Title

Tryptophan Operon Diversity Reveals Evolutionary Trends among Geographically Disparate Chlamydia trachomatis Ocular and Urogenital Strains Affecting Tryptophan Repressor and Synthase Function

Journal

mBio, 12(3)

ISSN

2161-2129

Authors

Bommana, Sankhya
Somboonna, Naraporn
Richards, Gracie
et al.

Publication Date

2021-06-29

DOI

10.1128/mbio.00605-21

Peer reviewed

AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE

Tryptophan Operon Diversity Reveals Evolutionary Trends among Geographically Disparate *Chlamydia trachomatis* Ocular and Urogenital Strains Affecting Tryptophan Repressor and Synthase Function

Sankhya Bommana,[a] Naraporn Somboonna,[b] Gracie Richards,[a] Maryam Tarazkar,[a] Deborah Dean[a,b,c,d]

[a]Department of Pediatrics, University of California San Francisco, Oakland, California, USA
[b]Department of Bioengineering, Joint Graduate Program, University of California San Francisco and University of California Berkeley, San Francisco, California, USA
[c]Department of Medicine, University of California San Francisco, San Francisco, California, USA
[d]Bixby Center for Global Reproductive Health, University of California San Francisco, San Francisco, California, USA

ABSTRACT The obligate intracellular pathogen *Chlamydia trachomatis* (*Ct*) is the leading cause of bacterial sexually transmitted infections and blindness globally. To date, *Ct* urogenital strains are considered tryptophan prototrophs, utilizing indole for tryptophan synthesis within a closed-conformation tetramer comprised of two α (TrpA)- and two β (TrpB)-subunits. In contrast, ocular strains are auxotrophs due to mutations in TrpA, relying on host tryptophan pools for survival. It has been speculated that there is strong selective pressure for urogenital strains to maintain a functional operon. Here, we performed genetic, phylogenetic, and novel functional modeling analyses of 595 geographically diverse *Ct* ocular, urethral, vaginal, and rectal strains with complete operon sequences. We found that ocular and urogenital, but not lymphogranuloma venereum, TrpA-coding sequences were under positive selection. However, vaginal and urethral strains exhibited greater nucleotide diversity and a higher ratio of nonsynonymous to synonymous substitutions [Pi(a)/Pi(s)] than ocular strains, suggesting a more rapid evolution of beneficial mutations. We also identified nonsynonymous amino acid changes for an ocular isolate with a urogenital backbone in the intergenic region between TrpR and TrpB at the exact binding site for YtgR—the only known iron-dependent transcription factor in *Chlamydia*—indicating that selective pressure has disabled the response to fluctuating iron levels. *In silico* effects on protein stability, ligand-binding affinity, and tryptophan repressor (TrpR) affinity for single-stranded DNA (ssDNA) measured by calculating free energy changes ($\Delta\Delta G$) between *Ct* reference and mutant tryptophan operon proteins were also analyzed. We found that tryptophan synthase function was likely suboptimal compared to other bacterial tryptophan prototrophs and that a diversity of urogenital strain mutations rendered the synthase nonfunctional or inefficient. The novel mutations identified here affected active sites in an orthosteric manner but also hindered α- and β-subunit allosteric interactions from distant sites, reducing efficiency of the tryptophan synthase. Importantly, strains with mutant proteins were inclined toward energy conservation by exhibiting an altered affinity for their respective ligands compared to reference strains, indicating greater fitness. This is not surprising as L-tryptophan is one of the most energetically costly amino acids to synthesize. Mutations in the tryptophan repressor gene (*trp*R) among urogenital strains were similarly detrimental to function. Our findings indicate that urogenital strains are evolving more rapidly than previously recognized with mutations that impact tryptophan operon function in a manner that is energetically beneficial, providing a novel host-pathogen evolutionary mechanism for intracellular survival.

Citation Bommana S, Somboonna N, Richards G, Tarazkar M, Dean D. 2021. Tryptophan operon diversity reveals evolutionary trends among geographically disparate *Chlamydia trachomatis* ocular and urogenital strains affecting tryptophan repressor and synthase function. mBio 12:e00605-21. https://doi.org/10.1128/mBio.00605-21.

Editor Nina R. Salama, Fred Hutchinson Cancer Research Center

Address correspondence to Deborah Dean, deborah.dean@ucsf.edu.

This article is a direct contribution from Deborah Dean, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Ashok Aiyar, Louisiana State University Health Sciences Center, and Wilhelmina Huston, University of Technology Sydney.

Received 11 March 2021
Accepted 19 March 2021
Published 11 May 2021

IMPORTANCE *Chlamydia trachomatis* (*Ct*) is a major global public health concern causing sexually transmitted and ocular infections affecting over 130 million and 260 million people, respectively. Sequelae include infertility, preterm birth, ectopic pregnancy, and blindness. *Ct* relies on available host tryptophan pools and/or substrates to synthesize tryptophan to survive. Urogenital strains synthesize tryptophan from indole using their intact tryptophan synthase (TS). Ocular strains contain a *trp*A frameshift mutation that encodes a truncated TrpA with loss of TS function. We found that TS function is likely suboptimal compared to other tryptophan prototrophs and that urogenital stains contain diverse mutations that render TS nonfunctional/inefficient, evolve more rapidly than previously recognized, and impact operon function in a manner that is energetically beneficial, providing an alternative host-pathogen evolutionary mechanism for intracellular survival. Our research has broad scientific appeal since our approach can be applied to other bacteria that may explain evolution/survival in host-pathogen interactions.

KEYWORDS *Chlamydia trachomatis*, evolution, phylogeny, protein modeling and docking, sexually transmitted infections, tryptophan biosynthesis, tryptophan operon

The obligate intracellular pathogen *Chlamydia trachomatis* (*Ct*) is the leading cause of bacterial sexually transmitted infections (STIs) and preventable blindness worldwide (1). Over 130 million cases of *Ct* STIs occur annually and can lead to severe reproductive complications such as chronic pelvic pain, tubal factor infertility, ectopic pregnancy, and preterm birth (2). *Ct* is also responsible for lymphogranuloma venereum (LGV), an invasive disease that can cause hemorrhagic proctitis and spread via regional lymphatics to the inguinal lymph nodes, forming buboes with subsequent suppuration (2). LGV is endemic in global tropical and subtropical regions as well as worldwide among men who have sex with men (3, 4). An estimated 232 million people are at risk of blindness from trachoma, a chronic ocular disease that occurs primarily in endemic tropical developing countries (5). Repeat and/or persistent infection can induce an immunopathogenic response that can lead to progression from conjunctival inflammation to tissue fibrosis, scarring, and blindness (6).

Molecular typing of *Ct ompA*, referred to as *ompA* genotyping, has been used to understand the tissue tropism and molecular epidemiology of *Ct* infections (7). In general, ocular *ompA* genotypes A, B, Ba, and C cause trachoma; urogenital genotypes D, Da, E, F, G, H, I, Ia, J, Ja, and K are linked to urogenital infections while only some of these genotypes cause rectal infections; and genotypes L1, L2, L2a, L2b, L2c, and L3 are responsible for LGV (6). The non-LGV strains are referred to as the trachoma noninvasive biological variant (biovar) while LGV strains are considered the invasive biovar (6).

With the advent of multilocus sequence typing (MLST) and partial and whole-genome sequencing (WGS), there has been a greater understanding of strain diversity, strain emergence from recombination (e.g., Da/ocular, B/urogenital, L2/D, Ja/E recombinant strains), and tissue tropism (8–17). We now know that there are *ompA* genotypes B, Ba, and C that are actually urogenital strains because they have a urogenital strain backbone (18, 28). Genes involved in apparent tissue tropism include *ompA*, *tarP*, *incA*, toxin-like genes, and *trpA*, among others (19).

Ct relies heavily on host nutrient supplies and biochemical processes to sufficiently proliferate within the host cell (20). Although urogenital strains are tryptophan prototrophs, auxotrophic ocular strains depend on the availability of host tryptophan pools for survival. Tryptophan synthesis is energetically costly, and, therefore, *Ct* ocular strains and other intracellular bacteria have evolved to scavenge this essential metabolite from their hosts (21). Urogenital strains are able to synthesize tryptophan from substrates such as indole, when tryptophan is scarce, using their intact functional tryptophan synthase (TS). Tryptophan scarcity occurs when there is competition for this amino acid (aa) from other microbes in the respective microbiome and as a result of the tryptophan catabolism by indoleamine 2,3-dioxygenase 1 (IDO1). IDO1 is induced by interferon gamma (IFN-γ), an

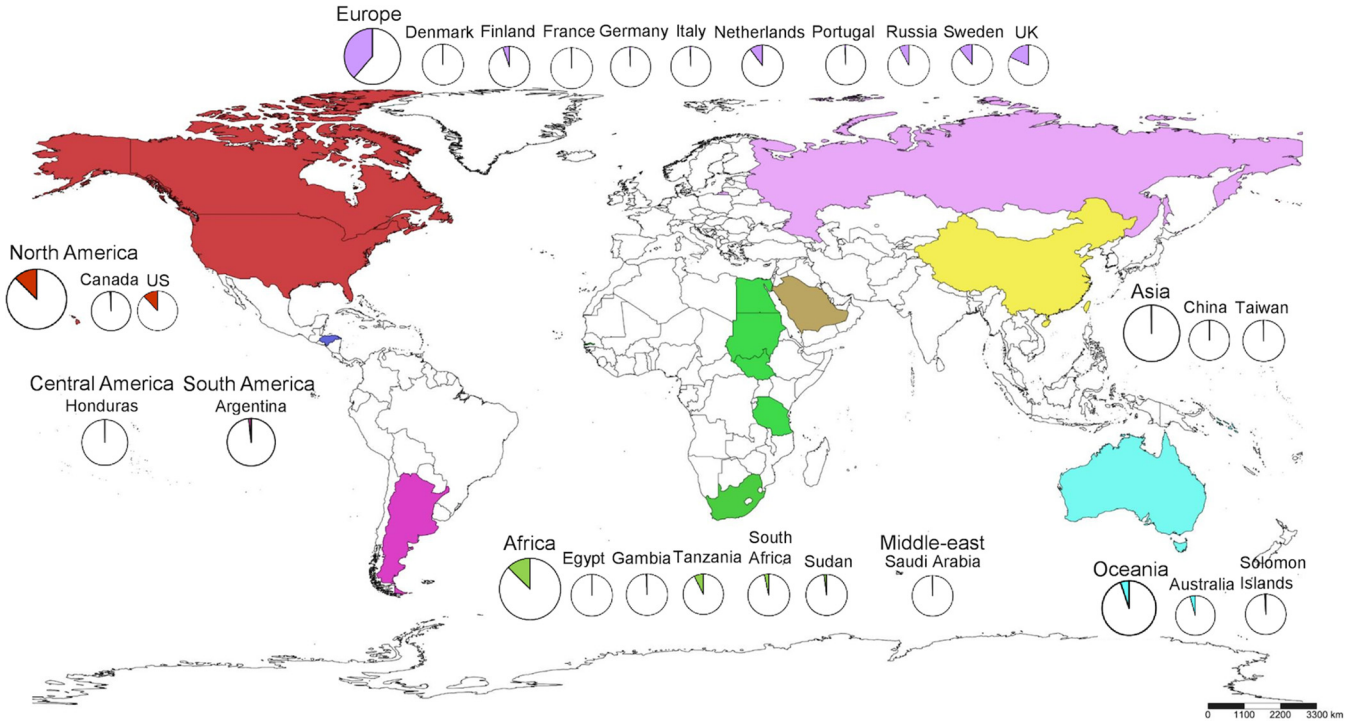


FIG 1 Global distribution of *Chlamydia trachomatis* strains harboring the tryptophan operon used in this study. Colors indicate the geographic areas or continents that were the sources of the samples as shown in the key. Circles within countries show the number of samples from that country. Countries neither sampled nor included in this study are shown in white. The scale bar underneath the map provides a visual indication of the size of features, and distance between features, on the map. Map was generated by SimpleMappr.

innate immune cytokine that is stimulated in response to *Ct* infection (22). An intact operon is thought, therefore, to enable *Ct* to escape immune surveillance and survive.

The operon consists of the tryptophan repressor, TrpR, and two β (TrpB)- and two α (TrpA)-subunits that form an $\alpha\beta\beta\alpha$ tetrad with an intramolecular tunnel. In the tunnel, indole is bound and converted to L-tryptophan (23, 24). Ocular strains, in contrast to urogenital strains, have lost this function due to a single nucleotide deletion and frameshift in *trp*A that encodes a truncated TrpA (25, 26). Initial studies revealed high homology among urogenital and ocular strains for the respective operon sequence (25). More recently, we discovered that long-term persistence of a urogenital clinical F strain occurred due to a novel mutation in *trp*A resulting in elongation of TrpA by 2 aa with an altered $\alpha\beta\beta\alpha$ structure and markedly decreased TS with lower uptake of tryptophan for metabolism compared to the reference strain (26). While it is thought that there is strong selective pressure for urogenital strains as opposed to ocular strains to maintain a functional operon (25), our recent data indicate that this may not be entirely true.

Current studies of tryptophan operons are based on *Ct* strains from Tanzania, Gambia, the United States, and Canada, often with a focus on TrpA and TrpB. We sought to expand this work by examining the entire tryptophan operon from 595 genomes representing ocular, urethral, vaginal, rectal, and LGV strains from 24 countries and all continents of the world except for Antarctica (Fig. 1). Here, we performed genetic, phylogenetic, and novel functional protein modeling analyses on these geographically diverse *Ct* strains to elucidate the evolution of the operon and its changing role in tissue tropism and disease pathogenesis.

RESULTS

Tryptophan operon phylogeny revealed distinct LGV and trachoma linages with prevalent, nonprevalent, and mixed urogenital clades with greater diversity for *trp*A. Tryptophan operon sequences consisting of *trp*R, intergenic region (IGR), *trp*B, and *trp*A, in that order, were available for 595 clinical and reference samples; 12 ocular

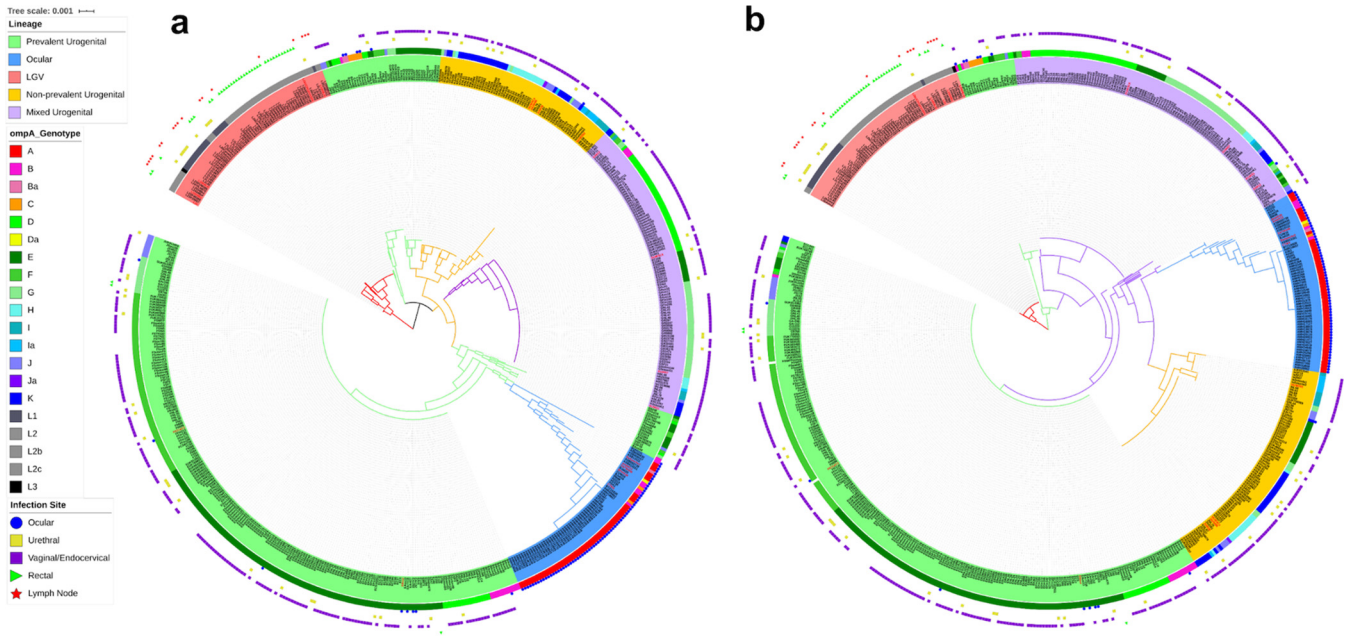


FIG 2 Interactive phylogenetic trees for *trp*RBA (a) (see https://itol.embl.de/tree/209360135466881582240570) and *trp*A (b) (see https://itol.embl.de/tree/761676910220385158992715 8) depicting *Ct* lineages: LGV, nonprevalent and prevalent urogenital, mixed urogenital, and a more recent ocular subclade. The relationship between these lineages with their *omp*A genotype and site or infection are represented as inner and outer ring color strips and as a binary data set. The tree was constructed by FastTree with a generalized time-reversible model based on a MAFFT alignment of concatenated *trp*RBA and *trp*A sequences for each of the 595 strains in the data set (see Materials and Methods). Both were executed in Geneious. *omp*A and site-of-infection data of the strains together with the phylogenetic tree were used to generate this figure in iTOL.

strains from Sudan (27) were excluded due to read quality (Fig. 1; see also Data Set S1 in the supplemental material). Phylogenetics revealed two distinct lineages representing the LGV and trachoma biovars. The latter branched into four distinct clades of prevalent, nonprevalent, and mixed urogenital strains with the ocular strains forming a more recent branch from an ancestral prevalent urogenital subclade (Fig. 2a, inner circle; https://itol.embl.de/tree/761676910220385158992715 8). In general, the prevalent urogenital strains consisted of D, Da, E, and F with some J and Ja strains; nonprevalent strains included all others (11).

The first branch of prevalent strains contained ocular *omp*A genotypes with urogenital backbones isolated from the eyes of Aboriginal children (18) and urogenital *omp*A genotypes with urogenital backgrounds isolated from the eyes of patients in Denmark and Argentina (Fig. 2a; blue dots at ~11 o'clock; Table S1).

*trp*A phylogeny similarly contained both biovar lineages, although the earliest ancestor after LGV was the ocular China B_QH111L strain with a D/G urogenital backbone (28) (Fig. 2b, inner circle; https://itol.embl.de/tree/761676910220385158992715 8). There was greater genetic diversity among nonprevalent urogenital and ocular branches compared to prevalent urogenital strains. Ocular A strains from Tanzania were a relatively recent clonal population compared to all other ocular strains with B_TZ1A828 as the ancestral strain (Fig. 2b). Overall, *trp*A phylogeny was somewhat congruent with operon phylogeny.

Tryptophan operon analyses revealed no recombination but hot spots of genetic variation and evidence of signatures for LGV, ocular, and urogenital tissue tropism. Using the MAFFT alignment and a series of programs run in RDP4, there was no evidence for recombination involving the operon (Tables S1 and S2).

Single nucleotide polymorphism (SNP) analysis of individual tryptophan operon coding sequences (CDS) was performed by segregating strains based on phylogenetic clustering with the reference strain sequences (Fig. S1) in addition to considering *omp*A genotype and anatomic origin.

*trp*R is considered highly conserved with only two common nucleotide (nt) positions of nonsynonymous substitutions confined to D, F, and J stains (Fig. 3a; blue

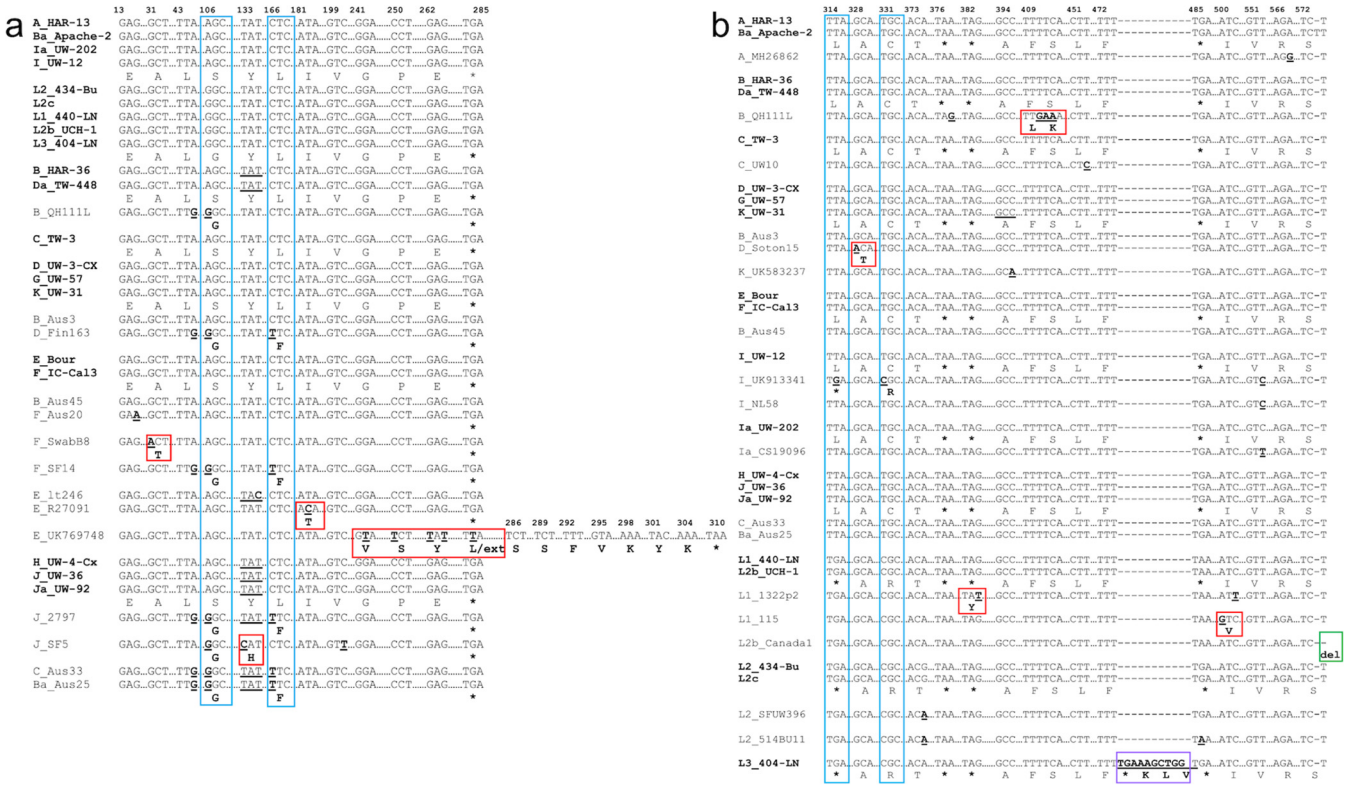


FIG 3 Partial nucleotide sequences of *trp*R (a) and IGR (b) with regions of single nucleotide polymorphism (SNPs) and indels in comparison to the 21 *Ct* reference strain sequences (in bold). Australian aboriginal strains Ba_Aus25, B_Aus3, B_Aus45, and C_Aus33 were included with the urogenital strains due to their "true" genome sequence identity as urogenital strains. Homologous regions in the sequence are not shown and are denoted as ".…." Dashes denote nucleotide deletion(s). Nucleotides in bold denote substitutions, and bold aa letters denote nonsynonymous aa substitutions compared to the reference strain. Nucleotide and/or aa changes that are unique to a single strain are boxed in red; changes common across several sequences/variants are boxed in blue. "ins," insertion (boxed in purple); "fs," frameshift; "ext," extension; "del," deletion (boxed in green).

boxes). Only reference strains and those with mutations are shown. All other strains were similar to reference or clinical strains. Unique nonsynonymous substitutions are also noted (Fig. 3a; red boxes). Strain E_UK769748 had four unique nonsynonymous substitutions that extended the protein by 7 aa (Fig. 3a; red box). The IGR had two common SNPs among a number of strains (Fig. 3b; blue boxes). L1_1322p2, L2b_Canada1, and L3_404-LN had indels. Mutations in B_QH111L occurred within the YtgR-binding site for YtgR, an iron-dependent transcription factor that binds to the IGR and regulates operon expression (29), which is conserved among all other strains (Fig. 3b).

In *trp*B, there were five common nt positions with nonsynonymous substitutions across urogenital strains (Fig. 4a; blue boxes) while six urogenital strains had unique nonsynonymous substitutions (Fig. 4a; red boxes) as did three ocular strains (Fig. 5a, green and purple boxes; Table S3). Two nonsynonymous mutations were common across the ocular strains (Fig. 5a, blue boxes; Table S3). The indel in B_QH111L resulted in a frameshift and early truncation at nt position 624. There were no SNPs for LGV strains.

*trp*A ocular and urogenital strains exhibited the highest overall nt diversity (Table S4). Seven urogenital and four ocular strains had nt positions with nonsynonymous substitutions across a number of strains (Fig. 4b and Fig. 5b, blue boxes; Table S4). Unique nonsynonymous substitutions were noted for 10 of these strains (Fig. 4b and Fig. 5b, red, green, and purple boxes; Table S4). We previously noted that F_SF11 (i.e., F_I-IV) had a nonsynonymous substitution and 2-aa extension exhibiting decreased operon function *in vitro* (26). Most ocular strains had a deletion at nt position 531 with frameshift and early truncation of the protein at 184 aa as previously

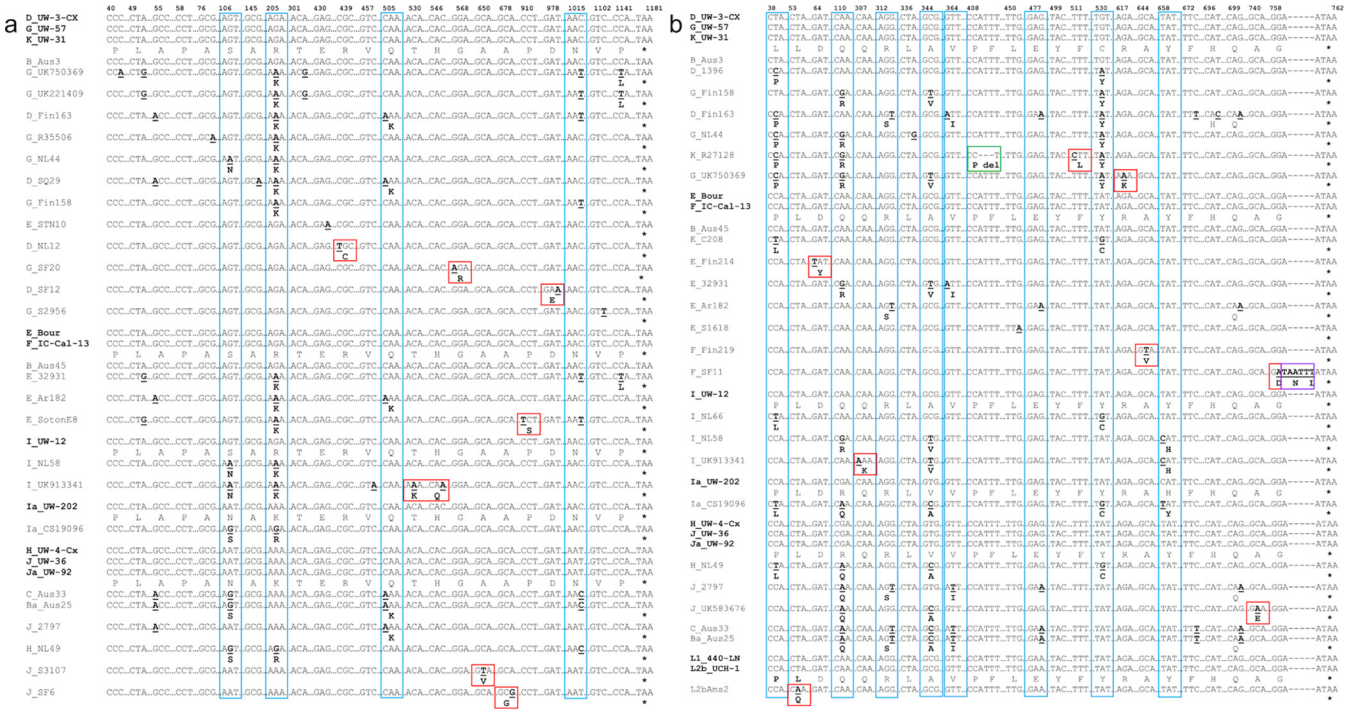


FIG 4 Partial nucleotide sequences of *trp*B (a) and *trp*A (b) urogenital strains with regions of single nucleotide polymorphism (SNPs) and indels in comparison to the 21 *Ct* reference strain sequences (in bold). Australian aboriginal strains Ba_Aus25, B_Aus3, B_Aus45, and C_Aus33 were included with the urogenital strains due to their "true" genome sequence identity as urogenital strains. Homologous regions in the sequence are not shown and are denoted as "…." Dashes denote nucleotide deletion(s). Nucleotides in bold denote substitutions, and bold aa letters denote nonsynonymous aa substitutions compared to the reference strain. Nucleotide and/or aa changes that are unique to a single strain are boxed in red; changes common across several sequences/variants are boxed in blue. "ins," insertion (boxed in purple); "fs," frameshift; "ext," extension; "del," deletion (boxed in green).

described (25). Interestingly, 410_412delATT was also commonly observed (Fig. 5b; Table S4). Four other strains also had indels. B_TZ1A828 was the only ocular strain with an intact TrpA-CDS of 762 nt (Fig. 5b; Table S4).

