Tetracycline Selective Pressure and Homologous Recombination Shape the Evolution of Chlamydia suis: A Recently Identified Zoonotic Pathogen

Sandeep J. Joseph1,†, Hanna Marti2,†, Xavier Didelot3, Timothy D. Read1,†, and Deborah Dean2,4,†,*

1Department of Medicine, Division of Infectious Diseases and Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA
2Center for Immunobiology and Vaccine Development, UCSF Benioff Children’s Hospital Oakland Research Institute, Oakland, CA, USA
3Department of Infectious Disease Epidemiology, Imperial College, London, United Kingdom
4Joint Graduate Program in Bioengineering, University of California, San Francisco and University of California, Berkeley, CA, USA

†These authors contributed equally to this work.
*Corresponding author: E-mail: ddean@chori.org.
Accepted: July 24, 2016
Data deposition: SRP076849.

Abstract

Species closely related to the human pathogen Chlamydia trachomatis (Ct) have recently been found to cause zoonotic infections, posing a public health threat especially in the case of tetracycline resistant Chlamydia suis (Cs) strains. These strains acquired a tet(C)-containing cassette via horizontal gene transfer (HGT). Genomes of 11 Cs strains from various tissues were sequenced to reconstruct evolutionary pathway(s) for tet(C) HGT. Cs had the highest recombination rate of Chlamydia species studied to date. Admixture occurred among Cs strains and with Chlamydia muridarum but not with Ct. Although in vitro tet(C) cassette exchange with Ct has been documented, in vivo evidence may require examining human samples from Ct and Cs co-infected sites. Molecular-clock dating indicated that ancestral clades of resistant Cs strains predated the 1947 discovery of tetracycline, which was subsequently used in animal feed. The cassette likely spread throughout Cs strains by homologous recombination after acquisition from an external source, and our analysis suggests Betaproteobacteria as the origin. Selective pressure from tetracycline may be responsible for recent bottlenecks in Cs populations. Since tetracycline is an important antibiotic for treating Ct, zoonotic infections at mutual sites of infection indicate the possibility for cassette transfer and major public health repercussions.

Key words: Chlamydia suis, tet(C)-containing cassette, tetracycline selective pressure, comparative genomics, genomic diversification.

Introduction

Chlamydia suis (Cs) is a common swine pathogen that represents a potential threat to human populations through zoonotic transmission. The organism has recently been identified in the conjunctivae of Nepalese trachoma patients (Dean et al. 2013) and Belgian slaughterhouse workers (De Puysselyr et al. 2014) by real-time PCR and microarray analyses as well as by isolation and Cs-specific PCR of nasal, pharyngeal, and stool samples obtained from pig farmers (De Puysselyr et al. 2015). The human eye, nasopharynx, and gastrointestinal tract are frequent sites of infection with the human pathogen Chlamydia trachomatis (Ct) (Somboonna et al. 2011; Dean et al. 2013; Li et al. 2015) and, in some cases, with other zoonotic Chlamydia species that now include Cs (Goldschmidt et al. 2006; Dean et al. 2008, 2013).

Pigs are one of the most important livestock animals in the world, accounting for up to 36% of global meat consumption (FAO, http://www.fao.org/ag/againfo/themes/en/meat/back gr_sources.html, Sources of Meat, last updated 2014). There is a high rate of contact between pigs and humans and, hence, the potential for zoonotic transmission of multidrug-resistant organisms, as illustrated by drug-resistant Staphylococcus aureus transfer from pigs to pig farmers.
(Oppliger et al. 2012; Wardyn et al. 2015). This example is especially relevant because of the high frequency of tetracycline resistance among Cs strains. Resistance was acquired via horizontal gene transfer (HGT) of the class C gene [tet(C)]-containing cassette most likely from another gut bacteria in response to selective pressure (Borel et al. 2012). This raises considerable concern because Cs shares 79.8% average nucleotide identity (ANI) with Ct, and co-infection in the human host presents an opportunity for cassette transfer to Ct. In vitro studies have already shown that the cassette can be transferred between Cs and Ct (Suchland et al. 2009).

