Whole-genome sequences of *Chlamydia trachomatis* directly from clinical samples without culture

Helena M.B. Seth-Smith,[1,11] Simon R. Harris,[1] Rachel J. Skilton,[2] Frans M. Radebe,[3] Daniel Golparian,[4] Elena Shipitsyna,[5] Pham Thanh Duy,[6] Paul Scott,[1] Lesley T. Cutcliffe,[2] Colette O'Neill,[2] Surendra Parmar,[7] Rachel Pitt,[8] Stephen Baker,[6] Catherine A. Ison,[8] Peter Marsh,[9] Hamid Jalal,[7] David A. Lewis,[3,10] Magnus Unemo,[4] Ian N. Clarke,[2] Julian Parkhill,[1] and Nicholas R. Thomson[1,11]

[1]Pathogen Genomics, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, United Kingdom; [2]Molecular Microbiology Group, Faculty of Medicine, University of Southampton, Southampton General Hospital, Southampton SO16 6YD, United Kingdom; [3]Centre for HIV and Sexually Transmitted Infections, National Institute for Communicable Diseases, National Health Laboratory Service, Johannesburg 2131, South Africa; [4]WHO Collaborating Centre for Gonorrhoea and Other STIs, National Reference Laboratory for Pathogenic Neisseria, Örebro University Hospital, Örebro SE-701 85, Sweden; [5]Laboratory of Microbiology, D.O. Ott Research Institute of Obstetrics and Gynaecology, 199034, St. Petersburg, Russia; [6]Oxford University Clinical Research Unit, Wellcome Trust Major Overseas List, The Hospital for Tropical Diseases, Ho Chi Minh City, Q5, Vietnam; [7]Clinical Microbiology and Public Health Laboratory, Health Protection Agency, Addenbrooke's Hospital, Cambridge CB2 0QQ, United Kingdom; [8]Sexually Transmitted Bacteria Reference Laboratory, Health Protection Agency, Colindale, London NW9 5HT, United Kingdom; [9]Health Protection Agency, Public Health Laboratory Southampton, Southampton General Hospital, Southampton SO16 6YD, United Kingdom; [10]Department of Internal Medicine, Faculty of Health Sciences, University of the Witwatersrand, Parktown 2193, Johannesburg, South Africa

The use of whole-genome sequencing as a tool for the study of infectious bacteria is of growing clinical interest. *Chlamydia trachomatis* is responsible for sexually transmitted infections and the blinding disease trachoma, which affect hundreds of millions of people worldwide. Recombination is widespread within the genome of *C. trachomatis*, thus whole-genome sequencing is necessary to understand the evolution, diversity, and epidemiology of this pathogen. Culture of *C. trachomatis* has, until now, been a prerequisite to obtain DNA for whole-genome sequencing; however, as *C. trachomatis* is an obligate intracellular pathogen, this procedure is technically demanding and time consuming. Discarded clinical samples represent a large resource for sequencing the genomes of pathogens, yet clinical swabs frequently contain very low levels of *C. trachomatis* DNA and large amounts of contaminating microbial and human DNA. To determine whether it is possible to obtain whole-genome sequences from bacteria without the need for culture, we have devised an approach that combines immunomagnetic separation (IMS) for targeted bacterial enrichment with multiple displacement amplification (MDA) for whole-genome amplification. Using IMS-MDA in conjunction with high-throughput multiplexed Illumina sequencing, we have produced the first whole bacterial genome sequences direct from clinical samples. We also show that this method can be used to generate genome data from nonviable archived samples. This method will prove a useful tool in answering questions relating to the biology of many difficult-to-culture or fastidious bacteria of clinical concern.

[Supplemental material is available for this article.]

Chlamydia trachomatis is a pathogen of global importance as the most common bacterial sexually transmitted infection (STI) (WHO 2011), and is also responsible for trachoma, the leading cause of infectious blindness worldwide (Mariotti et al. 2009; WHO 2012). Urogenital chlamydial infections usually manifest as urethritis and cervicitis, but are often asymptomatic and can result in severe complications and sequelae such as pelvic inflammatory disease, tubal damage, and infertility if untreated. In addition, some *C. trachomatis* infections are invasive, causing the disease lymphogranuloma venereum (LGV) (Burgoyne 1990).

[11]Corresponding authors
E-mail nrt@sanger.ac.uk
E-mail hss@sanger.ac.uk
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.150037.112.

C. trachomatis is an obligate intracellular pathogen with a specialized biphasic developmental cycle. The infectious elementary bodies (EBs) are taken up by the host cell into a cytoplasmic vacuole called an inclusion, where they differentiate into the actively replicating form, known as reticulate bodies (RBs). The developmental cycle is completed when RBs differentiate back into metabolically inert EB particles and are released from the host cell by lysis. Thus *C. trachomatis* requires tissue culture for in vitro growth, a technique which is technically challenging and time consuming. While cell culture used to be the "gold standard" for the laboratory diagnosis of *C. trachomatis* infections, this has been superseded by much more rapid and sensitive nucleic acid amplification tests (NAATs) (for review, see Skidmore et al. 2006).

For epidemiological surveillance, *C. trachomatis* strains have traditionally been classified into serovars based on the major outer membrane protein (MOMP), which represents the major surface

antigen (Stephens et al. 1982; Wang et al. 1985). Currently *ompA* genotyping, based on the gene encoding MOMP, is more commonly performed (for review, see Pedersen et al. 2009), with *ompA* genotypes A–C associated with trachoma, D–K with urogenital infections, and L1–L3 with LGV. Recent publications have confirmed previous findings that *ompA* is not a reliable marker of phylogeny (Millman et al. 2001; Gomes et al. 2004; Brunelle and Sensabaugh 2006), due to extensive recombination within the genomes of *C. trachomatis* strains (Jeffrey et al. 2010; Harris et al. 2012; Joseph et al. 2012), and that to fully understand the population structure and patterns of infection it is essential to determine the whole-genome sequence. The genome of *C. trachomatis* comprises a chromosome of 1.0 Mb and a plasmid of 7.5 kb which have been found to be highly conserved between strains, with few indels and no variably present genomic islands identified to date (Stephens et al. 1998; Carlson et al. 2005; Thomson et al. 2008; Seth-Smith et al. 2009; Jeffrey et al. 2010; Unemo et al. 2010; Somboonna et al. 2011; Harris et al. 2012). The use of full-genome sequence data in hospital settings has been demonstrated in recent studies (Köser et al. 2012b; Snitkin et al. 2012) and promises to revolutionize epidemiology and clinical microbiology (for review, see Köser et al. 2012a). Obtaining these data rapidly is a new challenge, particularly pertinent in the study of difficult-to-culture or fastidious bacteria.

Until now, cell culture has been necessary to generate sufficient *C. trachomatis* DNA for genome sequencing (Stephens et al. 1998; Carlson et al. 2005; Thomson et al. 2008; Seth-Smith et al. 2009; Jeffrey et al. 2010; Unemo et al. 2010; Somboonna et al. 2011; Harris et al. 2012; Joseph et al. 2012). Using clinical samples as a starting material, several months of passaging is often required. Not all strains of *C. trachomatis* culture equally well, with a subset of strains failing to grow in culture despite being detectable by other assays (for review, see Ridgway and Taylor-Robinson 1991). This implies that the growth of strains itself may impose a selective bias on the strains whose genomes we are able to sequence. It is clear that a rapid, simple, culture-independent technique for generating DNA for whole-genome sequencing is required.

