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

Sandeep J. Joseph[1,†], Hanna Marti[2,†], Xavier Didelot[3], Timothy D. Read[1,†], and Deborah Dean[2,4,†,*]

[1]Department of Medicine, Division of Infectious Diseases and Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA

[2]Center for Immunobiology and Vaccine Development, UCSF Benioff Children's Hospital Oakland Research Institute, Oakland, CA, USA

[3]Department of Infectious Disease Epidemiology, Imperial College, London, United Kingdom

[4]Joint 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 Puysseleyr 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 Puysseleyr 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/backgr_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 (L_2b/UCH-1, F/IC-Cal-13, C/TW-3/OT, D/UW-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 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 (ρ/θ) 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.

Table 1
Chlamydia suis Strain Information

Strain	Site/Disease	Location	Date of Isolation	Ref	Median Coverage
R19	Enteritis	Nebraska, USA	1992	Rogers and Andersen (1996)[a]	14.0×
R22	Conjunctivitis	Nebraska, USA	1992	Rogers et al. (1993)[a]	85.3×
R24	Respiratory	Nebraska, USA	1992	[a]	59.0×
R27	Enteritis	Nebraska, USA	1993	Rogers and Andersen (1996)[a]	22.38×
H5	Conjunctivitis	Iowa, USA	1994	NA[a]	37.06×
H7	Conjunctivitis	Iowa, USA	1994	Rogers and Andersen (1999)[a]	22.0×
S45	Feces	Austria	1960s	Kaltenboeck et al. (1993)[a]	19.5×
Rogers132	Intestine, lung, conjunctiva	Nebraska, USA	1996	Rogers et al. (1993)[a]	32.75×
R1	Enteritis	Midwest, USA	1994	NA[a]	21.7×
R16	Respiratory	Midwest, USA	1994	NA[a]	39.4×
R28	Enteritis	Midwest, USA	1995	NA[a]	60.47×
MD56	Conjunctivitis	Italy	2009	Donati et al. (2014)	160×[a]

[a]Samples obtained from Dr. Art Andersen's collected that is housed in and curated by Dr. Deborah Dean's lab.

To detect homologous intragenic recombination in the core genes, we implemented Pairwise Homoplasy Index (PHI); Neighbor Similarity Score (NSS); and Maximum χ^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 MoPn 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 (ϕ) (Nei and Li 1979) and Watterson's θ (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 (3.84×10^{-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

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. 1*A*, 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. 1*B*, 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 subpopulations 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. S1*A*, 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 ρ/θ 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 dataset, and these imports brought in new polymorphisms, clarifying the phylogenetic topology rather than disrupting it. The plasmid r/m and ρ/θ 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