Ct remains the leading bacterial cause of sexually transmitted diseases (STD) and preventable blindness worldwide (CDC, Chlamydia Statistics, last updated 2016, http://www.cdc.gov/std/chlamydia/stats.htm). As tetracycline drugs are commonly used for uncomplicated Ct infections and remain the drugs of choice for complicated infections (Hammerschlag 2013), acquisition of tetracycline resistance by Ct would be devastating.

Given the importance of Cs as a swine and recently emerged zoonotic pathogen, our aim was to provide the first genome sequences of multiple Cs strains, comparing these with other available Chlamydia genomes to reconstruct the evolutionary pathway for HGT of the tet(C)-containing cassette and the potential for transfer to other Chlamydia species that infect humans.

Materials and Methods

Whole Genome Sequencing, Alignments, Core Genes, and Phylogenetic Reconstruction

Table 1 lists the sequenced Cs strains and their clinical attributes. Each strain was individually prepared for genome sequencing as described (Somboonna et al. 2011). We also included one published draft genome of Cs MD56 (Donati et al. 2014) (see Supplementary Material online).

Random shotgun sequencing was performed using a GS-FLX instrument (454 Life Sequencing Inc., Branford, CT, USA). De novo assembly was performed using Newbler (Chaisson and Pevzner 2008) with additional genome processing as described (Joseph et al. 2012). We performed gap closure by PCR, Sanger sequencing and manual editing to close gaps between contigs, which were then mapped using CONTIGuator (Galardini et al. 2011) to the closely related species Chlamydia muridarum (Cm) (Read et al. 2003) to create concatenated, ordered “pseudocontigs”. The Prokka bacterial genome annotation pipeline (Seemann 2014) was used to identify genes and putative proteins. MAUVE (Darling et al. 2010) progressive alignments were computed for two datasets: 1) 12 Cs genomes (length 1038851 bps) and 2) these 12 Cs plus representative Ct genomes (L2b/UCH-1, FIC-Cal-13, C/TW-3/OT, D/UCH-3/CX) from each of the four disease-associated clades (Joseph et al. 2012) and Cm strain MoPn (941089 bp) (Read et al. 2003). The core Locally Collinear Blocks (LCBs) from all alignments were extracted and concatenated to form a super alignment for phylogenetics.

The predicted proteome from all 12 annotated Cs genomes was searched against itself using BLASTP (e value cutoff: 1e−05) with best scores analyzed by OrthoMCL (Li et al. 2003) to identify orthologous sequences. Core genes were defined as protein-coding genes shared by all Cs strains; unique genes per strain were also identified. Alignment of core genes by MUSCLE, filtering of protein alignments by GBLOCKS and generation of the whole genome protein alignment were performed as described (Joseph et al. 2015).

Core LCBs and core protein sequence alignments were concatenated to reconstruct the maximum likelihood (ML)-based phylogenetic tree, for principal component analysis (PCA) on the gene presence/absence matrix (panmatrix), along with the estimated branch lengths/evolutionary distances with statistical support as described (Joseph et al. 2015).

Recombination and Attribution of Origins to Recombination Events

ClonalFrame (version 1.2) (Didelot et al. 2010) was applied to the whole genome core nucleotide alignment to estimate recombination parameters as described (Joseph et al. 2015). Three independent and parallel runs of ClonalFrame showed high congruence for reconstructed phylogenies and recombination events. Mutation and recombination events were computed for each reconstructed branch substitution event introduced by mutation or recombination. The relative effect of recombination and mutation on genetic change (r/m) and the relative rate of mutation and recombination (μ/μr) were estimated as described (Read et al. 2013) (see Supplementary Material online).

