



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}

^aDepartment of Pediatrics, University of California San Francisco, Oakland, California, USA

^bDepartment of Bioengineering, Joint Graduate Program, University of California San Francisco and University of California Berkeley, San Francisco, California, USA

^cDepartment of Medicine, University of California San Francisco, San Francisco, California, USA

^dBixby 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 (*trpR*) 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

Copyright © 2021 Bommana et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

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 *trpA* 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 strains 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

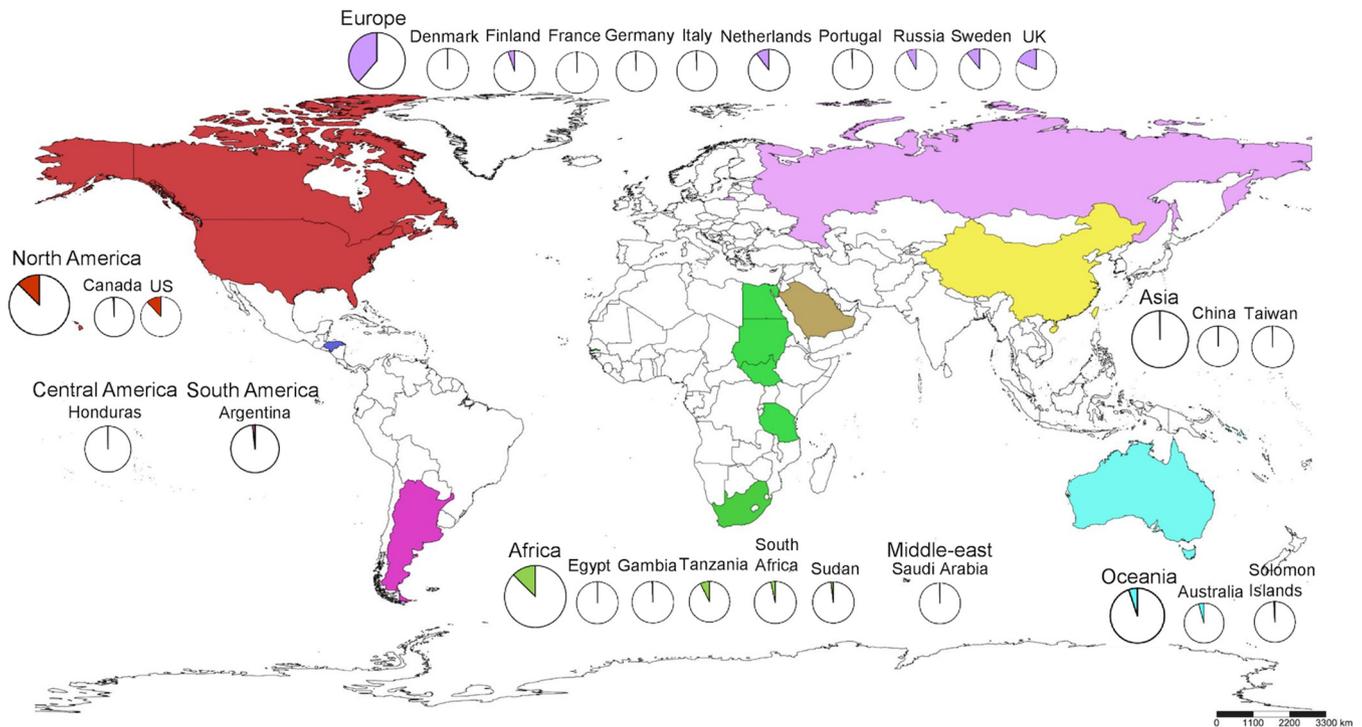


FIG 1 Global distribution of *Chlamydia trachomatis* strains harboring the tryptophan operon used in this study. Colors indicate the geographic areas or continents that were the sources of the samples as shown in the key. Circles within countries show the number of samples from that country. Countries 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 SimpleMapp.

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 *trpA* 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 *trpA* 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 lineages with prevalent, nonprevalent, and mixed urogenital clades with greater diversity for *trpA*. Tryptophan operon sequences consisting of *trpR*, intergenic region (IGR), *trpB*, and *trpA*, in that order, were available for 595 clinical and reference samples; 12 ocular

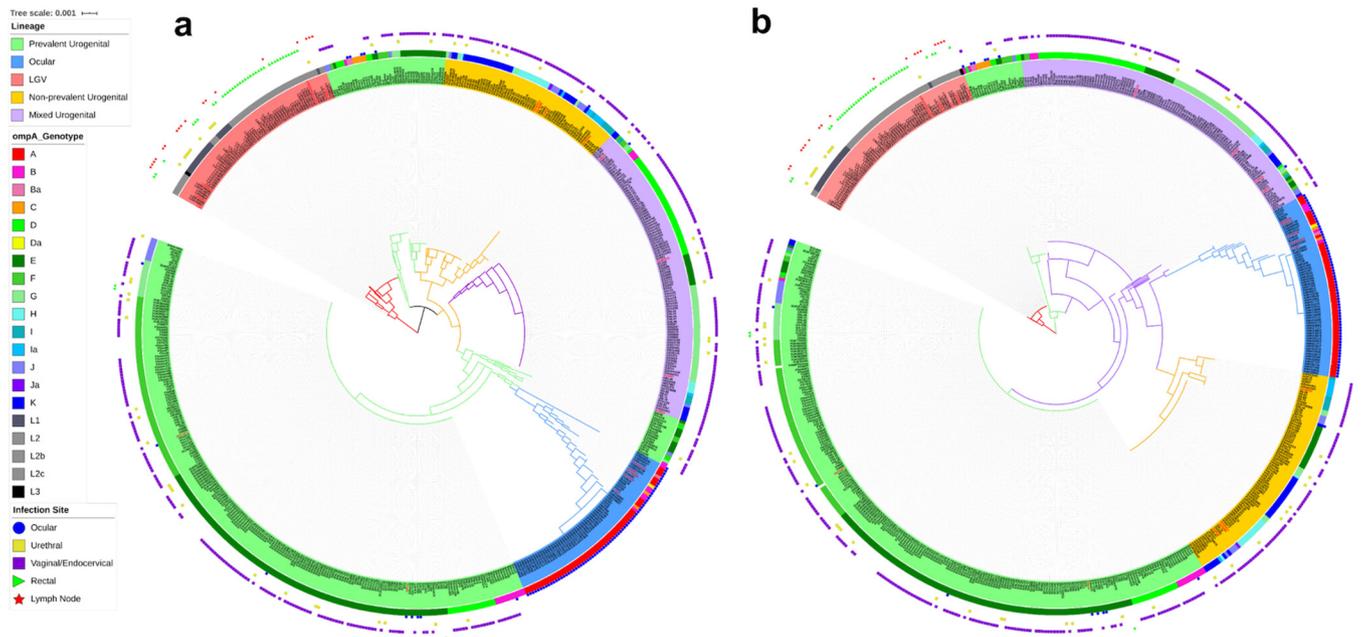


FIG 2 Interactive phylogenetic trees for *trpRBA* (a) (see <https://itol.embl.de/tree/209360135466881582240570>) and *trpA* (b) (see <https://itol.embl.de/tree/7616769102203851589927158>) depicting *Ct* lineages: LGV, nonprevalent and prevalent urogenital, mixed urogenital, and a more recent ocular subclade. The relationship between these lineages with their *ompA* 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 *trpRBA* and *trpA* sequences for each of the 595 strains in the data set (see Materials and Methods). Both were executed in Geneious. *ompA* 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, non-prevalent, 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/7616769102203851589927158>). 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 *ompA* genotypes with urogenital backbones isolated from the eyes of Aboriginal children (18) and urogenital *ompA* genotypes with urogenital backgrounds isolated from the eyes of patients in Denmark and Argentina (Fig. 2a; blue dots at ~11 o'clock; Table S1).

trpA 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/7616769102203851589927158>). 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, *trpA* 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 *ompA* genotype and anatomic origin.