Whole-genome amplification (WGA), in particular the technique of multiple displacement amplification (MDA), can be used on very low concentrations of starting material to generate the quantities of DNA required for sequencing and has been shown to provide complete genome coverage from bacterial DNA samples (Rodrigue et al. 2009; Chaparro et al. 2011). MDA is an isothermal DNA amplification process using ø29 polymerase and random hexamers with phosphorothioate modification, meaning that this technique can be used to amplify any DNA sample (Dean et al. 2001; Hosono et al. 2003). We aimed to determine whether it was possible to generate complete, accurate genome sequences of *C. trachomatis* from clinical samples with the use of MDA. To reduce the associated contamination and enrich for *C. trachomatis*, we combined MDA with immunomagnetic separation (IMS), an antibody-based approach for targeted enrichment of cells. IMS has been used to enrich for bacteria and remove PCR inhibitors in previous studies on genera including *Listeria* (Skjerve et al. 1990), *Mycobacterium* (Grant et al. 1998), *Escherichia coli* O157:H7 (Fratamico et al. 1992), *Actinobacillus* (Angen et al. 2001), and also *Chlamydia* (Niesters et al. 1991; Hedrum et al. 1992). Using this combined approach (IMS-MDA) we were able to generate whole-genome sequences of *C. trachomatis* directly from discarded clinical swabs, without the need for cell culture.

Results

Accurate genome sequences from amplified *C. trachomatis* DNA

To assess the sensitivity and accuracy of WGA of *C. trachomatis* DNA using MDA, a range of dilutions of *C. trachomatis* serovar L2 genomic DNA was used as substrate for amplification. DNA was extracted from a single well of a 24-well tissue culture tray (24WT) and, as a pre-dilution control, sequenced on an Illumina GAII machine following standard protocols (see Methods) using 11 indexed sequence adapters, or tags, per lane. The resulting data were assembled using Velvet and manual improvement to generate a reference sequence comprising two contigs, with an unassembled gap in the *tarp* (translocated actin recruiting phosphoprotein) gene, which contains a repetitive motif.

The same L2 genomic DNA was then subjected to a series of dilutions, and MDA reactions were performed on 1 μL of each dilution. The number of genome copies was determined using quantitative PCR (qPCR) with a TaqMan probe targeting the single-copy chromosomal *ompA* gene. Samples were sequenced on an Illumina HiSeq machine using nine tags per lane. To assess the coverage and accuracy of the resulting data, the sequencing reads were mapped in silico to the assembled reference L2 genome sequence, with the control reads self-mapped for comparison, and the variation between the reference and the sequenced samples was determined (Table 1).

The results show that complete coverage of the genome was achieved from all samples for which the starting material comprised at least 4800 input genome copies, giving 1,500,000 post-MDA copies. Below this, there was a dramatic drop-off in the number of post-MDA genome copies, resulting in a very low sequencing yield and only 26% coverage of the genome from an input of 3500 genome copies. This provides an approximate lower limit for the number of chromosomal copies required to generate accurate genome sequencing using MDA. In all cases, the complete plasmid was covered in large depth by sequence reads.

For the samples that mapped to the complete chromosome (from Dilution1 to Dilution7), mean chromosome coverage levels varied with yield, but in all cases were more than sufficient for accurate base calling. However, as the input DNA decreased, the variation in coverage level across the genome increased. This is illustrated by an increase in the coefficient of variation (CV = standard deviation/mean) of the coverage as the number of input genome copies decreased. These factors underline the need to use sequencing technologies that yield high numbers of reads per sample in order to generate sufficient depth of coverage to ensure accurate sequence generation. Coverage of the plasmid in the amplified samples was far greater than that of the chromosome, ranging from 79 to 213 times the chromosomal coverage (control = 10.5 times). This is due to preferential amplification of the small, circular plasmid, and means that it is not possible to estimate the plasmid copy number per chromosome from amplified samples.

The accuracy of the sequencing was determined through analysis of single nucleotide polymorphisms (SNPs), insertions, deletions, and heterogeneous sites after mapping against the reference assembly. All samples which mapped to 100% of the chromosome provided completely accurate genome sequences (chromosome and plasmid) with the exception of the sequence data derived from 4800 genome copies, for which eight chromosomal SNPs were identified (representing >99.99% accuracy). No insertions or deletions were identified in any of the samples giving 100% genome coverage, and only a single base insertion was identified in the data

Table 1. Analysis of sequence data from MDA-amplified dilutions of *C. trachomatis* DNA

Sample	Input genome copies (mean)	Post-MDA genome copies/μL (mean)	Read length (bp)	Yield (reads)[a]	Reads mapped[b]	Chromosome covered (%)	Chromosome coverage (mean)	Chromosome CV (std/mean)	Plasmid: chromosome coverage (ratio)	SNPs	Insertions	Deletions	Heterogeneous sites[c]
Control L2	195,000,000	N/A	54	6,897,022	4,503,991	100	217.1	0.16	10.5	0	0	0	5
Dilution1	3,000,000	370,000,000	75	27,763,984	18,807,755	100	862.8	0.34	79.4	0	0	0	10
Dilution2	2,200,000	230,000,000	75	27,926,608	16,381,245	100	536.6	0.36	166.6	0	0	0	14
Dilution3	315,000	350,000,000	75	36,055,636	23,407,856	100	1065.7	0.40	81.1	0	0	0	7
Dilution4	260,000	210,000,000	75	28,874,478	12,188,418	100	529.4	0.48	91.6	0	0	0	11
Dilution5	40,500	180,000,000	75	58,344,102	39,726,242	100	1477.8	0.47	130.2	0	0	0	9
Dilution6	23,500	48,000,000	75	39,659,067	22,769,406	100	645.9	0.58	213.8	0	0	0	21
Dilution7	4800	1,500,000	75	42,672,678	25,470,880	100	1117.9	1.92	89.2	8	0	0	170
Dilution8	3500	54,000	75	20,666	10,820	26	0.3	1.90	200.4	0	1	0	140

(N/A) Control sequencing without MDA.
[a]Yield after removal of reads mapping to human and PhiX.
[b]Reads mapping to L2 assembly, excluding matches to host cell mitochondria which fractionate with *Chlamydia* during DNA extraction from tissue culture.
[c]Number of sites for which a base call could not be made due to heterogeneity.

derived from Dilution8 with 3500 genome copies, although these data were not complete across the whole genome, with the insertion identified in the plasmid sequence. Heterogeneous sites are defined here as base positions at which clear base calling cannot be performed due to a lack of consensus in the mapped reads. Our analysis of these showed that lower levels of pre-MDA input DNA caused higher levels of heterogeneity, and also led to the incorrect base calls (SNPs) seen in the Dilution7 sample containing 4800 genome copies. These inaccuracies are located in regions of high read coverage with the variant base present in >75% of reads in all cases, indicating that an incorrect base may have been incorporated during amplification of these sites early in the MDA process.

Generation of completely accurate genome sequences can be achieved using MDA. An initial input of between 4800 and 23,500 genome copies is required, equivalent to between 1,500,000 and 48,000,000 copies post-MDA. These data also show that complete genome sequences can be derived from DNA extracted from a single well of a 24WT with a sufficient coverage (217×) to allow high quality de novo assembly. This observation greatly reduces the need for extended laboratory passage of strains prior to DNA extraction for genome sequencing.

Assessing the *C. trachomatis* content of clinical samples

To determine the feasibility of amplifying genomes directly from clinical samples using MDA we determined the DNA composition of *C. trachomatis* positive clinical samples. Addenbrooke's Hospital in Cambridge and Peterborough City Hospital provided samples from their routine diagnostic service. These included swabs and urines from males and females, with the diagnostic sample taken directly into lysis buffer for use with the Gen-probe Aptima or Abbott m2000 systems. DNA was extracted from these buffers using the Qiagen QIAmp DNA mini kit or Promega Wizard genomic DNA purification kit and analyzed for total DNA content and *C. trachomatis* DNA content.

Quantitation showed highly variable total DNA content in the samples from below the detection limit of 100 pg/μL to a maximum of 161 ng/μL (Table 2). *C. trachomatis* DNA was present at very low levels, often below the detection limit of 1000 genome copies/μL, comprising a maximum of 0.6% of the total DNA. This indicates that the samples include a large amount of non-chlamydial DNA, from the host or other resident microbiota.