For each branch of the tree reconstructed by ClonalFrame, recombined fragments were defined as genomic intervals with a posterior probability of recombination above 0.50 at every site, reaching 0.95 in at least one site, and >100 bps (Didelot et al. 2011). Each such recombined fragment was searched for using BLASTN against a database containing all public genomes and plasmid sequences of Chlamydia species (including the Cs genomes) minus strains of the clade affected by the import and against the NCBI NT BLAST database (last updated January 15, 2016). The hits with the highest normalized BLASTN score along with a 100% identity were kept. If best hits were to a single Cs strain or strains belonging to the same clade or subclades identified for Cs (see below), the origin of the event was attributed to that particular strain or internal node corresponding to those strains. Recombination events that failed to be assigned a putative origin were categorized as unsolved events.
To detect homologous intragenic recombination in the core genes, we implemented Pairwise Homoplasy Index (PHI); Neighbor Similarity Score (NSS); and Maximum $\chi^2$ using the PhiPack package (Bruen et al. 2006). Parameters, estimations of $P$ values and correction for multiple testing were performed as described (Joseph et al. 2011) (see Supplementary Material online).

**Genetic Diversity**

We selected 12 Cs and 12 Ct genomes with Cm strain MoPh as outgroup to assess genetic diversity. Ct genomes represented each of the four clades segregated by disease (Joseph et al. 2012). The predicted proteome from all 25 annotated Chlamydia genomes was searched against itself using BLASTP ($e$ value cutoff: 1e–05). The best BLASTP scores were utilized for identifying orthologous sequences using OrthoMCL as above. There were a total of 774 core genes shared among the three species. For each core gene, the codon alignment was generated by Pal2nal (Suyama et al. 2006) using protein and corresponding nucleotide sequences. For Cs and Ct, two measures of genetic diversity were calculated: average pairwise nucleotide diversity per site ($\phi$) (Nei and Li 1979) and Watterson’s $\theta$ (Watterson 1975). For both measurements, synonymous and nonsynonymous diversity were calculated using the corresponding codon alignment by specifying the two populations, excluding Cm, using the R package PopGenome (Pfeifer et al. 2014).

**Population Structure Analysis, Effective Population Size ($N_e$) and Demography for Cs and Ct**

The ChromoPainter algorithm was applied to genome-wide haplotype data generated from the second (12 Cs + 1 Cm + 4 Ct) core MAUVE alignment to delineate both population structure and investigate DNA transfer (admixtures) events among the three species. The ChromoPainter co-ancestry matrix output was used in fineSTRUCTURE to further explore population structure as described (Joseph et al. 2015) (see Supplementary Material online).

Bayesian Skyline Plot (BSP) analysis implemented in BEAST v.1.8.2 was used to co-estimate the genealogy and demographic history of the populations (Drummond and Rambaut 2007; Heled and Drummond 2008). Consequently, estimates of model parameters are integrated over phylogenetic and coalescent uncertainty. For BSP analysis, a core genome alignment for Ct and Cs was created separately after removing recombinant nucleotide sequences previously identified for Ct (Joseph et al. 2012) and Cs (this study); resultant alignments contained 296,984 and 791,030 nucleotide bases, respectively. A strict molecular clock model, the GTR nucleotide substitution model for phylogenetic reconstruction and tip dates defined as year of isolation for each strain in the two species, was used to estimate changes in population sizes for Ct and Cs, separately. We used a previously estimated mutation rate ($3.84 \times 10^{-10}$) for Ct (Borges et al. 2013) on an in vitro environment as a mean prior substitution rate for both species. Two separate runs of 50,000,000 Markov chain Monte Carlo (MCMC) iterations were performed for each species and adequate mixing was achieved; 10% burn-in was removed, and the sampling was done every 1,000 iterations. The remaining parameters were set as per Heled and Drummond (2008). Results were analyzed with Tracer v.1.5 and LogCombiner v.1.7.5 (Drummond and Rambaut 2007).