Ocular and urogenital but not LGV strains are under positive selection. Details on polymorphic sites, haplotype diversity, and Pi(a)/Pi(s) ratios for TrpR-, TrpB-, and TrpA-CDS are shown in Table 1. TrpA and TrpB ocular strains were under positive selection, as was TrpA for the urogenital strains (Table 1). While both vaginal and urethral TrpA-CDS were under positive selection, the latter exhibited greater nt diversity and a higher Pi(a)/Pi(s) ratio, although both had higher diversity and ratios compared to ocular strains (Table 1). TrpR- and TrpB-CDS appeared to be under negative selection for the urogenital strains. LGV strains had 0 to 1 segregation sites (Ss), and therefore, Pi(a)/Pi(s) values could not be calculated.

TrpR mutant strains reveal structural and functional changes in the DNA-binding motif and ligand-binding sites of the symmetric dimer. *Ct* TrpR forms a symmetric dimer of two chains each consisting of 94 aa and four identical ligand-binding sites (LBSs). Each subunit contains six α-helices (i.e., A to F; D and E; and DE turn, termed DNA-binding motif [DBS]) that undergo conformational change upon L-tryptophan binding to the LBSs, which is necessary for binding to the operator DNA (30). Three biologically fundamental interactions, protein-protein (dimerization), protein-ligand (corepressor binding), and protein-DNA (operator binding), are noted in Fig. S2.

Since there are no crystal structures of *Ct* TrpR, the protein was modeled against the best-hit homology to known crystal structure templates. Three TrpR urogenital strain mutants (i.e., E_R27091, F_SwabB8, and J_SF5) and their respective reference strains (i.e., E_Bour, F_ICCal3, and J_UW36 [Fig. 6 and Fig. S3 and S4; magenta]) had >95% correct folds and a GA341 MODELLER score of 1.0 to *Escherichia coli* variant T44L S88Y 6eni, indicating a highly reliable model (31, 32). Indole acetic acid (IAC) was found to be the bound ligand. The highest identity to these strains was 33% with an

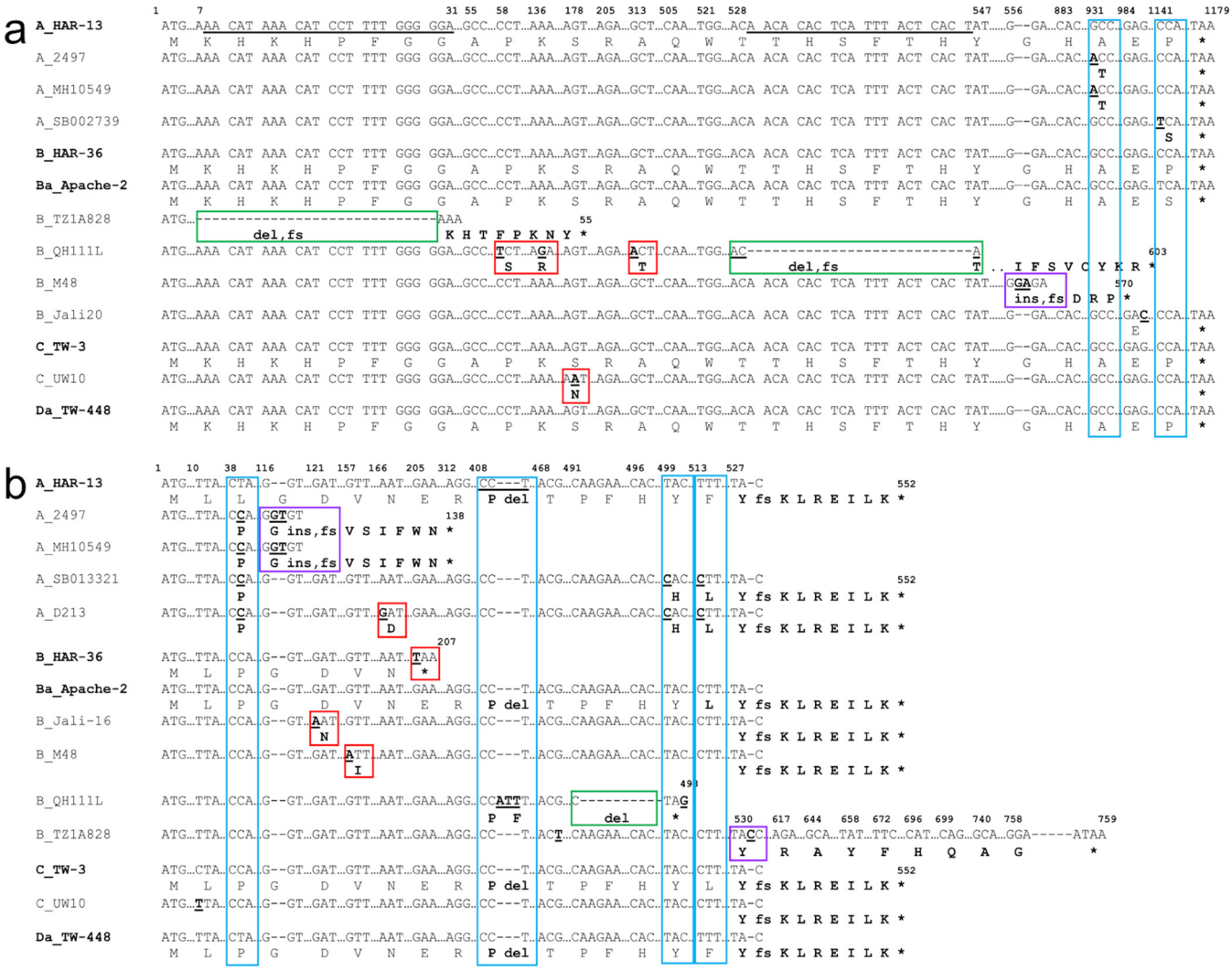


FIG 5 Partial nucleotide sequences of *trp*B (a) and *trp*A (b) ocular strains with regions of single nucleotide polymorphism (SNPs) and indels in comparison to the 21 *Ct* reference strain sequences (in bold). Australian aboriginal strains Ba_Aus25, B_Aus3, B_Aus45, and C_Aus33 were included with the urogenital strains due to their "true" genome sequence identity as urogenital strains. Homologous regions in the sequence are not shown and are denoted as ". . .." Dashes denote nucleotide deletion(s). Nucleotides in bold denote substitutions, and bold aa letters denote nonsynonymous aa substitutions compared to the reference strain. Nucleotide and/or aa changes that are unique to a single strain are boxed in red; changes common across several sequences/variants are boxed in blue. "ins," insertion (boxed in purple); "fs," frameshift; "ext," extension; "del," deletion (boxed in green).

overall root mean square deviation (RMSD) of 2.5 Å for mutant and 4.065 Å for reference strains. E_UK769748 had a GA341 MODELLER score of 0.99 with highest homology to *E. coli* 6fal TrpR with an identity of 33%. TrpR *Ct* strain alignments with the templates 6eni and 6fal are shown in Fig. 7a along with the LBSs and DBS residues. Phylogenetic reconstruction of *Ct* strains and the templates is shown in Fig. 7b.

E_R27091, J_SF5, F_SwabB8, and reference strains with template models are shown in Fig. 5 and Fig. S3 and S4. The E_R27091 I61T mutation was located in the DBM (Fig. 5a and b; charcoal gray), F_SwabB8 A11T was in helix A (Fig. S3; charcoal gray, Fig. 7a), and J_SF5 Y45H was proximal to LBS2 (Fig. S4; charcoal gray, Fig. 7a), while S36G (Fig. S4; charcoal gray) was located further upstream. E_UK769748 could not be modeled due to the aa elongation.

The effect of mutations on protein stability, interactions with single-stranded DNA (ssDNA) (operator-DNA), and protein-ligand affinity was predicted using mCSM and mCSM-lig, respectively (Table 2). There was a predicted decrease in affinity of LBS for IAC, except for E_R27091, but an increased affinity for ssDNA, except for J_SF5 S36G, which would affect its ability to prevent transcription initiation of *trp*BA. These

TABLE 1 Nucleotide diversity and Pi(a)/Pi(s) ratio for *C. trachomatis* TrpR-, TrpB-, and TrpA-coding sequences (CDS) for ocular, urogenital/rectal, and LGV strains based on phylogenetic lineages[a]

CDS	No. of haplotypes	Haplotype diversity (Hd)	*Ct* lineage	No. of sequences	No. of polymorphic segregating sites (Ss)	Synonymous nucleotide diversity (Pi(s); Jukes-Cantor corrected)	Nonsynonymous nucleotide diversity (Pi(a); Jukes-Cantor corrected)	Pi(a)/Pi(s) ratio	No. of synonymous substitutions	No. of nonsynonymous substitutions
*trp*R	10	0.2889	Ocular	79	1	0	0.00011	0	0	1
			Urogenital/rectal	455	14	0.00179	0.00101	0.564	4	9
			LGV	70	0	0	0	0	0	0
*trp*B	26	0.4844	Ocular	79	36	0.00075	0.00164	2.182	16	20
			Urogenital/rectal	455	20	0.00126	0.00092	0.726	9	11
			LGV	70	1	0	0	0	1	0
*trp*A	19	0.5883	Ocular	79	8	0.00074	0.00164	2.212	3	5
			Urogenital/rectal	455	16	0.00158	0.00209	1.317	6	10
			LGV	70	0	0	0	0	0	0

[a]Note that TrpR-, TrpB-, and TrpA-CDS were extracted from individual strains and segregated into ocular, urogenital, rectal, and LGV disease groups; TrpA-CDSs from urogenital strains were further segregated into vaginal and urethral strains. Diversity metrics were calculated in DnaSP6 using the MAFFT alignment (see Materials and Methods).

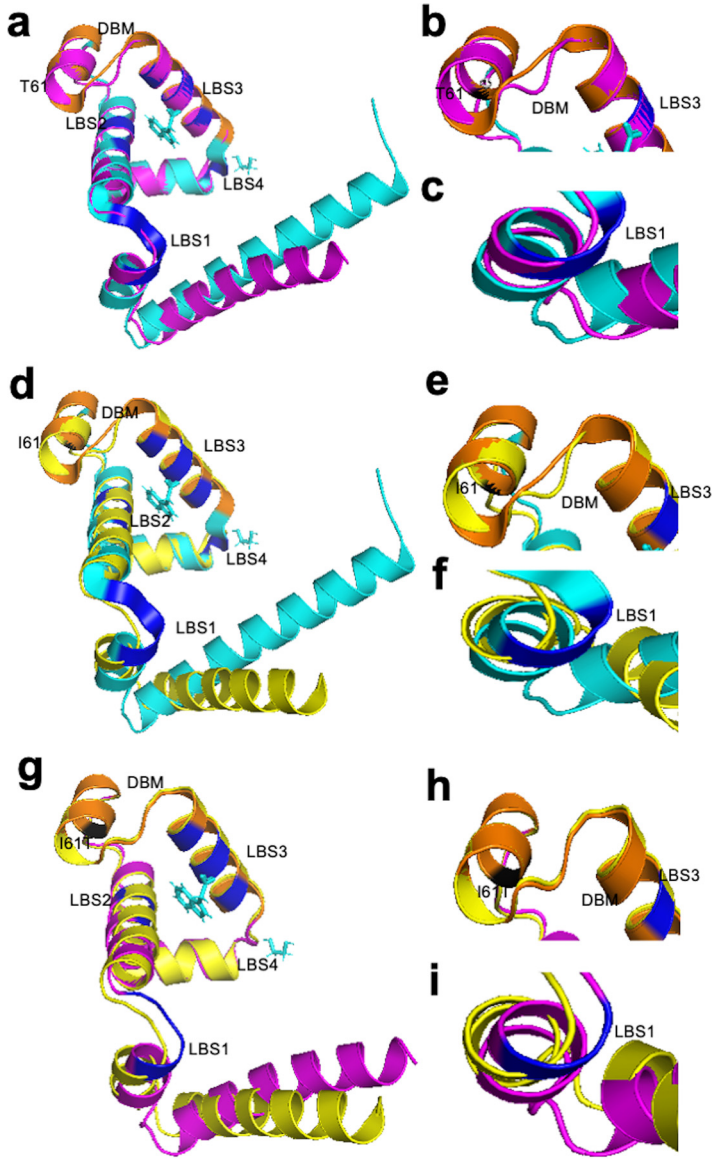


FIG 6 TrpR 3D predicted structures of *Ct* clinical E_R27091 and reference E_Bour strains (a to i). (a) TrpR structure of mutant strain E_R27091 (magenta) superimposed on the template 6eniA (cyan), with DNA-binding motif (DBM) in orange. The putative ligand-binding sites (LBS) 1, 2, and 3 are shown in dark blue, and the aa substitution in relation to E_Bour at T61 is in black. (b and c) Structural changes in DBM and LBS1, respectively, of the mutant are shown. (d) TrpR structure of E_Bour (yellow) superimposed on the template 6eniA (cyan) with annotations as per panels a, b, and c. (e and f) Structural changes in DBM and LBS1 of E_Bour. (g to i) TrpR 3D predicted structures of E_R27091 superimposed on E_Bour (g) with structural changes noted in relation to DBM based on I61T aa substitution (h) and E_Bour at LBS1 (i).

mutations are crucial in affecting repressor activity by altering its affinity with either the operator-DNA or the ligand IAC. There was also a predicted decrease in affinity of LBS for EDO (ethanediol), except E_R27091, suggesting that this mutation is neutral in terms of functional impact. The majority of the mutations were, therefore, predicted to decrease the overall stability of TrpR mutant proteins (Table 2).

TrpB and TrpA mutant strains reveal structural and functional changes in protein-ligand affinity affecting tryptophan synthase function. The $\alpha\beta\beta\alpha$ tetramer is responsible for two steps in tryptophan biosynthesis. The α-subunit catalyzes aldolytic cleavage of indole-3-glycerol phosphate (IGP) to glyceraldehyde 3-phosphate (GAP) and indole; indole is transferred via an intermolecular 25-Å-long hydrophobic

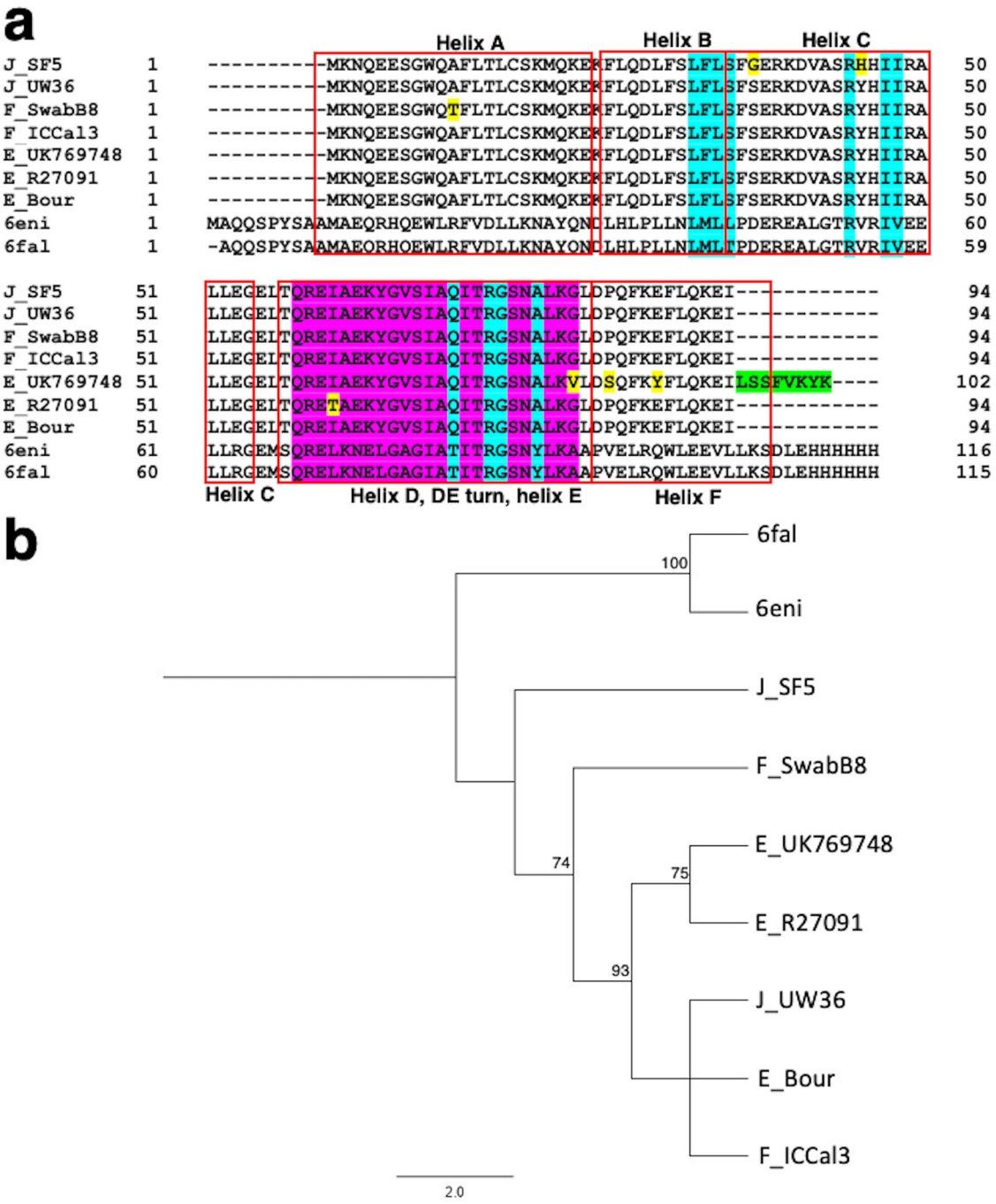


FIG 7 TrpR amino acid alignment (a) and phylogeny (b) of *Ct* mutant strains (E_R27091, E_UK769748, F_SwabB8, and J_SF5), reference strains (E_Bour, F_IC-Cal-3, and J_UW-36) and Protein Data Bank (PDB) templates (6fal and 6eni). (a) TrpR aa alignments were created using Clustal Omega, and the residues involved in putative ligand-binding sites (cyan), DNA-binding motif (magenta), and helices (box outlined in red) were annotated based on the crystal structures of *Escherichia coli* 6fal and 6eniA available from PDB and NCBI. Amino acid substitutions (yellow) and extension (green) in the mutant strains relative to their reference strain are annotated. (b) TrpR phylogeny of *Ct* strains and templates; the PDB aa templates were constructed using the approximate maximum likelihood phylogenetic tree in FastTree 2.1.11 with a generalized time-reversible model with 1,000-bootstrap sampling. Scale bar represents the number of differences or distance between the sequences.

tunnel to the active site of the β-subunit where it condenses with L-serine in a pyridoxal phosphate (PLP)-dependent reaction to produce L-tryptophan and water (24). The reactions occur when α- and β-subunits are in the closed conformation to mutually activate each other. Formation of the indole tunnel involves residues α176 to α196

TABLE 2 Predicted effects of urogenital *C. trachomatis* TrpR, TrpB, and TrpA mutant strains on protein stability, protein interaction with ssDNA, and ligand affinity noted as change in Gibbs free energy ($\Delta\Delta G$) and log (affinity fold change)[d]

Mutant	Mutation	Mutation location	Effect of mutation on protein stability ($\Delta\Delta G$)	Effect of mutation on protein interaction with ssDNA ($\Delta\Delta G$)	Effect of mutation on protein-ligand affinity (log[affinity fold change])	
TrpR[a]					EDO	IAC
E_UK769748	G81V	DNA-binding motif (DBM)	0.72	0.49	−1.83	−1.083
	P84S	Ligand-binding site (LBS)	0.49	1.096		−0.153
	E88Y	Helix F	−0.74	0.434	−2.38	−2.207
E_R27091	I61T	DNA-binding motif (DBM)	−3.19	1.078	0.36	0.345
F_SwabB8	A11T	Helix A	−2.05	1.566	−0.83	−0.682
J_SF5	S36G	Helix C	0.08	−1.538	−0.7	−0.239
	Y45H	Helix C	−0.34	0.38	−1.41	−1.404
TrpB[b]					PLP	Na^+
D_NL12	R147C	β-COMM domain	−0.885		−0.49	−1.19
D_SF12	D326E	β-subunit	−0.705		0.11	−1.016
E_SotonE8	R69K	β-subunit	−1.01		−0.47	−1.702
	P304S	β-subunit	−2.228		0.75	−0.339
	P381L	β-COMM domain	−0.677		0.19	−0.773
G_SF20	G190R	β-subunit	−0.634		−0.49	−0.798
I_UK913341	S36N	β-subunit	−0.239		0.39	−0.303
	R69K	β-subunit	−1.01		−0.47	−1.702
	T177K	Proximal to β-subunit interface residues	−0.771		0.5	−0.493
	H182Q	β-COMM domain	−1.852		−0.06	−0.176
J_SF5	A19T	TrpA interaction	−1.105		0.78	−0.285
	N36S	β-subunit	−0.41		−0.61	−1.307
J_S3107	N36S	β-subunit	−0.41		−0.61	−1.307
	A217V	β-subunit	−0.198		0.23	-0.438
TrpA[c]					MLI	FMT
E_Fin214	D22Y	α-subunit	−0.384		−0.56	−0.725
F_Fin219	A215V	α-subunit	−0.489		−0.39	−0.488
F_SF11	G253D	α-subunit	−0.478		2.05	1.481
G_UK750369	L13P	α-subunit	−1.334		−0.18	−0.442
	Q37R	α-subunit	0.147		−0.3	−0.447
	A115V	α-subunit	−0.435		−0.3	−0.436
	C177Y	α-L6	−0.473		−0.21	−0.366
	R206K	Catalytic domain	−1.396		−0.88	−0.88
I_UK913341	Q37R	α-subunit	0.227		−0.3	−0.454
	Q103K	α-subunit	−0.172		−0.29	−0.62
	A115V	α-subunit	−0.386		−0.3	−0.436
	E178Q	α-L6	−0.757		−0.724	−0.723
	Y220H	α-subunit	−1.576		−0.35	−0.532
J_UK583676	A247E	α-subunit	−0.606		0.71	0.454

[a]See Fig. 5 for location and position of the mutation in the TrpR amino acid sequence.
[b]See Fig. 6 for location and position of the mutation in the TrpB amino acid sequence.
[c]See Fig. 9 for location and position of the mutation in the TrpA amino acid sequence.
[d]Abbreviations: EDO, ethanediol; IAC, indole acetic acid; PLP, pyridoxal 5′-phosphate; Na^+, sodium ion; MLI, malonate ion; FMT, formic acid.

that trap indole and is dependent on TrpA α-Loop 6 (α-L6) mobility for allosteric interaction with TrpB beta communication (β-COMM) domain (33).

For TrpB, seven urogenital mutant strains and their reference strains had the highest homology with the last bacterial common ancestor (LBCA), 5ey5, with >95% correct folds and a GA341 MODELLER score of 1.0 with 57% identity to *Ct* strains and an overall RMSD of <1 Å; alignments including templates 5ey5, 5cgq, and 1x1q are shown in Fig. 8a. Sodium ion (Na^+) and PLP were identified as the bound ligands for the monovalent cation (MVC)-binding site. Phylogenetic reconstruction of *Ct* strains and the templates is shown in Fig. 8b. Models are shown for E_SotonE8 βP304S (MVC-binding site), I_UK913341 βT177K, and βH182Q (β-COMM domain with βT177K proximal to residues interacting with α-subunit) in Fig. 9a to c and Fig. S5A to C but not for D_NL12 βR147C in the β-COMM domain or mutations in D_SF12, G_SF20, J_S3107, and J_SF5 due to space limitations. All mutations predicted a decrease in protein-ligand affinity for Na^+,

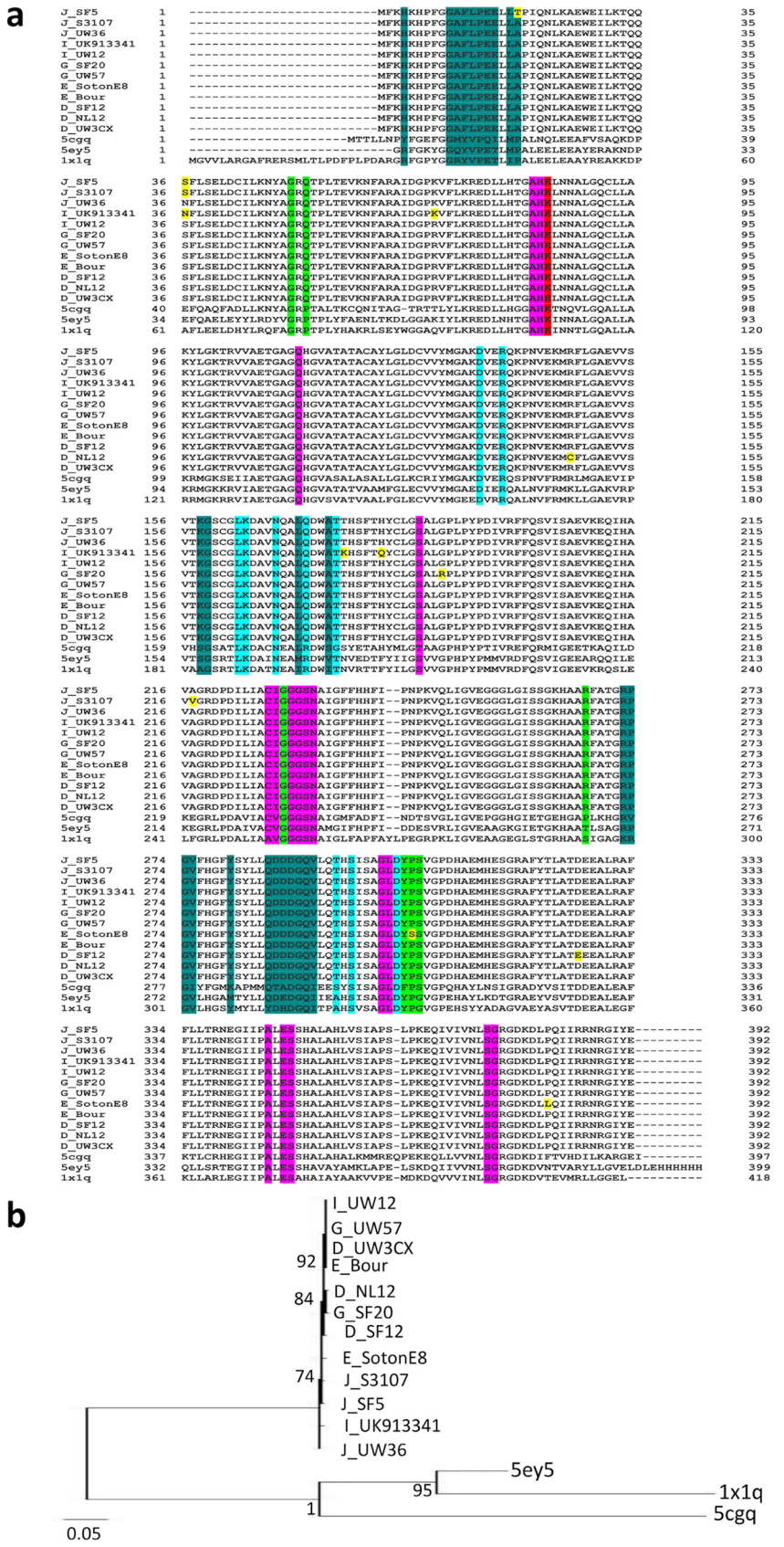


FIG 8 TrpB amino acid alignment (a) and phylogeny (b) of *Ct* mutant strains (D_NL12, D_SF12, E_SotonE8, G_SF20, I_UK913341, J_S3107, and J_SF5), reference strains (D_UW3CX, E_Bour, G_UW57,

(Continued on next page)

while seven mutations showed reduced affinity for PLP but resulted in reduced protein stability for all mutant strains (Table 2).

