Generating whole bacterial genome sequences of low-abundance species from complex samples with IMS-MDA

Helena M B Seth-Smith[1,6], Simon R Harris[1], Paul Scott[1], Surendra Parmar[2], Peter Marsh[3], Magnus Unemo[4], Ian N Clarke[5], Julian Parkhill[1] & Nicholas R Thomson[1]

[1]Pathogen Genomics, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridgeshire, UK. [2]Clinical Microbiology and Public Health Laboratory, Health Protection Agency, Addenbrooke's Hospital, Cambridge, UK. [3]Public Health England Laboratory Southampton, Southampton General Hospital, Southampton, UK. [4]World Health Organization (WHO) Collaborating Centre for Gonorrhoea and Other Sexually Transmitted Infections (STIs), National Reference Laboratory for Pathogenic Neisseria, Örebro University Hospital, Örebro, Sweden. [5]Molecular Microbiology Group, Faculty of Medicine, University of Southampton, Southampton General Hospital, Southampton, UK. [6]Present address: Functional Genomics Center Zürich, Universität Zürich, Zürich, Switzerland. Correspondence should be addressed to H.M.B.S.-S. (hss@fgcz.uzh.ch) or N.R.T. (nrt@sanger.ac.uk).

Published online 7 November 2013; doi:10.1038/nprot.2013.147

The study of bacterial populations using whole-genome sequencing is of considerable scientific and clinical interest. However, obtaining bacterial genomic information is not always trivial: the target bacteria may be difficult to culture or uncultured, and they may be found within samples containing complex mixtures of other contaminating microbes and/or host cells, from which it is very difficult to derive robust sequencing data. Here we describe our procedure to generate sufficient DNA for whole-genome sequencing from clinical samples and without the need for culture, as successfully used on the difficult-to-culture, obligate intracellular pathogen *Chlamydia trachomatis*. Our protocol combines immunomagnetic separation (IMS) for targeted bacterial enrichment with multiple displacement amplification (MDA) for whole-genome amplification (WGA), which is followed by high-throughput sequencing. Compared with other techniques that might be used to generate such data, IMS-MDA is an inexpensive, low-technology and highly transferable process that provides amplified genomic DNA for sequencing from target bacteria in under 5 h, with little hands-on time.

INTRODUCTION

Driven by plummeting sequencing costs, whole-genome sequencing has revolutionized the way we have been able to understand bacterial biology, pathogenesis, epidemiology, genetics and evolution. Sequencing technologies have been applied to track the temporal and geographic distribution of pathogens, and they offer the precision to understand the true nature of infection within a single person[1–7]. Much of our understanding of the nature and diversity of bacteria derives from species that can be easily, rapidly and selectively grown under laboratory conditions. It is a challenge to generate genomic DNA of sufficient quantity and quality for whole-genome sequencing from uncultured, fastidious or difficult-to-culture bacteria. Discarded clinical samples represent an important resource for studying bacteria, although such samples may be complex, often containing low levels of the species of interest among a multitude of contaminating bacteria and host cells.

We have developed an approach to enrich for specific bacteria and amplify sufficient genomic DNA for whole-genome sequencing directly from complex or nonviable samples without the need for culture[8]. Our methodology combines IMS with MDA and allows for access to the genomes of specific bacteria even when the target species is present at low levels and with other contaminating microbiota and host cells. To validate our methodology, we amplified and sequenced complete genomes of *C. trachomatis* strains present in discarded clinical samples, representing the first time that whole bacterial genome sequences have been generated directly from uncultured clinical samples[8].

C. trachomatis is a pathogen of global importance, causing more than 100 million cases of sexually transmitted chlamydial infection annually[9], as well as the blinding ocular disease trachoma[10]. There are several subspecies-level typing schemes for *C. trachomatis*, which group strains according to variation in the highly variable gene *ompA*, a panel of housekeeping genes (by multilocus sequence typing (MLST))[11–13] or fast-changing repetitive loci (by variable number tandem repeats (VNTRs))[14]. However, these methods lack the resolution required for detailed strain tracking in the case of *ompA* genotyping and MLST, and they can be vulnerable to stochastic change in the case of VNTRs. Moreover, recently, it has been shown that recombination in *C. trachomatis*, leading to genetic exchange and diversification unlinked to phylogeny, severely limits the interpretation of data from these typing methodologies[15–17]. For *C. trachomatis*, as for many other bacterial species, whole-genome sequencing is the only technology that can provide the resolution required to determine true relationships, in addition to the accuracy and specificity required to differentiate closely related isolates typical of monomorphic species[4].

C. trachomatis is an ideal model bacterium for testing non-culture-based sequencing protocols: as an obligate intracellular pathogen, it requires tissue culture for *in vitro* growth, a method that is technically challenging, expensive and time-consuming. Furthermore, some strains may be recalcitrant to culture, and discarded clinical samples are available for research. Although deep sequencing of a clinical sample has been shown to be able to give some information about an infecting *C. trachomatis* strain[18], this method is expensive and it is not scalable or able to provide sufficient strain resolution. Several other technologies exist or are being developed for the depletion of host cells in clinical

samples (including MolYsis Basic by Molzym and the GeneRead Bacterial DNA kit by Qiagen (not yet released)), particularly in blood samples (discussed in ref. 19), but these are of less use when other microbiota are present in the sample, masking data from the bacterium of interest. Also of note are DNA target enrichment technologies, in which the sequence of a reference strain can be used to develop baits covering the whole genome, allowing the specific hybridization of the target strain's DNA. Such methods have been shown to be effective in several studies on bacteriophage, bacteria and viruses[19–23], and they can be used on samples in lysis buffer, but they are highly expensive and time-consuming, requiring the initial design and purchase of a custom array of baits. In addition, if the genome of the species under study is variable, not all the variation will be captured by using these methods.