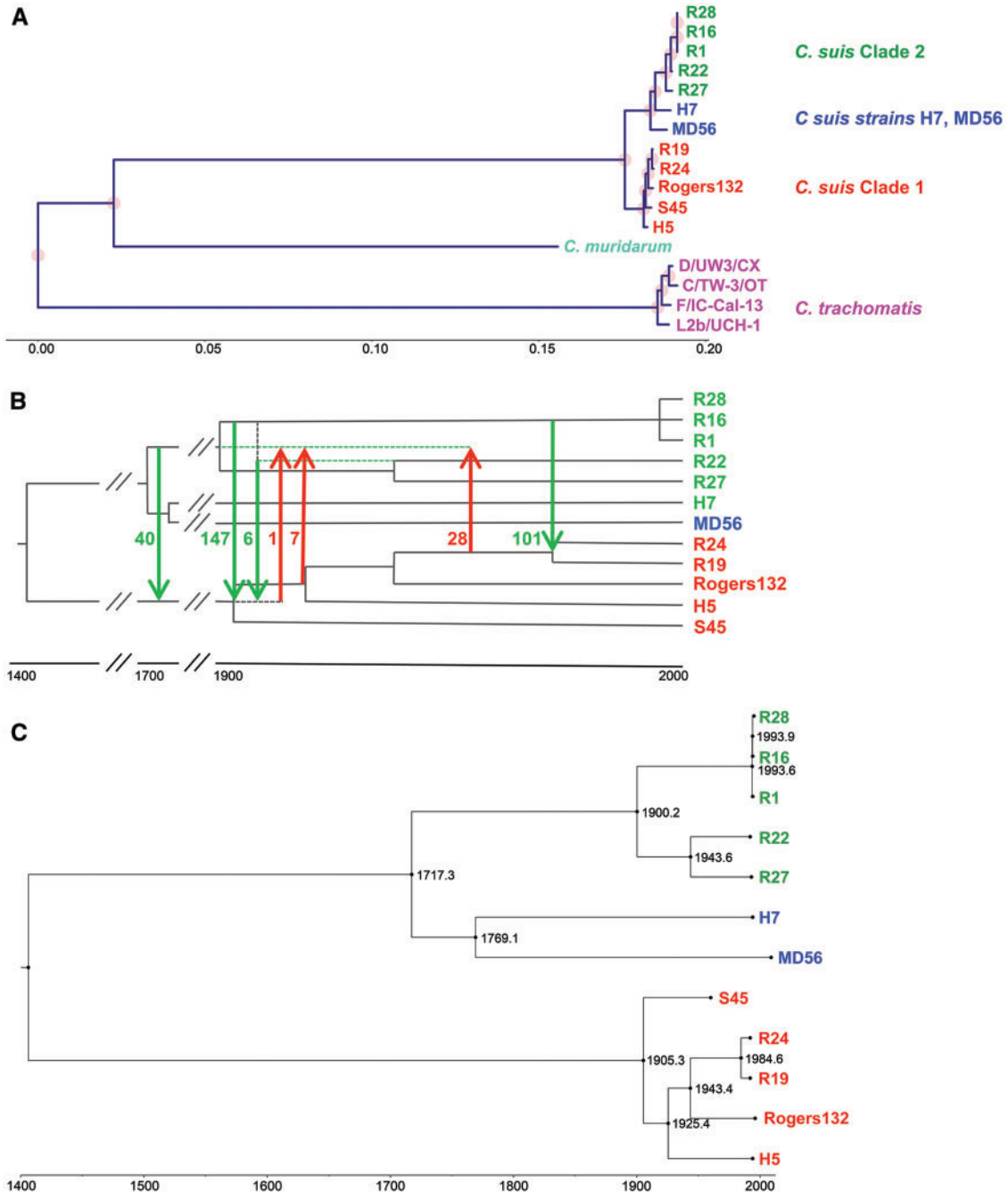

Fig. 1.—Phylogenies and interclade recombination events. (*A*) Whole-genome maximum likelihood phylogeny of *Chlamydia* species and strains. The tree was constructed using RAxML based on the whole-genome alignment (see Materials and Methods). The *Chlamydia trachomatis* (*Ct*) and *Chlamydia muridarum* (*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.

