The genome of *Onchocerca volvulus*, agent of river blindness

James A. Cotton¹, Sasisekhar Bennuru², Alexandra Grote³, Bhavana Harsha¹, Alan Tracey¹, Robin Beech⁴, Stephen R. Doyle¹, Matthew Dunn¹, Julie C. Dunning Hotopp⁵, Nancy Holroyd¹, Taisei Kikuchi⁶, Olivia Lambert⁷, Amruta Mhashilkar⁶, Prudence Mutowo⁷, Nirvana Nurismilu⁸,⁹, Jose M. C. Ribeiro¹⁰, Matthew B. Rogers¹¹, Eleanor Stanley¹, Lakshmipuram S. Swapna⁸, Isheng J. Tsai¹², Thomas R. Unnasch⁶, Denis Voronin¹², John Parkinson⁸,⁹,¹³, Thomas B. Nutman², Elodie Ghedin¹⁴*, Matthew Berriman¹⁵* and Sara Lustgarten¹²*

Human onchocerciasis is a serious neglected tropical disease caused by the filarial nematode *Onchocerca volvulus* that can lead to blindness and chronic disability. Control of the disease relies largely on mass administration of a single drug, and the development of new drugs and vaccines depends on a better knowledge of parasite biology. Here, we describe the chromosomes of *O. volvulus* and its *Wolbachia* endosymbiont. We provide the highest-quality sequence assembly for any parasitic nematode to date, giving a glimpse into the evolution of filarial parasite chromosomes and proteomes. This resource was used to investigate gene families with key functions that could be potentially exploited as targets for future drugs. Using metabolic reconstruction of the nematode and its endosymbiont, we identified enzymes that are likely to be essential for *O. volvulus* viability. In addition, we have generated a list of proteins that could be targeted by Federal-Drug-Agency-approved but repurposed drugs, providing starting points for anti-onchocerciasis drug development.

The filaria are a group of tissue-dwelling parasitic nematodes of vertebrates that are spread by blood-feeding arthropods. *Onchocerca volvulus* is the most pathogenic and is the agent of onchocerciasis (or ‘river blindness’), a leading cause of morbidity and socioeconomic loss for the world’s poorest populations¹. Approximately 17 million people are still infected with *O. volvulus*, predominantly in Africa². Infections are chronic and manifest clinically as debilitating skin disease and—in 1.2 million people—vision impairment or blindness. First-stage larvae, known as microfilariae (L1/Lf), are produced by fertile female worms residing within onchocercomata (nodules). They migrate to the skin and other organs (for example, the anterior chamber of the eye), where they induce inflammatory reactions that are responsible for most onchocerciasis-related pathology.

Onchocerciasis was identified by the World Health Organization (WHO) as a potential candidate for disease elimination through annual (or semiannual) mass drug administration (MDA) of ivermectin³, an approach that has eliminated onchocerciasis from all but two countries in the Americas⁴. Ivermectin is solely microfilaricidal, which means it must be given over decades, past the lifespan of the long-lived adult worms⁵,⁶. Moreover, in much of Central Africa where *Loa loa* is co-endemic with *O. volvulus*, ivermectin cannot be used due to the risk of *Loa*-associated irreversible neurological severe adverse events and death⁷. Despite its success in Latin America and small foci in Africa, elimination of onchocerciasis in Africa is unlikely to be achieved within the proposed timeframes solely through MDA with ivermectin⁸. Reliance on a single drug also increases the potential for the emergence of ivermectin-resistant *O. volvulus*⁹, making the development of new drugs or novel therapies imperative.

To gain better insight into this important but neglected pathogen, we generated a high-quality genome assembly of *O. volvulus*. Although draft whole-genome assemblies exist for other filarial nematode species¹⁰⁻¹⁺, we have, for the first time in any species of this group, reconstructed whole chromosomes, including a sex chromosome. We also generated a genome assembly for its obligate intracellular endosymbiont *Wolbachia* (wOv) and transcriptional data from eight life stages. Our analysis highlights the metabolic interplay between *O. volvulus* and wOv as a path to novel drug targets.

¹Welcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. ²Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland 20892, USA. ³Center for Genomics and Systems Biology, Department of Biology, New York University, New York, New York 10003, USA. ⁴Institute of Parasitology, McGill University, Montreal, Quebec H9X 3V9, Canada. ⁵Institute for Genome Sciences, Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. ⁶Global Health Infectious Disease Research Program, Department of Global Health, College of Public Health, University of South Florida, Tampa, Florida 33612, USA. ⁷European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK. ⁸Department of Computer Science, University of Toronto, Toronto M5S 3G4, Canada. ⁹Division of Molecular Structure and Function, Research Institute, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ¹⁰Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland 20892, USA. ¹¹Children’s Hospital of Pittsburgh, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15224, USA. ¹²New York Blood Center, New York, New York 10065, USA. ¹³Departments of Biochemistry and Molecular Genetics, University of Toronto, MSS 1A8, Canada. ¹⁴College of Global Public Health, New York University, New York, New York 10003, USA. ¹⁵Present addresses: Division of Parasitology, Faculty of Medicine, University of Miyazaki, Miyazaki, Japan (T.K.); Eagle Genomics Ltd, Babraham Hall, Babraham Research Campus, Babraham, Cambridgeshire CB22 3AT, UK (E.S.); Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan (I.J.T.). ²-e-mail: slustgiman@nybloodcenter.org; mb4@sanger.ac.uk; elodie.ghedin@nyu.edu
Results

O. volvulus genome structure and features. The 97 Mb nuclear genome of O. volvulus comprising three autosomes and a pair of sex chromosomes was assembled using a combination of sequencing, an optical map and manual improvement. Four large scaffolds (16–31 Mb) comprise 94% of the assembly and seven out of eight of their ends correspond to ends of optical maps or telomeric repeats (Fig. 1a and Supplementary Fig. 1). These scaffolds thus represent essentially complete chromosomes of O. volvulus. This is the highest-quality assembly for any parasitic nematode (Supplementary Tables 1 and 2) and only the fourth nematode species for which chromosome sequences are available. The assembly also includes the mitochondrial genome and complete wOv genome.

By analysing sequence data from male and female worms, we identified scaffold OM2 as the X chromosome (Supplementary Fig. 2a). For a long contiguous portion (22.2 of 25.5 Mb of the scaffold) the median depth of coverage of male sequence data was 50% that for the rest of the genome. The same region had a coverage over. Other scaffolds lack data from juvenile females and show low coverage in adult female libraries (Supplementary Fig. 2c,d), allowing us to identify ~1.2 Mb as the potentially Y-specific sequence. Only this small portion of the Y chromosome is present in our assembly as this chromosome is largely pseudo-autosomal (Fig. 1a) and so mostly assembled with the X chromosome. The small extent of sequence divergence between X and Y and the presence of an extensive PAR confirm that this evolved recently from an ancestral XO karyotype and contrasts with the situation in other nematodes that have X and Y chromosomes, where the Y is largely unique, repeat-rich and degenerate. Furthermore, one region of the PAR adjacent to the X-specific region shows an excess of heterozygous sites missing in the juvenile female sample (Supplementary Fig. 2b). This represents a region in which X and Y have begun to diverge, but where the two chromosomes are sufficiently similar that they are still represented by the same region of the assembly. We propose that this is a region of more recent divergence between X and Y chromosomes than the X- and Y-specific regions. This suggests a process of sex chromosome evolution similar to that observed in other systems, in which recombination suppression and subsequent divergence between sex chromosomes occurs in a patchy way, leading to different 'strata' of divergence.