The sequencing of genomes of single bacteria has been shown in several studies, isolated by micromanipulation, microfluidics or flow cytometry (reviewed in refs. 24,25). For some applications, such as comparison of sequence variation between individual cells of the same species within a complex mix, methodologies such as these may offer some advantages, depending on the specific research goal. However, these are highly time- and resource-consuming techniques, with the isolation stages so far not targeted toward particular bacterial species. Moreover, it is very difficult to produce full coverage of the genomes of interest, which is an aim of our protocol, such that detailed genomic epidemiology can be performed. Recently developed techniques also include a 'mini-metagenome' approach after automated cell sorting, and growth of individual bacteria from communities within gel microdroplets before amplification[26,27], from which better genome coverage is achieved because a larger number of bacteria are subject to amplification.

The need for methods to target the genomes of difficult-to-culture bacteria was confirmed in a very recent paper[28], in which IMS and MDA were also used in combination with sequenced *C. trachomatis* genomes from clinical samples. The method is largely equivalent to the one detailed here, although alternative magnetic beads were used, a DNase step was included to reduce host DNA contamination (a step which we found did not increase the success rate) and DNA extraction step was added before MDA.

Development of the protocol

IMS is an established technique for enriching target bacteria from complex mixtures by using antibodies attached to magnetic beads. In the past, IMS has been used with varied samples to concentrate the target bacteria and remove inhibitors from the sample, enabling clearer and more accurate nucleic acid amplification–based diagnosis[29–32]. We chose to test IMS for bacterial enrichment before genome sequencing.

Our experience of clinical swab samples for *C. trachomatis* diagnosis indicates that the vast majority of samples do not carry sufficient quantities of target bacteria and their DNA to allow genome sequencing from the sample either directly or after amplification[8]. Consequently, we chose to follow IMS with MDA, which amplifies high-molecular-weight DNA using ϕ29 polymerase and random hexamer primers[33,34]. This combined IMS-MDA protocol can provide sufficient high-quality genomic DNA for sequencing using high-throughput technologies[8]. The IMS-MDA protocol is rapid, flexible, low-technology and low-cost (for consumable reagents, excluding quantification and sequencing costs), with the potential for use with a multitude of sample types and bacterial species.

Overview of IMS-MDA

The IMS-MDA protocol is shown in **Figure 1** and includes the following stages: (i) primary and magnetic bead–conjugated secondary antibodies are mixed; (ii) after incubation, a magnet is used to retain the bound primary and secondary antibodies, and excess unbound primary antibody is removed; (iii) the bound antibodies are mixed with the clinical sample; (iv) after incubation, the samples are washed to remove contaminating microbes and cells that are not bound to the antibodies; (v) enriched target cells are retained by a magnet; and (vi) MDA is performed directly on the enriched bacterial sample. Subsequent to IMS-MDA, the following steps are used to generate genome sequence data: (vii) the amplified DNA is quantified, and samples with sufficient target DNA are sequenced using a high-throughput sequencing platform; and (viii) bioinformatics analysis is used to determine the quality of the sequence data and the nature of the targeted bacterial genome sequence.

Applications of the method

Our work has focused on obtaining complete genome sequences of the difficult-to-culture, sexually transmitted bacterium *C. trachomatis*[8]. The samples we first used to validate this approach were cultured and purified *C. trachomatis*, after which we tested the protocol on discarded urogenital swab samples that had returned a positive diagnosis for *C. trachomatis* by a routine diagnostic nucleic acid amplification test (NAAT)[8,35]. We have also successfully applied this technique to nonviable archived *C. trachomatis* samples, providing access to historical genome data that would have otherwise been lost, as the samples contained insufficient DNA even for sequencing at high coverage levels, without prior enrichment.

We believe that the IMS-MDA protocol would be equally applicable to any microorganism for which a suitable and

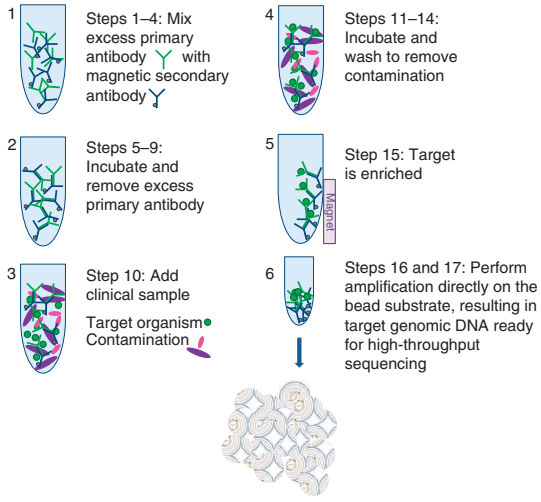

Figure 1 | Schematic of the immunomagnetic separation and multiple displacement amplification (IMS-MDA) procedure. In addition to the procedure shown, subsequent quantification, high-throughput sequencing and data analysis are described in Steps 18–25.

specific primary antibody or aptamer exists or can be developed, and for which suitable primary specimen material is available. The sample may be viable or nonviable, as long as the bacterium remains intact such that the antigen remains associated with the genomic DNA.

The low cost and high transportability of this protocol means that samples can be prepared even in resource-poor laboratories, and the crucial first stages of IMS (stages 1–5 in **Figure 1**; Steps 1–15 in the PROCEDURE) could even be performed in the field. IMS-MDA has advantages over DNA target enrichment techniques in situations when a reference genome sequence is not available, or when dealing with a highly variable genome, as *de novo* sequence assembly is possible by using the sequence data generated by IMS-MDA.

Experimental design

Sample choice. This technique is appropriate for samples that maintain intact bacteria. Our study used clinical samples placed directly into *Chlamydia* transport medium (CTM). Appropriate ethical permission may be required before work on clinical samples. To comply with the Human Tissue Act (2004) as it applies in England, Wales and Northern Ireland, human cells must be lysed on receipt, and sequence data from patient DNA must be discarded before sequence analysis (see details in MATERIALS).