We performed MDA directly on these DNA extracts (1 μL aliquot) to assess the potential for direct amplification of *C. trachomatis* genomes from a clinical sample. Although an increase in total DNA was seen in most samples and the number of *C. trachomatis* genome copies increased in some cases, the relative proportion of *C. trachomatis* DNA fell in many instances (Table 2), suggesting preferential amplification of the other DNA species within these complex samples. To test this phenomenon of a reduction in the proportion of *C. trachomatis* DNA after MDA, we performed an experiment with *C. trachomatis* DNA spiked into

Table 2. Analysis of DNA extracts from clinical samples

Sample	Source	DNA concentration (pg/μL)	*Ct* genome copies/μL	Initial % *Ct* DNA	Post-MDA DNA concentration (pg/μL)	Post-MDA *Ct* genome copies/μL	Post-MDA % *Ct* DNA
GQ1	Urine	252	<1000	ND	62.6[a]	<1000	ND
GQ2	Urine	166	<1000	ND	49.6[a]	<1000	ND
GQ3	Vaginal	<100	<1000	ND	63,800	<1000	ND
GQ4	Urine	576	<1000	ND	70.8[a]	<1000	ND
GQ5	Cervical	974	<1000	ND	336,000	50,500	0.015
GQ6	Cervical	746	<1000	ND	40[a]	<1000	ND
GQ7	Vaginal	1560	<1000	ND	284,000	<1000	ND
GQ8	Cervical	3280	1900	0.06	39.2[a]	<1000	ND
GQ9	Cervical	1280	<1000	ND	34.4[a]	<1000	ND
GQ10	Urine	454	<1000	ND	50.8[a]	<1000	ND
GW1	Urethral	15,600	81,800	0.5	452,000	429,000	0.09
GW2	Vaginal	139,000	87,500	0.6	528,000	13,100	0.002
GW3	Vaginal	<100	<1000	ND	3400	<1000	ND
GW4	Urethral	17,700	<1000	ND	496,000	<1000	ND
GW5	Vaginal	161,000	<1000	ND	508,000	<1000	ND
GW6	Urethral	1580	<1000	ND	474,000	<1000	ND
AQ1	Urine	<100	<1000	ND	2680	<1000	ND
AQ2	Urine	118	<1000	ND	738,000	<1000	ND
AQ3	Urine	1380	<1000	ND	640,000	<1000	ND
AQ4	Urine	<100	<1000	ND	742,000	<1000	ND
AQ5	Urine	<100	<1000	ND	30,200	<1000	ND
AQ6	Cervical	1330	<1000	ND	652,000	7600	0.001
AQ7	Urine	<100	<1000	ND	648,000	<1000	ND
AQ8	Urine	314	<1000	ND	734,000	<1000	ND
AQ9	Urine	9860	<1000	ND	854,000	<1000	ND
AQ10	Urine	106	<1000	ND	60,400	<1000	ND
AW1	Vag/Ure	<100	<1000	ND	570,000	<1000	ND
AW2	Vag/Ure	11,000	1000	0.009	486,000	5180	0.001
AW3	Urine	116	<1000	ND	604,000	<1000	ND
AW4	Vag/Ure	71,600	1050	0.001	440,000	1720	0.0004
AW5	Urine	33,000	<1000	ND	666,000	6540	0.0001
AW6	Urine	208	<1000	ND	121,000	<1000	ND

(*Ct*) *C. trachomatis*. (ND) Not determined. (GQ) Gen-probe samples extracted through Qiagen columns. (GW) Gen-probe samples extracted by Wizard. (AQ) Abbott samples extracted through Qiagen columns. (AW) Abbott samples extracted by Wizard.
[a]Samples appear to have failed to amplify.

carrier DNA at a range of concentrations. This showed that the percentage reduction occurs in all cases when the target DNA is present over the tested range of 0.1%–21% (Supplemental Table S1).

A second set of clinical samples was obtained from Addenbrooke's Hospital for direct sequencing. These were also discarded *C. trachomatis*-positive Gen-probe Aptima samples, from which the DNA was extracted using the Qiagen minikit. Total DNA concentrations showed that several of the samples did not contain sufficient input DNA required for Illumina sequencing (minimum input 1 μg). Of the samples which met this criterion, MDA was performed on 1 μL of the extracted DNA. Following quantification, both the pre- and post-amplification samples were sequenced. The resulting reads were mapped to a completed high-quality reference *C. trachomatis* chromosome (urogenital strain F/SW4, EMBL accession HE601804), and the results are shown in Table 3.

The results show that in all samples the proportion of reads matching regions within the human genome sequence was in excess of 92%. The depth of *C. trachomatis* coverage, as assessed by mapping to a reference genome, was very low and did not represent coverage of the entire genome, making it impossible to accurately call sequence bases. Amplification of the sample did not consistently or substantially improve the results, and in most cases only made the sequencing more uneven across the genome as indicated by the increased CV values. While it is possible that deeper sequencing of these samples would yield sufficient data to generate accurate sequences, this would be prohibitively expensive and not scalable. These data indicate that an enrichment step is needed to increase the relative percentage of *C. trachomatis* target DNA in order to obtain accurate and cost effective *C. trachomatis* genome sequences directly from clinical samples.

Developing IMS-MDA for *C. trachomatis*

We investigated an affinity-based technique, IMS, for targeted enrichment of *C. trachomatis* from clinical samples. This technique is appropriate for swabs taken directly into transport medium such as *Chlamydia* transport medium (CTM) or viral transport medium (VTM), which maintain EBs intact. IMS uses antibodies associated with magnetic beads to bind intact *Chlamydia*, with wash steps designed to remove contaminating material, enabling enrichment of the target species. We used a commercially available anti-*Chlamydia* mouse IgG primary antibody (IMAGEN *Chlamydia*, Oxoid) against *C. trachomatis* lipopolysaccharide (LPS), which binds to all serovars of *C. trachomatis* and has been tested for cross-reactivity against many other microbial species including *Lactobacillus lactis*, *Mycoplasma* spp., *Neisseria gonorrhoeae*, and *Gardnerella vaginalis* (IMAGEN *Chlamydia* booklet; Thornley et al. 1983, 1985). LPS is present at ~34,000 molecules per EB (Su et al. 1990), creating a high density target for antibody binding. This antibody was used with an anti-mouse IgG sheep secondary antibody conjugated to magnetic beads (Dynabead, Dynal, Invitrogen), using the manufacturer's protocol (Dynal). Briefly, the primary and secondary antibodies were first allowed to bind to each other, before the test sample was added and a series of washes was performed. MDA was performed directly after IMS, with the first incubation at 95°C used for bacterial lysis and denaturation of the genomic DNA.

To determine the efficacy of this approach, IMS-MDA was performed on a serial dilution of urogenital *C. trachomatis* strain D/314 from culture, diluted in CTM (Remel M4RT). The number of infectious particles in the input sample was estimated by infection of McCoy cells, and the total number of genome copies present in the input sample was determined by qPCR as an approximation of the number of *Chlamydia* within each sample. Perhaps unsurprisingly there was considerable disparity between these values, with the sample (100 μL) representing 10^{-3} dilution containing ~96 infectious particles and ~10,000,000 genome copies. This disparity is likely to reflect the inefficiency with which *C. trachomatis* infects McCoy cells and the presence of dead or noninfectious *Chlamydia* that register as genome copies.

Genome copies were assayed after the IMS procedure and after IMS-MDA. To compare with the values above, the number of genome copies in the 10^{-3} dilution sample after IMS was 3,805 ± 248, indicating the number of intact, DNA-containing bacteria recovered by the antibody binding. MDA generated sufficient DNA for sequencing (>5,000,000 copies/μL) from samples at dilutions 10^{-1} to 10^{-5} (Fig. 1), even though the number of genome copies recovered from the 10^{-5} sample after IMS was below the detection limit. Protocol variations, including use of an alternative primary antibody against MOMP, had no effect on the efficiency of this protocol (data not shown).