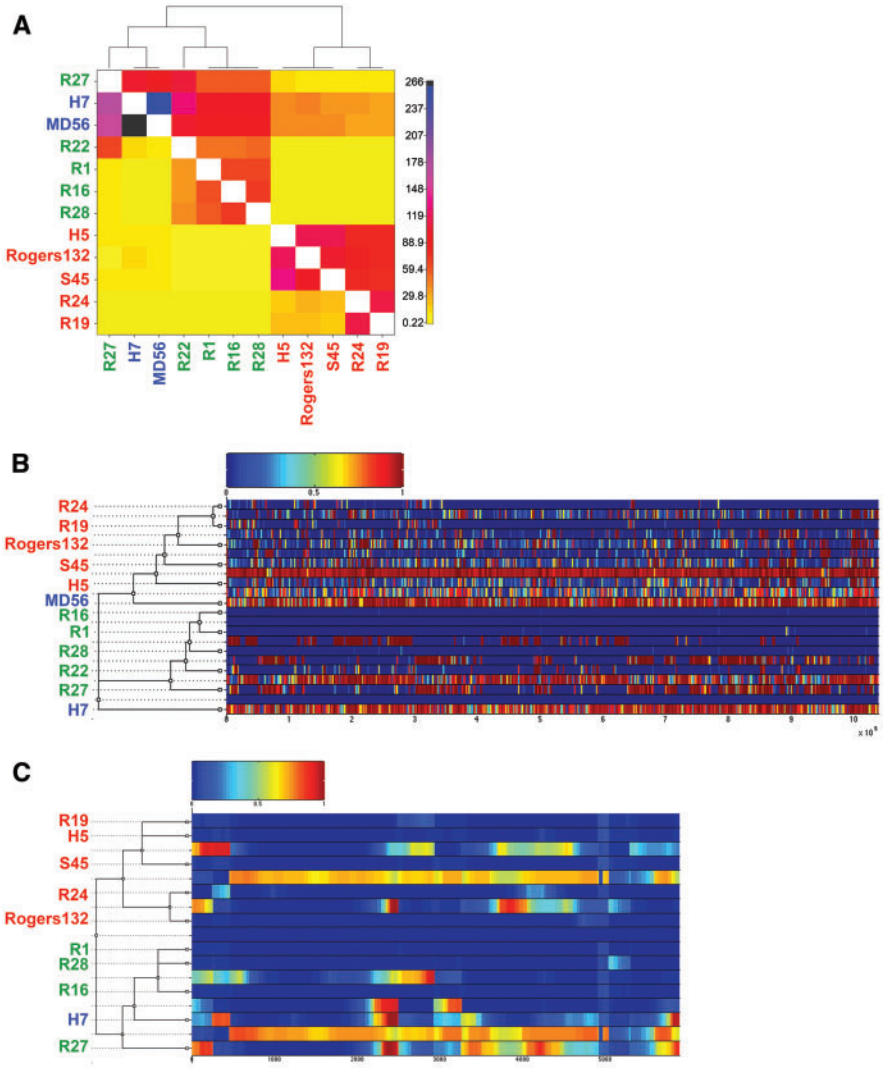

Fig. 2.—Frequent homologous recombination among *Chlamydia suis* (*Cs*) strains. (*A*) Abundant events of genetic exchange within the two *Cs* clades. ChromoPainter co-ancestry matrix with population structure was assignment based on fineSTRUCTURE analysis. The color of each cell of the matrix indicates the expected number of events (see vertical color bar on right) in which genetic material was copied from a donor genome (*x axis*) to a recipient genome (*y axis*). Clade 1 is shown in red, Clade 2 in green, and strains H7, and MD56 in blue; (*B*) Results of the ClonalFrame analysis on the chromosomes of 12 *Cs* genomes. Inferred clonal genealogy is shown on the left. Each branch of the tree corresponds to a row of the heat map, which is horizontally aligned according to the core MAUVE whole-genome alignment. Each row shows the posterior probability of recombination estimated by ClonalFrame on the corresponding branch (*y axis*) and along positions of the alignment (*x axis*); and *(C)* Results of the ClonalFrame analysis of 10 *Cs* plasmids. Strains R22 and MD56 were not included in this analysis because R22 does not possess a plasmid and plasmid contigs could not be distinguished from other contigs for MD56. Clade 1, Clade 2, and strains H7 and MD56 are shown in red, green and blue, respectively.

mean length of recombined fragments in *Ct* was only 357 bp (Read et al. 2013).

Based on PHI, NSS, and maximum χ^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 genes (*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*