Clinical samples that involve swabs or urine samples placed directly into lysis buffer are not suitable for this approach. With the general move of molecular diagnostic laboratories toward the use of commercial NAATs that supply bespoke collection vessels containing lysis/stabilization buffer, such as those for Gen-Probe Aptima or Abbott m2000rt, future studies will generally require prospective additional samples to be taken and collected in suitable transport medium as described above. Other possible samples could include any complex mixture with sufficient intact target material. Samples can be concentrated by centrifugation from larger volumes before resuspension in a suitable medium to keep the cells intact (e.g., CTM or PBS), or they can be homogenized if required. This method has so far been unsuccessful in generating complete genome sequences from urine samples, perhaps owing to the condition of the bacteria. The minimum number of bacteria required for a successful outcome was difficult to precisely establish using *C. trachomatis*, which is not easily quantifiable, but a concentration-dependent effect was observed during the validation of the procedure[8]. The affinity of the chosen antibody and the amount of contaminating material may also affect the required minimum input.

Choice of antibody. Selection of an appropriate antibody is key for IMS-MDA, as it is the crucial component for the specific enrichment of the target organism. Bacterial surface antigens such as lipopolysaccharides (LPS) and outer membrane proteins are common antibody targets. Previously developed commercial diagnostic antibodies are excellent reagents for IMS, as they are likely to have been extensively tested for cross-reactivity. For our study, we used the commercially available anti-*Chlamydia* mouse IgG primary antibody (see MATERIALS) that targets chlamydial LPS, present at ~34,000 molecules per bacterium[36]; it binds to all serovars of *C. trachomatis*. This reagent proved ideal, as it is inexpensive for the quantities used, and it was originally developed for diagnostic purposes and thus has been tested for cross-reactivity against many other microbial species possibly present in the urogenital tract including *Lactobacillus lactis*, *Mycoplasma* spp., *Neisseria gonorrhoeae* and *Gardnerella vaginalis* (IMAGEN *Chlamydia* booklet (http://www.oxoid.com/ifu (key-code TSMX7847))[37,38]. We also tested a range of other anti-*Chlamydia* antibodies that provided comparable results: mouse monoclonal IgG2b anti-*C. trachomatis* LPS, mouse monoclonal IgG2a anti-*C. trachomatis* MOMP and rabbit polyclonal IgG anti-*C. trachomatis* (see MATERIALS). Similarly, small peptide aptamers that bind to specific target molecules could be used for this purpose.

Magnetic bead selection. The primary antibodies can be attached to magnetic beads through the use of specific secondary, magnetic bead–conjugated antibodies, including anti-mouse or anti-rabbit IgG antibodies. The primary antibody used to validate this protocol is not directly linked to a magnetic bead, and thus we used a secondary anti-mouse IgG sheep antibody. Alternatively, primary antibodies can be directly conjugated to activated magnetic beads (e.g., Dynabeads MyOne carboxylic acid, Invitrogen).

Whole-genome amplification. We chose MDA as the most appropriate method for WGA. Since the development of this protocol, a new method of WGA has been developed, called multiple annealing and looping-based amplification cycles (MALBAC), which reduces the observed amplification bias of the isothermal MDA through cycles of primer extension[39]. Although the resulting genome coverage may be more even, MALBAC currently produces more errors in nucleotide incorporation through the use of DNA polymerases with lower fidelity than the ϕ29 polymerase. Therefore, we continue to recommend the use of MDA.

Bacterial genome quantification by quantitative PCR (qPCR). A specific qPCR system is required to enable the quantification of the target species' genomic DNA obtained by IMS-MDA. Our initial experiments used a SYBR Green system, but greater accuracy, sensitivity and reproducibility were obtained with the Taqman system described in Jalal *et al.*[35] for *C. trachomatis*, targeting the single-copy chromosomal *ompA* gene. The information derived from qPCR directs the choice of which samples are suitable for downstream sequencing. Quantification of the total DNA resulting from the MDA reaction, which includes amplified DNA from contaminating organisms, is unnecessary and uninformative for this purpose. For *C. trachomatis*, we found that a minimum of 1,500,000 genome copies per microliter was required to provide complete genome sequences with confidence in single-nucleotide polymorphism (SNP) calling.

Controls. Fresh buffer can be used as a negative control for contamination, and it can be processed in parallel with experimental samples, although we have never found contamination to occur. Buffer containing the target organism can be used as a positive control, if available. Positive and negative controls for MDA are described with the manufacturer's instructions. The final result is determined as being positive or negative after analysis of the sequencing data.

Genome sequencing. We used Illumina GAII and HiSeq platforms in our study, but the DNA produced should be equally

amenable to other sequencing technologies. For SNP analysis, high-throughput technologies such as Illumina or Ion Torrent are recommended because of their high and accurate sequence yield per sequencing run. We have used high levels of multiplexing (up to 96) with IMS-MDA samples with no deleterious effects.