Analyses performed in Arpeggio revealed that wild-type residue P304 forms a hydrogen bond with L284 and Y304 and repulsive van der Waals clash with Na^+ to activate the β-subunit (see Fig. 12a). The E_SotonE8 P304S results in loss of van der Waals clash with Na^+ and a weaker affinity of the β-subunit for Na^+ in addition to loss of hydrogen bonding with Y303 (see Fig. 12b), suggesting loss of β-subunit activation. Wild-type residue T177 is proximal to residues interacting allosterically with the α-subunit. T177 makes a polar van der Waals clash with S179, F180, and D173, weak polar van der Waals clash with T181 and D173, and a van der Waals clash with S179 and R102, in addition to a carbonyl and proximal bond with W174 (see Fig. 12c). W174 and D173 are β-subunit residues involved in allosteric interactions with the α-subunit (see Fig. 12c, cyan). The I_UK913341 T177K results in loss of interaction with R102 and all polar interactions with D173, S179, and T181 (see Fig. 12d), indicating suboptimal allosteric interactions with the α-subunit.

TrpA protein structures were modeled for seven urogenital strains and their respective reference strains, and all, except for G_UK750369 and G_UW57, had the closest homology to *Aquifex aeolicus* VF5 2ekc with >95% correct folds and a GA341 MODELLER score of 1.0, which had 32% identity to *Ct* strains; alignments with templates 5tch, 5ey5, and 2ekc are shown in Fig. 11a below. G_UK750369 and G_UW57 had the closest homology to *Mycobacterium tuberculosis* 5tch with a GA341 MODELLER score of 1.0 and sequence homology of 32% with an overall RMSD of 2.6 Å (Fig. 10a and d, magenta). Malonate ion (MLI) and formic acid (FMT) were identified as the bound ligands for the catalytic subunit. The phylogenetic tree is shown in Fig. 11b.

Modeling G_UK750369 αC177Y and αR206K located in α-L6 (Fig. 11a; also Fig. 10a and b, yellow) that are catalytic and binding sites for MLI (Fig. 11a; also Fig. 10a and c, yellow) resulted in structural changes in α-L6 and the catalytic site compared to G_UW57 (Fig. 10g to i) and 5tch (Fig. 10d, e, and f). I_UK913341 αE178Q located in α-L6 and mutations in E_Fin214, F_Fin219, F_SF11, and J_UK583676 were located elsewhere on the protein (Fig. 11a). All mutations resulted in reduced affinity for ligands MLI and FMT in addition to decreased protein stability, except for αQ37R, which had an increased/neutral effect on protein stability, and αG253D, which had increased affinity for MLI and FMT but reduced protein stability (Table 2).

Wild-type residue C177 forms polar contacts, weak polar contacts, and van der Waals clash interactions with the side chain of R181 (Fig. 12e). G_UK750369 C177Y resulted in loss of all interactions with R181 and any surrounding residues (Fig. 12f) with reduced affinity to MLI and FMT and decreased protein stability (Table 2). Interestingly, the Y177 residue was present in all *Ct* reference and mutant strains (Fig. 11a), suggesting that urogenital *Ct* strains inherently have compromised TS function.

Wild-type residue R206 located in the catalytic and binding site for MLI as per 5tch forms polar contacts (Fig. 12g; orange dotted lines) with MLI, hydrophobic van der Waals contact with the side chain of I208, polar van der Waals clash with D204, and polar contacts with the side chain of R181 located in α-L6 (Fig. 12g). R206K forms polar bonds and gained van der Waals clashes with MLI, lost the polar contact with R181 in α-L6 and D204, and gained van der Waals clashes and weak hydrogen bonds with I187 and T180. It also had weak polar and undefined van der Waals clashes with I208

FIG 8 Legend (Continued)

I_UW12, and J_UW36) and Protein Data Bank (PDB) templates (5cgq, 5ey5, and 1x1q). (a) TrpB aa alignments were created using Clustal Omega, and the residues involved in Na^+ ion and pyridoxal 5′-phosphate interaction (green and magenta), catalytic activity (red), TrpA interaction (dark green), and interface of α-Loop 6 (α-L6) and monovalent cation (MVC)-binding loop (cyan) were annotated based on the crystal structures and sequence of 5ey5 from PDB and NCBI. Amino acid substitutions (yellow) in the mutant strains relative to their reference strain are also annotated. (b) TrpA phylogeny of *Ct* strains and templates was constructed using the approximate maximum likelihood phylogenetic tree in FastTree 2.1.11 with a generalized time-reversible model with 1,000-bootstrap sampling.

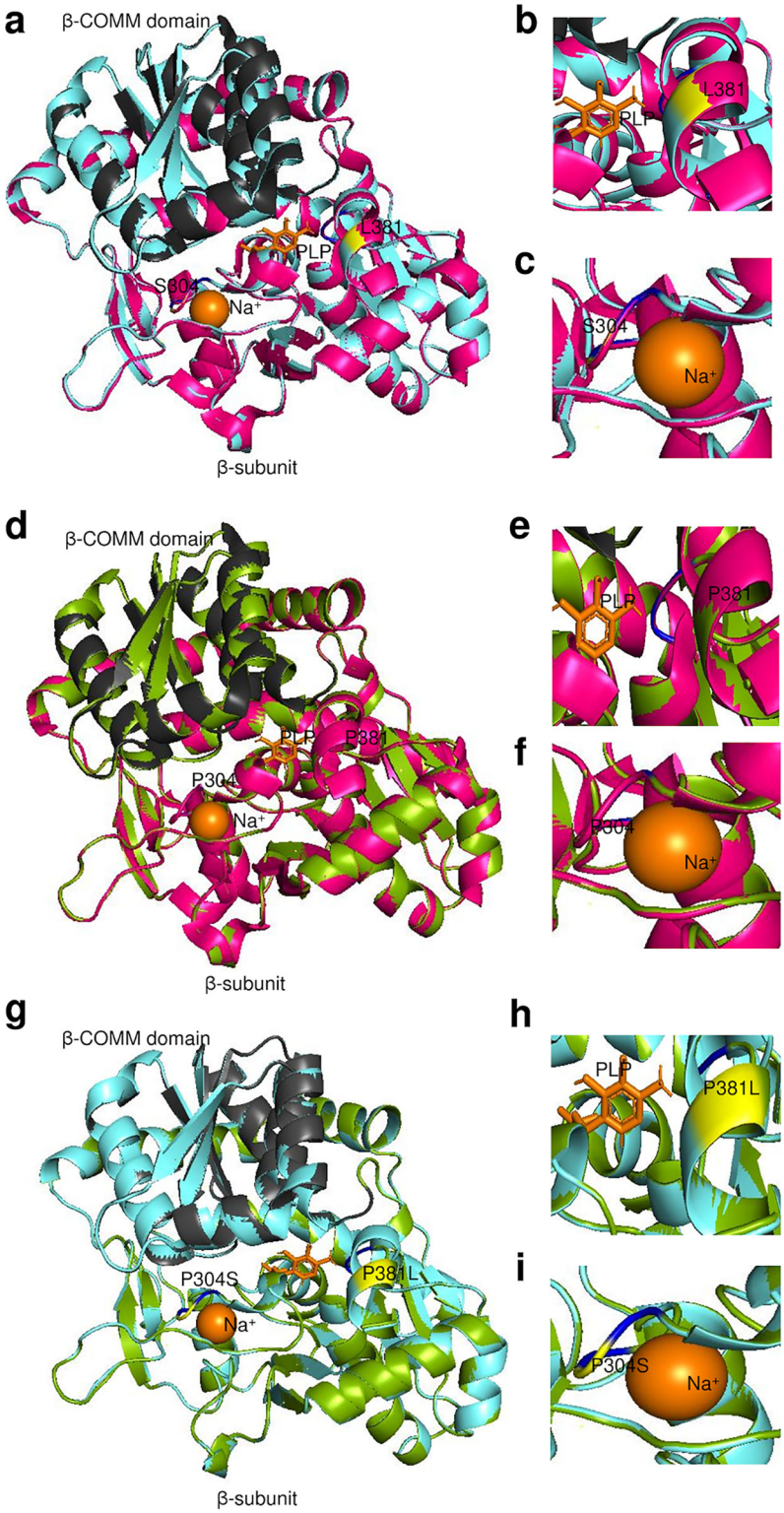


FIG 9 TrpB 3D predicted structures of *Ct* clinical E_SotonE8 and reference E_Bour strains (a to i). (a) TrpB structure of mutant strain E_SotonE8 (cyan) superimposed on the template 5ey5 (magenta), with the β-COMM domain in charcoal gray. Na^+ and the cofactor pyridoxal phosphate (PLP) interacting residues are shown in blue, and the aa substitution in relation to E_Bour at S304 and L381 is in yellow. (b and c) Structural changes in Na^+ and the cofactor PLP interacting residues of the mutant E_SotonE8 are shown. (d) TrpB structure of E_Bour (green) superimposed on the template 5ey5 (magenta) with annotations as per panels a, b, and c. (e and f) Structural changes in Na^+ and the cofactor PLP interacting residues of E_Bour. (g to i) TrpB predicted structures of E_SotonE8 superimposed om E_Bour (g) with structural changes noted in relation to PLP interacting residue on P381L (h) and Na^+ interacting residue on P304S (i).

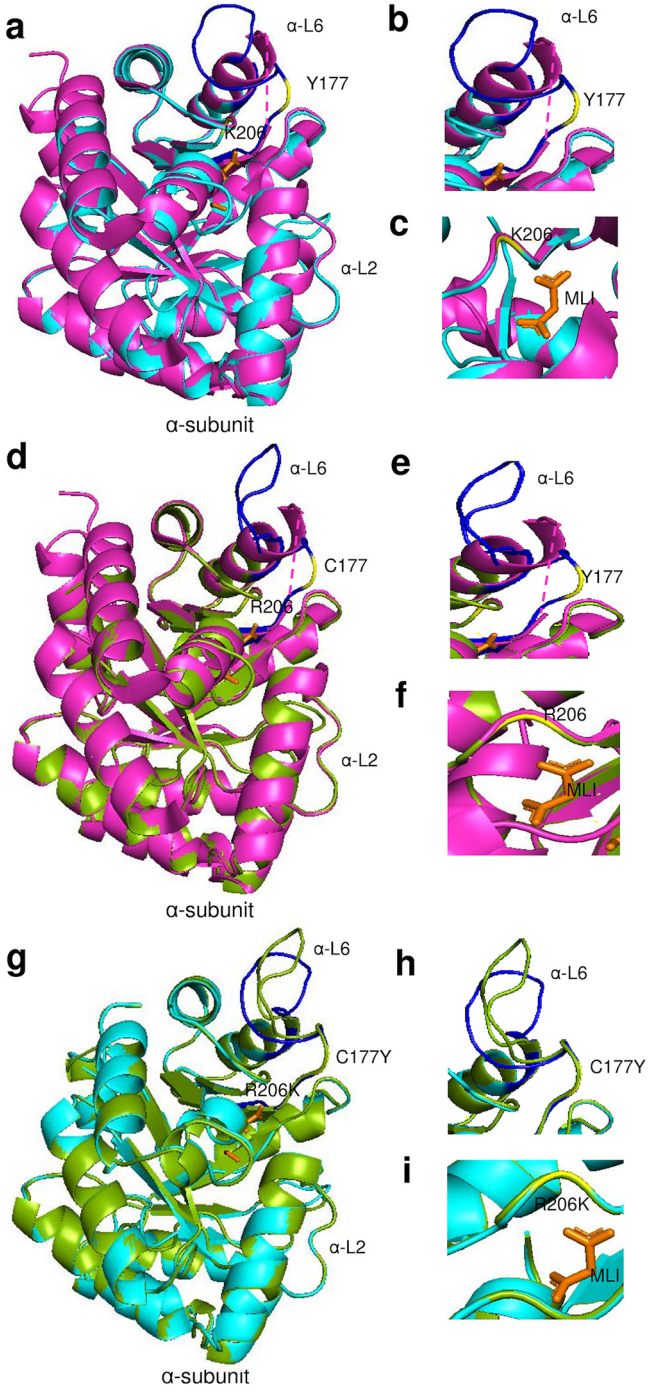


FIG 10 TrpA 3D predicted structures of *Ct* clinical G_UK750369 and reference G_UW57 strains (a to i). (a) TrpA structure of mutant strain G_UK750369 (cyan) superimposed on the template 5tch (magenta), with α-Loop 6 (α-L6) of the α-subunit in blue and malonate ion (MLI) in orange. The aa substitution in relation to G_UW57 at Y177 and K206 is in yellow. (b and c) Structural changes and location of the α-L6 and catalytic and subunit interaction regions of the mutant G_UK750369 are shown. (d) TrpA structure of G_UW57 (green) superimposed on the template 5tch (magenta) with annotations as per panels a, b, and c. (e and f) Structural changes in the α-L6 and catalytic and subunit interaction regions of the reference G_UW57. (g) TrpA predicted structures of G_UK750369 superimposed on G_UW57 (g) with structural changes noted in relation to α-L6 based on C177Y aa substitution (h) and catalytic and subunit interaction region based on R206K (i).

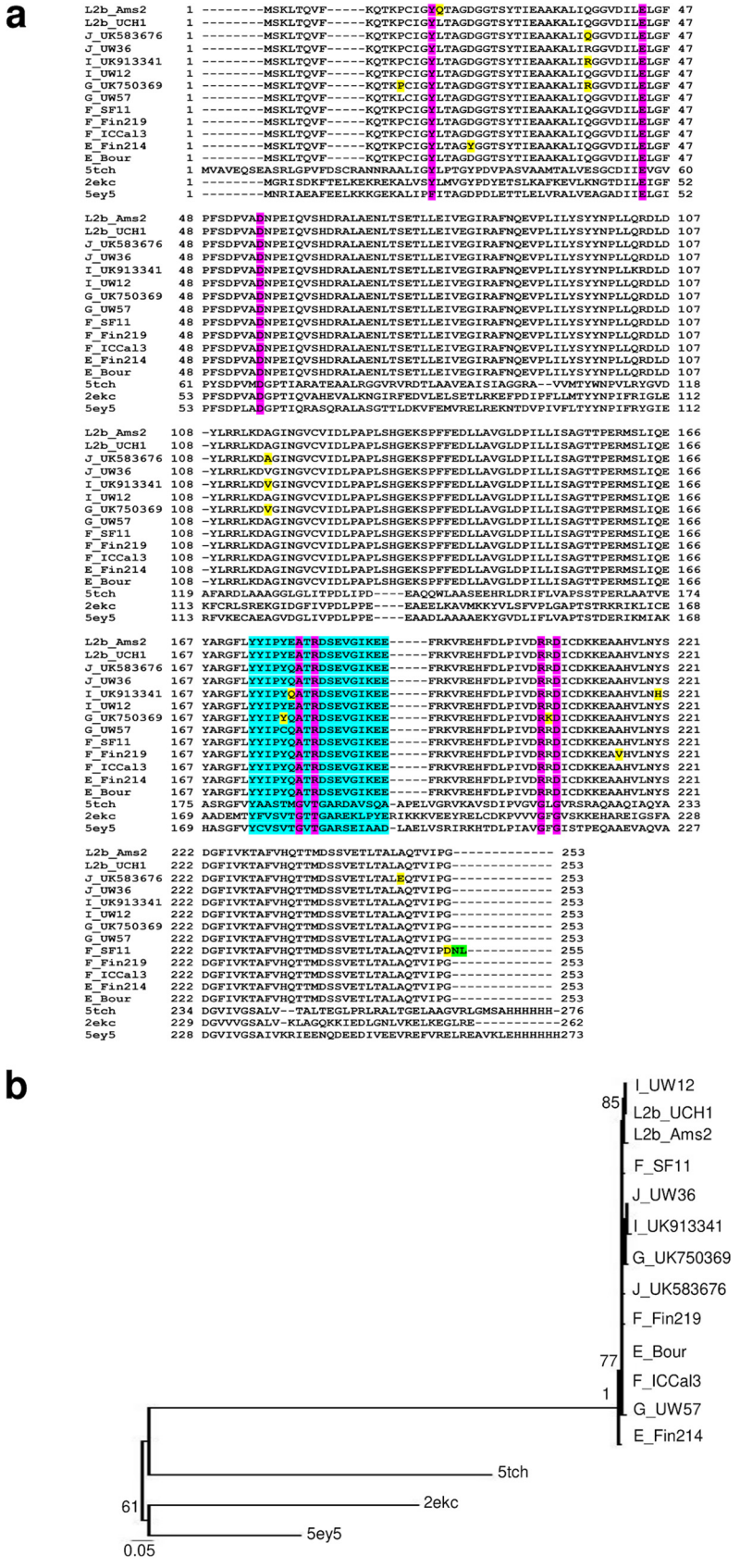


FIG 11 TrpA amino acid alignment (a) and phylogeny (b) of *Ct* mutant strains (E_Fin214, F_Fin219, F_SF11, G_UK750369, I_UK913341, J_UK583676, and L2b_Ams2), reference strains (E_Bour, F_ICCal3,

(Continued on next page)

(Fig. 12h). R206K resulted in reduced affinity to MLI and FMT and decreased protein stability (Table 2).

DISCUSSION

Human- and animal-pathogenic species of the *Chlamydiaceae* family such as *Ct*, *Chlamydia psittaci*, and *Chlamydia abortus* represent the most evolutionarily reduced genomes compared to family members of *Simkaniaceae*, *Parachlamydiaceae*, and *Waddliaceae* (34). *Chlamydiaceae* are the most recently evolved member with the smallest chromosome, likely reflecting recent adaptation to a relatively stable, homeostatic environment within mammalian and avian host cells. Other family members have a much wider range of eukaryotic hosts, including protozoans such as amoebae (34). Chlamydiae genomes contain a "plasticity zone" (PZ) (35), which has undergone increased levels of genetic variation compared to other chromosomal regions and includes virulence factors (36) such as the tryptophan operon. *Simkania negevensis* is the only species that has a complete operon comprised of *trp*R, *trp*Aa-Ab, *trp*B, *trp*D, *trp*C, *trp*Eb-Ea, and *aro*A that likely represents the ancestral operon for the phylum (37).

Comparative genomics of *Ct* strains have shown that ocular strains have emerged at a later time point in evolution than the nonprevalent urogenital strains that appeared ~10 to 15 million years ago (11). It is thought that, through reductive evolution, ocular strains lost their ability to synthesize tryptophan to maximize fitness and survive in a presumably indole-deficient ocular environment (26, 38). However, more recent ocular microbiome data suggest that indole-producing bacteria such as *Propionibacterium acnes* and *Escherichia coli* are commonly present on the ocular surface of the healthy conjunctiva (39, 40). Other mechanisms may therefore be at play.

Analysis of the expanded number of *Ct* strains in the present study revealed an increased diversification of the operon for the less studied *trp*R region and IGR as well as *trp*B and *trp*A. Mutations occurred in archival samples from the early 1900s to early 2000s for ocular samples while most of the mutations in urogenital strains were in the 2000s, spanning 59 years. Our sample set was biased toward prevalent genotypes E, F, and G with some rarer genotypes present as singletons or in smaller numbers (e.g., Ba, C, Da, Ja, L2c, L3). Increased sampling would likely reveal a greater diversity of genomes and delineate timeline-based mutation accumulation in ocular, urogenital, and rectal strains.

We found that the majority of genetic variation consisted of SNPs and indels with no evidence of recombination occurring within the operon, consistent with the relative congruence of the phylogenetic trees. The findings concur with our previous *Ct* genome analyses where, despite evidence for recombination occurring frequently at multiple sites throughout the chromosome (10, 11), it occurred less frequently than mutation (mean ρ/θ of 0.12; 95% credibility interval of 0.10 to 0.23) (16). Although it should be noted that the operon is part of a much larger recombinant fragment (18), our current data indicate that mutation is a significant driving force in the evolution of the operon.

Operon and *trp*A phylogenies were not entirely congruent with the well-described *Ct* whole-genome phylogeny (11, 16), except for the ocular strains. The operon phylogeny revealed a new clade consisting of an admixture of prevalent and nonprevalent urogenital strains with the prevalent clade as the earliest ancestor. Each clade appeared to be undergoing diversification and coevolution with polyphyletic branching of the nonprevalent clade. However, subclades arising from mixed and prevalent

FIG 11 Legend (Continued)
G_UW57, I_UW12, J_UW36, and L2b_UCH1), and Protein Data Bank (PDB) templates (5tch, 5ey5, and 2ekc). (a) TrpA aa alignments were created using Clustal Omega, and the residues involved in catalytic and subunit interaction (magenta), α-L6 residues α176 to 196, were annotated based on the crystal structures and sequence of 5tch from PDB and NCBI. Amino acid substitutions (yellow) and extension (green) in the mutant strains relative to their reference strain are also annotated. (b) TrpA phylogeny of *Ct* strains and templates was constructed using the approximate maximum likelihood phylogenetic tree in FastTree 2.1.11 with a generalized time-reversible model with 1,000-bootstrap sampling.

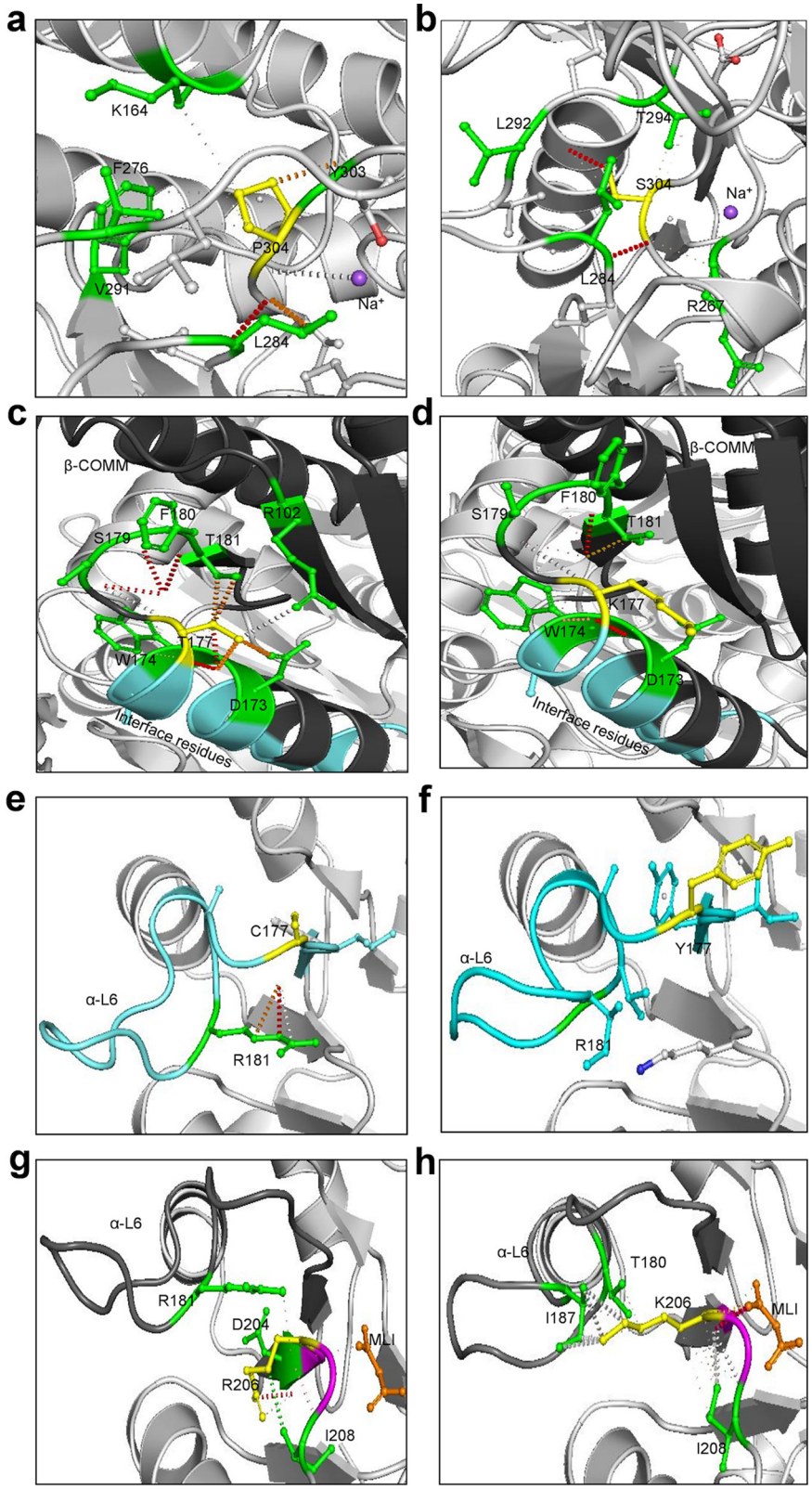


FIG 12 Effect of wild-type and mutant residues on interatomic interactions in TrpA and TrpB protein structures and small-molecule ligand binding. (a and b) The wild-type residue P304 (yellow) of E_Bour β-subunit (a) and the mutant residue S304 (yellow) of E_SotonE8 β-subunit (b) interactions with neighboring residues (green) and the ligand Na^+ (purple). (c and d) The wild-type residue T177 (yellow) of I_UW12 β-subunit (c) and the mutant residue K177 (yellow) of I_UK913341 β-subunit (d)

(Continued on next page)

urogenital clades were clonal. The ocular strains exhibited a more rapid monophyletic diversification, including, for example, clonal A strains from a single geographic site in Tanzania (16).

Operon SNP analysis revealed six urogenital strains with unique mutations in *trp*R. Strain E_UK769748 had five SNPs with an eight-residue extension at the C terminus of TrpR. These strains were isolated between 2003 and 2012, suggesting recent mutations, although no samples with complete operon sequences were available from these regions prior to 2003. Interestingly, the ocular and LGV TrpR-CDS had relatively high synteny.

B_QH111L had two nonsynonymous aa changes in the IGR at the exact binding site for YtgR, the only known iron-dependent transcription factor in *Chlamydia* (29). YtgR binds to the novel IRG promoter and regulates *trp*B and *trp*A expression during iron starvation. Simultaneously, YtgR binding promotes termination of transcripts from the primary promoter upstream of TrpR29. Although the B_QH111L *trp*BA mutations and resulting protein truncations suggest reductive evolution resulting in loss of TS function, the IGR mutations indicate selective pressure that disables the response to fluctuating iron levels with further loss of operon function.