We identified over 97% of a conserved set of eukaryotic genes in this assembly, with five of the six missing genes also missing from all other filarial genome assemblies, presumably reflecting ancestral gene losses from this group (Supplementary Table 4). O. volvulus shows large-scale variation in gene density, GC content and repeat density (Fig. 1b and Supplementary Table 5), establishing that these patterns are present in nematodes beyond Caenorhabditis spp. The pattern of variation on chromosomes 1 and X is more complex, with two peaks in gene density and GC. We propose that this pattern is due to the origins of these chromosomes as fusions of two ancestral chromosomes. For comparison of gene content, we have also generated sequence data for the related
NATURE MICROBIOLOGY ARTICLES

The majority (~91%) of genes had orthologues in other nematodes, with ~9% (1,173) being O. volvulus-specific, with little or no homology to genes annotated in other helminths (Supplementary Table 6). Predicted proteins were classified into functional categories21 (Supplementary Fig. 3a), although 44% of O. volvulus proteins have no predicted function. The distinctive biology of O. volvulus is likely to be underpinned by genes with potentially novel functions and with relatively few homologues in other helminth parasites. Of the O. volvulus-specific genes, 92% encode putative proteins of unknown function, of which 7% are potentially secreted (Supplementary Fig. 3b and Supplementary Table 6).

A total of 3,152 protein-coding genes were present on the large scaffolds showing that both genomes are predominantly co-linear, but not all genes have predicted one-to-one orthologues (Supplementary Fig. 4).

We further investigated the evolutionary history of O. volvulus using Ensembl Compara22. This produced high support for the traditional classification of this group as found in previous molecular work23 and confirmed the close relationship between O. volvulus and O. ochengi (Fig. 2). We observed many gene encoding putative G-protein coupled receptors, translational release factors, calcium-binding proteins and membrane proteins involved in cell–cell communication.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
duplication events of potential biological importance in the evolutionary history of *O. volvulus* (Supplementary Table 8 and Supplementary Data). For example, the gene family encoding the α subunit of the enzyme collagen prolyl 4-hydroxylase (C-4PHα), an important enzyme for collagen synthesis in the cuticle of nematodes, has been repeatedly duplicated in the lineage leading to *O. volvulus*, resulting in a paralogous family of 16 genes. Two copies are clearly truncated and may be pseudogenes, while at least 11 of these paralogues are full-length, have the conserved catalytic sites residues24 (Supplementary Fig. 5) and are predicted to be secreted. This drastic expansion probably reflects the requirement for changes in collagen composition during cuticle remodelling throughout *O. volvulus* development.

Similarly, three G-protein-coupled receptor (GPCR) gene families show a remarkable expansion in *O. volvulus* (Fig. 2 and Supplementary Table 8). In general, filarial genomes lack several GPCR gene families present in *C. elegans*, including Srr, Sra and Srb receptors9. However, unlike the other filariae, *O. volvulus* contains members of all other GPCR families represented in *C. elegans* (Fig. 3). For example, it retains the Str family involved in odorant detection, a family not found in other filariae but present in *Ascaris suum*, and the root-knot nematode *Meloidogyne hapla*, perhaps reflecting responses to specific cues in their environments9. The maintenance of Str and other families of GPCRs may suggest their importance throughout the developmental stages of *O. volvulus*, but could also reflect the high quality of the sequence assembly and annotation. Further underlining the importance of GPCRs to *O. volvulus* biology, the Srx and Srsx families are among the most duplicated genes in the *O. volvulus* genome, most striking of which is the expansion of the Srx family, with 35 duplications inferred to be specific to *O. volvulus*.

In contrast to GPCRs, the nuclear hormone receptor (NHR) gene family is particularly sparse. NHRs are important transcriptional modulators that regulate cellular differentiation, homeostasis and reproduction in nematodes and other animals25. *C. elegans* has over 270 NHRs—the highest number identified in any species—but the filarial nematodes have a far smaller repertoire. *Brugia malayi* has 50 and *O. volvulus* even fewer, with only 24 classical NHRs and 8 orphan receptors identified (Supplementary Table 9 and Supplementary Fig. 6). Despite this, *O. volvulus* has orthologues of all five *C. elegans* NHR genes that exclusively participate in molting and metamorphosis26,27 (Supplementary Table 10). Interestingly, the ecdysone receptor (OVOC9104) was gained at the base of the filarial lineage. Ecdysone and related hormones regulate molting and metamorphosis in arthropods, so it is tempting to speculate that this gain in NHRs may be an adaptation to the insect host.

Control of the neuromusculature is a principal target for many anthelmintic drugs, where they bind and activate the pantameric ligand-gated ion channels (pLGICs) that mediate fast-synaptic signalling. Like other filarial nematodes, *O. volvulus* encodes fewer pLGIC subunit genes (only 48 compared to 120 in *C. elegans*28 and other clad V nematodes29), but has retained genes encoding subunits of the glutamate- and GABA-gated chloride channels (Supplementary Fig. 7) that are putative targets for ivermectin30. In filarial nematodes, ivermectin paralyses or kills microfilariae gradually, but also suppresses microfilarial production from adult females. *O. volvulus* encodes an orthologue of the *avr-14* ivermectin target that in *B. malayi* is expressed within reproductive tissue; this may explain the effect of ivermectin on microfilarial production31.

Although the emergence of drug resistance in human parasites has been less rapid than in parasitic nematodes of livestock, cases of unexpected earlier reappearance of *O. volvulus* microfilariae following ivermectin treatment have been reported31. The mechanisms of ivermectin resistance remain unclear, but polymorphism of P-glycoprotein (Pgp) drug transporters in other nematodes has been linked to resistance, together with evidence that Pgp inhibitors can increase drug efficacy32,33. The Pgps are members of the ATP-binding cassette (ABC) transporter family. The genome of *O. volvulus* contains relatively few ABC transporter orthologues, with only 10 compared to 54 in *C. elegans*31.

**Metabolic reconstruction and identification of potential drug targets.** To examine the metabolic capabilities of *O. volvulus* and *L. loa*, a filarial parasite without a Wolbachia endosymbiont, we performed genome-scale metabolic reconstructions (Supplementary Tables 11–13). The resultant networks comprise 767 reactions (378 distinct enzymes) for *O. volvulus* and 648 reactions (301 enzymes) for *L. loa*. Each share a core of 628 reactions, with the majority (100 of 139) of additional reactions in the *O. volvulus* reconstruction contributed by wOv. Next, we performed flux balance analysis (FBA)34,35 to investigate the impact of single reaction knockouts on parasite growth. Metabolites were made freely available across reactions, an assumption that will minimize false-positive essential reaction predictions at the expense of false-negative predictions (for details see Methods).
For *O. volvulus*, FBA predicted 71 essential reactions (Table 1 and Supplementary Tables 11 and 12). For *L. loa*, 112 reactions were predicted to be essential (70 common to *O. volvulus* essential reactions), including 23 transport reactions (Supplementary Table 12). Only threonine transport was essential for *L. loa*; any reliance by *O. volvulus* on this alternative pathway could be exploited to selectively target it. Insights into possible repurposed drugs or new drug targets. We have also investigated potential *O. volvulus* targets of currently available Federal Drug Agency (FDA)-approved drugs. We excluded *O. volvulus* loci, as repurposing of antibacterial compounds for anti-*Wolbachia* chemotherapy has been extensively investigated. This strategy identified 51 *O. volvulus* proteins, mostly enzymes and proteins involved in ion transport and neurotransmission, that 85 drugs may target (Supplementary Table 14). We used the WHO Anatomical Therapeutic Classification (ATC) to remove from the set those drugs classed as anti-neoplastics (these are likely to have intolerable side effects for anthelmintic use), insights into possible repurposed drugs or new drug targets.