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

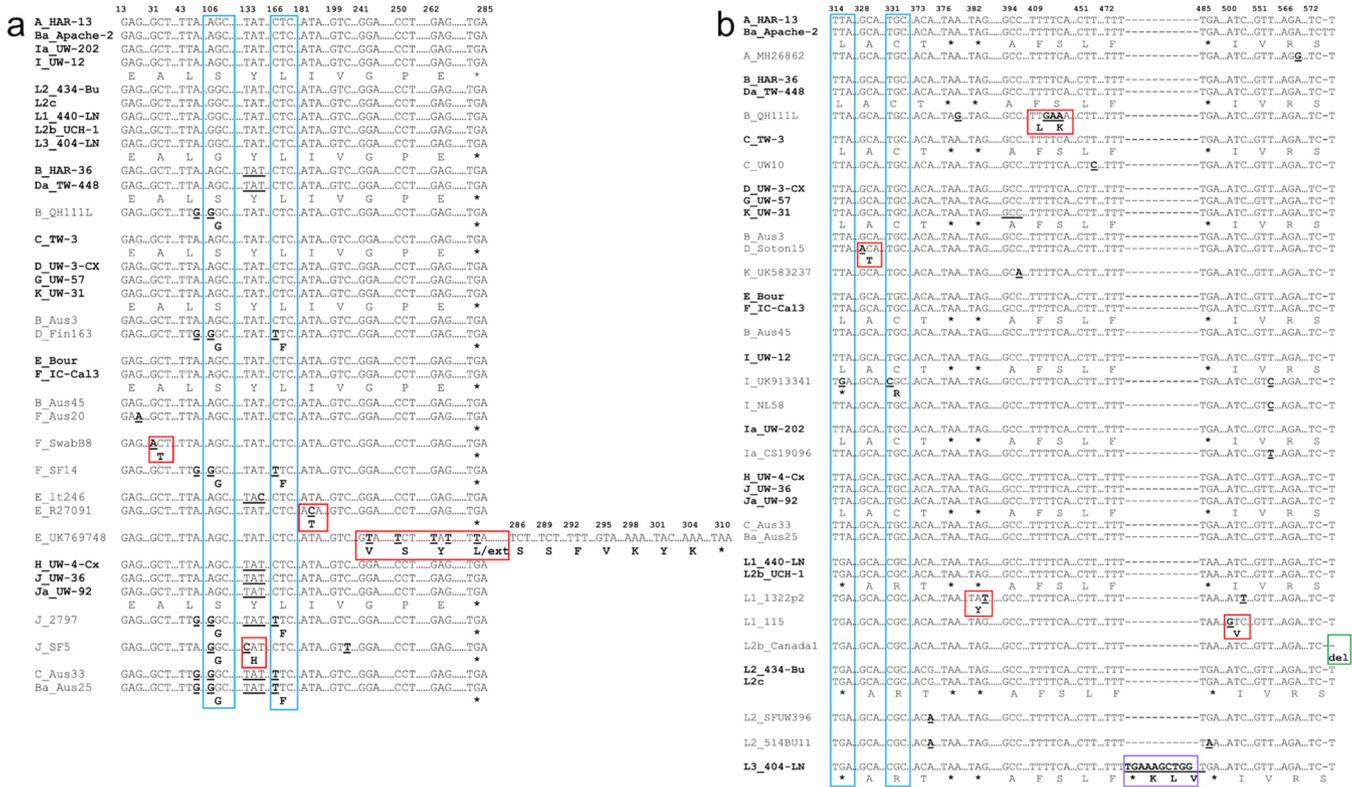


FIG 3 Partial nucleotide sequences of *trpR* (a) and IGR (b) with regions of single nucleotide polymorphism (SNPs) and indels in comparison to the 21 Ct reference strain sequences (in bold). Australian aboriginal strains Ba_Aus25, B_Aus3, B_Aus45, and C_Aus33 were included with the urogenital strains 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 *trpB*, 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.

trpA 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_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

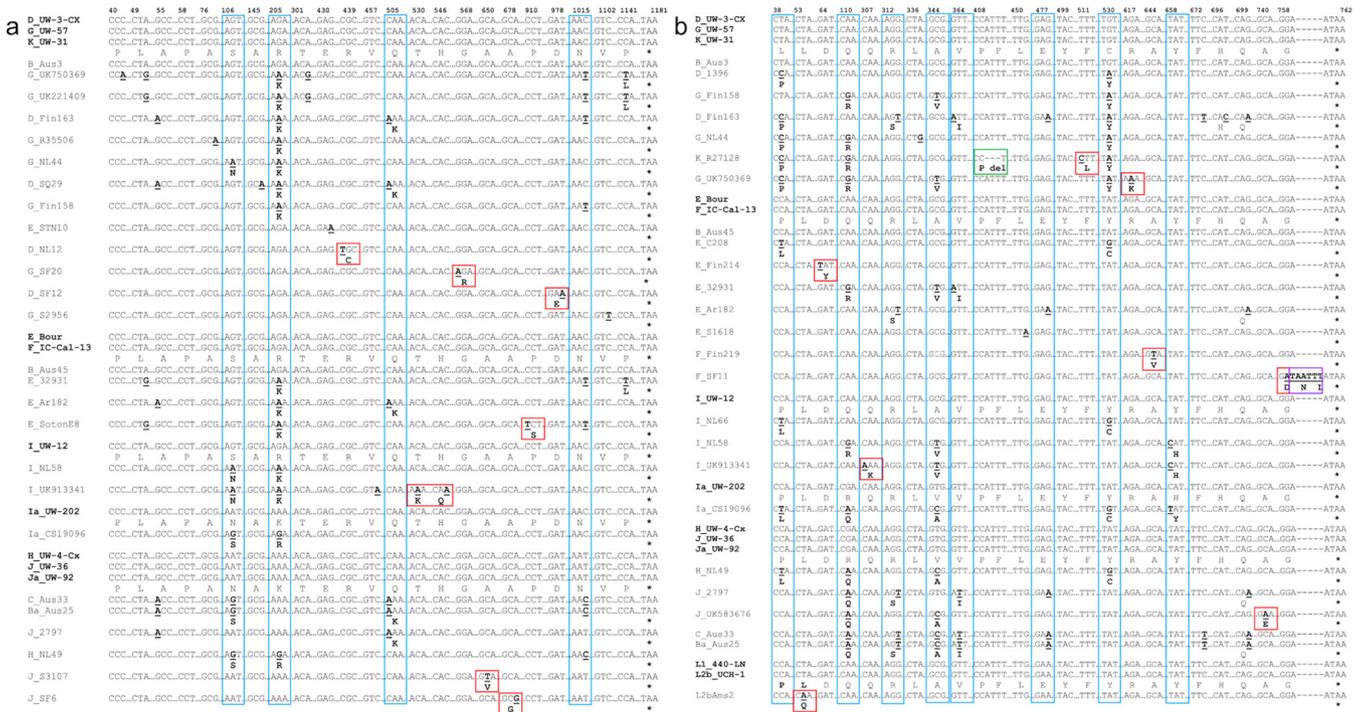


FIG 4 Partial nucleotide sequences of *trpB* (a) and *trpA* (b) urogenital strains with regions of single nucleotide polymorphism (SNPs) and indels in comparison to the 21 Ct reference strain sequences (in bold). Australian aboriginal strains Ba_Aus25, B_Aus3, B_Aus45, and C_Aus33 were included with 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

