the characterization of the *T. pallidum* proteome: i) we detected 557 *T. pallidum* proteins expressed during *in vivo* experimental rabbit infection using complementary mass spectrometry detection techniques; this is the first account of 499 proteins at the protein level using these methods, ii) protein abundance semi-quantified by spectral counting showed a low correlation with previous gene expression transcriptome data, iii) 116 predicted membrane localized proteins were detected, of which 16 have evidence supporting outer membrane localization and iv) a number of virulence factors were detected, including 8/12 Tpr proteins.

Acknowledgements

We are grateful to David L. Cox (Centers for Disease Control and Prevention, Atlanta, GA, USA) and Dr. Zákoucká (National Institute of Public Health/National Reference Laboratory for the Diagnostics of Syphilis, Prague, Czech Republic) for the DAL-1 strain bacteria samples.

Supporting Information (available online)

S1 Table Mass spectrometry data reports for biological (N=3) and technical runs (N=2)

S2 Table Extensive descriptions of all unique *T. pallidum* proteins identified in this study by mass spectrometry analysis

S3 Table Peptide identifications per biological (N=3) and MS run (N=2) and corresponding calculated spectral counting NSAF values

S4 Table *T. pallidum* membrane protein bioinformatic pipeline prediction analyses

References

1. Newman L, Rowley J, Vander Hoorn S, Wijesooriya NS, Unemo M, Low N, et al. Global Estimates of the Prevalence and Incidence of Four Curable Sexually Transmitted Infections in 2012 Based on Systematic Review and Global Reporting. PLoS One. 2015;10(12):e0143304.
2. Read P, Fairley CK, Chow EPF. Increasing trends of syphilis among men who have sex with men in high income countries. Sex Heal. 2015;12(2):155–63.
3. Newman L, Kamb M, Hawkes S, Gomez G, Say L, Seuc A, et al. Global estimates of syphilis in pregnancy and associated adverse outcomes: analysis of multinational antenatal surveillance data. [Internet]. Vol. 10, PLoS medicine. Public Library of Science; 2013. p. e1001396.
4. Cox DL, Riley B, Chang P, Sayahtaheri S, Tassell S, Hevelone J. Effects of molecular oxygen, oxidation-reduction potential, and antioxidants upon in vitro replication of Treponema pallidum subsp. pallidum. Appl Environ Microbiol. 1990;56(10):3063–72.
5. Cox CD, Barber MK. Oxygen uptake by Treponema pallidum. Infect Immun. 1974;10(1):123–7.
6. Cover WH, Norris SJ, Miller JN. The microaerophilic nature of Treponema pallidum: enhanced survival and incorporation of tritiated adenine under microaerobic conditions in the presence or absence of reducing compounds. Sex Transm Dis. 9(1):1–8.
7. Harman M, Vig DKK, Radolf JDD, Wolgemuth CWW. Viscous dynamics of lyme disease and syphilis spirochetes reveal flagellar torque and drag. Biophys J. 2013;105(10):2273–80.
8. Izard J, Renken C, Hsieh C-E, Desrosiers DC, Dunham-Ems S, La Vake C, et al. Cryo-electron tomography elucidates the molecular architecture of Treponema pallidum, the syphilis spirochete. J Bacteriol. 2009;191(24):7566–80.
9. Liu J, Howell JK, Bradley SD, Zheng Y, Zhou ZH, Norris SJ. Cellular architecture of Treponema pallidum: novel flagellum, periplasmic cone, and cell envelope as revealed by cryo electron tomography. J Mol Biol. 2010;403(4):546–61.
10. Radolf JD, Norgard M V, Schulz WW. Outer membrane ultrastructure explains the limited antigenicity of virulent Treponema pallidum. Proc Natl Acad Sci U S A. 1989;86(6):2051–5.
11. Walker EM, Zampighi GA, Blanco DR, Miller JN, Lovett MA. Demonstration of rare protein in the outer membrane of Treponema pallidum subsp. pallidum by freeze-fracture analysis. J Bacteriol. 1989;171(9):5005–11.
12. Norris SJ, Edmondson DG. Factors affecting the multiplication and subculture of Treponema pallidum subsp. pallidum in a tissue culture system [Internet]. Vol. 53, Infection and Immunity. 1986. p. 534–9.
13. Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, Dodson R, et al. Complete Genome Sequence of Treponema pallidum, the Syphilis Spirochete. Science (80-). 1998;281(5375):375–88.
14. Pětrošová H, Pospíšilová P, Strouhal M, Čejková D, Zobaníková M, Mikalová L, et al. Resequencing of Treponema pallidum ssp. pallidum Strains Nichols and SS14: Correction of Sequencing Errors Resulted in Increased Separation of Syphilis Treponeme Subclusters. Dong Q, editor. PLoS One. 2013;8(9):e74319.
15. Smajs D, Mckevitt M, Howell JK, Norris J, Cai W, Palzkill T, et al. Transcriptome of Treponema pallidum : Gene Expression Profile during Experimental Rabbit Infection Transcriptome of Treponema pallidum : Gene Expression Profile during Experimental Rabbit Infection †. 2005;
16. McGill MA, Edmondson DG, Carroll JA, Cook RG, Orkiszewski RS, Norris SJ. Characterization and serologic analysis of the Treponema pallidum proteome. Infect Immun. 2010;78(6):2631–43.
17. Norris SJ.. Polypeptides of Treponema pallidum : Progress toward Understanding Their Structural , Functional , and Immunologic Rolest. Microbiol Rev. 1993;57(3):750–79.
18. Catrein I, Herrmann R. The proteome of Mycoplasma pneumoniae, a supposedly "simple" cell. Proteomics. 2011;11(18):3614–32.
19. Lafond RE, Lukehart SA. Biological Basis for Syphilis. Clin Microbiol Rev. 2006;19(1):29–49.
20. Weinstock GM, Hardham JM, McLeod MP, Sodergren EJ, Norris SJ. The genome of Treponema pallidum: new light on the agent of syphilis. FEMS Microbiol Rev. 1998;22(4):323–32.
21. Schouls L. Recombinant DNA technology in syphilis research. In: Wright DJM, Archard L, editors. Molecular biology of sexually transmitted diseases. London: Chapman & Hall; 1992.
22. Smajs D, Mckevitt M, Wang L, Howell JK, Norris SJ, Palzkill T, et al. BAC library of T. pallidum

DNA in E. coli. Genome Res. 2002;12(3):515–22.
23. McKevitt M, Patel K, Smajs D, Marsh M, McLoughlin M, Norris SJ, et al. Systematic cloning of Treponema pallidum open reading frames for protein expression and antigen discovery. Genome Res. 2003;13(7):1665–74.
24. McKevitt M, Brinkman MB, McLoughlin M, Perez C, Howell JK, Weinstock GM, et al. Genome Scale Identification of Treponema pallidum Antigens. Infect Immun. 2005;73(7):4445–50.
25. Brinkman MB, Mckevitt M, Perez C, Howell J, George M, Norris SJ, et al. Reactivity of Antibodies from Syphilis Patients to a Protein Array Representing the Treponema pallidum Proteome Reactivity of Antibodies from Syphilis Patients to a Protein Array Representing the Treponema pallidum Proteome †. 2006;
26. Setubal JC, Reis M, Matsunaga J, Haake DA. Lipoprotein computational prediction in spirochaetal genomes. Microbiology. 2006;152(Pt 1):113–21.
27. Cox DL, Luthra A, Dunham-Ems S, Desrosiers DC, Salazar JC, Caimano MJ, et al. Surface immunolabeling and consensus computational framework to identify candidate rare outer membrane proteins of Treponema pallidum. Infect Immun. 2010;78(12):5178–94.
28. Naqvi AAT, Shahbaaz M, Ahmad F, Hassan MI. Identification of functional candidates amongst hypothetical proteins of Treponema pallidum ssp. pallidum. PLoS One. 2015;10(4):e0124177.
29. Mikalová L, Strouhal M, Čejková D, Zobaníková M, Pospíšilová P, Norris SJ, et al. Genome analysis of Treponema pallidum subsp. pallidum and subsp. pertenue strains: most of the genetic differences are localized in six regions. PLoS One. 2010;5(12):e15713.
30. Palmer G, Bankhead T, Lukehart S. "Nothing is permanent but change"†–antigenic variation in persistent bacterial pathogens. Cell Microbiol. 2009;11(12):1697–705.
31. LaFond RE, Centurion-Lara A, Godornes C, Van Voorhis WC, Lukehart SA. TprK sequence diversity accumulates during infection of rabbits with Treponema pallidum subsp. pallidum Nichols strain. Infect Immun. 2006;74(3):1896–906.
32. Giacani L, Brandt SL, Puray-Chavez M, Reid TB, Godornes C, Molini BJ, et al. Comparative investigation of the genomic regions involved in antigenic variation of the TprK antigen among treponemal species, subspecies, and strains. J Bacteriol. 2012;194(16):4208–25.
33. Leader BT, Hevner K, Molini BJ, Barrett LK, Van Voorhis WC, Lukehart SA. Antibody Responses Elicited against the Treponema pallidum Repeat Proteins Differ during Infection with Different Isolates of Treponema pallidum subsp. pallidum. Infect Immun. 2003;71(10):6054–7.
34. Morgan C a, Molini BJ, Lukehart S a, Van Voorhis WC. Segregation of B and T cell epitopes of Treponema pallidum repeat protein K to variable and conserved regions during experimental syphilis infection. J Immunol. 2002;169(2):952–7.
35. Reid TB, Molini BJ, Fernandez MC, Lukehart SA. Antigenic variation of TprK facilitates development of secondary syphilis. Infect Immun. 2014;82(12):4959–67.
36. Giacani L, Molini BJ, Kim EY, Godornes BC, Leader BT, Tantalo LC, et al. Antigenic variation in Treponema pallidum: TprK sequence diversity accumulates in response to immune pressure during experimental syphilis. J Immunol. 2010;184(7):3822–9.
37. Anand A, Luthra A, Dunham-Ems S, Caimano MJ, Karanian C, LeDoyt M, et al. TprC/D (Tp0117/131), a trimeric, pore-forming rare outer membrane protein of Treponema pallidum, has a bipartite domain structure. J Bacteriol. 2012;194(9):2321–33.
38. Anand A, LeDoyt M, Karanian C, Luthra A, Koszelak-Rosenblum M, Malkowski MG, et al. Bipartite Topology of Treponema pallidum Repeat Proteins C/D and I: Outer membrane insertion, trimerization and porin function require a C-terminal β-barrel domain. J Biol Chem. 2015;290(19):12313–31.
39. Weigel LM, Radolf JD, Norgard M V. The 47-kDa major lipoprotein immunogen of Treponema pallidum is a penicillin-binding protein with carboxypeptidase activity. Proc Natl Acad Sci U S A. 1994;91(24):11611–5.
40. Cha JY, Ishiwata A, Mobashery S. A novel beta-lactamase activity from a penicillin-binding protein of Treponema pallidum and why syphilis is still treatable with penicillin. J Biol Chem. 2004;279(15):14917–21.
41. Deka RK, Machius M, Norgard M V, Tomchick DR. Crystal structure of the 47-kDa lipoprotein of Treponema pallidum reveals a novel penicillin-binding protein. J Biol Chem. 2002;277(44):41857–64.

42. Radolf JD, Arndt LL, Akins DR, Curetty LL, Levi ME, Shen Y, et al. Treponema pallidum and Borrelia burgdorferi lipoproteins and synthetic lipopeptides activate monocytes/macrophages. J Immunol. 1995;154(6):2866–77.
43. Kelesidis T. The cross-talk between spirochetal lipoproteins and immunity. Front Immunol. 2014;5(JUN):85–8.
44. Semanjski M, Macek B. Shotgun proteomics of bacterial pathogens: advances, challenges and clinical implications. Expert Rev Proteomics. 2015;
45. Becher D, Hempel K, Sievers S, Zühlke D, Pané-Farré J, Otto A, et al. A proteomic view of an important human pathogen--towards the quantification of the entire Staphylococcus aureus proteome. PLoS One. 2009;4(12):e8176.
46. Soufi B, Krug K, Harst A, Macek B. Characterization of the E. coli proteome and its modifications during growth and ethanol stress. Front Microbiol. 2015;6:103.
47. Kruh NA, Troudt J, Izzo A, Prenni J, Dobos KM. Portrait of a pathogen: the Mycobacterium tuberculosis proteome in vivo. PLoS One. 2010;5(11):e13938.
48. Lukehart S a, Marra CM. Isolation and laboratory maintenance of Treponema pallidum. Curr Protoc Microbiol. 2007;Chapter 12(November):Unit 12A.1.
49. Hanff PA, Norris SJ, Lovett MA, Miller JN. Purification of Treponema pallidum, Nichols strain, by Percoll density gradient centrifugation. Sex Transm Dis. 11(4):275–86.
50. Zobaníková M, Mikolka P, Cejková D, Pospíšilová P, Chen L, Strouhal M, et al. Complete genome sequence of Treponema pallidum strain DAL-1. Stand Genomic Sci. 2012;7(1):12–21.
51. Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. J Proteome Res. 2006;5(9):2339–47.
52. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods. 2007;4(3):207–14.
53. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, et al. The PeptideAtlas project. Nucleic Acids Res. 2006;34(90001):D655–8.
54. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. [Internet]. Vol. 8, Nature methods. Nature Publishing Group; 2011. p. 785–6.
55. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci. 2003;12(8):1652–62.
56. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010;26(13):1608–15.
57. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567–80.
58. Bagos PG, Liakopoulos TD, Spyropoulos IC, Hamodrakas SJ. PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. Nucleic Acids Res. 2004;32(Web Server issue):W400-4.
59. Yu C-S, Chen Y-C, Lu C-H, Hwang J-K. Prediction of protein subcellular localization. Proteins. 2006;64(3):643–51.
60. Reynolds SM, Käll L, Riffle ME, Bilmes JA, Noble WS. Transmembrane topology and signal peptide prediction using dynamic bayesian networks. PLoS Comput Biol. 2008;4(11):e1000213.
61. Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res. 2007;35(Web Server issue):W429-32.
62. Viklund H, Bernsel A, Skwark M, Elofsson A. SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. Bioinformatics. 2008;24(24):2928–9.
63. Tusnády GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics. 2001;17(9):849–50.
64. Berven FS, Flikka K, Jensen HB, Eidhammer I. BOMP: a program to predict integral beta-barrel outer membrane proteins encoded within genomes of Gram-negative bacteria. Nucleic Acids Res. 2004;32(Web Server issue):W394-9.
65. Ou Y-Y, Chen S-A, Wu S-C. ETMB-RBF: discrimination of metal-binding sites in electron

transporters based on RBF networks with PSSM profiles and significant amino acid pairs. PLoS One. 2013;8(2):e46572.

66. Čejková D, Strouhal M, Norris SJ, Weinstock GM, Šmajs D. A Retrospective Study on Genetic Heterogeneity within Treponema Strains: Subpopulations Are Genetically Distinct in a Limited Number of Positions. PLoS Negl Trop Dis. 2015;9(10):e0004110.
67. Smajs D, Norris SJ, Weinstock GM. Genetic diversity in Treponema pallidum: implications for pathogenesis, evolution and molecular diagnostics of syphilis and yaws. Infect Genet Evol. 2012;12(2):191–202.
68. Stamm L V, Bergen HL. The sequence-variable, single-copy tprK gene of Treponema pallidum Nichols strain UNC and Street strain 14 encodes heterogeneous TprK proteins. Infect Immun. 2000;68(11):6482–6.
69. Centurion-Lara A, Godornes C, Castro C, Van Voorhis WC, Lukehart SA. The tprK gene is heterogeneous among Treponema pallidum strains and has multiple alleles. Infect Immun. 2000;68(2):824–31.
70. Centurion-Lara A, Giacani L, Godornes C, Molini BJ, Brinck Reid T, Lukehart SA. Fine analysis of genetic diversity of the tpr gene family among treponemal species, subspecies and strains. PLoS Negl Trop Dis. 2013;7(5):e2222.
71. Giacani L, Brandt SL, Ke W, Reid TB, Molini BJ, Iverson-Cabral S, et al. Transcription of TP0126, Treponema pallidum putative OmpW homolog, is regulated by the length of a homopolymeric guanosine repeat. Infect Immun. 2015;83(6):2275–89.
72. Desrosiers DC, Anand A, Luthra A, Dunham-Ems SM, LeDoyt M, Cummings M a D, et al. TP0326, a Treponema pallidum β-barrel assembly machinery A (BamA) orthologue and rare outer membrane protein. Mol Microbiol. 2011;80(6):1496–515.
73. Luthra A, Anand A, Radolf JD. Treponema pallidum in Gel Microdroplets: A Method for Topological Analysis of BamA (TP0326) and Localization of Rare Outer Membrane Proteins. Methods Mol Biol. 2015;1329:67–75.
74. Cameron CE, Lukehart S a, Castro C, Molini B, Godornes C, Van Voorhis WC. Opsonic potential, protective capacity, and sequence conservation of the Treponema pallidum subspecies pallidum Tp92. J Infect Dis. 2000;181(4):1401–13.
75. Cameron CE, Brown EL, Kuroiwa JMY, Schnapp LM, Brouwer NL. Treponema pallidum fibronectin-binding proteins. J Bacteriol. 2004;186(20):7019–22.
76. Giacani L, Sambri V, Marangoni A, Cavrini F, Storni E, Donati M, et al. Immunological evaluation and cellular location analysis of the TprI antigen of Treponema pallidum subsp. pallidum. Infect Immun. 2005;73(6):3817–22.
77. Salazar JC, Pope CD, Moore MW, Pope J, Kiely TG, Radolf JD. Lipoprotein-dependent and -independent immune responses to spirochetal infection. Clin Diagn Lab Immunol. 2005;12(8):949–58.
78. Zybailov BL, Florens L, Washburn MP. Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. Mol Biosyst. 2007;3(5):354–60.
79. Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. FEBS Lett. 2009;583(24):3966–73.
80. Giacani L, Molini B, Godornes C, Barrett L, Van Voorhis W, Centurion-Lara A, et al. Quantitative analysis of tpr gene expression in Treponema pallidum isolates: Differences among isolates and correlation with T-cell responsiveness in experimental syphilis. Infect Immun. 2007;75(1):104–12.
81. Kim KH, Aulakh S, Paetzel M. The bacterial outer membrane β-barrel assembly machinery. Protein Sci. 2012;21(6):751–68.
82. Pagès J-M, James CE, Winterhalter M. The porin and the permeating antibiotic: a selective diffusion barrier in Gram-negative bacteria. Nat Rev Microbiol. 2008;6(12):893–903.
83. Knowles TJ, Scott-Tucker A, Overduin M, Henderson IR. Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. Nat Rev Microbiol. 2009;7(3):206–14.
84. Selkrig J, Leyton DL, Webb CT, Lithgow T. Assembly of β-barrel proteins into bacterial outer membranes. Biochim Biophys Acta. 2014;1843(8):1542–50.
85. Lyu ZX, Zhao XS. Periplasmic quality control in biogenesis of outer membrane proteins. Biochem Soc Trans. 2015;43(2):133–8.
86. Luthra A, Anand A, Hawley KL, LeDoyt M, La Vake CJ, Caimano MJ, et al. A Homology Model

Reveals Novel Structural Features and an Immunodominant Surface Loop/Opsonic Target in the Treponema pallidum BamA Ortholog TP_0326. J Bacteriol. 2015;197(11):1906–20.
87. Tomson FL, Conley PG, Norgard M V, Hagman KE. Assessment of cell-surface exposure and vaccinogenic potentials of Treponema pallidum candidate outer membrane proteins. Microbes Infect. 2007;9(11):1267–75.
88. Bunikis I, Denker K, Ostberg Y, Andersen C, Benz R, Bergström S. An RND-type efflux system in Borrelia burgdorferi is involved in virulence and resistance to antimicrobial compounds. PLoS Pathog. 2008;4(2):e1000009.
89. Eshghi A, Cullen P a, Cowen L, Zuerner RL, Cameron CE. Global Proteome Analysis of Leptospira interrogans research articles. 2009;4564–78.
90. Babolin C, Amedei A, Ozolins D, Zilevica A, D'Elios MM, de Bernard M. TpF1 from Treponema pallidum activates inflammasome and promotes the development of regulatory T cells. J Immunol. 2011;187(3):1377–84.
91. Weinstock GM, Smajs D, Hardham J, Norris SJ. From microbial genome sequence to applications. Res Microbiol. 2000;151(2):151–8.
92. Giacani L, Lukehart S, Centurion-Lara A. Length of guanosine homopolymeric repeats modulates promoter activity of subfamily II tpr genes of Treponema pallidum ssp. pallidum. FEMS Immunol Med Microbiol. 2007;51(2):289–301.
93. Brinkman MB, McGill MA, Pettersson J, Rogers A, Matejková P, Smajs D, et al. A novel Treponema pallidum antigen, TP0136, is an outer membrane protein that binds human fibronectin. Infect Immun. 2008;76(5):1848–57.
94. Cahoon LA, Seifert HS. An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. Science. 2009;325(5941):764–7.
95. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
96. Kolpakov R, Bana G, Kucherov G. mreps: Efficient and flexible detection of tandem repeats in DNA. Nucleic Acids Res. 2003;31(13):3672–8.
97. Moxon R, Bayliss C, Hood D. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. Annu Rev Genet. 2006;40:307–33.
98. Jerome JP, Bell JA, Plovanich-Jones AE, Barrick JE, Brown CT, Mansfield LS. Standing genetic variation in contingency loci drives the rapid adaptation of Campylobacter jejuni to a novel host. PLoS One. 2011;6(1):e16399.
99. Chu Q, Ma J, Saghatelian A. Identification and characterization of sORF-encoded polypeptides. Crit Rev Biochem Mol Biol. 2015;50(2):134–41.
100. Wadler CS, Vanderpool CK. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. Proc Natl Acad Sci. 2007;104(51):20454–9.
101. Lluch-Senar M, Delgado J, Chen W-H, Lloréns-Rico V, O'Reilly FJ, Wodke JA, et al. Defining a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium. Mol Syst Biol. 2015;11(1):780.
102. Kelkar YD, Ochman H. Genome reduction promotes increase in protein functional complexity in bacteria. Genetics. 2013;193(1):303–7.
103. Jeffery CJ. Moonlighting proteins: old proteins learning new tricks. Trends Genet. 2003;19(8):415–7.
104. Henderson B, Martin A. Bacterial moonlighting proteins and bacterial virulence. [Internet]. Vol. 358, Current topics in microbiology and immunology. Springer Berlin Heidelberg; 2013. p. 155–213.
105. Henderson B. An overview of protein moonlighting in bacterial infection. Biochem Soc Trans. 2014;42(6):1720–7.
106. Khan I, Chen Y, Dong T, Hong X, Takeuchi R, Mori H, et al. Genome-scale identification and characterization of moonlighting proteins. Biol Direct. 2014;9(1):30.
107. Khan IK, Kihara D. Computational characterization of moonlighting proteins. Biochem Soc Trans. 2014;42(6):1780–5.
108. Strugnell RA, Drummond L, Faine S. Secondary lesions in rabbits experimentally infected with Treponema pallidum. Genitourin Med. 1986;62(1):4–8.
109. Carlson JA, Dabiri G, Cribier B, Sell S. The immunopathobiology of syphilis: the manifestations

and course of syphilis are determined by the level of delayed-type hypersensitivity. Am J Dermatopathol. 2011;33(5):433–60.

110. Doherty MK, Hammond DE, Clague MJ, Gaskell SJ, Beynon RJ. Turnover of the human proteome: determination of protein intracellular stability by dynamic SILAC. J Proteome Res. 2009;8(1):104–12.

111. Tsiatsiani L, Heck AJR. Proteomics beyond trypsin. FEBS J. 2015;282(14):2612–26.

112. Maaß S, Becher D. Methods and applications of absolute protein quantification in microbial systems. J Proteomics. 2016;136:222–33.

Chapter 5

Needle lost in the haystack: Multiple Reaction Monitoring fails to detect *Treponema pallidum* candidate protein biomarkers in plasma and urine samples from individuals with syphilis

Authors: Kara K. Osbak[1*]; Geert A. Van Raemdonck[1,2*]; Xaveer Van Ostade[2&]; Chris R. Kenyon[1,3&]

**/&- These authors contributed equally to this work*

Affiliations

1. HIV/STI Unit, Institute of Tropical Medicine, Antwerp, Belgium
2. Laboratory for Protein Science, Proteomics and Epigenetic Signalling (PPES) and Centre for Proteomics (CFP), University of Antwerp, Wilrijk, Belgium
3. Division of Infectious Diseases and HIV Medicine, University of Cape Town, Cape Town, South Africa

I contributed to the experimental work related to estimating the MS limit of detection for *T. pallidum* protein detection, sample collection (cohort study), data interpretation and manuscript preparation.

Conflicts of Interest and Sources of Funding

This work was supported by a grant from the Flanders Research Foundation, SOFI-B Grant (#757003) to CRK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

A test that could directly detect *Treponema pallidum* antigens in biofluid samples from individuals with syphilis would be of considerable clinical utility. In this study, 11 candidate *T. pallidum* biomarker proteins were chosen according to their physiochemical characteristics, *T. pallidum* specificity and predicted abundance. Thirty isotopically labelled proteotypic surrogate peptides (hPTPs) were synthesized and incorporated into a scheduled multiple reaction monitoring assay. Protein extracts from undepleted/unenriched plasma (N= 18) and urine (N= 4) samples from 18 individuals with syphilis in various clinical stages were tryptically digested, spiked with the hPTP mixture and analysed with a triple quadrupole mass spectrometer. No endogenous PTPs corresponding to the eleven candidate biomarkers were detected in any samples analysed. To estimate the Limit of Detection (LOD) of a comparably sensitive mass spectrometer (LTQ-Orbitrap), two dilution series of rabbit cultured purified *T. pallidum* were prepared in PBS. Polyclonal anti-*T. pallidum* antibodies coupled to magnetic Dynabeads were used to enrich one sample series; no LOD improvement was found compared to the unenriched series. The estimated LOD of MS instruments is 300 *T. pallidum*/ml in PBS. Biomarker protein detection likely failed due to the low (femtomoles/liter) predicted concentration of *T. pallidum* proteins. Alternative sample preparation strategies may improve the detectability of *T. pallidum* proteins in biofluids.