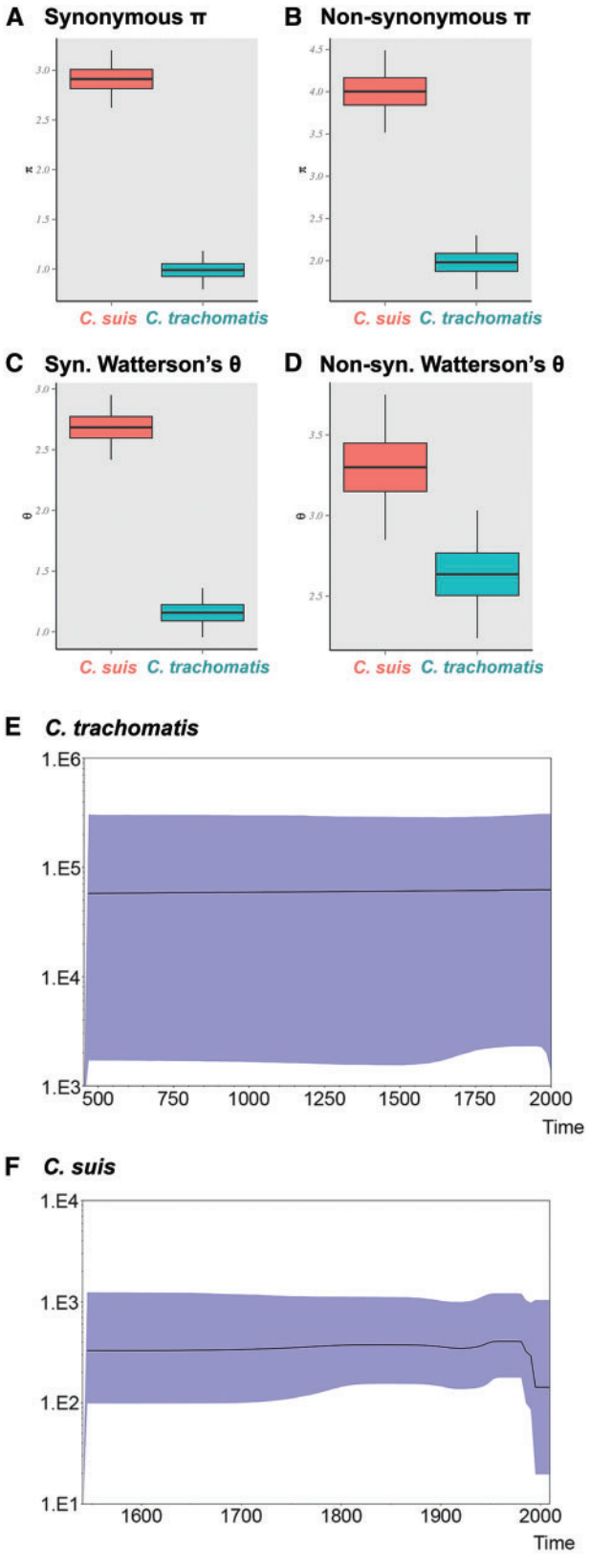

Fig. 3.—*Chlamydia suis* (*Cs*) is more genetically diverse than *Chlamydia trachomatis* (*Ct*). Shown is the mean pairwise genetic difference between *Cs* strains in a multiple sequence alignment (MSA) (ϕ) for (*A*) synonymous and (*B*) nonsynonymous sites compared with *Ct*. The measure of diversity based on the number of segregating mutations in a MSA (θ_w) for synonymous and nonsynonymous sites is shown in panels (*C*) and (*D*), respectively. The Bayesian Skyline Plot (BSP) analysis implemented in the BEAST shows the effective population size of (*E*) *Ct* and (*F*) *Cs* over time. The black line indicates the posterior mean whereas the blue shapes show the true demographic function. The black line indicates the posterior mean whereas the blue shapes show the true demographic function.

harbored more pairwise genetic diversity (measured by ϕ) than *Ct*. For nonsynonymous sites, the mean estimate for *Cs* for ϕ 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 θ_w: 2.68 and 1.15, respectively) and nonsynonymous (mean estimate θ_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* (1.872×10^{-5} per site/year; 95% HPD: 1.3712×10^{-5}–2.348×10^{-5}) than for *Ct* (3.39×10^{-7} per site/year). The effective population size was calculated using both synonymous measures of genetic diversity (θ_w and ϕ) (Kimura and Crow 1964). The effective population size for *Cs* was estimated to be smaller (between 7.7×10^4 and 8.6×10^4) than *Ct* (8.5×10^5 and 9.7×10^5) 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.