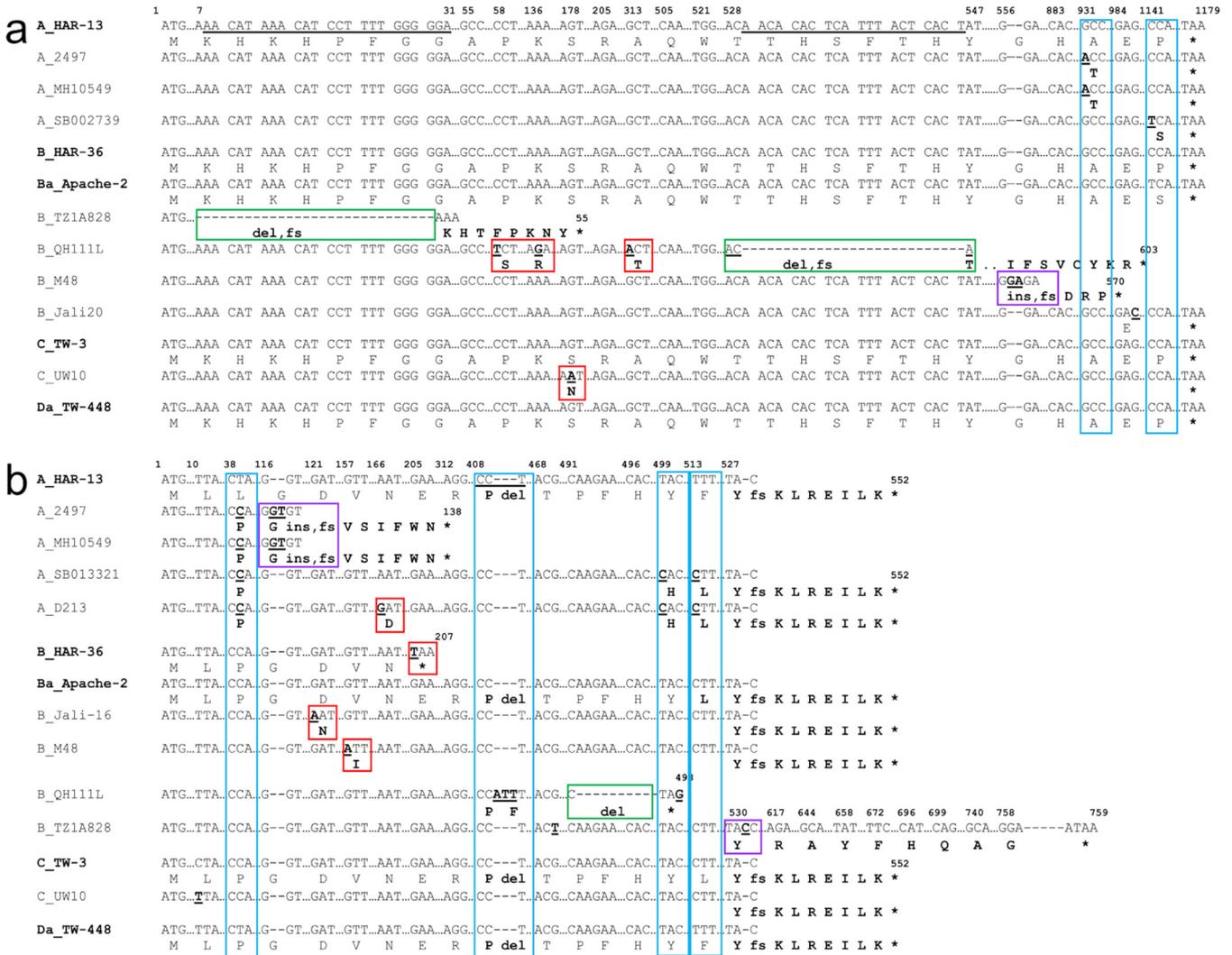


FIG 5 Partial nucleotide sequences of *trpB* (a) and *trpA* (b) ocular strains with regions of single nucleotide polymorphism (SNPs) and indels in comparison to the 21 *Ct* reference strain sequences (in bold). Australian aboriginal strains Ba_Aus25, B_Aus3, B_Aus45, and C_Aus33 were included with the 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 *trpBA*. 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
trpR	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
trpB	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
trpA	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

^aNote 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).

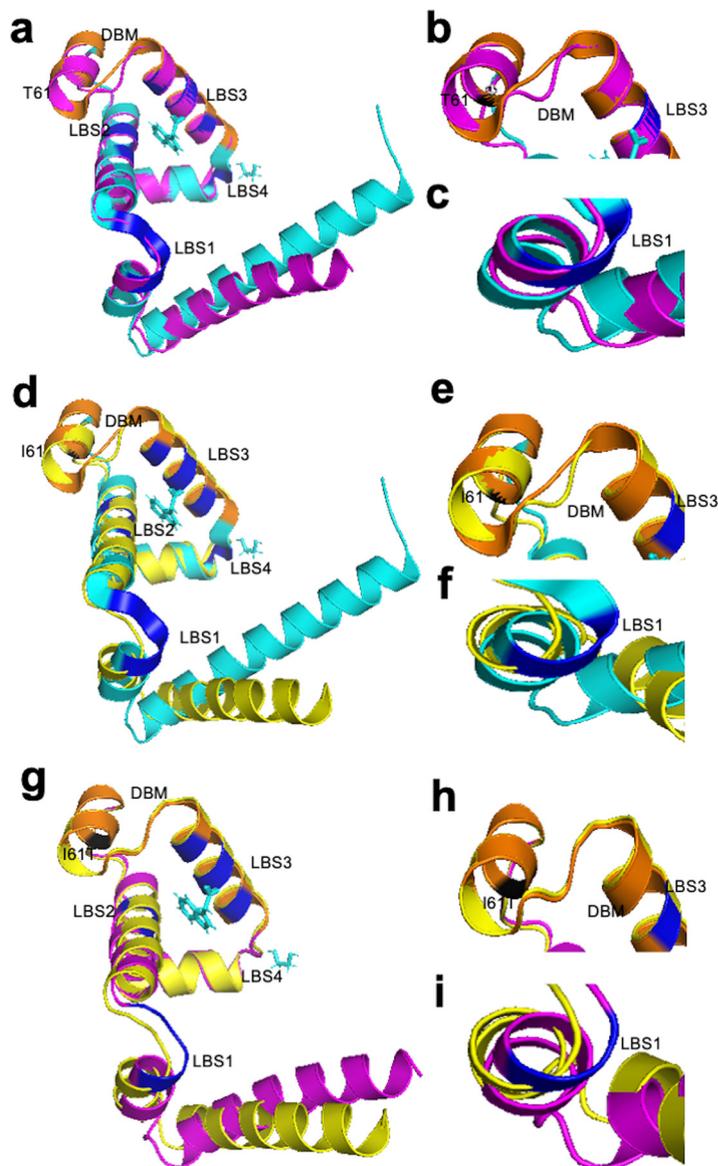


FIG 6 TrpR 3D predicted structures of *Ct* clinical E_R27091 and reference E_Bour strains (a to i). (a) TrpR structure of mutant strain E_R27091 (magenta) superimposed on the template 6eniA (cyan), with DNA-binding motif (DBM) in orange. The putative ligand-binding sites (LBS) 1, 2, and 3 are shown in 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