Data analysis. We mapped the sequence data to a known reference genome to determine the extent, depth and evenness of coverage. This was performed using SMALT (http://www.sanger.ac.uk/resources/software/smalt/), although similar programs are available, including bwa and SOAP[40,41]. After IMS-MDA, we have found that bases and SNPs can be called with accuracy when the mean depth of coverage is greater than 35×, as long as the coverage coefficient of variation (CV = s.d./mean) is not >1. MDA can produce uneven coverage, and with CV values of >0.5 manual checking of base calling is recommended. The *C. trachomatis* genome is extremely stable in terms of gene content and overall genome architecture[16,42–45], meaning that mapping-based approaches allowed us to accurately reconstruct more than 99% of the genome of samples meeting these coverage criteria. For other bacteria, in which the genome may contain accessory regions that would not be accessible by mapping-based methods, *de novo* assembly would be required. In such cases, it is important to note that many assembly algorithms make the assumption of relatively constant read-coverage levels across the sequenced genome, and they may produce poor results when used with sequence reads from samples that have been subjected to IMS-MDA. We successfully assembled large contigs from *C. trachomatis* IMS-MDA data by using SPAdes[46], an assembly program designed for use with sequence reads from MDA-amplified single cell–derived DNA. For *de novo* assembly of IMS-MDA data, we would therefore recommend the use of an algorithm designed for sequence data from amplified DNA, such as velvet-sc[47], IDBA-UD[48] or SPAdes[46]. These programs also account for chimeric DNA rearrangements that may occur during MDA, and which again may confound analyses with assembly programs not designed for this specific purpose. To remove assembled regions derived from contaminating DNA, we aligned the resulting contigs against a reference *C. trachomatis* genome sequence with ABACAS[49]. This process is enabled by the extremely conserved nature of the *C. trachomatis* genome, and it allows the identification and resolution of inverted repeats, which may otherwise result in artifacts during assembly. The resulting *C. trachomatis* genomes in our study were finally assembled into 2–21 contigs[8]. We were also able to identify one mixed infection within the samples tested, and to separate out the sequences of both strains based on the relative coverage of the two genomes[8].

Limitations of the method

IMS-MDA allows access to the genomic information in otherwise inaccessible samples, and our studies have shown that we can generate accurate whole-genome sequences from 15–30% of discarded clinical swab samples that have tested positive by a routine diagnostic NAAT[8]. The reasons for the lack of success in some samples is unclear, but they were unrelated to choice of antibody or incubation conditions. Therefore, it is assumed that the load of the target bacterium and the integrity of the sample are keys to the success of the technique. Clearly, samples in which the genomic DNA is dissociated from the targeted antigen will not provide sufficient DNA of interest to sequence. Careful selection of samples is essential in order to allow the greatest likelihood of success.

MATERIALS

REAGENTS

- $NaH_2PO_4 \cdot H_2O$ (Sigma-Aldrich, cat. no. S9638)
- $Na_2HPO_4 \cdot 2H_2O$ (Sigma-Aldrich, cat. no. S3264)
- NaCl (Sigma-Aldrich, cat. no. S3014)
- Tween 20 (Sigma-Aldrich, cat. no. P9416)
- Primary antibodies: IMAGEN *Chlamydia* (Oxoid, Thermo Fisher, cat. no. K610111-2), mouse monoclonal anti-*C. trachomatis* LPS IgG2b (MyBioSource, clone no. M4020310), mouse monoclonal anti-*C. trachomatis* MOMP IgG2a (AbCam, Cambridge, UK, clone no. BIOD166) and rabbit polyclonal anti-*C. trachomatis* IgG (MyBioSource, cat. no. MBS221885)
- Dynabeads M-280 sheep anti-mouse IgG (Invitrogen, cat. no. 11201D) or other appropriate secondary antibody–coupled or activated Dynabeads
- Illustra GenomiPhi V2 DNA amplification kit (GE Healthcare, cat. no. 25-6600-30) or RepliG (Qiagen, cat. no. 150023)
- Clinical samples: we obtained discarded urogenital clinical samples in Remel M4RT transport medium (Thermo Fisher) **! CAUTION** Appropriate ethical permission might be required for the use of discarded clinical samples. Our study was approved by the appropriate National Research Ethics Service Committee. In addition, we automatically mapped all sequence reads to the human genome directly after the sequence generation, and discarded human sequence data without further analysis **! CAUTION** Clinical samples will very likely contain viable bacteria; therefore, samples should be handled under appropriate containment conditions (for *C. trachomatis* under containment level 2, preferably in a class 2 biological safety cabinet) until the bacteria have been heat-killed in Step 15 or 17.
- TaqMan fast advanced master mix (Applied Biosystems, Invitrogen, cat. no. 4444556) or alternative qPCR reagents
- qPCR primers and probes: for *C. trachomatis* quantification, we used the Taqman primer set targeting the chromosomal single-copy *ompA* gene: F primer, HJ-MOMP-1: 5′-GACTTTGTTTTCGACCGTGTT-3′; R primer, HJ-MOMP-2: 5′-ACARAATACATCAAARCGATCCCA-3′; probe, MOMP: 5′- VIC-ATGTTTACVAAYGCYGCTT-3′ (ref. 35) (Sigma-Aldrich custom oligos) **▲ CRITICAL** Ideally, the qPCR target should be a single-copy locus on the bacterial chromosome. The largest chromosome should be targeted if there is more than one in the bacterium of interest as MDA amplifies smaller circular molecules such as plasmids to higher copy numbers than larger molecules. If no genomic information on the bacterium exists, it could be attempted to target specific 16S rRNA gene sequences.

EQUIPMENT

- DynaMag-2 magnet (Invitrogen, cat. no. 12321D)
- Shaking incubator (Innnova 42, New Brunswick Scientific, Eppendorf, cat. no. M1335-0002 or equivalent) or platform rocker (STR6, Stuart Scientific, ScienceLab.com, cat. no. 65-286-674 or equivalent). A tube rotator can also be used (SB2 with SB3/1, Stuart Scientific, cat. nos. SB2 and SB3/1 or equivalent)
- Microcentrifuge (Eppendorf 5418, cat. no. FA-45-18-11 or equivalent)
- Safe-lock tubes, 2 ml (Eppendorf, cat. no. 0030 120.094)
- Vortex (Vortex-Genie 2, Scientific Industries, cat. no. SI-0266 or equivalent)
- PCR tubes or plates and sealer, 0.2 ml (thin-walled dome-capped strips, Thermo Scientific, cat. no. AB-0451; Eppendorf Twin.tec, Eppendorf, cat. no. 0030133374 and Microseal A film, Bio-Rad, cat no. MSA5002 or equivalent)
- PCR machine (e.g., MJ Research Tetrad 2 DNA Engineer cycler, Bio-Rad, cat. no. PTC–0240G)
- Real-time qPCR machine (StepOne real-time PCR System, Life Technologies; Invitrogen, cat. no. 4376600 or equivalent)
- MicroAmp fast optical 96-well qPCR plate (Applied Biosystems, cat. no. 4311971)