**Table 1 | Overview of the metabolic pathways included in the metabolic reconstructions.**

| Superpathway | Metabolic pathway | Number of reactions in pathway | Number of reactions predicted essential* |
|--------------|-------------------|-------------------------------|-----------------------------------------|
|               |                   | *L. loa* | *Ov and wOv* | *wOv only* | *L. loa* | *Ov and wOv* | *wOv only* |
| Amino acid metabolism | Alanine, aspartate and glutamate metabolism | 14 | 15 | 3 | 1 | 0 |
|               | Arginine and proline metabolism | 26 | 25 | 2 | 3 | 3 |
|               | Biosynthesis of amino acids | 37 | 41 | 9 | 3 | 0 |
|               | Cysteine and methionine metabolism | 17 | 19 | 2 | 4 | 0 |
|               | Glycine, serine and threonine metabolism | 17 | 17 | 3 | 1 | 0 |
| Biosynthesis of other secondary metabolites | Streptomyacin biosynthesis | 4 | 4 | 0 | 2 | 0 |
| Carbohydrate metabolism | Amino sugar and nucleotide sugar metabolism | 23 | 23 | 3 | 4 | 0 |
|               | Butanoate metabolism | 8 | 8 | 0 | 1 | 0 |
|               | Inositol phosphate metabolism | 12 | 12 | 0 | 3 | 3 |
|               | Pentose phosphate pathway | 17 | 18 | 0 | 2 | 2 |
|               | Propanoate metabolism | 14 | 14 | 1 | 1 | 0 |
|               | Pyruvate metabolism | 20 | 20 | 1 | 1 | 0 |
| Energy metabolism | Methane metabolism | 16 | 16 | 0 | 2 | 2 |
| Lipid metabolism | Arachidonic acid metabolism | 16 | 14 | 0 | 3 | 3 |
|               | Biosynthesis of unsaturated fatty acids | 14 | 15 | 2 | 7 | 2 |
|               | Fatty acid biosynthesis | 3 | 15 | 32 | 1 | 0 |
|               | Fatty acid degradation | 32 | 32 | 0 | 19 | 0 |
|               | Fatty acid elongation | 28 | 28 | 0 | 24 | 0 |
|               | Fatty acid metabolism | 48 | 57 | 32 | 31 | 2 |
|               | Glycerolipid metabolism | 9 | 9 | 0 | 2 | 2 |
|               | Glycerophospholipid metabolism | 22 | 22 | 1 | 6 | 6 |
|               | Sphingolipid metabolism | 16 | 16 | 0 | 5 | 5 |
| Metabolism of cofactors and vitamins | Nicotinate and nicotinamide metabolism† | 11 | 12 | 1 | 1 | 0 |
|               | One carbon pool by folate | 10 | 10 | 6 | 1 | 0 |
| Metabolism of other amino acids | Pantothenate and CoA biosynthesis | 11 | 11 | 1 | 3 | 3 |
|               | Beta-alanine metabolism | 9 | 9 | 0 | 1 | 1 |
|               | Glutathione metabolism | 19 | 18 | 0 | 2 | 2 |
| Metabolism of terpenoids and polyketides | Terpenoid backbone biosynthesis | 12 | 12 | 2 | 8 | 0 |
|               | Tetracycline biosynthesis | 1 | 1 | 0 | 1 | 0 |
| Nucleotide metabolism | Purine metabolism | 53 | 63 | 9 | 7 | 3 |
|               | Pyrimidine metabolism | 47 | 51 | 7 | 7 | 5 |
| Transport reactions | Extracellular transport | 43 | 43 | 0 | 23 | 18 |
|               | Total (non-redundant) number of reactions in pathways | 428 | 455 | 72 | 86 | 49 |
|               | Total number of reactions in reconstructions | 777 | 796 | 100 | 112 | 70 |

*One reaction was essential either from *L. loa* (with or without genetic evidence); *O. volvulus* (Ov) and Wolbachia (wOv) with reactions in the *O. volvulus* reconstruction (both from gene annotation and added in pathway gap-filling) not being contributed by wOv only; or wOv with reactions in the *Ov* reconstruction supported by genetic evidence only from wOv. Only a single reaction, in the nicotinate and nicotinamide pathway, is essential and uniquely provided by wOv. Transport reactions for both nematode and Wolbachia lack genetic evidence therefore making wOv only and Ov and wOv transporters indistinguishable. Note that of the 18 transport reactions essential to *O. volvulus*, all are essential to *L. loa* as well except for threonine transport. Also reported is the total number of reactions belonging to the pathways listed here, as well as the total number of reactions in the reconstructions.

Insights into possible repurposed drugs or new drug targets. We have also investigated potential *O. volvulus* targets of currently available Federal Drug Agency (FDA)-approved drugs. We excluded *O. volvulus* loci, as repurposing of antibacterial compounds for anti-*Wolbachia* chemotherapy has been extensively investigated. This strategy identified 51 *O. volvulus* proteins, mostly enzymes and proteins involved in ion transport and neurotransmission, that 85 drugs may target (Supplementary Table 14). We used the WHO Anatomical Therapeutic Classification (ATC) to remove from the set those drugs classed as anti-neoplastics (these are likely to have intolerable side effects for anthelmintic use), Insights into possible repurposed drugs or new drug targets.
resulting in a set of 42 drugs. By screening these targets based on sequence features, we identified 16 *O. volvulus* proteins likely to be good drug targets (Table 2).

In a second approach, we leveraged known mechanisms of action of particular drug classes on specific nematode protein families to prioritize from a wider set of potential targets. Among the *O. volvulus* protein families with at least one drug target, those comprising zinc finger C4, NHR, dopamine neurotransmitter, GPCR, serotonin receptors, GABA receptor and ion channels could be considered privileged based on the extensive list of drugs by which they may be targeted. The genome sequence reveals the full repertoire of these families, providing a rich resource of potential leads for future drug development efforts (Supplementary Fig. 8). Finally, there is the possibility that existing antiparasitic drugs such as levamisole and morantel, which each target distinct classes of acetylcholine-gated cation channels (AcHR), should be revisited as antifilarial agents. Levamisole targets unc-38 and unc-63 (duplicated in *O. volvulus*), while morantel targets acr-26 and acr-27 (OVOC7603; a distinct class of AcHR found only in parasitic nematodes). Importantly, these targets are all conserved between *O. volvulus* and *B. malayi*.

The availability of the complete genome sequence will enhance the opportunity to investigate the mechanisms of action of existing classes of anthelmintics.

**wOv and lateral gene transfer.** In addition to the filarial genome, we assembled the complete 956 kb genome of wOv (Supplementary Table 15). The *O. volvulus* genome was examined for evidence of lateral gene transfers (LGTs) from wOv to its host. These are termed nuclear Wolbachia transfers or nuwts. In total, 531 putative nuwts were identified. Interestingly, one LGT event (OVOC_OM22,596,036–22,595,692) was found to have a best match to *Medichloria mitochondrii*, the endosymbiont of tick mitochondria (Supplementary Table 16). Seven nuwts were larger than 1 kb (maximum of 8.1 kb), four of which were confirmed by amplification, cloning and sequence verification of the Wolbachia–nematode junction (Supplementary Table 17). In a gene-based analysis aimed at detecting more divergent nuwts, a majority (98%) of the 576 regions that contain nuwt open reading frames were fragmented, but four were full-length, potentially functional genes, with relatively few overall mutations, suggesting they may be full-length merely because they are recent transfers (Supplementary Table 16). Overall, nuwts appear to be largely nonfunctional in this genome.

**Insights into host–parasite interactions.** *O. volvulus*, like the other filarial parasites, interacts with its definitive human host and its intermediate arthropod host (*Simulium* spp.) during its life cycle. These parasites must have their own innate immune system to protect them from microbial pathogens, but they also are thought to have evolved mechanisms to subvert both human and insect host defence mechanisms. Although molecules such as immunoglobulins or Toll-like receptors (TLRs) are absent in all filarial nematodes, *O. volvulus* and other filariae sequenced to date have homologues of some downstream proteins from the TLR signalling pathway (Supplementary Table 18). The innate immune system encoded by the *O. volvulus* genome also includes C-type lectins, galectins, jacobins and scavenger receptors. Similar to the other filariae, *O. volvulus* does not appear to produce the antibacterial peptides seen in *C. elegans* (and in other non-filarial nematodes), although it is possible that different peptides are produced by filariae.

Analysis of the putative proteome of *O. volvulus* identified a number of human cytokine and chemokine mimics and/or antagonists (Supplementary Table 18). In addition, the *O. volvulus* genome encodes 12 serine protease inhibitors (SPIs) including serpins and small SPIs, and five cysteine protease inhibitors including cystatins. Both protein families have been shown to interfere with antigen processing and presentation and are potentially involved in immune regulation and in parasite interference with the host immune response. The *O. volvulus* genome also encodes proteins with sequences similar to those of human autoantigens, some of which have been implicated in inducing cross-reactive antibodies that have been connected with the pathogenesis of posterior eye disease and the nodding syndrome.