Eight urogenital strains had novel mutations in *trp*B as did some ocular strains (Fig. 4). Gambian strain B_M48 had an indel with a frameshift and early truncation of the protein. Previously described B_QH111L from China and B_TZ1A828 from Tanzania both had frameshifts and early truncation of the protein (25, 28, 41). These strains were isolated between 1998 and 2016, similar to the time frame for the *trp*R mutations.

*trp*A SNP analyses revealed mutations and indels with frameshifts and early truncation of the CDS similar to previous studies (25, 42) but also novel SNPs for both ocular and urogenital strains. The former included indels at nt 529 and 138 with TrpA truncation at 184 aa and 46 aa, respectively, and no evidence for B strain operons as previously reported (25). Novel SNPs produced a stop codon in B_HAR36, a frameshift and early truncation in B_QH111L, and an intact *trp*A sequence for B_TZ1A828, which was remarkable and similar to urogenital strains. Except for B_TZ1A828, these findings were not unexpected in that ocular strains have uniformly evolved toward a loss of operon function.

Seven urogenital strains and a single LGV strain from 2003 to 2012 had unique *trp*A mutations. We previously reported on serial isolates of strain F_SF11 (i.e., F_I-IV) isolated in 2010 that had a 2-aa TrpA extension with decreased function *in vitro* (26). These data were surprising as the operon is thought to be highly conserved and functional among urogenital strains. To further analyze evolution among the strains, we evaluated evidence for selective pressure. Both ocular and urogenital strains were under positive selection with virtually no mutations in the LGV strains, suggesting that these latter strains are under strong stabilizing pressure where the operon has likely become fixed. Indeed, LGV strains are known to be markedly more resistant to IFN-γ inhibition than non-LGV strains. Further segregation of the strains by anatomic site showed that urethral strains were under greater positive selection than vaginal strains. The higher Pi(a)/Pi(s) for these strains compared to the ocular strains suggests a more rapid evolution of beneficial mutations. Interestingly, the former were from populations in Argentina, Italy, Russia, Sweden, the United Kingdom, and the United States with diverse genotypes (e.g., D, E, F, G, H, J, and K) and include populations of men who have sex with men (MSM). The high frequency of sexual activity with risky sexual

FIG 12 Legend (Continued)
interacting with $\alpha\beta$ interface (cyan) and neighboring residues (green). (e and f) The wild-type residue C177 (yellow) of G_UW57 α-subunit (e) and mutant residue Y177 (yellow) of G_UK750369 α-subunit (f) interactions with residue R181 (green) located in the α-L6 (cyan). (g and h) The wild-type residue R206 (yellow) of G_UW57 α-subunit (g) and mutant residue K206 (yellow) of G_UK750369 α-subunit (h) interactions with residues (green) in the α-L6 and malonate ion (MLI; orange). The α-L6 region shown in charcoal gray and catalytic site in magenta. Polar contacts (red), weak polar contacts (orange), van der Waals clash interactions (gray), and hydrophobic van der Waals contact (green) between residues and ligands are shown as dotted lines.

behavior may drive urethral strain selection and mutations in these latter populations (3, 43–45). Selection could also be enhanced by additional pressure from host immune responses elicited by exposure to microbes in various anatomic sites that might have been visited during sex (e.g., oropharynx, rectum). This would indicate that diversification might be favored toward increasing gene fitness (46), but could also lead to deleterious mutations as in Muller's ratchet (47).

To further explore operon functionality of the mutations, we used a three-dimensional (3D) protein homology modeling approach to assess $\alpha\beta\beta\alpha$ structural variations as we previously described (26) in addition to TrpR modeling and the effect on protein stability, ligand-binding affinity, and TrpR affinity for ssDNA measured by calculating free energy changes ($\Delta\Delta G$) between reference and mutant operon proteins. Similar to many other Gram-negative bacteria, expression of chlamydial *trp*BA genes is tightly regulated by TrpR. When in contact with its corepressor tryptophan, TrpR prevents *trp*BA transcription by binding to the *trp* operator-DNA (48). Tryptophan starvation, in contrast, induces transcription of *trp*BA to restore tryptophan levels in the intracellular environment (25, 26). Several studies with *E. coli* TrpR have reported that indole derivatives such as IAC derepress the operon by forming a "pseudorepressor" that displaces tryptophan and has poor affinity for the operator-DNA, thereby allowing *trp*BA transcription (30, 49, 50). However, if indole is not available as a substrate for synthesis when TS has been produced, ammonia will be produced that is detrimental to *Ct* as it has antibacterial properties (26, 38). This has also been shown to be the case with other intracellular pathogenic bacteria such as *Legionella pneumophila* and *M. tuberculosis* (51, 52).

We modeled TrpR mutants and their *Ct* reference strains against the closest homology template, *E. coli* 6eni. Strains E_UK769748, F_SwabB8, J_SF5, and J_SF6 exhibited a decreased affinity for ligand IAC, and E_UK769748, E_R27091, E_R29005, and F_SwabB8 had increased affinity for ssDNA (*trp* operator-DNA). Since binding of TrpR at the operator-DNA interferes with RNA polymerase binding, initiation of *trp*BA transcription would be prevented (53), eliminating any risk of ammonia production if indole was in low supply. All but E_UK769748 had a net decrease in protein stability with a reduced energy expenditure. Finally, G81V and P84S in E_UK769748 and S35G in J_SF5 had near-neutral effects on protein stability and entropy, suggesting that these mutations arise at no/low fitness cost.

The closest protein model for *Ct* TrpB mutants and their reference strains was *E. coli* 5ey5 with cofactor PLP and Na^+ as the ligand. In TS catalysis and regulation, cation activation is achieved through an allosteric linkage connecting the PLP active site (54). MVC consists of a backbone of carbonyls that can accommodate cations of different size and incorporate ligated water molecules to satisfy electrostatic requirements of cation-bound MVC. MVC-bound Na^+, in contrast to the MVC-free form (i.e., without ion binding), creates a conformational change in the β-COMM domain allowing allosteric interactions with α-L6 in TrpA to a closed conformation with indole tunnel formation. *E. coli* and hence *Ct* can bind only Na^+. However, *in vitro* studies of other bacteria have found that cations such as Cs^+ more effectively induce a closed conformation that enhances TS function compared to Na^+ (54), indicating that *Ct* strains likely do not have an optimally functional TS.

Urogenital strain D_NL12, E_SotonE8 and I_UK913341 mutations were located within the β-COMM domain and predicted to reduce Na^+ ligand affinity and therefore TS activity as a result of no/low catalytic activation and perturbed allosteric interactions. βP304S—a nonpolar hydrophobic proline substituted for a polar hydrophilic serine—in the MVC-binding site of E_SotonE8 may have also impacted catalysis and allosteric interactions. Interestingly, mutations D326E in D_SF12, T177K in I_UK913341, and A217V in J_S3107 had increased affinity to PLP, which appeared to be detrimental to the Na^+ interaction with MVC (Table 2). All of these mutations resulted in decreased protein stability due to the reduced affinity to Na^+ with reduced energy expenditures.

The TrpA protein model for *Ct* mutants and reference strains was *M. tuberculosis* 5tch with MLI and FMT as the ligands. G_UK750369 αC177Y and I_UK913341 αE178Q

mutations located within α-L6 and G_UK750369 αR206K in the catalytic and subunit interaction domain exhibited reduced ligand affinity for both with reduced protein stability and energy expenditure. Only the αQ37R mutation in G_UK750369 and I_UK913341 predicted an increased protein stability with a lower energy cost, suggesting greater fitness (Table 2). Interestingly, the αC177Y mutation was present in many urogenital reference and mutant strains, indicating that urogenital strains in general have reduced protein stability and ligand affinity for MLI and FMT (Fig. 11a). Furthermore, mutations present elsewhere on the α-subunit of E_Fin214, F_Fin219, and J_UK583676 had reduced ligand affinity and protein stability, indicating a reduction in synthase efficiency through suboptimal allosteric interactions. Our previously reported F_SF11 mutant (26) showed a predicted increase in affinity for MLI and FMT but reduced protein stability. This mutation likely increased energy utilization in interacting with MLI and FMT, which would have had a destabilizing effect on protein structure. This is not surprising in that F_SF11 had markedly decreased tryptophan synthesis and lower uptake of tryptophan for metabolism *in vitro* (26). While it would have been advantageous to perform *in vitro* studies on additional strains in this data set, the majority were not available or not available as viable organisms.

It has been proposed that *Ct* evolved from an ancestor that colonized the gastrointestinal tract (55), a site with abundant indole from indole-producing bacteria. Consistent with this, we found that rectal strains belonging to the LGV and urogenital lineages had intact functional operons while mutations that render the operon nonfunctional were found only among urethral and vaginal/endocervical strains, suggesting that the former strains have reached a high degree of fitness in this anatomic site. *trp*R, *trp*A, and *trp*B have undergone multiple mutation events in the latter strains, indicating disparate evolutionary strategies to either maintain the functional operon or lose function while preserving its ability to scavenge the intracellular host environment for tryptophan. Our findings dispel the dogma in the field that there is strong selective pressure for all urogenital strains to maintain a functional operon. Niche specificity appears to dictate maintenance or loss of function. While the ability to utilize indole seems advantageous for this bacterium, loss of function through evolution and selective pressure as in the case of ocular strains seems to have benefited the organism by allowing it not to have to synthesize tryptophan, which is energetically costly (56). Clinical benefit for this loss of function was also demonstrated in our previous study for sexually transmitted urogenital strain F_SF11 (26). Based on the present study, we are seeing this same trend in urogenital strains across the globe: strains with mutant proteins were inclined toward energy conservation by exhibiting a decreased affinity for their respective ligands. This is not surprising because L-tryptophan is the one of the most energetically costly amino acids to synthesize (56). Our data, therefore, indicate a novel host-pathogen evolutionary mechanism for intracellular survival whereby urogenital strains are evolving more rapidly with mutations that impact tryptophan operon function in a manner that is energetically beneficial for the organism.

MATERIALS AND METHODS

Tryptophan operon sequences, genome assembly, and/or operon sequencing. Of 562 genomes, 486 were available from NCBI SRA and 76 from GenBank. Three unpublished *Ct* genomes and 30 unpublished *Ct* clinical samples with complete operon sequences from the Dean Laboratory were also available, totaling 595 high-quality operon sequences (Fig. 1; see also Table S1 in the supplemental material).

The 486 SRA genomes were imported and assembled in Geneious Prime 2019.2.1 (57); the 21 *Ct* reference genomes were used for mapping SRA reads using the built-in Geneious assembler at low sensitivity/fastest setting with up to 5 iterations. The 76 nearly complete and three unpublished *Ct* genomes were also imported into Geneious. Tryptophan operons consisting of *trp*R, intergenic region (IGR), *trp*B, and *trp*A, in that order, were extracted in Geneious after checking operon mapped reads that were at a minimum coverage depth of 10×. In addition, operons for the 30 clinical samples were sequenced using techniques and primers as previously described (26).

Alignment, phylogeny, recombination, and mutation analysis. The operons from 595 reference and clinical samples were concatenated and aligned using MAFFT v7.450 (58). The alignment for each sample was used to construct an approximate maximum likelihood phylogenetic tree using FastTree 2.1.11 (59) with a generalized time-reversible model. Both MAFFT and FastTree were built-in and executed in Geneious; 1,000-bootstrap sampling was used for statistical support of the tree. The tree

together with the metadata on *ompA* genotype and anatomic site of infection was used to generate the phylogenetic tree in Tree of Life v5.5 (60).

The MAFFT alignment of the operon sequences was also used to detect recombination events by RDP (61), GENECONV (62), BOOTSCAN (63), MAXCHI (64), CHIMAERA (65), SISCAN (66), PHYLPRO (67), LARD (68), and 3Seq (69). All software was built-in and executed in RDP4. The full exploratory scan was performed using each of the methods followed by a scan using the individual software. Additionally, a secondary recombination analysis was performed on putative recombinants identified from the RDP4 analysis to further tease out the breakpoint analysis of any recombination events noted. Briefly, nucleotide MAFFT alignments of the putative recombinants and their potential parents were submitted to Recco (70).

Operon sequences from the 21 reference strains (A_HAR13, B_HAR36, Ba_Apache2, C_TW3OT, D_UW3Cx, Da_TW448, E_Bour, F_IC-Cal13, G_UW57Cx, H_UW4Cx, I_UW12Ur, Ia_UW202, J_UW36Cx, Ja_UW92, K_UW31Cx, L_1_440, L_2_434, L_2a_UW396, L_2b_UCH1proctitis, L_2c, and L_3_404) were aligned using MAFFT to build a phylogenetic tree (Fig. S1). These data were used to determine *ompA* genotype and generate SNPs in the Geneious variant finder for the 595 genome sequences. Amino acid sequences with substitutions, premature stop codons, extensions, or frameshifts compared to respective reference sequences were designated mutants; their functionality was further studied by three-dimensional (3D) homology and functional protein modeling (see below).

DnaSP 6.0 (71) was used to analyze operon sequence polymorphisms to determine evidence for selection by calculating the ratio of nonsynonymous to synonymous substitutions [Pi(a)/Pi(s)] (Jukes-Cantor corrected), number of polymorphic (segregating) sites, and haplotype diversity.

3D protein structure homology modeling of *Ct* operon mutant and reference strains. TrpR, TrpB, and TrpA protein structure homology modeling was performed as previously described (26) with few modifications. Briefly, *Ct* mutant and reference operon protein sequences were uploaded to MODBASE, and MODELLER (31, 72) was used to calculate comparative 3D protein structures using PSI-BLAST (73), pairwise sequence alignment, sequence-sequence, and sequence-profile. PyMOL was used for structural visualization and imaging (The PyMOL Molecular Graphics System, version 1.2r3pre; Schrödinger, LLC). Predictions from MODELLER for both structure and functional residues were verified using the Simple Modular Architecture Research Tool (SMART) (74), SignalIP (75), Protein family (Pfam) (76), Conserved Domain (CD) (77), and Transmembrane helix prediction (TMHMM) (78). With a cutoff of 0.01 E value to perform protein BLAST against the Protein Data Bank (PDB) in addition to Phyre (Protein Homology/analogy Recognition Engine) (79), we searched for other homologs for structuring the tryptophan operon proteins.

Predicted models were used to manually identify and visualize nonsynonymous mutations, TrpR ligand-binding sites (LBSs), and DNA-binding motifs (DBM), TrpA α-Loop 6 (α-L6) and catalytic site, and TrpB beta communication (β-COMM) domain and monovalent cation (MVC)-binding sites, and indole tunnel based on both our prior research and crystal structures of respective bacterial strains (24, 30).

Molecular docking, binding site analysis, and effect of mutations on TrpR, TrpB and TrpA protein structure stability and ligand-binding affinity. Molecular docking and binding site analyses were performed using AutoDock (80) and PyMOL (81). AutoDock performs automated docking by simulated annealing, local gradient search, and genetic algorithm. Final atomic conformations were the result of minimizing the energy in a force field using multiple possible schemes. The combination of a genetic algorithm with inheritance of local optimizations yielded a Lamarckian genetic algorithm.

Ligand structural information was obtained from PubChem (82) where 3D structures were drawn in Chimera (83); geometry was optimized using the steepest descent method followed by conjugate gradient algorithms. Calculations with 50 intervals were used to ensure the highest accuracy of energy optimization. Since ligands are small molecules, the Gasteiger force field was assigned for the calculations.

Docking involved the prediction of ligand conformation and orientation within a targeted binding site. The Gibbs free energy correlation derived from molecular conformations was obtained using AutoDock. The free energy of binding (ΔG) is related to binding affinity and the equilibrium constant in the equation $\Delta G = -\mathrm{RT}\ \ln K_A$.

Modeling the *Ct* operon mutations was performed as previously described (84) with modifications. To predict the effect of mutations on protein-structure stability, ssDNA affinity, and ligand affinity binding, mCSM (http://biosig.unimelb.edu.au/mcsm/stability), mCSM-NA (http://biosig.unimelb.edu.au/mcsm_na/prediction), and mCSM-lig (http://biosig.unimelb.edu.au/mcsm_lig/) were used. PDB files of *Ct* reference protein templates and text files containing sets of mutations in mutant strains were provided as inputs to the respective program to estimate the free energy difference ($\Delta\Delta G$) between reference and mutant protein forms as a measure of thermodynamic stability and estimate the affinity change of small-molecule ligands for proteins as log-affinity fold change. The PyMOL plugin Arpeggio (http://biosig.unimelb.edu.au/arpeggioweb/) was used to study and visualize interatomic interactions.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.1 MB.

FIG S2, PDF file, 0.1 MB.

FIG S3, PDF file, 0.2 MB.

FIG S4, PDF file, 0.2 MB.

FIG S5, PDF file, 0.2 MB.

TABLE S1, PDF file, 0.1 MB.

TABLE S2, PDF file, 0.1 MB.
TABLE S3, PDF file, 0.1 MB.
TABLE S4, PDF file, 0.1 MB.
DATA SET S1, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We thank Dr. Elaine C. Meng (UCSF, Department of Pharmaceutical Chemistry) for her technical assistance with the MODELLER, Chimera, and Docking. We thank Dr. Noa Ziklo for initial discussions regarding the manuscript.

This work was supported in part by Public Health Service grants from the National Institutes of Health R01AI098843 and R01AI151075 (to D.D.).

REFERENCES

1. WHO. 2018. Report on global sexually transmitted infection surveillance. World Health Organization, Geneva, Switzerland.
2. Bennett JE, Dolin R, Blaser MJ. 2016. Douglas and Bennett's infectious diseases essentials. Elsevier Inc, Philadelphia, PA.
3. de Vrieze NHN, de Vries HJC. 2014. Lymphogranuloma venereum among men who have sex with men. An epidemiological and clinical review. Expert Rev Anti Infect Ther 12:697–704. https://doi.org/10.1586/14787210.2014.901169.
4. Harris SR, Clarke IN, Seth-Smith HMB, Solomon AW, Cutcliffe LT, Marsh P, Skilton RJ, Holland MJ, Mabey D, Peeling RW, Lewis DA, Spratt BG, Unemo M, Persson K, Bjartling C, Brunham R, de Vries HJC, Morré SA, Speksnijder A, Bébéar CM, Clerc M, de Barbeyrac B, Parkhill J, Thomson NR. 2012. Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. Nat Genet 44:413–419. https://doi.org/10.1038/ng.2214.
5. WHO. 2019. Trachoma epidemiologic situation. World Health Organization, Geneva, Switzerland.
6. Dean D. 2013. Chlamydia trachomatis pathogenicity and disease, p 25–60. *In* Black CM (ed), Chlamydial infection: a clinical and public health perspective. Issues in infectious diseases, vol 7. Karger, Basel, Switzerland.
7. Millman K, Black CM, Johnson RE, Stamm WE, Jones RB, Hook EW, Martin DH, Bolan G, Tavaré S, Dean D. 2004. Population-based genetic and evolutionary analysis of *Chlamydia trachomatis* urogenital strain variation in the United States. J Bacteriol 186:2457–2465. https://doi.org/10.1128/JB.186.8.2457-2465.2004.
8. Gomes JP, Bruno WJ, Nunes A, Santos N, Florindo C, Borrego MJ, Dean D. 2007. Evolution of *Chlamydia trachomatis* diversity occurs by widespread interstrain recombination involving hotspots. Genome Res 17:50–60. https://doi.org/10.1101/gr.5674706.
9. Herrmann B, Isaksson J, Ryberg M, Tångrot J, Saleh I, Versteeg B, Gravningen K, Bruisten S. 2015. Global multilocus sequence type analysis of *Chlamydia trachomatis* strains from 16 countries. J Clin Microbiol 53:2172–2179. https://doi.org/10.1128/JCM.00249-15.
10. Joseph SJ, Didelot X, Gandhi K, Dean D, Read TD. 2011. Interplay of recombination and selection in the genomes of *Chlamydia trachomatis*. Biol Direct 6:28. https://doi.org/10.1186/1745-6150-6-28.
11. Joseph SJ, Didelot X, Rothschild J, de Vries HJC, Morré SA, Read TD, Dean D. 2012. Population genomics of *Chlamydia trachomatis*: insights on drift, selection, recombination, and population structure. Mol Biol Evol 29:3933–3946. https://doi.org/10.1093/molbev/mss198.
12. Klint M, Fuxelius H-H, Goldkuhl RR, Skarin H, Rutemark C, Andersson SGE, Persson K, Herrmann B. 2007. High-resolution genotyping of *Chlamydia trachomatis* strains by multilocus sequence analysis. J Clin Microbiol 45:1410–1414. https://doi.org/10.1128/JCM.02301-06.
13. Smelov V, Vrbanac A, van Ess EF, Noz MP, Wan R, Eklund C, Morgan T, Shrier LA, Sanders B, Dillner J, de Vries HJC, Morre SA, Dean D. 2017. *Chlamydia trachomatis* strain types have diversified regionally and globally with evidence for recombination across geographic divides. Front Microbiol 8:2195. https://doi.org/10.3389/fmicb.2017.02195.
14. Somboonna N, Wan R, Ojcius DM, Pettengill MA, Joseph SJ, Chang A, Hsu R, Read TD, Dean D. 2011. Hypervirulent *Chlamydia trachomatis* clinical strain is a recombinant between lymphogranuloma venereum (L2) and D lineages. mBio 2:e00045-11. https://doi.org/10.1128/mBio.00045-11.
15. Dean D, Bruno WJ, Wan R, Gomes JP, Devignot S, Mehari T, de Vries HJC, Morré SA, Myers G, Read TD, Spratt BG. 2009. Predicting phenotype and emerging strains among *Chlamydia trachomatis* infections. Emerg Infect Dis 15:1385–1394. https://doi.org/10.3201/eid1509.090272.
16. Hadfield J, Harris SR, Seth-Smith HMB, Parmar S, Andersson P, Giffard PM, Schachter J, Moncada J, Ellison L, Vaulet MLG, Fermepin MR, Radebe F, Mendoza S, Ouburg S, Morré SA, Sachse K, Puolakkainen M, Korhonen SJ, Sonnex C, Wiggins R, Jalal H, Brunelli T, Casprini P, Pitt R, Ison C, Savicheva A, Shipitsyna E, Hadad R, Kari L, Burton MJ, Mabey D, Solomon AW, Lewis D, Marsh P, Unemo M, Clarke IN, Parkhill J, Thomson NR. 2017. Comprehensive global genome dynamics of *Chlamydia trachomatis* show ancient diversification followed by contemporary mixing and recent lineage expansion. Genome Res 27:1220–1229. https://doi.org/10.1101/gr.212647.116.
17. Millman KL, Tavaré S, Dean D. 2001. Recombination in the ompA gene but not the omcB gene of *Chlamydia* contributes to serovar-specific differences in tissue tropism, immune surveillance, and persistence of the organism. J Bacteriol 183:5997–6008. https://doi.org/10.1128/JB.183.20.5997-6008.2001.
18. Andersson P, Harris SR, Seth Smith HMB, Hadfield J, O'Neill C, Cutcliffe LT, Douglas FP, Asche LV, Mathews JD, Hutton SI, Sarovich DS, Tong SYC, Clarke IN, Thomson NR, Giffard PM. 2016. *Chlamydia trachomatis* from Australian Aboriginal people with trachoma are polyphyletic composed of multiple distinctive lineages. Nat Commun 7:10688. https://doi.org/10.1038/ncomms10688.
19. Abdelsamed H, Peters J, Byrne GI. 2013. Genetic variation in *Chlamydia trachomatis* and their hosts: impact on disease severity and tissue tropism. Future Microbiol 8:1129–1146. https://doi.org/10.2217/fmb.13.80.
20. Steele S, Brunton J, Kawula T. 2015. The role of autophagy in intracellular pathogen nutrient acquisition. Front Cell Infect Microbiol 5:51. https://doi.org/10.3389/fcimb.2015.00051.
21. Bastidas RJ, Elwell CA, Engel JN, Valdivia RH. 2013. Chlamydial intracellular survival strategies. Cold Spring Harb Perspect Med 3:a010256. https://doi.org/10.1101/cshperspect.a010256.
22. Ziklo N, Huston WM, Hocking JS, Timms P. 2016. *Chlamydia trachomatis* genital tract infections: when host immune response and the microbiome collide. Trends Microbiol 24:750–765. https://doi.org/10.1016/j.tim.2016.05.007.
23. Merkl R. 2007. Modelling the evolution of the archeal tryptophan synthase. BMC Evol Biol 7:59. https://doi.org/10.1186/1471-2148-7-59.
24. Busch F, Rajendran C, Heyn K, Schlee S, Merkl R, Sterner R. 2016. Ancestral tryptophan synthase reveals functional sophistication of primordial enzyme complexes. Cell Chem Biol 23:709–715. https://doi.org/10.1016/j.chembiol.2016.05.009.
25. Caldwell HD, Wood H, Crane D, Bailey R, Jones RB, Mabey D, Maclean I, Mohammed Z, Peeling R, Roshick C, Schachter J, Solomon AW, Stamm WE, Suchland RJ, Taylor L, West SK, Quinn TC, Belland RJ, McClarty G. 2003. Polymorphisms in *Chlamydia trachomatis* tryptophan synthase genes differentiate between genital and ocular isolates. J Clin Invest 111:1757–1769. https://doi.org/10.1172/JCI17993.
26. Somboonna N, Ziklo N, Ferrin TE, Suh JH, Dean D. 2019. Clinical persistence of *Chlamydia trachomatis* sexually transmitted strains involves novel mutations in the functional $\alpha\beta\beta\alpha$ tetramer of the tryptophan synthase operon. mBio 10:e01464-19. https://doi.org/10.1128/mBio.01464-19.
27. Alkhidir AAI, Holland MJ, Elhag WI, Williams CA, Breuer J, Elemam AE, El Hussain KMK, Ournasseir MEH, Pickering H. 2019. Whole-genome