**Results and Discussion**

Two Clades predominate in the Cs Phylogeny and Population Structure

The whole genome alignment of the 12 Cs strains allowed us to perform the first comparative genomic analysis of this species. The alignment comprised 102 homologous core LCBs

---

**Table 1**

**Chlamydia suis Strain Information**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Site/Disease Location</th>
<th>Date of Isolation</th>
<th>Ref</th>
<th>Median Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>R19</td>
<td>Enteritis Nebraska, USA 1992</td>
<td>Rogers and Andersen (1996)*</td>
<td>14.0x</td>
<td></td>
</tr>
<tr>
<td>R22</td>
<td>Conjunctivitis Nebraska, USA 1992</td>
<td>Rogers et al. (1993)*</td>
<td>85.3x</td>
<td></td>
</tr>
<tr>
<td>R24</td>
<td>Respiratory Nebraska, USA 1992</td>
<td>*</td>
<td>59.0x</td>
<td></td>
</tr>
<tr>
<td>R27</td>
<td>Enteritis Nebraska, USA 1993</td>
<td>Rogers and Andersen (1996)*</td>
<td>22.38x</td>
<td></td>
</tr>
<tr>
<td>H5</td>
<td>Conjunctivitis Iowa, USA 1994</td>
<td>Na*</td>
<td>37.06x</td>
<td></td>
</tr>
<tr>
<td>H7</td>
<td>Conjunctivitis Iowa, USA 1994</td>
<td>Rogers and Andersen (1999)*</td>
<td>22.0x</td>
<td></td>
</tr>
<tr>
<td>S45</td>
<td>Feces Austria 60s</td>
<td>Kaltenboeck et al. (1993)*</td>
<td>19.5x</td>
<td></td>
</tr>
<tr>
<td>Rogers132</td>
<td>Intestine, lung, conjunctiva Nebraska, USA 1996</td>
<td>Rogers et al. (1993)*</td>
<td>32.75x</td>
<td></td>
</tr>
<tr>
<td>R1</td>
<td>Enteritis Midwest, USA 1994</td>
<td>Na*</td>
<td>21.7x</td>
<td></td>
</tr>
<tr>
<td>R16</td>
<td>Respiratory Midwest, USA 1994</td>
<td>Na*</td>
<td>39.4x</td>
<td></td>
</tr>
<tr>
<td>R28</td>
<td>Enteritis Midwest, USA 1995</td>
<td>Na*</td>
<td>60.47x</td>
<td></td>
</tr>
<tr>
<td>MD56</td>
<td>Conjunctivitis Italy 2009</td>
<td>Donati et al. (2014)</td>
<td>160x*</td>
<td></td>
</tr>
</tbody>
</table>

*Samples obtained from Dr. Art Andersen’s collected that is housed in and curated by Dr. Deborah Dean’s lab.
representing 1,038,851 bp of the 1.09 Mb average Chlamydia genus genome size with an average of 922 protein-coding genes and a conserved plasmid of ~7.5 kb, except for strain R22 that lacked a plasmid (supplementary table S1, sheet A, Supplementary Material online).

In common with findings from other Chlamydia studies (Joseph et al. 2011, 2012, 2015; Harris et al. 2012), most of the genome encoded conserved core functions. A total of 861 core genes (grouped into families by orthology relationships) were present among all 12 strains, constituting ~93% of the genes per genome. All except R22 contained genes not present in any other genome (supplementary table S1, sheets A and B, Supplementary Material online).

Whole-genome phylogenies inferred from the whole-genome DNA alignment and concatenated core proteome of the 12 Cs, 4 Ct, and 1 Cm genomes agreed in stratifying Cs into two clades (fig. 1A). The PCA tree similarly clustered the genomes except that those in Clade 2 and MD56 formed a distinct group (supplementary fig. 1A, Supplementary Material online). Unlike Ct (Joseph et al. 2011, 2012), the clades did not resolve along disease or anatomic demarcations. The protein tree also formed the same two clades (supplementary fig. 1B, Supplementary Material online). This phylogeny was likely affected by frequent homologous recombination events described below.