This experiment indicates that IMS-MDA is successful at recovering *C. trachomatis* from CTM, and could thus potentially generate DNA for genome sequencing directly from clinical

Table 3. Analysis of sequencing data from DNA extracts from clinical samples, with and without MDA

Sample	Amplified	DNA concentration (pg/μL)	*Ct* genome copies/μL	Total yield for tag (kb)	Reads matching *Ct* genome (%)	Chromosome coverage (%)	Depth of coverage (mean)	Depth of coverage (std)
GA3	N	1300	1300	1,698,872	1.1	49.4	0.9	11.7
	Y	8000	<1000	1,834,073	0.8	20.7	0.3	3.3
GA4	N	1300	<1000	1,697,929	0.3	44.7	1.1	14.0
	Y	27,000	<1000	2,250,286	0.1	16.1	0.6	5.3
GA5	N	1900	2300	1,422,125	2.8	80.0	2.1	13.3
	Y	24,000	<1000	1,610,920	0.3	8.3	0.1	2.6
GA9	N	8200	79,000	1,812,022	3.0	94.5	3.5	10.7
	Y	37,000	10,000	1,771,792	5.2	80.6	1.9	3.6
GA10	N	1000	<1000	NS				
	Y	6000	<1000	1,742,050	0.1	7.0	0.1	3.2
GA11	N	3700	2500	1,957,726	0.1	9.2	0.3	17.0
	Y	55,000	<1000	1,811,183	0.2	9.8	0.2	3.1
GA12	N	3500	<1000	1,115,405	0.1	7.0	0.2	6.0
	Y	<100	<1000	1,518,803	4.7	78.2	1.8	3.6

All samples were run on Illumina HiSeq with a read length of 100 bp and 24 samples per lane. Sample sources are unknown. (*Ct*) *C. trachomatis*.

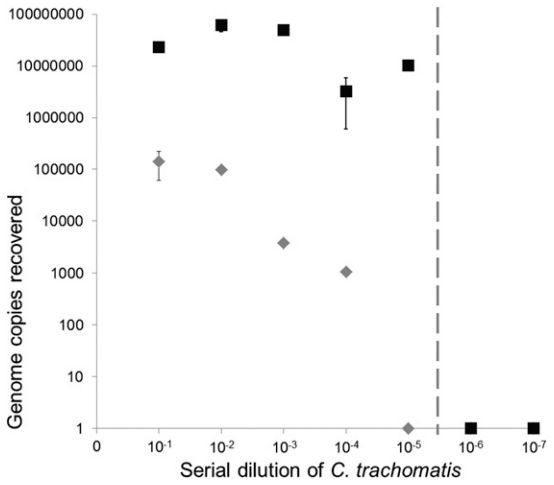

Figure 1. Recovery of *C. trachomatis* DNA following IMS and IMS-MDA on a serial dilution of *C. trachomatis* EBs. Post-IMS values are indicated by the gray diamonds and represent total genome copies recovered by IMS. The DNA recovered from the −5 dilution sample was below the detection limit of 1000 genome copies/microliter. Post-IMS-MDA values are indicated by black squares and represent genome copies/microliter. Error bars indicate standard deviation from duplicate experiments. The dotted gray line indicates the cut-off load, below which IMS-MDA produces insufficient DNA for sequencing.

samples. The success of the approach appears to be dependent on the initial load of *C. trachomatis* present in the sample.

Complete genome sequences directly from clinical samples

To test whether IMS-MDA can be applied to clinical samples containing high levels of human cells and other microbiota, we applied the above protocol to discarded routine *C. trachomatis*-positive swab samples. Eighteen samples in CTM (Remel M4RT), all diagnosed positive by qPCR (Jalal et al. 2006), were obtained from Addenbrooke's Hospital in two batches. All samples were processed by IMS-MDA and quantified by qPCR to estimate the output number of genome copies. From the first batch (seven samples; swab1–3, 5–8), all samples were sequenced, following standard multiplex Illumina sample preparation and sequencing protocols. For the subsequent batch (11 samples; swabB1–11), four samples with the highest amount of *C. trachomatis* DNA were sequenced. For the analysis of the resultant sequence data all reads were mapped against the complete genome of STI strain F/SW4 (EMBL accession HE601804).

The mean depth of coverage was highly variable between samples, with samples containing fewer *C. trachomatis* genome copies post IMS-MDA providing less *C. trachomatis* genome data (Table 4). Of the five samples generating reads covering >99% of the chromosome, all showed a mean depth of coverage of >38× (Table 4), allowing confident base calling. These samples all contained >10,000,000 genome copies/μL post-IMS-MDA, whereas swab7 and swabB11 produced 5000 and 300,000 genome copies/μL post-IMS-MDA, respectively, which was found to be insufficient for production of a complete genome. In agreement with the previous data, it seems that an excess of 1,500,000 copies/μL is required to successfully generate genomic data. Use of alternative primary antibodies did not improve recovery of *C. trachomatis* DNA after IMS-MDA (data not shown). The diagnostic qPCR CT (cycle threshold) value was not found to be indicative of IMS-MDA success, indicating that this clinical assessment of chlamydial load is not relevant in this context. IMS-MDA is dependent on the presence of intact *Chlamydia* whereas CT value reflects the total amount of chlamydial DNA present in the sample. The integrity of the samples in this respect may depend on sample type, patient-to-patient variation, how the sample was collected, and how the sample was stored in the clinic.

From 18 clinical samples, five complete genome sequences were generated. These data indicate that IMS-MDA can be used on clinical samples to produce a reliable genome sequence. These genomes represent the first bacterial genomes produced directly from clinical samples, without culture.

Assembly of novel genome sequences generated by IMS-MDA

Sequence data generated using MDA has a more uneven depth of coverage than that generated from cultured samples, as seen above, through semi-random over-amplification of regions of the genome (Pinard et al. 2006). This is reflected in chromosome CV values (Table 4), as defined above, and can be visualized by mapping read data from a cultured control strain (Table 1, top row) against a reference genome of *C. trachomatis* and comparing this with mapped read data from the five successfully sequenced clinical samples above (Supplemental Fig. S1).

While genome variation can be accurately determined by mapping against an appropriate reference and calling SNPs, uneven genomic coverage is potentially problematic for de novo genome assembly as it violates the assumptions of many assembly algorithms. Assembly of data sequenced following IMS-MDA of clinical samples is further complicated by the potential presence of low-level contaminant DNA either from the sample or from contaminated reagents (Blainey and Quake 2010).

With the recent rapid advances in the fields of single cell sequencing (for review, see Lasken 2012), it is becoming more commonplace to assemble data produced by the sequencing of MDA-amplified DNA. This has led to the development of informatics techniques that allow improved assembly of such sequence data with uneven coverage, including the SPAdes assembly software (Bankevich et al. 2012). We applied the single cell option of the SPAdes assembler to produce de novo assemblies of the five clinical samples sequenced after IMS-MDA (above). For comparison we also assembled the above data from the cultured control L2 strain using the same program, but without the single cell option (Table 5).

IMS enriches samples for *C. trachomatis* but does not produce pure samples of the target bacterium, which is apparent within the assembly as a number of small contigs of contaminant species were derived from all five clinical samples. These include contigs with similarity to *Lactobacillus* and *Gardnerella* spp., two common commensals of the vagina (Ravel et al. 2011), and *Ralstonia* spp., a known kit contaminant (Kulakov et al. 2002; Pride et al. 2012). The high level of conservation of the *C. trachomatis* genome across the species allows these contaminant contigs to be easily identified and removed; we achieved this using abacas (Assefa et al. 2009) to order the contigs against a reference *C. trachomatis* genome (F/SW4).

