Short Communication

A preliminary investigation of microsatellite-based genotyping in *Trichomonas vaginalis*.

Marianna Prokopi[a,1], Theodoros Chatzitheodorou[a,2], John P. Ackers[a], C. Graham Clark[a]

[a]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine

Present Addresses: 1 - Department of Cardiology, King's College, University of London; 2 - Microbiology Laboratory, Evangelismos Hospital, Athens, Greece

Corresponding Author and guarantor:

C. Graham Clark, PhD. (graham.clark@lshtm.ac.uk), Department of Pathogen Molecular Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT. Telephone: +44-20-7927-2351; FAX: +44-20-7636-8739

Running Title: Microsatellites in *Trichomonas vaginalis*

Abstract.

The genetic epidemiology of *Trichomonas vaginalis* is poorly understood at present. The recent release of the organism's genome sequence opens the way to investigation of polymorphic markers allowing strain identification. We here report a preliminary analysis of microsatellite loci in *T. vaginalis* and show that this approach holds promise for future studies of infection transmission and organism diversity.

Key Words: *Trichomonas vaginalis*; microsatellite; diversity.

1. Introduction

Trichomonas vaginalis is a widespread cause of human vaginitis and urethritis. It is frequently found in mixed infections with other sexually transmitted diseases[1] and it is associated with an increased risk of acquiring HIV infection in carriers[2,3]. The gold standard for *T. vaginalis* diagnosis is culture, but a number of PCR-based detection methods have been described[4]. In most instances infection is easily diagnosed and treated with metronidazole, although laboratory confirmed cases of clinical metronidazole resistance are well known and on the increase[4].

Given the prevalence and importance of trichomoniasis, it is remarkable how little is still known of the pathogen's genetic diversity and molecular epidemiology. We do not know how many strains circulate in a population at any one time, whether some strains are more likely to cause symptoms, or how often metronidazole resistance appears independently. The presence of *T. vaginalis* is highly suggestive of sexual abuse when present in preadolescent children[5] and is one of the more common

infections acquired by women through sexual assault[6]. However we do not currently have a rapid and reliable way of forensically linking an infection to a potential source. For this, polymorphic markers of high discriminatory value are required.

Initial investigations into *T. vaginalis* genetic diversity used RAPD analysis (Rapid Amplification of Polymorphic DNA) [7]. Drawbacks of this method are a lack of reproducibility between laboratories and a requirement for DNA from axenic organisms, since DNA from any source can act as a template for amplification when using random primer annealing. Subsequently, a method relying on rare restriction enzyme sites to identify polymorphisms among pulsed-field gel separated trichomonad DNA samples was described[8] but it also requires successful cultivation, significant numbers of organisms, and specialised equipment. Recently, methods making use of restriction fragment length polymorphism in PCR amplified genes have been described[9,10] both of which revealed substantial levels of genetic diversity among isolates but are not really adaptable to high throughput analyses.

In 2007, the *T. vaginalis* genome was published[11]. This resource is a potential goldmine for identifying new targets for detecting polymorphism among *T. vaginalis* strains. The most widely used methodologies for detecting variation in eukaryotes are probably those relying on microsatellites. We have undertaken a bioinformatic survey of potential microsatellite loci in the genome data and here report a preliminary experimental analysis of their variability.

2. Materials and Methods.

The *T. vaginalis* genome data deposited in GenBank were searched via the NCBI BLAST tool for all possible combinations of di-, tri-and tetra-nucleotide repeats. Three di-, nine tri-, and eleven tetra-nucleotide containing loci were selected for initial experimental analysis and primers were designed that amplified the chromosomal region containing the repeats (Table 1). A panel of nine *T. vaginalis* isolates of diverse origin was used for initial screening and these were later supplemented by DNA from an additional eight isolates (courtesy of Dr S. Kilvington, University of Leicester). Isolates were grown in medium TYM pH6.0 supplemented with 10% adult bovine serum[12]. DNA was purified using the Puregene Core Kit A (QIAGEN, Hilden, Germany). PCR amplification was performed under standard conditions (30 cycles of denaturation at 94 ˚C for 30sec, annealing at the appropriate temperature for 30sec, and extension at 72 ˚C for 30sec, followed by a final extension of 2min at 72 ˚C) using BioTaq polymerase (Bioline, UK). Products were separated in 2% agarose gels in Tris-borate-EDTA buffer and stained with ethidium bromide. A selection of products was gel purified (Qiagen, UK) and sequenced using Big Dye Terminator sequencing reagents and an ABI3730 DNA Analyzer (Applied Biosystems, UK). None of the final panel of primers used showed identity to sequences in the human genome,