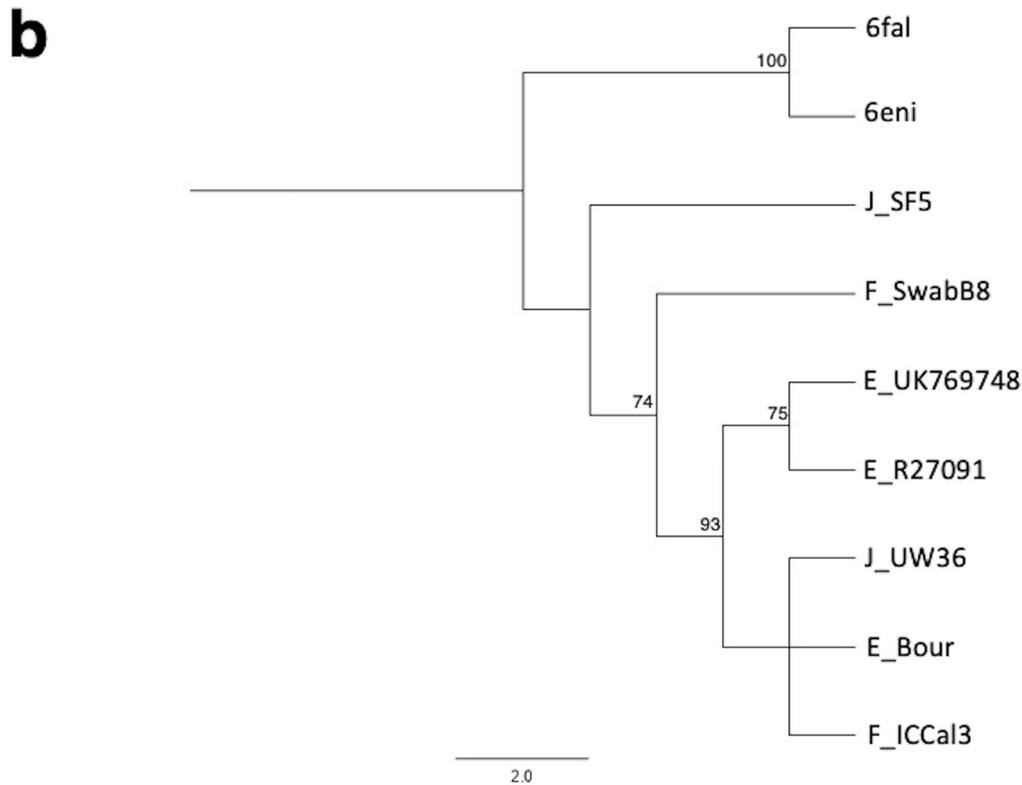
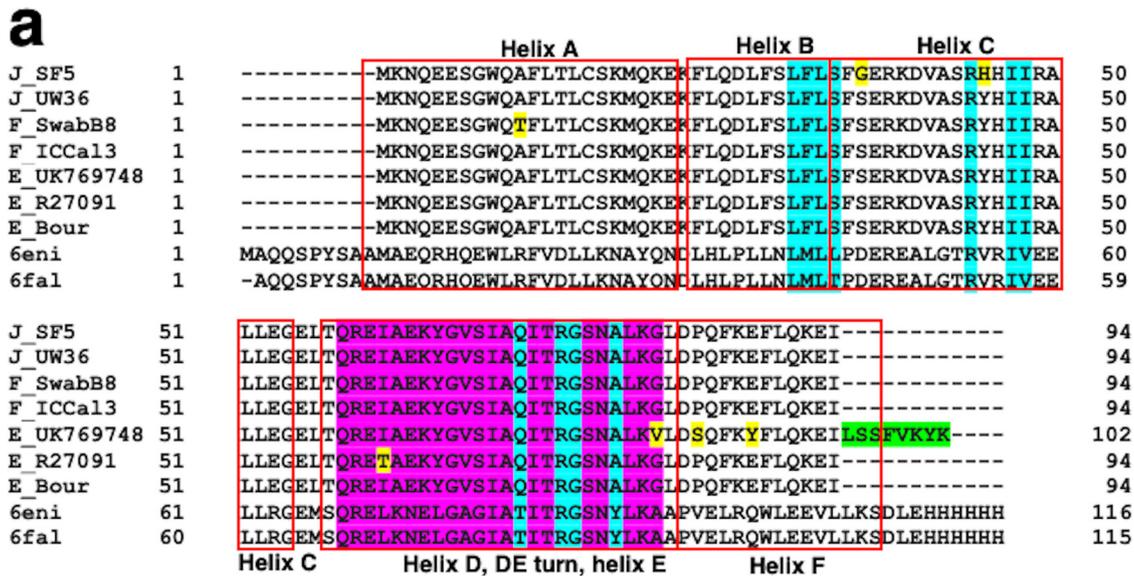


FIG 7 TrpR amino acid alignment (a) and phylogeny (b) of *Ct* mutant strains (E_R27091, E_UK769748, F_SwabB8, and J_SF5), reference strains (E_Bour, F_IC-Cal-3, and J_UW-36) and Protein Data Bank (PDB) templates (6fal and 6eni). (a) TrpR aa alignments were created using Clustal Omega, and the residues 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 α 176 to α 196

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

^aSee Fig. 5 for location and position of the mutation in the TrpR amino acid sequence.

^bSee Fig. 6 for location and position of the mutation in the TrpB amino acid sequence.

^cSee Fig. 9 for location and position of the mutation in the TrpA amino acid sequence.

^dAbbreviations: 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 β P304S (MVC-binding site), I_UK913341 β T177K, and β H182Q (β -COMM domain with β T177K proximal to residues interacting with α -subunit) in Fig. 9a to c and Fig. S5A to C but not for D_NL12 β R147C 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⁺,