**Discussion**

The *O. volvulus* genome assembly represents the highest-quality genomic data available for any non-model nematode species. It will be a critical resource for research on other filarial nematodes.
that reside in different niches within the host, for which only draft genomes are available. Comparative analysis with other nematode genomes has confirmed the recent evolution of an XY karyotype suggesting ancient chromosomal fusions led to the formation of the *Onchocerca* chromosomes. It should be possible to confirm this hypothesis when chromosome-scale assemblies for filarial nematodes without these fusions become available.

The genome data constitute an invaluable and comprehensive resource for the development of new and urgently needed interventions against onchocerciasis and other filariases. In particular, we describe the orthologues and paralogues of known or suspected targets of existing anthelmintic compounds and identify targets of other licensed compounds that could show activity against *O. volvulus*. Our analysis of the targets of these drugs highlights several protein families that could guide further drug discovery in *Onchocerca*. Finally, we performed metabolic reconstructions of this hypothesis when chromosome-scale assemblies for *O. volvulus* became available.

### Methods

#### Parasite material for genome sequencing

All *O. volvulus* parasite material used for genome sequencing was collected in the research facility at the Tropical Medicine Research Station, Kumba, Cameroon. Written informed consent was obtained. In cases of illiteracy, a literate witness signed and a thumbprint was made by the participant. Institutional Review Board (IRB) approvals were obtained from both the New York Blood Center and from the Tropical Medicine Research Station, Kumba (Protocols 321 and 01, respectively). The individuals who consented to participate in the study were born or had resided for more than ten years in villages around Kumba. They were confirmed to have microfilariemia in their skin snips and clinical symptoms of disease, such as dermatitis, nodules and oculare lesions. None of the subjects had received ivermectin treatment before the study. The adult worm samples were obtained as part of a noductectomy campaign conducted in villages surrounding Kumba in 1996–1998 and in 2006. Nodules were excised under sterile conditions and were treated with collagenase overnight, following the protocol of Schulz-Key and colleagues. Briefly, cleaned individual nodules were immersed in 0.5% collagenase (Sigma grade IV) in RPMI 1640 containing 10% FCS + 200 units of penicillin and 200 μg ml⁻¹ streptomycin. The flat tubes containing the nodule were then placed in a rocking water bath and incubated at 35 °C until the tissue was completely digested. Once digested, the liberated worms were unraveled from residual tissue with mounted needles under a dissecting scope and then washed in several changes of RPMI. Individual female worms were snap-frozen in Eppendorf tubes with liquid nitrogen. They were then stored and shipped in liquid nitrogen and, upon arrival in New York, stored at −80 °C until shipment on dry ice to the Wellcome Trust Sanger Institute.

Freshly dissected *O. volvulus* L3s were also cryopreserved according to the method described by Cupp and colleagues and were shipped to the New York Blood Center in liquid nitrogen and, upon arrival in New York, were stored in liquid nitrogen. To collect sexed juvenile adult worms (40 days in culture), thawed and washed L3s were cultured for 14 days in the presence of peripheral blood in achieving timely elimination of onchocerciasis, and these targets are now important candidates for experimental testing and validation.

#### Table 2 The top 16 *O. volvulus* targets and their predicted drugs.

| Ov gene ID | Function of family | Drugs (WHO name) | ATC level 3 classification |
|------------|-------------------|------------------|---------------------------|
| OVOCC723 | Ribonucleotide reductase, small chain | Fludarabine phosphate, clofarabine, gemcitabine, (hydroxyurea), gallium nitrate | Antimetabolites, other antineoplastic agents. |
| OVOCC1146 | Calcineurin-like phosphoesterase | Tiagabine | Antiepileptics |
| OVOCC15 | Sodium neurotransmitter symporter family | Impiramine, clomipramine, amitriptyline, nortriptyline, protriptyline, amoxapine, fluoxetine, citalopram, paroxetine, sertraline, fluvoxamine, escitalopram, trazodone, nefazodone, venlafaxine, milnacipran, duloxetine, desvenlafaxine, vilazodone, vortioxetine, (levomilnacipran, chlorpheniramine) | Antidepressants (unclassified) |
| OVOCC1974 | Ion transport protein | Dronedarone, (nimodipine, felodipine) | Antiarrhythmics, class I and III (selective calcium channel blockers with mainly vascular effects) |
| OVOCC1986 | KQT-1 potassium channel | Dronedarone | Antiarrhythmics, class I and III |
| OVOCC111 | Receptor family ligand binding region | Baclofen, oxybate | Muscle relaxants, central acting agents, null |
| OVOCC1975 | G-protein coupled GABA receptor activity | Metformin | Blood glucose lowering drugs excluding insulin |
| OVOCC12632 | NADH-ubiquinone oxidoreductase chains 1, 4 and 5; probable NADH dehydrogenase [ubiquinone] iron-sulfur protein 7 | Metformin | Blood glucose lowering drugs excluding insulin |
| OVOCC12635 | 3-oxo-5-α-steroid 4-dehydrogenase | Dutasteride | Drugs used in benign prostatic hypertrophy |
| OVOCC10592 | ERG2 and sigma receptor like protein | Pentazocine, (dextromethorphan) | Opioids, cough suppressants (other nervous system drugs) |
| OVOCC8585 | C-terminal tandem repeated domain in type 4 procollagen | Xiaflex, (ocripasmin) | Other drugs for disorders of the musculoskeletal system (other ophthalmologicals) |
| OVOCC10695 | Glycosyl transferase family 21 | Miglustat, eliglustat | Other alimentary tract and metabolism products |
| OVOCC1410 | Amidase | Acetaminophen | Other analgesics and antipyretics |

The table shows anatomical therapeutic classification groups for drugs identified as potential repurposing candidates. For drugs in parentheses, the mechanism is shown in similar style parentheses in the ‘ATC level 3 classification’ column.
mononuclear cells (PBMCs) (1.5 × 10^5 per well of a 96-well plate) and then over a period of optimal internal Ebroflats (plate) (S. Lustigman, unpublished data). On day 40, they were separated into juvenile female and male worms based on their size (males being much smaller than females) and the morphology of the posterior end51. They were then frozen individually, stored at –80 °C until shipment on dry ice to the Wellcome Trust Sanger Institute.

Cows that grazed in northern Cameroon, where O. ochengi is highly endemic, were brought to abattoirs located in Douala, Cameroon. Subcutaneous nodules containing adult O. ochengi worms were identified on the umbilical skin of slaughtered infected cows. Adult worm masses containing one viable adult female and two viable adult males were then harvested from the punctured skins by dissection of the nodule with a sterile razor blade and then snap-frozen in liquid nitrogen. The material was transported to the USA in liquid nitrogen and, upon arrival in New York, stored at –80 °C.

Parasite material for transcriptomics. All parasite material was prepared in the Tropical Medicine Research Station, Kumba, Cameroon, between 1993 and 1999, except for female worm samples, which were from both Cameroon and Ecuador, where samples were collected as part of a previous study52. L3 were obtained from Simulium damnosum flies 7 days after infection with skin microfilariae, as described previously52,53. To obtain moulting larvae, freshly dissected L3s were cultured in vitro in groups of ten larvae in 96-well plates containing a 1:1 mixture of Icovex's modified Dulbecco's medium and NCTC-135, 20% fetal calf serum and antibiotic-antimycotic solution (Life Technologies) for 3 days at 37 °C. Larvae were collected immediately after dissection (L3) or after 1, 2 or 3 days in culture, washed with Tris-HCl buffered saline and snap-frozen in liquid nitrogen. A random sample (400) of the optical contigs, two additional contigs could be made between four of the six superscaffolds, with evidence of these contigs coming from read pairs joining the two ends in one case sequence similarity between the two ends, and synteny with C. elegans and B. malayi optical map (unpublished). Supporting that, in each case, the two scaffolds being joined belong to the same chromosome. The v4 genome assembly includes just these two additional joins, with minimum sizes for the introduced gaps (approximately 100 and 300 kb) estimated as the lengths of unjoined 'overhangs' in the optical mapping molecules. The final assembly thus has all four chromosome pairs represented by large contiguous scaffolds (Supplementary Table 2).