Introduction

Treponema pallidum ssp. *pallidum (T. pallidum)*, a non-culturable microaerophilic spirochete, is responsible for more than 8 million new cases of syphilis per year [1]. There has been a resurgence of syphilis in a number of world regions over the last two decades [1–3]. In Europe [2] and North America [3], this increase has been most marked in men who have sex with men (MSM). A striking feature of these outbreaks has been the increasing proportion of cases that are occurring in patients with a previous diagnosis of syphilis [4,5]. Patients with reinfections are more likely to present with asymptomatic or less symptomatic disease [4], hence the diagnosis of reinfection is wholly dependent on subtle changes in serological tests [6]. Two types of serological tests are used to diagnose syphilis: treponemal tests detect antibodies to *T. pallidum* and non-treponemal tests, such as the rapid plasma reagin (RPR) test, detect agglutination secondary to the presence of anti-lipoidal antibodies reactive to material released from damaged host cells and possibly cardiolipin released from *T. pallidum* [7]. Treponemal tests remain positive for life and are therefore of no use in the diagnosis of reinfection. Non-treponemal tests are used for syphilis post-treatment follow-up and diagnosis of reinfection. A wide range of factors can result in increases in test titres, causing syphilis to be over-diagnosed and unnecessarily treated [6,8–10]. Direct *T. pallidum* detection techniques, including various nucleic acid amplification tests (NAATs), have been developed, but apart from testing of primary ulcer specimens the sensitivity of these techniques is low [11]. Even in the setting of secondary syphilis when there is a high *T. pallidum* load in the blood [12] the sensitivity of PCR reaches only 52 % on serum specimens [11,13].

The *T. pallidum* genome, through evolutionary reduction, is one of the smallest of the human bacterial pathogens with a predicted 1044 ORFs [14]. Approximately half of the predicted proteins have been detected through MS techniques [15,16], including the semi-quantification of *T. pallidum* proteins using spectral counting [16]. A *T. pallidum* transcriptome study demonstrated that almost all genes were expressed during peak rabbit experimental infection [17]. This maximum utilization of the genome, well characterized proteome, and swift invasion of the organism into the bloodstream (within 24 hours after infection [18]) make this pathogen an ideal candidate for antigen diagnostic assay development. A variety of antigen tests against other pathogens have been designed for clinical samples such as blood, cerebrospinal fluid, faeces and urine and these have proven their utility in the diagnosis and assessment of therapeutic response in a number of infections, including *Helicobacter pylori* [19], *Cryptococcus neoformans* [20], *Cryptosporidium* ssp. [21], *Entamoeba histolytica* [22], Ebola virus [23] and

Mycobacteria tuberculosis [24]. If a highly sensitive and specific test could be developed that is able to confirm the presence or absence of *T. pallidum* in the body then this would be of considerable utility in the diagnosis of syphilis reinfections and in assessing therapeutic response. It could also be useful for the diagnosis of neuro- and congenital syphilis – two diagnoses where contemporary tests are suboptimal [25].

During the last decade, advanced MS-based proteomics platforms have emerged as mainstay bioanalytical tools for a broad range of clinical applications, including targeted protein identification [26] and bacteria identification and typing [27]. Particularly the AQUA workflow [28,29], with its use of stable isotopically labelled standard proteotypic peptides (henceforth referred to as 'heavy' PTPs or hPTPs) and selected/multiple reaction monitoring-mass spectrometry (SRM/MRM MS), has emerged as a powerful technique for the fast determination of multiple protein concentrations in highly complex sample matrixes such as urine (reviewed by Mermelekas *et al*. [30]) and plasma (reviewed by Pernemalm and Lehtiö [31]). Precise quantitation of proteins is possible by using hPTPs as internal standards that correspond to endogenous peptides created during the enzymatic digestion of the sample of interest. When combined, the endogenous and synthetic peptides elute together chromatographically and ionize with the same efficiency. Since the quantity of the labelled peptide is known, the absolute quantity of the targeted native protein can be determined by comparing MRM hPTP/endogenous peak areas. The precision and utility of this highly sensitive multiplexed method has been demonstrated on undepleted/ unenriched plasma for the detection of a panel of human cardiovascular disease [32] and cancer [33] biomarkers with a detection capability of four orders of magnitude (10^3–10^4 range in protein concentration) and up to femtomolar level sensitivity in plasma [34]. Recently, a panel of 136 cancer candidate biomarkers was interrogated in unenriched urine samples using MRM, revealing detection limits of up to 25 picogram/ml urine [35].

With regards to infectious disease biomarker studies, MS-based approaches identified candidate biomarkers in urine for *Leishmania* sp. [36], which has led to the development of a urine capture ELISA diagnostic test [37]. Considerable progress has also been made in *Mycobacterium tuberculosis* [38–40] biomarker studies; recent advancements include the detection of *M. tuberculosis* in urine using IgG capture, immunodepletion and MRM methods [41] and MRM assay of exosomes isolated from serum samples from patients with tuberculosis [38].

In this study, we investigated if *T. pallidum* proteins could be detected in plasma and urine samples from individuals with syphilis using a targeted proteomics (MRM) approach. Successful development of a *T. pallidum* antigen test will most likely be contingent upon the simultaneous detection of multiple protein biomarkers to comprehensively cover different stages of disease. Eleven *T. pallidum* protein biomarkers were chosen based on a predicted specificity, high predicted abundance and physiochemical properties. Thirty surrogate hPTPs were synthesized corresponding to eleven candidate *T. pallidum* biomarkers. Analysis of eighteen plasma and four urine samples revealed no detectable MRM signal for the endogenous peptides from the biomarkers of interest. This is likely due to the extremely low (femtomoles per liter) predicted concentration of bacterial proteins in the samples of interest, or the fact that the biomarkers are not expressed during infection. *T. pallidum* spiking experiments established a MS detection limit of 300 bacteria/ml in PBS; polyclonal anti-*T. pallidum* magnetic bead enrichment did not improve the protein detectability.

Materials and Methods

Ethics statement

The prospective observational cohort study (SeTPAT ClinicalTrials.gov #NCT02059525) that provided the clinical samples used in this study was approved by the Institutional Review Board of the Institute of Tropical Medicine Antwerp (ITM) and the Ethics Committee of the University of Antwerp (13/44/426), Belgium. The *T. pallidum* ssp. *pallidum* DAL-1 strain used in this study was propagated in rabbits at the Veterinary Research Institute in Brno, Czech Republic. The handling of animals in the study was performed in accordance with the current Czech legislation (Animal Protection and Welfare Act No. 246/1992 Coll. of the Government of the Czech Republic). These specific experiments were approved by the Ethics Committee of the Veterinary Research Institute (Permit Number 20– 2014).

Study Participants

Between January 2014 and August 2015, 120 patients attending the ITM clinic, over the age of 17 years and in whom a new diagnosis of syphilis was made and had not received antibiotics in the preceding thirty days were recruited into the study. Thirty HIV-positive controls, in whom the diagnosis of syphilis was excluded via serological and PCR testing, were also recruited. The diagnosis and staging of syphilis was according to the Centers for Disease Control and Prevention classification [42] and treatment was administered according to European guidelines [43]. All patient sera

were tested for syphilis using a RPR test (BD Macro-Vue RPR card test, Becton, Dickinson and Co., Sparks, MD, United States (US)) and an antibody detection *Treponema pallidum* Particle Agglutination test (SERODIA-TPPA Fujirebio Inc., Tokyo, Japan). A PCR test targeting *T. pallidum polA* was also performed on serum [44] and whole blood samples were tested for multiple gene targets [45], as previously described. Selection criteria of participants for the MRM assay analysis included a range of syphilis clinical stages and prioritized predicted high bacterial loads, as demonstrated by positive PCR tests and/or high RPR titres. Patients with early stage syphilis (primary, secondary, early latent) that were plasma and/or whole blood PCR positive for *T. pallidum* were expected to have the highest bacterial load [11,12].

Plasma and urine sample processing

Plasma was collected immediately before Benzathine Penicillin G intramuscular injection using 7.5 ml EDTA-coated blood collection tubes (Sarstedt Monovette, Nümbrecht, Germany). We refer to these samples as the *pre-penicillin* samples. A selection of randomly selected patients participated in an additional blood draw three hours after penicillin treatment since studies have shown penicillin to be fast acting on *T. pallidum,* leading to consequent cell lysis and antigen release [46]. These samples are termed the *post-penicillin* samples. Plasma was chosen for the MRM assay according to HUPO guidelines [47]. Protease inhibitors were not added to the plasma samples since previous studies did not demonstrate a significant higher protein yield with treated samples [48] and peptides could inadvertently be modified [49]. Plasma were subjected to dual centrifugation in an effort to minimize cellular contamination: whole blood was centrifuged at 2000 g for 10 minutes at ambient temperature, followed by transfer of the plasma fraction to a 50 ml falcon tube and centrifugation at 2400 g for 15 minutes. All plasma were processed and aliquoted into cryovials for storage at -80 °C until further testing. Mid-stream random-void urine samples were collected and processed following HUPO guidelines [50], including centrifugation for 10 minutes at 2000 g at ambient temperature in order to remove insoluble contents such as cells and casts. Urine was aliquoted into 15 ml falcon tubes and stored at -80 °C until further testing. All plasma and urine samples were processed within three hours of collection and were only subjected to one freeze thaw cycle.

T. pallidum protein biomarker selection

In a previous descriptive study we used non-gel based complementary MS techniques to characterize the proteome of *in vivo* rabbit cultured *T. pallidum* [16]. Candidate *T. pallidum* biomarker proteins for the MRM assay were chosen based on the following specific criteria: relative protein abundance (based on semi-quantitative spectral

counting techniques [16]), Clusters of Orthologous Groups (COG) functional categorization, microarray transcriptome data [17], protein size, physicochemical properties (i.e. previously detected by MS), predicted subcellular localization [16] and literature review. Each of the candidate biomarkers were digested *in silico* by subjecting the FASTA-formatted sequences to tryptic digestion, assuming 100 % digestion efficiency. Proteotypic peptides (PTPs) corresponding to these proteins were determined using ESPPredictor [51] and pBLAST [52]; analysis of the protein and PTPs was performed to determine possible homology with other bacterial species and human proteins. After PTP selection was finalized, isotopically labelled synthetic peptide standards (hPTPs) corresponding to the selected PTPs were synthesized (Heavy Peptide™ *AQUA* Basic with >95% purity; Thermo Fisher Scientific, Ulm, Germany).

Plasma and urine sample preparation for MRM assay analysis

Protein concentrations of urine and plasma samples were determined based on the area under the curve at 214 nm using a RP-C4 column (Vydac 214TP5415; 4.6×150 mm, particle size 5 μm; Alltech Associates Inc., Lokeren, Belgium) coupled to an Alliance e2695 HPLC system equipped with a 996 PDA detector (Waters Corporation, Milford, MA, United States (US)). For each sample, 250 μg of protein was precipitated by adding six volumes of ice cold LC-MS grade acetone (Biosolve, Valkenswaard, the Netherlands) followed by overnight incubation at -20 °C. In all cases, lo-bind Eppendorf tubes (Eppendorf, Hamburg, Germany) were used to ensure high recovery rates of proteins and peptides. Protein pellets were re-suspended in 50 mM Tris-HCl/6 M urea/5 mM DTT/10 % beta-mercaptoethanol (25 μL/100 μg protein) at pH 8.7. For the denaturation and reduction process all samples were incubated at 65 °C for 1 hour. Subsequently, proteins in all fractions were diluted in 50 mM Tris-HCl/ 1 mM $CaCl_2$ (75 μL/100 μg protein) and alkylated by adding 200 mM iodoacetamide (10 μL/100 μg protein) during 1 hour at ambient temperature and protected from light. Proteomics-grade modified trypsin (Promega, Madison, WI, US) was added at a 30:1 protein-to-enzyme ratio. After incubation at 37 °C for 18 hours the digestion was stopped by freezing the samples. Protein digests were desalted by SPE using GracePure SPE C18-Max (50 mg) (W. R. Grace & Co., Columbia, MD, US) RP cartridges and a vacuum manifold. SPE cartridges were conditioned with 100 % methanol and equilibrated with 100 % LC/MS grade H_2O and 0.1 % formic acid (FA). After loading the complete acidified (0.1 % FA) tryptic digest, peptides were washed with 10 % methanol and eluted with 40 % methanol/ 40 % acetonitrile (ACN) and 0.1 % FA. Eluted peptides were lyophilized and frozen at -20 °C until further analysis. Immediately before

analysis, lyophilized digests were resuspended in 5 % ACN/0.1 % FA and spiked with a mixture of all hPTPs.

MRM assay optimization and mass spectrometric analysis

Optimization of each PTP was performed on a triple quadrupole mass spectrometer (Waters Xevo TQ) (Waters Corporation, Milford, MA, US) in order to obtain the most intense transitions. The capillary voltage was tuned to approximately 2 kV with a source temperature of 150 °C. Desolvation temperature was set at 400 °C with a nitrogen gas flow of 800 L/h. Cone voltage, collision energy and dwell times were optimized for each of the PTPs. All PTPs were dissolved in mobile phase A (MP-A), containing 5 % ACN (LC/MS grade) and 0.1 % FA. For each of the peptides individually, the Limit of Detection (LOD) was determined by performing a dilution series in MP-A. Based on these concentrations, a mixture of all hPTPs was made. A balanced hPTP mixture has been shown to increase quantification accuracy and reproducibility compared to an equimolar mixture in previous studies [34]. To check for possible suppressive effects of the plasma matrix, the hPTP mixture was spiked into plasma from a control study subject. Therefore a balanced mixture of hPTP (concentrations detailed in *S1 Table*) was spiked into 50 μg of plasma digest. Chromatographic separation of the plasma and urine samples was performed on a RP-C18 UPLC column (Waters, CSH 150 x 2.1 mm, 1.7 μm at 35 °C) connected to an Acquity UPLC system (Waters Corporation, Milford, MA, US). In order to separate all peptides as best as possible, an optimized linear gradient of Mobile Phase B (MP-B) (0.1 % FA in 100 % ACN) was applied: 5 % MP-B during 1 min and from 5 to 35 % MP-B in 5 min, followed by a steep increase to 100 % MP-B in 1 min, all at a flow rate of 300 μL/min. Based on the specific retention times of each peptide, three scheduled MRM runs of 10 minutes were generated, each of them containing 20 MS1 channels (10 endogenous (*T. pallidum)* PTPs without isotopic label and 10 channels with a synthetic hPTP equivalent). At least three transitions (ion pairs) were selected for each peptide of interest. For each scheduled MRM analysis, 50 μg of peptides (injection loop of 5 μL) per plasma/urine sample were loaded onto the analytical column. In addition to an extensive needle wash after each injection, a blank run was performed between two subsequent clinical samples to prevent carry-over effects. Data acquisition was controlled by MassLynx version 4.1, while targeted datasets were analysed by TargetLynx, which is part of MassLynx (Waters Corporation, Milford, MA, US). All Xevo TQ MS raw spectral files are available at PeptideAtlas (http://www.peptideatlas.org/) [53] with the identifier PASS00978.

Magnetic bead antibody-based enrichment of *T. pallidum* proteins and approximation of the MS LOD for *T. pallidum* protein detection

T. pallidum protein enrichment was performed using magnetic beads (Dynabeads® M-270, Life Technologies, CA, US) coated with biotin-conjugated polyclonal *T. pallidum*-specific antibodies (PA1-73103, Thermo Fisher Scientific, CA, US) through streptavidin-biotin conjugation. According to the manufacturer's protocol, 10 μg of antibody was used to bind 1 mg of beads (approximately 5 x 10^7 beads).

In vivo rabbit cultured purified *T. pallidum* DAL-1 strain extracts [54,55] were kindly provided by the group of David Šmajs from the Masaryk University, Czech Republic. The original concentration of the *T. pallidum* extract was approximately 10^6 bacteria/ml as quantified under dark-field microscopy. Samples were stored in 1 ml phosphate buffered saline (PBS) and only subjected to one freeze-thaw cycle. Two dilution series of *T. pallidum* were prepared, each time starting in 1 ml of PBS and finally equating to eight approximate bacterial concentrations: 10^4, 10^3, 300, 100, 33, 10, 3 and 0 bacteria/ml.

For one dilution series, each of the eight fractions were incubated with a constant amount (~10^5) of magnetic beads coated with polyclonal anti-*T. pallidum* antibodies. After incubation for two hours at 4° C and magnetic separation, the supernatant was discarded and beads were washed three times with PBS. To lyse the antibody bound bacteria, 1 ml of PBS was added to each bead sample, these were sonicated on ice using a Vibracell sonicator (two times 30 seconds with an amplitude of 50 %). The bead fraction was retained (retentant) after sonication by using magnetic separation. Released proteins were precipitated adding ice-cold acetone and incubated overnight at -20 °C. Tryptic digestion was performed, following the aforementioned procedure, on both the precipitated proteins (supernatant) and directly "on-bead" (retentate), to test for possible unreleased proteins during sonication. For the second dilution series (unenriched), 1 ml was directly drawn from each of the eight samples. The samples from this series were also sonicated on ice (two times 30 seconds with an amplitude of 50 %) to lyse the bacteria. Released proteins were then acetone precipitated and subsequently digested, in conformance with the other parallel series procedure.

Liquid chromatography- electrospray ionization- LTQ-Orbitrap mass spectrometry analysis of enriched and non-enriched serially diluted *T. pallidum* samples

Peptide mixtures were separated by RPLC on a Waters nano-UPLC system using a nanoACQUITY BEH C18 Trap column (100 Å, 5 µm, 180 µm X 20 mm) connected to a nanoACQUITY BEH C18 analytical Column (130 Å, 1.7 µm, 100 µm X 100 mm) (Waters Corporation, Milford, MA, US). Peptides were dissolved in MP-A, containing 2 % ACN and 0.1 % FA and spiked with 20 fmol [Glu1]-fibrinopeptide B, which serves as an internal calibrant. A linear gradient of MP-B (0.1 % FA in 98 % ACN) from 2 to 45% MP-B in 45 min, followed by a steep increase to 95 % MP-B in 2 min at a flow rate of 400 nl/min. The nano-LC was coupled online with a LTQ Orbitrap Velos (Thermo Scientific, San Jose, CA, US) mass spectrometer using a PicoTip Emitter (New Objective, Woburn, MA, US) linked to a nanospray ion source. The mass spectrometer was set up in a data dependent acquisition MS/MS mode where a full scan spectrum (350–2500 m/z, resolution of 60.000) was followed by a maximum of ten CID tandem mass spectra (100 to 2000 m/z). Peptide ions were selected as the twenty most intense peaks of the MS scan. CID scans were acquired in the LTQ IT part of the mass spectrometer with normalized collision energy of 32 %.

Obtained spectra were screened against the *T. pallidum* reference and resequenced databases (UniProt ID proteome UP000014259 [14] and UP000000811 [56] using the MASCOT search engine (Matrix Science; version 2.1.03) based on the digestion enzyme trypsin. Carbamidomethylation of cysteines was listed as a fixed modification, while methionine oxidation was set as a variable modification. A maximum of one missed cleavage was tolerated. Mass tolerance was set to 10 ppm for the precursors and 0.8 Da for the fragment ions. False discovery rate was set at 5 %. Scaffold Q+ (version 4.6.2, Proteome Software Inc., Portland, OR, US) was used to validate MS/MS-based peptide and protein identifications. Protein identifications were accepted if they could be established at greater than 95.0 % probability according to the protein prophet algorithm [57].

All LTQ-Orbitrap MS/MS raw spectral data is available at PeptideAtlas [53] with the identifier PASS00978.

Results

Study subject inclusion

Eighteen syphilis-positive study participants were selected for the MRM assay analyses (*Table 1*). All participants were male and identified as MSM. A third of the participants (6/18; 33 %) were HIV positive. Five (28 %) presented with primary, eleven secondary (61 %) and two early latent (11 %) stage disease. Thirteen participants were confirmed *T. pallidum*-positive by serum and/or whole blood PCR testing. Four participants had indeterminate PCR results, meaning their sample was weakly positive. A second confirmatory PCR was not performed on these samples. One patient was negative for both whole blood and serum PCR. All participants tested positive with both the RPR and TPPA tests. The median RPR value was 1/64 (IQR: 1/16- 1/128). In total, 22 samples were analysed, including N=12 pre-penicillin treatment plasma, N=6 post-penicillin treatment plasma and N=4 pre-penicillin treatment urine samples.

Table 1 *Summary of the clinical and laboratory characteristics of study subjects included in this study*

Study Nr.	HIV status	Syphilis stage	Sample type	Pre or post-treatment sampling[#]	PCR Whole Blood	PCR Serum	RPR titre	TPPA titre
1	Positive	Secondary	Plasma	Pre	Positive	Indet.	1/512	>1/20480
2	Negative	Primary	Plasma	Pre	Negative	Indet.	1/4	1/160
3	Positive	Early latent	Plasma	Pre	Negative	Positive	1/1	1/1280
4	Positive	Secondary	Plasma	Pre	Positive	Positive	1/128	1/20480
5	Positive	Secondary	Plasma	Pre	Positive	Positive	1/128	>1/20480
6	Negative	Secondary	Plasma	Pre	Negative	Positive	1/128	>1/20480
7	Positive	Early Latent	Plasma	Pre	Positive	Positive	1/64	1/10240
8	Positive	Secondary	Plasma	Pre	Positive	Indet.	1/32	1/1280
9	Positive	Secondary	Plasma Urine	Pre	Positive	Positive	1/512	>1/20480
10	Negative	Primary	Plasma Urine	Pre	Positive	Indet.	1/16	1/5120
11	Positive	Secondary	Plasma Urine	Pre	ND	Indet.	1/128	>1/20480
12	Negative	Secondary	Plasma	Pre	Positive	Negative	1/32	>1/20480
13	Positive	Secondary	Plasma	Post	ND	Indet.	1/128	>1/20480
14	Negative	Primary	Plasma Urine	Post	Positive	Indet.	1/16	1/5120
15	Negative	Primary	Plasma	Post	Negative	Indet.	1/8	1/1280
16	Positive	Secondary	Plasma	Post	Positive	Negative	1/64	1/20480
17	Positive	Primary	Plasma	Post	Positive	Negative	1/64	>1/20480
18	Positive	Secondary	Plasma	Post	Negative	Negative	1/128	>1/20480

Legend: [#]- patients were treated with intramuscular injection with 2.4 MU Benzathine penicillin G; Indet.- indeterminate PCR result, second confirmatory PCR was not performed; ND- not done

T. pallidum protein biomarker selection

Eleven *T. pallidum* proteins were selected as candidate biomarkers (*Table 2*). Most selected biomarkers had high normalized spectral abundance factor (NSAF) scores according to our previous study [16] (median 4.02; IQR: 1.97-6.97) and high microarray signal ratios [17] (median 3.05; IQR: 0.74-6.8). The median protein molecular weight was 39 kDa (IQR: 28- 81). Two proteins were predicted to be located in the flagellum (TP_0249 and TP_0792), two in the ribosome (TP_0250b and TP_0244) and the subcellular localization of five proteins was unknown. Protein TP_0326, a BamA

orthologue, has been experimentally shown [58–60] to be localized in the outer membrane. A typical target for PCR assays is *polA*, coding protein TP_0105 [61]. One protein, Peptidyl-prolyl cis-trans isomerase (TP_0862) was found in a previous proteomics study where it demonstrated moderate reactivity during immunoblot experiments with human and rabbit *T. pallidum* infected serum [15]. Protein TprG (TP_0317) is part of the paralogous *tpr* gene family that encodes candidate virulence factors [62] and is partially homologous to Tpr E/J. According to pBLAST analysis, all chosen biomarker proteins and corresponding PTPs did not demonstrate high homology with other pathogens, non-pathogenic commensal bacterial or human proteins (data not shown). One to three corresponding well-suited PTPs were selected for each biomarker, for a total of 30 PTPs. Details of these are provided in *Table 2.*

Table 2 *List of* T. pallidum *protein biomarker candidates and their corresponding proteotypic peptides (PTPs) tested in MRM analyses of urine and plasma samples from individuals with syphilis*

Number	UniProt Accession Number[a]	TP Nr./ gene	Protein Name	Peptide Number	Peptide Sequence[g]	Protein Weight (kDa)	Predicted Subcellular Location[&]	COG category function	Spectral Count NSAF value in *T. pallidum* [16]	cDNA/ DNA signal ratio [17]
				16	TSAVSGAIP**I**ENR					
				17	MALNTQIQSSAADI**V**K					
1	R9US76	TP_0105/ *polA*	DNA-directed DNA polymerase	18	VHTSFVQIGT**A**TGR	112	NK	L	2.55	0.283
				19	TEAGGVVVQFT**I**QEGK					
				20	EQWASSPGLAES**F**R					
				21	LAFANTFTSPGG**I**PK					
				20	EQWASSPGLAES**F**R					
			Putative outer membrane protein	21	LAFANTFTSPGG**I**PK		Outer			
2	O83346	TP_0326/ tp92	assembly factor	23	DESVL**I**DFAK	94	Membrane	M	1.77	0.682
				22	LATEVGFTPSGG**A**QR					
3	R9UVD9	TP_0249/ flaA1	Flagellar filament outer layer protein	23	DESVL**I**DFAK	39	NK	N	7.47	16.05
				24	GTLLDGTVFD**A**SR					
				25	KPGVQVTSSGLQYEVV**K**					
4	O83834	TP_0862/ fklB	Peptidyl-prolyl cis-trans isomerase	27	FYVPSSLGYGE**R**	28	NK	O	NF	5.29
				26	MPPSPC**A**VLR					
				30	VASVVVISVDN**R**					
				28	YFLPGEC**A**GR					
				30	VASVVVISVDN**R**					
5	O83892[&]	TP_0922	Uncharacterized Protein	28	YFLPGEC**A**GR	33	NK	None	7.80	1.599
6	R9USJ3	TP_0250b/ rpsT	30S ribosomal protein S20	29	LYNGVFSSPEVV**R**	11	Ribosome	None	5.28	3.39
				10	TGEEPLPV**F**TK					
				9	ATAVGIMYDC**L**ER					
7	R9UU30	TP_0244/rpsG	30S ribosomal protein S7	11	LAAEILDAYHSTGT**A**FK	18	Ribosome	J	6.78	3.053
				1	VLDAVTAATETALQS**R**					
				8	GNPMSLFNLPDQQ**K**					
8	O83337	TP_0317	Tpr protein G	2	LTGSATLEWGISYG**K**[s]	81	NK	None	1.32	0.743
				6	ELSVQAANGIYS**A**EDR					
				7	DAGDESVMNIDSPE**K**					
9	P21991	TP_0792/ flaB1	Flagellar filament core protein FlaB	12	AYIGTMTAVAMG**I**R	31	Flagellum	N	6.97	13.82

Table 2 *List of* T. pallidum *protein biomarker candidates and their corresponding proteotypic peptides (PTPs) tested in MRM analyses of urine and plasma samples from individuals with syphilis*