Population structure analysis using fineStructure confirmed two main clades similar to those identified by ML and six sub-populations of which two were within Clade 1 and four within Clade 2 (fig. 2A). ClonalFrame (Didelot and Falush 2007), which takes the effects of recombination into consideration, also identified the same two clades, up to the placement of the root (fig. 2B); as the ML tree benefited from inclusion of non-Cs outgroup genomes it was better able to root the tree. The plasmid phylogeny similarly resolved into two clades (fig. 2C).

Chlamydia suis Undergoes Homologous Recombination and Genetic Exchange with Other Chlamydia Species, and has a Higher Rate of Intraspecific Recombination Compared to Ct

ClonalFrame does not model the origin of recombination events, but postprocessing of the output can determine the source of the origins of those events (Didelot et al. 2011). Of 1,593 events identified, 345 could be assigned a putative origin, either from another Cs strain or from an internal node. We could not assign origins to the rest because a source with a normalized sequence identity of >95% was not identified. The most probable reason for numerous unassigned recombination events is the small number of available Cs genomes. A larger sample size, including Chlamydia populations causing unapparent infection, although difficult to obtain, would likely yield more putative origins.

Of the 345 recombination events with assigned putative origins (fig. 2B), 15 occurred within the two clades, while 330 were interclade exchanges (fig. 1B). There were 147 recombination events identified on an internal branch of Clade 1, originating from the ancestor of R16, R1, and R28 in Clade 2. Similarly, there were 101 DNA exchange events to the internal/ancestral node of R24 and R19 in Clade 1 from the internal node above R16, R1, and R28 in Clade 2. Such high rates of intraspecific recombination have also been documented in other gut bacteria such as Salmonella enterica and Helicobacter pylori (Suerbaum and Josenhans 2007; Vos and Didelot 2009).

The fineSTRUCTURE coancestry matrix further confirmed the extent of genetic exchange within and across the two clades (fig. 2A). Outgroup strains H7 and MD56 appeared to have received DNA imports from all Cs strains. Similarly, admixture was evident in Clade 1 strains H5, Rogers132 and S45, receiving imports from Clade 2 strains, and from MD56 and H7. The lineage that included R27 appeared to be a major source of DNA transfer to all Clade 2 strains, MD56 and H7 as well as the Clade 1 strains, H5, Rogers132, S45 and R24. We did not find admixture between Cs and Ct, whereas Cm appeared to have received DNA from both species (supplementary fig. S1A, Supplementary Material online). While these data suggest a barrier to recombination for species other than Cm, recent laboratory experiments producing tetracycline resistant Ct strains indicate otherwise (Suchland et al. 2009). In vivo evidence of cassette exchange may require larger sample sizes of both species from co-infected human populations.

The impact of recombination on Cs genomes was quantified by applying ClonalFrame to the whole-genome alignment (fig. 2B). ClonalFrame estimated the 95% credibility interval of \( \rho/\theta \) at 0.339–1.49 (mean = 0.89), indicating that mutation and recombination have been approximately equally frequent during the evolutionary process in Cs. The 95% credibility interval of \( r/m \) was 4.619–17.513 (mean = 10.829), indicating that recombination introduced approximately ten times more substitutions than mutations (supplementary table S1, sheet C, Supplementary Material online). When we compared genomes that are closely related to each other (e.g., R19 and R24), many recombination events separated them. The expectation for genomes that are more distantly related (e.g., H7 and MD56) is that most/all of the genome has recombined. Many imports were coming from sources external to the data-set, and these imports brought in new polymorphisms, clarifying the phylogenetic topology rather than disrupting it. The plasmid \( r/m \) and \( \rho/\theta \) estimates are noted in supplementary table S1, sheet C, Supplementary Material online.