To assess the completeness of the assemblies, we ran CEGMA v2 (ref. 20), which reports the percentage of 248 highly conserved eukaryotic gene families that are present as full or partial genes in the assembly. For most eukaroytes, we would expect to see nearly 100% of CEGMA families represented by a full gene in the genome. Thus, CEGMA provides a measure of the completeness of the assembly for a species.

The Wolbachia endosymbiont genome sequence was assembled by first identifying scaffolds from the initial V5V assembly containing similarity to published sequences of five published lambda phage clones containing Wolbachia sequence from Fenn and colleagues54 (a total of 70.8 kb). Nuclermo identified a set of eight contigs totalling 113 kb that all had >99% similarity to the published clones over at least 300 bp and that all had very similar coverage (43–46x) in the published data. These contigs had more dispersed and would not be functionally linked (Brugia malayi Wolbachia sequence55). Applying a cutoff of depth between 42 and 47x coverage and PROMER similarity of at least 75% over 250 bp with the B. malayi Wolbachia sequence identified 44 contigs covering 926 kb that were putatively assigned to the O. volvulus Wolbachia sequence. Manual finishing using GAPFiner allowed us to order, extend and link these contigs to complete the Wolbachia genome assembly.

Transcriptome sequencing and analyses. High-throughput transcriptome data were generated from the RNA of O. volvulus stage-specific parasites: nodular microfilariae, skin microfilariae, L2, L3, L3D1, L3D3, adult male and adult female worms. For all larval stages and adult worms, RNA was prepared using TRIzol and lysing matrix D (1.4 mm ceramic matrix) and a Fastprep24 (MP Biomedicals). RNAseq libraries were prepared following the RNAseq protocols of the Illumina mRNASeq Sample Prep kit and the Illumina TruSeq kit (Illumina). Transcriptome libraries were sequenced on Illumina HiSeq 2000 or MiSeq machines (Supplementary Table 19).

Gene prediction. Gene predictions were conducted by various methods available in MAKER version 2.2.8 (ref. 72). The MAKER annotation pipeline consists of four major steps that generate high-quality annotations by taking into account evidence from multiple sources. First, assembled contigs are filtered against RepeatMasker73, RepBase74 and a species specific repeat library generated by Repeat Modeler (http://www.repeatmasker.org/RepeatModeler.html) using RepeatMasker (http://www.repeatmasker.org/) to identify and mask repetitive elements in the genome. Second, gene prediction is conducted by using self-trained76 and SNAP 2013-02-16 (ref. 77) are employed to generate de novo 46x) assembly using the unaligned reads (see above), before the adapter ligation stage. Fresh beads were then used for size selection. O. volvulus genomic DNA was used to generate a 3 kb mate pair library using a modified SOLiD 5500 protocol adapted for Illumina sequencing52.

Genome sequencing for assembly. Whole-genome sequence libraries (Supplementary Table 19) were generated from genomic DNA extracted from a single adult female O. volvulus worm and a single adult female O. ochengi worm, both of which were collected in 0.1 M EDTA. DNA extracted using the QIAamp DNA mini kit (Qiagen). The size of the genomic-tic tip. The ratio of host to parasite DNA was examined by qPCR using single-copy genes as markers. PCR-free 400–550 bp paired-end Illumina libraries were produced (one library per species) using a protocol based on a previously described method but75 then using Agencourt AMPure XP beads for sample clean-up and size selection. Genomic DNA was precipitated onto beads after each enzymatic stage with an equal volume of 20% polyethylene glycol 6000 and 2.5 M sodium chloride solution. Beads were not separated from the sample throughout the process until after the adapter ligation stage. Fresh beads were then used for size selection. O. volvulus genomic DNA was used to generate a 3 kb mate pair library using a modified SOLiD 5500 protocol adapted for illumina sequencing52.

Genome sequencing for assembly. Whole-genome sequence libraries (Supplementary Table 19) were generated from genomic DNA extracted from a single adult female O. volvulus worm and a single adult female O. ochengi worm, both of which were collected in 0.1 M EDTA. DNA extracted using the QIAamp DNA mini kit (Qiagen). The size of the genomic-tic tip. The ratio of host to parasite DNA was examined by qPCR using single-copy genes as markers. PCR-free 400–550 bp paired-end Illumina libraries were produced (one library per species) using a protocol based on a previously described method but75 then using Agencourt AMPure XP beads for sample clean-up and size selection. Genomic DNA was precipitated onto beads after each enzymatic stage with an equal volume of 20% polyethylene glycol 6000 and 2.5 M sodium chloride solution. Beads were not separated from the sample throughout the process until after the adapter ligation stage. Fresh beads were then used for size selection. O. volvulus genomic DNA was used to generate a 3 kb mate pair library using a modified SOLiD 5500 protocol adapted for illumina sequencing52.

Genome resequencing of sexed juvenile female and male worms. DNA was extracted from sexed juvenile male and female worms using the Promega Wizard kit, following the manufacturer's instructions. Then, 400 bp fragment Illumina libraries were produced (Supplementary Table 19) following the protocol used for genome sequencing (see above), but using 12 cycles of amplification.

Genome assemblies. The assembly of the genome was produced using a combination of sequencing, a de novo optical restriction map (Supplementary Fig. 1a) and extensive manual assembly improvement. The initial assembly for O. volvulus was produced from a combination of short-fragment paired-end and mate-pair Illumina libraries. Short paired-end sequence reads were first corrected and initially assembled using SGA v0.9.7.9 (ref. 58). This draft assembly was then used to calculate the distribution of k-mer s for all odd values of k between 41 and 81, using GenomeTools v1.3.7 (ref. 29). The k-mer length for which the maximum number of unique k-mers were present in the SGA assembly was then used as the k-mer setting for de Bruijin graph construction in a second assembly with Velvet v1.2.03 (ref. 60). The mate-pair library was then used to further scaffold this Velvet assembly using SPASPACE (ref. 61), followed by gap closing using Gapfiller (ref. 62) and IMAGE (ref. 63). A 'bin' assembly using the unaligned reads was incorporated into the main assembly and this assembly was scaffolded using SPASPACE. The O. ochengi genome assembly was produced using the above method except that only a short-fragment paired-end (PCR-free) library was used and therefore the SPASPACE scaffolding step was excluded. The O. ochengi assembly was found to be contaminated with sequences derived from the host (cow) genome. To identify contaminating contigs, PROMER (ref. 64) was used to compare the Bos taurus UMD 3.1 (ref. 65) assembly against the initial O. ochengi assembly and contigs of greater than 1000 bp were identified. Those representing 90% of the contig length were removed. This process removed ~20 Mb of the O. ochengi assembly and this cleaned version was used for all comparisons reported here.

For O. volvulus, the output of the automated assembly process above produced an 'interim' assembly, which was improved by an extensive manual finishing effort, using GAPS (ref. 66); scaffolds were extended, linked and, where possible, errors detected by REAPR (ref. 67) were fixed. The de novo optical map guided this process.

The improved sequence scaffolds (totalling 90.9 Mb) were aligned against the optical contigs using MapSolver. Any miss-joints were resolved and new potential joints were investigated and executed where possible. Further automated gap closure was undertaken during this finishing process using IMAGE and Gapfiller, and the accuracy of the consensus sequence was improved using TCOON2 (ref. 68) with both short-fragment and mate-pair Illumina reads. Finally, REAPR was run to detect and break missassembled regions based on the mate-pair library data. The v3 genome assembly resulting from this work was used for gene finding (Supplementary Table 2).

Following the assembly, a set of expressed sequence tags (ESTs) were produced (Supplementary Table 19) following the protocol used for genome assembly. The v4 genome assembly includes just these two additional joins, with minimum sizes for the introduced gaps (approximately 100 and 300 kb) estimated as the lengths of unjoined 'overhangs' in the optical mapping molecules. The final assembly thus has all four chromosome pairs represented by large contiguous scaffolds (Supplementary Table 2).