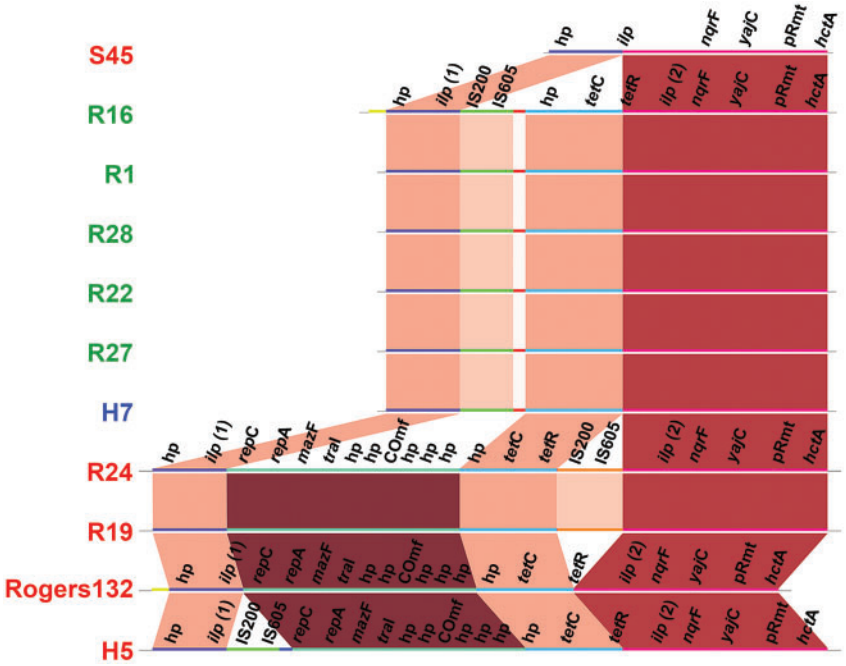

Fig. 4.—The variable structure of the *tet*(C) cassette among the *Chlamydia suis* (*Cs*) strains. The plot represents the genes within the tetracycline resistance [*tet*(C)] cassette and its neighboring genes for each of the ten newly sequenced, tetracycline resistant strains as well as the gene order of the same region in the tetracycline sensitive strain S45. Clade 1 strains (S45, R19, R24, Rogers132, H5) are shown in red, Clade 2 strains (R1, R16, R22, R27, R28) in green and H7 is highlighted in blue. The key to gene abbreviations is as follows: hp, hypothetical protein; *ilp*, invasion gene; IS200, transposase; IS605, putative transposase; *rep*C, replication protein C; *rep*A, replication protein A; *maz*F, Toxin MazF; *tra*I, conjugal transfer relaxase TraI; COmt, CO dehydrogenase maturation factor; *tet*C, tetracycline resistance gene class C; *tet*R, tetracycline repressor gene; *nqr*F, Na(+)-translocating NADH-quinone reductase subunit F; *yaj*C, preprotein translocase subunit YajC; pRmt, putative RNA methyltransferase; and *hct*A, 18 kDa histone analog.

Previously, Dugan *et al.* (2004) demonstrated that the cassettes 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, 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), *tet*R(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