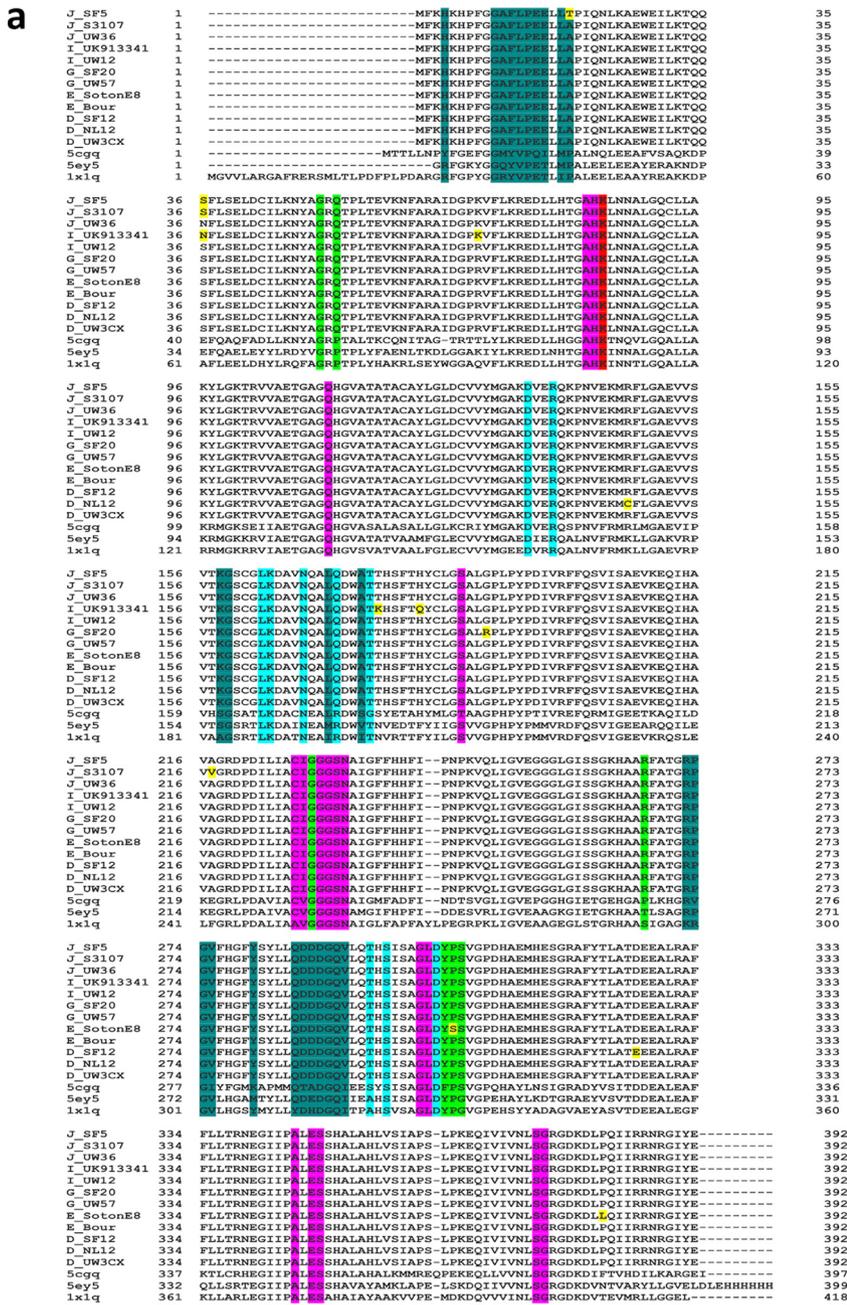


FIG 8 TrpB amino acid alignment (a) and phylogeny (b) of *Ct* mutant strains (D_NL12, D_SF12, E_SotonE8, G_SF20, I_UK913341, J_S3107, and J_SF5), reference strains (D_UW3CX, E_Bour, G_UW57, LKARLEGIIPAL, ESHALAHVLSIAPS, LPKEQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 392 J_S3107 392 FLLTRNEGIIIPAL ESHALAHVLSIAPS LPKEQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 392 J_UW36 392 FLLTRNEGIIIPAL ESHALAHVLSIAPS LPKEQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 392 I_UK913341 392 FLLTRNEGIIIPAL ESHALAHVLSIAPS LPKEQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 392 G_SF20 392 FLLTRNEGIIIPAL ESHALAHVLSIAPS LPKEQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 392 G_UW57 392 FLLTRNEGIIIPAL ESHALAHVLSIAPS LPKEQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 392 E_SotonE8 392 FLLTRNEGIIIPAL ESHALAHVLSIAPS LPKEQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 392 E_Bour 392 FLLTRNEGIIIPAL ESHALAHVLSIAPS LPKEQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 392 D_SF12 392 FLLTRNEGIIIPAL ESHALAHVLSIAPS LPKEQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 392 D_NL12 392 FLLTRNEGIIIPAL ESHALAHVLSIAPS LPKEQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 392 G_UW3CX 392 FLLTRNEGIIIPAL ESHALAHVLSIAPS LPKEQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 392 S_Cgq 397 KTLCRNEGIIIPAL ESHALAHVLSIAPS LPKEQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 397 5ey5 392 QLLSRTNEGIIIPAL ESHAVAYAMKLAPE LSKDQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 399 1x1q 361 LKLARLEGIIPAL ESHALAHVLSIAPS LPKEQIVIVNLSGRGDKDLQPI IRRNRGIYE----- 418

(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 α C177Y and α R206K 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 α E178Q 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 α Q37R, which had an increased/neutral effect on protein stability, and α G253D, 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) TrpA 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.

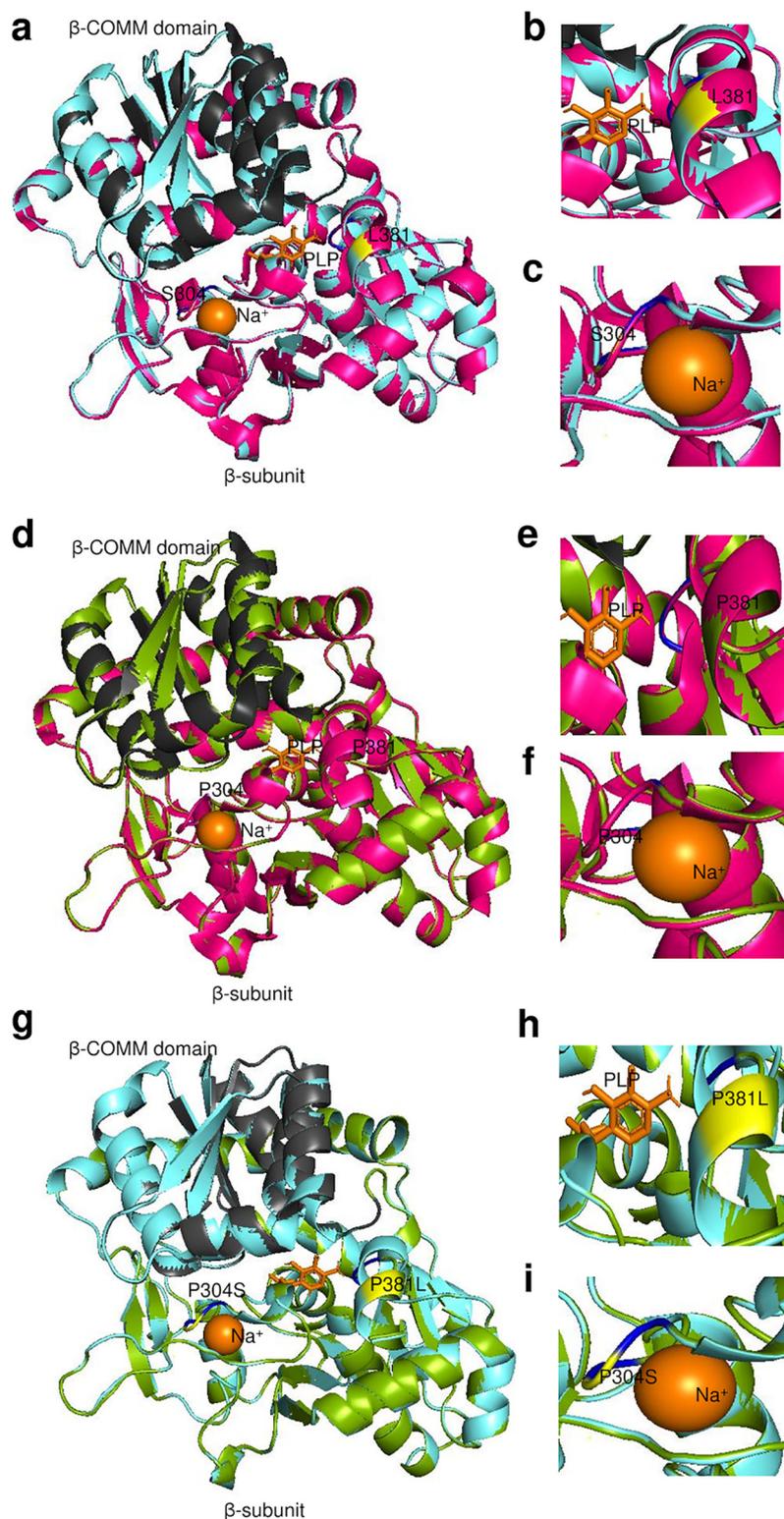


FIG 9 TrpB 3D predicted structures of Cr clinical E_SotonE8 and reference E_Bour strains (a to i). (a) TrpB structure of mutant strain E_SotonE8 (cyan) superimposed on the template 5ey5 (magenta), with the β -COMM domain in charcoal gray. Na^+ and the cofactor pyridoxal phosphate (PLP) interacting residues 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 on E_Bour (g) with structural changes noted in relation to PLP interacting residue on P381L (h) and Na^+ interacting residue on P304S (i).