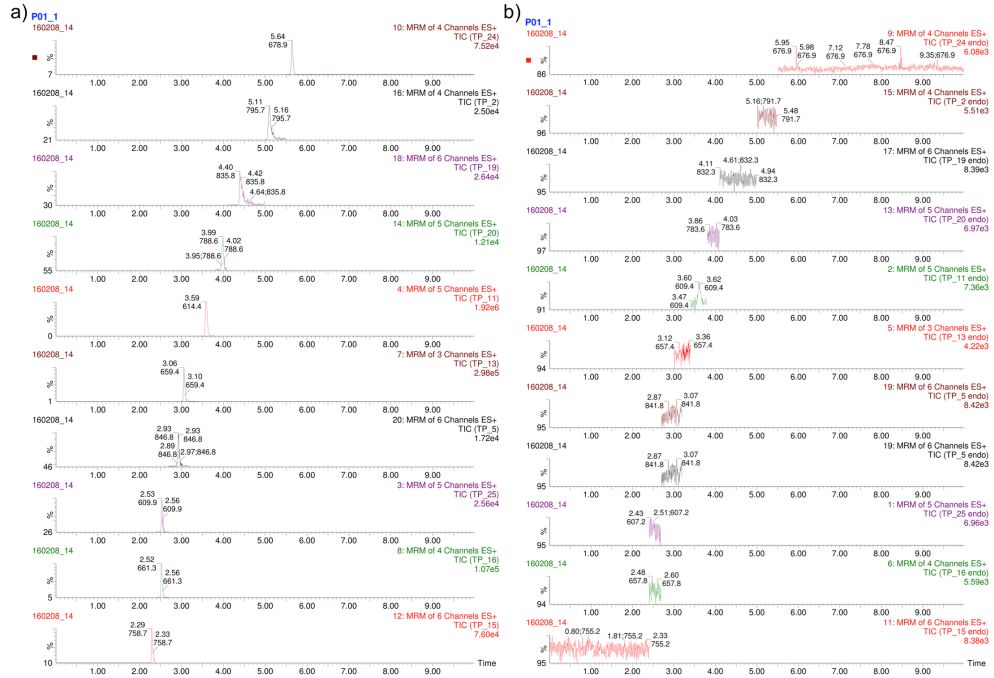
Number	UniProt Accession Number*	TP Nr./ gene	Protein Name	Peptide Number	Peptide Sequence#	Protein Weight (kDa)	Predicted Subcellular Location&	COG category function	Spectral Count NSAF value in *T. pallidum* [16]	cDNA/DNA signal ratio [17]
10	R9UTS8	TP_0748/ cfpA/ tpn83	Cytoplasmic filament protein A	3	GVNELETHTNSL**L**R	79	Cytoplasm	S	2.75	6.79
				4	ADIGQSFASDGS**A**DQK					
				5	EYDDTDISNLPDE**R**					
11	O83417	TP_0402/ fliI	Flagellum-specific- ATP synthase	13	EIGLASGELP**A**TR	48	Flagellum; Cytoplasm	NU	1.97	1.241
				14	SVIVSATSDESPL**A**R					
				15	VGAYQQGSDAE**L**DR					

*Legend: *- UniProt proteome ID UP000014259; &- ORF was not annotated in the re-sequenced Nichols strain genome due to its length below the 150 bp limit [14]; #- underlined/bold amino acids indicate stable isotope labelled residues; $- peptide is homologous in Tpr E/G/J protein sequences; @- subcellular location as reported in Osbak* et al. *[16]; NK- not known; NSAF- normalized spectral abundance factor; COG- clusters of orthologous groups; COG categories: L- Replication, recombination and repair, M- Cell wall/membrane/envelope biogenesis; N- Cell motility; O- Posttranslational modification, protein turnover, chaperones; J- Translation, ribosomal structure and biogenesis; S- Function unknown; U- Intracellular trafficking, secretion, and vesicular transportation*

Multiple reaction monitoring assay optimization

The LOD for each peptide was determined individually by performing a dilution series of MP-A whereby the median LOD was 68.5 (IQR 14.2-176.7) picomoles. Once the peptide mixture composition was optimized based on the LOD, 2 μL of this mixture (*S1 Table*) was spiked into 50 μg plasma from a control patient whereby no significant variations in the signal of the hPTP transitions could be detected, indicating that there was no evidence of transition interference from the plasma. After optimizing each of the PTPs, three different sets of transitions were combined in an MRM assay based on their chromatographic retention time, as detailed in *S1 Table*. The experiments contained a total of 141 targeted ion pairs (transitions) corresponding to 30 PTPs from eleven *T. pallidum* proteins. Ten of the eleven proteins were represented by two or more (h)PTPs (*Table 2*/*S1*). In total, three scheduled MRM assays of 10 minutes, each containing 20 peptides (10 endogenous (*T. pallidum*) peptides and 10 hPTP standards) were developed. These assays were evaluated based on a balanced mixture of all 30 hPTPs standards. Unfortunately, although each of the 30 spiked hPTPs could be detected, none of the selected endogenous *T. pallidum* peptides could be identified in any of the MRM assays (*Fig 1*/*S1*).

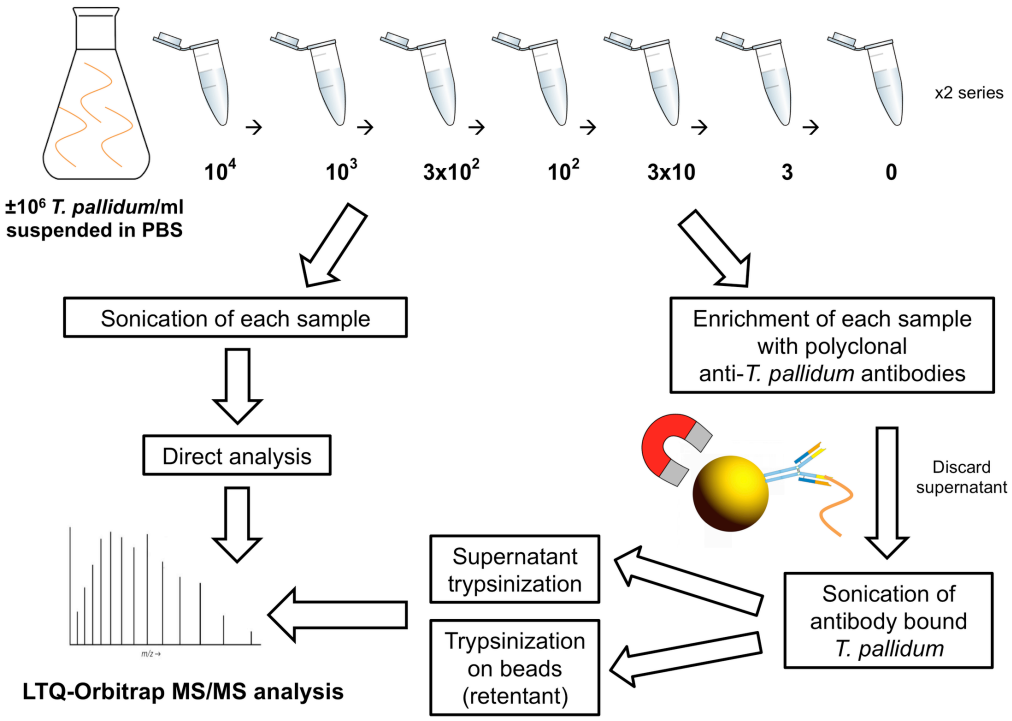
Fig 1 *Examples of intensity plots displaying selected multiple reaction monitoring experiments for a plasma sample from a secondary stage syphilis-infected individual: (a) synthetic hPTPs, even numbers and (b) endogenous (*T. pallidum*) PTPs, odd numbers; gradient 1 of 3. For each peptide the number of selected transitions (channels) is reported. The x-axis shows the chromatographic retention time of the corresponding peptide while the y-axis shows the relative intensity of the MS2 signal. Note: Signal fluctuations present in the 'endogenous' PTP chromatogram are always the result of just one transition, often coupled with a shift in retention time and differing m/z-values differ from the hPTP run, thus these are considered to be noise.*



Estimation of mass spectrometry LOD and ineffective *T. pallidum* protein enrichment using magnetic bead coupled polyclonal anti-*T. pallidum* antibodies

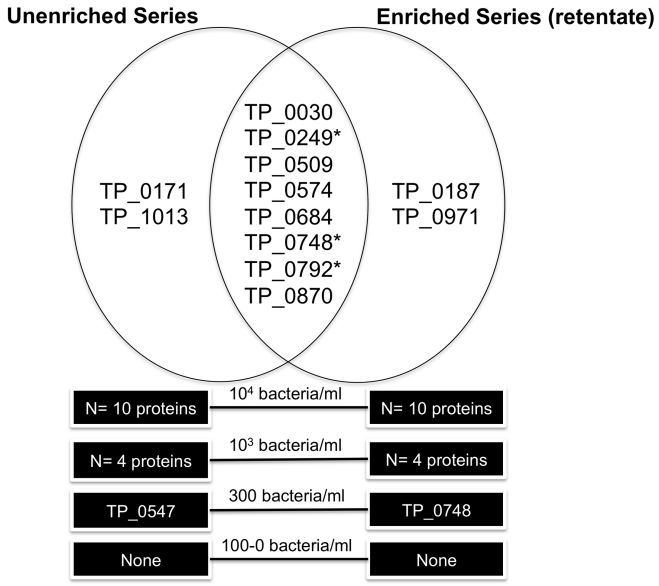
Two *T. pallidum* spiking dilution series were prepared in PBS and subjected to LTQ-Orbitrap MS/MS analysis in order to estimate the LOD of MS detection. One of the series was subjected to an additional polyclonal antibody coupled magnetic bead enrichment step, including sonication of the beads and subsequent separate measurement of the lysate and on-bead digestion retentate (*Fig* 2).

Fig 2 ***Work-flow diagram describing the estimation of* T. pallidum *protein MS LOD experiments conducted in this study.*** In total, eight different concentrations of *T. pallidum* (from 10^4 to 0 bacteria/ml PBS) were treated in three different ways i) *T. pallidum* was enriched using magnetic beads coated with polyclonal anti-*T. pallidum* antibodies, and lysed by sonication for release of *T. pallidum* proteins in the supernatant. Acetone precipitated proteins were trypsinized; ii) In order to detect any remaining protein on the beads, the beads were also trypsinized (retentant on-bead trypsinization); iii) As a control, non-enriched samples were sonicated and immediately trypsinized. *-proteins selected as candidate biomarkers in this study. All samples were analysed by an LTQ-Orbitrap mass spectrometer.



Two unique *T. pallidum* proteins, Cytoplasmic filament protein A (TP_0748) and Lipoprotein antigen Tp47 (TP_0547), were found in the 300 bacteria/ml fraction in the enriched and unenriched samples, respectively (*Fig 3; S2 Table*). Therefore, the LOD based on a high-resolution LTQ- Orbitrap instrument was approximately 300 bacteria/ml PBS for both the antibody enriched and unenriched samples, meaning there was no significant improvement in LOD using bead enrichment. No proteins were detected in any sample concentrations for the enriched bacterial lysate (supernatant) fraction. Possibly, the sonication conditions were not harsh enough to lyse the bacteria on the beads and lysis was mainly the results of trypsin treatment under denaturing conditions. In total, eight unique *T. pallidum* proteins were found in both the unenriched and enriched retentate dilution series: 60 kDa chaperonin (TP_0030), Flagellar filament outer layer protein flaA1 (TP_0249), Alkyl hydroperoxide reductase (TP_0509), Lipoprotein antigen Tp47 (TP_0574), Galactose ABC superfamily ATP binding cassette transporter, binding protein (TP_0684), Cytoplasmic filament protein A (TP_0748) and the Flagellar filament core proteins flaB1/B3 (TP_0792/TP_0870). Four proteins, Lipoprotein, 15 kDa (TP_0171), 10 kDa chaperonin (TP_1013), Elongation factor Tu (TP_0187) and Tp34 lipoprotein (TP_0971) were only found in the unenriched and enriched series, respectively. Ten unique *T. pallidum* proteins were found in the highest concentration (10^4 bacteria/ml) for both the enriched retentate sample (N= 10) and non-enriched sample (N=10). Five unique *T. pallidum* proteins were found in the 10^3 bacteria/ml sample, including N= 4 in the unenriched and N=4 in the retentate fractions. A peptide (LSGGVAVIK) related to 60 kDa chaperonin (TP_0030) was detected in the low concentration (100/ 33/ 10/ 3 bacteria/ml) and in the negative control samples of the enriched sample series. This was likely a false-positive non-specific peptide secondary to rabbit protein contamination since this short peptide sequence is closely homologous to the *Oryctolagus cuniculus* (rabbit) 60 kDa heat shock protein, or could have originated from the beads or antibodies. As a result it has been excluded from the analysis. Three *T. pallidum* proteins detected in both the enriched and unenriched sample series were also biomarker candidates tested in the MRM assay experiments: Flagellar filament core protein flaB2 (TP_0792), Cytoplasmic filament protein A (TP_0748) and the Flagellar filament outer layer protein flaA1 (TP_0249). Detailed information about the identified proteins, peptides, coverage and search parameters can be found in *S2 Table*. Rough concentration calculations estimated that our target PTPs would be present in the femtomoles per liter range in human *T. pallidum* infection (calculations presented in *S3 Table*).

Fig 3 T. pallidum *proteins detected in LOD magnetic bead coupled polyclonal anti-*T. pallidum *antibody enrichment experiments (protein and peptide identification threshold of 95 %).*



Discussion

The *T. pallidum* MRM assay designed in this study failed to detect any of the 30 targeted proteotypic peptides related to eleven candidate *T. pallidum* protein biomarkers in eighteen plasma and four urine samples from individuals with syphilis. A number of explanations are possible. The foremost is the extremely low predicted concentration of bacterial proteins compared to host proteins. To a large extent our estimates of *T. pallidum* bacterial load in blood are based on molecular studies. In one of the largest studies, Tipple *et al.* found that median copy numbers of Lipoprotein antigen Tp47 (TP_0574) DNA detectable per milliliter of whole blood was 127, 516 and 70 in primary, secondary and latent syphilis, respectively [12]. Other studies have produced comparable results [46,63,64], with the exception of a recent study that found a median of 1.4 x 10^5 *T. pallidum*/ml in whole blood from patients with secondary syphilis [65].

The concentration of *T. pallidum* in blood according to these PCR-based studies is lower compared to our estimated LOD in a shotgun experiment on diluted samples (300 *T. pallidum*/ml) since we would need a 500x higher concentration (same amount of proteins from 300 *T. pallidum* in 1 ml vs. 2 μl) to detect the 300 *T. pallidum*/ml (see *S3 Table*). Despite this outcome, we were hoping to detect *T. pallidum* proteins in the

plasma or urine of some syphilis patients because i) MRM measurements are generally more sensitive than shotgun experiments since scanning times are drastically reduced and ii) the amounts from Tipple *et al*. [12] were averages so we hypothesized that some patients (especially those with secondary syphilis) might have high *T. pallidum* levels detectable by MRM. These results could then motivate us to develop an (immuno)assay, capable of detecting the proteins even at low concentrations.

Little difference in *T. pallidum* abundance has been found between whole blood, plasma or serum [11]. Not much is known about the persistence of *T. pallidum* in the human urinary tract and to our knowledge no studies have quantified *T. pallidum* in the urine of syphilis-infected patients. However, even if *T. pallidum* does not consistently persist in the urinary tract, bacterial proteins present in the blood could be filtered through the glomerulus, ending up in the urine either intact or as peptide fragments, depending on the size of the protein and state of proteolysis [66].

These considerations suggest that detection of *T. pallidum* proteins in human biofluids may not be possible without additional steps such as front-end immunoaffinity depletion [67], two-dimensional LC separation [68] and/or selective enrichment of target proteins/peptides (as reviewed by Shi *et al*. [69]). These techniques, or combinations thereof, have allowed the detection of low abundance proteins up to the low- to sub-nanogram/ml level [69,70] in clinical samples. For example, to reduce the wide dynamic range of plasma proteins, multicomponent single-step immunoaffinity depletion of high-abundant (host) proteins can allow up to a 10-20-fold enrichment of low-abundant proteins due to the depletion of 90–95 % of the total protein mass [67]. However, of particular concern with this approach is the possibility of concomitant removal of low-abundance proteins due to protein binding to the antibodies or high-abundant proteins, as shown in a study that systematically analysed the antibody bound (high-abundant) protein fraction which found that this fraction contained 101 proteins at a high degree of confidence [71]. *T. pallidum* has a high binding affinity for constituents of serum and host cells, including laminin [72], fibronectin [73,74] and albumin [75], which may lead to unintentional depletion of targeted proteins if human protein specific immunodepletion would be applied. Furthermore, targeted mass spectrometric immunoassays (MSIA) that use surface-immobilized antibodies to affinity retrieve proteins from biological samples have proven their utility for clinical applications [76–78]. In our study, magnet bead coupled polyclonal anti- *T. pallidum* antibodies failed to significantly detect more *T. pallidum* proteins compared to the unenriched dilution series. Antibody effectivity is dictated by binding affinity; we used

commercial antibodies that were to our knowledge not previously characterized as to their binding affinity or targeted proteins. Furthermore, it is unlikely that the polyclonal antibodies would bind a large range of proteins since few (<5 %) *T. pallidum* proteins are immunogenic [15,79]. The fact that *T. pallidum* can remain in 'plain sight' without invoking immune defences [80], together with the very low amount of outer membrane proteins compared to other human pathogens [81], also suggests that antibody enrichment of whole organisms and/or proteins would probably not be an effective strategy. Peptide-level immunoenrichment, also known as the 'Stable Isotope Standards and Capture by Anti-Peptide Antibodies' (SISCAPA) method developed by Anderson *et al.* [82] has shown considerable promise as a high-throughput, automated, highly multiplexed approach for protein biomarker quantification, with MRM application detection limits in the low picogram/ml range of protein concentration in plasma [83]. If a selection of *T. pallidum* peptides could be definitively demonstrated to be present in plasma or urine, then this could be an attractive analytical approach with a strong potential for yielding the detection capabilities and precision needed for clinical applications.

However, apart from the low abundance in plasma or urine, other factors could explain why the *T. pallidum* proteins were not detected in our MRM experiments: 1) The LOD *T. pallidum* spiking experiments were performed in PBS buffer as opposed to a highly complex plasma or urine matrix background. 2) Variations in gene expression and structural components of proteins could also account for the lack of *T. pallidum* protein detection. Fluctuations in gene expression may explain why we did not find TprG, a protein implicated in phase variation which has been shown to be expressed at varying levels during infection due to changes in the number of guanine nucleotide repeats immediately upstream of its transcriptional start site [84]. Heterogeneous *T. pallidum* protein sequence sites [14,16,85] could also confound rigid MRM assay detection parameters. Such heterogeneity has been shown [16] to be present in one candidate biomarker, TP_0922, although this variable site was not present in the PTPs incorporated in this MRM assay. Poor proteolytic cleavage can stem from structural features of the protein, different digestion kinetics and post-translational modifications. For example, phosphorylated residues within two amino acids of the point of cleavage can hinder proteolysis [86]. Little is known about the extent of *T. pallidum* protein post-translational modification aside from a study that demonstrated glycosylation of the Flagellar core proteins (FlaBs) as reported by antibody and glycan staining techniques [87], however, the exact modification sites and extent of modification remain unknown. Other proteomics studies of *L. interrogans* have demonstrated likely roles for protein

acetylation and methylation in virulence mechanisms [88,89]. 3) Furthermore, we only tested eleven out of more than a thousand predicted proteins in the *T. pallidum* proteome [56], a selection largely based on spectral counting [16] as an estimation of protein abundance. We cannot assume, however, that this indirect manner of quantifying *T. pallidum* protein levels in a rabbit testicle model directly recapitulates *T. pallidum* protein expression levels in plasma samples of syphilis-infected patients. One of the reasons for this is that protein expression may vary according to host and disease stage. Antigen detection during latent stage disease will be especially challenging since *T. pallidum* has been shown to sequester itself in protected niches such as eyes, hair follicles and nerves [90]. Other *T. pallidum* proteins may be more suitable diagnostic biomarkers, given that they are reflective of the disease stages studied and that they are consistently present in the biofluids of interest. For example, Lipoprotein Tp47, which could still be identified in the most diluted *T. pallidum* sample (300 *T. pallidum*/ml) in this study, could be an interesting biomarker for future studies. 4) Various technical limitations such as a possible suboptimal chromatographic gradient length, modifiable proteotypic residues and protein degradation secondary to sample processing could have impeded biomarker detection. Other studies have reported chromatographic gradient lengths of 30 minutes or longer [32,33,35,38], thus implementation of longer gradients could be considered in future studies in order to improve peptide resolution. In this study, chromatographic separations were performed in triple using shorter 10-minute gradients in order to optimize the sample throughput without the loss of MS sensitivity due to overlapping transition windows. Therefore, co-eluting peptides were split over different chromatographic runs since plasma protein availability was not a limiting factor. Oxidizable proteotypic residues, namely cysteine, methionine and tryptophan, can cause artifactual modifications during processing or storage resulting multiple forms of targeted peptides. With this said, the PTP selection process also requires a necessary balance between many different parameters, whereby selection of peptides containing suboptimal amino acid residues can sometimes remain the most favourable option. Ribosomal protein TP_0250b was only represented by one PTP, which may have limited detectability, thus future assays could ideally incorporate more than one peptide per protein. 5) Sample processing may have also contributed to protein degradation; therefore prompt analysis of fresh non-frozen biological specimens, if possible, is recommended. Moreover, alternative sample processing procedures, such as the use of molecular weight cut off filters to concentrate urine could improve protein detectability [39]. 6) Lastly, only a limited amount of clinical samples were analysed, especially urine and the study was a single-centre study with only MSM participants, therefore it is not generalizable. An improvement for future

studies would be the incorporation of isotopically labelled (non-*T. pallidum*) reference standards, which have been shown to improve analytical precision, detect variations in instrument performance and aid in detecting chemical interferences [91].

Targeted MS approaches are only able to search for a limited amount of pre-selected biomarker candidates. A more comprehensive approach would be to take a step backwards to conduct broader shotgun proteomics in plasma and urine samples of individuals with syphilis. Shotgun approaches identifying *M. tuberculosis* antigens in urine have been previously successful [39,40]. A compelling study from Eyford *et al.* used a 'deep-mining' proteomics approach and were able to detect 254 *Typanosoma brucei rhodesiense* proteins in plasma from African sleeping sickness patients [92]. Quantitative data independent acquisition (DIA) modes of MS analysis, including SWATH-MS [93], are also very promising avenues for clinical applications [94,95].

In an effort to identify promising *T. pallidum* diagnostic biomarkers, we designed a scheduled MRM assay incorporating 141 MRM ions pairs correlated to 30 PTPs/ 11 *T. pallidum* proteins. Factors such as the extremely low (femtomoles per liter) predicted *T. pallidum* protein concentration in biofluids, possible variable protein expression according to host/disease stage and potential presence of protein post-translational modifications likely contributed to the lack of signal detection for all candidate biomarkers investigated. Since the proteins targeted in this study were likely buried in the proverbial haystack of plasma proteins, alternative samples preparation and analysis strategies are warranted. With the rapidly progressing innovations of MS applications and technology, we believe clinical proteomics is far from its pinnacle of potential.