sequencing of ocular *Chlamydia trachomatis* isolates from Gadarif State, Sudan. Parasites Vectors 12:518. https://doi.org/10.1186/s13071-019-3770-7.
28. Feng L, Lu X, Yu Y, Wang T, Luo S, Sun Z, Duan Q, Wang N, Song L. 2016. Survey, culture, and genome analysis of ocular *Chlamydia trachomatis* in Tibetan boarding primary schools in Qinghai Province. Front Cell Infect Microbiol 6:207. https://doi.org/10.3389/fcimb.2016.00207.
29. Pokorzynski ND, Brinkworth AJ, Carabeo R. 2019. A bipartite iron-dependent transcriptional regulation of the tryptophan salvage pathway in *Chlamydia trachomatis*. Elife 8:e42295. https://doi.org/10.7554/eLife.42295.
30. Komeiji Y, Uebayasi M, Yamato I. 1994. Molecular dynamics simulations of *trp* apo-and holorepressors: domain structure and ligand–protein interaction. Proteins 20:248–258. https://doi.org/10.1002/prot.340200305.
31. Pieper U, Eswar N, Braberg H, Madhusudhan MS, Davis FP, Stuart AC, Mirkovic N, Rossi A, Marti-Renom MA, Fiser A, Webb B, Greenblatt D, Huang CC, Ferrin TE, Sali A. 2004. MODBASE, a database of annotated comparative protein structure models, and associated resources. Nucleic Acids Res 32:D217–D222. https://doi.org/10.1093/nar/gkh095.
32. Melo F, Sánchez R, Sali A. 2002. Statistical potentials for fold assessment. Protein Sci 11:430–448. https://doi.org/10.1002/pro.110430.
33. Ngo H, Kimmich N, Harris R, Niks D, Blumenstein L, Kulik V, Barends TR, Schlichting I, Dunn MF. 2007. Allosteric regulation of substrate channeling in tryptophan synthase: modulation of the L-serine reaction in stage I of the β-reaction by α-site ligands. Biochemistry 46:7740–7753. https://doi.org/10.1021/bi7003872.
34. Collingro A, Tischler P, Weinmaier T, Penz T, Heinz E, Brunham RC, Read TD, Bavoil PM, Sachse K, Kahane S, Friedman MG, Rattei T, Myers GSA, Horn M. 2011. Unity in variety—the pan-genome of the Chlamydiae. Mol Biol Evol 28:3253–3270. https://doi.org/10.1093/molbev/msr161.
35. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K, Bass S, Linher K, Weidman J, Khouri H, Craven B, Bowman C, Dodson R, Gwinn M, Nelson W, DeBoy R, Kolonay J, McClarty G, Salzberg SL, Eisen J, Fraser CM. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res 28:1397–1406. https://doi.org/10.1093/nar/28.6.1397.
36. Taylor LD, Nelson DE, Dorward DW, Whitmire WM, Caldwell HD. 2010. Biological characterization of *Chlamydia trachomatis* plasticity zone MACPF domain family protein CT153. Infect Immun 78:2691–2699. https://doi.org/10.1128/IAI.01455-09.
37. Bonner CA, Byrne GI, Jensen RA. 2014. *Chlamydia* exploit the mammalian tryptophan-depletion defense strategy as a counter-defensive cue to trigger a survival state of persistence. Front Cell Infect Microbiol 4:17. https://doi.org/10.3389/fcimb.2014.00017.
38. Sherchand SP, Aiyar A. 2019. Ammonia generation by tryptophan synthase drives a key genetic difference between genital and ocular *Chlamydia trachomatis* isolates. Proc Natl Acad Sci U S A 116:12468–12477. https://doi.org/10.1073/pnas.1821652116.
39. Dong Q, Brulc JM, Iovieno A, Bates B, Garoutte A, Miller D, Revanna KV, Gao X, Antonopoulos DA, Slepak VZ, Shestopalov VI. 2011. Diversity of bacteria at healthy human conjunctiva. Invest Ophthalmol Vis Sci 52:5408–5413. https://doi.org/10.1167/iovs.10-6939.
40. Willcox MDP. 2013. Characterization of the normal microbiota of the ocular surface. Exp Eye Res 117:99–105. https://doi.org/10.1016/j.exer.2013.06.003.
41. Seth-Smith HMB, Harris SR, Skilton RJ, Radebe FM, Golparian D, Shipitsyna E, Duy PT, Scott P, Cutcliffe LT, O'Neill C, Parmar S, Pitt R, Baker S, Ison CA, Marsh P, Jalal H, Lewis DA, Unemo M, Clarke IN, Parkhill J, Thomson NR. 2013. Whole-genome sequences of *Chlamydia trachomatis* directly from clinical samples without culture. Genome Res 23:855–866. https://doi.org/10.1101/gr.150037.112.
42. Fehlner-Gardiner C, Roshick C, Carlson JH, Hughes S, Belland RJ, Caldwell HD, McClarty G. 2002. Molecular basis defining human *Chlamydia trachomatis* tissue tropism. A possible role for tryptophan synthase. J Biol Chem 277:26893–26903. https://doi.org/10.1074/jbc.M203937200.
43. Christerson L, Bom RJM, Bruisten SM, Yass R, Hardick J, Bratt G, Gaydos CA, Morré SA, Herrmann B. 2012. *Chlamydia trachomatis* strains show specific clustering for men who have sex with men compared to heterosexual populations in Sweden, the Netherlands, and the United States. J Clin Microbiol 50:3548–3555. https://doi.org/10.1128/JCM.01713-12.
44. Jeffrey BM, Suchland RJ, Quinn KL, Davidson JR, Stamm WE, Rockey DD. 2010. Genome sequencing of recent clinical *Chlamydia trachomatis* strains identifies loci associated with tissue tropism and regions of apparent recombination. Infect Immun 78:2544–2553. https://doi.org/10.1128/IAI.01324-09.
45. Unemo M, Seth-Smith HMB, Cutcliffe LT, Skilton RJ, Barlow D, Goulding D, Persson K, Harris SR, Kelly A, Bjartling C, Fredlund H, Olcén P, Thomson NR, Clarke IN. 2010. The Swedish new variant of *Chlamydia trachomatis*: genome sequence, morphology, cell tropism and phenotypic characterization. Microbiology (Reading) 156:1394–1404. https://doi.org/10.1099/mic.0.036830-0.
46. Bush RM. 2001. Predicting adaptive evolution. Nat Rev Genet 2:387–392. https://doi.org/10.1038/35072023.
47. Andersson DI, Hughes D. 1996. Muller's ratchet decreases fitness of a DNA-based microbe. Proc Natl Acad Sci U S A 93:906–907. https://doi.org/10.1073/pnas.93.2.906.
48. Akers JC, Tan M. 2006. Molecular mechanism of tryptophan-dependent transcriptional regulation in *Chlamydia trachomatis*. J Bacteriol 188:4236–4243. https://doi.org/10.1128/JB.01660-05.
49. Marmorstein RQ, Sigler PB. 1989. Stereochemical effects of L-tryptophan and its analogues on trp repressor's affinity for operator-DNA. J Biol Chem 264:9149–9154. https://doi.org/10.1016/S0021-9258(18)60507-1.
50. Lawson CL, Sigler PB. 1988. The structure of trp pseudorepressor at 1.65Å shows why indole propionate acts as a trp 'inducer'. Nature 333:869–871. https://doi.org/10.1038/333869a0.
51. Negatu DA, Liu JJJ, Zimmerman M, Kaya F, Dartois V, Aldrich CC, Gengenbacher M, Dick T. 2018. Whole-cell screen of fragment library identifies gut microbiota metabolite indole propionic acid as antitubercular. Antimicrob Agents Chemother 62:e01571-17. https://doi.org/10.1128/AAC.01571-17.
52. Grossowicz N. 1990. Phytohormones as specific inhibitors of *Legionella pneumophila* growth. Isr J Med Sci 26:187–190.
53. Otwinowski Z, Schevitz RW, Zhang R-G, Lawson CL, Joachimiak A, Marmorstein RQ, Luisi BF, Sigler PB. 1988. Crystal structure of trp repressor/operator complex at atomic resolution. Nature 335:321–329. https://doi.org/10.1038/335321a0.
54. Peracchi A, Mozzarelli A, Rossi GL. 1995. Monovalent cations affect dynamic and functional properties of the tryptophan synthase $\alpha 2\beta 2$ complex. Biochemistry 34:9459–9465. https://doi.org/10.1021/bi00029a022.
55. Rank RG, Yeruva L. 2014. Hidden in plain sight: chlamydial gastrointestinal infection and its relevance to persistence in human genital infection. Infect Immun 82:1362–1371. https://doi.org/10.1128/IAI.01244-13.
56. Barton MD, Delneri D, Oliver SG, Rattray M, Bergman CM. 2010. Evolutionary systems biology of amino acid biosynthetic cost in yeast. PLoS One 5:e11935. https://doi.org/10.1371/journal.pone.0011935.
57. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199.
58. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
59. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
60. Letunic I, Bork P. 2019. Interactive tree of life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. https://doi.org/10.1093/nar/gkz239.
61. Martin D, Rybicki E. 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16:562–563. https://doi.org/10.1093/bioinformatics/16.6.562.
62. Padidam M, Sawyer S, Fauquet CM. 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265:218–225. https://doi.org/10.1006/viro.1999.0056.
63. Martin DP, Posada D, Crandall KA, Williamson C. 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res Hum Retroviruses 21:98–102. https://doi.org/10.1089/aid.2005.21.98.
64. Smith JM. 1992. Analyzing the mosaic structure of genes. J Mol Evol 34:126–129. https://doi.org/10.1007/BF00182389.
65. Posada D, Crandall KA. 2001. Selecting the best-fit model of nucleotide substitution. Syst Biol 50:580–601. https://doi.org/10.1080/106351501750435121.
66. Gibbs MJ, Armstrong JS, Gibbs AJ. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16:573–582. https://doi.org/10.1093/bioinformatics/16.7.573.

67. Weiller GF. 1998. Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. Mol Biol Evol 15:326–335. https://doi.org/10.1093/oxfordjournals.molbev.a025929.
68. Holmes EC, Worobey M, Rambaut A. 1999. Phylogenetic evidence for recombination in dengue virus. Mol Biol Evol 16:405–409. https://doi.org/10.1093/oxfordjournals.molbev.a026121.
69. Boni MF, Posada D, Feldman MW. 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176:1035–1047. https://doi.org/10.1534/genetics.106.068874.
70. Maydt J, Lengauer T. 2006. Recco: recombination analysis using cost optimization. Bioinformatics 22:1064–1071. https://doi.org/10.1093/bioinformatics/btl057.
71. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302. https://doi.org/10.1093/molbev/msx248.
72. Šali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815. https://doi.org/10.1006/jmbi.1993.1626.
73. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.
74. Letunic I, Bork P. 2018. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res 46:D493–D496. https://doi.org/10.1093/nar/gkx922.
75. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786. https://doi.org/10.1038/nmeth.1701.
76. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432. https://doi.org/10.1093/nar/gky995.
77. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45:D200–D203. https://doi.org/10.1093/nar/gkw1129.
78. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315.
79. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858. https://doi.org/10.1038/nprot.2015.053.
80. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. 2009. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30:2785–2791. https://doi.org/10.1002/jcc.21256.
81. Schrodinger LLC. 2015. The PyMOL molecular graphics system, version 1.8.
82. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE. 2019. PubChem 2019 update: improved access to chemical data. Nucleic Acids Res 47:D1102–D1109. https://doi.org/10.1093/nar/gky1033.
83. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. 2006. Tools for integrated sequence-structure analysis with UCSF Chimera. BMC Bioinformatics 7:339.
84. Munir A, Kumar N, Ramalingam SB, Tamilzhalagan S, Shanmugam SK, Palaniappan AN, Nair D, Priyadarshini P, Natarajan M, Tripathy S, Ranganathan UD, Peacock SJ, Parkhill J, Blundell TL, Malhotra S. 2019. Identification and characterization of genetic determinants of isoniazid and rifampicin resistance in *Mycobacterium tuberculosis* in southern India. Sci Rep 9:10283. https://doi.org/10.1038/s41598-019-46756-x.

Supplementary Information

Tryptophan operon diversity reveals evolutionary trends among geographically disparate *Chlamydia trachomatis* ocular and urogenital strains affecting tryptophan repressor and synthase function

Sankhya Bommana[a], Naraporn Somboonna[b], Gracie Richards[a], Maryam Tarazkar[a], Deborah Dean[a, b, c, d*]

[a]Department of Pediatrics, University of California San Francisco, Oakland, CA, USA

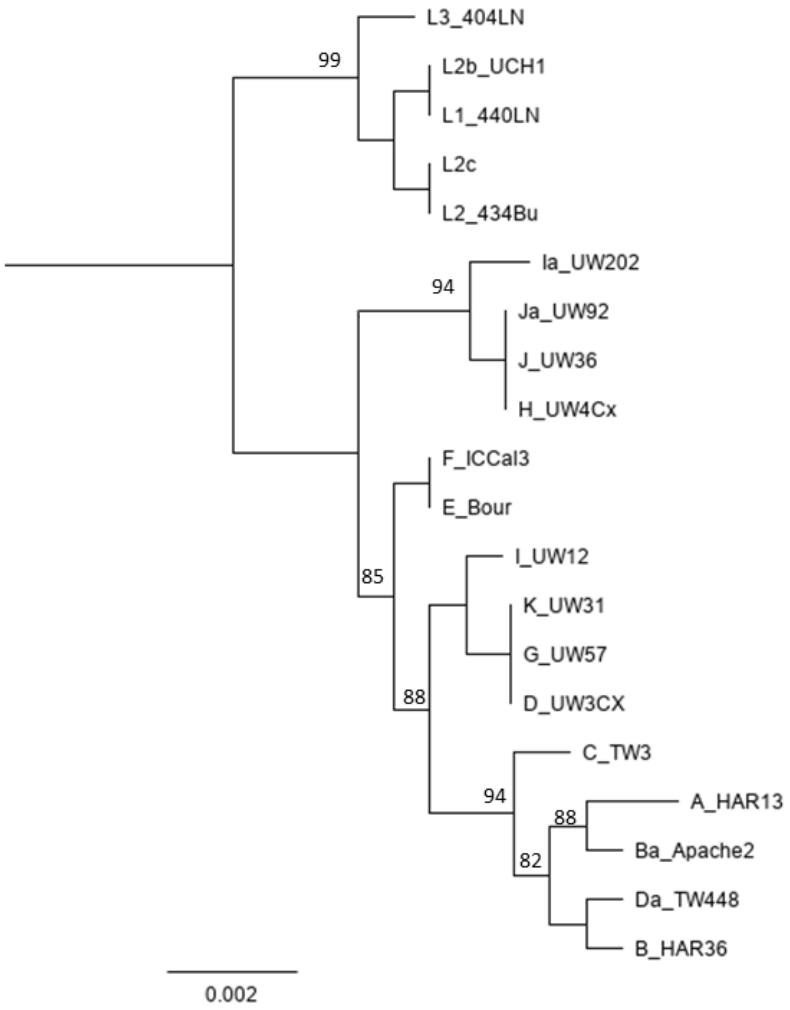
[b]Department of Bioengineering, Joint Graduate Program, University of California San Francisco and University of California Berkeley, San Francisco, CA, USA

[c]Department of Medicine, University of California San Francisco, San Francisco, CA, USA

[d]Bixby Center for Global Reproductive Health, University of California San Francisco, San Francisco, CA, USA

Table of contents

Supplementary Figures



Supplementary Fig 1. Concatenated *trp*RBA phylogenetic tree of 21 *Ct* reference strains constructed by FastTree with a Generalized Time-Reversible model based on a MAFFT alignment of *trp*RBA sequences (see Methods). Both were executed in Geneious (https://www.geneious.com). The tree scale indicates the distance between the sequences and the branch length indicates the number of substitutions that have occurred in that branch.

Chain A

mAQQSPYSAAMAEQRHQEWLRFVDLLKNAYQNDLHLPLLNLMLLPDEREALGTRVRIVEELLRGEMSQRELKNELGAGIATITRGSNYLKAAPVELRQWLEEVLLksdlehhhhhh

IAC2 R IV T RG Y

IAC1 LMLL

DBS1 QRELKNELGAGIATITRGSNYLKA

EDO PVE R

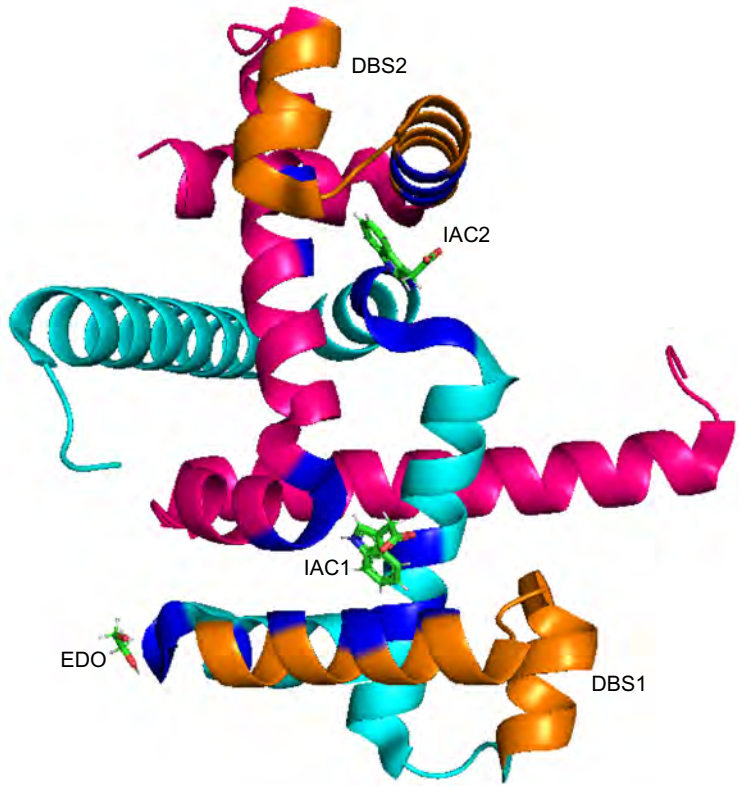
Chain B

mAQQSPYSAAMAEQRHQEWLRFVDLLKNAYQNDLHLPLLNLMLLPDEREALGTRVRIVEELLRGEMSQRELKNELGAGIATITRGSNYLKAAPVELRQWLEEVLLksdlehhhhhh

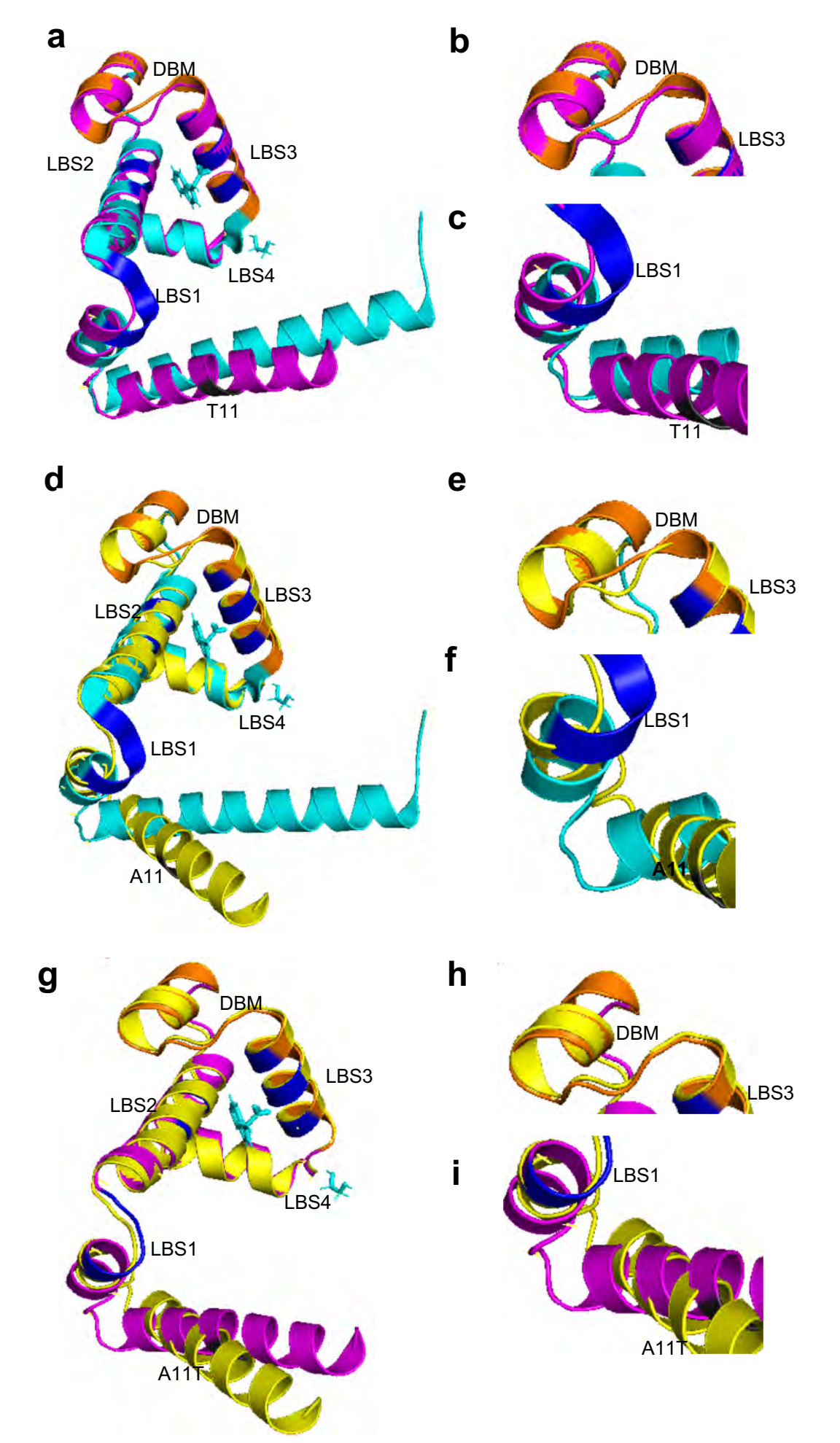
IAC1 R IV T RG Y

IAC2 LMLL

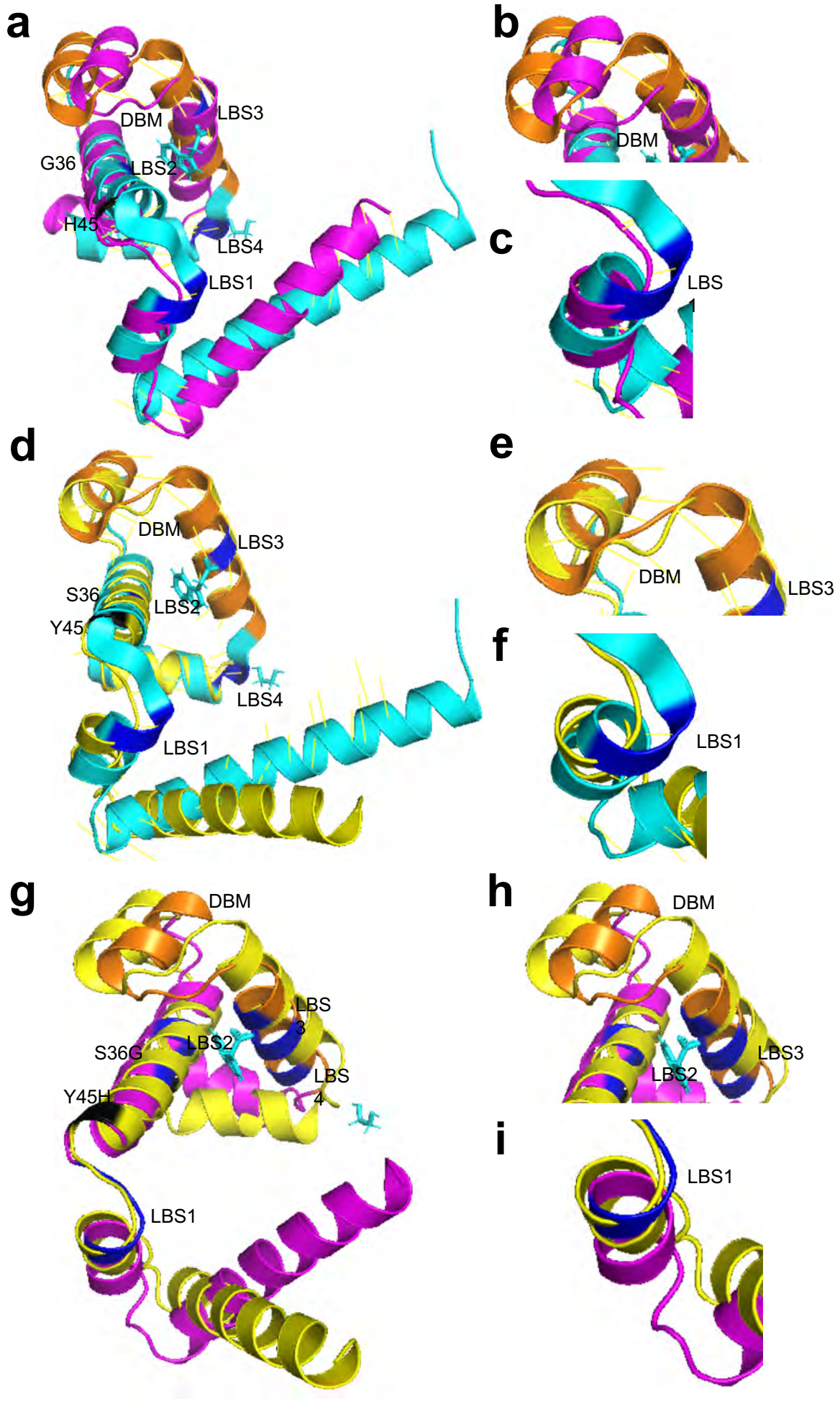
DBS2 QRELKNELGAGIATITRGSNYLKA



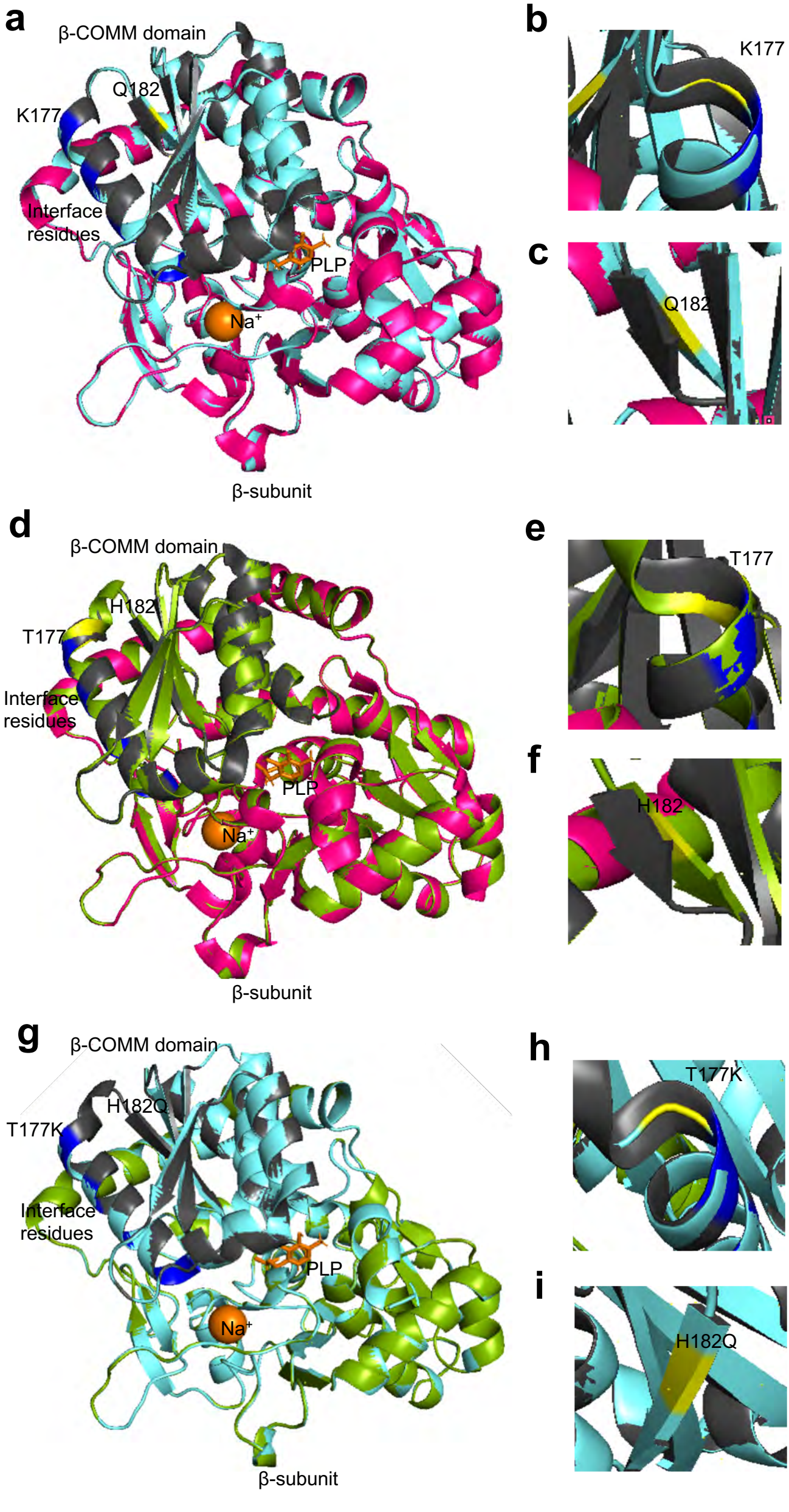
Supplementary Fig 2. Crystal structure of tryptophan repressor in *E.coli*, PDB: 6ENI. The two monomer subunits of the symmetric dimer are shown as cartoon diagrams, chain A in cyan and chain B in pink. Two molecules of indol-3-acetic acid (IAC) is bound at chain A and chain B, the 1,2-ethanediol (EDO) is bound at chain A only. The DNA binding motif (DBS) involved in the protein-DNA (operator binding) interactions is shown in orange and is present in chain A and B. The putative ligand binding sites for IAC and EDO are shown in blue. The reference sequence of *E.coli* TrpR 6eni with the LBS (blue), DBS (orange) and residues involved in dimerization (bold) are shown as well. Each molecule of the ligand IAC has LBS on chain A and chain B, respectively.