Recombination affected segments with a mean length of 1,117 bp (95% credibility interval of 902–1,391 bp). Similar estimated mean track lengths of recombined fragments were reported in Chlamydia psittaci (1,116 bp), but the
Phylogenies and interclade recombination events. (A) Whole-genome maximum likelihood phylogeny of Chlamydia species and strains. The tree was constructed using RAxML based on the whole-genome alignment (see Materials and Methods). The Chlamydia trachomatis (Ct) and Chlamydia muridarum (Cm) strains are highlighted in magenta and aqua, respectively. Chlamydia suis (Cs) strains are highlighted in red for Clade 1 strains (R19, R24, Rogers132, H5, S45), green for Clade 2 (R1, R16, R28, R22, R27) and blue for MD56 and H7. The light red circles indicate 100% bootstrap estimates; (B) High frequency of interclade recombination events originating from the ancestral nodes of Cs strains. A putative origin was assigned to each recombination event using ClonalFrame analysis. Out of the 345 recombination events detected, 330 were interclade exchanges from Clade 2 to Clade 1 or Clade 1 to Clade 2, respectively. Clade 1 and recombination events originating from Clade 1 are marked in red; Clade 2 and recombination events originating from Clade 2 are marked in green. H7 and MD56 are highlighted in blue. For example, six recombination events to Clade 1 originated from a Clade 2 ancestral node consisting of R28, R16, R1 and R22; 28 recombination events to Clade 2 originated from the ancestral node of R24 R19; (C) Bayesian phylogenetic reconstruction of all Cs genomes indicates early separation of the two distinct Cs clades. BEAST software based on an alignment of core genes lacking recombinant regions was used to estimate the dates of the Last Common Ancestors (LCA) to the Cs clades. Clade 1 strains are marked in red, Clade 2 strains in green, and MD56 and H7 are marked in blue.

Fig. 1.—Phylogenies and interclade recombination events. (A) Whole-genome maximum likelihood phylogeny of Chlamydia species and strains. The tree was constructed using RAxML based on the whole-genome alignment (see Materials and Methods). The Chlamydia trachomatis (Ct) and Chlamydia muridarum (Cm) strains are highlighted in magenta and aqua, respectively. Chlamydia suis (Cs) strains are highlighted in red for Clade 1 strains (R19, R24, Rogers132, H5, S45), green for Clade 2 (R1, R16, R28, R22, R27) and blue for MD56 and H7. The light red circles indicate 100% bootstrap estimates; (B) High frequency of interclade recombination events originating from the ancestral nodes of Cs strains. A putative origin was assigned to each recombination event using ClonalFrame analysis. Out of the 345 recombination events detected, 330 were interclade exchanges from Clade 2 to Clade 1 or Clade 1 to Clade 2, respectively. Clade 1 and recombination events originating from Clade 1 are marked in red; Clade 2 and recombination events originating from Clade 2 are marked in green. H7 and MD56 are highlighted in blue. For example, six recombination events to Clade 1 originated from a Clade 2 ancestral node consisting of R28, R16, R1 and R22; 28 recombination events to Clade 2 originated from the ancestral node of R24 R19; (C) Bayesian phylogenetic reconstruction of all Cs genomes indicates early separation of the two distinct Cs clades. BEAST software based on an alignment of core genes lacking recombinant regions was used to estimate the dates of the Last Common Ancestors (LCA) to the Cs clades. Clade 1 strains are marked in red, Clade 2 strains in green, and MD56 and H7 are marked in blue.
The mean length of recombined fragments in Ct was only 357 bp (Read et al. 2013).

Based on PHI, NSS, and maximum $\chi^2$, 77 Cs core genes showed significant evidence for intragenic recombination ($P < 0.05$) by at least one method (supplementary table S1, sheet D, Supplementary Material online); 36 genes were identified as having undergone recombination by all three, including genes known to be recombinogenic in Ct (Gomes et al. 2004, 2007; Joseph et al. 2012): The major outer membrane gene ($ompA$), polymorphic membrane protein gene ($pmpE$, $pmpG$), secretion system apparatus protein gene ($ssaV$), inclusion membrane protein A gene ($incA$) and the lipoprotein-releasing system transmembrane protein ($lolE$).