References

1. Newman L, Rowley J, Vander Hoorn S, Wijesooriya NS, Unemo M, Low N, et al. Global Estimates of the Prevalence and Incidence of Four Curable Sexually Transmitted Infections in 2012 Based on Systematic Review and Global Reporting. PLoS One. 2015;10(12):e0143304.
2. Van de Laar M, Spiteri G. Increasing trends of gonorrhoea and syphilis and the threat of drug-resistant gonorrhoea in Europe. Euro Surveill. 2012;17(29):2011–3.
3. Peterman TA, Su J, Bernstein KT, Weinstock H. Syphilis in the United States: on the rise? Expert Rev Anti Infect Ther. 2015;13(2):161–8.
4. Kenyon C, Lynen L, Florence E, Caluwaerts S, Vandenbruaene M, Apers L, et al. Syphilis reinfections pose problems for syphilis diagnosis in Antwerp, Belgium - 1992 to 2012. Euro Surveill. 2014;19(45):20958.
5. Ogilvie GS, Taylor DL, Moniruzzaman A, Knowles L, Jones H, Kim PH-J, et al. A population-based study of infectious syphilis rediagnosis in British Columbia, 1995-2005. Clin Infect Dis. 2009;48(11):1554–8.
6. Seña ACC, White BLL, Sparling PFF. Novel Treponema pallidum serologic tests: a paradigm shift in syphilis screening for the 21st century. Clin Infect Dis. 2010;51(6):700–8.
7. Belisle JT, Brandt ME, Radolf JD, Norgard M V. Fatty acids of Treponema pallidum and Borrelia burgdorferi lipoproteins. J Bacteriol. 1994;176(8):2151–7.
8. Joyanes P, Borobio M V, Arquez JM, Perea EJ. The association of false-positive rapid plasma reagin results and HIV infection. Sex Transm Dis. 1998;25(10):569–71.
9. Monath TP, Frey SE. Possible autoimmune reactions following smallpox vaccination: the biologic false positive test for syphilis. Vaccine. 2009;27(10):1645–50.
10. Sena AC, Wolff M, Martin DH, Behets F, Van Damme K, Leone P, et al. Predictors of serological cure and Serofast State after treatment in HIV-negative persons with early syphilis. Clin Infect Dis. 2011;53(11):1092–9.
11. Gayet-Ageron A, Lautenschlager S, Ninet B, Perneger T V, Combescure C. Sensitivity, specificity and likelihood ratios of PCR in the diagnosis of syphilis: a systematic review and meta-analysis. Sex Transm Infect. 2013;89(3):251–6.
12. Tipple C, Hanna MOF, Hill S, Daniel J, Goldmeier D, McClure MO, et al. Getting the measure of syphilis: qPCR to better understand early infection. Sex Transm Infect. 2011;87(6):479–85.
13. Castro R, Prieto E, Aguas MJ, Manata MJ, Botas J, Santo I, et al. Detection of Treponema pallidum sp pallidum DNA in latent syphilis. Int J STD AIDS. 2007;18(12):842–5.
14. Pětrošová H, Pospíšilová P, Strouhal M, Čejková D, Zobaníková M, Mikalová L, et al. Resequencing of Treponema pallidum ssp. pallidum Strains Nichols and SS14: Correction of Sequencing Errors Resulted in Increased Separation of Syphilis Treponeme Subclusters. Dong Q, editor. PLoS One. 2013;8(9):e74319.
15. McGill MA, Edmondson DG, Carroll JA, Cook RG, Orkiszewski RS, Norris SJ. Characterization and serologic analysis of the Treponema pallidum proteome. Infect Immun. 2010;78(6):2631–43.
16. Osbak KK, Houston S, Lithgow K V, Meehan CJ, Strouhal M, Šmajs D, et al. Characterizing the Syphilis-Causing Treponema pallidum ssp. pallidum Proteome Using Complementary Mass Spectrometry. PLoS Negl Trop Dis. 2016;10(9):e0004988.
17. Smajs D, McKevitt M, Howell JK, Norris SJ, Cai W-W, Palzkill T, et al. Transcriptome of Treponema pallidum: gene expression profile during experimental rabbit infection. J Bacteriol. 2005;187(5):1866–74.
18. Salazar JC, Rathi A, Michael NL, Radolf JD, Jagodzinski LL. Assessment of the kinetics of Treponema pallidum dissemination into blood and tissues in experimental syphilis by real-time quantitative PCR. Infect Immun. 2007;75(6):2954–8.
19. Vaira D, Malfertheiner P, Mégraud F, Axon AT, Deltenre M, Hirschl AM, et al. Diagnosis of Helicobacter pylori infection with a new non-invasive antigen-based assay. HpSA European study group. Lancet. 1999;354(9172):30–3.
20. Jarvis JN, Percival A, Bauman S, Pelfrey J, Meintjes G, Williams GN, et al. Evaluation of a Novel Point-of-Care Cryptococcal Antigen Test on Serum, Plasma, and Urine From Patients With HIV-

Associated Cryptococcal Meningitis. Clin Infect Dis. 2011;53(10):1019–23.
21. Parisi MT, Tierno PM. Evaluation of new rapid commercial enzyme immunoassay for detection of Cryptosporidium oocysts in untreated stool specimens. J Clin Microbiol. 1995;33(7):1963–5.
22. Haque R, Ali IK, Akther S, Petri WA. Comparison of PCR, isoenzyme analysis, and antigen detection for diagnosis of Entamoeba histolytica infection. J Clin Microbiol. 1998;36(2):449–52.
23. Cross RW, Boisen ML, Millett MM, Nelson DS, Oottamasathien D, Hartnett JN, et al. Analytical Validation of the ReEBOV Antigen Rapid Test for Point-of-Care Diagnosis of Ebola Virus Infection. J Infect Dis. 2016;214(suppl 3):S210–7.
24. Flores LL, Steingart KR, Dendukuri N, Schiller I, Minion J, Pai M, et al. Systematic Review and Meta-Analysis of Antigen Detection Tests for the Diagnosis of Tuberculosis. Clin Vaccine Immunol. 2011;18(10):1616–27.
25. Tramont, E; Mandell, GL; Bennett, JE; Dolin R. Syphilis. In: Principles and Practice of Infectious Diseases. 8th ed. Churchill Livingstone Inc.; 2015.
26. Sabbagh B, Mindt S, Neumaier M, Findeisen P. Clinical applications of MS-based protein quantification. PROTEOMICS - Clin Appl. 2016;10(4):323–45.
27. Cheng K, Chui H, Domish L, Hernandez D, Wang G. Recent development of mass spectrometry and proteomics applications in identification and typing of bacteria. PROTEOMICS - Clin Appl. 2016;10(4):346–57.
28. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A. 2003;100(12):6940–5.
29. Kettenbach AN, Rush J, Gerber SA. Absolute quantification of protein and post-translational modification abundance with stable isotope–labeled synthetic peptides. Nat Protoc. 2011;6(2):175–86.
30. Mermelekas G, Vlahou A, Zoidakis J. SRM/MRM targeted proteomics as a tool for biomarker validation and absolute quantification in human urine. Expert Rev Mol Diagn. 2015;15(11):1441–54.
31. Pernemalm M, Lehtiö J. Mass spectrometry-based plasma proteomics: state of the art and future outlook. Expert Rev Proteomics. 2014;11(4):431–48.
32. Domanski D, Percy AJ, Yang J, Chambers AG, Hill JS, Freue GVC, et al. MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. Proteomics. 2012;12(8):1222–43.
33. Percy AJ, Chambers AG, Yang J, Borchers CH. Multiplexed MRM-based quantitation of candidate cancer biomarker proteins in undepleted and non-enriched human plasma. Proteomics. 2013;13(14):2202–15.
34. Kuzyk MA, Smith D, Yang J, Cross TJ, Jackson AM, Hardie DB, et al. Multiple Reaction Monitoring-based, Multiplexed, Absolute Quantitation of 45 Proteins in Human Plasma. Mol Cell Proteomics. 2009;8(8):1860–77.
35. Percy AJ, Yang J, Hardie DB, Chambers AG, Tamura-Wells J, Borchers CH. Precise quantitation of 136 urinary proteins by LC/MRM-MS using stable isotope labeled peptides as internal standards for biomarker discovery and/or verification studies. Methods. 2015;81:24–33.
36. Abeijon C, Kashino SS, Silva FO, Costa DL, Fujiwara RT, Costa CHN, et al. Identification and diagnostic utility of Leishmania infantum proteins found in urine samples from patients with visceral leishmaniasis. Clin Vaccine Immunol. 2012;19(6):935–43.
37. Abeijon C, Campos-Neto A. Potential Non-invasive Urine-Based Antigen (Protein) Detection Assay to Diagnose Active Visceral Leishmaniasis. PLoS Negl Trop Dis. 2013;7(5):5–8.
38. Kruh-Garcia NA, Wolfe LM, Chaisson LH, Worodria WO, Nahid P, Schorey JS, et al. Detection of Mycobacterium tuberculosis Peptides in the Exosomes of Patients with Active and Latent M. tuberculosis Infection Using MRM-MS. Koomen JM, editor. PLoS One. 2014;9(7):e103811.
39. Young BL, Mlamla Z, Gqamana PP, Smit S, Roberts T, Peter J, et al. The identification of tuberculosis biomarkers in human urine samples. Eur Respir J. 2014;43(6):1719–29.
40. Kashino SS, Pollock N, Napolitano DR, Rodrigues V, Campos-Neto A, Campos-Neto A. Identification and characterization of Mycobacterium tuberculosis antigens in urine of patients with active pulmonary tuberculosis: An innovative and alternative approach of antigen discovery of useful microbial molecules. Clin Exp Immunol. 2008;153(1):56–62.
41. Kim SH, Lee N-E, Lee JSJY, Shin JH, Lee JSJY, Ko J-H, et al. Identification of Mycobacterial Antigens in Human Urine by Use of Immunoglobulin G Isolated from Sera of Patients with Active Pulmonary

Tuberculosis. Land GA, editor. J Clin Microbiol. 2016;54(6):1631–7.
42. Workowski KA, Berman S, Centers for Disease Control and Prevention (CDC). Sexually transmitted diseases treatment guidelines, 2010. MMWR Recomm reports Morb Mortal Wkly report Recomm reports. 2010;59(RR-12):1–110.
43. French P, Gomberg M, Janier M, Schmidt B, van Voorst Vader P, Young H. IUSTI: 2008 European Guidelines on the Management of Syphilis. Int J STD AIDS. 2009;20(5):300–9.
44. Liu H, Rodes B, Chen C-Y, Steiner B. New Tests for Syphilis: Rational Design of a PCR Method for Detection of Treponema pallidum in Clinical Specimens Using Unique Regions of the DNA Polymerase I Gene. J Clin Microbiol. 2001;39(5):1941–6.
45. Flasarová M, Pospíšilová P, Mikalová L, Vališová Z, Dastychová E, Strnadel R, et al. Sequencing-based molecular typing of treponema pallidum strains in the Czech Republic: all identified genotypes are related to the sequence of the SS14 strain. Acta Derm Venereol. 2012;92(6):669–74.
46. Tipple C, Jones R, McClure M, Taylor G. Rapid Treponema pallidum Clearance from Blood and Ulcer Samples following Single Dose Benzathine Penicillin Treatment of Early Syphilis. Vinetz JM, editor. PLoS Negl Trop Dis. 2015;9(2):1–10.
47. Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Yi J, Schuchard MD, et al. HUPO Plasma Proteome Project specimen collection and handling: Towards the standardization of parameters for plasma proteome samples. Proteomics. 2005;5(13):3262–77.
48. Aguilar-Mahecha A, Kuzyk MA, Domanski D, Borchers CH, Basik M. The effect of pre-analytical variability on the measurement of MRM-MS-based mid- to high-abundance plasma protein biomarkers and a panel of cytokines. Krauss-Etschmann S, editor. PLoS One. 2012;7(6):e38290.
49. Schuchard MD, Mehigh RJ, Cockrill SL, Lipscomb GT, Stephan JD, Wildsmith J, et al. Artifactual isoform profile modification following treatment of human plasma or serum with protease inhibitor, monitored by 2-dimensional electrophoresis and mass spectrometry. Biotechniques. 2005;39(2):239–47.
50. Human Proteome Organizaton; Human Kidney and Urine Proteome Project. Standard Protocol for Urine Collection and Storage [Internet].
51. Fusaro VA, Mani DR, Mesirov JP, Carr SA. Prediction of high-responding peptides for targeted protein assays by mass spectrometry. Nat Biotechnol. 2009;27(2):190–8.
52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
53. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, et al. The PeptideAtlas project. Nucleic Acids Res. 2006;34(90001):D655–8.
54. Lukehart S a, Marra CM. Isolation and laboratory maintenance of Treponema pallidum. Curr Protoc Microbiol. 2007;Chapter 12(November):Unit 12A.1.
55. Hanff PA, Norris SJ, Lovett MA, Miller JN. Purification of Treponema pallidum, Nichols strain, by Percoll density gradient centrifugation. Sex Transm Dis. 11(4):275–86.
56. Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, Dodson R, et al. Complete Genome Sequence of Treponema pallidum, the Syphilis Spirochete. Science (80-). 1998;281(5375):375–88.
57. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem. 2002;74(20):5383–92.
58. Desrosiers DC, Anand A, Luthra A, Dunham-Ems SM, LeDoyt M, Cummings M a D, et al. TP0326, a Treponema pallidum β-barrel assembly machinery A (BamA) orthologue and rare outer membrane protein. Mol Microbiol. 2011;80(6):1496–515.
59. Cameron CE, Lukehart S a, Castro C, Molini B, Godornes C, Van Voorhis WC. Opsonic potential, protective capacity, and sequence conservation of the Treponema pallidum subspecies pallidum Tp92. J Infect Dis. 2000;181(4):1401–13.
60. Luthra A, Anand A, Hawley KL, LeDoyt M, La Vake CJ, Caimano MJ, et al. A Homology Model Reveals Novel Structural Features and an Immunodominant Surface Loop/Opsonic Target in the Treponema pallidum BamA Ortholog TP_0326. J Bacteriol. 2015;197(11):1906–20.
61. Gayet-Ageron A, Laurent F, Schrenzel J, Charton B, Jimenez-Getaz G, Tangomo M, et al. Performance of the 47-kilodalton membrane protein versus DNA polymerase I genes for detection of Treponema pallidum by PCR in ulcers. J Clin Microbiol. 2015;53(3):976–80.
62. Centurion-Lara A, Giacani L, Godornes C, Molini BJ, Brinck Reid T, Lukehart SA. Fine analysis of

genetic diversity of the tpr gene family among treponemal species, subspecies and strains. PLoS Negl Trop Dis. 2013;7(5):e2222.
63. Zhou L, Gong R, Lu X, Zhang Y, Tang J. Development of a Multiplex Real-Time PCR Assay for the Detection of Treponema pallidum, HCV, HIV-1, and HBV. Jpn J Infect Dis. 2015;68(6):481–7.
64. Cruz ARA, Pillay A, Zuluaga A V., Ramirez LG, Duque JE, Aristizabal GEG, et al. Secondary Syphilis in Cali, Colombia: New Concepts in Disease Pathogenesis. Lukehart S, editor. PLoS Negl Trop Dis. 2010;4(5):e690.
65. Pinto M, Antelo M, Ferreira R, Azevedo J, Santo I, Borrego MJ, et al. A retrospective cross-sectional quantitative molecular approach in biological samples from patients with syphilis. Microb Pathog. 2017;104:296–302.
66. Lawn SD. Point-of-care detection of lipoarabinomannan (LAM) in urine for diagnosis of HIV-associated tuberculosis: a state of the art review. BMC Infect Dis. 2012;12(1):103–103.
67. Wu C, Duan J, Liu T, Smith RD, Qian W-J. Contributions of immunoaffinity chromatography to deep proteome profiling of human biofluids. J Chromatogr B. 2016;1021:57–68.
68. Percy AJ, Simon R, Chambers AG, Borchers CH. Enhanced sensitivity and multiplexing with 2D LC/MRM-MS and labeled standards for deeper and more comprehensive protein quantitation. J Proteomics. 2014;106:113–24.
69. Shi T, Su D, Liu T, Tang K, Camp DG, Qian W-JJ, et al. Advancing the sensitivity of selected reaction monitoring-based targeted quantitative proteomics. Proteomics. 2012;12(8):1074–92.
70. Keshishian H, Addona T, Burgess M, Kuhn E, Carr SA. Quantitative, Multiplexed Assays for Low Abundance Proteins in Plasma by Targeted Mass Spectrometry and Stable Isotope Dilution. Mol Cell Proteomics. 2007;6(12):2212–29.
71. Yadav AK, Bhardwaj G, Basak T, Kumar D, Ahmad S, Priyadarshini R, et al. A systematic analysis of eluted fraction of plasma post immunoaffinity depletion: implications in biomarker discovery. PLoS One. 2011;6(9):e24442.
72. Cameron CE. Identification of a Treponema pallidum laminin-binding protein. Infect Immun. 2003;71(5):2525–33.
73. Cameron CE, Brown EL, Janelle MY, Schnapp LM, Brouwer NL, Kuroiwa JMY. Treponema pallidum Fibronectin-Binding Proteins Treponema pallidum Fibronectin-Binding Proteins. 2004;186(20):7019–22.
74. Brinkman MB, McGill MA, Pettersson J, Rogers A, Matejkova P, Smajs D, et al. A novel Treponema pallidum antigen, TP0136, is an outer membrane protein that binds human fibronectin. Infect Immun. 2008;76(5):1848–57.
75. Penn CW, Cockayne A, Bailey MJ. The Outer Membrane of Treponema pallidum: Biological Significance and Biochemical Properties. Microbiology. 1985;131(9):2349–57.
76. Krastins B, Prakash A, Sarracino DA, Nedelkov D, Niederkofler EE, Kiernan UA, et al. Rapid development of sensitive, high-throughput, quantitative and highly selective mass spectrometric targeted immunoassays for clinically important proteins in human plasma and serum. Clin Biochem. 2013;46(6):399–410.
77. Nelson RW, Krone JR, Bieber AL, Williams P. Mass Spectrometric Immunoassay. Anal Chem. 1995;67(7):1153–8.
78. Madian AG, Rochelle NS, Regnier FE. Mass-Linked Immuno-Selective Assays in Targeted Proteomics. Anal Chem. 2013;85(2):737–48.
79. Brinkman MB, Mckevitt M, McLoughlin M, Perez C, Howell J, Weinstock GM, et al. Reactivity of Antibodies from Syphilis Patients to a Protein Array Representing the Treponema pallidum Proteome. J Clin Microbiol. 2006;44(3):888–91.
80. Salazar JC, Hazlett KRO, Radolf JD. The immune response to infection with Treponema pallidum, the stealth pathogen. Microbes Infect. 2002;4(11):1133–40.
81. Lafond RE, Lukehart S a. Biological Basis for Syphilis Biological Basis for Syphilis. 2006;19(1):29–49.
82. Anderson NLG, Anderson NLG, Haines LR, Hardie DB, Olafson RW, Pearson TW. Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). J Proteome Res. 3(2):235–44.
83. Whiteaker JR, Zhao L, Anderson L, Paulovich AG. An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass

spectrometry-based quantification of protein biomarkers. Mol Cell Proteomics. 2010;9(1):184–96.

84. Giacani L, Lukehart S, Centurion-Lara A. Length of guanosine homopolymeric repeats modulates promoter activity of subfamily II tpr genes of Treponema pallidum ssp. pallidum. FEMS Immunol Med Microbiol. 2007;51(2):289–301.
85. Giacani L, Molini BJ, Kim EY, Godornes BC, Leader BT, Tantalo LC, et al. Antigenic variation in Treponema pallidum: TprK sequence diversity accumulates in response to immune pressure during experimental syphilis. J Immunol. 2010;184(7):3822–9.
86. Molina H, Horn DM, Tang N, Mathivanan S, Pandey A. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. Proc Natl Acad Sci U S A. 2007;104(7):2199–204.
87. Wyss C. Flagellins, but not endoflagellar sheath proteins, of Treponema pallidum and of pathogen-related oral spirochetes are glycosylated. Infect Immun. 1998;66(12):5751–4.
88. Eshghi A, Pinne M, Haake DA, Zuerner RL, Frank A, Cameron CE. Methylation and in vivo expression of the surfaceexposed Leptospira interrogans outer-membrane protein OmpL32. Microbiology. 2012;158(3):622–35.
89. Witchell TD, Eshghi A, Nally JE, Hof R, Boulanger MJ, Wunder EA, et al. Post-translational Modification of LipL32 during Leptospira interrogans Infection. Small PLC, editor. PLoS Negl Trop Dis. 2014;8(10):e3280.
90. Sell S, Salman J, Norris SJ. Reinfection of chancre-immune rabbits with Treponema pallidum. I. Light and immunofluorescence studies. Am J Pathol. 1985;118(2):248–55.
91. Percy AJ, Chambers AG, Smith DS, Borchers CH. Standardized Protocols for Quality Control of MRM-based Plasma Proteomic Workflows. J Proteome Res. 2013;12(1):222–33.
92. Eyford BA, Ahmad R, Enyaru JC, Carr SA, Pearson TW. Identification of Trypanosome Proteins in Plasma from African Sleeping Sickness Patients Infected with T. b. rhodesiense. PLoS One. 2013;8(8).
93. Anjo SI, Santa C, Manadas B. SWATH-MS as a tool for biomarker discovery - from basic research to clinical applications. Proteomics. 2017;1600278.
94. Nigjeh EN, Chen R, Brand RE, Petersen GM, Chari ST, von Haller PD, et al. Quantitative Proteomics Based on Optimized Data-Independent Acquisition in Plasma Analysis. J Proteome Res. 2017;16(2):665–76.
95. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics. 2012;11(6):O111.016717.

Chapter 6

Evaluation of an automated quantitative latex immunoturbidimetric non-treponemal assay for diagnosis and follow-up of syphilis: a prospective cohort study

Authors: Kara K. Osbak[1]; Said Abdellati[1]; Achillaes Tsoumanis[1], Marjan Van Esbroeck[1], Luc Kestens[2,3], Tania Crucitti*[1], Chris R. Kenyon*[1,4]
**these authors contributed equally to this work*

Affiliations:

1. Institute of Tropical Medicine, Department of Clinical Sciences,
2. Institute of Tropical Medicine, Department of Biomedical Sciences, Antwerp, Belgium
3. University of Antwerp, Department of Biomedical Sciences, Antwerp, Belgium
4. University of Cape Town, Cape Town, South Africa

I contributed to the sample collection (cohort study), data analysis, interpretation and manuscript preparation.

Published in Journal of Medical Microbiology, July 2017
doi: 10.1099/jmm.0.000559

Conflicts of Interest and Sources of Funding
This work was part of Project ID: 757003 funded by the Flemish Government-Department of Economy, Science & Innovation granted to CRK. The authors declare no conflicts of interest. The sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results. Sekisui Diagnostics lent the SK500 Clinical Chemistry System to the ITM and provided the Sekure RPR kits.

Abstract

Purpose

The Sekure Rapid Plasma Reagin (RPR-S) (Sekisui Diagnostics) automated quantitative latex immunoturbidimetric assay performed on the SK500 Clinical Chemistry Platform was evaluated for clinical appropriateness.

Methodology

Syphilis-infected individuals and controls were recruited into a prospective cohort study conducted at a Sexually Transmitted Infection clinic in Antwerp, Belgium. Sera collected at diagnosis (baseline) and at 3, 6, 9 and 12 months post-treatment were tested with RPR-S and Macro-Vue RPR card (RPR-C) (Becton Dickinson) assays; RPR-C was considered the reference test. IgG/IgM enzyme immunoassay and *Treponema pallidum polA* serum PCR results were consulted by discordancy at baseline. Categorical analyses were performed and correlations were assessed with (non)-linear regression. Post-treatment longitudinal serological evolution was evaluated.

Results

A total of 463 samples from 120 new syphilis cases from a variety of clinical stages and 30 syphilis-negative controls were tested. Initially there was a weak correlation between quantitative RPR-C/S (R = 0.15). In 70 samples there was a strong suspicion of hook effect. Of these, 57/70 sera were retested with an extra dilution step, resulting in an average 12-fold increase in quantitative RPR-S results. After the extra dilution, the overall qualitative RPR-C/S agreement was 78.89 %, (κ–coefficient 0.484). Of the 92 discordant samples, nine were from the baseline visit (RPR-C: 1-8), which could have led to possible missed diagnoses using the RPR-S.

Conclusions

Sensitivity and accuracy of the RPR-S test requires improvement before it can be used to diagnose syphilis and evaluate treatment efficacy in clinical practice.

Introduction

Syphilis, a multi-stage chronic disease transmitted sexually or vertically through the placenta, remains a substantial public health burden with more than 8 million new infections per year worldwide [1]. The past 15 years has witnessed a reemergence of syphilis in Belgium. An increasing proportion of cases are reinfections [2,3] which more frequently present asymptomatically [3], making the diagnosis dependent on serological test results.

Serological tests to diagnose syphilis fall into two main categories: 1) treponemal tests (TTs) that measure antibodies directed against *Treponema pallidum* ssp. *pallidum* (henceforth referred to as *T. pallidum*), the etiological agent of syphilis, using recombinant or wild-type antigen–antibody reactions, and 2) non-treponemal (NTT) tests [4] which measure nonspecific antibodies directed against cardiolipin released from damaged human cells and possibly present in *T. pallidum's* cell wall [5,6]. There are two main diagnostic algorithms for syphilis screening: a 'traditional algorithm' that screens with a NTT and confirms with a TT [7] and a 'reverse algorithm' that uses the TTs for screening and the quantitative NTT [4] or a second TT assay (followed by a NTT assay to measure infection activity) for confirmation. The latter has been promoted in low-prevalence settings with the advent of automated low-cost *T. pallidum* chemiluminescent/luminescent immunoassays [4]. Only NTTs, such as the Rapid Plasma Reagin Card test, (henceforth referred to as RPR-C) or the Venereal Disease Research Laboratory (VDRL) test, can be used for post-treatment follow-up to determine *T. pallidum* eradication and for the diagnosis of reinfection [8]. A four-fold decrease in RPR-C titer, considered as a clinically significant difference between two NTT results, is necessary to qualify a response as adequate six to twelve months post-treatment. A recent systematic review estimated that the median proportion of adequately treated patients classified as 'serofast' or 'non-responders' was 20.5 % at 6-month and 11.2 % at 12-months post-treatment, for all disease stages [9]. To diagnose a syphilis reinfection, a four-fold NTT titre increase, using the same test, should be demonstrated. However, increases in RPR-C titres can be secondary to a large variety of (auto)-inflammatory conditions such as HIV and hepatitis [8,10–12], leading to false positive diagnoses. Inaccurately low or false negative RPR assay results have been attributed to the hook or prozone effect during syphilis testing [13–15], an undesirable phenomenon observed in immunoassays when an unbalanced high antibody to antigen ratio interferes with the antigen-antibody reaction, which is typically alleviated by sample dilution.

In most laboratories, NTTs are performed manually and are thus subject to variation in inter-individual interpretation and reagent batch variation [16,17]. Automation of NTTs could contribute in various ways to improve syphilis diagnosis, for example, by increasing test accuracy with the use of an internal standardized calibrant, higher throughput capability and reporting of more accurate continuous values instead of titre values. In Korean and Japanese laboratories, automated quantitative latex turbidimetric immunoassays (LTIA) for syphilis diagnostics have been widely implemented [18]. However, large-scale comparative studies, in particular in high sexually transmitted infection (STI) prevalence settings, are lacking. Currently, two automated RPR tests are commercially available, the Sekure RPR LTIA (henceforth referred to as RPR-S), also marketed as the 'Mediace RPR assay' (Sekisui Chemical Co. Ltd., Osaka, Japan) and the HISens Auto RPR LTIA (HBI Co., Ltd., Korea). The RPR-S test utilizes latex particles coated with purified cardiolipin-lecithin antigens extracted from bovine heart and measures agglutination secondary to antibody binding. In previous studies, the Mediace RPR assay performed on other analytical platforms was shown to have a high sensitivity (100 %) for early stage disease [19], however, low qualitative correlation (83.8 %) with the traditional RPR-C test, low sensitivity for latent stage syphilis (55.6 %) [19,20] and problems with prozone effect at high RPR-C titre values [20] have been reported.

In this study, we conducted a clinical laboratory evaluation to investigate the clinical appropriateness of the Sekure RPR assay for syphilis diagnosis and post-treatment serological follow-up. The SK500 Clinical Chemistry Platform was used to perform RPR-S testing on sera from 120 syphilis-infected patients and 30 controls collected during the course of a prospective cohort study.