3. Results and Discussion.

A number of the microsatellite loci initially investigated yielded two or more bands in agarose gels when amplified, suggesting the existence of multiple related sequences in the genome, and these were discounted from further analysis. The loci that gave a single band were studied further. Four that gave reliable amplification and exhibited

visible product size variation in agarose gels were selected for sequence analysis (Table 1). In all cases sequence polymorphism was found and variation in microsatellite repeat number was the basis for the size differences observed. It is notable, however, that in some cases polymorphisms of other types or in other locations than just the anticipated microsatellite copy number was observed (Table 1). In most cases these represented single nucleotide polymorphisms (SNPs) or variant microsatellite repeats, but in one case a second set of tandem repeats was also present that exhibited unexpected polymorphism.

No two samples investigated showed the same genotype. Indeed, at locus GAA-1 alone only two shared the same repeat number. The *T. vaginalis* samples investigated were from 4 countries but included six isolated in the same location (Leicester, UK) within two years, indicating that local diversity can also be high.

Over 400 potential microsatellite loci were identified when screening the genome data, so even if only ultimately 5% proved to be useful for polymorphism detection this would be more than adequate to provide the basis for a microsatellite-based strain identification scheme. However, because of our detection of mixed-location length variation in the same products, and given that high throughput microsatellite typing methods rely on detection of PCR product size rather than sequencing, it is possible that some allele assignments based only on size will actually lead to distinct sequences being assigned the same allele number. Sequencing of PCR products is probably not a viable alternative for analysing all loci and all strains in large studies, but we suggest that a pilot sequencing investigation be undertaken to better understand the basis of the size variation before loci are selected for incorporation

into typing schemes. Although only a preliminary investigation, our study does confirm the existence of substantial genetic diversity within *T. vaginalis* and underlines the promise of microsatellite-based typing systems for this species.

After this work was completed a more comprehensive study of microsatellite and SNP variation in *T. vaginalis* was published[13], which showed that microsatellites in this organism are stable under long term culture However, we note that an investigation of the basis of the microsatellite variation by sequencing was not reported in that study.

Authors' contributions

CGC and JPA conceived the experimental design of the study. CGC analysed the data and wrote the first draft of the manuscript. TC carried out preliminary experiments to validate the approach; this formed the basis of a MSc project submitted in 2006. MP undertook the experimental work presented above and analysed the data; these formed the basis of a MSc Project submitted in 2007. JPA and CGC supervised the MSc projects. All authors read and approved the final manuscript. CGC is guarantor of the paper.

Acknowledgments.

We thank Dr Simon Kilvington for providing some of the DNA samples used in this study.

Funding: None.

Conflicts of interest: None declared.

Ethical approval: Not required.

References

1. Allsworth JE, Ratner JA, Peipert JF. Trichomoniasis and other sexually transmitted infections: results from the 2001-2004 National Health and Nutrition Examination Surveys. *Sex Transm Dis* 2009, **36**: 738-44.

2. Mavedzenge SN, Pol BV, Cheng H, Montgomery ET, Blanchard K, de Bruyn G, et al. Epidemiological synergy of *Trichomonas vaginalis* and HIV in Zimbabwean and South African women. *Sex Transm Dis* 2010, **37:** 460-6.

3. McClelland RS. *Trichomonas vaginalis* infection: can we afford to do nothing? *J Infect Dis* 2008, **197:** 487-9.

4. Shafir SC, Sorvillo FJ, Smith L. Current issues and considerations regarding trichomoniasis and Human Immunodeficiency Virus in African-Americans. *Clin Microbiol Rev* 2009; **22**, 37-45.

5. Lewin, LC. Sexually transmitted infections in preadolescent children. *J Pediatr Health Care* 2007, **21**: 153-61.

6. Davies AG, Clay JC. Prevalence of sexually transmitted disease infection in women alleging rape. *Sex Transm Dis* 1992, **19:** 298-300.

7. Rojas L, Fraga J, Sariego I. Genetic variability between *Trichomonas vaginalis* isolates and correlation with clinical presentation. *Infect Genet Evol* 2004, **4:** 53-8.

8. Upcroft JA, Delgadillo-Correa MG, Dunne RL, Sturm AW, Johnson PJ, Upcroft P. Genotyping *Trichomonas vaginalis*. *Int J Parasitol* 2006, **36:** 821-8

9. Crucitti T, Abdellati S, Van Dyck E, Buvé A. Molecular typing of the actin gene of *Trichomonas vaginalis* isolates by PCR-restriction fragment length polymorphism. *Clin Microbiol Infect* 2008, **14:** 844-52.