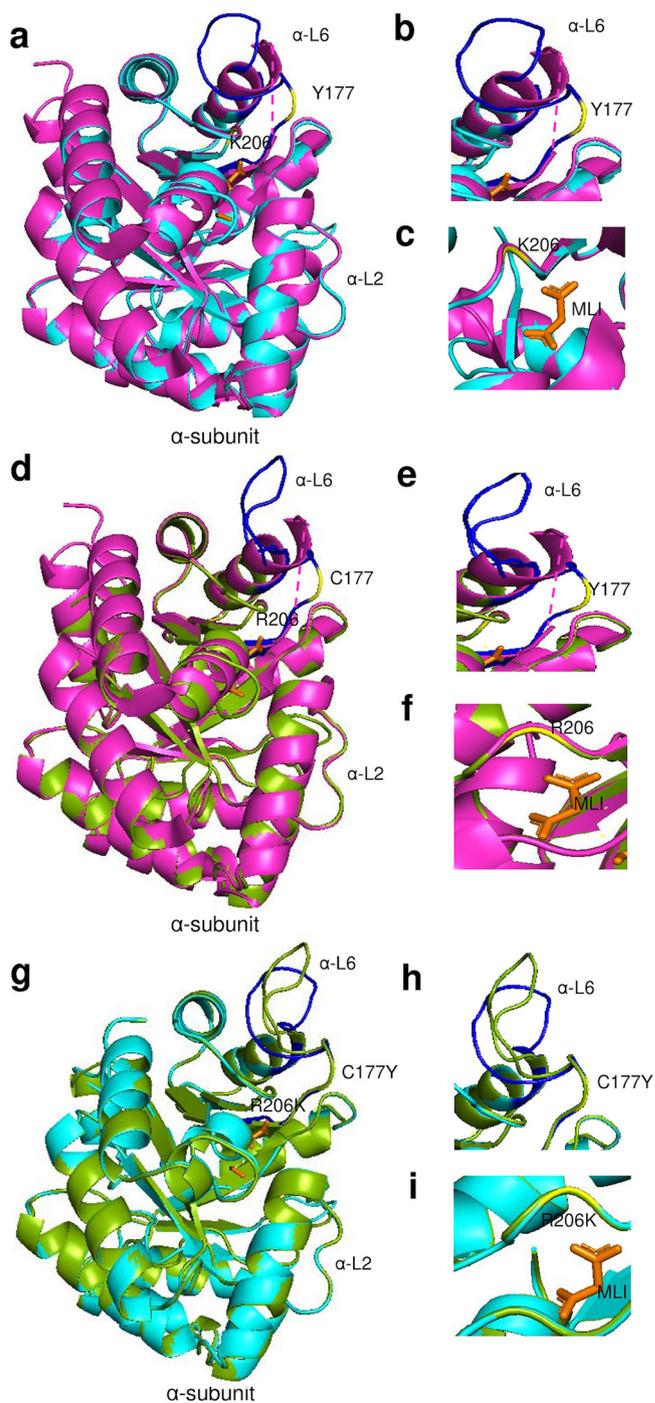


FIG 10 TrpA 3D predicted structures of *Ct* clinical G_UK750369 and reference G_UW57 strains (a to i). (a) TrpA structure of mutant strain G_UK750369 (cyan) superimposed on the template 5tch (magenta), with α -Loop 6 (α -L6) of the α -subunit in blue and malonate ion (MLI) in orange. The aa substitution 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).

a

L2b_Ams2	1	-----MSKLTQVF-----KQTKPCIGVLTAGDGGTSYTI EAAKALIQGGVDIL ELGF	47
L2b_UCH1	1	-----MSKLTQVF-----KQTKPCIGVLTAGDGGTSYTI EAAKALIQGGVDIL ELGF	47
J_UK583676	1	-----MSKLTQVF-----KQTKPCIGVLTAGDGGTSYTI EAAKALIQGGVDIL ELGF	47
J_UW36	1	-----MSKLTQVF-----KQTKPCIGVLTAGDGGTSYTI EAAKALIRGGVDIL ELGF	47
I_UK913341	1	-----MSKLTQVF-----KQTKPCIGVLTAGDGGTSYTI EAAKALIRGGVDIL ELGF	47
I_UW12	1	-----MSKLTQVF-----KQTKPCIGVLTAGDGGTSYTI EAAKALIQGGVDIL ELGF	47
G_UK750369	1	-----MSKLTQVF-----KQTKPCIGVLTAGDGGTSYTI EAAKALIRGGVDIL ELGF	47
G_UW57	1	-----MSKLTQVF-----KQTKPCIGVLTAGDGGTSYTI EAAKALIQGGVDIL ELGF	47
F_SF11	1	-----MSKLTQVF-----KQTKPCIGVLTAGDGGTSYTI EAAKALIQGGVDIL ELGF	47
F_Fin219	1	-----MSKLTQVF-----KQTKPCIGVLTAGDGGTSYTI EAAKALIQGGVDIL ELGF	47
F_ICCa13	1	-----MSKLTQVF-----KQTKPCIGVLTAGDGGTSYTI EAAKALIQGGVDIL ELGF	47
E_Fin214	1	-----MSKLTQVF-----KQTKPCIGVLTAGDGGTSYTI EAAKALIQGGVDIL ELGF	47
E_Bour	1	-----MSKLTQVF-----KQTKPCIGVLTAGDGGTSYTI EAAKALIQGGVDIL ELGF	47
5tch	1	MVAVEQSEASRLGPVDFSCRANRRAALGLPTGYPDVFPASVAAMTALVESGDCII EIVGV	60
2ekc	1	-----MGRISDKFTLKKKREKALVSLMVGYPDYETSLKAFKVELKNGDIL ELGF	52
5ey5	1	-----MNRIAEAFELKKKGEKALVSLMVGYPDYETSLKAFKVELKNGDIL ELGF	52
L2b_Ams2	48	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	107
L2b_UCH1	48	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	107
J_UK583676	48	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	107
J_UW36	48	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	107
I_UK913341	48	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	107
I_UW12	48	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	107
G_UK750369	48	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	107
G_UW57	48	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	107
F_SF11	48	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	107
F_Fin219	48	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	107
F_ICCa13	48	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	107
E_Fin214	48	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	107
E_Bour	48	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	107
5tch	61	PYSDPVMGPTIARATEAALRGVRRVDTLAAVEA ^I S ^I AGGRA ⁻ -VVMYWN ^P VLRYG ^V D	118
2ekc	53	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	112
5ey5	53	PFSDPVA ^N PEIQVSHDRALAE ^N LTSETLLEIVEGIRAFNQE ^V PLLILSYN ^N PLLRD ^L	112
L2b_Ams2	108	-YLRRKLDAGINGVCVIDLPAPLSHGEEKS ^F FFEDLLAVGLD ^P ILLISAGT ^T PERMSLIQ ^E	166
L2b_UCH1	108	-YLRRKLDAGINGVCVIDLPAPLSHGEEKS ^F FFEDLLAVGLD ^P ILLISAGT ^T PERMSLIQ ^E	166
J_UK583676	108	-YLRRKLDAGINGVCVIDLPAPLSHGEEKS ^F FFEDLLAVGLD ^P ILLISAGT ^T PERMSLIQ ^E	166
J_UW36	108	-YLRRKLDAGINGVCVIDLPAPLSHGEEKS ^F FFEDLLAVGLD ^P ILLISAGT ^T PERMSLIQ ^E	166
I_UK913341	108	-YLRRKLDAGINGVCVIDLPAPLSHGEEKS ^F FFEDLLAVGLD ^P ILLISAGT ^T PERMSLIQ ^E	166
I_UW12	108	-YLRRKLDAGINGVCVIDLPAPLSHGEEKS ^F FFEDLLAVGLD ^P ILLISAGT ^T PERMSLIQ ^E	166
G_UK750369	108	-YLRRKLDAGINGVCVIDLPAPLSHGEEKS ^F FFEDLLAVGLD ^P ILLISAGT ^T PERMSLIQ ^E	166
G_UW57	108	-YLRRKLDAGINGVCVIDLPAPLSHGEEKS ^F FFEDLLAVGLD ^P ILLISAGT ^T PERMSLIQ ^E	166
F_SF11	108	-YLRRKLDAGINGVCVIDLPAPLSHGEEKS ^F FFEDLLAVGLD ^P ILLISAGT ^T PERMSLIQ ^E	166
F_Fin219	108	-YLRRKLDAGINGVCVIDLPAPLSHGEEKS ^F FFEDLLAVGLD ^P ILLISAGT ^T PERMSLIQ ^E	166
F_ICCa13	108	-YLRRKLDAGINGVCVIDLPAPLSHGEEKS ^F FFEDLLAVGLD ^P ILLISAGT ^T PERMSLIQ ^E	166
E_Fin214	108	-YLRRKLDAGINGVCVIDLPAPLSHGEEKS ^F FFEDLLAVGLD ^P ILLISAGT ^T PERMSLIQ ^E	166
E_Bour	108	-YLRRKLDAGINGVCVIDLPAPLSHGEEKS ^F FFEDLLAVGLD ^P ILLISAGT ^T PERMSLIQ ^E	166
5tch	119	AFARD ^L AAAGGLGLIT ^P DLIPD ⁻ ---EAQW ^L AA ^E EHRLDR ^I FLVAP ^S ST ^T PERLAAT ^V E	174
2ekc	113	KFCRL ^S REKIGD ^G IVDPD ^L LPPE ⁻ ---EAE ^L KA ^V MK ^K YV ^L SF ^V PLGAP ^T STR ^K IK ^L ICE	168
5ey5	113	RFVKE ^C AEAGV ^D GLIVDPD ^L LPPE ⁻ ---EAAD ^L AAA ^A E ^E YGV ^D LFLVAP ^T ST ^T DERIK ^M IAK	168
L2b_Ams2	167	YARGFL ^Y YI ^P YEA ^T RDSEVGI ^K EE ⁻ ---FRK ^V REH ^F DL ^P IVDR ^R IC ^D CK ^K EA ^A HV ^L NY ^S	221
L2b_UCH1	167	YARGFL ^Y YI ^P YEA ^T RDSEVGI ^K EE ⁻ ---FRK ^V REH ^F DL ^P IVDR ^R IC ^D CK ^K EA ^A HV ^L NY ^S	221
J_UK583676	167	YARGFL ^Y YI ^P YQA ^T RDSEVGI ^K EE ⁻ ---FRK ^V REH ^F DL ^P IVDR ^R IC ^D CK ^K EA ^A HV ^L NY ^S	221
J_UW36	167	YARGFL ^Y YI ^P YQA ^T RDSEVGI ^K EE ⁻ ---FRK ^V REH ^F DL ^P IVDR ^R IC ^D CK ^K EA ^A HV ^L NY ^S	221
I_UK913341	167	YARGFL ^Y YI ^P YQA ^T RDSEVGI ^K EE ⁻ ---FRK ^V REH ^F DL ^P IVDR ^R IC ^D CK ^K EA ^A HV ^L NH ^S	221
I_UW12	167	YARGFL ^Y YI ^P YEA ^T RDSEVGI ^K EE ⁻ ---FRK ^V REH ^F DL ^P IVDR ^R IC ^D CK ^K EA ^A HV ^L NY ^S	221
G_UK750369	167	YARGFL ^Y YI ^P YQA ^T RDSEVGI ^K EE ⁻ ---FRK ^V REH ^F DL ^P IVDR ^R IC ^D CK ^K EA ^A HV ^L NY ^S	221
G_UW57	167	YARGFL ^Y YI ^P YQA ^T RDSEVGI ^K EE ⁻ ---FRK ^V REH ^F DL ^P IVDR ^R IC ^D CK ^K EA ^A HV ^L NY ^S	221
F_SF11	167	YARGFL ^Y YI ^P YQA ^T RDSEVGI ^K EE ⁻ ---FRK ^V REH ^F DL ^P IVDR ^R IC ^D CK ^K EA ^A HV ^L NY ^S	221
F_Fin219	167	YARGFL ^Y YI ^P YQA ^T RDSEVGI ^K EE ⁻ ---FRK ^V REH ^F DL ^P IVDR ^R IC ^D CK ^K EA ^A HV ^L NY ^S	221
F_ICCa13	167	YARGFL ^Y YI ^P YQA ^T RDSEVGI ^K EE ⁻ ---FRK ^V REH ^F DL ^P IVDR ^R IC ^D CK ^K EA ^A HV ^L NY ^S	221
E_Fin214	167	YARGFL ^Y YI ^P YQA ^T RDSEVGI ^K EE ⁻ ---FRK ^V REH ^F DL ^P IVDR ^R IC ^D CK ^K EA ^A HV ^L NY ^S	221
E_Bour	167	YARGFL ^Y YI ^P YQA ^T RDSEVGI ^K EE ⁻ ---FRK ^V REH ^F DL ^P IVDR ^R IC ^D CK ^K EA ^A HV ^L NY ^S	221
5tch	175	ASRG ^F V ^A AST ^M GV ^N CARD ^A VSQA ⁻ AP ^E L ^V GR ^V RV ^A SD ^I PV ^G GL ^V SR ^A QA ^Q IA ^Q YA ^S	233
2ekc	169	AAD ^E MT ^F Y ^V SV ^T GT ^T GARE ^K LPYER ^I KK ^K VEY ^R EL ^C DK ^P VV ^V GV ^G SK ^K EHARE ^I GS ^F A	228
5ey5	169	HAS ^F V ^Y CV ^S VT ^G V ^T GARSE ^T AAD ⁻ LAEL ^V SR ^I RK ^H TDL ^P IAV ^G FG ^I ST ^T PEQA ^A EVA ^Q VA	227
L2b_Ams2	222	DGFIVK ^T AFV ^H Q ^T MDSSVET ^L TALAQ ^T VI ^P G ⁻ -----	253
L2b_UCH1	222	DGFIVK ^T AFV ^H Q ^T MDSSVET ^L TALAQ ^T VI ^P G ⁻ -----	253
J_UK583676	222	DGFIVK ^T AFV ^H Q ^T MDSSVET ^L TALAQ ^T VI ^P G ⁻ -----	253
J_UW36	222	DGFIVK ^T AFV ^H Q ^T MDSSVET ^L TALAQ ^T VI ^P G ⁻ -----	253
I_UK913341	222	DGFIVK ^T AFV ^H Q ^T MDSSVET ^L TALAQ ^T VI ^P G ⁻ -----	253
I_UW12	222	DGFIVK ^T AFV ^H Q ^T MDSSVET ^L TALAQ ^T VI ^P G ⁻ -----	253
G_UK750369	222	DGFIVK ^T AFV ^H Q ^T MDSSVET ^L TALAQ ^T VI ^P G ⁻ -----	253
G_UW57	222	DGFIVK ^T AFV ^H Q ^T MDSSVET ^L TALAQ ^T VI ^P G ⁻ -----	253
F_SF11	222	DGFIVK ^T AFV ^H Q ^T MDSSVET ^L TALAQ ^T VI ^P G ⁻ -----	255
F_Fin219	222	DGFIVK ^T AFV ^H Q ^T MDSSVET ^L TALAQ ^T VI ^P G ⁻ -----	253
F_ICCa13	222	DGFIVK ^T AFV ^H Q ^T MDSSVET ^L TALAQ ^T VI ^P G ⁻ -----	253
E_Fin214	222	DGFIVK ^T AFV ^H Q ^T MDSSVET ^L TALAQ ^T VI ^P G ⁻ -----	253
E_Bour	222	DGFIVK ^T AFV ^H Q ^T MDSSVET ^L TALAQ ^T VI ^P G ⁻ -----	253
5tch	234	DGVVGS ^A L ^V -TAL ^T EG ^L FR ^I AL ^T GE ^L AG ^V RL ^G MSA ^H HHHH ^H	276
2ekc	229	DGVVGS ^A L ^V -KLAG ^Q K ^I ED ^L GN ^L V ^K EL ^R EGL ^R E ⁻ -----	262
5ey5	228	DGVVGS ^A I ^V K ^R I ^E N ^Q DE ^E D ^I VE ^V RE ^F VR ^E L ^R EAV ^K LE ^H HHHH ^H 273	