Materials and Methods

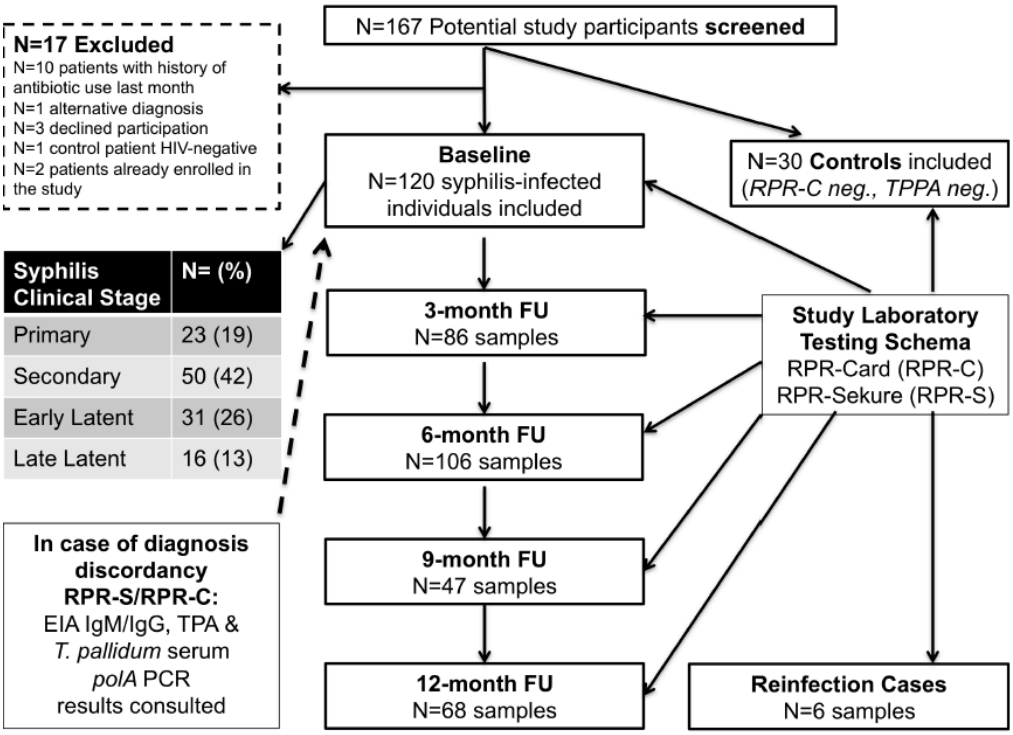
Study participants

This study was conducted as a sub-study of the "*Treponema pallidum*-specific Proteomic Changes in Patients With Incident Syphilis Infection (SeTPAT)" study (ClinicalTrials.gov Registration #: NCT02059525) investigating novel biomarkers for syphilis-infection. Potentially eligible study participants over the age of 17 in whom a new syphilis diagnosis was made were screened and recruited between January 2014 and August 2015 at the STI clinic of the ITM in Antwerp, Belgium. Additionally, HIV-infected controls with negative syphilis serological tests (RPR-C and *T. pallidum* particle agglutination test (TPPA) negative) were included during the same time as the study

recruitment. Exclusion criteria were the use of beta-lactam, doxycycline or macrolide antibiotics during the 28 days preceding enrollment. Syphilis diagnosis and disease staging was done according to the Centers for Disease Control guidelines [21]. Stage-appropriate treatment was administered according to European guidelines [22] in the form of intramuscular injection with Benzathine Penicillin G, or oral doxycycline in the case of penicillin allergy or penicillin unavailability. All syphilis-infected participants were followed-up at 3, 6, 9 and 12 months post-treatment and in the cases of suspicion of syphilis reinfection. The study physician systematically recorded clinical and laboratory details in a standardized fashion during each study visit.

After screening 167 potential study participants, a total of 150 individuals were included into the study, 120 with an active episode of syphilis (RPR-C and TPPA positive) and 30 controls (*Fig 1*/ *Table 1*), with the exception of one case of secondary syphilis with a negative RPR-C but positive TPPA and serum *T. pallidum* PCR at baseline. The median age of the syphilis-infected participants was 40 years (IQR 31.5-48) and of the controls 37.5 years (IQR 32-45). Only one syphilis-infected participant was a woman. The diagnosed syphilis stages were: N=23 primary, N=50 secondary, N=31 early latent and N=16 late latent stage disease. Of the syphilis-infected individuals, 103/120 (85.8 %) were HIV-infected, of which 91/103 (88.4 %) were taking antiretroviral therapy. All except one self-reported being men who have sex with men (MSM), 43 % reported having more than 10 sexual partners during the last year and almost two-thirds had a previous history of syphilis infection, indicating this is a generally high-risk population. The median RPR-C titer at the time of syphilis diagnosis was 64 (IQR 32- 128). During follow-up, participants provided a total of 86 3-month samples, 108 6-month samples, 47 9-month samples and 68 12-month samples. Six 'reinfection' samples were provided for analysis (*Fig 1*). These were collected in the case of syphilis reinfection between regular study visits, or after the 12-month sub-study period.

Fig 1 *Prospective cohort sub-study overview and laboratory testing work-flow description. A total of 150 study participants, 120 with a new syphilis diagnosis and 30 syphilis negative controls were recruited at the ITM STI clinic in Antwerp, Belgium. Legend: FU= follow-up*



Syphilis Clinical Stage	N= (%)
Primary	23 (19)
Secondary	50 (42)
Early Latent	31 (26)
Late Latent	16 (13)

Table 1 *Clinical cohort characteristics of participants in this study*

	Syphilis-infected cases N=120	Controls N=30
Characteristics	**N=; (%/IQR)**	**N=; (%/IQR)**
Gender (male)	119 (99)	30 (100)
MSM	117 (98)	24 (80)
HIV-infected	103 (86)	30 (100)
Taking ART	91 (88)	24 (80)
Benzathine Penicillin G treatment	117 (98)	NA
RPR-C titre at baseline	64 (32-128)	0
Age	40 (31.5-48)	37.5 (32-45)
Number of Previous Syphilis Episodes		
0	48 (40)	30 (100)
1	41 (34)	0
2	17 (14)	0
3	9 (8)	0
≥4	5 (4)	0

Legend: ART: antiretroviral therapy; NA: not applicable

Routine clinical serological testing

Blood was drawn into serum gel tubes (Sarstedt Monovette, Nümbrecht, Germany) and separated at 2000 g for 10 minutes at ambient temperature. Serum samples were divided and subsequently either 1) refrigerated at 4 °C until routine clinical syphilis serological testing (as detailed below) within four days or 2) aliquoted into 2 mL cryovials and stored at -80 °C within three hours for later evaluative testing. All fresh sera were analyzed with Macro-Vue RPR Card (RPR-C) (Becton Dickinson, Sparks, MD, United States of America (USA)) and a TPA assay (Ortho-Clinical Diagnostics, Rochester, NY, USA) according to manufacturer's instructions. If the sera tested RPR-C positive, the titre was determined to an endpoint. Since these tests were performed during regular clinical routine testing, different experienced laboratory technicians performed the tests during the course of the study. All baseline fresh sera were also tested with an *in-house T. pallidum* PCR directed against *polA* [23] and a SERODIA-TPPA (Fujirebio Inc., Tokyo, Japan) assay.

Implementation of the SK500 Clinical Chemistry System for RPR-S testing

In collaboration with Sekisui Diagnostics, LLC, a SK500 analytical system was installed in the ITM STI reference laboratory. Two laboratory technicians followed a practical two-day on-site training, including trial-run analysis of non-study related reference samples. Evaluative testing of the RPR-S assay took place after the samples of interest were collected (maximum 974 days), stored at -20 °C until analysis, which is outside of the RPR-S test insert stability claim. Samples were subsequently thawed, batched analyzed and stored at 4 °C until all sub-study testing was completed. Sera testing was conducted on all available study samples from baseline, 3, 6, 9 and 12 months post-treatment visits and on samples from serologically (RPR-C) confirmed cases of reinfection following manufacturer's instructions (Version KI486616.EN.02, June 2016). Sera were thawed at ambient temperature and centrifuged at 15,000 g for 10 minutes prior to testing. For the RPR-S testing, 180 µL of Reagent 1 and 20 µL of serum were added to the assay tube, incubated for 5 minutes, then 60 µL of Reagent 2 was added and incubated for 5 minutes. Absorbance at 700 nm was measured at regular intervals up to 7.5 minutes. Results are expressed in RPR units (R.U.), whereby 1 R.U. is equivalent to a 1-fold RPR-C titer according to the manufacturer. A measurement of 1 R.U. or higher indicates that the sample is antibody positive and measurements of 2 R.U. and 4 R.U. are equivalent to the RPR-C dilutions of 1/2 and 1/4 respectively. During the first-run analysis of all samples, the analyzer was programmed to automatically subject samples with an initial R.U. score of ≥8.0 to a 10x dilution step with 0.9 % NaCl solution and retest again. In addition, the analyzer was programmed to detect the prozone phenomenon.

Intermittent statistical analysis was performed after the first-run RPR-S testing of all samples. The results showed a possible hook effect, after which the analyzer was programmed to automatically dilute (10x) samples with an initial RU score of ≥4.0 and retest.

All RPR-S testing was performed in single-run analyses by the same study technician, with exception for the second analysis of samples showing a possible hook effect, which were tested by the company technicians at the ITM. Industry assistants were available for troubleshooting consultations through mail correspondence, telephonically and visited the facility during the testing period.

Quality control

The ITM laboratories in Antwerp where these samples were analyzed are ISO15189 accredited. The RPR-S reagent kit and respective controls used in this study were from the same lot number. Before each batch analysis, the RPR-S assay was calibrated with the RPR Calibrator Set and control samples provided; control sample results were compared to the sample validation range provided by the manufacturer. The reagents used for the RPR-C, TPPA and TPA tests were from different lot numbers due to the prolonged testing period and nature of the routine clinical testing setting. All testing reagents were stored in the same conditions according to manufacturer's instructions. The reanalysis of the suspect hook effect samples were first re-examined without the extra dilution step to assess possible testing variation after prolonged (defined as longer than 2 weeks) storage at 4 °C. Details of the freeze-thaw cycle frequency was recorded for each sample.

Operator feedback

After the two-day on-site training provided by Sekisui Diagnostics technical experts, the study laboratory technicians were able to perform the test first under supervision and coaching and after one week independently without major problems. Overall the technician found the SK500 platform easy to use. However, a comprehensive understanding of the analyzer is required to run the samples independently and one must diligently focus on the test progress and reaction dynamics during each run to ensure that the testing runs smoothly and efficiently.

Statistical analysis

Henceforth, the '*first-run*' analysis refers to the initial data set with the R.U. cutoff of 8 triggering 10x dilution. The '*second-run*' analysis refers to the same data as those used in the *first-run*, except for the 57 suspected-hook-effect-samples that were re-run at a R.U. 10x dilution cutoff of 4. For the comparative categorical and continuous data analysis the results were substantively equivalent when analyzed per visit or when all visits were pooled. For the longitudinal serological evaluation of RPR-S/C results, the baseline, 6- and 12-month and reinfection time points were considered individually.

The RPR-C test was the reference test. Categorical analysis involved the construction of two-by-two contingency tables (positive/negative) for the RPR-S/C result comparisons. Percentage agreement and Cohen's kappa (κ)-coefficient value [24] were calculated to estimate agreement between both tests. The sensitivity, specificity, positive (PPV) and negative predictive values (NPV) with 95 % confidence intervals (CIs) were calculated using the baseline visit samples. Results of Enzyme Immuno Assays (EIAs) and PCR

testing were consulted in the case of discordancy between the RPR-S/C results at the baseline visit. The PCR analyses were performed in the context of the main study and the EIAs (Anti-Treponema-pallidum-ELISA IgM/IgG; Euroimmun, Lübeck, Germany; recomWell Treponema ELISA IgM/IgG, Mikrogen, Neuried, Germany) were performed in the context of another sub-study evaluating two different IgG and IgM EIAs.

Continuous variables are expressed with the median value and interquartile range (IQR). Histograms were plotted and Shapiro–Wilk testing [25] was conducted to assess data normality. (Non)-linear regression was performed after scatterplots were constructed to evaluate the relationship of the transformed results. Initial *first-run* RPR-S/C results were assessed using Pearson's correlation analysis (R). Non-normal continuous variables from the *second-run* dataset were transformed in order to achieve more linear relationships for correlation analyses; the linear agreement was assessed by Passing and Bablok regression analysis [26] within the linear range test values.

To assess the potential effects of prolonged storage at 4 °C, sample means were compared with a Wilcoxon- signed rank test [27]. Serofast and seroreversion classification frequency between RPR-S/C testing was evaluated with a chi-squared (χ^2) test [28] and κ-coefficient value calculations.

All analyses were performed in Stata 12.1 (StataCorp LP, College Station, TX, USA) and MedCalc version 17.1 (MedCalc Software, Mariakerke, Belgium). A P-value of less than 0.05 was considered statistically significant.

Serofast, seroreversion and syphilis reinfection definitions

We defined 'serofast' state [21,29] as the failure of the RPR-C titre to decline four-fold between the baseline and the 6- (6-month serofast) and 12-month visit (12-month serofast), respectively. Seroreversion was defined as RPR-C titre conversion to negative at the 6 (6-month seroreversion) or 12-month (12-month seroreversion) time points. Reinfection was defined as an episode whereby the RPR titre increased ≥four-fold after a previous syphilis episode demonstrated a serological response to therapy (≥four-fold decrease in RPR-C titre). We assumed a 1:1 relationship between the RPR-C/S values as dictated by the manufacturer's testing specifications. As a result we extrapolated the definitions mentioned above to the RPR-S testing. Thus, a four-fold change in RPR-S result was calculated by dividing (decrease) or multiplying (increase) the R.U value by four.

Results

Summary of the quantitative RPR-S serological test evaluations

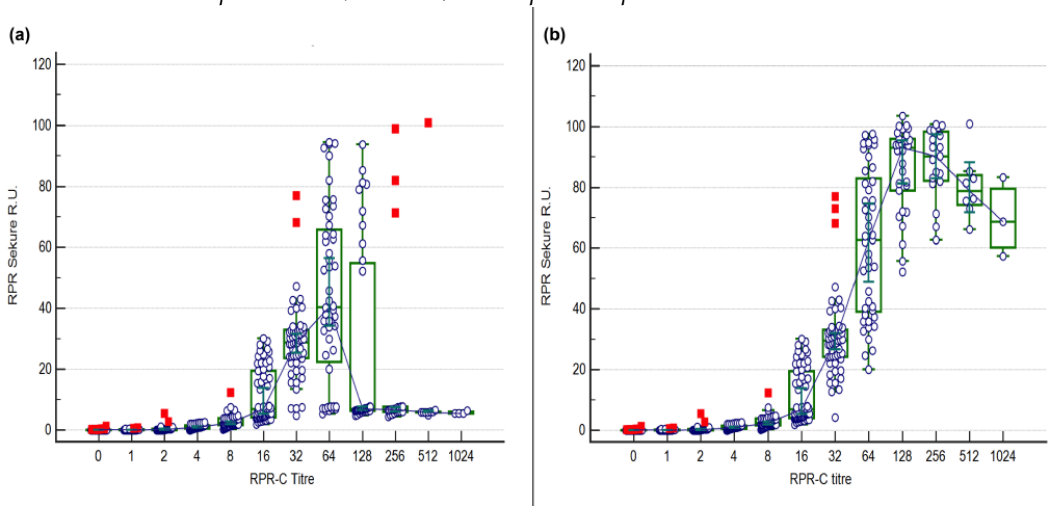
Serological testing of 463 sera with the RPR-S kit was performed over the course of a two-month period in late 2016 (*first-run* analysis). Once all RPR-S testing was complete, an intermittent quantitative statistical analysis was performed to examine test correlations with RPR-C results. The initial RPR-S/C correlation was R = 0.15, whereby 70 samples had a strong suspicion for hook effect (low RPR-S R.U./high-RPR-C titre) as identified by scatterplot analyses, corresponding to a median RPR-C titre of 128 (IQR 128-256) and RPR-S median R.U. value of 6.6 (IQR 5.9-7.2). Fifty-seven of these samples were available for retesting which was performed after a median of 40 days (IQR 27-45) after thawing and continuous storage at 4°C. The 57 samples were first reanalyzed without 10x dilution as a 'control check'; these results did not significantly differ from the original *first-run* testing (P = 0.22). After implementation of a sample dilution step (10x) at RPR cutoff 4 R.U., the RPR-S values for the 57 samples increased significantly (P < 0.001) (*Table 2; Fig 2(b)*), resulting in a median R.U. value of 90.2 (IQR 81.2-96), equaling an average 12-fold increase compared to the first run. Only one RPR-S result did not increase after retesting, this corresponded to a RPR-C of 32, RPR-S 4.7/4.3 *first/second-run* result, respectively. A total of 450 samples were included in the *second-run* dataset, thereby including 393 samples from the *first-run* dataset (no suspicion of prozone or hook effect), the 57 samples subjected to the 4 R.U. cutoff 10x dilution and excluding the 13 potential hook effect samples which were not available for retesting (*Fig S1*). All samples with an RPR-C titre of 1 (N=29) tested RPR-S negative whilst 11.6 % (5/43), 48.8 % (21/43), and 98.4 % (42/47) of RRP-C titre 2, 4 and 8, respectively, tested RPS-S positive. A breakdown of the results per RPR-C titre level and schematic graphical overview are provided in (*Fig 2(a/b)/Table 2)*.

Table 2 *RPR-Sekure R.U. result breakdown per RPR-C titre for the* first- *and* second-run *analyses*

RPR-C Titre	RPR-Sekure *FIRST-RUN* ANALYSIS					RPR-Sekure *SECOND-RUN* ANALYSIS				
	Neg.	Pos.	Total	Median R.U.	IQR	Neg.	Pos.	Total	Median R.U.	IQR
0	71	1	72	0	0-0	71	1	72	0	0-0
1	29	0	29	0	0-0.2	29	0	29	0	0-0.1
2	38	5	43	0.1	0-0.4	38	5	43	0.1	0-0.4
4	22	21	43	0.9	0.5-1.5	22	21	43	0.9	0.5-1.5
8	5	42	47	2.3	1.6-3.8	5	42	47	2.3	1.6-3.8
16	0	57	57	6.9	4.2-19.4	0	57	57	6.9	4.1-19.4
32	0	55	55	28.9	23.6-32.9	0	54	54	29.7	24.2-33.1
64	0	48	48	40.5	22.4-65.8	0	45	45	62.7	39.5-81.9
128	0	35	35	6.9	6.5-55.7	0	30	30	92.9	79-96
256	0	21	21	6.8	6.1-7.6	0	19	19	90.2	81.9-98.8
512	0	9	9	5.9	5.8-6	0	8	8	78.9	74.3-84.05
1024	0	4	4	5.6	5.5-6.05	0	3	3	68.8	57.4-83.2
Total	165	298	463			165	285	450		

Legend: IQR: interquartile range; R.U.: RPR units; RPR-C: manual RPR-card testing

Fig 2 *Box and Whisker Plots depicting RPR-S R.U. results per RPR-C titre for the (a)* First *and (b) Second-run analyses. RPR-S results are shown as median and interquartile range (box), with 10th and 90th percentiles (whiskers). Red squares represent outliers.*



(Non-linear) correlation analysis RPR-S performance for second-run analysis

There was a non-normal data distribution (RPR-S: W = 0.77, P< 0.001; RPR-C: W = 0.45, P < 0.001), thus the RPR-S *second-run* and RPR-C values were $\log_e(x)$ transformed. *Fig S2* is a graphical depiction of the sigmoidal relationship that was fitted with the regression equation. When considering the RPR ranges 4-64 (N=246 samples) for the Passing Bablok linear regression analysis, the Intercept A was -2.27 (95 % CI: -2.54- 2.04), representing the systemic differences, and the proportional difference (Slope B) was 1.60 (95 % CI: 1.52- 1.70), yielding the equation "y=-2.27+1.60x" (*Fig S2 & S3*). The RPR-S/C linear correlation range was estimated to be between RPR-C titres 8-32, corresponding to a median RPR-S 2.3-29.7 (IQR 1.6-32.9) R.U..

Categorical assessment of the RPR-S test

With respect to the RPR-S *first* and *second-run analyses*, in both runs 94/463 and 94/450 samples were classified as false negative when compared with RPR-C testing, respectively (*Table 3*). There was a 78.48 % agreement for the *first-run* and a 78.89 % agreement for the *second-run*, together with a moderate [30] κ-coefficient of 0.488 and 0.484, respectively. The quantitative agreement ±1 RPR-C dilution (margin of error) corresponding to a ±2-fold RPR-S range was 42.98 % for the *first-run* and 50.44 % for the *second-run* analyses. When considering the RPR-C titre breakdown, the false negative RPR-S results occurred in the low RPR-C titre ranges 1-8 (*Table 2*) for both runs. Only one RPR-S sample was false positive, with a weakly positive value of 1.3. The overall

sensitivity was 93.3 % for the *first-run* (baseline visit; N=150) and 92.5 % for the *second-run* (baseline visit; N=138) and the specificity was 100 % for both runs (*Table 4*).

Table 3 *Cross-tabulations of the RPR-S* (first *and* second-run) *by the results of the RPR-C testing*

		RPR-C		
		Positive	**Negative**	**Total**
First-run Analysis ALL				
RPR-S	Positive	297	1[#]	298
	Negative	94[$]	71	165
	Total	391	72	463
Second-run Analysis ALL				
RPR-S	Positive	284	1[#]	285
	Negative	94[$]	71	165
	Total	378	72	450
First-run Analysis Baseline				
RPR-S	Positive	111	0	111
	Negative	8	31	39
	Total	119	31	150
Second-run Analysis Baseline				
RPR-S	Positive	99	0	99
	Negative	8	31	39
	Total	107	31	138

Legend: #- RPR-S value was weakly positive 1.3; $- RPR-S tested negative values corresponded to RPR-C titre ranges 1-8 (see Table 2)

Table 4 *Categorical analysis of RPR-S test versus reference test (RPR-C)*

	RPR-C *vs.* RPR-S First-run N=463 (95 % CI)	**RPR-C *vs.* RPR-S Second-run N=450 (95 % CI)**
K- coefficient	0.488 (0.409-0.568)	0.484 (0.405-0.564)
% Agreement Qualitative@	79.48 %	78.89 %
% Agreement Quantitative*	42.98 %	50.44 %
	N=150 (95 % CI)	**N=138 (95 % CI)**
Sensitivity&	93.3 % (87.2- 97.1)	92.5 % (85.8-96.7)
Specificity&	100 % (88.8-100)	100 % (88.8-100)
PPV&	100 % (96.7-100)	100 % (96.3-100)
NPV&	79.5 % (63.5-91.8)	79.5 % (63.5-90.7)

*Legend: PPV- Positive Predictive Value; NPV- Negative predictive value; CI- confidence interval; @- qualitative agreement positive/negative; *- quantitative ± two-fold RPR-C/S agreement equals ± one dilution RPR-C/ ± two-fold RPR-S R.U. result; &- calculated using baseline samples only*

Potential clinical misclassifications at baseline according to RPR-S testing

Of the 120 syphilis-infected individuals included in this study, nine would have been considered syphilis-negative according to the RPR-S testing at baseline [7]. Additional laboratory information, such as TPA, EIAs IgM/IgG and serum *polA* PCR results were consulted in order to serologically characterize these nine cases (full details are provided in *Table S1*).

All nine individuals were HIV-infected males and two individuals had a previous history of syphilis with positive RPR-C and TPPA tests. All except one had negative RPR-C results available from before the study period. Two individuals presented with primary (ulcers), one with secondary (maculopapular rash), four with early latent and two with late latent disease stages. All had positive RPR-C titres at baseline in the low range (2-4), except for one secondary syphilis case that was RPR-C negative, serum PCR positive, EIA IgM positive at baseline, without a history of syphilis. His RPR-C became positive at his 9-month visit. Eight of the nine were classified as definite new syphilis infections, based on seroconversion of RPR-C results within the prior 12 months. In addition, three were PCR positive, three exhibited convincing clinical signs of syphilis and five were IgM positive. The ninth individual classified with probable syphilis presented with recent high risk behavior, no history of syphilis symptoms or treatment and a new diagnosis of HIV, RPR-C of 2, TPA of 151, positive IgG but negative IgM and PCR tests.

Six and twelve- month post-treatment serofast and seroreversion cases according to RPR-C/S testing

Details of the 12-month RPR-S/C follow-up for the syphilis positive participants with available longitudinal data are reported in *Table S2*. Serological follow-up results at the 6-month and 12-month time points were available for 96/120 and 63/120 individuals, respectively. Samples were excluded from the longitudinal analysis if they were taken after an episode of re-infection during the course of the study (N=6) and if they had negative RPR-C (N=1) or RPR-S (N=9) result at baseline. All analyses were performed with the *second-run* analysis.

According to the RPR-S testing there were significantly more 6-month serofast cases (22.6 % *versus* 10.2 %), P = 0.001); at 12-months there was no significant difference between the two testing methods (14.8 % *versus* 5.6 %, P = 0.353). There were significantly more seroreversions at both 6-month (42.1 % *versus* 9.1 %, P < 0.001) and 12-month (46.3 % *versus* 20.4 %, P = 0.001) time points according to the RPR-S testing. There was a general moderate to low percentage agreement and minimal agreement according to the K-coefficient calculations for all time points and classifications (serofast/seroreversion) (*Table 5*).

Table 5 *Contingency tables of the 6- and 12-month serofast and seroreversion cases according to the RPR-S and RPR-C testing*

		RPR-C				
		No N=(%)	**Yes N=(%)**	**Total N=(%)**	**Agreement (%)**	**K-coefficient (95% CI)**
Serofast 6-month						
RPR-S	No	65	3	68 (77.3)	80.68 %	0.318 (0.081-0.554)
	Yes	14	6	20 (22.6)		
	Total	79 (89.8)	9 (10.2)	88		
Serofast 12-month						
RPR-S	No	44	2	46 (85.2)	83.33 %	0.110 (-0.202-0.421)
	Yes	7	1	8 (14.8)		
	Total	51 (94.4)	3 (5.6)	54		
Seroreversion 6-month						
RPR-S	No	51	0	51 (58.0)	67.05 %	0.242 (0.095-0.390)
	Yes	29	8	37 (42.1)		
	Total	80 (90.9)	8 (9.1)	88		
Seroreversion 12-month						
RPR-S	No	28	1	29 (53.7)	70.37 %	0.380 (0.167-0.593)
	Yes	15	10	25 (46.3)		
	Total	43 (79.6)	11 (20.4)	54		

Reinfections during the course of the study

A total of nine reinfection cases could be evaluated with both the RPR-C and RPR-S tests (*second-run* analysis) during the total course of the study follow-up. Seven of nine were asymptomatic and serologically confirmed (RPR-C). Only 6/9 cases would have been classified as reinfections according to the RRP-S test (*Table S2*).

Discussion

The objective of this study was to assess the clinical applicability of the Sekure RPR test performed on the SK500 Clinical Chemistry platform, specifically for the applications of syphilis diagnosis and post-treatment serological follow-up. A novel high-throughput, sensitive and precise NTT would be a welcome improvement of syphilis diagnostics. We found two predominant problems with the RPR-S test. Firstly, 'hook effect' and/or ineffectual sample dilution presumably affected samples with high antibody levels, corresponding to RPR-C titre ≥128, leading to lower than expected RPR-S results. This problem was partially rectified when samples were re-run at R.U. cut-off of 4.0 to trigger a 10x dilution step. Secondly, the RPR-S test was not sufficiently sensitive at

RPR-C titres ≤8 – missing nine cases of syphilis at the baseline visit. Consequently, we estimated the limited linear RPR-C/S correlation range to be between RPR 8-32 based on the *second-run* analysis. This lack of RPR-S test sensitivity and accuracy would have negative clinical outcome consequences, including potential misdiagnoses and misclassified treatment outcomes, such as an overestimation of serofast status and resulting antibiotic overtreatment.