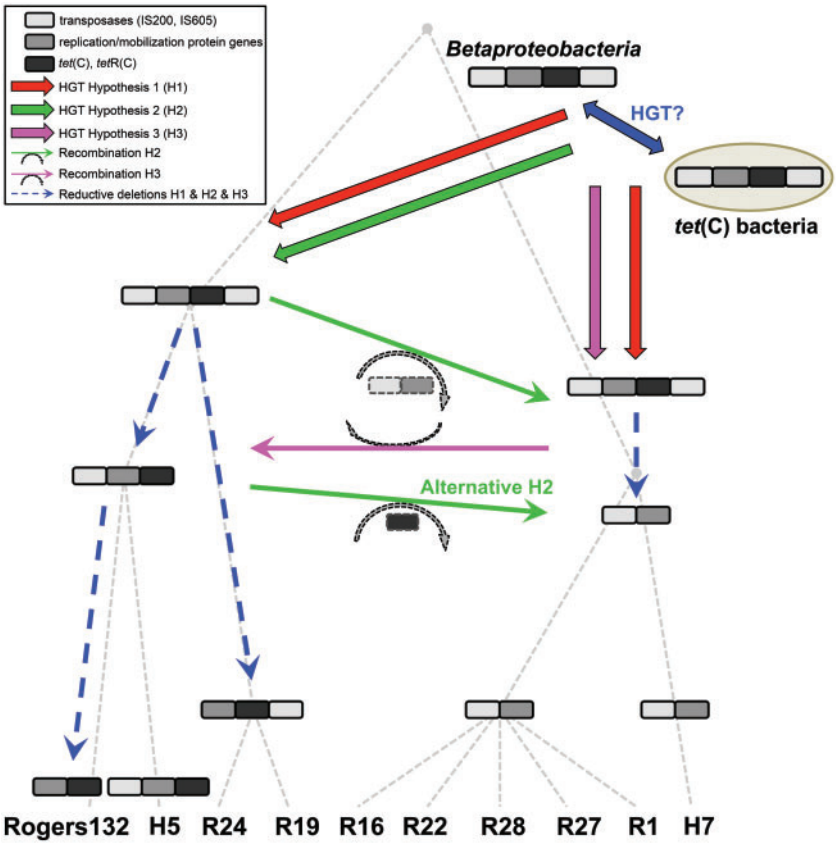

Fig. 5.—Proposed hypotheses for *tet*(C)-cassette acquisition within and across *Chlamydia suis* (*Cs*) clades 1 and 2. The *Cs tet*(C) cassette either originated from *Betaproteobacteria* or from a tetracycline resistant bacteria (upper right cartoons). Based on BEAST molecular dating showing that the two clades predated the introduction of the *tet*(C) cassette, we propose three hypotheses: Hypothesis 1 (H1, red): both clades acquired the *tet*(C) cassette through horizontal gene transfer (HGT) independently from each other and then lost parts of the cassette through reductive deletions (blue); or Hypothesis 2 (H2, green): a single HGT event led to the acquisition of the cassette in Clade 1 before it moved to Clade 2 through homologous recombination, with or without losing the replication/mobilization protein genes in the process; or Hypothesis 3 (H3, magenta): a single HGT event led to the acquisition of the cassette in Clade 2 before it moved to Clade 1 through homologous recombination and then lost parts of the cassette through reductive deletions (blue).

by an index strain and exchanged between *Cs* strains by intra-species 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 Vrbanac 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

Borel N, et al. 2012. Selection for tetracycline-resistant *Chlamydia suis* in treated pigs. Vet Microbiol. 156:143–146.

Borges V, et al. 2013. Effect of long-term laboratory propagation on *Chlamydia trachomatis* genome dynamics. Infect Genet Evol. 17:23–32.

Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172: 2665–2681.

CDC—Chlamydia Statistics. [cited 2016 March 22]. Available from: http://www.cdc.gov/std/chlamydia/stats.htm.

Chaisson MJ, Pevzner PA. 2008. Short read fragment assembly of bacterial genomes. Genome Res. 18:324–330.

Darling AE, Mau B, Perna NT. 2010. Progressive mauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147.

Dean D, Kandel RP, Adhikari HK, Hessel T. 2008. Multiple Chlamydiaceae species in trachoma: implications for disease pathogenesis and control. PLoS Med 5:e14.

Dean D, Rothschild J, Ruettger A, Kandel RP, Sachse K. 2013. Zoonotic Chlamydiaceae species associated with trachoma, Nepal. Emerg. Infect. Dis 19:1948–1955.

De Puysseleyr K, et al. 2014. Evaluation of the presence and zoonotic transmission of *Chlamydia suis* in a pig slaughterhouse. BMC Infect Dis. 14:560.

De Puysseleyr L, et al. 2015. Assessment of *Chlamydia suis* infection in pig farmers. Transbound Emerg Dis. doi: 10.1111/tbed.12446.

Didelot X, et al. 2011. Recombination and population structure in *Salmonella enterica*. PLoS Genet. 7:e1002191.

Didelot X, Falush D. 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics 175:1251–1266.

Didelot X, Lawson D, Darling A, Falush D. 2010. Inference of homologous recombination in bacteria using whole-genome sequences. Genetics 186:1435–1449.

Donati M, et al. 2014. Genome sequence of *Chlamydia suis* MD56, isolated from the conjunctiva of a weaned piglet. Genome Announc. 2(3):1 doi: 10.1128/genomeA.00425-14.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7:214.

Dugan J, Rockey DD, Jones L, Andersen AA. 2004. Tetracycline resistance in *Chlamydia suis* mediated by genomic islands inserted into the chlamydial inv-like gene. Antimicrob Agents Chemother. 48:3989–3995.

FAO's Animal Production and Health Division: Meat & Meat Products [Cited 2016b March 22]. Available from: http://www.fao.org/ag/againfo/themes/en/meat/backgr_sources.html.

Galardini M, Biondi EG, Bazzicalupo M, Mengoni A. 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. Source Code Biol Med. 6:11.

Gillings MR. 2013. Evolutionary consequences of antibiotic use for the resistome, mobilome and microbial pangenome. Front Microbiol. 4:4.

Goldschmidt P, et al. 2006. Detection by broad-range real-time PCR assay of Chlamydia species infecting human and animals. Br J Ophthalmol. 90:1425–1429.

Gomes JP, et al. 2007. Evolution of *Chlamydia trachomatis* diversity occurs by widespread interstrain recombination involving hotspots. Genome Res. 17:50–60.

Gomes JP, Bruno WJ, Borrego MJ, Dean D. 2004. Recombination in the genome of *Chlamydia trachomatis* involving the polymorphic membrane protein C gene relative to ompA and evidence for horizontal gene transfer. J Bacteriol. 186:4295–4306.