The assembly statistics for the control and clinical samples are shown in Table 5. Assembly of the sequence data from the cultured control sample produced two chromosomal contigs totaling 1,032,628 bp. The only contig breaks coincided with the two identical rRNA operons, which collapse into one contig during assembly of short read data. Despite the uneven coverage in the

Table 4. Performance of IMS-MDA and whole-genome sequencing on clinical samples

Sample	Source	CT value (MOMP)[a]	Genome copies post-IMS-MDA (per mL)[b]	Machine	Read length (bp)	Tags per lane	Total yield for tag (kb)	*Ct* chromosome covered (%)	Depth of coverage (mean)	Chromosome CV (std/mean)	Reads mapping to human genome (%)	*ompA* genotype
Swab1	Cervical	21.4	4500	GAII	54	12	19,577	**4.3**	**0.0**	**4.85**	**77.8**	
Swab2	Vaginal	17.2	<1000	GAII	54	12	227,744	**86.6**	**2.5**	**0.79**	**94.4**	
Swab3	Vaginal	19.7	<1000	GAII	54	12	192,239	**8.0**	**0.1**	**4.24**	**78.5**	
Swab5	Vaginal	17.8	77,000,000	GAII	54	12	250,571	100.0	98.3	0.46	40.6	F
Swab6	Vaginal	19.7	44,000,000	GAII	54	12	157,199	99.9	63.7	0.51	40.2	E
Swab7	Cervical	18.5	5000	GAII	54	12	287,815	**92.7**	**3.5**	**0.75**	**26.5**	
Swab8	Cervical	23.6	<1000	GAII	54	12	18,536	**0.4**	**0.0**	**15.05**	**92.5**	
SwabB1	Vaginal	23.1	110,000,000	HiSeq	75	21	871,541	100.0	86.4	0.54	10.4	F
SwabB2	Cervical	24.5	<1000	NS								
SwabB3	Vaginal	25.9	<1000	NS								
SwabB4	Female	23.1	11,000,000	GAII	54	12	331,977	99.9	35.9	0.63	38.8	E
SwabB5	Cervical	17.6	<1000	NS								
SwabB6	Vaginal	22.4	2700	NS								
SwabB7	Cervical	20.4	18,000	NS								
SwabB8	Vaginal	22.1	15,000,000	GAII	54	12	294,782	100.0	60.9	0.54	32.0	F
SwabB9	Female	25.2	<1000	NS								
SwabB10	Male	23.5	<1000	NS								
SwabB11	Vaginal	14.8	300,000	HiSeq	75	21	688,148	**78.0**	**1.6**	**0.85**	**97.5**	

(*Ct*) *C. trachomatis*. (CV) Coefficient of variation. (NS) Not sequenced. Values in bold indicate insufficient coverage of the *C. trachomatis* genome to allow SNP calling or perform further analysis.
[a]CT value is the value from the diagnostic qPCR. Higher numbers represent lower bacterial loads.
[b]Estimated from sybrgreen qPCR values.

Table 5. Assembly of IMS-MDA clinical swab sequence data

Sample	Control L2	Swab5	Swab6	SwabB1	SwabB4	SwabB8[a]	
SPAdes single cell option used	N	Y	Y	Y	Y	Y	
Total length	1,396,321	2,050,692	1,170,823	3,994,769	8,644,732	38,233,648	Before contamination removal
Number of contigs	4597	2002	282	6419	13,597	48,345	
Total length	1,032,628	1,042,675	1,037,506	1,044,130	1,036,854	1,037,223	Contigs matching the *C. trachomatis* chromosome
Number of contigs	2	2	2	9	21	5	
Mean length	516,314	521,338	518,753	116,014	49,374	207,445	
Std dev lengths	172,783	499,242	496,657	166,290	55,143	206,570	
Max length	689,097	1,020,579	1,015,410	453,473	192,642	539,668	
Min length	343,531	22,096	22,096	198	248	284	
N50	689,097	1,020,579	1,015,410	326,211	83,787	539,668	

[a]One chromosomal contig contained a misassembly forming a large inverted repeat which was manually corrected.

clinical samples, assemblies of these samples produced between two and 21 contigs covering the chromosomes. Total assembly lengths, following the removal of contaminant contigs, were between 1,032,628 bp and 1,044,130 bp, consistent with the expected size of the *C. trachomatis* chromosome. The most fragmented chromosome assembly was produced from the swabB4 data, which assembled into 21 contigs of between 248 and 192,642 bp. This sample had the lowest mean depth of coverage of the five clinical samples, and the highest chromosome CV (0.63; Table 4), indicating that these data have the most variable coverage within this data set. In all cases the plasmid was present in the assembly, although often represented as a number of contigs, due to the extremely high plasmid coverage resulting from MDA (data not shown). Following assembly it was possible to derive genotypes for these five strains, which indicated that swab6 and swabB5 have *ompA* genotype E and swab5, swabB1, and swabB8 have *ompA* genotype F (Table 4). These strains are most closely related to the sequenced strains E/SotonE4 (EMBL accession HE601802) and F/SW5 (EMBL accession HE601805), respectively.

Additional applications of IMS-MDA

The full IMS-MDA protocol can be performed on samples within 5 h with very little hands-on time. Therefore, this technique can be used on samples from cell culture to generate DNA for sequencing as an alternative to the standard protocol, which involves differential centrifugation to remove host cell debris and concentrate the *Chlamydia* followed by genomic DNA extraction and resuspension. We tested both protocols in parallel on eight strains of *C. trachomatis* from urogenital clinical samples. These were grown in tissue culture to infect five wells of a 24WT, with two wells from each pooled to undergo IMS-MDA, and three wells pooled for genomic DNA extraction, prior to Illumina sequencing and mapping to a reference genome.

We found that all samples produced 100% genome coverage, with the extracted samples producing a minimum 37.9× mean depth of coverage (Supplemental Table S2). Comparison of the sequenced IMS-MDA samples found no SNPs between samples of the same strains, even down to a mean depth of coverage of 14.4× post IMS-MDA. The sequence data were used to determine genotype (Supplemental Table S2), with these samples characterized as *ompA* genotype E (four samples), G (three samples), and K (one sample). Again, heterogeneous sites were analyzed and were found to be relatively similar across both sampling methods of the same strain. Strain R3059 showed an increased level of heterogeneity after IMS-MDA, which may be attributed to the low output level of *C. trachomatis* genome copies, indicating a low input level of DNA pre-MDA as seen above. One sample (R33512) showed very high levels of heterogeneity across the genome (>4500 sites) in sequence data after both extraction and IMS-MDA, seemingly indicative of a mixed infection. In this sample, a major variant appears to be present in ~85% of the reads, and a minor variant in ~15% of the reads. While mapping or assembly will give information on the dominant strain present, we can also resolve mixed infections bioinformatically when there is sufficient bias between the strains. In this case, it was possible to separate out the variants based on their relative frequency (data not shown), which indicates that the two strains fall in the different trachoma clades T2 (*ompA* genotype K) and T1 (*ompA* genotype E), which are separated by 4860 SNPs (Harris et al. 2012). Together, these data indicate that IMS-MDA is a rapid way of generating genomic DNA which gives accurate and interpretable genome sequence data.

IMS-MDA can also be used on *C. trachomatis* samples that have been archived at −80°C: both original clinical samples in CTM that have been stored since diagnosis and passaged samples that have been archived and are no longer viable. We have obtained complete genome sequences from seven LGV strains of *C. trachomatis*, which had been stored as diagnostic samples for up to 8 yr (47 samples assayed by IMS-MDA), and three LGV strains, which had been passaged one to five times prior to storage, up to 16 yr previously (15 samples assayed by IMS-MDA; Supplemental Table S3). These strains were determined to be *ompA* genotypes L1 and L2b, indicating the wide applicability of this technique through the use of a characterized, diagnostic antibody. While the viability of the latter samples was not tested, the former clinical samples were known to be nonviable through treatment at 56°C for viral inactivation prior to NAATs diagnosis. The IMS-MDA protocol gave rapid access to the genomic information within these archived samples, providing data from otherwise inaccessible samples.

Practical considerations for the use of IMS-MDA

IMS-MDA has potential for use in non-laboratory settings, as the IMS reagents are stable, only simple equipment is necessary, and the incubations can be carried out within the temperature range of 4°C and 30°C (data not shown). The MDA component, however, requires storage at −80°C and more complex and precise incubations, preferably using a PCR machine. By performing IMS on replicate samples and storing the resulting washed beads at a range of temperatures for 7 and 14 d prior to MDA, we determined that the *C. trachomatis* DNA is stable post-IMS over these time periods. Recovery of sufficient DNA for sequencing was possible after storage at 20°C for 14 d, although maximum recovery was achieved

after storage at −20°C (Supplemental Fig. S2). Additionally, for high-throughput sample processing, IMS-MDA can be carried out in 96-well format with equivalent yield (data not shown). The cost of IMS-MDA reagents works out at approximately US$5 per sample with low up-front equipment costs.