• MicroAmp fast optical adhesive film (Applied Biosystems, cat. no. 4346906)
• Class 2 biological safety cabinet for handling clinical samples or other samples that may contain live containment level 2 organisms (where required, depending on the organisms processed)

REAGENT SETUP

Isotonic PBS Mix 0.16 grams per liter $NaH_2PO_4 \cdot H_2O$, 0.98 grams per liter $Na_2HPO_4 \cdot 2H_2O$ and 8.10 grams per liter NaCl; autoclave and divide isotonic PBS into aliquots into sterile 100-ml bottles.

PBS-Tween (PBST) wash buffer Add Tween 20 to isotonic PBS to 0.05% (vol/vol). Freshly prepare the buffer before use. ▲ **CRITICAL** Lack of Tween 20 in the wash buffer can cause clumping of the beads and lower the efficacy of the procedure.

qPCR standards Serially diluted standards should be used alongside experimental samples for approximating the number of target organism genome copies. For this purpose, either use a known quantity of pure genomic DNA or, where necessary, generate, purify and quantify a PCR product covering the locus targeted by the qPCR assay. For the *C. trachomatis ompA* Taqman assay, we generated a standard PCR product by using the forward primer: 5′-CGGAATTGTGCATTTACGTG-3′ and the reverse primer: 5′-CTACGC TGAGGACGGTAAGC-3′.

PROCEDURE

IMS of bacteria from complex samples ● TIMING minimum 2.5 h

▲ **CRITICAL** This protocol can be used to process between 1 and 16 samples simultaneously. Beads for all samples are prepared in one pool in Steps 1–9, before aliquotting them to the individual samples.

1| Vortex the stock of Dynabeads, remove 2 µl for each sample to be processed and place it in a 2-ml tube.

2| Add 0.5 ml of PBST and resuspend the beads with a 1-s vortex or by flicking the tube. Place the tube in the DynaMag-2 magnet and leave it for 2 min. Remove the liquid while the tube is still held in the magnet, retaining the Dynabeads in the tube.
? TROUBLESHOOTING

3| Remove the tube from the magnet and repeat Step 2.

4| Remove the tube from the magnet and add 0.5 ml of PBST. For each sample to be processed, add primary antibody to the equivalent of ~10^{12} IgG molecules (e.g., 0.25 µl of IMAGEN *Chlamydia* per sample).

5| Incubate the tube at 20 °C with shaking at 200 r.p.m. for at least 1 h. Note that we observe no decrease in performance using temperatures in the range of 4–30 °C, incubation time of up to 24 h or with alternative mixing methods including rotation and rocking.

6| During the incubation, prepare the clinical samples by defrosting them if necessary.
! CAUTION Samples containing live bacteria must be handled under appropriate containment conditions (see MATERIALS).

7| Perform a pulse-spin in the microcentrifuge on the tube from Step 5.

8| Place the tube in the magnet and leave it for 2 min. Remove the liquid while the tube is still held in the magnet, retaining the Dynabeads in the tube.
? TROUBLESHOOTING

9| To remove any unbound primary antibody from the solution, perform Step 2 twice and finally resuspend the Dynabeads in 50 µl of PBST per sample to be processed.

10| Place aliquots of the clinical samples (10–200 µl depending on the available volume) in 2-ml tubes. Add 50 µl of the antibody-bound Dynabeads from Step 9 to each sample and mix by flicking the tubes.
? TROUBLESHOOTING

11| Perform Step 5.

12| Perform a pulse-spin on the tubes.

13| Place the tubes in the magnet and leave them for 2 min. Remove the liquid while the tubes are still held in the magnet, retaining the Dynabeads in the tubes.
? TROUBLESHOOTING

14| To remove contaminating material from the target bacteria, perform Step 2 twice.

15| After the final wash, retain the tubes in the magnet and remove as much of the wash buffer as possible, leaving just the Dynabeads in the tubes.
■ **PAUSE POINT** The material can be stored at −20 °C for at least 14 d before amplification. The material can be transported at this stage. To inactivate biological material for transport, the samples can be heated to 95 °C for 5 min and subsequently stored at −20 °C for several months.

MDA ● TIMING 2.5 h
16| Perform MDA by using the Illustra GenomiPhi V2 DNA amplification kit. Resuspend each of the Dynabeads pellets from Step 15, which carry the template DNA, in 9 µl of sample buffer, and transfer the full volume to a 0.2-ml well of a PCR plate. The Dynabeads can remain in the samples throughout the reaction.

17| Complete the MDA according to the manufacturer's instructions, using the 95 °C denaturation step to release DNA from the bacteria.

DNA quantification and genome sequencing ● TIMING variable
18| Quantify the amount of target bacterium DNA produced by performing Taqman qPCR. Make sufficient master mix for the required number of samples and standards by using TaqMan Fast Advanced Master Mix in a reaction volume of 20 µl, with each primer at a final concentration of 300 nM and probe at a final concentration of 150 nM.

19| Add 1 µl of each sample from Step 17 to 19 µl of master mix in a 96-well qPCR plate and seal it with adhesive film.