b

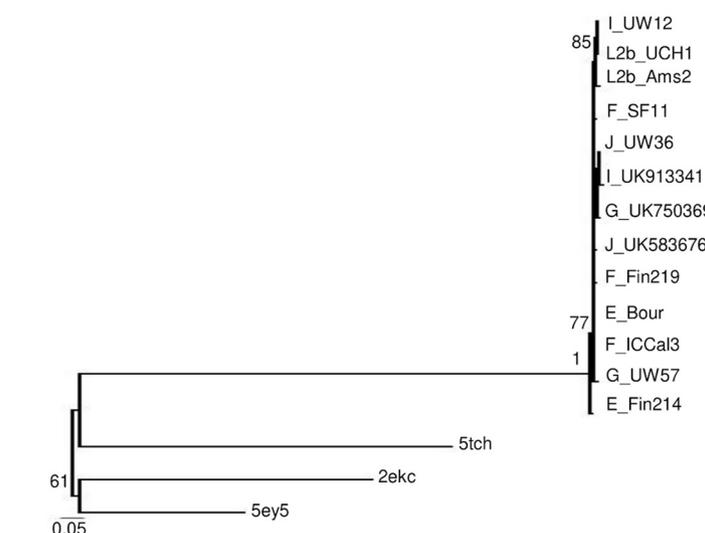


FIG 11 TrpA amino acid alignment (a) and phylogeny (b) of *Ct* mutant strains (E_Fin214, F_Fin219, F_SF11, G_UK750369, I_UK913341, J_UK583676, and L2b_Ams2), reference strains (E_Bour, F_ICCa13, (Continued on next page)

(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 *trpR*, *trpAa*-*Ab*, *trpB*, *trpD*, *trpC*, *trpEb*-*Ea*, and *aroA* 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 *trpR* region and IGR as well as *trpB* and *trpA*. 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 *trpA* 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 α 176 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.