Our results are concordant with a previous study investigating the Mediace RPR LTIA assay, which found a poor correlation between the RPR-S and RPR-C testing and evidence of hook effects at high RPR-C titres [20]. Sample dilution presumably (partially) corrected the hook effect, as demonstrated in the present study where 56/57 attained higher values closer to the expected range after dilution. The maximum RPR-S value in our study was 103.5 (corresponding to a RPR-C titre 128), lower than a previously reported maximum result of 384 R.U. [31], although it was not specified in the article if the samples were diluted and on which analytical platform the tests were performed. These findings could be partially due to the low maximum value of the calibration curve (8.0 R.U.), whereby 10x dilution only results in a maximum measurement 80 R.U. within the linear range. Per procedure the extra dilution step did not affect the qualitative analysis and agreement. Notably, few previous studies of the RPR-Mediace test have consistently evaluated samples with RPR-C titre values higher than 128 [20,31,32]. This oversight is important since many syphilis patients present with high RPR titres at diagnosis, as seen in this study where 44/120 (36.7 %) of individuals presented with RPR titres ≥128 at baseline.

A limited RPR-S test sensitivity was observed at the lower antibody level range, corresponding to an RPR titre ≤8. Overall, the RPR-S had an overall good sensitivity of 93.3 % and specificity of 100 %, however, these values must be interpreted with caution since this cohort biomarker discovery study was not designed to adequately address the performance of the test as a population screening tool. The attained sensitivity is higher than a previous study of 101 serum samples, reporting a RPR-S sensitivity of 60.5 % compared to the RPR-C test [33]. Further, if the 'low titre' false negative results from the other study visits would be taken into account and extrapolated to theoretical disease events, the clinical consequences would be important.

Regarding clinical diagnosis discrepancies between RPR-S/C testing, 9/120 (7.5 %) syphilis diagnoses would have been missed at baseline and 3/9 (33 %) cases of reinfection would have been missed if only the RPR-S results were taken into account.

It is important to note that these reinfection cases could have been misclassified, since syphilitic reactivation due to subcurative treatment and treponemal antibiotic resistance [34,35] can also cause significant increases in NTT results, although clinically significant high-level resistance has only been reported for macrolides. Specifically, 2/9 (22 %) of these missed cases were from primary stage whereby antibody levels are often low or insufficient to be detected in the serological tests, as reflected by the low TPA/TPPA results for these cases and one case was deemed 'probable' since we cannot rule out that the syphilis infection could have been indirectly treated by antibiotics long before diagnosis.

The appropriateness of RPR-S testing for post-treatment syphilis follow-up is a concern due to the aforementioned suboptimal performance at high antibody titres and low quantitative agreement between the tests (43 % *first-run* and 50 % *second-run*), therefore it may not be possible to deduce a correct delta RPR change after treatment. This is reflected in the significantly higher amount of serofast 6-month cases according to the RPR-S testing, which when translated into clinical practice could result in unnecessary antibiotic treatment. Nevertheless, pending RPR-S optimization, the definition of successful treatment (serofast/non-responders) could change with the introduction of a test providing continuous values instead of titres.

There are a number of caveats related to biological, technical and analytical biases, which may have influenced the results, thereby limiting generalizability. For example, we did not control for the possible effects of HIV and the participants were almost all MSM. The RPR-C test was used as 'reference test', however, it is by no means a perfect test [16,17] and there is a possibility that some RPR-C results could be interpreted incorrectly beyond the error margin of ± 1 RPR-C dilution. Some samples were subjected to suboptimal testing conditions, namely storage at 4 °C for longer than two weeks. Statistical testing however revealed that this did not produce a significant difference in the RPR-S results. RPR-C testing was performed on fresh sera as opposed to thawed samples for the batched RPR-S testing; ideally the RPR S and C testing would have been run in parallel to show the effect of the extended frozen storage period on the samples. Advanced statistical modeling was not performed to assess the possible role of biological bias since samples originated from multiple testing time points from the same individuals. Finally, despite efforts to transform the data, we could not achieve a sufficiently linear distribution (all RPR-C ranges) to warrant a Bland Altman analysis, which is typically recommended for the comparison of two methods measuring the concentration of an analyte [36].

In this study, we implemented an automatic dilution of samples with an initial R.U. value of 8 as set by the company. Dilution at a cut-off of 4 R.U. would likely alleviate the 'hook effects' since the *first-run* analysis samples with suspected hook effect had median initial value of 6.6 (IQR 5.9-7.2). Moreover, further dilution steps (higher than 10x) are likely warranted and adjusting the calibration curve to measure higher than 8.0 R.U and quantitate higher than 4.0 R.U. could improve test performance. Further studies could also investigate if lowering the RPR-S test cutoff level to less than 1.0 R.U. which is closer to the limit of detection of 0.2 R.U. would increase sensitivity, albeit this may increase the amount of false-positive results secondary to noise or cross-reactions at lower cut-off levels. For example, a study by Noh *et al.* implemented a 0.5 R.U. cut-off level and 45.4 % of positive results were rated as false positive compared to RPR-C/TPLA/EIA testing [19]. This is the first study investigating the testing efficacy of the RPR-S reagent on the SK500 Clinical Chemistry platform. Further optimization of the platform settings may improve test performance.

Conclusions

This study exemplifies how an industry independent evaluation of a diagnostic test can be beneficial to improving laboratory technology through independent testing of well-characterized clinical samples. Although we identified a number of important problems with the RPR-S test, our improved results with a dilution step at a lower cut-off value suggests a remedy to the suboptimal performance. Further test optimization is required by the manufacturer.

Acknowledgements

We would like to thank the ITM laboratory teams for their involvement in the study and the authors gratefully acknowledge the individuals who participated in this study.

Supplementary Materials

Table S1 (available online). Clinical misclassification discrepancies according to RPR-Sekure and RPR-Card testing at baseline

Table S2 (available online). Longitudinal follow-up of syphilis-infected patients including serofast and seroreversion classifications and reinfection cases

Fig S1 *Flow-chart for* first- *and* second-run *RPR-S analyses*

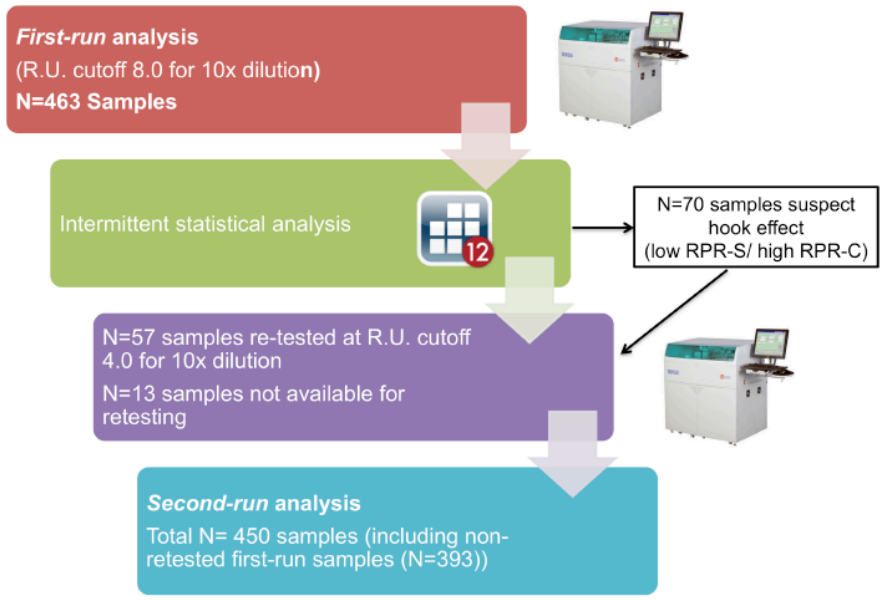
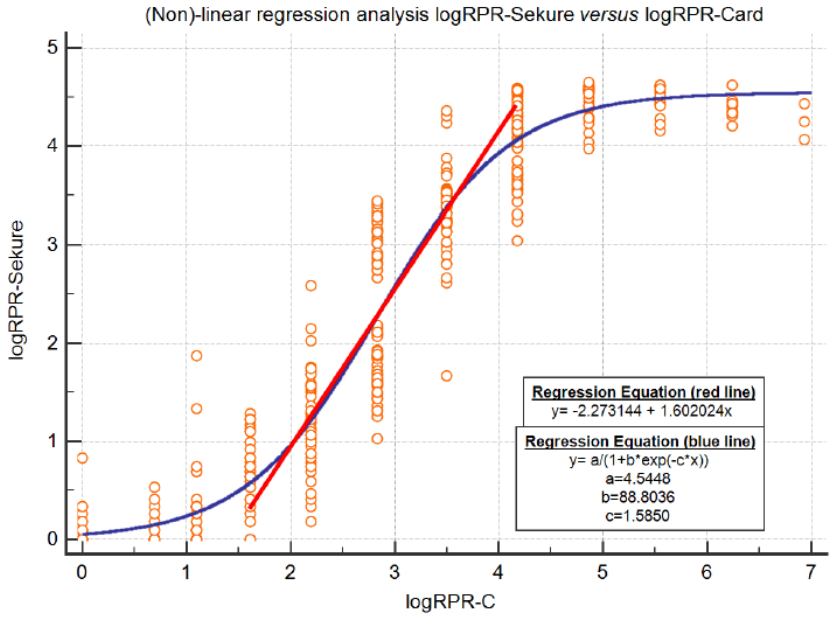
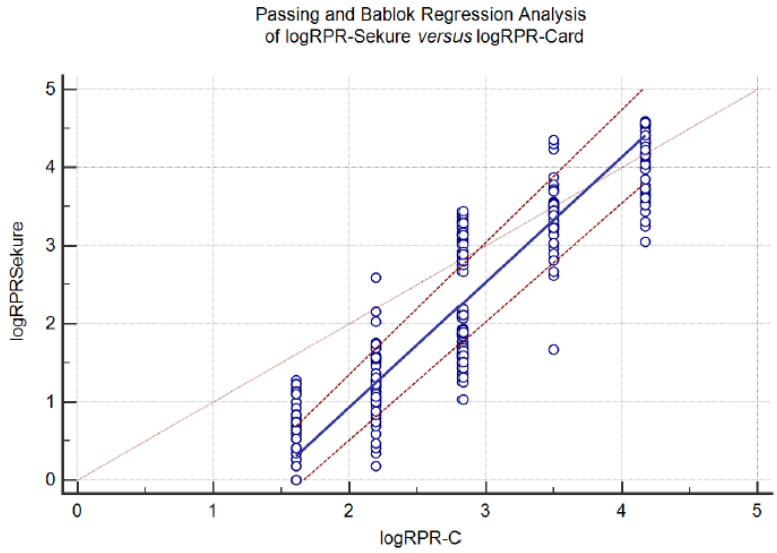


Fig S2 *Non-linear regression plot with best-fit fitted Passing-Bablok regression line of RPR-S plotted values corresponding to RPR-C titre range 4-64*



There was a non-normal data distribution (RPR-S: W= 0.77, P<0.001; RPR-C: W=0.45, P<0.001), thus the RPR-S second-run and RPR-C values were $\log_e(x)$ transformed. *Fig 2* is a graphical depiction of sigmoidal relationship that was fitted with the regression equation: "Y=a/(1+b*exp(-c*x))", whereby "a" (maximum asymptote $\log_e$RPR-S)= 4.5448 (95% CI: 4.43- 4.66), "b" (infection point)= 88.80 (95% CI: 63.43- 114.17), "c" (Hill's slope)= 1.59 (95% CI- 1.47- 1.70). When considering the RPR titre ranges 4-64 (N=246 samples) for the Passing Bablok linear regression analysis, the Intercept A was -2.27 (95% CI: -2.54- 2.04), representing the systemic differences, and the proportional differences (Slope B) was 1.60 (95 % CI: 1.52- 1.70), yielding the equation "y=-2.273144+1.602024x"). The RPR-C correlation linear range was determined to be between RPR-C titres 8-32, corresponding to a median RPR-S 2.3-29.7 R.U. (IQR 1.6-32.9).

Fig S3 *Passing and Bablok linear regression analysis of RRP-S* versus *RPR-C (RPR titre range 4-64)*



y = -2.273144 + 1.602024 x	
Systematic differences	
Intercept A	-2.2731
95% CI	-2.5391 to -2.0395
Proportional differences	
Slope B	1.6020
95% CI	1.5194 to 1.6954
Random differences	
Residual Standard Deviation (RSD)	0.2905
± 1.96 RSD Interval	-0.5693 to 0.5693
Linear model validity	
Cusum test for linearity	No significant deviation from linearity (P=0.18)

References

1. Newman L, Rowley J, Vander Hoorn S, Wijesooriya NS, Unemo M, Low N, et al. Global Estimates of the Prevalence and Incidence of Four Curable Sexually Transmitted Infections in 2012 Based on Systematic Review and Global Reporting. PLoS One. 2015;10(12):e0143304.
2. Apers L, Crucitti T, Verbrugge R, Vandenbruaene M. Sexually transmitted infections: what's new? Acta Clin Belg. 67(3):154–9.
3. Kenyon C, Lynen L, Florence E, Caluwaerts S, Vandenbruaene M, Apers L, et al. Syphilis reinfections pose problems for syphilis diagnosis in Antwerp, Belgium - 1992 to 2012. Euro Surveill. 2014;19(45):20958.
4. Seña ACC, White BLL, Sparling PFF. Novel Treponema pallidum serologic tests: a paradigm shift in syphilis screening for the 21st century. Clin Infect Dis. 2010;51(6):700–8.
5. Matthews HM, Yang TK, Jenkin HM. Unique lipid composition of Treponema pallidum (Nichols virulent strain). Infect Immun. 1979;24(3):713–9.
6. Belisle JT, Brandt ME, Radolf JD, Norgard M V. Fatty acids of Treponema pallidum and Borrelia burgdorferi lipoproteins. J Bacteriol. 1994;176(8):2151–7.
7. Report MW. Discordant results from reverse sequence syphilis screening--five laboratories, United States, 2006-2010. MMWR Morb Mortal Wkly Rep. 2011;60(5):133–7.
8. Lafond RE, Lukehart SA. Biological Basis for Syphilis. Clin Microbiol Rev. 2006;19(1):29–49.
9. Seña AC, Zhang X-H, Li T, Zheng H-P, Yang B, Yang L-G, et al. A systematic review of syphilis serological treatment outcomes in HIV-infected and HIV-uninfected persons: rethinking the significance of serological non-responsiveness and the serofast state after therapy. BMC Infect Dis. 2015;15(1):479.
10. Nandwani R, Evans DT. Are you sure it's syphilis? A review of false positive serology. Int J STD AIDS. 1995;6(4):241–8.
11. Hernández-Aguado I, Bolumar F, Moreno R, Pardo FJ, Torres N, Belda J, et al. False-positive tests for syphilis associated with human immunodeficiency virus and hepatitis B virus infection among intravenous drug abusers. Valencian Study Group on HIV Epidemiology. Eur J Clin Microbiol Infect Dis. 1998;17(11):784–7.
12. Liu F, Liu L-LL, Guo X-JJ, Xi Y, Lin L-RR, Zhang H-LL, et al. Characterization of the classical biological false-positive reaction in the serological test for syphilis in the modern era. Int Immunopharmacol. 2014;20(2):331–6.
13. Liu L-L, Lin L-R, Tong M-L, Zhang H-L, Huang S-J, Chen Y-Y, et al. Incidence and Risk Factors for the Prozone Phenomenon in Serologic Testing for Syphilis in a Large Cohort. Clin Infect Dis. 2014;59(3):384–9.
14. Haslett P, Laverty M. The prozone phenomenon in syphilis associated with HIV infection. Arch Intern Med. 1994;154(14):1643–4.
15. Jurado RL, Campbell J, Martin PD. Prozone Phenomenon in Secondary Syphilis. Arch Intern Med. 1993;153(21):2496.
16. Gupta SM, Bala M, Muralidhar S, Ray K. Evaluation of test results of microbiology laboratories of North India for standard tests for syphilis under an external quality assurance scheme. Eur J Clin Microbiol Infect Dis. 2009;28(5):461–8.
17. Muller I, Brade V, Hagedorn H-J, Straube E, Schorner C, Frosch M, et al. Is Serological Testing a Reliable Tool in Laboratory Diagnosis of Syphilis? Meta-Analysis of Eight External Quality Control Surveys Performed by the German Infection Serology Proficiency Testing Program. J Clin Microbiol. 2006;44(4):1335–41.
18. Huh HJ, Chung J-W, Park SY, Chae SL. Comparison of Automated Treponemal and Nontreponemal Test Algorithms as First-Line Syphilis Screening Assays. Ann Lab Med. 2016;36(1):23.
19. Noh J, Ko HH, Yun Y, Choi YS, Lee SG, Shin S, et al. Evaluation of Performance and False Positivity of Mediace RPR Test that Uses a Chemistry Autoanalyzer. Korean J Lab Med. 2008;28(4):312.
20. Kim YS, Lee J, Lee HK, Kim H, Kwon HJ, Min KO, et al. Comparison of Quantitative Results among Two Automated Rapid Plasma Reagin (RPR) Assays and a Manual RPR Test. Korean J Lab Med. 2009;29(4):331.

21. Workowski KA, Berman S, Centers for Disease Control and Prevention (CDC). Sexually transmitted diseases treatment guidelines, 2010. MMWR Recomm reports Morb Mortal Wkly report Recomm reports. 2010;59(RR-12):1–110.
22. French P, Gomberg M, Janier M, Schmidt B, van Voorst Vader P, Young H. IUSTI: 2008 European Guidelines on the Management of Syphilis. Int J STD AIDS. 2009;20(5):300–9.
23. Liu H, Rodes B, Chen C-Y, Steiner B. New Tests for Syphilis: Rational Design of a PCR Method for Detection of Treponema pallidum in Clinical Specimens Using Unique Regions of the DNA Polymerase I Gene. J Clin Microbiol. 2001;39(5):1941–6.
24. Cohen J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas. 1960;20(1):37–46.
25. Shapiro SS, Wilk MB. An Analysis of Variance Test for Normality (Complete Samples). Biometrika. 1965;52(3/4):591.
26. Passing H, Bablok. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. J Clin Chem Clin Biochem. 1983;21(11):709–20.
27. Wilcoxon F. Individual Comparisons by Ranking Methods. Biometrics Bull. 1945;1(6):80–3.
28. Yates F. Contingency Tables Involving Small Numbers and the χ 2 Test. Suppl to J R Stat Soc. 1934;1(2):217.
29. Rolfs RT, Joesoef MR, Hendershot EF, Rompalo AM, Augenbraun MH, Chiu M, et al. A Randomized Trial of Enhanced Therapy for Early Syphilis in Patients with and without Human Immunodeficiency Virus Infection. N Engl J Med. 1997;337(5):307–14.
30. McHugh ML. Interrater reliability: the kappa statistic. Biochem medica. 2012;22(3):276–82.
31. Osato, Kazuhisa; Nagao, Tomoe; Inuzumi, Kiyomi; Araki, Hisami; Kawai K. Clinical Evaluation of Latex Agglutination Test Kits for Detecting Anti-syphilitic Lipoidal Antibodies and Anti-treponemal Antibodies. Japanese J Sex Transm Dis. 2002;13(1):124–30.
32. Onoe T, Honde M, Matsuo K, Sasaki H, Sawamura M, Onoe Y, et al. Examination of the correlation between the manual and automated serological testing methods for syphilis. J Dermatol. 2012;39(4):355–61.
33. Kinjyo, Tohru, Nago, Tamami; Sakiyama, Kensin; Oshiro, Masaru; Nagamine, Tatumi; Yamane N. Laboratory -based evaluation of Latex Agglutination Turbidimetric Assay by Mediace RPR on P Module of Hitachi Auto analyzer 7600 to Quantitatively Determine Serum RPR Antibody. Japanese J Clin Lab Autom. 2005;30(3):257–62.
34. Stamm, L V. Syphilis: antibiotic treatment and resistance. Epidemiol Infect. 2014;143(8):1567–74.
35. Ghinsberg RC, Nitzan Y. Is syphilis an incurable disease? Med Hypotheses. 1992;39(1):35–40.
36. Martin Bland J, Altman D. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. Lancet. 1986;327(8476):307–10.

Chapter 7
General Discussion

The WHO estimates that everyday 16 000 people become infected with *T. pallidum* [1] worldwide, with record high syphilis incidence rates reported recently in countries such as the United States [2] and England [3]. At the time of writing this thesis, elimination of mother-to-child transmission of HIV and syphilis had already been achieved in Cuba, Thailand and Belarus [4]. Ambitious goals have been set by the WHO aimed at ending STI epidemics as public health concerns by 2030, including a 90% reduction in syphilis [5]. Now more than ever, robust diagnostic techniques and enhanced syphilis prevention strategies rooted in strong scientific support are needed to achieve these goals.

This thesis approaches syphilis from two perspectives: i) epidemiological, through the description of the relationship between syphilis and HIV at two different levels and ii) diagnostics, firstly by implementing a 'back to the basics' approach to *T. pallidum* protein biomarker discovery using MS-based proteomics approaches, in addition to the laboratory evaluation of an automated nontreponemal assay for clinical appropriateness.

In this general discussion, insights from these studies will be discussed in a broader context of STI prevention and diagnostic test development, together with future perspectives.

Lessons learnt through epidemiological studies of syphilis and HIV

The elucidation of STI epidemiological trends through the periodic analyses of datasets on multiple levels can help guide public health responses and resource allocation. These findings may also be useful predictors of potential future epidemics. Underlying determinants of HIV/STI spread, such as sexual networks and behaviour, evolve over time, influenced by factors such as AIDS mortality [6,7], condom-use [8] and societal norms [9]. There has been considerable debate as to what explains variations in HIV spread [10].

A recent country-level study showed that antenatal syphilis-prevalence was positively associated with HSV-2 prevalence during two different time periods (1990-1999 & 2008) indicating that similar risk-factor determinants influenced variations in prevalence for both STIs. Because HSV-2 is an incurable infection, variations in STI screening and treatment efficacy cannot explain variations in HSV-2 prevalence. Consequently, if variations in HSV-2 and syphilis prevalence are determined by the same factor, then this is less likely to be STI treatment efficacy [11]. In **chapter 2**, we built on this work by assessing if there was a country-level association between

antenatal syphilis from the early HIV epidemics (1900-1999) and peak HIV prevalence. Linear regression analyses of data from 76 countries revealed a strong association-syphilis prevalence in the 1990s predicted approximately half of the variation in peak HIV prevalence. We therefore concluded that underlying determinants played a role in the spread of both syphilis and HIV and/or syphilis was an important cofactor in the spread of HIV. These findings were concordant with studies of other population levels, such as individual [12,13], ethnic/racial group [14,15] and world regions [16]. Taken together with previous studies, which showed that at three levels of aggregation (ethnic groups [14], countries [11] and world regions [16]) prior to the HIV epidemic, populations with higher rates of HSV-2 had higher rates of syphilis and went on to experience higher HIV prevalences. These findings are important, as they are compatible with a parsimonious explanation that high STI prevalence populations had more densely connected sexual networks which put them at risk for higher prevalence of all/most STIs [17]. Individuals in these denser higher-risk sexual networks could be targeted for enhanced screening, educational campaigns (e.g. STI awareness/testing) and promotion of biomedical STI prevention strategies.

The widespread implementation of effective and accessible ART and new HIV prevention strategies could have an indirect impact on syphilis incidence. For example, mathematical studies have theorized that ART may contribute to increasing syphilis incidence by impairing host immunity to *T. pallidum* [18]. Some fears have been also raised that strategies such as treatment as prevention (TaSP), referring to a population strategy of starting ART shortly after HIV diagnosis and irrespective of $CD4^+$ T-cell count in order to suppress the viral load and prevent transmission to sexual partners and Pre-exposure prophylaxis (PrEP), a highly effective regime of ART taken to prevent HIV acquisition during period of regular high-risk exposure [19–21], could unintentionally result in behavioural disinhibition [22], analogous to earlier fears about penicillin and syphilis [23]. Currently, there is no evidence for risk compensation in sexual practises [24], however, it was pointed out by Alexei *et al.*, that initial PrEP efficacy studies were conducted with a placebo arm, thus participants might not have been willing to take risks [25]. With the recent roll-out of PrEP for use in regular clinical setting in many countries, including Belgium, continued studies are needed to assess STI outcomes in 'real world' circumstances.

Molecular phylogenetics is a useful adjuvant tool to investigate STI trends in clinical cohorts

Epidemiological studies supplemented with molecular phylogenetic analyses for transmission cluster inference can give unprecedented insights into disease

transmission patterns [26–28], with recent studies approaching near real-time transmission mapping capabilities depending on the (sampling/population) density of the dataset and analytical methods used [29,30]. Transmission clusters may explain as much as 75% of HIV incidence in MSMs, with one infection leading up to an estimated 10 to 13 onward transmissions [31,32]. The detection of subpopulations exposed to high(er) rates of transmission, together with improved understanding of links between clusters and underlying communities, could facilitate more impactful and cost-effective deployments of public health resources and contribute to the individualization of care.

In the study described in **chapter 3** we assessed if we could use the HIV phylogenetic tree of HIV clients in follow-up at a single-centre HIV cohort in Antwerp, Belgium to see if there was clustering of incident STIs, namely syphilis, chlamydia and gonorrhoea. Phylogenetic analyses of HIV-sequences from 1169 individuals using clustering software Cluster Picker [33], were performed using commonly used stringency criteria. Laboratory STI incidence, behavioural and demographic data were overlaid to investigate trends between the inferred clusters. In total, 37.5% of study patients were present in transmission clusters, which considering the 17-year sampling time span and high population diversity, was surprisingly high.

Interestingly, we found a high proportion (24/136 clusters; 92 individuals) of mixed sexual transmission clusters (MSM/heterosexual), these findings were concordant with other studies [34,35]. There was no evidence of STI clustering when statistical analyses were stratified for sexual preference, except for a group of heterosexual males with a history of syphilis. This lead us to conclude that there were six probable cases of sexual orientation misclassification, based on the close genetic proximity of these individuals to other MSM in the HIV phylogram and the history of syphilis, a disease prevalent in the local MSM population [36]. These probable misclassified individuals could represent a bridge population to lower risk groups. Our conclusions need to be interpreted with caution, however, due the possibility of unsampled female partners and software limitations that could potentially bias results.