We also determined whether *C. trachomatis* remains viable after being subjected to the IMS protocol. Infection onto a McCoy cell monolayer was attempted using post-IMS bead-bound *C. trachomatis*, and samples eluted from the beads after IMS. While viability was greatly reduced in both these samples compared with the control sample pre-IMS, inclusions were seen under both conditions (Supplemental Fig. S3), indicating that some infectious ability remains.

Discussion

Current diagnostic techniques for many bacterial pathogens provide information on the presence or absence of a target species, occasionally generating limited subtyping information. While this is generally sufficient for patient treatment and basic epidemiology, it provides little detailed resolution relating to the causal agent. In order to study evolution and epidemics, investigate outbreaks, trace sources, and potentially track the spread of emergent drug resistant strains, greater resolution is required. For maximum phylogenetic resolution and strain identification, it is necessary to have the full genomic sequence (Harris et al. 2012). In the post-genomic age of clinical medicine, it is becoming possible as well as desirable to derive the maximum level of information from infectious agents (Gardy et al. 2011; Eyre et al. 2012; Köser et al. 2012b; Snitkin et al. 2012). While many clinically relevant bacteria are routinely cultured, some pathogens are difficult and time consuming to grow in vitro, and many environmental bacteria remain uncultured. Methods able to generate genomic information while circumventing the need for culture have great benefits in the applicability to all bacteria, and in the speed of clinical response. We have developed a rapid, culture-independent method for generating DNA for genome sequencing from a targeted species, direct from clinical samples. We have shown IMS-MDA to be an accurate, high-throughput, inexpensive, and low-tech methodology with high potential for transfer to other bacteria.

C. trachomatis, with its small genome and obligate intracellular developmental cycle, is an excellent model with which to validate this approach. With appropriate ethical consent, discarded clinical samples can be used for research purposes post-diagnosis. These samples contain very low levels of *C. trachomatis* and high levels of other material including human cells as well as other microbiota inhabiting the same body site. We have shown that it is possible to generate sufficient DNA for accurate whole-genome sequencing by combining targeted enrichment using IMS and WGA using MDA. In addition we have shown that IMS-MDA can be put to alternative uses including the rapid generation of whole genomes from limited culture volumes and from irreplaceable archival samples. While this approach does not generate data from every sample, it currently provides a respectable success rate of 15%–30%, which in many cases will offer the only possible way to generate genome sequence from complex or delicate samples.

This approach opens up new avenues of investigation for genomic research into difficult-to-culture and fastidious bacteria. Using alternative antibodies or aptamers, this technique has great potential for use with other organisms that may be present at low load in clinical and environmental samples which present culture challenges, or require specialist growth conditions such as a high containment level. We anticipate that this technique will be modified for successful use with many other organisms.

Methods

C. trachomatis strains, samples, and cell culture

Clinical samples and archived samples of *C. trachomatis* for analysis and IMS were kindly provided by Addenbrooke's Hospital, Cambridge, UK; the WHO Collaborating Center for Gonorrhoea and Other STIs, Örebro, Sweden; the Health Protection Agency, Colindale, UK; and the Center for HIV and Sexually Transmitted Infections at the National Institute for Communicable Diseases, Johannesburg, South Africa. Strain D/314 is from a cervical sample, provided by Dr. Harry Mallinson Ph.D., Consultant Clinical Scientist, University Hospital Aintree NHS Foundation Trust, Liverpool, and strain C/TW3 was obtained from American Type Culture Collection (ATCC VR-1477).

The study was approved by the National Research Ethics Service Committee East of England, Cambridge South (REC reference: 11/EE/0166). All discarded clinical samples used in this research underwent bead beating or treatment with Triton (Sigma-Aldrich) to 1% to lyse any human cells present, on arrival at the Wellcome Trust Sanger Institute, Cambridge, UK. Any human sequence data produced were removed within the sequencing system; therefore, no analysis or archiving of human sequence data was carried out. Thus there was no requirement for ethical approval for the use of extracted DNA even though the sample originally contained human DNA, due to the fact that human sequence data were removed prior to genetic analysis. All the work done in this study fell outside the requirements of the Human Tissue Act as it applies in England, Wales, and Northern Ireland.

Cell culture was performed using McCoy cells as described (Seth-Smith et al. 2009), in culture volumes of 1 mL in 24-well tissue culture trays (24WT) and 10 mL in T75 tissue culture flasks.

DNA extraction, amplification, and quantification

Genomic DNA was prepared as previously described (Seth-Smith et al. 2009). WGA was performed using the illustra Genomiphi V2 kit (GE Healthcare), using 1 μL input DNA as per the manufacturer's instructions. Quantification of total DNA was performed using a Qubit fluorometer (Life Technologies), with broad range or high specificity reagents as appropriate. To determine the level of *C. trachomatis* genomic DNA present, a Taqman qPCR method targeting the single-copy chromosomal *ompA* gene was used (Jalal et al. 2006), performed on a StepOne Plus real-time PCR system (Applied Biosystems) in 96-well format. Taqman Fast Advanced reagents (Applied Biosystems) were used according to the manufacturer's instructions with 1 μL input DNA in a total reaction volume of 20 μL. Samples were heated to 50°C for 2 min, then 95°C for 20 sec, followed by 40 cycles of 95°C for 1 sec and 60°C for 20 sec. Standards were created using a serially diluted PCR product covering the assay region (f: 5′-CGGAATTGTGCATTTACGTG-3′; r: 5′-CTACGCTGAGGACGGTAAGC-3′), quantified by Qubit as above. Readings were analyzed as genome copies per microliter. Prior to this, qPCR used Fast Sybrgreen master mix (Applied Biosystems) and the primers HJ-Plasmid-1: 5′-AACCAAGGTCGATG TGATAG-3′ and HJ-Plasmid-2: 5′-TCAGATAATTGGCGATTCTT-3′ (Jalal et al. 2006).

Sample sequencing

When sequencing samples following MDA, the volume of the MDA samples remaining after DNA quantification, and *C. trachomatis*

DNA quantification (18 μL) underwent Illumina sequencing. All samples were sequenced on GAII or HiSeq machines, paired-end, with read lengths of 54, 75, or 100 bp.

Sequence data analysis

For clinical samples, reads mapping to human DNA were enumerated and then removed by an automated pipeline prior to data release to researchers and public archives. Human data are determined as any sequence template for which either the forward or reverse read aligns to the phase one 1000 Genome Human reference sequence (human_g1k_v37) using Burrows-Wheeler Alignment (Li and Durbin 2009) run with the -q15 parameter. The sequence data were submitted to ENA with the accession numbers given in Supplemental Table S4.

Mapping of the resulting data against a relevant reference *C. trachomatis* genome was then carried out using SMALT (http://www.sanger.ac.uk/resources/software/smalt/) as previously described (Harris et al. 2010). As the rRNA operon is present in two copies in the *C. trachomatis* genome, reads will not map uniquely to these regions; thus, to accurately calculate the proportion of the reference covered by sequence reads, we employed the SMALT option to randomly map reads that have two or more best mappings with the same alignment score. Assignment to *ompA* genotypes was performed by mapping reads against a database of known *ompA* types.

De novo assembly was carried out using the SPAdes genome assembler (Bankevich et al. 2012). For the control sample, for which DNA was extracted from a cultured sample, the assembly was carried out without the single cell option. For clinical samples the single cell option was used to account for the uneven coverage distribution of reads relative to the genome, caused by MDA. To remove contigs representing contamination in the assemblies, contigs were aligned to a reference *C. trachomatis* genome (F/SW4 EMBL Accession HE601804) using ABACAS (Assefa et al. 2009). A single misassembly was identified in the assembly of swabB8, and corrected manually. This misassembly formed a large inverted repeat of an entire contig >300 kb, for which there was no supporting evidence in the raw data. The cause of this misassembly is being investigated for a bug fix by the authors of SPAdes.