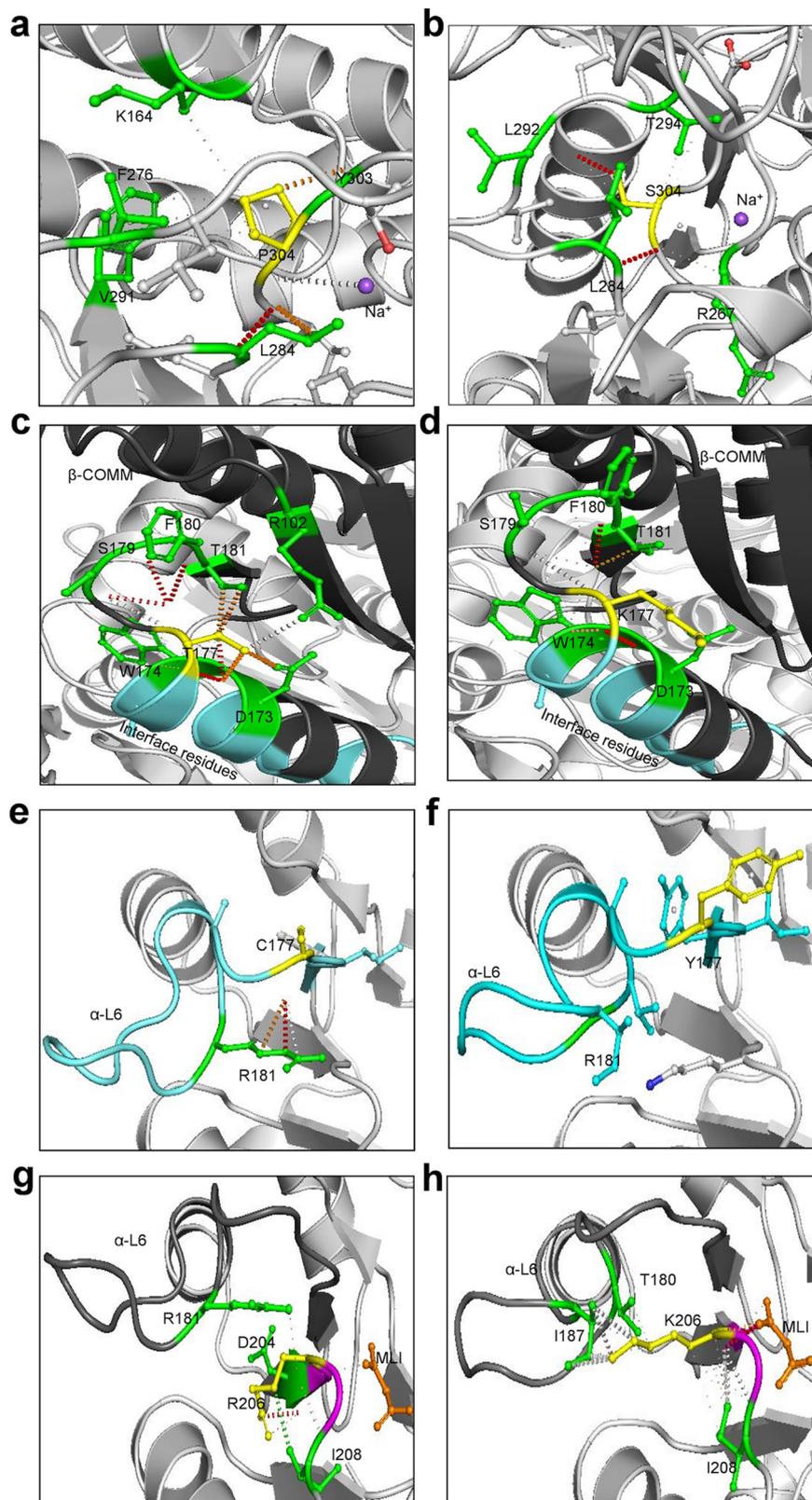


FIG 12 Effect of wild-type and mutant residues on interatomic interactions in TrpA and TrpB protein structures and small-molecule ligand binding. (a and b) The wild-type residue P304 (yellow) of E_Bour β -subunit (a) and the mutant residue S304 (yellow) of E_SotonE8 β -subunit (b) interactions with neighboring 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_UW12 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 *trpR*. 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 *trpB* and *trpA* expression during iron starvation. Simultaneously, YtgR binding promotes termination of transcripts from the primary promoter upstream of TrpR29. Although the B_QH111L *trpBA* 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 *trpB* 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 *trpR* mutations.

trpA 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 *trpA* 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 *trpA* 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 $P_i(a)/P_i(s)$ 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 *trpBA* genes is tightly regulated by TrpR. When in contact with its corepressor tryptophan, TrpR prevents *trpBA* transcription by binding to the *trp* operator-DNA (48). Tryptophan starvation, in contrast, induces transcription of *trpBA* 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 *trpBA* 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 *trpBA* 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. β P304S—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 α C177Y and I_UK913341 α E178Q

mutations located within α -L6 and G_UK750369 α R206K in the catalytic and subunit interaction domain exhibited reduced ligand affinity for both with reduced protein stability and energy expenditure. Only the α Q37R mutation in G_UK750369 and I_UK913341 predicted an increased protein stability with a lower energy cost, suggesting greater fitness (Table 2). Interestingly, the α C177Y 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 non-functional 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. *trpR*, *trpA*, and *trpB* 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 *trpR*, intergenic region (IGR), *trpB*, and *trpA*, in that order, were extracted in Geneious after checking operon mapped reads that were at a minimum coverage depth of 10 \times . 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, L₁_440, L₂_434, L_{2a}_UW396, L_{2b}_UCH1proctitis, L_{2c}, and L₃_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), SignallP (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 $\Delta G = -RT \ln K_d$.

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

- WHO. 2018. Report on global sexually transmitted infection surveillance. World Health Organization, Geneva, Switzerland.
- Bennett JE, Dolin R, Blaser MJ. 2016. Douglas and Bennett's infectious diseases essentials. Elsevier Inc, Philadelphia, PA.
- de Vriese 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>.
- 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, Morr  SA, Speksnijder A, B b ar 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>.
- WHO. 2019. Trachoma epidemiologic situation. World Health Organization, Geneva, Switzerland.
- 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.
- Millman K, Black CM, Johnson RE, Stamm WE, Jones RB, Hook EW, Martin DH, Bolan G, Tavar  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>.
- 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>.
- Herrmann B, Isaksson J, Ryberg M, T ngrot J, Saleh I, Versteeg B, Gravingen 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>.
- 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>.
- Joseph SJ, Didelot X, Rothschild J, de Vries HJC, Morr  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>.
- 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>.
- Smelov V, Vrbanc A, van Ess EF, Noz MP, Wan R, Eklund C, Morgan T, Shrier LA, Sanders B, Dillner J, de Vries HJC, Morr  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>.
- 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>.
- Dean D, Bruno WJ, Wan R, Gomes JP, Devignot S, Mehari T, de Vries HJC, Morr  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>.
- Hadfield J, Harris SR, Seth-Smith HMB, Parmar S, Andersson P, Giffard PM, Schachter J, Moncada J, Ellison L, Valet MLG, Fermepin MR, Radebe F, Mendoza S, Ouburg S, Morr  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>.
- Millman KL, Tavar  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>.
- 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>.
- 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>.
- 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>.
- 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>.
- 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>.
- Merkel R. 2007. Modelling the evolution of the archeal tryptophan synthase. *BMC Evol Biol* 7:59. <https://doi.org/10.1186/1471-2148-7-59>.
- Busch F, Rajendran C, Heyn K, Schlee S, Merkel 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>.
- 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>.
- 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>.
- 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 holo-repressors: 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. Fehlnner-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](https://doi.org/10.1016/S0021-9258(18)60507-1).
 50. Lawson CL, Sigler PB. 1988. The structure of *trp* pseudorepressor at 1.65 Å 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, Yeru 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>.