Sexual identity is an important determinant of clinical risk profiling, thus thorough and non-judgemental sexual history taking and reporting in clinical practice is essential to guide appropriate screening and counselling. Studies have shown that patients are generally open to disclosing their sexual identity to their care givers [37–40], nevertheless, some clients could be reluctant to disclose their sexual identity to health care workers for fears of privacy intrusion and stigmatization. Even though

Belgium has strong anti-discriminatory laws [41] and scored low on internalized homonegativity [42] compared to other countries in Europe, negative sentiments towards sexual minority groups have been reported [43]. Moreover, a recent study described increased transmission between heterosexuals and MSMs in Nordic countries [34]. In Belgium, bisexual/heterosexual behavior prevalence among MSM is estimated to be 8.38% [44]. It is important to adapt STI counseling accordingly for these individuals to include their lower-risk female partners, if applicable.

Study limitations and future perspectives

Studies have indicated that most genetic clustering methods are systematically biased to detect variations among subpopulations in sampling rate. As a result analyses tend to collect individuals who were sampled soon after infection, whether or not those infections were from higher rates of transmission [45,46]. Consequently, from a public health perspective, subpopulations with less access to care who are also burdened by higher rates of transmission could be potentially neglected in targeting campaigns if targeting is based on these observations. Moreover, studies have had limited success in detecting (sexual) context network communities from phylogenetics [47] mostly because population dynamics strongly shape phylogenies and therefore have the potential to confound inferences of network structure [48,49].

Complex analytical methods such as spatial and/or temporal analyses were not employed in this study, yet these could be considered in future to add additional dimensions to our cohort. For example, the application of Kulldorff's scan statistical analyses [50] conducted by the SaTScan software could be applied in future studies. This method is useful [51–53] for the geographical surveillance of disease and for use in prospective, real-time or periodic disease surveillance for early detection of disease outbreaks. Sophisticated phylogenetics analyses such as molecular clock methods (reviewed by Ho and Duchêne [54]) and assessments of overlap between transmission clusters [47], could also be considered, however, these methods require the skills of an bioinformatician, hence they would be beyond the capacity of most health-care facilities.

Perspectives on STI surveillance

The age-old saying, recently reiterated by Margret Chan (Former Director General WHO)- "*What gets measured gets done*" resonates strongly for disease surveillance. One of the milestones for the 2020 WHO STI goals aims at having STI surveillance systems in place in 70% of countries that are able to monitor progress towards STI reduction

and elimination targets [5]. Accurate disease surveillance is contingent on representative population sampling, good reporting practices, appropriate diagnostic testing and background knowledge of important sociodemographic and behavioural factors prevalent in the setting(s) of interest. This can be achieved by following established recommendations [55].

Many methodological caveats can be present in the surveillance cascade, with the potential to misguide prevention efforts. Underreporting of disease incidence to public health authorities is a common occurrence due to gaps in reporting from laboratories and clinicians. With regards to syphilis surveillance, type of diagnostic testing used in screening programs have the potential to influence prevalence estimates. Particularly in low resource settings, syphilis testing is often only based on one test result without a second confirmatory test, a situation that can lead to over-diagnosis and treatment. A meta-analysis study by Ham *et al.* [56] took note of diagnostic and reporting discrepancies and inferred new correction factor estimates to improve syphilis seroprevalence estimates. These correction factors were applied in the **chapter 2** study. It is important to note that these had yet to been validated in separate studies, therefore it cannot be ruled out that they might have lead to inaccurate prevalence estimates, which is an alternative explanation for the observed HIV/syphilis association.

Advanced modelling and statistical methods can be used to increase the resolution of observations and can help extrapolate national survey data to smaller geographical areas. For example, a recent study [57] estimated chlamydia prevalence in small UK jurisdictions by modelling the infection, testing, and treatment processes and found that the positivity (proportion of positive tests) may not be a reliable proxy for local population prevalence. Moreover, data on sexual behaviour are often not routinely collected by large-scale surveys, which can lead to unrepresentative denominators for calculation of HIV and STI prevalence and incidence. In areas where behavioural data are scarce, these parameters can, in some circumstances, be indirectly estimated by extrapolation from other datasets, as demonstrated in a recent US study's estimation of small-area and rural MSM population sizes [58].

Depending on ones research objectives, it can also be important to account for societal contexts that could influence health outcomes. For example, studies of structural stigma, referring to societal policies, conditions and norms that compromise the resources, opportunities and well-being of a socially marginalized groups [59], can provide insights into health-seeking behaviour and STI determinants. Results from an

European MSM Internet Survey (EMIS) administered across 38 European countries (N = 174 209) revealed that MSM in European countries with high levels of stigma were more vulnerable to HIV and STI infections due to lower access to HIV/STI prevention services, higher odds of sexual risk behaviours, unmet prevention needs and decreased STI testing practices [9] and were more likely to report bisexual or heterosexual behaviours and attractions [44].

Future perspectives on STI prevention

Disease prevention can only be effective if it targets the correct population with the most suitable interventions. In 2015, the ECDC issued a report suggesting a number of specific key components that should be considered for inclusion into public health programs directed at MSM in Europe (detailed in *Table 1*). The promotion of these core services and interventions is expected to effectively reduce HIV and STI transmission among MSM [60].

Table 1 *Overview of key components that should be considered for effective HIV and STI prevention among MSM according to the ECDC. Reproduced from reference* [60]

Key Component	Description
Vaccinations	Promotion and delivery of vaccinations to protect against hepatitis A and B. Consider vaccination for human papilloma virus.
Condoms	Easily accessible condoms and condom-compatible lubricants and promotion of their effective use.
HIV and STI testing	Voluntary and confidential HIV and STI counselling and testing via a variety of modalities that are easy to access for the target group. Voluntary partner referral can support the early diagnosis and treatment of contacts.
Treatment	Timely provision of treatment for HIV, viral hepatitis and STIs should be ensured. Preventive benefits of treatment are significant.
Health promotion	Provide accurate and accessible information that enables men to understand and assess sexual health-related risks and prevention efficacy, and that promotes awareness of one's own HIV and STI status
MSM-competent health services	MSM-competent points of care offering a comprehensive sexual health programme including health promotion, counselling, peer support, prevention, adequate diagnostics and treatment will increase service uptake. Ensure target group involvement and training for providers on how to offer comprehensive care for MSM.
Targeted care for MSM living with HIV	Provide antiretroviral treatment for HIV and vaccination; regular STI screening using adequate diagnostics; treatment for STIs; individual counselling, sexual health promotion and peer-support groups for men living with HIV.

As highlighted in **chapter 3**, some individuals could be reluctant to disclose their sexual identity or behaviour with caregivers. When triaged for STI screening as a heterosexual they might not be offered a suitable STI laboratory work-up, including HIV antigen testing to detect early infection and syphilis screening. Moreover, they may also be less inclined to participate in regular STI screening. An outreach and online testing campaign, Swab2know, was shown to be feasible and effective manner of HIV testing of saliva among MSM in Belgium [61]. Such low-barrier and anonymous testing could be an attractive alternative for hard-to-reach populations. Health-seeking behaviour can be also encouraged through geosocial and sexual networking apps (eg. Grinder, Hornet and Scruff) that are increasingly being used by MSM to meet (anonymous) partners [62]. Recent studies have found this platform to acceptable and feasible for providing sexual health information and HIV/STD testing referrals [63,64].

Diagnostic screening can play a large role in controlling syphilis epidemics as long as there is suitable coverage and/or frequency [65,66], as demonstrated by the effectivity of the antenatal screening programs [67]. According to mathematical studies of simulated MSM populations, strategies that focus on higher frequency of testing in smaller higher-risk fractions of the population are likely to be most effective in reducing syphilis incidence [66]. However, screening actually coverage could be an additional mechanism for syphilis persistence in MSM. In a modelling study by Tuite *et al.* increased screening in populations with initially low levels of coverage may paradoxically lead to an initially higher equilibrium infection incidence than that observed in the absence of intervention due to treatment-induced loss of immunity [65]. Either way, without improved syphilis diagnostics, such as an antigen test, screening strategies will remain suboptimal.

Lastly, with regards to pre-exposure prophylaxis for bacterial STIs, a recent randomized control pilot study of high-risk MSM (N=30) demonstrated that daily doxycycline taken prophylactically over 48 weeks led to a 75% reduction in incident STIs (*N. gonorrhoeae*, *C. trachomatis*, or *T. pallidum)*. However, concerns have been raised about potential negative effects of prolonged low-dose antibiotic exposure on the microbiome and effects on antibiotic resistance [68,69].

Insights gained from MS-based *T. pallidum* biomarker discovery studies

Studies of human infectious diseases such as tuberculosis [70] and schistosomiasis [71] have benefited greatly during the last years from the application of MS-based proteomics methods. We investigated *T. pallidum* proteome expression during rabbit infection using complementary mass spectrometric (MS) methods, resulting in the detection of approximately half of the predicted proteome and relative quantification of the detected proteins through spectral counting techniques. Candidate biomarker proteins (N=11) were selected from these analyses and scrutinized for diagnostic test suitability via targeted MRM assay with heavy peptide surrogates and analysis of urine and plasma from *T. pallidum*-infected individuals. Unfortunately, none of the endogenous peptides were detected in the samples investigated. Polyclonal antibody enrichment of serially diluted *T. pallidum* in PBS and subsequent ESI-LTQ-Orbitrap MS analysis alluded to the insufficient detection capability of MS instruments to detect treponemal proteins without alternative sample preparation steps. These studies are described in **chapters 4** and **5** of this thesis.

Knowledge of protein expression patterns during infection can yield insights into *T. pallidum* pathogenesis and can aid in biomarker selection. We detected 116 predicted membrane proteins, 16 of them are predicted to be outer membrane proteins. Moreover, 114 uncharacterized/ hypothetical *T. pallidum* proteins were also uncovered, some even at high abundance, which shows that that there is still a long way to go in elucidating the unique biology of *T. pallidum*. Future studies could consider isolating treponemes from human syphilitic lesions (chancres, skin lesions) in order to obtain a better reflection of the human state of infection, which would likely differ from that of the rabbit milieu in terms of immune pressures. Albeit proteome analyses usually require a minimum of 10^6 bacteria/ml for a comprehensive analysis, innovations in MS technology and methods could make this feasible with lower pathogen concentrations. Obtaining more absolute and accurate *T. pallidum* protein concentrations could be achieved by using such methods as Isobaric tags for relative and absolute quantitation iTRAQ [72] or data independent acquisition techniques [73]. Post-translational modifications such as methylations, phosphorylations and alkylations are known to play an important role in bacterial function and pathogenesis, such as found in another spirochete *Leptospira interrogans* [74,75]. Further studies could elucidate (or be extra vigilant to) possible PTMs in *T. pallidum*, since PTMs can interfere with diagnostic applications by interfering with epitope binding by antibodies and/or MS database recognition.

Our pioneering study using MRM for *T. pallidum* protein detection in blood and urine samples described in **chapter 5** illustrates the inherent difficulty of detecting pathogen proteins in the proverbial haystack of host proteins. One of the main learning points of the works related to *T. pallidum* biomarker discovery was the realization that direct translation of bacterial *in vivo* expression studies may not be entirely translatable to the human diseased condition (and hence suitable as biomarkers) for various reasons, such as: i) protein expression can easily differ between rabbit and humans; ii) protein expression may have been influenced by stressful sample processing conditions (eg. temperature variations, centrifugal forces) and iii) spectral counting techniques were approximate and relative, therefore the abundance might not be reflective of the human infectious state. All in all, it was impossible to discern if the lack of *T. pallidum* protein detection was secondary to very low or lack of protein expression during infection, or rather if the limiting factor was the detection capability of the MS platform used.

There might have been subtle genetic differences between the database strains and those circulating in our population. Proteomics analyses are only as strong as the identification databases supporting them, thus if the genetic composition of the reference strains is no longer reflective of the circulating strains then this will have a negative effect on the identification accuracy. This is not of particular concern for *T. pallidum* since it evolves very slowly, indicative of its continued susceptibility to penicillin. However, there is experimental evidence that *T. pallidum* strains participate in evolutionary processes [76] and intra-strain variability has also been highlighted in studies [77], including our own analysis that found intra-strain differences at the protein level for four proteins (TP0082, TP0248, TP0922 and TP0692). Therefore, future *T. pallidum* protein studies should periodically refresh their sequence databases to reflect current (submitted) circulating strains and/or screen multiple strains in the database.

Our crude estimate of *T. pallidum* protein concentrations in plasma (femtomoles/litres) would likely only be applicable to the early stages of infection when the bacterial load in the blood is expected to be the highest. We determined that the limit of detection of MS instruments is 300 *T. pallidum*/ml in PBS based measurements on an ESI-Orbitrap LTQ mass spectrometer; this sensitivity could be higher with more sensitive MRM techniques. It is important to keep in mind, however, that serum has a very high protein concentration range, thus the detection limit could be lower than the predicted 300 treponemes/ml from the PBS-based experiment.

One of the study limitations was our reliance on the resolving capacity of MRM without extensive prefractionation or enrichment/depletion strategies. This decision not to apply depletion or enrichment techniques was based on previous studies that were successful in detecting cardiovascular [78] and cancer biomarkers [79] with a detection capability of four orders of magnitude (10^3–10^4 range in protein concentration) and up to femtomolar level sensitivity in plasma [80]. Depletion of high abundant proteins has disadvantages [81], the foremost could be the inadvertent depletion of the *T. pallidum* biomarkers since some proteins are know to bind to host proteins [82–84]. In the future, one could consider using protein and/or peptide enrichment using anti-peptide/ anti-protein (monoclonal/polyclonal) antibodies via the SISCAPA technique [85]. This method has shown considerable promise as a high-throughput, automated, highly multiplexed approach for protein biomarker quantification and validation [86] with MRM detection limits in the low picogram/ml range of protein concentration in plasma [87]. Another advantage of this method is the minimization, or in some cases elimination, of the liquid chromatography step owing to the high purity of the peptide analytes enriched by antibody capture, enabling the high-throughput needed for clinical studies.

In light of complexity of the plasma proteome, urine may be more of an ideal biofluid for syphilis diagnostics since it is less complex and can be easily produced and collected in a non-invasive manner. The human urinome has been rigorously interrogated during recent years using state-of-the-art mass spectrometric techniques, resulting in the identification of approximately 2500 unique proteins, many of which are modified by post-translational modifications such as glycosylation and phosphorylations [62–64]. Previous syphilis studies on urine have witnessed low sensitivity of PCR in urine (29–44%) during primary and secondary syphilis [88], yet, a promising finding was the positivity of secondary stage syphilis samples without chancres, indicating dissemination through blood filtration by the kidneys or bladder. Alternatively, shotgun proteomics techniques such as used in a tuberculosis study by Young *et al.* [89] and DIA methods [90,91] could be used to uncover new candidate *T. pallidum* biomarkers in urine.

Lessons learnt through the evaluation of the Sekure RPR test

To determine whether a test is useful for a given setting or application, evaluations must be conducted in the appropriate laboratory, clinical or field setting. There are many factors that can influence test performance, including differences in the infectious agent or population (e.g. genetic variability, disease prevalence) and different testing methodologies (automated versus manual; reagent types). Therefore,

test evaluations should be performed under a range of conditions that are likely to be encountered in practice and ideally through multicenter trials. A comprehensive report published by the TDR (Special Programme for Research and Training in Tropical Diseases) Diagnostics Evaluation Expert Panel [92] outlines the general principles of the design and conduct of trials to evaluate diagnostic tests for infectious diseases.

Currently, nontreponemal serological testing is the only method available for post-treatment serological follow-up of syphilis. In most laboratories this method is performed manually and thus subject to numerous shortcomings. In 2016, the Sekure RPR-test (RPR-S) was Conformité Européene (CE) marked and launched for European distribution. Studies were lacking on large-scale evaluations of the RPR-S test and the SeTPAT cohort could provide a unique longitudinal set of well-characterized clinical samples from a high syphilis-prevalence setting. During the test evaluation performance deficiencies were uncovered at both ends of the RPR titre spectrum, which in clinical practice would have led to probable missed diagnoses and over-diagnosis of the serofast state. These insights are useful for future refinements of this test.

The diagnostic test development pipeline is a high-cost endeavour. In the current economic milieu it is unlikely that an academic institute would have enough resources available to see the development of a diagnostic test through to validation and distribution. Industry partners benefit from academic expertise, independency and access to well characterized samples. In a review by Phillips *et al.*, [93] money and lack of samples for validating and testing biomarkers stood out as major reasons for the decrease in the rate novel diagnostics are reaching the market, described as the 'pipeline problem'. A shift to co-development models is expected to relieve some of these problems, and as shown in our study these efforts result in important findings and that could lead to better diagnostic test performance.

Future perspective on syphilis diagnostics

An excellent review by Caliendo *et al.* [94] describes the current state of affairs of infectious disease diagnostics, including unmet needs and technologies such as MALDI-TOF/TOF MS that are increasingly being used by clinical laboratories for pathogen detection. Time will tell if cost and the need for suitable laboratory facilities will be hurdles to the wide-spread dissemination of proteomics-based technologies to smaller laboratories and lower resource settings. With the current rapid pace of diagnostic technology development, I believe we can remain optimistic.

As mentioned before, antigen testing using specific antibodies is already widely used for the detection of many pathogens in various biofluids [94]. A definite advantage to these assays is the ability to distinguish new infections from old and the lack of reliance on cultivation of viable organisms (which in the case of *T. pallidum* is not feasible). In many cases antigen tests can be developed in a point of care format that are capable of providing rapid results to guide clinical decision making during the same encounter, or for use in home testing [95]. These are particularly useful in low resource settings where laboratory access can be limited [96]. Pending the identification and validation of syphilis biomarkers, it is hoped a syphilis antigen could become available in this format.

Conclusions

My original contributions to knowledge are:

- the description of a strong country-level association between peak HIV prevalence and antenatal syphilis prevalence in the 1990s, which likely explains approximately half of the variation in peak HIV prevalences. This alludes to the presence of similar underlying determinants, such as sexual network connectivity, that helped drive these epidemics.
- the finding that there are probable cases of sexual identity misclassification of MSM as heterosexuals in a HIV-cohort in Antwerp, Belgium. These individuals could represent a high-risk bridge population.
- the most comprehensive description of the *T. pallidum* proteome to date. These data give important insights into *in vivo* protein expression, which can be used to elucidate mechanisms of pathogenesis and yield potential protein biomarkers for diagnostic applications.
- the discovery that the limit of detection of MS instruments may be too high for *T. pallidum* protein detection in human biofluids without extra sample preparation steps such as antibody enrichment of target proteins/peptides. Alternative candidate *T. pallidum* protein biomarkers, besides the eleven examined in our study, could be investigated for their diagnostic potential.
    - such alternative biomarkers can be uncovered using a shotgun proteomics approach with MS-based DDA and DIA analyses of urine protein extracts from individuals with syphilis. Such experiments were initiated, but not yet completed, at the time of writing this thesis. Preliminary results confirmed the presence of four unique *T. pallidum* proteins in urine from individuals with primary and secondary stage syphilis; these could represent promising biomarkers.

- test performance deficiencies were uncovered during a laboratory evaluation of a syphilis nontreponemal assay. These insights can be helpful for improving future test design and function.

The worldwide human and economic burden of STIs is staggering. For this reason, policy makers need to be convinced by study results that investments in clinical (including diagnostics) and public health strategies can improve STI control [97]. As shown in our studies, (molecular) epidemiological studies can shed light on possible determinants of STI spread, which can be useful for guiding prevention efforts. Moreover, molecular phylogenetic methods applied to clinical cohort descriptions can highlight transmission trends, such as high-risk bridge populations, that might not be possible with conventional analyses.

With regards to syphilis biomarker discovery, these preliminary efforts will hopefully pave the way towards the realization of an antigen test. Latent stage syphilis, when *T. pallidum* levels are at their lowest and sequestered away, will be the most challenging disease stage to diagnose directly. If this is not feasible, diagnosis of early stage syphilis via sensitive direct detection methods will no doubt still be a substantial improvement in the syphilis diagnostic arsenal.

References

1. Newman L, Rowley J, Vander Hoorn S, Wijesooriya NS, Unemo M, Low N, et al. Global Estimates of the Prevalence and Incidence of Four Curable Sexually Transmitted Infections in 2012 Based on Systematic Review and Global Reporting. PLoS One. 2015;10(12):e0143304.
2. Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2015. Atlanta; 2016.
3. England PH. Sexually Transmitted Infections and Chlamydia Screening in England, 2016. 2017.
4. World Health Organization. WHO | Elimination of mother-to-child transmission (EMTCT) of HIV and syphilis. Who. 2017;
5. WHO | Global health sector strategy on Sexually Transmitted Infections, 2016-2021. WHO. 2016;
6. Chesson HW, Dee TS, Aral SO. Aids mortality may have contributed to the decline in syphilis rates in the United States in the 1990s. Sex Transm Dis. 30(5):419–24.
7. Kenyon CR, Osbak K, Buyze J, Chico RM. The changing relationship between bacterial STIs and HIV prevalence in South Africa ? an ecological study. Int J STD AIDS. 2015;26(8).
8. Abara WE, Oraka E, Jeffries WL, Chavez P, Nasrullah M, DiNenno E. Prevalence and correlates of condom use among sexually active men who have sex with men in the United States: findings from the National Survey of Family Growth, 2002, 2006?10 and 2011?13. Sex Health. 2017;
9. Pachankis JE, Hatzenbuehler ML, Hickson F, Weatherburn P, Berg RC, Marcus U, et al. Hidden from health: structural stigma, sexual orientation concealment, and HIV across 38 countries in the European MSM Internet Survey. AIDS. 2015;29(10):1239–46.
10. Buvé A, Laga M. Epidemiological research in the HIV field. AIDS. 2012;26(10):1203–4.
11. Kenyon CR, Tsoumanis A, Osbak K. Strong Country Level Correlation between Syphilis and HSV-2 Prevalence. J Sex Transm Dis. 2016;2016:1–6.
12. Nelson KE, Eiumtrakul S, Celentano D, Maclean I, Ronald A, Suprasert S, et al. The association of herpes simplex virus type 2 (HSV-2), Haemophilus ducreyi, and syphilis with HIV infection in young men in northern Thailand. J Acquir Immune Defic Syndr Hum Retrovirol. 1997;16(4):293–300.
13. Buvé A, Weiss HA, Laga M, Van Dyck E, Musonda R, Zekeng L, et al. The epidemiology of gonorrhoea, chlamydial infection and syphilis in four African cities. AIDS. 2001;15 Suppl 4:S79-88.
14. Kenyon CR, Osbak K, Buyze J. The Prevalence of HIV by Ethnic Group Is Correlated with HSV-2 and Syphilis Prevalence in Kenya, South Africa, the United Kingdom, and the United States. Interdiscip Perspect Infect Dis. 2014;2014:284317.
15. Morris M, Kurth AE, Hamilton DT, Moody J, Wakefield S. Concurrent partnerships and HIV prevalence disparities by race: linking science and public health practice. Am J Public Health. 2009;99(6):1023–31.
16. Kenyon C, Buyze J, Colebunders R. Classification of incidence and prevalence of certain sexually transmitted infections by world regions. Int J Infect Dis. 2014;18:73–80.
17. Kenyon CR, Buyze J, Klebanoff M, Brotman RM. Incident Trichomonas vaginalis Is Associated With Partnership Concurrency. Sex Transm Dis. 2017;1.
18. Rekart ML, Ndifon W, Brunham RC, Dushoff J, Park SW, Rawat S, et al. A double-edged sword: does highly active antiretroviral therapy contribute to syphilis incidence by impairing immunity to *Treponema pallidum* ? Sex Transm Infect. 2017;93(5):374–8.
19. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure Chemoprophylaxis for HIV Prevention in Men Who Have Sex with Men. N Engl J Med. 2010;363(27):2587–99.
20. Molina J-M, Capitant C, Spire B, Pialoux G, Cotte L, Charreau I, et al. On-Demand Preexposure Prophylaxis in Men at High Risk for HIV-1 Infection. N Engl J Med. 2015;373(23):2237–46.
21. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. Lancet (London, England). 2016;387(10013):53–60.
22. Blumenthal J, Haubrich RH. Will risk compensation accompany pre-exposure prophylaxis for HIV? Virtual Mentor. 2014;16(11):909–15.
23. Quetel C. History of Syphilis. Cambridge, UK: Polity Press; 1990.