20| Perform the qPCR with the cycling conditions shown in the table below:

Cycle number	Uracil-N glycosylase incubation	Polymerase activation	Denature	Anneal/extend
1	50 °C, 2 min	95 °C, 20 s		
2–41			95 °C, 1 s	60 °C, 20 s

21| Analyze the results in the qPCR software; use the cycle threshold (C_t) values from the standards to calculate the amount of target DNA in the IMS-MDA samples.
▲ **CRITICAL STEP** To obtain complete, accurate bacterial genome sequences, >1,500,000 genome copies per µl are required (data from *C. trachomatis*). The minimum requirement for other bacteria and samples should be empirically determined. It is not recommended to proceed to sequencing with samples with <500,000 genome copies per microliter.
? TROUBLESHOOTING

22| Use the remaining material from Step 17 (19 µl) for high-throughput sequencing, up to 96-plex. (Illumina GAII or HiSeq, or Ion Torrent technology is suitable to generate a high yield of accurate sequence data.)

23| Analyze the resulting sequence reads by mapping them to a known reference genome with appropriate software to determine the depth and evenness of coverage (see Experimental Design). At least 99% of the genome should have reads to map against it, with the mean depth of coverage greater than 35×, although these values should be verified for each organism and sequencing method used. Calculate the coverage CV = s.d./mean. If this value is <0.5, the coverage should be sufficiently even to allow confident base calling. If the CV value is between 0.5 and 1, manual checking of base calling is recommended. If the CV value is >1, some regions of the genome may not have sufficient coverage to allow nucleotide variance analysis.

24| Use the coverage values to correlate the qPCR results with sequencing success, such that future sequencing can be performed on samples with an appropriate number of genome copies to allow the greatest chance of success.

25| Where mapping is not possible or advisable (when no referenced genome exists, or when a large amount of novel accessory DNA is suspected), perform *de novo* assembly using SPAdes[46] or an equivalent program designed for use with data obtained from a sample that has undergone MDA. If required, align the resulting contigs against a reference sequence by using ABACAS[49] to determine core genomics regions, and to identify possible accessory or contaminant-derived data.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 1**.

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Solution
2, 8, 13	Loss of beads in the reaction tube	Clumping of beads, which leads to poor retention on the magnet	Ensure that Tween is added to PBST. Alternative buffers can also be used (see Dynabeads manufacturer's instructions) Take care when removing buffers and ensure visually that the pellet of beads remains in the tube
		Beads may be lost in the tube or pipette tip	Pulse-spin the tube if buffer is splashed up the side of tube. Avoid pipetting to mix beads: vortex or flick the tube
10	Non mixing of sample and bead mixture during incubation	The sample and bead mixture exist as separate droplets in the tube and do not come into contact during the incubation	Ensure that the bead-containing buffer is pipetted directly into the sample. If the sample is viscous, addition of further PBST may be advisable to a maximum of 500 µl
21	Inefficient qPCR	Primers or probe concentrations and cycling conditions are inappropriate for assay	Use concentrations and conditions recommended for the specific assay Use the standard dilutions to test a variety of conditions according to the manufacturer's instructions
	Insufficient amplification of target DNA	An inappropriate primary antibody, or insufficient quantities of it, may have been used in Step 4	This can be determined by retaining the supernatant in Step 8 and performing qPCR to determine whether unbound target bacterium remains in the supernatant. Increased amounts of antibody or an alternative antibody can then be tested
		Too much contaminating DNA or MDA reaction inhibitors remain in the samples	Perform two or more additional washes in PBST at Step 14. A balance between removal of contamination and loss of or damage to the material of interest must be found
		Excessive PBST buffer remains after IMS and is transferred to the MDA reaction	Ensure that as much PBST as possible is removed after the final wash (Step 15). Ideally 1 µl should remain, although MDA has been found to be effective even when 5 µl buffer is added to the sample buffer with the beads
		Poor sample quality	Select input (clinical) samples that are better preserved or of higher quality, or use a higher sample volume

● TIMING

Steps 1–15, IMS of bacteria from complex samples: minimum 2.5 h
Steps 16 and 17, MDA: 2.5 h
Steps 18–25, DNA quantification, genome sequencing and analysis: several days up to weeks, depending on the method used

ANTICIPATED RESULTS

This protocol should yield complete bacterial genome sequences, derived through mapping and variant nucleotide calling or *de novo* assembly. The ultimate success of the protocol can only be assessed after analysis of the sequence read data, although qPCR results can give an interim indication as to the likely outcome. Sequence reads mapping to the target bacterium may constitute only 9% of the output per sample, with the remaining reads representing contaminating DNA, either human or microbial (**Table 2**). This is because IMS is used for enrichment of the target organism through depletion of contaminating DNA as opposed to absolute purification of the target. The additional sequencing read data may therefore provide information on the other organisms in the sample under analysis.

The success rate of the protocol is likely to depend on the nature and the load of the target organism in the input sample. Indeed, repetition of the protocol on additional sample material is unlikely to yield improved results and is not specifically recommended. With *C. trachomatis*, we achieved a success rate of 15–30% from urogenital samples and archived diagnostic samples[8]. Rates were considerably higher for more recent, viable samples archived after growth (100%) (ref. 8). By using IMS-MDA, we have also been able to generate complete genomes from older samples (up to 30 years old), which had been

TABLE 2 | Performance of IMS-MDA with whole-genome sequencing on clinical samples.