### Higher Genetic Diversity and Lower Effective Population Size for Cs Compared to Ct

We examined how the genetic diversity and demography of Cs compared with Ct. Using the MSA of the 774 core genes shared by Cs and Ct, we estimated that, on average, Cs...
harbored more pairwise genetic diversity (measured by $\phi$) than Ct. For nonsynonymous sites, the mean estimate for Cs for $\phi$ was 4.00, compared to 1.98 for Ct ($P$ value $<0.001$); for synonymous sites, estimates were 2.91 and 1.00 ($P$ value $<0.001$), respectively. Cs and Ct were also significantly different for both synonymous (mean estimate $\theta_w$: 2.68 and 1.15, respectively) and nonsynonymous (mean estimate $\theta_w$: 3.30 and 2.63, respectively) sites (fig. 3A–D).

We further compared effective population sizes using BSP analysis. There was a higher estimated substitution rate for Cs (1.872 x $10^{-5}$ per site/year; 95% HPD: 1.3712 – 2.348 x $10^{-5}$) than for Ct (3.39 x $10^{-7}$ per site/year). The effective population size was calculated using both synonymous measures of genetic diversity ($\theta_w$ and $\phi$) (Kimura and Crow 1964). The effective population size for Cs was estimated to be smaller (between 7.7 x $10^4$ and 8.6 x $10^4$) than Ct (8.5 x $10^5$ and 9.7 x $10^5$) despite a significantly higher genetic diversity, which may be due to the higher mutation rate in Cs, which was similar to that reported for C. psittaci (Read et al. 2013). For Ct, BSP analysis showed stable population sizes in the past decades while, for Cs, the trajectory of the skyline plot indicated population size reduction in the past 30–50 years (fig. 3E and F). This potential recent bottleneck in the Cs population might be associated with indiscriminate use of tetracycline in pig farming since the 1950s (Roberts 2003). This hypothesis is supported by the fact that antibiotics are well known inducers of selective pressure, raising the general rates of recombination, mutation, and HGT in bacteria (Gillings 2013). Antibiotic use has also been shown to impact the population structure of other pathogens. For example, a decline in the population size of Neisseria gonorrhoeae, a sexually transmitted bacterial pathogen, was observed in the 1980s/1990s after introduction of penicillin and emergence of penicillinase-producing N. gonorrhoeae (Pérez-Losada et al. 2007).

Insertion of Tetracycline Resistance Transposon into the Cs Genome by HGT

The most remarkable feature of the Cs genomes is the acquisition of a tetracycline resistance cassette by HGT. Of the 11 strains we sequenced, all but S45 contained a class C tetracycline resistance gene [tet(C)] within a variable cassette.
Previously, Dugan et al. (2004) demonstrated that the cassette of seven Cs strains were inserted at an identical site within the invasin-like (inv-like or ilp) gene in a hypervariable segment between a hypothetical protein and Na(+)−translocating NADH-quinone reductase subunit F (nqrF). We confirmed this insertion site for these strains and four Clade 2 strains, R1, R16, R22, R27, R28, sequenced here. The two clades identified by ClonalFrame correlated with the structural variants described by Dugan et al. (2004). Clade 2 strains did not contain replication factor or mobilization protein genes, and the tet(C)/tetR(C) segment were positioned downstream of IS200 and putative IS605 transposases (fig. 4). Clade 1 strain H5 contained these transposases on the distal side of the enterobacterial plasmid whereas Rogers132 was missing the transposases entirely.