24. Freeborn K, Portillo CJ. Does Pre-exposure prophylaxis (PrEP) for HIV prevention in men who have sex with men (MSM) change risk behavior? A systematic review. J Clin Nurs. 2017;
25. Alaei K, Paynter CA, Juan S-C, Alaei A. Using preexposure prophylaxis, losing condoms? Preexposure prophylaxis promotion may undermine safe sex. AIDS. 2016;30(18):2753–6.
26. Leigh Brown AJ, Lycett SJ, Weinert L, Hughes GJ, Fearnhill E, Dunn DT, et al. Transmission network parameters estimated from HIV sequences for a nationwide epidemic. J Infect Dis. 2011;204(9):1463–9.
27. Brenner BG, Roger M, Stephens D, Moisi D, Hardy I, Weinberg J, et al. Transmission clustering drives the onward spread of the HIV epidemic among men who have sex with men in Quebec. J Infect Dis. 2011;204(7):1115–9.
28. Chalmet K, Staelens D, Blot S, Dinakis S, Pelgrom J, Plum J, et al. Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. BMC Infect Dis. 2010;10(1):262.
29. Poon AFY, Gustafson R, Daly P, Zerr L, Demlow SE, Wong J, et al. Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. lancet HIV. 2016;3(5):e231-8.
30. Little SJ, Kosakovsky Pond SL, Anderson CM, Young JA, Wertheim JO, Mehta SR, et al. Using HIV Networks to Inform Real Time Prevention Interventions. Harrigan PR, editor. PLoS One. 2014;9(6):e98443.
31. Brenner BG, Wainberg MA. Future of phylogeny in HIV prevention. J Acquir Immune Defic Syndr. 2013;63 Suppl 2(0 2):S248-54.
32. Leventhal GE, Günthard HF, Bonhoeffer S, Stadler T. Using an epidemiological model for phylogenetic inference reveals density dependence in HIV transmission. Mol Biol Evol. 2014;31(1):6–17.
33. Ragonnet-Cronin M, Hodcroft E, Hué S, Fearnhill E, Delpech V, Brown AJL, et al. Automated analysis of phylogenetic clusters. BMC Bioinformatics. 2013;14:317.
34. Esbjörnsson J, Mild M, Audelin A, Fonager J, Skar H, Bruun Jørgensen L, et al. HIV-1 transmission between MSM and heterosexuals, and increasing proportions of circulating recombinant forms in the Nordic Countries. Virus Evol. 2016;2(1):vew010.
35. Hué S, Brown AE, Ragonnet-Cronin M, Lycett SJ, Dunn DT, Fearnhill E, et al. Phylogenetic analyses reveal HIV-1 infections between men misclassified as heterosexual transmissions. AIDS. 2014;28(13):1967–75.
36. Kenyon C, Lynen L, Florence E, Caluwaerts S, Vandenbruaene M, Apers L, et al. Syphilis reinfections pose problems for syphilis diagnosis in Antwerp, Belgium - 1992 to 2012. Euro Surveill. 2014;19(45):20958.
37. Verhoeven V, Colliers A, Verster A, Avonts D, Peremans L, Van Royen P. Collecting data for sexually transmitted infections (STI) surveillance: what do patients prefer in Flanders? BMC Health Serv Res. 2007;7(1):149.
38. Haider AH, Schneider EB, Kodadek LM, Adler RR, Ranjit A, Torain M, et al. Emergency Department Query for Patient-Centered Approaches to Sexual Orientation and Gender Identity. JAMA Intern Med. 2017;177(6):819.
39. Rosenberg ES, Sullivan PS, DiNenno EA, Salazar LF, Sanchez TH. Number of casual male sexual partners and associated factors among men who have sex with men: Results from the National HIV Behavioral Surveillance system. BMC Public Health. 2011;11(1):189.
40. Coleman TA, Bauer GR, Pugh D, Aykroyd G, Powell L, Newman R. Sexual Orientation Disclosure in Primary Care Settings by Gay, Bisexual, and Other Men Who Have Sex with Men in a Canadian City. LGBT Heal. 2017;4(1):42–54.
41. Control EC for DP and. hematic report: Stigma and discrimination . Monitoring implementatio n of the Dublin Declaration on Partnership to F ight HIV/AIDS in Europe and Central Asia: 2012 Progress Report. Stockholm; 2013.
42. Berg RC, Ross MW, Weatherburn P, Schmidt AJ. Structural and environmental factors are associated with internalised homonegativity in men who have sex with men: Findings from the European MSM Internet Survey (EMIS) in 38 countries. Soc Sci Med. 2013;78:61–9.
43. Hooghe M, Claes E, Harell A, Quintelier E, Dejaeghere Y. Anti-Gay Sentiment Among Adolescents in Belgium and Canada: A Comparative Investigation into the Role of Gender and Religion. J

Homosex. 2010;57(3):384–400.
44. Pachankis JE, Hatzenbuehler ML, Mirandola M, Weatherburn P, Berg RC, Marcus U, et al. The Geography of Sexual Orientation: Structural Stigma and Sexual Attraction, Behavior, and Identity Among Men Who Have Sex with Men Across 38 European Countries. Arch Sex Behav. 2016;
45. Poon AFY. Impacts and shortcomings of genetic clustering methods for infectious disease outbreaks. Virus Evol. 2016;2(2):vew031.
46. Volz EM, Koopman JS, Ward MJ, Brown AL, Frost SDW. Simple Epidemiological Dynamics Explain Phylogenetic Clustering of HIV from Patients with Recent Infection. Fraser C, editor. PLoS Comput Biol. 2012;8(6):e1002552.
47. Villandre L, Stephens DA, Labbe A, Günthard HF, Kouyos R, Stadler T, et al. Assessment of Overlap of Phylogenetic Transmission Clusters and Communities in Simple Sexual Contact Networks: Applications to HIV-1. Braunstein LA, editor. PLoS One. 2016;11(2):e0148459.
48. Robinson K, Fyson N, Cohen T, Fraser C, Colijn C. How the Dynamics and Structure of Sexual Contact Networks Shape Pathogen Phylogenies. Kosakovsky Pond SL, editor. PLoS Comput Biol. 2013;9(6):e1003105.
49. Grabowski MK, Redd AD. Molecular tools for studying HIV transmission in sexual networks. Curr Opin HIV AIDS. 2014;9(2):126–33.
50. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. Stat Med. 1995;14(8):799–810.
51. Gesink DC, Sullivan AB, Norwood T a, Serre ML, Miller WC. Does core area theory apply to sexually transmitted diseases in rural environments? Sex Transm Dis. 2013;40(1):32–40.
52. Petersen J, Gibin M, Sile B, Simms I. Identifying and interpreting spatiotemporal variation in diagnoses of infectious syphilis among men, England: 2009 to 2013. Sex Transm Infect. 2016;92(5):380–6.
53. van Aar F, den Daas C, van der Sande MAB, Soetens LC, de Vries HJC, van Benthem BHB. Outbreaks of syphilis among men who have sex with men attending STI clinics between 2007 and 2015 in the Netherlands: a space-time clustering study. Sex Transm Infect. 2016;1–6.
54. Ho SYW, Duchêne S. Molecular-clock methods for estimating evolutionary rates and timescales. Mol Ecol. 2014;23(24):5947–65.
55. Centers for Disease Control and Prevention (CDC). Recommendations for Public Health Surveillance of Syphilis in the United States. Atlanda; 2003.
56. Ham DC, Lin C, Newman L, Wijesooriya NS, Kamb M. Improving global estimates of syphilis in pregnancy by diagnostic test type: A systematic review and meta-analysis. Int J Gynecol Obstet. 2015;130:S10–4.
57. Lewis J, White PJ. Estimating Local Chlamydia Incidence and Prevalence Using Surveillance Data. Epidemiology. 2017;28(4):492–502.
58. Grey JA, Bernstein KT, Sullivan PS, Purcell DW, Chesson HW, Gift TL, et al. Estimating the Population Sizes of Men Who Have Sex With Men in US States and Counties Using Data From the American Community Survey. JMIR Public Heal Surveill. 2016;2(1):e14.
59. Hatzenbuehler ML, Link BG. Introduction to the special issue on structural stigma and health. Soc Sci Med. 2014;103:1–6.
60. Control EC for DP and. HIV and STI prevention among men who have sex with men. Stockholm; 2015.
61. Platteau T, Fransen K, Apers L, Kenyon C, Albers L, Vermoesen T, et al. Swab2know: An HIV-Testing Strategy Using Oral Fluid Samples and Online Communication of Test Results for Men Who Have Sex With Men in Belgium. J Med Internet Res. 2015;17(9):e213.
62. Goedel WC, Duncan DT. Geosocial-Networking App Usage Patterns of Gay, Bisexual, and Other Men Who Have Sex With Men: Survey Among Users of Grindr, A Mobile Dating App. JMIR Public Heal Surveill. 2015;1(1):e4.
63. Sun CJ, Stowers J, Miller C, Bachmann LH, Rhodes SD. Acceptability and Feasibility of Using Established Geosocial and Sexual Networking Mobile Applications to Promote HIV and STD Testing Among Men Who Have Sex with Men. AIDS Behav. 2015;19(3):543–52.
64. Jenkins Hall W, Sun CJ, Tanner AE, Mann L, Stowers J, Rhodes SD. HIV-Prevention Opportunities With GPS-Based Social and Sexual Networking Applications for Men Who Have Sex With Men. AIDS Educ Prev. 2017;29(1):38–48.

65. Tuite A, Fisman D. Go big or go home: impact of screening coverage on syphilis infection dynamics. Sex Transm Infect. 2016;92(1):49–54.
66. Tuite AR, Fisman DN, Mishra S. Screen more or screen more often? Using mathematical models to inform syphilis control strategies. BMC Public Health. 2013;13(1):606.
67. Hawkes S, Matin N, Broutet N, Low N. Effectiveness of interventions to improve screening for syphilis in pregnancy: a systematic review and meta-analysis. Lancet Infect Dis. 2011;11(9):684–91.
68. Dubourg G, Raoult D. The challenges of preexposure prophylaxis for bacterial sexually transmitted infections. Clin Microbiol Infect. 2016;22(9):753–6.
69. Golden MR, Handsfield HH. Preexposure Prophylaxis to Prevent Bacterial Sexually Transmitted Infections in Men Who Have Sex With Men. Sex Transm Dis. 2015;42(2):104–6.
70. Haas CT, Roe JK, Pollara G, Mehta M, Noursadeghi M. Diagnostic "omics" for active tuberculosis. BMC Med. 2016;14(1):37.
71. Driguez P, McManus DP, Gobert GN. Clinical implications of recent findings in schistosome proteomics. Expert Rev Proteomics. 2016;13(1):19–33.
72. Eshghi A, Cullen P a, Cowen L, Zuerner RL, Cameron CE. Global Proteome Analysis of Leptospira interrogans research articles. 2009;4564–78.
73. Malmström L, Bakochi A, Svensson G, Kilsgård O, Lantz H, Petersson AC, et al. Quantitative proteogenomics of human pathogens using DIA-MS. J Proteomics. 2015;129:98–107.
74. Eshghi A, Pinne M, Haake DA, Zuerner RL, Frank A, Cameron CE. Methylation and in vivo expression of the surfaceexposed Leptospira interrogans outer-membrane protein OmpL32. Microbiology. 2012;158(3):622–35.
75. Witchell TD, Eshghi A, Nally JE, Hof R, Boulanger MJ, Wunder EA, et al. Post-translational Modification of LipL32 during Leptospira interrogans Infection. Small PLC, editor. PLoS Negl Trop Dis. 2014;8(10):e3280.
76. Giacani L, Chattopadhyay S, Centurion-Lara A, Jeffrey BM, Le HT, Molini BJ, et al. Footprint of positive selection in Treponema pallidum subsp. pallidum genome sequences suggests adaptive microevolution of the syphilis pathogen. Picardeau M, editor. PLoS Negl Trop Dis. 2012;6(6):e1698.
77. Čejková D, Strouhal M, Norris SJ, Weinstock GM, Šmajs D. A Retrospective Study on Genetic Heterogeneity within Treponema Strains: Subpopulations Are Genetically Distinct in a Limited Number of Positions. PLoS Negl Trop Dis. 2015;9(10):e0004110.
78. Domanski D, Percy AJ, Yang J, Chambers AG, Hill JS, Freue GVC, et al. MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. Proteomics. 2012;12(8):1222–43.
79. Percy AJ, Chambers AG, Yang J, Borchers CH. Multiplexed MRM-based quantitation of candidate cancer biomarker proteins in undepleted and non-enriched human plasma. Proteomics. 2013;13(14):2202–15.
80. Kuzyk MA, Smith D, Yang J, Cross TJ, Jackson AM, Hardie DB, et al. Multiple Reaction Monitoring-based, Multiplexed, Absolute Quantitation of 45 Proteins in Human Plasma. Mol Cell Proteomics. 2009;8(8):1860–77.
81. Pernemalm M, Lehtiö J. Mass spectrometry-based plasma proteomics: state of the art and future outlook. Expert Rev Proteomics. 2014;11(4):431–48.
82. Cameron CE, Brown EL, Kuroiwa JMY, Schnapp LM, Brouwer NL. Treponema pallidum fibronectin-binding proteins. J Bacteriol. 2004;186(20):7019–22.
83. Cameron CE. Identification of a Treponema pallidum laminin-binding protein. Infect Immun. 2003;71(5):2525–33.
84. Penn CW, Cockayne A, Bailey MJ. The Outer Membrane of Treponema pallidum: Biological Significance and Biochemical Properties. Microbiology. 1985;131(9):2349–57.
85. Anderson NLG, Anderson NLG, Haines LR, Hardie DB, Olafson RW, Pearson TW. Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). J Proteome Res. 3(2):235–44.
86. Hsiao YC, Chi LM, Chien KY, Chiang WF, Chen SF, Chuang YN, et al. Development of a multiplexed assay for oral cancer candidate biomarkers using peptide immunoaffinity enrichment and targeted mass spectrometry. Mol Cell Proteomics. 2017;mcp.000147.2017.
87. Whiteaker JR, Zhao L, Anderson L, Paulovich AG. An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass

spectrometry-based quantification of protein biomarkers. Mol Cell Proteomics. 2010;9(1):184–96.
88. Engelkens HJ, ten Kate FJ, Vuzevski VD, van der Sluis JJ, Stolz E, Piguet V, et al. Primary and secondary syphilis: a histopathological study. Int J STD AIDS. 2009;2(4):280–4.
89. Young BL, Mlamla Z, Gqamana PP, Smit S, Roberts T, Peter J, et al. The identification of tuberculosis biomarkers in human urine samples. Eur Respir J. 2014;43(6):1719–29.
90. Muntel J, Xuan Y, Berger ST, Reiter L, Bachur R, Kentsis A, et al. Advancing Urinary Protein Biomarker Discovery by Data-Independent Acquisition on a Quadrupole-Orbitrap Mass Spectrometer. J Proteome Res. 2015;14(11):4752–62.
91. Nigjeh EN, Chen R, Brand RE, Petersen GM, Chari ST, von Haller PD, et al. Quantitative Proteomics Based on Optimized Data-Independent Acquisition in Plasma Analysis. J Proteome Res. 2017;16(2):665–76.
92. TDR Diagnostics Evaluation Expert Panel, Banoo S, Bell D, Bossuyt P, Herring A, Mabey D, et al. Evaluation of diagnostic tests for infectious diseases: general principles. Nat Rev Microbiol. 2010;8(12 Suppl):S17-29.
93. Phillips KA, Van Bebber S, Issa AM. Diagnostics and biomarker development: priming the pipeline. Nat Rev Drug Discov. 2006;5(6):463–9.
94. Caliendo AM, Gilbert DN, Ginocchio CC, Hanson KE, May L, Quinn TC, et al. Better tests, better care: improved diagnostics for infectious diseases. Clin Infect Dis. 2013;57 Suppl 3(Suppl 3):S139-70.
95. Tucker JD, Bien CH, Peeling RW. Point-of-care testing for sexually transmitted infections. Curr Opin Infect Dis. 2013;26(1):73–9.
96. Marks M, Mabey DC. The introduction of syphilis point of care tests in resource limited settings. Expert Rev Mol Diagn. 2017;17(4):321–5.
97. Unemo M, Bradshaw CS, Hocking JS, de Vries HJC, Francis SC, Mabey D, et al. Sexually transmitted infections: challenges ahead. Lancet Infect Dis. 2017;17(8):e235–79.

Curriculum Vitae

PERSONAL INFORMATION

Birth date:	December 2, 1984
Birth place:	Edmonton, Alberta, Canada
Nationality:	Dutch

EDUCATION

2018-2023	Medical microbiology residency programme Erasmus MC, Rotterdam, the Netherlands
2014-2017	Graduate studies PhD programme University of Antwerp/ Institute of Tropical Medicine
2005-2012	Medicine Universiteit van Amsterdam, Amsterdam, the Netherlands
2005	Academic Dutch Language, Staatsexamen NT2-II diploma
2002-2004	Bachelor of Science programme University of Alberta, Edmonton, Canada
2002	International Baccalaureate diploma

RESEARCH EXPERIENCE

PhD candidate: 'Treponema pallidum-*specific proteomic changes in patients with incident syphilis infection: an observational study in Belgium'* SeTPAT project
Supervisors: Prof. dr. C.R. Kenyon & Prof. dr. Xaveer van Ostade

2013- Research assistant '*Prediction and Pathogenesis of Immune Reconstitution Inflammatory Syndrome'* study
Institute: Centre de Recherché Medicales de Lambaréné, Gabon
Supervisor: Prof. dr. M.P. Grobusch

2009- Research internship MSc degree: '*Glucokinase Double Mutant Project*'
Institute: Oxford Centre for Diabetes, Endocrinology en Metabolism, Oxford University, United Kingdom
Supervisors: Prof. dr. A.L. Gloyn & Prof. dr. J.B.L. Hoekstra

PUBLICATIONS

Osbak KK, Abdellati S, Tsoumanis A, Van Esbroeck M, Kestens L, Crucitti T, et al. Evaluation of an automated quantitative latex immunoturbidimetric non-treponemal assay for diagnosis and follow-up of syphilis: a prospective cohort study. J Med Microbiol. 2017.

Kenyon CR, **Osbak KK**, Van Esbroek M, Lynen L, Crucitti T. What Is the Role of Paired Rapid Plasma Reagin Testing (Simultaneous Testing of Acute and Convalescent Samples) in the Diagnosis of Repeat Syphilis and the Follow-up of Syphilis? Sex Transm Dis. 2017; 1.

Kenyon, CR; Tsoumanis, A; **Osbak, KK**; van Esbroeck, M; Florence, E; Crucitti, T; Kestens L. Repeat syphilis has a different immune response compared to initial syphilis. Sex Transm Infect. 2017. *in press*.

Kenyon C, **Osbak KK**, Crucitti T, Kestens L. The immunological response to syphilis differs by HIV status; a prospective observational cohort study. BMC Infect Dis. 2017.

Janssen S, **Osbak KK**, Holman R, Hermans S, Moekotte A, Knap M, et al. Low incidence of the immune reconstitution inflammatory syndrome among HIV-infected patients starting antiretroviral therapy in Gabon: a prospective cohort study. Infection. 2017.

Mikalová L, Grillová L, **Osbak KK**, Strouhal M, Kenyon C, Crucitti T, et al. Molecular Typing of Syphilis-Causing Strains Among Human Immunodeficiency Virus-Positive Patients in Antwerp, Belgium. Sex Transm Dis. 2017.

Osbak KK, Houston S, Lithgow K V, Meehan CJ, Strouhal M, Šmajs D, et al. Characterizing the Syphilis-Causing Treponema pallidum ssp. pallidum Proteome Using Complementary Mass Spectrometry. PLoS Negl Trop Dis. 2016.

Kenyon CR, **Osbak KK**, Vandenbruane M. What is the appropriate therapy for asymptomatic gonorrhoea in men who have sex with men? Int J STD AIDS. 2016.

Kenyon CR, **Osbak KK**, Tsoumanis A. The Global Epidemiology of Syphilis in the Past Century - A Systematic Review Based on Antenatal Syphilis Prevalence. PLoS NTD. 2016.

Kenyon CR, Tsoumanis A, **Osbak KK**. Strong Country Level Correlation between Syphilis and HSV-2 Prevalence. J Sex Transm Dis. 2016.

Osbak KK, Rowley JT, Kassebaum NJ, Kenyon CR. The Prevalence of Syphilis From the Early HIV Period Is Correlated With Peak HIV Prevalence at a Country Level. Sex Transm Dis. 2016.

Kenyon CR, **Osbak KK**. How many MSM in Europe could benefit from PrEP--a 9 billion Euro question? Int J STD AIDS. 2015.

Kenyon C, **Osbak KK**. Sexual networks, HIV, race and bacterial vaginosis. AIDS. 2015.

Bélard S, Janssen S, **Osbak KK**, Adegnika AA, Ondounda M, Grobusch MP. Limited access to drugs for resistant tuberculosis: a call to action. J Public Health (Oxf). 2015.

PUBLICATIONS (cont.)

Kenyon CR, **Osbak KK**, Chico RM. What underpins the decline in syphilis in Southern and Eastern Africa? An exploratory ecological analysis. Int J Infect Dis. 2014.

Kenyon CR, **Osbak KK**, Buyze J, Johnson S, van Lankveld J. Variations of Sexual Scripts Relating to Concurrency by Race, Class, and Gender in South Africa. J Sex Res. 2015.

Kenyon CR, **Osbak KK**, Buyze J. The Prevalence of HIV by Ethnic Group Is Correlated with HSV-2 and Syphilis Prevalence in Kenya, South Africa, the United Kingdom, and the United States. Interdiscip Perspect Infect Dis. 2014.

Kenyon CR, **Osbak KK**. Recent progress in understanding the epidemiology of bacterial vaginosis. Curr Opin Obstet Gynecol. 2014.

Kenyon CR, **Osbak KK**, Buyze J, Chico RM. The changing relationship between bacterial STIs and HIV prevalence in South Africa - an ecological study. Int J STD AIDS. 2015.

Kenyon C, **Osbak KK**. Certain attributes of the sexual ecosystem of high-risk MSM have resulted in an altered microbiome with an enhanced propensity to generate and transmit antibiotic resistance. Med Hypotheses. 2014.

Osbak KK*, Beer NL*, van de Bunt M, Tribble ND, Steele AM, Wensley KJ, et al. Insights Into the Pathogenicity of Rare Missense GCK Variants From the Identification and Functional Characterization of Compound Heterozygous and Double Mutations Inherited in Cis. Diabetes Care. 2012. **shared first authorship*

Osbak KK, Colclough K, Saint-Martin C, Beer NL, Bellanné-Chantelot C, Ellard S, et al. Update on mutations in glucokinase (GCK), which cause maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemic hypoglycemia. Hum Mutat. 2009.

INTERNATIONAL POSTER & ORAL PRESENTATIONS

STI & HIV World Congress, 2017, Rio de Janeiro, Brazil- *poster*

Gordon Research Seminar/Conference "Biology of Spirochetes" 2016, Ventura, United States- *poster*

Research Seminar: "*T. pallidum* Biomarker Discovery", 2015, University of Victoria, Victoria, Canada- *invited speaker*

Human Proteome Organization World Congress 2015, Vancouver, Canada- *poster*

European Congress of Clinical Microbiology and Infectious Diseases 2015, Copenhagen, Denmark- *poster*

SCHOLARSHIPS & PRIZES

Erasmus Scholarship 2009

Amsterdam University Fund Spinoza Scholarship 2009

Novo Nordisk Funding Award 2009

'Outstanding Poster Award' & travel grant- Gordon Research Seminar 2016

Acknowledgements

I would like to begin with expressing my gratitude for the opportunity to conduct scientific research. Pursuing a PhD degree has been by far one of my most challenging and rewarding endeavors. A line from Dr. Seuss aptly describes some sections of this journey: *"Onward up many a frightening creek, though your arms may get sore and your sneakers may leak...Oh! The places you'll go!"*. Being on the forefront of the knowledge unknown is exhilarating. This process has made me a better person.

It was a privilege to work with and learn from many talented individuals during the last four years. The successful completion of this thesis would not have been possible without the unfaltering support of my supervisors, colleagues, family, friends and partner. The following are short accounts of individuals who played a direct role in this process. My gratitude, however, extends beyond this list- those who are not mentioned by name are not forgotten.

First, to my PhD supervisors Chris Kenyon & Xaveer van Ostade: I sincerely thank you for your continuous guidance, enthusiasm and support. It was an ambitious project studying fascinating pathogens with a steep learning curve. I appreciate that you trusted me with important scientific decisions. I learned an incredible amount from your diverse areas of expertise and the many opportunities you both provided to me. As a study clinician, four years ago I would have never expected that my role would blossom into going into the depths of proteomics and STI epidemiology. I will look back at these formative years fondly.

It was a pleasure to work at ITM's Clinical Sciences Department- such a diverse team of talented individuals. Special thanks to the attending doctors and supporting staff (nurses/reception/administration) at the clinic for your diligent support of the SeTPAT study. Also many thanks to Marjan van Esbroek and Tania Crucitti and the laboratory staff Said, Irith, Hilda, Vicky, Benedicte and Nikki who went above and beyond by expertly assisting us with the sample preparation and testing. We could not have done it without you!

Thanks to the proteomics group at UAntwerp for the many analyses performed and valuable contributions. Geert van Raemdonck: my-oh-my how far I have come since your initial teachings of mass spectrometry with an analogy of smashing a Kinder Suprise egg on a wall. You are a fantastic teacher and your commitment to the project was admirable. It was also a pleasure to work with Martin Dom, Mops and Reuben.

Thanks also to Maarten Dhaenens from Ghent University for your dedication to tackling the almost impossible task of detecting *T. pallidum* proteins in urine samples.

I was fortunate to be able to collaborate with two foremost groups of *T. pallidum* experts. From the University of Victoria, Canada: Prof. Caroline Cameron, Dr. Simon Houston & Karen Lithgow and from the University of Masaryk, Czech Republic: Prof. David Šmajs and Dr. Michal Strouhal. You helped us immensely- thanks for sharing your expertise and for your enthusiasm for our project. Our collaborative interactions were inspiring and I appreciate your invitations to visit your labs.

Thanks to the funders who facilitated this project and to the dedicated study participants who made time to attend the study visits.

It will be very difficult to beat the extent of happy comradery I experienced with fellow students and colleagues at the ITM. Wednesday night 'Beer Time' at the Heilig Huisken with the 'Friends of Fermentation' was our Wednesday night social oasis and often the highlight of my week. We laughed, we celebrated and most importantly we were there for one another when we needed it the most. Laura Kuijpers—I met you early on and we hit it off immediately- thank you for the many coffee, café and lunch chats that boosted our morale. The KFC union with Conor Meehan and Franck Dumetz- through thick and thin, you both were there. Many good times were had, almost every day I have a silent chuckle to myself thinking of our adventures together. I learned a lot from you two, scientifically and otherwise. Adam Hendy- one of my favorite PhD memories is that fateful June afternoon we spontaneously went to Cargo Bar to sit in the sun. Charles, the best office mate—thanks for the cookies, coffee and inspiring chats. Patricia, Achilles, Vera, Nico, Irina, Marlene, Sofia and Aya- I will not forget the memories we made.

I am thankful for my teachers and mentors, old and new, who have helped steer me in the right direction along the way: Hugh Ross (Sal Comp), Gerben Schaaf (AMC), Anna Gloyn (Oxford), Martin Grobusch & Saskia Janssen (AMC) and Bouke de Jong & Lut Lynen (ITM).

Thanks to my dear old friends for always being there: my med school friendship trio Shahrzad Deliran, Berber Binnerts & Ling Lu; Jack Chen & Marianne Lammersen; Fay(bert) Leggat & Andrea Becker. I can't wait to see where life will bring us!

I dedicate this thesis to my loving family: my mother Gloria, Kevin & Sandra, Shaelene & family, my Dad Gary & partner Linda and the extended Osbak family in Breton- Grandma Elsie Osbak, Uncle Lanny, Tracy & Rena. From little onwards, they indulged my insatiable curiosity and encouraged me to be myself, to learn and explore in every way they could make possible. Kitchen science experiments, space & science camps, courses, a plethora of books (to name a few!)… endless enthusiasm for all of my chosen paths- even when the wings you gave me flew me across the ocean. Without your emotional and financial support I would not be where I am. I am also thankful for my extended families- from Amsterdam: Anke van der Hoeven, Frans Bus & Carma Tonks and my Belgian family: Pascale Didelez, Bruno, Jana & Jelle Peeters. Thank you for welcoming me into your warm homes.

Last but not least, my fiancé Didier Didelez (aka ).You give me so much joy- thank you for being you. Sharing my days with you is the greatest gift I could wish for.