Urogenital swab	Genome copies after IMS-MDA (per µl)[a]	Illumina	Read length (bp)	Tags per lane	Total yield for tag (kb)	Reads mapping to *Ct* genome (%)	Reads mapping to *Ct* plasmid (%)	*Ct* chromosome covered (%)[b]	Depth of coverage (mean)	Chromosome CV (s.d./mean)	Reads mapping to human genome (%)	Reads mapping to contaminating microbes (%)[c]	No. of assembled contigs
Swab1	4,500	GAII	54	12	19,577	0.3	0.3	4.3	0.0	4.85	77.8	0.0	
Swab2	<1,000	GAII	54	12	227,744	0.9	0.6	86.6	2.5	0.79	94.4	0.0	
Swab3	<1,000	GAII	54	12	192,239	<0.2	0.0	8.0	0.1	4.24	78.5	1.7	
Swab5	77,000,000	GAII	54	12	**250,571**	**40.6**	**34.6**	**100.0**	**98.3**	**0.46**	**16.3**	**2.0**	**2**
Swab6	44,000,000	GAII	54	12	**157,199**	**40.2**	**47.6**	**99.9**	**63.7**	**0.51**	**5.4**	**0.8**	**2**
Swab7	5,000	GAII	54	12	287,815	1.3	5.1	92.7	3.5	0.75	26.5	20.3	
Swab8	<1,00	GAII	54	12	18,536	<0.2	0.0	0.4	0.0	15.05	92.5	0.0	
SwabB1	110,000,000	Hiseq	75	21	**871,541**	**9.6**	**1.5**	**100.0**	**86.4**	**0.54**	**10.4**	**13.6**	**9**
SwabB2	<1,000	NS											
SwabB3	<1,000	NS											
SwabB4	11,000,000	GAII	54	12	**331,977**	**10.9**	**14.7**	**99.9**	**35.9**	**0.63**	**38.8**	**1.4**	**21**
SwabB5	<1,000	NS											
SwabB6	2,700	NS											
SwabB7	18,000	NS											
SwabB8	15,000,000	GAII	54	12	**294,782**	**21.4**	**38.1**	**100.0**	**60.9**	**0.54**	**32.0**	**0.8**	**5**
SwabB9	<1,000	NS											
SwabB10	<1,000	NS											
SwabB11	300,000	Hiseq	75	21	688,148	0.2	0.0	78.0	1.6	0.85	97.5	0.0	

NS, not sequenced. *Ct*, *Chlamydia trachomatis*.

[a]Estimated from qPCR values.

[b]Overamplification of small circular molecules can occur during MDA.

[c]Contaminating microbe genomes mapped against were *Gardnerella vaginalis* (GenBank accession number CP001849), *Lactobacillus johnsonii* (GenBank accession numbers FN298497 and AE017198), *Lactobacillus reuteri* (GenBank accession number CP000705), *Lactobacillus gasseri* (GenBank accession number CP000413), *Lactobacillus crispatus* (GenBank accession number FN692037), *Mycoplasma genitalium* (GenBank accession number L43967) and *Mobiluncus curtisii* (GenBank accession number CP001992). Data in **bold** indicate the samples that provided sufficient coverage of the *C. trachomatis* genome to allow SNP calling and further genomic analysis, after IMS-MDA and Illumina sequencing. Some of these data were previously published[8].

through several rounds of freezing and thawing and were no longer viable, with a 7% success rate (H.M.B.S.-S., L. Ellison, C. Labiran, N.R.T. and I.N.C., unpublished data). For samples that do not produce complete genome sequences, partial sequence data may still provide useful information for bioinformatics analysis.

ACKNOWLEDGMENTS This work was supported by the Wellcome Trust (grant no. 098051). We thank L. Ellison for useful comments on the manuscript.

AUTHOR CONTRIBUTIONS N.R.T., P.M., I.N.C., H.M.B.S.-S. and S.R.H. conceived of the project. H.M.B.S.-S. designed and performed the experiments, researched and optimized the protocol, and wrote the manuscript. N.R.T. supervised the development of the protocol, analyzed the data and wrote the manuscript. S.R.H. performed key data analysis, helped in the development of the protocol and wrote the manuscript. P.S. and S.P. contributed valuable information to the development of the protocol. S.P. provided discarded clinical samples. P.M., M.U., S.P., P.S., I.N.C. and J.P. provided insight into the experimental design and progress of the protocol, and on the results and implications of the work.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