Analysis of clinical samples

Discarded positive clinical samples, taken into either Gen-probe Aptima (Hologic Gen-Probe) or Abbott multi-collect (Abbott), lysis buffers were obtained from Addenbrooke's Hospital and Peterborough City Hospital. For initial analysis, buffer exchange was either performed using Qiagen DNA blood mini kit (Qiagen), where the elution volume was half that of the input volume leading to a doubling in DNA concentration, or prepared using the Wizard genomic DNA purification kit (Promega), where the DNA was resuspended in a small volume to create a 12.5× concentration compared with the clinical sample. One microliter of these samples were used in DNA quantification or amplification as above.

MDA was performed on a series of samples in which *C. trachomatis* genomic DNA was spiked at varied concentrations into *Erwinia carotovora* carrier genomic DNA. The accurate input (0.4%–21% *C. trachomatis* DNA) and post-MDA DNA quantities were calculated using Qubit and qPCR quantification, as described above.

Prior to sequencing, a second set of Genprobe Aptima samples was obtained. From these, one aliquot of 500 μL was used in an ethanol precipitation and resuspended in 50 μL autoclaved purified water, and a second aliquot of 400 μL was used with the Qiagen DNA blood mini kit with elution in 100 μL autoclaved purified water. These were analyzed or amplified as above, and the remaining volumes were sequenced as above.

Immunomagnetic separation and MDA

Where clinical, cultured, and archived samples underwent IMS-MDA, samples were harvested and bead-beaten prior to IMS, as per the DNA extraction protocol. Clinical samples from Addenbrooke's Hospital were supplied in Remel Microtest M4RT (Thermo Fisher), those from the HPA were supplied in a range of collection buffers including Remel M4RT. Where clinical samples were used, samples were routinely heated to 56°C for 15 min prior to use for viral inactivation. The magnetic beads (3 μL per sample) (Dynabeads M-280 Sheep Anti-Mouse IgG, Invitrogen) were washed twice in isotonic PBS with 0.05% Tween20 (PBST) as per the manufacturer's instructions. Primary antibody (0.3 μL per sample of IMAGEN *Chlamydia*, Oxoid) was added to the Dynabeads, with 10 volumes of PBST. The binding of primary antibody to beads was performed at 20°C shaking at 200 rpm for 1–20 h. Subsequently excess primary antibody was removed with two washes of the beads in PBST with the use of a Dynamag (Invitrogen), and the beads resuspended in PBST to 50 μL per sample. To each aliquot of 50 μL beads, between 20 and 200 μL of sample was added, depending on availability of the sample. Binding was performed at 20°C shaking at 200 rpm for 1–20 h. Beads were washed twice with PBST, using the Dynamag, in order to remove any contamination present. After removal of the final buffer wash, Genomiphi amplification was performed on the full amount of remaining magnetic beads, without elution. Where samples were stored and shipped after IMS but prior to amplification, the dry beads were treated at 95°C for 5 min and then stored at −20°C and shipped with ice packs.

Where variations to this standard protocol were used, they are indicated in Supplemental Table S5.

Spike experiment details

C. trachomatis strain D/314 (urogenital) was grown in tissue culture in a full 24WT for 3 d. These wells were harvested, pooled, bead-beaten, centrifuged at 1500 rpm for 5 min to remove the cell debris, and the final pellet resuspended in 200 μL M4RT buffer (Remel, Thermo Fisher Scientific). A 10-fold serial dilution was performed in duplicate in M4RT buffer down to 10^{-7}. For each sample, 100 μL was used to infect a sub-confluent layer of McCoy cells in one well of a 24WT (cell numbers quantified through cell counting of trypsinized duplicate wells) to determine numbers of infectious *Chlamydia*, and 100 μL was extracted and quantified by qPCR. IMS was performed on two samples of 100 μL each, of which one was amplified immediately as above, and one was analyzed by qPCR. Additional replicates were incubated at 4°C for 48 h and then stored at −80°C for 2 wk to simulate possible sample treatment within a clinic.

Stability experiment

C. trachomatis strain D/314 (urogenital) was grown in tissue culture in a full 24WT for 3 d. As above, these wells were harvested, pooled, bead-beaten, centrifuged to remove the cell debris, and the final pellet resuspended in 2SP buffer. Samples were diluted to 10^{-2} and 10^{-4} dilutions. Aliquots of 100 μL of these dilutions were inoculated into a sub-confluent layer of McCoy cells (cell numbers quantified through cell counting of trypsinized duplicate wells) to determine numbers of infectious *Chlamydia*. Further aliquots of 100 μL were subjected to the standard IMS protocol in replicates of 28. Two were immediately frozen at −80°C for subsequent qPCR

analysis of recovered number of genome copies. Two were immediately amplified and the resulting DNA was frozen at −80°C for subsequent qPCR analysis. The remaining aliquots were stored at 4°C, −20°C, and room temperature (~20°C) for 7 and 14 d, after which samples were treated as above in duplicates.

Growth in tissue culture after IMS

Strains of *C. trachomatis* D/314 (urogenital) and ATCC strain C/TW3 (ocular) were grown in tissue culture as above and underwent IMS in duplicate. After the final two washes in PBST, one replicate was used for direct infection into one well of a 24WT using centrifugation. The second replicate for each strain was eluted from the magnetic beads using 100 µL 0.1 M citrate pH3, with the eluate and the beads used for infection as above. Positive controls for each strain and negative controls were performed. Infections were followed by phase contrast microscopy over 3 d. Images were captured using a Zeiss AxioObserver A1 system.

Data access

The sequence data were submitted to the European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena/) with the accession numbers given in Supplemental Table S4.

Acknowledgments

This work was supported by the Wellcome Trust (grant number 098051).

References

Angen Ø, Heegaard PMH, Lavritsen DT, Sørensen V. 2001. Isolation of *Actinobacillus pleuropneumoniae* serotype 2 by immunomagnetic separation. *Vet Microbiol* **79:** 19–29.

Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. 2009. ABACAS: Algorithm based automatic contiguation of assembled sequences. *Bioinformatics* **25:** 1968–1969.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19:** 455–477.

Blainey PC, Quake SR. 2010. Digital MDA for enumeration of total nucleic acid contamination. *Nucleic Acids Res* **39:** e19.

Brunelle BW, Sensabaugh GF. 2006. The *ompA* gene in *Chlamydia trachomatis* differs in phylogeny and rate of evolution from other regions of the genome. *Infect Immun* **74:** 578–585.

Burgoyne RA. 1990. Lymphogranuloma venereum. *Prim Care* **17:** 153–157.

Carlson JH, Porcella SF, McClarty G, Caldwell HD. 2005. Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains. *Infect Immun* **73:** 6407–6418.

Chaparro PJP, McCulloch JA, Cerdeira LT, Al-Dilaimi A, Canto de Sá LL, de Oliveira R, Tauch A, de Carvalho Azevedo VA, Cruz Schneider MP, da Silva AL. 2011. Whole genome sequencing of environmental *Vibrio cholerae* O1 from 10 nanograms of DNA using short reads. *J Microbiol Methods* **87:** 208–212.

Dean FB, Nelson JR, Giesler TL, Lasken RS. 2001. Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res* **11:** 1095–1099.

Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, Ip CL, Wilson DJ, Didelot X, O'Connor L et al. 2012. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open* **2:** e001124.

Fratamico PM, Schultz FJ, Buchanan RL. 1992. Rapid isolation of *Escherichia coli* O157:H7 from enrichment cultures of foods using an immunomagnetic separation method. *Food Microbiol* **9:** 105–113.

Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, Rempel S, Moore R, Zhao Y, Holt R, et al. 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* **364:** 730–739.

Gomes JP, Bruno WJ, Borrego MJ, Dean D. 2004. Recombination in the genome of *Chlamydia trachomatis* involving the polymorphic membrane protein C gene relative to *ompA* and evidence for horizontal gene transfer. *J Bacteriol* **186:** 4295–4306.

Grant IR, Ball HJ, Rowe MT. 1998. Isolation of *Mycobacterium paratuberculosis* from milk by immunomagnetic separation. *Appl Environ Microbiol* **64:** 3153–3158.