10. Meade JC, de Mestral J, Stiles JK, Secor WE, Finley RW, Cleary JD, et al. Genetic diversity of *Trichomonas vaginalis* clinical isolates determined by *Eco*RI restriction fragment length polymorphism of heat-shock protein 70 genes. *Am J Trop Med Hyg* 2009, **80:** 245-51.

11. Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao Q, et al. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science* 2007, **315:** 207-12.

12. Clark CG, Diamond LS. Methods for cultivation of luminal parasitic protists of clinical importance. *Clin Microbiol Rev* 2002, **15**: 329-41.

13. Conrad M, Zubacova Z, Dunn LA, Upcroft J, Sullivan SA, Tachezy J, et al. Microsatellite polymorphism in the sexually transmitted human pathogen *Trichomonas vaginalis* indicates a genetically diverse parasite. *Mol Biochem Parasitol* 2011, **175**: 30-8.

Locus Strain	GAA-1	GAA-2 #	GACC	TAC	origin
G3 *	(GAA)97	(GAA)5(AAA)(GAA)6	(-A)...(GACC)5...(TATT)4	(TAC)(TAA)(TAC)6	UK
CDC337	(GAA)55	(GAA)8	(-A)...(GACC)4...(TATT)3	(TAC)8	USA
RU393	(GAA)36	(GAA)4	(-A)...(GACC)5...(TATT)4	(TAC)10	USA
97F1106	(GAA)12	(GAA)5(AAA)(GAA)6	(-A)...(GACC)4...(TATT)3	(TAC)(TAA)(TAC)6	UK$
LUMP1910	(GAA)39	(GAA)6	(-A)...(GACC)5...(TATT)5	(TAC)(TAA)(TAC)6	UK
96F1926	(GAA)66	(GAA)9	(-A)...(GACC)5...(TATT)5	(TAC)(TAA)(TAC)6	UK$
FB2911	(GAA)22	(GAA)2(AAA)(GAA)5(AAA)(GAA)5	(-A)...(GACC)5...(TATT)4		UK
797	(GAA)27		(TT)...(GACC)5...(TATT)5	(TAC)(TAA)(TAC)6	Unknown
94F0264	(GAA)66	(GAA)5(AAA)(GAA)3	(TT)...(GACC)5...(TATT)5	(TAC)8	UK$
296	(GAA)54	(GAA)5(AAA)(GAA)3	(TT)...(GACC)5...(TATT)5	(TAC)8	Unknown
TV/PHL/3	(GAA)78	(GAA)4(AAA)(GAA)3	(TT)...(GACC)5...(TATT)5		UK$
TV/PHL/4	(GAA)74	(GAA)5(AAA)(GAA)3	(TT)...(GACC)5...(TATT)5		UK$
Saudi1	(GAA)53		(-A)...(GACC)4...(TATT)4		Saudi Arabia
Saudi3	(GAA)13		(-A)...(GACC)5...(TATT)4		Saudi Arabia
94M2	(GAA)69	(GAA)3			Israel
1910L	(GAA)53	(GAA)2(AAA)(GAA)4	(TT)...(GACC)4...(TATT)4		UK
TV/PHL/2		(GAA)5(AAA)(GAA)3	(TT)...(GACC)5...(TATT)4		UK$

Table 1. Structure of the microsatellite loci studied. The sequences obtained for selected isolates are shown with the unit sequence in parentheses followed by the copy number where greater than one. The primer pairs used were: GAA-1F: TATGGTCCTTTCCCTTCCTG, GAA-1R: GCTGCAGCATGAAGACTTTG; GAA-2F: GATCAGGAAACTCTGGAA, GAA-2R: TCTTCTTCTTCAAAAGCATACTTC; GACCF: CGGCTAGGTGAGAAACTAGG, GACCR: AATTTCTCACAGAAGTGACGC; TACF: GCAAAAGGCGACTATTATGC, TACR: CACGTTTAAAGCCTACCATAGTG. The annealing temperature in all cases was 57 ˚C.

* G3 is the genome sequence reference strain and the sequence is taken from the following accession numbers: NW_001581816 (GAA-1), NW_001581816 (GAA-2), NW_001579018 (GACC), NW_001820766 (TAC).

All start with (GAA)4(AAA)(GAA)2(AAA) followed by the structures listed

$ All isolated in 1996-7 in Leicester, UK