Hammerschlag MR. 2013. Treatment of *Chlamydia trachomatis* infections. Basel, Switzerland: Karger Medical and Scientific Publishers.

Harris SR, et al. 2012. Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. Nat Genet. 44(413):9–S1.

Heled J, Drummond AJ. 2008. Bayesian inference of population size history from multiple loci. BMC Evol Biol. 8:289.

Joseph SJ, et al. 2012. Population genomics of *Chlamydia trachomatis*: insights on drift, selection, recombination, and population structure. Mol Biol Evol. 29:3933–3946.

Joseph SJ, et al. 2015. Chlamydiaceae genomics reveals interspecies admixture and the recent evolution of *Chlamydia abortus* infecting lower mammalian species and humans. Genome Biol Evol. 7:3070–3084.

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.

Kaltenboeck B, Kousoulas KG, Storz J. 1993. Structures of and allelic diversity and relationships among the major outer membrane protein (ompA) genes of the four chlamydial species. J Bacteriol. 175:487–502.

Kimura M, Crow JF. 1964. The number of alleles that can be maintained in a finite population. Genetics 49:725–738.

Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178–2189.

Li Y, et al. 2015. The clinical characteristics and genotype distribution of *Chlamydia trachomatis* infection in infants less than six months of age hospitalized with pneumonia. Infect Genet Evol. 29:48–52.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A. 76:5269–5273.

Oppliger A, et al. 2012. Antimicrobial resistance of *Staphylococcus aureus* strains acquired by pig farmers from pigs. Appl Environ Microbiol. 78:8010–8014.

Pérez-Losada M, et al. 2007. Distinguishing importation from diversification of quinolone-resistant *Neisseria gonorrhoeae* by molecular evolutionary analysis. BMC Evol Biol. 7:84.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol. 31:1929–1936.

Read TD, et al. 2013. Comparative analysis of *Chlamydia psittaci* genomes reveals the recent emergence of a pathogenic lineage with a broad host range. MBio 4:(2):e00604–e00612. doi: 10.1128/mBio.00604-12.

Read TD, et al. 2003. Genome sequence of *Chlamydophila caviae* (*Chlamydia psittaci* GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaceae. Nucleic Acids Res. 31:2134–2147.

Roberts MC. 2003. Tetracycline therapy: update. Clin Infect Dis. 36:462–467.

Roberts MC. 2005. Update on acquired tetracycline resistance genes. FEMS Microbiol Lett. 245:195–203.

Rogers DG, Andersen AA. 1996. Intestinal lesions caused by two swine chlamydial isolates in gnotobiotic pigs. J Vet Diagn Invest. 8:433–440.

Rogers DG, Andersen AA. 1999. Conjunctivitis caused by a swine *Chlamydia trachomatis*-like organism in gnotobiotic pigs. J Vet Diagn Invest. 11:341–344.

Rogers DG, Andersen AA, Hogg A, Nielsen DL, Huebert MA. 1993. Conjunctivitis and keratoconjunctivitis associated with chlamydiae in swine. J Am Vet Med Assoc. 203:1321–1323.

Schautteet K, et al. 2013. Tetracycline-resistant *Chlamydia suis* in cases of reproductive failure on Belgian, cypriote and Israeli pig production farms. J Med Microbiol. 62:331–334.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

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

Suchland RJ, Sandoz KM, Jeffrey BM, Stamm WE, Rockey DD. 2009. Horizontal transfer of tetracycline resistance among *Chlamydia* spp. in vitro. Antimicrob Agents Chemother. 53:4604–4611.

Suerbaum S, Josenhans C. 2007. Helicobacter pylori evolution and phenotypic diversification in a changing host. Nat Rev Microbiol. 5:441–452.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609–W612.

Tian B, Fadhil NH, Powell JE, Kwong WK, Moran NA. 2012. Long-term exposure to antibiotics has caused accumulation of resistance determinants in the gut microbiota of honeybees. MBio 3(6):e00377–e00312. doi: 10.1128/mBio.00377-12

Vos M, Didelot X. 2009. A comparison of homologous recombination rates in bacteria and archaea. ISME J. 3:199–208.

Wardyn SE, et al. 2015. Swine farming is a risk factor for infection with and high prevalence of carriage of multidrug-resistant *Staphylococcus aureus*. Clin Infect Dis. 61:59–66.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 7:256–276.

Associate editor: Bill Martin