a
DBM
LBS2
LBS3
LBS4
LBS1
T11
b
DBM
LBS3
c
LBS1
T11
d
DBM
LBS3
LBS2
LBS4
LBS1
A11
e
DBM
LBS3
f
LBS1
A11
g
DBM
LBS3
LBS2
LBS4
LBS1
A11T
h
DBM
LBS3
i
LBS1
A11T

Supplementary Fig 3. TrpR 3D predicted structures of *Ct* clinical F/SwabB8 and reference F/IC-Cal-3 strains (A-I). (A) TrpR structure of *Ct* mutant strain F/SwabB8 (magenta) superimposed on the template 6eniA (cyan), with DBM in orange. The LBS 1, 2, 3 and 4 are shown in dark blue, and the aa substitution in relation to F/IC-Cal-3 at T11 is in black. (B and C) Structural changes in DBM and LBS1, respectively, of the mutant are shown. (D) TrpR structure of F/IC-Cal-3 (yellow) superimposed on the template 6eniA (cyan) with annotations as per A, B and C. (E and F) Structural changes in DBM and LBS1 of E/Bour. (G) TrpR 3D predicted structures of F/SwabB8 superimposed on F/IC-Cal-3 with structural changes noted in relation to (H) DBM based and (I) LBS1.


a
DBM
LBS3
G36
LBS2
H45
LBS4
LBS1
b
DBM
c
LBS
1
d
DBM
LBS3
S36
LBS2
Y45
LBS4
LBS1
e
DBM
LBS3
f
LBS1
g
DBM
LBS
3
S36G
LBS2
LBS
4
Y45H
LBS1
h
DBM
LBS3
LBS2
i
LBS1

Supplementary Fig 4. TrpR 3D predicted structures of *Ct* clinical J/SF5 and reference J/UW-36 strains (A-I). (A) TrpR structure of *Ct* mutant strain J/SF5 (magenta) superimposed on the template 6eniA (cyan), with DBM in orange. The LBS 1, 2, 3 and 4 are shown in dark blue, and the aa substitution in relation to J/UW-36 at G36 and H45 is in black. (B and C) Structural changes in DBM and LBS1, respectively, of the mutant are shown. (D) TrpR structure of J/UW-36 (yellow) superimposed on the template 6eniA (cyan) with annotations as per A, B and C. (E and F) Structural changes in DBM and LBS1 of J/UW-36. (G) TrpR 3D predicted structures of J/SF5 superimposed on J/UW-36 with structural changes noted in relation to J/UW-36 at LBS1 and DBM based on S36G and Y45H aa substitution (H and I).


a
β-COMM domain
Q182
K177
Interface residues
PLP
Na+
β-subunit
b
K177
c
Q182
d
β-COMM domain
H182
T177
Interface residues
PLP
Na+
β-subunit
e
T177
f
H182
g
β-COMM domain
H182Q
T177K
Interface residues
PLP
Na+
β-subunit
h
T177K
i
H182Q

Supplementary Fig 5. TrpB 3D predicted structures of *Ct* clinical I_UK913341 and reference I_UW12 strains (A-I). (A) TrpB structure of mutant strain I_UK913341 (cyan) superimposed on the template 5ey5 (magenta), with the COMM domain in charcoal grey. TrpB interface residues that interact with TrpA are shown in blue and the aa substitution in relation to E_Bour at K177 and Q182 are in yellow. (B and C) Structural changes in β subunit interface and COMM domain residues of the mutant E_SotonE8 are shown. (D) TrpB structure of I_UW12 (green) superimposed on the template 5ey5 (magenta) with annotations as per A, B and C. (E and F) Structural changes in β subunit interface and COMM domain residues of the reference I_UW12 are shown. (G) TrpB predicted structures of I_UK913341 superimposed on I_UW12 with structural changes noted in relation to (H) β subunit interface residues on P381L and (I) COMM domain on P304S.

Supplementary Tables

Supplementary Table 1. Primary recombination analysis in the tryptophan operon for global *C. trachomatis* strains (n=595) using RDP4.

Event#	Daughter	Major parent	Minor parent	Beginning	Ending	Uncorrected	MC corrected	Identified by (Av. P-Val)	Recombinant score
1	**B9**	AHAR13	L3404LN	1090	1450	$6.163X10^{-10}$	$7.396X10^{-8}$	GENECONV ($1.123X10^{-2}$)	0.549 (B9)
								Bootscan ($1.124X10^{-3}$)	0.196 (AHAR13)
								Maxchi ($2.381X10^{-4}$)	0.255 (L3404LN)
								3Seq ($3.223X10^{-5}$)	
2	**A8**	AHAR13	L3404LN	971	1555	$8.421X10^{-6}$	$1.011X10^{-3}$	GENECONV ($1.123X10^{-2}$)	0.549 (B9)
								Bootscan ($1.124X10^{-3}$)	0.196 (AHAR13)
								Maxchi ($2.381X10^{-4}$)	0.255 (L3404LN)
								3Seq ($3.223X10^{-5}$)	
		AHAR13	B9	971	1555	$8.431X10^{-6}$	$1.011X10^{-3}$	MaxChi ($2.629X10^{-3}$)	0.549 (B9)
								3Seq ($1.360X10^{-2}$)	0.196 (AHAR13)
									0.255 (L3404LN)
	A15	AHAR13 (99.8%)	L3404LN	1012	1438	$5.575X10^{-8}$	$6.908X10^{-6}$	GENECONV ($1.123X10^{-2}$)	0.549 (B9)
								Bootscan ($1.124X10^{-3}$)	0.196 (AHAR13)
								Maxchi ($2.381X10^{-4}$)	0.255 (L3404LN)
								3Seq ($3.223X10^{-5}$)	
		AHAR13 (99.8%)	B9 (92.9%)	1090	1206	$5.604X10^{-5}$	$6.725X10^{-3}$	MaxChi ($2.629X10^{-3}$)	0.446 (A15)
								3Seq ($1.360X10^{-2}$)	0.215 (AHAR13)
									0.339 (B9)
	S45	AHAR13	B9 (99.6%)	1339	1206	$1.567X10^{-4}$	$1.881X10^{-2}$	MaxChi ($2.629X10^{-3}$)	0.446 (S45)
								3Seq ($1.360\ X10^{-2}$)	0.215 (AHAR13)
									0.339 (B9)
3	**BQH111L**	DFin163 (99.3%)	EUK769748	332	1912	$1.182X10^{-4}$	$1.418X10^{-2}$	3Seq ($1.418X10^{-2}$)	0.452 (BQH111L)
									0.305 (DFin163)
									0.242 (EUK769748)
4	**IUK913341**	L3404LN (99.3%)	EUK769748	433	2486	$4.014X10^{-4}$	$4.817X10^{-2}$	3Seq ($7.766X10^{-1}$)	0.537 (IUK913341)
									0.253 (L3404LN)
									0.210 (EUK769748)
	IaSF16	L3404LN (98.9%)	EUK769748 (99.8%)	433	2486	$7.87\ X\ 10^{-3}$	0.944	3Seq ($7.766X10^{-1}$)	0.537 (IaSF16)
									0.253 (L3404LN)
									0.210 (EUK769748)
	IaSF27	L3404LN (98.9%)	EUK769748 (99.8%)	433	2486	$7.87\ X\ 10^{-3}$	0.944	3Seq ($7.766X10^{-1}$)	0.537 (IaSF27)
									0.253 (L3404LN)
									0.210 (EUK769748)
	INL58	L3404LN (98.9%)	EUK769748 (99.8%)	433	2486	$6.597\ X\ 10^{-3}$	0.791	3Seq ($7.766X10^{-1}$)	0.537 (INL58)
									0.253 (L3404LN)
									0.210 (EUK769748)
	INL63	L3404LN (98.9%)	EUK769748 (99.8%)	433	2486	$6.597\ X\ 10^{-3}$	0.791	3Seq ($7.766X10^{-1}$)	0.537 (INL63)
									0.253 (L3404LN)
									0.210 (EUK769748)
	INL69	L3404LN (98.9%)	EUK769748 (99.8%)	433	2486	$6.597\ X\ 10^{-3}$	0.791	3Seq ($7.766X10^{-1}$)	0.537 (INL69)
									0.253 (L3404LN)
									0.210 (EUK769748)
	INL70	L3404LN (98.9%)	EUK769748 (99.8%)	433	2486	$6.597\ X\ 10^{-3}$	0.791	3Seq ($7.766X10^{-1}$)	0.537 (INL70)
									0.253 (L3404LN)
									0.210 (EUK769748)
	IS2459	L3404LN (98.9%)	EUK769748 (99.8%)	433	2486	$6.597\ X\ 10^{-3}$	0.791	3Seq ($7.766X10^{-1}$)	0.537 (IS2459)
									0.253 (L3404LN)
									0.210 (EUK769748)
	Ia20-97	L3404LN (98.9%)	EUK769748 (99.8%)	433	2486	$6.597\ X\ 10^{-3}$	0.791	3Seq ($7.766X10^{-1}$)	0.537 (Ia20-97)
									0.253 (L3404LN)
									0.210 (EUK769748)
	IaSoton1	L3404LN (98.9%)	EUK769748 (99.8%)	433	2486	$6.597\ X\ 10^{-3}$	0.791	3Seq ($7.766X10^{-1}$)	0.537 (IaSoton1)
									0.253 (L3404LN)
									0.210 (EUK769748)
	IaSoton3	L3404LN (98.9%)	EUK769748 (99.8%)	433	2486	$6.597\ X\ 10^{-3}$	0.791	3Seq ($7.766X10^{-1}$)	0.537 (IaSoton2)
									0.253 (L3404LN)
									0.210 (EUK769748)
	IaUW202	L3404LN (98.9%)	EUK769748 (99.8%)	433	2486	$8.33\ X\ 10^{-3}$	1	3Seq ($7.766X10^{-1}$)	0.537 (IaUW202)
									0.253 (L3404LN)
									0.210 (EUK769748)

Note: Tryptophan operon was extracted from individual strains, and recombination analysis was performed in RDP4 using the MAFFT alignment of 595 strains (see Methods).

Supplementary Table 2. Secondary analysis of putative *C. trachomatis* recombinants involving the tryptophan operon identified by RDP4 using Recco.

Sequence name	Start*	End*	Savings^	Aln pv#	Seq pv~	overall p-value	Mutation Cost
B_QH111L	1935	2291	2	1	0.23077	1	12
B_QH111L	107	245	1	1	0.93507		11
I_UK913341	246	314	2	1	0.2997	1	10
I_UK913341	332	531	2	1	0.2997		10
I_UK913341	2472	2473	0.4	1	1		9.2

*Range of nucleotide positions where a recombination breakpoint could occur.
^Amount of mutation cost saved by introducing this recombination.
#p-value of this recombination event regarding savings distribution over the entire alignment - a very conservative measure for recombination.
~p-value of this recombination event regarding savings distribution over recombinant.

Supplementary Table 3. Single nucleotide polymorphism in *trpB* for *C. trachomatis* clinical and reference ocular strains.

Lineage	Site of infection	Unique *trpB* strain variants	Year of isolation	Geography	Number of variants	Variant ID	Year of isolation	Geography	Changes to *trpB*	Length of *trpB* (bp)
Ocular	Ocular	**A_HAR13**	1958	Egypt	7	A_D213, A_D230, A_SA1,	1957, 2001	Gambia, Saudi A	None	1179
Ocular	Ocular	A_2497	2000	Tanzania	47	A_363, A_5291, A_7249, A_MH858, A_MH1364, A_MH2145, A_MH3234, A_MH2497, A_MH4510, A_MH5368, A_MH5786, A_MH6446, A_MH7205, A_MH8910, A_MH9922, A_MH10549, A_MH10648, A_MH10901, A_MH11715, A_MH11979, A_MH12023, A_MH13849, A_MH14553, A_MH15048, A_MH15741, A_MH16005, A_MH16170, A_MH16665, A_MH17127_A_MH18843, A_MH18876, A_MH19676, A_MH20130, A_MH20933, A_MH21571, A_MH23527, A_MH24519, A_MH24640, A_MH24673, A_MH25256, A_MH25883, A_MH27137, A_MH35739, A_MH47300, A_MH53658, A_MH19657, A_MH26862	2000	Tanzania	931G>A; Nonsynonymous substitution	1179
Ocular	Ocular	A_SB002739	2013	Solomon Island	4	A_SB006930, A_SB008107, A_SB13112, A_SB13321	2013	Solomon Island	1141C>T; Nonsynonymous substitution	1179
Ocular	Ocular	**B_HAR36***	1969	Saudi Arabia	1	B_Jali16	1985	Gambia	None	1179
Ocular	Ocular	B_TZ1A828	1998	Tanzania	None	None	None	None	8-28del; frameshift and early truncation	33
Urogenital	Ocular	B_QH111L	2016	China	None	None	None	None	58C>T, 137A>G; Synonymous and nonsynonymous substitution	603
Ocular	Ocular	B_M48	2007	Gambia	None	None	None	None	557-558insGA; early truncation	570
Ocular	Ocular	B_Jali20	1985	Gambia	None	None	None	None	986G>C	1179
Ocular	Ocular	**Ba_Apache2**	1960	USA	None	None	None	None	1141C>T; Nonsynonymous substitution	1179
Ocular	Ocular	**C_TW3**	1959	Taiwan	None	None	None	None	None	1179
Ocular	Ocular	C_UW10	1964	Canada	None	None	None	None	179G>A; synonymous substitution	1179
Ocular	Ocular	**Da_TW448**	1985	Taiwan	None	None	None	None	None	1179

Note: trpB truncations are seen in B_TZ1A828, B_QH111L, B_M48.
*Since *Ct* reference strain B_UW5OT lacks trpB gene, B/HAR-36 was used as the reference strain for B genotype strains.

Supplementary Table 4. Single nucleotide polymorphism in *trpA* for *C. trachomatis* clinical and reference ocular strains.

Lineage	Site of infection	Unique *trpA* strain variants	Year of isolation	Geography	No. of variants	Variant ID	Year of isolation	Geography	Changes to *trpA*	Length of *trpA* (bp)
Ocular	Ocular	**A_HAR13**	1958	Egypt	9	A_SB002739, A_SB006930, A_SB008107, A_SB13112, A_SB13321	2013	Soloman Islands	529delC; frameshift	552
Ocular	Ocular	A_2497	2000	Tanzania	47	A_363, A_5291, A_7249, A_MH858, A_MH1364, A_MH2145, A_MH3234, A_MH2497, A_MH4510, A_MH5368, A_MH5786, A_MH6446, A_MH7205, A_MH8910, A_MH9922, A_MH10549, A_MH10648, A_MH10901, A_MH11715, A_MH11979, A_MH12023, A_MH13849, A_MH14553, A_MH15048, A_MH15741, A_MH16005, A_MH16170, A_MH16665, A_MH17127, A_MH18843, A_MH18876, A_MH19679, A_MH20933, A_MH20130, A_MH21571, A_MH23527, A_MH24519, A_MH24640, A_MH24673, A_MH25256, A_MH25883, A_MH27137, A_MH35739, A_MH 47300, A_MH53658, A_MH26862, A_MH19657	2000, 2001	Tanzania	117-118insGT; frameshift	138
Ocular	Ocular	A_D213	2001	Gambia	1	A_230	2001	Gambia	116A>G; nonsynonymous substitution	552
Ocular	Ocular	**Ba_Apache2**	1960	USA	None	None	None	None	529delC; frameshift	552
Ocular	Ocular	**B_HAR36***	1969	Saudi Arabia	None	None	None	None	205G>T; early truncation	207
Ocular	Ocular	B_Jali16	1985	Gambia	None	None	None	None	121G>A; nonsynonymous substitution; 529delC; frames	552
Ocular	Ocular	B_Jali20	1985	Gambia	None	None	None	None	529delC; frameshift	552
Ocular	Ocular	B_M48	2007	Gambia	None	None	None	None	157G>A; synonymous substitution; 529delC; frameshift	552
Urogenital	Ocular	B_QH111L	2016	China	None	None	None	None	410-412insATT; 492-604del; frameshift & early truncatio	498
Ocular	Ocular	B_TZ1A828	1998	Tanzania	None	None	None	None	None	759
Ocular	Ocular	**C_TW-3**	1959	Taiwan	None	None	None	None	529delC; frameshift	552
Ocular	Ocular	C_UW10	1964	Canada	None	None	None	None	10C>A, 529delC; frameshift	552
Ocular	Ocular	**Da_TW448**	1985	Taiwan	None	None	None	None	529delC; frameshift	552

Note: All ocular strains except for B_QH111L have the 410-412 deletion. Reference strains are highlighted in bold.

*Since *Ct* reference strain B_UW5OT lacks *trpA* gene, B/HAR-36 is used as the reference strain for B genotype strains.

Supplementary Data 1. List of *C. trachomatis* strains used in this study and associated metadata for the tryptophan operon nucleotide sequence.

Name	Lineage	*omp* A Genotyp	Country	Gender	Year	Source	operon (bp)	ENA ERR	ENA ERS	NCBI	Study accession	Genome Reference
A_A8	ocular	A	Sudan	M	2016-2019	ocular	2563				PRJEB32246	Alkidir 2019
A_A15	ocular	A	Sudan	F	2016-2019	ocular	2563				PRJEB32246	Alkidir 2019
A_363	ocular	A	Tanzania	F	2000	ocular	2565	ERR034213	ERS017900	-		Harris 2012
A_2497	ocular	A	Tanzania	F	2000	ocular	2565	-	-	FM872306		Harris 2012
A_5291	ocular	A	Tanzania	F	2000	ocular	2565	ERR034214	ERS017901	-		Harris 2012
A_7249	ocular	A	Tanzania	M	2000	ocular	2565	ERR034215	ERS017902	-		Harris 2012
A_D213	ocular	A	Gambia	M	2001	ocular	2562	ERR175652	ERS177838	-		
A_D230	ocular	A	Gambia	F	2001	ocular	2562	ERR111554	ERS075177	-		
A_HAR13	ocular	A	Egypt	F	1958	conjunctiv	2356	-	-	CP000051		Carlson 2005
A_MH858	ocular	A	Tanzania	M	2000	ocular	2565	ERR175560	ERS177716	-		
A_MH1364	ocular	A	Tanzania	F	2000	ocular	2565	ERR175575	ERS177731	-		
A_MH2145	ocular	A	Tanzania	F	2000	ocular	2565	ERR175561	ERS177717	-		
A_MH2497	ocular	A	Tanzania	F	2000	ocular	2565	ERR175562	ERS177718	-		
A_MH3234	ocular	A	Tanzania	F	2000	ocular	2565	ERR175576	ERS177732	-		
A_MH4510	ocular	A	Tanzania	F	2000	ocular	2565	ERR175577	ERS177733	-		
A_MH5368	ocular	A	Tanzania	F	2000	ocular	2565	ERR175563	ERS177719	-		
A_MH5786	ocular	A	Tanzania	M	2000	ocular	2565	ERR175564	ERS177720	-		
A_MH6446	ocular	A	Tanzania	M	2000	ocular	2565	ERR175578	ERS177734	-		
A_MH7205	ocular	A	Tanzania	M	2000	ocular	2565	ERR175579	ERS177735	-		
A_MH8910	ocular	A	Tanzania	M	2000	ocular	2565	ERR175580	ERS177736	-		
A_MH9922	ocular	A	Tanzania	F	2000	ocular	2565	ERR175581	ERS177737	-		
A_MH10549	ocular	A	Tanzania	M	2000	ocular	2565	ERR175582	ERS177738	-		
A_MH10648	ocular	A	Tanzania	F	2000	ocular	2565	ERR175583	ERS177739	-		
A_MH10901	ocular	A	Tanzania	M	2000	ocular	2565	ERR175584	ERS177740	-		
A_MH11715	ocular	A	Tanzania	M	2000	ocular	2565	ERR175585	ERS177741	-		
A_MH11979	ocular	A	Tanzania	M	2000	ocular	2565	ERR175586	ERS177742	-		
A_MH12023	ocular	A	Tanzania	F	2000	ocular	2565	ERR175565	ERS177721	-		
A_MH13849	ocular	A	Tanzania	M	2000	ocular	2565	ERR175566	ERS177722	-		
A_MH14553	ocular	A	Tanzania	F	2000	ocular	2565	ERR175567	ERS177723	-		
A_MH15048	ocular	A	Tanzania	F	2000	ocular	2565	ERR175587	ERS177743	-		
A_MH15741	ocular	A	Tanzania	F	2000	ocular	2565	ERR175588	ERS177744	-		
A_MH16005	ocular	A	Tanzania	M	2000	ocular	2565	ERR175568	ERS177724	-		
A_MH16170	ocular	A	Tanzania	M	2000	ocular	2565	ERR175589	ERS177745	-		
A_MH16665	ocular	A	Tanzania	M	2000	ocular	2565	ERR175590	ERS177746	-		
A_MH17127	ocular	A	Tanzania	M	2000	ocular	2565	ERR175591	ERS177747	-		
A_MH18843	ocular	A	Tanzania	F	2000	ocular	2565	ERR175569	ERS177725	-		
A_MH18876	ocular	A	Tanzania	F	2000	ocular	2565	ERR175592	ERS177748	-		
A_MH19657	ocular	A	Tanzania	M	2000	ocular	2565	ERR175570	ERS177726	-		

A_MH19679	ocular	A	Tanzania	F	2000	ocular	2565	ERR175593	ERS177749	-		
A_MH20130	ocular	A	Tanzania	F	2000	ocular	2565	ERR175594	ERS177750	-		
A_MH20933	ocular	A	Tanzania	F	2000	ocular	2565	ERR175595	ERS177751	-		
A_MH21571	ocular	A	Tanzania	F	2000	ocular	2565	ERR175596	ERS177752	-		
A_MH23527	ocular	A	Tanzania	F	2000	ocular	2565	ERR175571	ERS177727	-		
A_MH24519	ocular	A	Tanzania	M	2000	ocular	2565	ERR175597	ERS177753	-		
A_MH24640	ocular	A	Tanzania	F	2000	ocular	2565	ERR175598	ERS177754	-		
A_MH24673	ocular	A	Tanzania	M	2000	ocular	2565	ERR175572	ERS177728	-		
A_MH25256	ocular	A	Tanzania	M	2000	ocular	2565	ERR175573	ERS177729	-		
A_MH25883	ocular	A	Tanzania	F	2000	ocular	2565	ERR175599	ERS177755	-		
A_MH26862	ocular	A	Tanzania	F	2000	ocular	2565	ERR175600	ERS177756	-		
A_MH27137	ocular	A	Tanzania	F	2000	ocular	2565	ERR175601	ERS177757	-		
A_MH35739	ocular	A	Tanzania	M	2000	ocular	2565	ERR175603	ERS177759	-		
A_MH47300	ocular	A	Tanzania	F	2000	ocular	2565	ERR175574	ERS177730	-		
A_MH53658	ocular	A	Tanzania	F	2000	ocular	2565	ERR175604	ERS177760	-		
A_SA1	ocular	A	Saudi Arabia	F	1957	ocular	2562	ERR558498	ERS177777	-		
A_SB002739	ocular	A	Solomon Island		2013	ocular	2562	-	-	CP016418		Butcher 2016
A_SB006930	ocular	A	Solomon Island		2013	ocular	2562	-	-	CP016420		Butcher 2016
A_SB008107	ocular	A	Solomon Island		2013	ocular	2562	-	-	CP016422		Butcher 2016
A_SB013112	ocular	A	Solomon Island		2013	ocular	2562	-	-	CP016424		Butcher 2016
A_SB013321	ocular	A	Solomon Island		2013	ocular	2562	-	-	CP016426		Butcher 2016
A_B1	ocular	A	Sudan	M	2016-2019	ocular	2562			-	PRJEB32246	Alkidir 2019
A_B9	ocular	A	Sudan	F	2016-2019	ocular	2563			-	PRJEB32246	Alkidir 2019
A_B13	ocular	A	Sudan	F	2016-2019	ocular	2563			-	PRJEB32246	Alkidir 2019
A_S45	ocular	A	Sudan	M	2016-2019	ocular	2563			-	PRJEB32246	Alkidir 2019
A_S59	ocular	A	Sudan	F	2016-2019	ocular	2562			-	PRJEB32246	Alkidir 2019
A_T94	ocular	A	Sudan	F	2016-2019	ocular	2562			-	PRJEB32246	Alkidir 2019
A_TF16	ocular	A	Sudan	F	2016-2019	ocular	2562			-	PRJEB32246	Alkidir 2019
A_TF34	ocular	A	Sudan	F	2016-2019	ocular	2562			-	PRJEB32246	Alkidir 2019
A_TF54	ocular	A	Sudan	M	2016-2019	ocular	2562			-	PRJEB32246	Alkidir 2019
A_J52	ocular	A	Sudan	M	2016-2019	ocular	2562			-	PRJEB32246	Alkidir 2019
Ba_Apache2	ocular	Ba	USA		1960	ocular	2563	ERR140762	ERS095032	-		
B_Aus2	genital	B-Ba	Australia			unknown	2566	ERR189742	ERS153019	-		Andersson 2016
B_Aus3	genital	B-Ba	Australia		1986-1989	unknown	2566	ERR189743	ERS153020	-		Andersson 2016
B_Aus4	genital	B-Ba	Australia		1986-1989	unknown	2566	ERR189744	ERS153021	-		Andersson 2016
B_Aus5	genital	B-Ba	Australia		1986-1989	unknown	2566	ERR189745	ERS153022	-		Andersson 2016
B_Aus6	genital	B-Ba	Australia		1986-1989	unknown	2566	ERR189766	ERS153043	-		Andersson 2016
Ba_Aus25	genital	Ba	Australia		1988	ocular	2566	ERR386232	ERS351392	-		Andersson 2016
Ba_Aus28	genital	Ba	Australia		1988	ocular	2566	ERR386222	ERS351377	-		Andersson 2016
B_Aus36	genital	B-Ba	Australia		1989	ocular	2566	ERR386225	ERS351385	-		Andersson 2016
B_Aus40	genital	B-Ba	Australia			endocervic	2566	ERR999730	ERS747489	-		Andersson 2016