To assess the completeness of the assemblies, we ran CEGMA v2 (ref. 20), which reports the percentage of 248 highly conserved eukaryotic gene families that are present as full or partial genes in the assembly. For most eukaroytes, we would expect to see nearly 100% of CEGMA families represented by a full gene in the genome. Thus, CEGMA provides a measure of the completeness of the assembly for a species.
Comparison of Fig. 2. Using the Compara Perl API to query our custom Ensembl Compara database, TopHat2 (ref. 83). The TopHat output includes the inferred positions of intron default KOGs. The RNAseq data were mapped to the genome assembly using Augustus and SNAP were trained using CEGMA protein evidence gained from the taxonomy-specific of MAKER was performed using the est2genome and protein2genome option with respectively. Gene models obtained from the genome using BLASTN and BLASTX (ref. 81), respectively, and these alignments (ESTs), cDNAs and proteins from related organisms were aligned against the NATURE MICROBIOLOGY ARTICLES O. volvulus different MAKER gene models overlapped in their coding sequence, the gene model encoded proteins of fewer than 30 amino acids were discarded. Third, if two not overlap with any Augustus, genBlastG or RATT gene models were discarded, as was evidence score fi ned with respect to splice sites using Exonerate82. Finally, the EST

Identification of Wolbachia insertions in O. volvulus genome. The O. volvulus Wolbachia 956 kb genome contains 785 predicted protein-coding genes. Comparison of wOv with the published Wolbachia of O. ochengi (wOv) revealed the two genomes to be nearly identical (99.51%) but wOv to be 1,937 bp larger, predominantly due to a small number of large indels. Approximately 4,400 single nucleotide polymorphisms (SNPs) were also identified, of which 57 variants may result in coding differences, including a few that could be involved in the interaction of the symbionts with their hosts. The O. volvulus assembly was searched against the wOv genome with NUMCER v3.06 using MAXMATCH, revealing a total of 531 matches. Of those, 486 were >100 bp and 7 were ≥1 kb. Subsequently, the predicted wOv proteins were used to search for regions with protein homology in the O. volvulus genome using BLASTX as implemented in NCBI BLAST 2.2.22, with results reported in tabular format and an e-value threshold of 1e-15. The regions with matches in the O. volvulus assembly were extracted and searched against protein NT with BLASTN and NR with BLASTX as implemented in NCBI BLAST 2.2.21. Matches not meeting an e-value threshold of 1e-15 were discarded, as well as those without best matches to bacterial gene/protein (which would include mitochondrial and nuclear mitochondrial sequences), yielding a list of putative nuwts in the O. volvulus genome (Supplementary Table 16). These regions were also searched with PRAZE (http://ber.sourceforge.net/) against the predicted wOv proteins to identify frameshifts, nonsense mutations and truncations.

Comparative analysis of the predicted O. volvulus gene set. To help us examine the evolution of gene families and to identify the families of genes lost or expanded in O. volvulus, we ran the Ensembl Comparer pipeline21. Ensembl Comparer is a pipeline that clusters genes into families based on all-against-all blast scores, produces multiple alignments and phylogenies for each gene family, and predicts paralogy and orthology relationships by reconciling reconstructed gene trees with the phylogeny for the taxa included. Along with O. volvulus and O. ochengi, we included published filarial nematode genomes (Dirofilaria immitis, Wuchereria bancrofti, B. malayi and L. loa), Ascariis suum as the only other class III nematode species for which genome data has been published and for which the model free-living stage is included, observe sequence with a single class I parasitic nematode Trichuris muris as an outgroup to all of these. The phylogenetic tree of these species used as input to Comparer was constructed from existing evidence on relationships between filarial species and was identical to topology in that shown to Fig. 2. Using the Comparer Perl API to query our custom Ensembl Comparer database, we identified gene families with multiple expansions or losses in O. volvulus relative to the other comparator species. We removed all the ‘dubious’ duplications (those with a duplication confidence score ≤50) and defined a gene family as expanded if there were at least two gene duplication events reported in O. volvulus. To help us in identifying gene families that are expanded in O. volvulus, we looked at both the gene count and the total protein count of the gene families, which, as the latter statistic is less sensitive to fragmented gene models. A phylogeny of the nine nematode species was generated from 3,148 single-copy Compara gene families that have genes from at least seven species. Sequences for each gene family were aligned using Mafft v7.205 (ref. 90) in automatic (–auto) mode or were manually aligned when trimmings were present. The default parameters, and trimmed alignments were then concatenated to produce a single global data matrix. The phylogeny was then inferred using RAxML v8.0.24 (ref. 82), where each gene family was treated as a single taxon, and the tree was the best-fitting model for that alignment (minimum corrected Akaike information criterion (AIC) from the empirical amino acid substitution models available in that version of RAxML. The preferred phylogeny was estimated with 10 random addition-sequene replicates and support for splits on the tree estimated using 100 bootstrap replicates.

Functional annotation of genes and gene families. Gene ontology (GO) terms were assigned to genes by transferring GO terms from C. elegans orthologues based on the Ensemble Compara approach for transferring GO terms to orthologues in vertebrate species, but modified for improved sensitivity in transferring GO terms across phyla. Manually curated GO annotations were downloaded from the GO Consortium website3 and, for a particular predicted protein in the present study, the manually curated GO terms were obtained for its C. elegans orthologues and then transferred to our predicted protein. GO terms of the three possible types (molecular function, cellular component and biological process) were assigned to predicted proteins in this way. Additional GO terms were identified using InterProScan4. We developed two pipelines to annotate the Compara gene families: one for the assignment of GO terms to genes and another to assign product description to the gene families.

Chromogonics screening. A computational target-based approach was used to screen FDA-approved drugs across all World Health Organization Anatomical Therapeutic Classes (WHO ATC)6 to identify drugs with potential for use as anthelmintics. For protein target identification, we built a Compara database containing all the proteins in release WBPS1 of WormBase ParSite together with new evidence that the 2010-07 Ensemble database contain a genome assembly GRCh37. An initial broad stroke approach identified as a potential drug target any O. volvulus protein belonging to the same gene family in this database as a human protein annotated as a drug target in CHEMBL. A more stringent approach identified O. volvulus proteins as targets only if they were one-to-one orthologues of a human target protein in the same Compara database. Further filtering of the identified proteins was based on molecular weight, number of transmembrane domains, number of disulfide bonds and life-cycle stage. We identified approved drugs with annotated targets in CHEMBL for targets with O. volvulus homologues identified by both the stringent one–one orthologue ‘individual’ protein approach and the more comprehensive approach using Compara families. Both drug sets were filtered on mode of administration (that is, oral, topical and parental). Drugs administered solely via an intravenous route were removed from the set. Further filtering was carried out to restrict the drug set to those with a single protein target, thus excluding drugs targeting protein complexes. We created a drug–protein family interaction network to represent family-level drug identification using the network visualization tool Cytoscape10. The edges were built on the human reference genome assembly GRCh37. An initial broad stroke approach identified as having at least one drug target and those with particularly high numbers to be highlighted as potentially the most ‘drugable’ targets.

Ion channels and other anthelmintic drug targets. Nucleotide sequences for selected ion-channel genes were aligned as codons using MAFFT (ref. 98) and regions with uncertainty caused by high levels of sequence divergence and differing sequence length were removed. Maximum-likelihood phylogenies were inferred using PhyML (v2.102413) with branch node significance determined from Shimodaira–Hasegawa (SH) statistics and from 100 bootstrap replicates.