Dugan et al. (2004) found by BLASTN that the Aeromonas salmonicida plasmid pRAS3.2 had the highest homology to the entire tet(C) cassette (query cover: 83%, identity cover: 99%) but no homology between the pRAS3.2 plasmid and the IS200/IS605 transposases. Interestingly, the consecutive sequences of tet(C), tetR(C), IS200, and IS605 were 100% identical (query cover: 91–100%) to Snodgrassella species [e.g., Snodgrassella alvi wkB2 (accession no. CP007446.1), Candidatus Snodgrassella sp. T4_34144 (JQ966977.1)] in the family Neisseriaceae (Tian et al. 2012). These tetracycline resistant commensal bacteria are found in the gut of honeybees. Honeybees are important agricultural pollinators that are extensively treated with antibiotics, including tetracycline, particularly in commercial beekeeping in the U.S. where the presence of resistance genes is ubiquitous in these animals (Tian et al. 2012). This is similarly the case for pigs in the U.S., Europe, and the Middle East (Dugan et al. 2004; Borel et al. 2012; Schautteet et al. 2013), highlighting a common pathway for tet(C) acquisition under parallel antibiotic selective pressure in two very different, economically important animal species occurring after the discovery of tetracycline in 1947 (Roberts 2003, 2005).

The exact source of the cassette remains an enigma although it is likely of Betaproteobacteria origin. The close sequence identity of the cassette across Cs strains argues for a recent single ancestral insertion of the cassette. However, BEAST molecular dating identified AD1905 and AD1717 as the mean estimated dates for the MRCA of Clades 1 and 2, respectively (fig. 1C), suggesting that they predated the introduction of tetracycline into pig farming. Two explanations for this pattern are possible: there were multiple independent acquisitions of the cassette through interspecies transfer (fig. 5, Hypothesis 1); and/or the cassette was acquired...
by an index strain and exchanged between Cs strains by intraspecies recombination (fig. 5, Hypotheses 2 and 3). Given our findings that recombination is frequent within Cs but relatively uncommon between species in Chlamydia based on the data to date, we suggest that the latter scenario is more plausible.

**Conclusions**

Our study was limited by the small number of Cs samples available for genome sequencing, although they came from various tissues in different pigs. We were also limited in our analyses of admixture with Ct because the public databases contained few if any strains isolated from sites that could be co-infected with Cs. To more rigorously evaluate recombination and mechanisms of exchange between Cs and Ct, isolates of both species from human ocular, pharyngeal, and gastrointestinal sites would be needed. Nonetheless, we were able to show that Cs is highly recombinogenic with admixture involving all Cs strains and with their close phylogenetic relative Cm. The evolutionary impact of tetracycline selective pressure has fostered high rates of tetracycline resistance among Cs populations. Since tetracycline is an important antibiotic for treating mild to severe Ct infections, Cs zoonoses at sites co-infected with Ct represent the possibility for cassette transfer and potential for major public health repercussions.

**Supplementary Material**

Supplementary table S1 and figures S1 and S2 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

**Acknowledgments**

The authors are grateful to Dr Arthur Andersen for providing his collection of Chlamydiaceae species and strains to the Dean lab, for archiving and curating, without which this research
would not have been possible. We thank the Dean Lab members Manu Sharma and Alison Vrbancic for excellent technical assistance. This work was supported in part by Public Health Service grant from the National Institute of Health (R01 AI098843 to D.D. and T.D.R.), and the National Science Foundation (NSF) (2009-65109-05760 to D.D.) and an Early Postdoctoral Mobility Fellowship grant from the Swiss National Science Foundation (SNSF) (P2ZHP3_158590 to H.M.).

**Literature Cited**


Tetracycline Selective Pressure and Homologous Recombination Shape the Evolution of *Chlamydia suis*


Somboonna N, et al. 2011. Hypervirulent *Chlamydia trachomatis* clinical strain is a recombinant between lymphogranuloma venereum (L2) and D lineages. MBio 2:e00045–e00011.


**Associate editor:** Bill Martin