Harris SR, Feil EJ, Holden MTG, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, et al. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327:** 469–474.

Harris SR, Clarke IN, Seth-Smith HMB, Solomon AW, Cutcliffe LT, Marsh P, Skilton RJ, Holland MJ, Mabey D, Peeling RW, et al. 2012. Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. *Nat Genet* **44:** 364–366.

Hedrum A, Lundeberg J, Påhlson C, Uhlén M. 1992. Immunomagnetic recovery of *Chlamydia trachomatis* from urine with subsequent colorimetric DNA detection. *Genome Res* **2:** 167–171.

Hosono S, Faruqi AF, Dean FB, Du Y, Sun Z, Wu X, Du J, Kingsmore SF, Egholm M, Lasken RS. 2003. Unbiased whole-genome amplification directly from clinical samples. *Genome Res* **13:** 954–964.

Jalal H, Stephan H, Curran MD, Burton J, Bradley M, Carne C. 2006. Development and validation of a rotor-gene Real-Time PCR assay for detection, identification, and quantification of *Chlamydia trachomatis* in a single reaction. *J Clin Microbiol* **44:** 206–213.

Jeffrey BM, Suchland RJ, Quinn KL, Davidson JR, Stamm WE, Rockey DD. 2010. Genome sequencing of recent clinical *Chlamydia trachomatis* strains identifies loci associated with tissue tropism and regions of apparent recombination. *Infect Immun* **78:** 2544–2553.

Joseph SJ, Didelot X, Rothschild J, De Vries HJC, Morré SA, Read TD, Dean D. 2012. Population genomics of *Chlamydia trachomatis*: Insights on drift, selection, recombination, and population structure. *Mol Biol Evol* **29:** 3933–3946.

Köser CU, Ellington MJ, Cartwright EJP, Gillespie SH, Brown NM, Farrington M, Holden MTG, Dougan G, Bentley SD, Parkhill J, et al. 2012a. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog* **8:** e1002824.

Köser CU, Holden MT, Ellington MJ, Cartwright EJ, Brown NM, Ogilvy-Stuart AL, Hsu LY, Chewapreecha C, Croucher NJ, Harris SR, et al. 2012b. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med* **366:** 2267–2275.

Kulakov LA, McAlister MB, Ogden KL, Larkin MJ, O'Hanlon JF. 2002. Analysis of bacteria contaminating ultrapure water in industrial systems. *Appl Environ Microbiol* **68:** 1548–1555.

Lasken RS. 2012. Genomic sequencing of uncultured microorganisms from single cells. *Nat Rev Microbiol* **10:** 631–640.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25:** 1754–1760.

Mariotti SP, Pascolini D, Rose-Nussbaumer J. 2009. Trachoma: Global magnitude of a preventable cause of blindness. *Br J Ophthalmol* **93:** 563–568.

Millman KL, Tavare S, Dean D. 2001. Recombination in the *ompA* gene but not the *omcB* gene of *Chlamydia* contributes to serovar-specific differences in tissue tropism, immune surveillance, and persistence of the organism. *J Bacteriol* **183:** 5997–6008.

Niesters HGM, Quint WGV, Herbrink P. 1991. Antigen capture directed polymerase chain reaction for detection of *Chlamydia trachomatis*. In *Techniques in diagnostic pathology* (ed. G Bullock et al.), Vol. 2. Academic Press Limited, MA.

Pedersen LN, Herrman B, Moller JK. 2009. Typing *Chlamydia trachomatis*: From egg yolk to nanotechnology. *FEMS Immunol Med Microbiol* **55:** 120–130.

Pinard R, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, Egholm M, Rothberg JM, Leamon JH. 2006. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics* **7:** 216.

Pride DT, Salzman J, Relman DA. 2012. Comparisons of clustered regularly interspaced short palindromic repeats and viromes in human saliva reveal bacterial adaptations to salivary viruses. *Environ Microbiol* **14:** 2564–2576.

Ravel J, Gaker P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, et al. 2011. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci* (Suppl 1) **108:** 4680–4687.

Ridgway GL, Taylor-Robinson D. 1991. Current problems in microbiology: 1. Chlamydial infections: Which laboratory test? *J Clin Pathol* **44:** 1–5.

Rodrigue S, Malmstrom RR, Berlin AM, Birren BA, Henn MR, Chisholm SW. 2009. Whole genome amplification and *de novo* assembly of single bacterial cells. *PLoS ONE* **4:** e6864.

Seth-Smith HMB, Harris SR, Persson K, Marsh P, Barron A, Bjartling C, Clark L, Cutcliffe LT, Lambden PR, Lennard N, et al. 2009. Co-evolution of genomes and plasmids within *Chlamydia trachomatis* and the emergence in Sweden of a new variant strain. *BMC Genomics* **10:** 239.

Skidmore S, Horner P, Mallinson H. 2006. Testing specimens for *Chlamydia trachomatis*. *Sex Transm Infect* **82:** 272–275.
Skjerve E, Rørvik LM, Olsvik O. 1990. Detection of *Listeria monocytogenes* in foods by immunomagnetic separation. *Appl Environ Microbiol* **56:** 3478–3481.
Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Program NCS, Henderson DK, Palmore TN, Segre JA. 2012. Tracking a hospital outbreak of carbapenem-resistant klebsiella pneumoniae with whole-genome sequencing. *Sci Transl Med* **4:** 148ra116.
Somboonna N, Wan R, Ojcius DM, Pettengill MA, Joseph SJ, Chang A, Hsu R, Read TD, Dean D. 2011. Hypervirulent *Chlamydia trachomatis* clinical strain is a recombinant between lymphogranuloma venereum (L_2) and D lineages. *MBio* **2:** e00045-11.
Stephens RS, Tam MR, Kuo C-C, Nowinski RC. 1982. Monoclonal antibodies to *Chlamydia trachomatis*: Antibody specificities and antigen characterisation. *J Immunol* **128:** 1083–1089.
Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q, et al. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282:** 754–759.
Su H, Watkins NG, Zhang YX, Caldwell HD. 1990. *Chlamydia trachomatis*-host cell interactions: Role of the chlamydial major outer membrane protein as an adhesin. *Infect Immun* **58:** 1017–1025.
Thomson NR, Holden MT, Carder C, Lennard N, Lockey SJ, Marsh P, Skipp P, O'Connor CD, Goodhead I, Norbertzcak H, et al. 2008. *Chlamydia trachomatis*: Genome sequence analysis of lymphogranuloma venereum isolates. *Genome Res* **18:** 161–171.
Thornley MJ, Lusher M, Scott ML, Coombs RRA, Evans RT, Thomas BJ, Taylor-Robinson D. 1983. Characterization of a monoclonal antibody to the group antigen of *Chlamydia* spp. and its use for antigen detection by reverse passive haemagglutination and indirect immunofluorescence. *FEMS Microbiol Lett* **17:** 45–49.
Thornley MJ, Zamza SE, Byrne MD, Lusher M, Evans RT. 1985. Properties of monoclonal antibodies to the genus-specific antigen of *Chlamydia* and their use for antigen detection by reverse passive haemagglutination. *J Gen Microbiol* **131:** 7–15.
Unemo M, Seth-Smith HMB, Cutcliffe LT, Skilton RJ, Barlow D, Goulding D, Persson K, Harris SR, Kelly A, Bjartling C, et al. 2010. The Swedish new variant of *Chlamydia trachomatis*: Genome sequence, morphology, cell tropism and phenotypic characterization. *Microbiology* **156:** 1394–1404.
Wang S-P, Kuo C-C, Barnes RC, Stephens RS, Grayston JT. 1985. Immunotyping of *Chlamydia trachomatis* with monoclonal antibodies. *J Infect Dis* **152:** 791–800.
WHO. 2011. *Prevalence and incidence of selected sexually transmitted infections*. World Health Organization, Geneva.
WHO. 2012. Global WHO alliance for the elimination of blinding trachoma by 2020. *Wkly Epidemiol Rec* **87:** 161–168.

Received September 28, 2012; accepted in revised form February 12, 2013.