Metabolic reconstruction and FBA. For each organism, an initial set of EC predictions was obtained from several methods: (1) DETECT v2.0 (cutoff ILS ≥ 0.9), (2) BLASTp (e-value 1e-10 against SWISSPROT115 enzymes), (3) PRIAM100, enzyme rel. Feb-2014 (minimum probability ≥ 0.5, sequence identity ≥70, check catalytic - TRUE), (4) KAAS103, (5) EFICAz104 and (6) EC assignments derived as follows from BRENDA, DETECT and reactions identified by both PRIAM and KAAS. Reaction assignments to metabolic pathways and pathway loci filling were performed using Pathway Tools v18.0 (ref. 106). Those predictions with support from either EFICAz, PRIAM or BLASTp were used to augment the high-confidence set of predictions. In addition, novel Pathway Tool predictions to genes that were previously unannotated were also incorporated. From our draft networks, we completed our metabolic reconstruction, starting with our high-confidence set of enzymes for both O. volvulus and L. loa. In the case of O. volvulus, we also added 100 enzymes (ECs) unique to its Wolbachia endosymbiont. We used KEGG as a guide to find the reactions catalysed by individual enzymes. Reactions that use macromolecules to derive certain metabolites, specifically those reactions that contain the same compound (such as DNA) on both sides of the equation, were considered uninformative and filtered. We further modified some reactions from their original KEGG formulation as they contained a glycan identified in KEGG as equivalent to a compound also present in the reconstruction. Reaction directionalities were defined with reference to KEGG, with reactions marked as either reversible (lower bound ≤ 1,000 and upper bound 1,000 mol (mol h)−1 or not known) or irreversible (lower bound 0, upper bound 1,000 mmol (mol h)−1 and lower bound = 0, upper bound 0 mol (mol h)−1).
depending on directionality). We added a non-growth-associated maintained (NGAM) 10 mmol (g_DW h)^{-1} equation (correcting for the organism's needs outside of growth) to be at least 5 mmol (g_DW h)^{-1}. We allowed for glucose uptake at a maximum of 10 mmol (g_DW h)^{-1} and initially allowed for diffusion of water, oxygen, carbon dioxide, ammonia, diprophosphate, phosphate and ethanol. We also allowed, by default, the transport of all amino acids into the system. Model biomass was defined with reference to a previous metabolite reconstruction of the parasites Toxoplasma gondii^{107} and Leishmania major^{108} (Supplementary Table 13). Pathway gap-filling for both *O. volvulus* and *L. loa* was first performed to ensure production of biomass components and subsequently to complete metabolic pathways missing a limited number of reactions. In this process, we identified from the literature additional metabolites as possibly important for growth and added them to the part of the biomass (biomass components 46–49, Supplementary Table 13)^{109–111}. We also added as part of the biomass such important cofactors as NADH, NADP+ and FAD (biomass components 50–52). In all, we added 68 KEGG reactions for *O. volvulus* and 50 for *L. loa*. In addition, two reactions previously added in gap-filling experiments for *O. volvulus* were found to have some evidence through reciprocal BLAST against *L. loa* (R04230 and R04231 corresponding to EC 6.3.2.5). To enable arachidonic acid metabolism from linoleic acid, two reactions (common to both reconstructions and with gene evidence: R03814_1 (1.14.19.3) and R00390_1 (6.2.1.35)) were modified from their KEGG formulacions, while one reaction (common to both reconstructions: N03681) was custom-made. Further modifications involving changing the directionalities of R01665 (2.7.4.14) and R01664 (3.1.3.5) to ensure CMP was not a dead end; R00086 (ATP phosphohydrolase; linked to many ECs) was modified to allow proton export. For both models, we ensured that the same utility reactions were added so that the models were comparable. Thus, in an attempt to reconstruct consists of 63 NC biotransformation reactions, namely 43 transport reactions, 8 sinks, 4 demand reactions and 8 diffusion reactions (a diffusion reaction was added when allowing for the movement of biceranate). Moreover, both reconstructions contain six irreversible reactions to ensure that specific metabolites (hexanoyl-CoA, octanoyl-CoA, decanoyl-CoA, lauroyl-CoA, tetradecanoyl-CoA and palmitoyl-CoA, with KEGG IDs C05270, C01944, C05274, C01832, C02593 and C00040, respectively) are considered as contenders for the more general compound acyl-CoA (KEGG ID: C00040) involved in such pathways as glycerophospholipid metabolism, for example. Consequently, the reaction catalysed by R03814_1 (1.14.19.3) to ensure CMP was not a dead end; R00086 (ATP phosphohydrolase; linked to many ECs) was modified to allow proton export. For both models, we ensured that the same utility reactions were added (that is, setting the upper and lower bounds as zero) and then measuring the yield through the biomass function. Growth ratios were found by dividing the yield following the block by the yield in an unrestrained model (that is, with the reaction unblocked). Reactions were defined as essential if their growth ratio was predicted to be less than 0.1 upon removal of that reaction. Further details of the metabolic reconstruction are provided in Supplementary Table 13.

In these analyses no attempt was made to partition reactions due to the challenge of accurately assigning reactions to distinct intracellular compartments (for example, mitochondrion, golgi and wOv), as well as determining the specific small-molecule transporter activities that allow metabolites to move between compartments. Instead, metabolites were assumed to be freely available across compartments, an assumption, as previously mentioned, that is expected to minimize false-positive essential reaction predictions at the expense of increasing the likelihood false-negative predictions.

Data availability. Sequence data have been deposited in the European Nucleotide Archive (ENA). Assemblies and annotation are available at WormBase and WormBase-ParaSite (http://parasite.wormbase.org/Onchocerca_volvulus_prjeb5131/index.html and http://parasite.wormbase.org/Onchocerca_ochengi_prjeb1204/index.html). All have been submitted to GenBank under BioProjects PRIEB513 (O. volvulus) and PRIEB1204 (O. ochengi). RNASeq data are available under BioProject PRIEB2965 (O. volvulus). All accession numbers are available in Supplementary Table 19. The wOv version described here differs slightly from the GenBank version and is available from ftp://ftp.sanger.ac.uk/pub/project/pathogens/Onchocerca_volvulus/.

The Supplementary Tables provide information that supports the data presented. Supplementary Data are also provided for all the Ensemble Compare analyses.

Received 6 April 2016; accepted 26 September 2016; published 21 November 2016; corrected 14 July 2017
31. Li, B. W., Rush, A. C. & Weil, G. J. High level expression of a glutamate-gated chloride channel in recombinant tissues of Brugia malayi may explain the stabilizing effect of ivermectin on filarial worms. Int. J. Parasitol. Drugs Drug Res. 4, 71–76 (2014).

32. Ardelli, B. F. Transport proteins of the ABC systems superfamily and their role in drug action and resistance in nematodes. Parasitol. Int. 62, 639–646 (2013).

33. Kotze, A. C. et al. Recent advances in candidate-gene and whole-genome approaches to the discovery of anthelmintic resistance markers and the description of drug/receptor interactions. Int. J. Parasitol. Drugs Drug Res. 4, 164–184 (2014).

34. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? Nat. Biotechnol. 28, 245–248 (2010).

35. Varma, A. & Palsson, B. O. Stoichiometric flux balance models quantitatively predict metabolic by-product secretion in wild-type Escherichia coli W3110. Appl. Environ. Microbiol. 60, 3724–3731 (1994).

36. Chen, S. N. & Howells, R. E. Brugia pahangi: uptake and incorporation of adenosine and thymidine. Exp. Parasitol. 47, 209–221 (1979).

37. Chen, S. N. & Howells, R. E. Brugia pahangi: uptake and incorporation of nucleic acid precursors by microfilarete and macrofilarete in vitro. Exp. Parasitol. 57, 405–412 (1986).

38. Reijman, D. et al. N-phosphonocarboxylpolyribonucleic derivatives of guanine: a new class of bi-substrate inhibitors of human purine nucleoside phosphorylases. J. Med. Chem. 55, 1612–1621 (2012).

39. Taylor, M. J., Hoerauf, A., Townsend, S., Slak, B. E. & Ward, S. A. Anti-Brugia discovery and development: safe macrofilaricides for onchocercal filariasis. Parasitology 141, 119–127 (2014).

40. Boulin, T. et al. Eight genes are required for functional reconstitution of the Caenorhabditis elegans levamisole-sensitive acetylcholine receptor. Proc. Natl Acad. Sci. USA 105, 18590 (2008).