1. Mutreja, A. *et al.* Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* **477**, 462–465 (2011).
2. Bryant, J.M. *et al.* Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: a retrospective cohort study. *Lancet* **381**, 1551–1560 (2013).
3. Harrison, E.M. *et al.* Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel *mecA* homologue *mecC*. *EMBO Mol. Med.* **5**, 509–515 (2013).
4. Croucher, N.J., Harris, S.R., Grad, Y.H. & Hanage, W.P. Bacterial genomes in epidemiology—present and future. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120202 (2013).
5. Holden, M.T. *et al.* A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res.* **23**, 653–664 (2013).
6. He, M. *et al.* Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat. Genet.* **45**, 109–113 (2013).
7. Eyre, D.W. *et al.* Detection of mixed infection from bacterial whole genome sequence data allows assessment of its role in *Clostridium difficile* transmission. *PLoS Comput. Biol.* **9**, e1003059 (2013).
8. Seth-Smith, H.M.B. *et al.* Whole-genome sequences of *Chlamydia trachomatis* directly from clinical samples without culture. *Genome Res.* **23**, 855–866 (2013).
9. World Health Organization. *Prevalence and Incidence of Selected Sexually Transmitted Infections* (World Health Organization, Geneva, 2011).
10. Global WHO alliance for the elimination of blinding trachoma by 2020. in *Weekly Epidemiological Report* Vol. 17 161–168 (World Health Organisation, Geneva, 2012).
11. Dean, D. *et al.* Predicting phenotype and emerging strains among *Chlamydia trachomatis* infections. *Emerg. Infect. Dis.* **15**, 1385–1394 (2009).
12. Klint, M. *et al.* High-resolution genotyping of *Chlamydia trachomatis* strains by multilocus sequence analysis. *J. Clin. Microbiol.* **45**, 1410–1414 (2007).
13. Pannekoek, Y. *et al.* Multi locus sequence typing of Chlamydiales: clonal groupings within the obligate intracellular bacteria *Chlamydia trachomatis*. *BMC Microbiol.* **8**, 42 (2008).
14. Pedersen, L.N., Pødenphant, L. & Møller, J.K. Highly discriminative genotyping of *Chlamydia trachomatis* using omp1 and a set of variable number tandem repeats. *Clin. Microbiol. Infect.* **14**, 644–652 (2008).
15. Jeffrey, B.M. *et al.* Genome sequencing of recent clinical *Chlamydia trachomatis* strains identifies loci associated with tissue tropism and regions of apparent recombination. *Infect. Immun.* **78**, 2544–2553 (2010).
16. Harris, S.R. *et al.* Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. *Nat. Genet.* **44**, 364–366 (2012).
17. Joseph, S.J., Didelot, X., Gandhi, K., Dean, D. & Read, T.D. Interplay of recombination and selection in the genomes of *Chlamydia trachomatis*. *Biol. Direct.* **6**, 28 (2011).
18. Andersson, P., Klein, M., Lilliebridge, R.A. & Giffard, P.M. Sequences of multiple bacterial genomes and a *Chlamydia trachomatis* genotype from direct sequencing of DNA derived from a vaginal swab diagnostic specimen. *Clin. Microbiol. Infect.* **19**, E405–E408 (2013).
19. Melnikov, A. *et al.* Hybrid selection for sequencing pathogen genomes from clinical samples. *Genome Biol.* **12**, R73 (2011).
20. Depledge, D.P. *et al.* Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS ONE* **6**, e27805 (2011).
21. Bos, K.I. *et al.* A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506–510 (2011).
22. Kent, B.N. *et al.* Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) determined by targeted genome capture. *Genome Biol. Evol.* **3**, 209–218 (2011).
23. Geniez, S. *et al.* Targeted genome enrichment for efficient purification of endosymbiont DNA from host DNA. *Symbiosis* **58**, 201–207 (2012).
24. Lasken, R.S. Genomic sequencing of uncultured microorganisms from single cells. *Nat. Rev. Microbiol.* **10**, 631–640 (2012).
25. Yilmaz, S. & Singh, A.K. Single cell genome sequencing. *Curr. Opin. Biotechnol.* **23**, 437–443 (2012).
26. Fitzsimons, M.S. *et al.* Nearly finished genomes produced using gel microdroplet culturing reveal substantial intraspecies genomic diversity within the human microbiome. *Genome Res.* **23**, 878–888 (2013).
27. McLean, J.S. *et al.* Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum. *Proc. Natl. Acad. Sci. USA* **110**, E2390–E2399 (2013).
28. Putman, T.E., Suchland, R.J., Ivanovitch, J.D. & Rockey, D.D. Culture-independent sequence analysis of *Chlamydia trachomatis* in urogenital specimens identifies regions of recombination and in-patient sequence mutations. *Microbiology* **159**, 2109–2117 (2013).
29. Skjerve, E., Rørvik, L.M. & Olsvik, O. Detection of *Listeria monocytogenes* in foods by immunomagnetic separation. *Appl. Environ. Microbiol.* **56**, 3478–3481 (1990).
30. Niesters, H.G.M., Quint, W.G.V. & Herbrink, P. Antigen capture directed polymerase chain reaction for detection of *Chlamydia trachomatis*. in *Techniques in diagnostic pathology* Vol. 2 (eds. Bullock, G., Van Velzen, D. & Warhol, M.J.) (Academic Press, 1991).
31. Fratamico, P.M., Schultz, F.J. & Buchanan, R.L. Rapid isolation of *Escherichia coli* O157:H7 from enrichment cultures of foods using an immunomagnetic separation method. *Food Microbiol.* **9**, 105–113 (1992).
32. Hedrum, A., Lundeberg, J., Påhlson, C. & Uhlén, M. Immunomagnetic recovery of *Chlamydia trachomatis* from urine with subsequent colorimetric DNA detection. *Genome Res.* **2**, 167–171 (1992).
33. Dean, F.B., Nelson, J.R., Giesler, T.L. & Lasken, R.S. Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* **11**, 1095–1099 (2001).
34. Hosono, S. *et al.* Unbiased whole-genome amplification directly from clinical samples. *Genome Res.* **13**, 954–964 (2003).
35. Jalal, H. *et al.* 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 (2006).
36. Su, H., Watkins, N.G., Zhang, Y.X. & Caldwell, H.D. *Chlamydia trachomatis*-host cell interactions: role of the chlamydial major outer membrane protein as an adhesin. *Infect. Immun.* **58**, 1017–1025 (1990).
37. Thornley, M.J. *et al.* 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 (1983).
38. Thornley, M.J., Zamza, S.E., Byrne, M.D., Lusher, M. & Evans, R.T. 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 (1985).
39. Zong, C., Lu, S., Chapman, A.R. & Xie, X.S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626 (2012).
40. Li, R., Li, Y., Kristiansen, K. & Wang, J. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**, 713–714 (2008).
41. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
42. Carlson, J.H., Porcella, S.F., McClarty, G. & Caldwell, H.D. Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains. *Infect. Immun.* **73**, 6407–6418 (2005).
43. Thomson, N.R. *et al. Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. *Genome Res.* **18**, 161–171 (2008).
44. Stephens, R.S. *et al.* Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**, 754–759 (1998).
45. Unemo, M. *et al.* The Swedish new variant of *Chlamydia trachomatis*: Genome sequence, morphology, cell tropism and phenotypic characterization. *Microbiology* **156**, 1394–1404 (2010).
46. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
47. Chitsaz, H. *et al.* Efficient *de novo* assembly of single-cell bacterial genomes from short-read data sets. *Nat. Biotechnol.* **29**, 915–921 (2011).
48. Peng, Y., Leung, H.C.M., Yiu, S.M. & Chin, F.Y.L. IDBA-UD: A *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
49. Assefa, S., Keane, T.M., Otto, T.D., Newbold, C. & Berriman, M. ABACAS: algorithm based automatic contiguation of assembled sequences. *Bioinformatics* **25**, 1968–1969 (2009).