B_Aus41	genital	B-Ba	Australia			endocervic	2566	ERR999731	ERS747490	-	Andersson 2016
B_Aus42	genital	B-Ba	Australia			endocervic	2566	ERR999732	ERS747491	-	Andersson 2016
B_Aus43	genital	B-Ba	Australia			endocervic	2566	ERR999733	ERS747492	-	Andersson 2016
B_Aus44	genital	B-Ba	Australia			endocervic	2566	ERR999734	ERS747493	-	Andersson 2016
B_Aus45	genital	B-Ba	Australia			endocervic	2566	ERR999735	ERS747494	-	Andersson 2016
B_Fin101	genital	B-Ba	Finland	F	2009	endocervix	2566	ERR278185	ERS200114	-	
B_Fin203	genital	B-Ba	Finland	F	2011	endocervix	2566	ERR278154	ERS200083	-	
B_HAR36	ocular	B-Ba	Saudi Arabia		1969	ocular	2562	ERR189736	ERS153013	-	
B_Jali16	ocular	B-Ba	Gambia		1985	ocular	2562	ERR189738	ERS153015	-	
B_Jali20	ocular	B-Ba	Gambia		1985	ocular	2562	-	-	FM872308	Seth-Smith 2009
B_M48	ocular	B-Ba	Gambia	M	2007	ocular	2564	ERR175631	ERS177817	-	
B_NL2	genital	B-Ba	Netherlands	F	2001	endocervix	2566	ERR211016	ERS161066	-	
B/QH111L		B	China	M	2016	ocular	2434			CP018052	Feng Le 2016
B_Sou42	genital	B-Ba	UK		1985	Unknown	2566	ERR210997	ERS161047	-	
B_TZ1A828	ocular	B-Ba	Tanzania	F	1998	ocular	2541	-	-	FM872307	Seth-Smith 2009
C_Aus8	genital	C	Australia		1986-1989	unknown	2566	ERR210992	ERS161042	-	Andersson 2016
C_Aus9	genital	C	Australia		1986-1989	unknown	2566	ERR189768	ERS153045	-	Andersson 2016
C_Aus10	genital	C	Australia		1986-1989	unknown	2566	ERR189746	ERS153023	-	Andersson 2016
C_Aus30	genital	C	Australia	F	1988	ocular	2566	ERR386223	ERS351383	-	Andersson 2016
C_Aus33	genital	C	Australia	M	1988	ocular	2566	ERR386224	ERS351384	-	Andersson 2016
C_TW3	ocular	C	Taiwan		1959	ocular	2562	-	-	CP006945	
C_UW10	ocular	C	Canada	M	1964	ocular	2562	ERR175630	ERS177816	-	-
Da_TW448	ocular	Da	Taiwan		1985	Conjunctiv	2562				This study
D/13-96		D	USA, Seattle	NA	1996	endocervix		-	-	CP006676	Putman 2013
D/14-96		D	USA, Seattle	NA	1996	endocervix		-	-	CP006677	Putman 2013
D_Aus11	genital	D	Australia		1986-1989	unknown	2566	ERR189747	ERS153024	-	Andersson 2016
D_Aus12	genital	D	Australia		1986-1989	unknown	2566	ERR189767	ERS153044	-	Andersson 2016
D_C32	genital	D	UK	F	2011	Cx/Urethra	2566	ERR175621	ERS177807	-	-
D/CS637/11		D		F		endocervix				CP007131	Borges 2015
D_Fin163	genital	D	Finland	F	2010	endocervix	2566	ERR278143	ERS200072	-	-
D_Fin178	genital	D	Finland	F	2010	endocervix	2566	ERR278147	ERS200076	-	-
D_Fin187	genital	D	Finland	F	2010	endocervix	2566	ERR278150	ERS200079	-	-
D_HonLC4	genital	D	Honduras			unknown	2566	ERR658366	ERS151260	-	-
D_HPA314	genital	D	UK	F		endocervix	2566	ERR108297	ERS082979	-	-
D_NL4	genital	D	Netherlands	F	2001	endocervix	2566	ERR211017	ERS161067	-	-
D_NL5	genital	D	Netherlands	F	2001	endocervix	2566	ERR211018	ERS161068	-	-
D_NL6	genital	D	Netherlands	F	2001	endocervix	2566	ERR164672	ERS133259	-	-
D_NL8	genital	D	Netherlands	F	2001	endocervix	2566	ERR164674	ERS133261	-	-
D_NL10	genital	D	Netherlands	F	2001	endocervix	2566	ERR164676	ERS133263	-	-
D_NL11	genital	D	Netherlands	F	2001	endocervix	2566	ERR164677	ERS133264	-	-
D_NL12	genital	D	Netherlands	F	2001	endocervix	2566	ERR164678	ERS133265	-	-

D_NL13	genital	D	Netherlands	F	2001	endocervix	2566	ERR164679	ERS133266	-	-
D_NL14	genital	D	Netherlands	F	2001	endocervix	2566	ERR164680	ERS133267	-	-
D_NL15	genital	D	Netherlands	F	2001	endocervix	2566	ERR164681	ERS133268	-	-
D_NL16	genital	D	Netherlands	F	2001	endocervix	2566	ERR210970	ERS161020	-	-
D_NL17	genital	D	Netherlands	F	2001	endocervix	2566	ERR164682	ERS133269	-	-
D_NL19	genital	D	Netherlands	F	2001	endocervix	2566	ERR164684	ERS133271	-	-
D_NL32	genital	D	Netherlands	F	2001	endocervix	2566	ERR189763	ERS153040	-	-
D_NL59	genital	D	Netherlands	F	2001	endocervix	2566	ERR189765	ERS153042	-	-
D_NL71	genital	D	Netherlands	F	2001	endocervix	2566	ERR210988	ERS161038	-	-
D_S276I	genital	D	Sweden	F	2011	endocervix	2566	ERR140754	ERS095024	-	-
D_S1736	genital	D	Sweden	F	2010	endocervix	2566	ERR108286	ERS082968	-	-
D_S1879	genital	D	Sweden	M	2010	urethra	2566	ERR108287	ERS082969	-	-
D_S2130	genital	D	Sweden	F	2010	endocervix	2566	ERR111586	ERS075209	-	-
D_S3257	genital	D	Sweden	F	2010	endocervix	2566	ERR111590	ERS075213	-	-
D_S3489	genital	D	Sweden	F	2010	endocervix	2566	ERR111589	ERS075212	-	-
D_S3929	genital	D	Sweden	F	2010	endocervix	2566	ERR111578	ERS075201	-	-
D_S4093	genital	D	Sweden	F	2010	endocervix	2566	ERR140804	ERS095074	-	-
D_S4828	genital	D	Sweden	F	2010	endocervix	2566	ERR108294	ERS082976	-	-
D_SF2	genital	D	USA	F	2003	endocervix	2566				This study
D_SF12	genital	D	USA	F	2003	endocervix	2566				This study
D_Soton15	genital	D	UK	F	2009	endocervix	2566	ERR026568	ERS013807	-	-
D_Soton42	genital	D	UK	F	2009	endocervix	2566	ERR140819	ERS095089	-	-
D_Soton47	genital	D	UK	F	2009	endocervix	2566	ERR026573	ERS013812	-	-
D_Soton49	genital	D	UK	F	2009	endocervix	2566	ERR140821	ERS095091	-	-
D_Soton54	genital	D	UK	F	2009	endocervix	2566	ERR140822	ERS095092	-	-
D_Soton128	genital	D	UK	F	2009	endocervix	2566	ERR140826	ERS095096	-	-
D_Soton150	genital	D	UK	F	2009	endocervix	2566	ERR026586	ERS013823	-	-
D_SotonD1	genital	D	UK	F	2009	endocervix	2566	ERR027327	ERS008761	-	Harris 2012
D_SotonD2	genital	D	UK	F	2009	endocervix	2566	ERR026542	ERS013784	-	
D_SotonD3	genital	D	UK	F	2009	endocervix	2566	ERR026545	ERS013785	-	
D_SotonD4	genital	D	UK	F	2009	endocervix	2566	ERR026546	ERS013786	-	
D_SotonD5	genital	D	UK	F	2009	endocervix	2566	ERR026547	ERS013787	-	Harris 2012
D_SotonD6	genital	D	UK	F	2009	endocervix	2566	ERR027328	ERS008762	-	Harris 2012
D_SQ29	genital		USA		1991	unknown	2566			CP017731	
D_SQ32	genital		USA		1995	unknown	2566			CP017730	
D_STN101	genital	D	UK		1985	unknown	2566	ERR658584	ERS208565	-	-
D_STN113	genital	D	UK		1985	unknown	2566	ERR658586	ERS208567		
D_T9p	genital	D	UK		2012	unknown	2566	ERR140781	ERS095051	-	-
D_UK466322	genital	D	UK	F	2012	Cx/Urethra	2566	ERR658635	ERS208362	-	-
D_UK663610	genital	D	UK	F	2012	Cx/Urethra	2566	ERR658666	ERS208393	-	-
D_UK750376	genital	D	UK	M	2012	urethra	2566	ERR658498	ERS160257	-	-

D_UK750523	genital	D	UK	F	2012	endocervix	2566	ERR658501	ERS160260	-	-
D_UK912432	genital	D	UK	F	2012	Cx/Urethra	2566	ERR658673	ERS208400	-	-
D_UW3CX	genital	D	USA	F	1965	endocervix	2566	-	-	AE001273	Stephens 1998
E-103	genital	E	Germany	F	1992	unknown	2566			CP015294	Eder 2017
E_150	genital	E	0	M		rectum	2566	-	-	CP001886	Jeffrey 2010
E_160	genital	E	Germany	F	1995	unknown	2566			CP001886	Eder 2017
E_547	genital	E	Germany	F	1991	unknown	2566			CP015298	Eder 2017
E_940U470	genital	E	USA	M		urethra	2566	ERR348845	ERS248053	-	-
E_12-94	genital	E	USA			unknown	2566	-	-	CP006675	Putman 2013
E_8873	genital	E	Germany	F	1998	unknown	2566			CP015300	Eder 2017
E_11023	genital	E		F		endocervix	2566			CP001890	Jeffrey 2010
E_32931	genital	E	Germany	F		vagina	2566			CP015302	Eder 2017
E_Ar5	genital	E	Argentina	F	2006	endocervix	2566	ERR111634	ERS082951	-	-
E_Ar152	genital	E	Argentina	F	2005	ocular	2566	ERR111636	ERS082953	-	-
E_Ar182	genital	E	Argentina	M	2008	urethra	2566	ERR108273	ERS082955	-	-
E_Ar250	genital	E	Argentina	M	2006	ocular	2566	ERR658410	ERS160297		
E_Ar427	genital	E	Argentina	F	2011	endocervix	2566	ERR108279	ERS082961	-	-
E_Ar7218	genital	E	Argentina	M	2004	ocular	2566	ERR108280	ERS082962	-	-
E_Aus13	genital	E	Australia		1986-1989	unknown	2566	ERR189761	ERS153038	-	Andersson 2016
E_Bour	genital	E	USA	M	1959	ocular	2566	ERR008578	ERS001401	HE603212	Harris 2012
E_C5	genital	E	UK	F	2011	vagina	2566	ERR175618	ERS177804	-	-
E_C37	genital	E	UK	F	2011	vagina	2566	ERR175623	ERS177809	-	-
E_C58	genital	E	UK	F	2011	vagina	2566	ERR175626	ERS177812	-	-
E_C194	genital	E	UK	F	2011	vagina	2566	ERR175638	ERS177824	-	-
E_C208	genital	E	UK	M	2011	Urine	2566	ERR175640	ERS177826	-	-
E_C236	genital	E	UK	F	2011	Ur/Cx	2566	ERR175643	ERS177829	-	-
E_C258	genital	E	UK		2011	unknown	2566	ERR175646	ERS177832	-	-
E_C599	genital	E	USA	M	1996	urethra	2566	ERR027326	ERS008760	-	-
E_CC15	genital	E	UK	F	2009	endocervix	2566	ERR140842	ERS095112	-	-
E_CC35	genital	E	UK	M	2010	urethra	2566	ERR140843	ERS095113	-	-
E_CLIN1	genital	E	USA			endocervix	2566				This study
E_CLIN3	genital	E	USA			endocervix	2566				This study
E_CS102511	genital	E	NA	F	2015	endocervix				CP010567	Borges 2015
E_DK20	genital	E	Denmark		1967	conjunctiv	2566	ERR558495	ERS177774	-	-
E_Fin127	genital	E	Finland	F	2010	endocervix	2566	ERR278134	ERS200063	-	-
E_Fin129	genital	E	Finland	F	2010	endocervix	2566	ERR278135	ERS200064	-	-
E_Fin142	genital	E	Finland	F	2010	endocervix	2566	ERR278136	ERS200065	-	-
E_Fin155	genital	E	Finland	F	2010	endocervix	2566	ERR278141	ERS200070	-	-
E_Fin159	genital	E	Finland	F	2010	endocervix	2566	ERR278187	ERS200116	-	-
E_Fin172	genital	E	Finland	F	2010	endocervix	2566	ERR278145	ERS200074	-	-
E_Fin184	genital	E	Finland	F	2010	endocervix	2566	ERR278148	ERS200077	-	-

E_Fin185	genital	E	Finland	F	2010	endocervix	2566	ERR278149	ERS200078	-	-
E_Fin194	genital	E	Finland	F	2010	endocervix	2566	ERR278152	ERS200081	-	-
E_Fin198	genital	E	Finland	F	2010	endocervix	2566	ERR278153	ERS200082	-	-
E_Fin214	genital	E	Finland	F	2011	endocervix	2566	ERR278158	ERS200087	-	-
E_Fin220	genital	E	Finland	F	2011	endocervix	2566	ERR278161	ERS200090	-	-
E_It246	genital	E	Italy	F	2010	endocervix	2566	ERR658535	ERS208511	-	-
E_It363	genital	E	Italy	F	2010	endocervix	2566	ERR658534	ERS208510	-	-
E_It769	genital	E	Italy	F	2011	endocervix	2566	ERR658536	ERS208512	-	-
E_It807	genital	E	Italy			unknown	2566	ERR658407	ERS160294	-	-
E_IU824	genital	E	USA	F		endometria	2566	ERR558500	ERS177779	HF562298	O'Neill 2013
E_IU888	genital	E	USA	F		endometria	2566	ERR558501	ERS177780	HF562300	O'Neill 2013
E_NL21	genital	E	Netherlands	F	2001	endocervix	2566	ERR211019	ERS161069	-	-
E_NL23	genital	E	Netherlands	F	2001	endocervix	2566	ERR189756	ERS153033	-	-
E_NL24	genital	E	Netherlands	F	2001	endocervix	2566	ERR189757	ERS153034	-	-
E_NL26	genital	E	Netherlands	F	2001	endocervix	2566	ERR164687	ERS133274	-	-
E_NL28	genital	E	Netherlands	F	2001	endocervix	2566	ERR164689	ERS133276	-	-
E_NL29	genital	E	Netherlands	F	2001	endocervix	2566	ERR189758	ERS153035	-	-
E_R526	genital	E	Russia	F	2011	endocervix	2566	ERR111568	ERS075191	-	-
E_R1430	genital	E	Russia	M	2011	urethra	2566	ERR111621	ERS082938	-	-
E_R4159	genital	E	Russia	M	2011	urethra	2566	ERR111615	ERS082932	-	-
E_R4195	genital	E	Russia	F	2011	endocervix	2566	ERR140796	ERS095066	-	-
E_R4528	genital	E	Russia	F	2011	endocervix	2566	ERR140835	ERS095105	-	-
E_R16965	genital	E	Russia	F	2011	endocervix	2566	ERR140837	ERS095107	-	-
E_R25114	genital	E	Russia	F	2010	endocervix	2566	ERR111596	ERS075219	-	-
E_R26833	genital	E	Russia	F	2010	endocervix	2566	ERR111601	ERS082918	-	Seth-Smith 2013
E_R27091	genital	E	Russia	F	2010	endocervix	2566	ERR111566	ERS075189	-	Seth-Smith 2013
E_R28017	genital	E	Russia	F	2010	endocervix	2566	ERR111565	ERS075188	-	Seth-Smith 2013
E_R28044	genital	E	Russia	F	2010	endocervix	2566	ERR111594	ERS075217	-	-
E_R29005	genital	E	Russia	F	2010	endocervix	2566	ERR111630	ERS082947	-	-
E_R30444	genital	E	Russia	F	2010	endocervix	2566	ERR111600	ERS075223	-	-
E_R32100	genital	E	Russia	F	2010	endocervix	2566	ERR111563	ERS075186	-	Seth-Smith 2013
E_R33420	genital	E	Russia	F	2010	endocervix	2566	ERR164655	ERS133242	-	-
E_R35067	genital	E	Russia	M	2010	urethra	2566	ERR111567	ERS075190	-	-
E_Rb1392	genital	E	UK	F	2010	endocervix	2566	ERR071994	ERS066957	-	-
E_Rb10387	genital	E	UK	M	2010	Unknown	2566	ERR034219	ERS017906	-	-
E_Rb10392	genital	E	UK	F	2010	endocervix	2566	ERR175633	ERS177819	-	-
E_S581	genital	E	Sweden	F	2011	endocervix	2566	ERR189737	ERS153014	-	-
E_S1019	genital	E	Sweden	F	2010	endocervix	2566	ERR140808	ERS095078	-	-
E_S1148	genital	E	Sweden	F	2010	endocervix	2566	ERR140810	ERS095080	-	-
E_S1227	genital	E	Sweden	F	2010	endocervix	2566	ERR140805	ERS095075	-	-
E_S1528	genital	E	Sweden	F	2010	endocervix	2566	ERR140809	ERS095079	-	-

E_S1613	genital	E	Sweden	F	2010	endocervix	2566	ERR140812	ERS095082	-	-
E_S1618	genital	E	Sweden	F	2010	endocervix	2566	ERR108290	ERS082972	-	-
E_S1886	genital	E	Sweden	F	2010	endocervix	2566	ERR108289	ERS082971	-	-
E_S2384	genital	E	Sweden	F	2010	endocervix	2566	ERR140760	ERS095030	-	-
E_S2491	genital	E	Sweden	F	2010	endocervix	2566	ERR108284	ERS082966	-	-
E_S2699	genital	E	Sweden	F	2011	endocervix	2566	ERR140831	ERS095101	-	-
E_S2713	genital	E	Sweden	F	2011	endocervix	2566	ERR140832	ERS095102	-	-
E_S3024	genital	E	Sweden	M	2010	conjunctiv	2566	ERR111579	ERS075202	-	-
E_S3066	genital	E	Sweden	F	2011	endocervix	2566	ERR140841	ERS095111	-	-
E_S3073	genital	E	Sweden	F	2010	endocervix	2566	ERR140801	ERS095071	-	-
E_S3085	genital	E	Sweden	F	2011	endocervix	2566	ERR140833	ERS095103	-	-
E_S3122	genital	E	Sweden	F	2010	endocervix	2566	ERR111582	ERS075205	-	-
E_S3695	genital	E	Sweden	F	2006	endocervix	2566	ERR140815	ERS095085	-	-
E_S3711	genital	E	Sweden	F	2011	endocervix	2566	ERR140834	ERS095104	-	-
E_S3724	genital	E	Sweden	F	2006	endocervix	2566	ERR140816	ERS095086	-	-
E_S3732	genital	E	Sweden	F	2010	Unknown	2566	ERR108282	ERS082964	-	-
E_S4007	genital	E	Sweden	F	2006	endocervix	2566	ERR140817	ERS095087	-	-
E_S4106	genital	E	Sweden	F	2010	endocervix	2566	ERR140803	ERS095073	-	-
E_S4247	genital	E	Sweden	M	2010	urethra	2566	ERR108293	ERS082975	-	-
E_S4324	genital	E	Sweden	M	2010	urethra	2566	ERR108288	ERS082970	-	-
E_S4471	genital	E	Sweden	F	2010	endocervix	2566	ERR108285	ERS082967	-	-
E_SF4	genital	E	USA	F	2003	endocervix	2566	-	-	-	This study
E_SF9	genital	E	USA	F	2003	endocervix	2566	-	-	-	This study
E_SF13	genital	E	USA	F	2003	endocervix	2566	-	-	-	This study
E_SF14	genital	E	USA	F	2003	endocervix	2566	-	-	-	This study
E_SF15	genital	E	USA	F	2003	endocervix	2566	-	-	-	This study
E_SF17	genital	E	USA	F	2003	endocervix	2566	-	-	-	This study
E_SF18	genital	E	USA	F	2003	endocervix	2566	-	-	-	This study
E_SF23	genital	E	USA	F	2003	endocervix	2566	-	-	-	This study
E_SF24	genital	E	USA	F	2003	endocervix	2566	-	-	-	This study
E_SF26	genital	E	USA	F	2003	endocervix	2566	-	-	-	This study
E_SF28	genital	E	USA	F	2003	endocervix	2566	-	-	-	v
E_Soton17	genital	E	UK	F	2009	endocervix	2566	ERR111555	ERS075178	-	-
E_Soton53	genital	E	UK	F	2009	endocervix	2566	ERR111556	ERS075179	-	-
E_Soton83	genital	E	UK	F	2009	endocervix	2566	ERR140823	ERS095093	-	-
E_Soton107	genital	E	UK	F	2009	endocervix	2566	ERR111557	ERS075180	-	-
E_Soton116	genital	E	UK	F	2009	endocervix	2566	ERR111558	ERS075181	-	-
E_Soton120	genital	E	UK	F	2009	endocervix	2566	ERR111559	ERS075182	-	-
E_Soton145	genital	E	UK	F	2009	endocervix	2566	ERR111560	ERS075183	-	-
E_Soton159	genital	E	UK	F	2009	endocervix	2566	ERR111561	ERS075184	-	-
E_SotonE2	genital	E	UK	F	2009	endocervix	2566	ERR175612	ERS013789	-	-

E_SotonE4	genital	E	UK	F	2009	endocervix	2566	ERR026551	ERS013791	HE603232	Harris 2012
E_SotonE6	genital	E	UK	F	2009	endocervix	2566	ERR026543	ERS013793	-	-
E_SotonE8	genital	E	UK	F	2009	endocervix	2566	ERR027329	ERS008763	HE603233	Harris 2012
E_Sou60	genital	E	UK		1985	unknown	2566	ERR210999	ERS161049	-	-
E_Sou75	genital	E	UK		1985	unknown	2566	ERR211004	ERS161054	-	-
E_Sou102	genital	E	UK	F	1985	unknown	2566	ERR211010	ERS161060	-	-
E_STN2	genital	E	UK		1985	unknown	2566	ERR658574	ERS208554	-	-
E_STN10	genital	E	UK		1985	unknown	2566	ERR658575	ERS208555	-	-
E_STN11	genital	E	UK		1985	unknown	2566	ERR658576	ERS208556	-	-
E_STN12	genital	E	UK		1985	unknown	2566	ERR658577	ERS208557	-	-
E_STN47	genital	E	UK		1985	unknown	2566	ERR658581	ERS208562	-	-
E_STN68	genital	E	UK		1985	unknown	2566	ERR658582	ERS208563	-	-
E_STN92	genital	E	UK		1985	unknown	2566	ERR658583	ERS208564	-	-
E_STN115	genital	E	UK		1985	unknown	2566	ERR658587	ERS208568	-	-
E_STN119	genital	E	UK		1985	unknown	2566	ERR658588	ERS208569	-	-
E_SW2	genital	E	Sweden	M	2006	urethra	2566	ERR008596	ERS001397	FM865439	Unemo 2010
E_SW3	genital	E	Sweden	F	2001	endocervix	2566	ERR008589	ERS001406	FM865440	Harris 2012
E_Swab6	genital	E	UK	F	2010	vagina	2566	ERR175608	ERS177764	-	Seth-Smith 2013
E_SwabB4	genital	E	UK	F	2010	unknown	2566	ERR034345	ERS015770	-	Seth-Smith 2013
E_UK34334	genital	E	UK	M	2012	urethra	2566	ERR658682	ERS208409	-	-
E_UK466129	genital	E	UK	F	2012	Cx/Urethra	2566	ERR658591	ERS208318	-	-
E_UK466546	genital	E	UK	F	2012	Cx/Urethra	2566	ERR658639	ERS208366	-	-
E_UK582206	genital	E	UK	F	2012	Cx/Urethra	2566	ERR658642	ERS208369	-	-
E_UK582263	genital	E	UK	F	2012	Cx/Urethra	2566	ERR658595	ERS208322	-	-
E_UK584031	genital	E	UK	F	2012	Cx/Urethra	2566	ERR658602	ERS208329	-	-
E_UK663813	genital	E	UK	F	2012	Cx/Urethra	2566	ERR658606	ERS208333	-	-
E_UK663968	genital	E	UK	M	2012	urethra	2566	ERR658668	ERS208395	-	-
E_UK664394	genital	E	UK	M	2012	urethra	2566	ERR658669	ERS208396	-	-
E_UK769748	genital	E	UK	F	2012	Cx/Urethra	2566	ERR658670	ERS208397	-	-
E_UK769852	genital	E	UK	M	2012	urethra	2566	ERR658671	ERS208398	-	-
E_UK913723	genital	E	UK	F	2012	Cx/Urethra	2566	ERR658617	ERS208344	-	-
E_UK913953	genital	E	UK	F	2012	vagina	2566	ERR658619	ERS208346	-	-
F_70	genital	F				unknown	2566	-	-	ABYF01000001	Jeffrey 2010
F_1-93	genital	F	USA	F	1993	endocervix	2566	-	-	CP006671	Putman 2013
F_2-93	genital	F	USA	F	1993	endocervix	2566	-	-	CP006672	Putman 2013
F_6-94	genital	F	USA	F	1994	endocervix	2566	-	-	CP006673	Putman 2013
F_11-96	genital	F	USA	F	1996	endocervix	2566	-	-	CP006674	Putman 2013
F_6068	genital	F	Germany	F		urethra				CP015306	Eder 2017
F_AddT9	genital	F	UK			unknown	2566	ERR658365	ERS151259	-	-
F_Aus20	genital	F	Australia		1986-1989	unknown	2566	ERR189754	ERS153031	-	Andersson 2016
F_C55	genital	F	UK	F	2011	vagina	2566	ERR175625	ERS177811	-	-

F_CS84708	genital	F		F		2015	endocervix				CP010569	Borges 2015
F_Fin106	genital	F	Finland	F	2010		endocervix	2566	ERR278132	ERS200061	-	-
F_Fin152	genital	F	Finland	F	2010		endocervix	2566	ERR278139	ERS200068	-	-
F_Fin175	genital	F	Finland	F	2010		endocervix	2566	ERR278146	ERS200075	-	-
F_Fin181	genital	F	Finland	F	2010		endocervix	2566	ERR278162	ERS200091	-	-
F_Fin213	genital	F	Finland	F	2011		endocervix	2566	ERR278190	ERS200119	-	-
F_Fin219	genital	F	Finland	F	2011		endocervix	2566	ERR278160	ERS200089	-	-
F_IC-Cal-3	genital	F	USA		1960		ocular (nec	2566	ERR278181	ERS200110	-	Joseph 2012
F_It686	genital	F	Italy	F	2011		endocervix	2566	ERR658532	ERS208508	-	-
F_It688	genital	F	Italy	M	2011		urethra	2566	ERR140764	ERS095034	-	-
F_NI1	genital	F	Unknown				unknown	2566	ERR175650	ERS177836	-	-
F_NL30	genital	F	Netherlands	F	2001		endocervix	2566	ERR164690	ERS133277	-	-
F_NL31	genital	F	Netherlands	F	2001		endocervix	2566	ERR189762	ERS153039	-	-
F_NL35	genital	F	Netherlands	F	2001		endocervix	2566	ERR211022	ERS161072	-	-
F_NL36	genital	F	Netherlands	F	2001		endocervix	2566	ERR189726	ERS153003	-	-
F_NL38	genital	F	Netherlands	F	2001		endocervix	2566	ERR278184	ERS200113	-	-
F_R4663	genital	F	Russia	F	2011		endocervix	2566	ERR140795	ERS095065	-	-
F_R7369	genital	F	Russia	F	2011		endocervix	2566	ERR140798	ERS095068	-	-
F_R12921	genital	F	Russia	F	2011		endocervix	2566	ERR140799	ERS095069	-	-
F_R28312	genital	F	Russia	M	2010		urethra	2566	ERR111597	ERS075220	-	-
F_S1470	genital	F	Sweden	F	2010		endocervix	2566	ERR108292	ERS082974	-	-
F_S1494	genital	F	Sweden	F	2010		endocervix	2566	ERR140811	ERS095081	-	-
F_S2430	genital	F	Sweden	M	2010		urethra	2566	ERR111583	ERS075206	-	-
F_S2526	genital	F	Sweden	F	2010		unknown	2566	ERR111577	ERS075200	-	-
F_S2595	genital	F	Sweden	F	2010		endocervix	2566	ERR108283	ERS082965	-	
F_S3948	genital	F	Sweden	F	2010		endocervix	2566	ERR140800	ERS095070	-	-
F_S4410	genital	F	Sweden	F	2010		endocervix	2566	ERR111593	ERS075216	-	-
F_SF7	genital	F	USA	F	2003		endocervix	2566				Samboonna 2019
F_SF8	genital	F	USA	F	2003		endocervix	2566				Samboonna 2019
F_SF10	genital	F	USA	F	2003		endocervix	2574				Samboonna 2019
F_SF11	genital	F	USA	F	2003		endocervix	2580				Samboonna 2019
F_SF14	genital	F	USA	F	2003		endocervix	2574				Samboonna 2019
F_SF16	genital	F	USA	F	2003		endocervix	2566				Samboonna 2019
F_SF19	genital	F	USA	F	2003		endocervix	2566				Samboonna 2019
F_SF21	genital	F	USA	F	2003		endocervix	2574				Samboonna 2019
F_Soton18	genital	F	UK	F	2009		endocervix	2566	ERR026569	ERS013808	-	-
F_Soton48	genital	F	UK	F	2009		endocervix	2566	ERR140820	ERS095090	-	-
F_Soton88	genital	F	UK	F	2009		endocervix	2566	ERR026581	ERS013818	-	-
F_Soton106	genital	F	UK	F	2009		endocervix	2566	ERR140824	ERS095094	-	-
F_Soton118	genital	F	UK	F	2009		endocervix	2566	ERR140825	ERS095095	-	-
F_Soton137	genital	F	UK	F	2009		endocervix	2566	ERR140827	ERS095097	-	-