41. Courtot, E. et al. Functional characterization of a novel class of morantel-sensitive acetylcholine receptors in nematodes. PLoS Pathogens 11, e1005267 (2015).

42. Mostafa, E. et al. Transient effects of levamisole on Brugia malayi microfilariae. Invert. Neurosci. 15, 5 (2015).

43. Gregory, W. F. & Mazels, R. M. Cystatins from nematodes: evolution, sensitive acetylcholine receptors in nematodes. J. Med. Chem. 296, 28, 9 (2007).

44. Molehin, A. J., Gobert, G. N. & McManus, D. P. Serine protease inhibitors of onchocerciasis: history of progress and current feasibility using ivermectin monotherapy. J. Infect. Dis. 256, 1047–1048 (2015).

45. Gebauer, J. & Kozarewa, I. A genome-scale database and reconstruction of the Onchocerca volvulus genome: comparison of the nucleotide sequences between the two species. PloS Pathogens 2, e47 (2006).

46. Kurtz, S. et al. Entwistle and open software for comparing large genomes. Genome Biol. 5, R12 (2004).

47. Foster, J. et al. The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode. PLoS Biol. 3, e121 (2005).

48. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12, 491 (2011).

49. Smith, C. D. et al. Improved repeat identification and masking in Dipterans. Gene 389, 1–9 (2007).

50. Kohany, O., Gentles, A. J., Hankus, L. & Jurka, J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 7, 47 (2006).

51. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34, W345–W349 (2006).

52. Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O. & Borodovsky, M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 18, 1979–1990 (2008).

53. Korf, I. Gene finding in novel genomes. BMC Bioinformatics 5, 59 (2004).

54. She, R. et al. Gemblast: using BLAST searches to build homologous gene models. Bioinformatics 21, 214–215 (2011).

55. Yook, K. et al. Wormbase 2012: more genomes, more data, new website. Nucleic Acids Res. 40, D735–D741 (2012).

56. Otto, T. D., Dillon, G. P., Degrave, W. S. & Berryman, M. RATT: Rapid Annotation Transfer Tool. Nucleic Acids Res. 39, e57 (2011).

57. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402 (1997).

58. Slater, G. S. & Birney, E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6, 31 (2005).

59. Kim, D. et al. TopHat2: accurate alignment of transcripts in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013).

60. Nakamura, Y., Cochrane, G. & Karsch-Mizrachi, I. The International Nucleotide Sequence Database Collaboration. Nucleic Acids Res. 41, D21–D24 (2013).

61. Lizotte-Waniek, M. et al. Identification of potential vaccine and drug target candidates by expressed sequence tag analysis and immunoscreening of Onchocerca volvulus larval cDNA libraries. Infect. Immun. 68, 3499–3501 (2000).

62. Williams, S. A., Laney, S. J., Lizotte-Waniek, M., Bierwerta, L. A. & Unnasch, T. R. The river blindness genome project. Trends Parasitol. 18, 86–90 (2002).

63. UniProt, Consortium. Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Res. 41, D43–D47 (2013).

64. Logan-Klumper, F. J. et al. GeneDB—an annotation database for pathogens. Nucleic Acids Res. 40, D98–108 (2012).

65. Darby, A. C. et al. Analysis of gene expression from the Wolbachia genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. Genome Res. 22, 2467–2477 (2012).

66. Kato, H. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780 (2013).

67. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552 (2000).
92. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313 (2014).
93. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res. 43, D1049–D1056 (2015).
94. Jones, P. T. et al. Interproscan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240 (2014).
95. US FDA. Approved Drug Products with Therapeutic Equivalence Evaluations 33rd edn (US Department of Health and Human Services, 2013).
96. WHO Collaborating Centre for Drug Statistics Methodology. Guidelines for ATC Classification and DDD Assignment 19th edn (World Health Organization, 2015).
97. Shannon, P. et al. Cytoscape: a software environment for integrated models of biological networks. Genome Res. 13, 2498–2504 (2003).
98. Katoh, K., Misawa, K., Kuma, K. & Miyata. T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066 (2002).
99. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704 (2003).
100. Hung, S. S., Wasmuth, J., Sanford, C. & Parkinson, J. DETECT—a density estimation tool for enzyme classification and its application to Plasmidium falciparum. Bioinformatics 26, 1690–1698 (2010).
101. UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res. 43, D204–D212 (2015).
102. Claudel-Renard, C., Chevalet, C., Faraut, T. & Kahn, D. Enzyme-specific profiles for genome annotation: PRIAM. Nucleic Acids Res. 31, 6633–6639 (2003).
103. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182–W185 (2007).
104. Tian, W., Arakaki, A. K. & Skolnick, J. EFFiCA: a comprehensive approach for accurate genome-scale enzyme function inference. Nucleic Acids Res. 32, 6226–6239 (2004).
105. Schomburg, I. et al. BREND A: a resource for enzyme data and metabolic information. Trends Biochem. Sci. 27, 54–56 (2002).
106. Karp, P. D. et al. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. Brief. Bioinform. 11, 40–79 (2010).
107. Song, C. et al. Metabolic reconstruction identifies strain-specific regulation of virulence in Toxoplasma gondii. Mol. Syst. Biol. 9, 708 (2013).
108. Chavali, A. K., Whittomore, J. D., Eddy, J. A., Williams, K. T. & Papin, J. A. Systems analysis of metabolism in the pathogenic trypanosomatid Leishmania major. Mol. Syst. Biol. 4, 177 (2008).
109. Foster, J. M., Zhang, Y., Kumar, S. & Carlow, C. K. Parasitic nematodes have two distinct chitin synthases. Mol. Biochem. Parasitol. 142, 126–132 (2005).
110. Rojo-Arroola, L. et al. Chemical and genetic validation of the statin drug target to treat the helminth disease, schistosomiasis. PLoS ONE 9, e87594 (2014).
111. Wittich, R. M. & Walter, R. D. Putrescine N-acetyltransferase in the wolbachia pathogen informatics for informatics support and for providing the pipeline to annotate the Wolbachia genome, T. Truico of the New York Blood Center (NYBC) for the generation of O. volvulus D40 juvenile female and male worms, and J. Liu and A. Contreras from the NYBC for preparing all the DNA and RNA samples used in these studies. C.J.D.H. was supported by NIAID (U19 AI110820) as well as an NIH Director’s New Innovator Award (DP2-OD007372), J.P., N.N. and L.S.S. were supported by the Canadian Institutes of Health Research (MOP# 84556) and the Natural Sciences and Engineering Research Council of Canada (RGPIN 2014-06664). Additional computing resources were provided through Compute Canada by the University of Toronto SciNet HPC Consortium. S.L. was supported in part by intramural funding from the NYBC, The Edna McConnell Clark Foundation and NIH/NIAID (R01 AI42382). The authors thank the people of the villages around Kumba, Marumba I, Marumba II, Boa Bakundu, Bombanda and Bombele, Cameroon, who were part of the field studies that enabled the collection of O. volvulus parasite material. The authors also thank P. Enyong and other personnel at the Tropical Medicine Research Station, Kumba, for their assistance in collecting the O. volvulus and O. ochengi parasite material.

Author contributions
All authors read and approved the manuscript. N.H., T.K., J.T., A.T., O.L., E.S., M.B.R., M.D., I.T.T. and S.B. carried out sample preparation, sequencing, genome assembly and/or gene finding and annotation. S.R.D., J.C.D.H. and D.V. analysed Wolbachia sequence data. J.P., N.N. and L.S.S. performed metabolic reconstructions and constraints-based analyses. P.M. performed the chemogenomics screening. R.B., B.H., A.G., A.M., T.R.U., S.B. and J.M.C.R. performed data analyses. J.A.C., T.R.N., E.G., M.B. and S.L. analysed data and wrote the manuscript.

Additional information
Supplementary information is available for this paper.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to E.G., M.B. and S.L.
How to cite this article: Cotton, J. A. et al. The genome of Onchocerca volvulus, agent of river blindness. Nat. Microbiol. 2, 16216 (2016).

Competing interests
The authors declare no competing financial interests.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/