F_SotonF1	genital	F	UK	F	2009	endocervix	2566	ERR026556	ERS013796	-	-
F_SotonF2	genital	F	UK	F	2009	endocervix	2566	ERR026557	ERS013797	-	-
F_SotonF3	genital	F	UK	F	2009	endocervix	2566	ERR027330	ERS008764	HE603234	Harris 2012
F_SotonF4	genital	F	UK	F	2009	endocervix	2566	ERR026558	ERS013798	-	-
F_Sou9	genital	F	UK		1985	unknown	2566	ERR189771	ERS153048	-	-
F_Sou87	genital	F	UK		1985	unknown	2566	ERR211005	ERS161055	-	-
F_Sou89	genital	F	UK	F		unknown	2566	ERR211006	ERS161056	-	-
F_Sou100	genital	F	UK	M		unknown	2566	ERR211009	ERS161059	-	-
F_STN15	genital	F	UK		1985	unknown	2566	ERR658578	ERS208558	-	-
F_STN22	genital	F	UK		1985	unknown	2566	ERR658580	ERS208560	-	-
F_STN110	genital	F	UK		1985	unknown	2566	ERR658585	ERS208566	-	-
F_SW4	genital	F	Sweden	F	2002	endocervix	2566	ERR008588	ERS001414	FM865441	Harris 2012
F_SW5	genital	F	Sweden	F	2002	endocervix	2566	ERR008582	ERS001415	FM865442	Harris 2012
F_Swab5	genital	F	UK	F	2010	vagina	2566	ERR024700	ERS013115	-	Seth-Smith 2013
F_SwabB1	genital	F	UK	F	2010	vagina	2566	ERR173901	ERS177788	-	Seth-Smith 2013
F_SwabB8	genital	F	UK	F	2010	vagina	2566	ERR034346	ERS015772	-	Seth-Smith 2013
F_SWFP	genital	F	Sweden	F	2009	endocervix	2566	ERR008598	ERS001400	-	-
F_UK35155	genital	F	UK	F	2012	Cx/Urethra	2566	ERR658626	ERS208353	-	-
F_UK220521	genital	F	UK	M	2012	urethra	2566	ERR658399	ERS160285	-	-
F_UK465966	genital	F	UK	F	2012	Cx/Urethra	2566	ERR658632	ERS208359	-	-
F_UK466273	genital	F	UK	F	2012	Cx/Urethra	2566	ERR658592	ERS208319	-	-
F_UK583012	genital	F	UK	F	2012	Cx/Urethra	2566	ERR658656	ERS208383	-	-
F_UK583072	genital	F	UK	F	2012	Cx/Urethra	2566	ERR658652	ERS208379	-	-
F_UK583468	genital	F	UK	F	2012	Cx/Urethra	2566	ERR658649	ERS208376	-	-
F_UK584026	genital	F	UK	F	2012	Cx/Urethra	2566	ERR658647	ERS208374	-	-
F_UK663442	genital	F	UK	M	2012	urethra	2566	ERR658664	ERS208391	-	-
F_UK770010	genital	F	UK	F	2012	Cx/Urethra	2566	ERR658672	ERS208399	-	-
G_9301	genital	G	0	M		urethra	2566	-	-	CP001930	Jeffrey 2010
G_9768	genital	G	0	M		rectum	2566	-	-	CP001887	Jeffrey 2010
G_11074	genital	G	0	M		rectum	2566	-	-	CP001889	Jeffrey 2010
G_11222	genital	G	0	F		endocervix	2566	-	-	CP001888	Jeffrey 2010
G_Ar112	genital	G	Argentina	M	2007	urethra	2566	ERR111635	ERS082952	-	-
G_Ar246	genital	G	Argentina	M	2007	urethra	2566	ERR108276	ERS082958	-	-
G_Aus1	genital	G	Australia		1986-1989	unknown	2566	ERR189741	ERS153018	-	Andersson 2016
G_Aus16	genital	G	Australia		1986-1989	unknown	2566	ERR189751	ERS153028	-	Andersson 2016
G_Aus17	genital	G	Australia		1986-1989	unknown	2566	ERR189752	ERS153029	-	Andersson 2016
G_Aus18	genital	G	Australia		1986-1989	unknown	2566	ERR189759	ERS153036	-	Andersson 2016
G_Aus19	genital	G	Australia		1986-1989	unknown	2566	ERR189760	ERS153037	-	Andersson 2016
G_Fin144	genital	G	Finland	F	2010	endocervix	2566	ERR278137	ERS200066	-	-
G_Fin153	genital	G	Finland	F	2010	endocervix	2566	ERR278140	ERS200069	-	-
G_Fin158	genital	G	Finland	F	2010	endocervix	2566	ERR278142	ERS200071	-	-

G_Fin205	genital	G	Finland	F	2011	endocervix	2566	ERR278188	ERS200117	-	-
G_NL39	genital	G	Netherlands	F	2001	endocervix	2566	ERR189764	ERS153041	-	-
G_NL40	genital	G	Netherlands	F	2001	endocervix	2566	ERR189729	ERS153006	-	-
G_NL41	genital	G	Netherlands	F	2001	endocervix	2566	ERR210971	ERS161021	-	-
G_NL42	genital	G	Netherlands	F	2001	endocervix	2566	ERR210972	ERS161022	-	-
G_NL43	genital	G	Netherlands	F	2001	endocervix	2566	ERR189730	ERS153007	-	-
G_NL44	genital	G	Netherlands	F	2001	endocervix	2566	ERR278214	ERS200143	-	-
G_NL45	genital	G	Netherlands	F	2001	endocervix	2566	ERR189732	ERS153009	-	-
G_NL46	genital	G	Netherlands	F	2001	endocervix	2566	ERR189733	ERS153010	-	-
G_NL47	genital	G	Netherlands	F	2001	endocervix	2566	ERR210973	ERS161023	-	-
G_NL48	genital	G	Netherlands	F	2001	endocervix	2566	ERR189734	ERS153011	-	-
G_R297	genital	G	Russia	F	2011	endocervix	2566	ERR111576	ERS075199	-	-
G_R459	genital	G	Russia	F	2011	endocervix	2566	ERR111617	ERS082934	-	-
G_R2247	genital	G	Russia	F	2011	endocervix	2566	ERR111620	ERS082937	-	-
G_R3059	genital	G	Russia	F	2011	endocervix	2566	ERR111612	ERS082929	-	Seth-Smith 2013
G_R4175	genital	G	Russia	F	2011	endocervix	2566	ERR111614	ERS082931	-	Seth-Smith 2013
G_R9069	genital	G	Russia	F	2011	endocervix	2566	ERR111625	ERS082942	-	-
G_R9892	genital	G	Russia	M	2011	urethra	2566	ERR111619	ERS082936	-	-
G_R15108	genital	G	Russia	M	2011	urethra	2566	ERR140756	ERS095026	-	-
G_R23736	genital	G	Russia	F	2010	endocervix	2566	ERR111595	ERS075218	-	-
G_R27757	genital	G	Russia	F	2010	endocervix	2566	ERR111623	ERS082940	-	-
G_R30591	genital	G	Russia	F	2010	endocervix	2566	ERR111569	ERS075192	-	-
G_R31458	genital	G	Russia	M	2010	urethra	2566	ERR111605	ERS082922	-	-
G_R35506	genital	G	Russia	F	2010	endocervix	2566	ERR111610	ERS082927	-	-
G_R36176	genital	G	Russia	F	2010	endocervix	2566	ERR111611	ERS082928	-	Seth-Smith 2013
G_S1471	genital	G	Sweden	F	2010	endocervix	2566	ERR108291	ERS082973	-	-
G_S1824	genital	G	Sweden	F	2010	endocervix	2566	ERR111587	ERS075210	-	-
G_S1846	genital	G	Sweden	F	2010	endocervix	2566	ERR140807	ERS095077	-	-
G_S2477	genital	G	Sweden	F	2010	endocervix	2566	ERR140814	ERS095084	-	-
G_S2956	genital	G	Sweden	F	2010	conjunctiv	2566	ERR111581	ERS075204	-	-
G_S3270	genital	G	Sweden	F	2010	endocervix	2566	ERR111588	ERS075211	-	-
G_S3344	genital	G	Sweden	F	2010	endocervix	2566	ERR111591	ERS075214	-	-
G_S4641	genital	G	Sweden	F	2010	endocervix	2566	ERR140813	ERS095083	-	-
G_S4658	genital	G	Sweden	M	2010	conjunctiv	2566	ERR108295	ERS082977	-	-
G_SF20	genital	G	USA	F	2003	endocervix	2566				This study
G_SF21	genital	G	USA	F	2003	endocervix	2566				This study
G_Soton144	genital	G	UK	F	2009	endocervix	2566	ERR140828	ERS095098	-	-
G_SotonG1	genital	G	UK	F	2009	endocervix	2566	ERR026560	ERS013800	-	Harris 2012
G_UK22140	genital	G	UK			unknown		ERR658523	ERS160283		
G_UK58250	genital	G	UK	M	2012	urethra	2566	ERR658658	ERS208385	-	-
G_UK75036	genital	G	UK	F	2012	endocervix	2566	ERR658369	ERS160255	-	-

G_UK913362	genital	G	UK	F	2012		Cx/Urethra	2566	ERR658677	ERS208404	-	-
G_UW57	genital	G	USA		1971		endocervix	2566	ERR164658	ERS133245	-	Joseph 2012
H_Fin109	genital	H	Finland	F	2010		endocervix	2566	ERR278213	ERS200142	-	-
H_NL49	genital	H	Netherlands	F	2001		endocervix	2566	ERR210974	ERS161024	-	-
H_NL50	genital	H	Netherlands	F	2001		endocervix	2566	ERR210975	ERS161025	-	-
H_NL51	genital	H	Netherlands	F	2001		endocervix	2566	ERR210976	ERS161026	-	-
H_NL53	genital	H	Netherlands	F	2001		endocervix	2566	ERR210978	ERS161028	-	-
H_NL54	genital	H	Netherlands	F	2001		endocervix	2566	ERR210979	ERS161029	-	-
H_NL56	genital	H	Netherlands	F	2001		endocervix	2566	ERR210981	ERS161031	-	-
H_R13670	genital	H	Russia	M	2011		urethra	2566	ERR164654	ERS133241	-	-
H_R25308	genital	H	Russia	F	2010		endocervix	2566	ERR111616	ERS082933	-	-
H_R27887	genital	H	Russia	F	2010		endocervix	2566	ERR111599	ERS075222	-	-
H_R31975	genital	H	Russia	F	2010		endocervix	2566	ERR111606	ERS082923	-	-
H_S269	genital	H	Sweden	F	2011		unknown	2566	ERR164646	ERS133233	-	-
H_S1026	genital	H	Sweden	F	2010		endocervix	2566	ERR108301	ERS082983	-	-
H_S1314	genital	H	Sweden	F	2010		endocervix	2566	ERR108296	ERS082978	-	-
H_S1432	genital	H	Sweden	F	2010		endocervix	2566	ERR108303	ERS082985	-	-
H_S4377	genital	H	Sweden	F	2010		endocervix	2566	ERR164648	ERS133235	-	-
H_UW4	genital	H	USA	F	1965		endocervix	2566	ERR558497	ERS177776	-	Joseph 2012
H_UW4	genital	H	USA	F	1965		endocervix	2566	ERR558497	ERS177776	-	Joseph 2012
Ia_20-97	genital	I-Ia	USA				endocervix	2566	-	-	CP006678	Putman 2013
Ia/CS190/96		I-Ia		F	2015		endocervix				CP010571	Borges 2015
Ia_SF16	genital	Ia	USA	F	2003		endocervix	2566				
Ia_SF25	genital	Ia	USA	F	2003		endocervix	2566				
Ia_SF27	genital	Ia	USA	F	2003		endocervix	2566				
Ia_SotonIa1	genital	I-Ia	UK	F	2009		endocervix	2566	ERR026555	ERS013804	HE603236	Harris 2012
Ia_SotonIa3	genital	I-Ia	UK	F	2009		endocervix	2566	ERR026565	ERS013805	HE603237	Harris 2012
Ia_UW202	genital	I-Ia	USA				unknown					This study
I_NL58	genital	I-Ia	Netherlands	F	2001		endocervix	2566	ERR210982	ERS161032	-	-
I_NL63	genital	I-Ia	Netherlands	F	2001		endocervix	2566	ERR210985	ERS161035	-	-
I_NL66	genital	I-Ia	Netherlands	F	2001		endocervix	2566	ERR211026	ERS161076	-	-
I_NL67	genital	I-Ia	Netherlands	F	2001		endocervix	2566	ERR210986	ERS161036	-	-
I_NL69	genital	I-Ia	Netherlands	F	2001		endocervix	2566	ERR211023	ERS161073	-	-
I_NL70	genital	I-Ia	Netherlands	F	2001		endocervix	2566	ERR210987	ERS161037	-	-
I_NL72	genital	I-Ia	Netherlands	F	2001		endocervix	2566	ERR211027	ERS161077	-	-
I_S2459	genital	I-Ia	Sweden	F	2010		endocervix	2566	ERR111592	ERS075215	-	-
I_UK913341	genital	I-Ia	UK	F		2012	vagina	2566	ERR658613	ERS208340		
I_UW12	genital	I-Ia	USA		1966		urethra	2566	ERR278215	ERS200144	-	-
Ja/UW-92	genital	Ja	Seattle, USA		1992		unknown	2566				This study
J_27-97	genital	J	USA				endocervix	2566	-	-	CP006679	Putman 2013
J_31-98	genital	J	USA				endocervix	2566	-	-	CP006680	Putman 2013

J_6276	genital	J	0	F		endocervix	2566	-	-	ABYD01000001	Jeffrey 2010
J 6276s	genital					endocervix	2566			ABYD01000002	Jeffrey 2010
J_C114	genital	J	UK	F	2011	vagina	2566	ERR175629	ERS177815	-	-
J_NL55	genital	J	Netherlands	F	2001	endocervix	2566	ERR210980	ERS161030	-	-
J_NL76	genital	J	Netherlands	F	2001	endocervix	2566	ERR210989	ERS161039	-	-
J_NL78	genital	J	Netherlands	F	2001	endocervix	2566	ERR210990	ERS161040	-	-
J_S42	genital	J	Sweden	F	2011	endocervix	2566	ERR140818	ERS095088	-	-
J_S178	genital	J	Sweden	F	2011	endocervix	2566	ERR140829	ERS095099	-	-
J_S1254	genital	J	Sweden	F	2010	endocervix	2566	ERR108302	ERS082984	-	-
J_S3107	genital	J	Sweden	F	2010	endocervix	2566	ERR108299	ERS082981	-	-
J_S4281	genital	J	Sweden	F	2010	endocervix	2566	ERR108300	ERS082982	-	-
J_S4821	genital	J	Sweden	F	2010	endocervix	2566	ERR164649	ERS133236	-	-
J_SF5	genital	J	USA	F	2003	endocervix	2566				This study
J_SF6	genital	J	USA	F	2003	endocervix	2566				This study
J_Sou106	genital	J	UK	F	1985	unknown	2566	ERR211013	ERS161063	-	-
J_UK583676	genital	J	UK	M	2012	urethra	2566	ERR658657	ERS208384	-	-
J_UK913454	genital	J	UK	M	2012	urethra	2566	ERR658679	ERS208406	-	-
J_UW36	genital	J	USA	F	1971	endocervix	2566	ERR558504	ERS177783	-	-
K_Ar74	genital	K	Argentina	F	2005	ocular	2566	ERR658565	ERS208545		
K_Ar650	genital	K	Argentina	M	2006	urethra	2566	ERR658566	ERS208546	-	-
K_Fin128	genital	K	Finland	F	2010	endocervix	2566	ERR278186	ERS200115	-	-
K_Fin139	genital	K	Finland	F	2010	endocervix	2566	ERR278163	ERS200092	-	-
K_Fin202	genital	K	Finland	F	2011	endocervix	2566	ERR278164	ERS200093	-	-
K_Fin204	genital	K	Finland	F	2011	endocervix	2566	ERR278165	ERS200094	-	-
K_NL81	genital	K	Netherlands	F	2001	endocervix	2566	ERR211028	ERS161078	-	-
K_NL82	genital	K	Netherlands	F	2001	endocervix	2566	ERR211024	ERS161074	-	-
K_NL83	genital	K	Netherlands	F	2001	endocervix	2566	ERR211029	ERS161079	-	-
K_NL84	genital	K	Netherlands	F	2001	endocervix	2566	ERR211025	ERS161075	-	-
K_NL85	genital	K	Netherlands	F	2001	endocervix	2566	ERR210991	ERS161041	-	-
K_NL87	genital	K	Netherlands	F	2001	endocervix	2566	ERR211030	ERS161080	-	-
K_R2084	genital	K	Russia	F	2011	endocervix	2566	ERR111622	ERS082939	-	-
K_R11642	genital	K	Russia	F	2011	endocervix	2566	ERR140839	ERS095109	-	-
K_R13207	genital	K	Russia	F	2011	endocervix	2566	ERR140836	ERS095106	-	-
K_R14876	genital	K	Russia	M	2011	urethra	2566	ERR140838	ERS095108	-	-
K_R15212	genital	K	Russia	F	2011	endocervix	2566	ERR140840	ERS095110	-	-
K_R26881	genital	K	Russia	F	2010	endocervix	2566	ERR111626	ERS082943	-	-
K_R27128	genital	K	Russia	F	2010	endocervix	2563	ERR111629	ERS082946	-	-
K_R32840	genital	K	Russia	F	2010	endocervix	2566	ERR111627	ERS082944	-	-
K_R34345	genital	K	Russia	F	2010	endocervix	2566	ERR111609	ERS082926	-	-
K_R34962	genital	K	Russia	M	2010	urethra	2566	ERR111608	ERS082925	-	-
K_R35248	genital	K	Russia	M	2010	urethra	2566	ERR111628	ERS082945	-	-

K_S143	genital	K	Sweden	F	2011	endocervix	2566	ERR108304	ERS082986	-	-
K_S4034	genital	K	Sweden	F	2010	endocervix	2566	ERR164647	ERS133234	-	-
K_S4229	genital	K	Sweden	F	2010	urethra	2566	ERR111584	ERS075207	-	-
K_SotonK1	genital	K	UK	F	2009	endocervix	2566	ERR026559	ERS013799	HE603238	Harris 2012
K_UK583237	genital	K	UK	F	2012	Cx/Urethra	2566	ERR658600	ERS208327	-	-
K_UK66306(	genital	K	UK	F	2012	Cx/Urethra	2566	ERR658662	ERS208389	-	-
K_UK66312	genital	K	UK	F	2012	Cx/Urethra	2566	ERR658660	ERS208387	-	-
K_UK76907	genital	K	UK	F	2012	Urine	2566	ERR658561	ERS208541	-	-
K_UW31	genital	K	USA	F	1973	endocervix	2566	ERR558505	ERS177784	-	-
L1_L82	LGV	L1	South Africa	M	1985	urethra	2566	ERR211057	ERS161107	-	-
L1_L115p10	LGV	L1	South Africa	M	1986	urethra	2566	ERR211062	ERS161112	-	-
L1_L146	LGV	L1	South Africa	M	1986	ulcer	2566	ERR211044	ERS161094	-	-
L1_L165	LGV	L1	South Africa	M	1986	ulcer	2566	ERR211045	ERS161095	-	-
L1_L224	LGV	L1	South Africa	M	1986	urethra	2566	ERR211058	ERS161108	-	-
L1_L232	LGV	L1	South Africa	M	1987	urethra	2566	ERR211059	ERS161109	-	-
L1_L246	LGV	L1	South Africa	M	1987	urethra	2566	ERR211046	ERS161096	-	-
L1_L867	LGV	L1	South Africa	M	1993	urethra	2566	ERR211060	ERS161110	-	-
L1_L942	LGV	L1	South Africa	M	1994	urethra	2566	ERR211061	ERS161111	-	-
L1_L1034	LGV	L1	South Africa	M	1994	urethra	2566	ERR211035	ERS161085	-	-
L1_LGV98	LGV	L1	South Africa	M		unknown	2566	ERR071990	ERS066953	-	Seth-Smith 2013
L1_LGV913	LGV	L1	South Africa	M		unknown	2566	ERR071991	ERS066954	-	Seth-Smith 2013
L1_SA160	LGV	L1	South Africa	M	1986	ulcer	2566	ERR658551	ERS208531	-	-
L1_SA409	LGV	L1	South Africa	M	1990	ulcer	2566	ERR658553	ERS208533	-	-
L1_SABY21(	LGV	L1	South Africa	M	1999	ulcer	2566	ERR658549	ERS208529	-	-
L1_115	LGV	L1	South Africa			unknown	2566	ERR008593	ERS001411	HE603218	Harris 2012
L1_224	LGV	L1	South Africa			unknown	2566	ERR008580	ERS001404	HE603220	Harris 2012
L1_440	LGV	L1	USA	M	1968	lymph nod	2566	ERR008595	ERS001396		Harris 2012
L1_1333p2	LGV	L2	South Africa	M		genital ulc	2566			HE601951	
L1_Ur58380(	LGV	L1	UK	M	2012	Urine	2566	ERR658557	ERS208537	HE603228	Harris 2012
L2b_795	LGV	L2b	France	M	2004	rectum	2566	ERR008586	ERS001409		Harris 2012
L2b_8200	LGV	L2b	Sweden	M	2007	proctitis	2566	ERR021952	ERS004108		Harris 2012
L2b_8200-07	LGV	L2b	Sweden	M	2007	proctitis	2566	ERR021952	ERS004108		Harris 2012
L2b_Ams1	LGV	L2b	Netherlands	M	2004	penile ulce	2566			HE601959	Harris 2012
L2b_Ams2	LGV	L2b	Netherlands	M	2005	anus	2566			HE601961	Harris 2012
L2b_Ams3	LGV	L2b	Netherlands	M	2004	anus	2566			HE601962	Harris 2012
L2b_Ams4	LGV	L2b	Netherlands	M	2005	anus	2566			HE601964	Harris 2012
L2b_Ams5	LGV	L2b	Netherlands	M	2004	anus	2566			HE601965	Harris 2012
L2b_C1	LGV	L2b	Canada	M	2004	rectum	2566	ERR008579	ERS001398		Harris 2012
L2b_C2	LGV	L2b	Canada	M	2005	rectum	2566	ERR008592	ERS001399		Harris 2012
L2b_Canada1	LGV	L2b	Canada	M	2004	rectum	2566			HE601963	Harris 2012
L2b_Canada2	LGV	L2b	Canada	M	2005	rectum	2566			HE601957	Harris 2012

L2b_CC37	LGV	L2b	UK	M	2011	rectum	2566	ERR140847	ERS095117		Harris 2012
L2b_CS1908	LGV	L2b	Portugal	M	2015	anorectal	2566			CP009923	Borges 2015
L2b_CS7840	LGV	L2b	Portugal		2015	proctitis	2566			CP009925	Borges 2016
L2b_CV204	LGV	L2b	France	M	2006	rectum	2566	ERR019531	ERS003307		Harris 2012
L2b_H17IMS	LGV	L2b	UK	M	2008	rectum	2566	ERR140766	ERS095036		
L2b_HPA1	LGV	L2b	UK	M	2005	rectum	2566	ERR164659	ERS133246		
L2b_HPA21	LGV	L2b	UK	M	2009	rectum	2566	ERR140767	ERS095037		
L2b_HPA27	LGV	L2b	UK	M	2005	rectum	2566	ERR164661	ERS133248		
L2b_HPA29	LGV	L2b	UK	M	2004	rectum	2566	ERR164662	ERS133249		
L2b_HPA31	LGV	L2b	UK	M	2005	rectum	2566	ERR164663	ERS133250		Thomson 2008
L2b_HPA34	LGV	L2b	UK	M	2008	rectum	2566	ERR164665	ERS133252		Harris 2012
L2b_LST	LGV	L2b	France	M	2008	rectum	2566	ERR019528	ERS003315		Somboonna 2011
L2b_s11	LGV	L2b	Netherlands	M	2004	penile ulce	2566	ERR008590	ERS001408	HE603228	Harris 2012
L2b_s121	LGV	L2b	Netherlands	M	2005	anal swab	2566	ERR008597	ERS001402		
L2b_s300	LGV	L2b	Netherlands	M	2004	anal swab	2566	ERR008599	ERS001413		
L2b_s750	LGV	L2b	Netherlands	M	2004	anal swab	2566	ERR008594	ERS001410		
L2b_s906	LGV	L2b	Netherlands	M	2005	anal swab	2566	ERR008584	ERS001395		
L2b_SF4180	LGV	L2b	USA	M	1984	rectum	2566	ERR348840	ERS248048		
L2b_SF4644	LGV	L2b	USA	M	1985	rectum	2566	ERR348841	ERS248049		
L2b_SF1565	LGV	L2b	USA	M	2001	rectum	2566	ERR516390	ERS373084		
L2b_SF1567	LGV	L2b	USA	M	2003	rectum	2566	ERR516391	ERS373085		
L2b_SF1567	LGV	L2b	USA	M	2003	rectum	2566	ERR348858	ERS248080		
L2b_UCH1	LGV	L2b	UK	M	2006	proctitis	2566	ERR008581	ERS001407		
L2b_UCH2	LGV	L2b	UK			proctitis	2566	ERR008587	ERS001405		
L2c	LGV	L2	USA	M	2000s	rectum	2566			CP002024	Somboonna 2011
L2_L198	LGV	L2	South Africa	M	1986	ulcer	2566	ERR211050	ERS161100		Seth-Smith 2013
L2_L694	LGV	L2	South Africa	M	1993	urethra	2566	ERR211064	ERS161114		Seth-Smith 2014
L2_LGV173	LGV	L2	South Africa	M		NA	2566	ERR071989	ERS066952		Seth-Smith 2015
L2_SF25667	LGV	L2	USA	M	1981	rectum	2566	ERR351533	ERS248067		Seth-Smith 2016
L2_SF40369	LGV	L2	USA	M	1984	rectum	2566	ERR348839	ERS248047		Seth-Smith 2017
L2_SFUW39	LGV	L2	USA	M	1980s	rectum	2566	ERR348842	ERS248050		Seth-Smith 2018
L2_434Bu	LGV	L2	USA	M	1968	bubo	2566			AM884176	Harris 2012
L2_434Bu(f)	LGV	L2	USA	M	1968	inguinal bu	2566			CP003963	
L2_434Bu(i)	LGV	L2	USA	M	1968	inguinal bu	2566			CP003965	
L2_470LN87	LGV	L2	USA	M	1968	lymph nod	2566	ERR348847	ERS248055		Harris 2012
L2_514BU11	LGV	L2	USA	M	1968	bubo	2566	ERR348844	ERS248052		
L2_526BU5	LGV	L2	USA	M	1968	bubo	2566	ERR516392	ERS373086		Harris 2012
L3_404	LGV	L3	USA	M	1967	lymph nod	2576	ERR008583	ERS001416	HE601955	