Directional Evolution of *Chlamydia trachomatis* towards Niche-Specific Adaptation

Vitor Borges, Alexandra Nunes, Rita Ferreira, Maria J. Borrego, and João P. Gomes

Department of Infectious Diseases, National Institute of Health, Lisbon, Portugal

On behalf of the host-pathogen “arms race,” a cutting-edge approach for elucidating genotype-phenotype relationships relies on the identification of positively selected loci involved in pathoadaptation. We studied the obligate intracellular bacterium *Chlamydia trachomatis*, for which same-species strains display a nearly identical core and pan genome, while presenting a wide range of tissue tropism and ecological success. We sought to evaluate the evolutionary patterns underlying species separation (divergence) and *C. trachomatis* serovar radiation (polymorphism) and to establish genotype-phenotype associations. By analyzing 60 *Chlamydia* strains, we detected traces of Muller’s ratchet as a result of speciation and identified positively selected genes and codons hypothetically involved in the infection of different human cell types (e.g., columnar epithelial cells of ocular or genital mucosae and mononuclear phagocytes) and also events likely driving pathogenic and ecological success dissimilarities. In general, these genes code for proteins involved in immune response elicitation, proteolysis, and the subversion of host-cell functions, and also for proteins with unknown function(s). Several genes are potentially involved in more than one adaptive process, suggesting multiple functions or a distinct *modus operandi* for a specific function, and thus should be considered as crucial research targets. In addition, six of the nine genes encoding the putative antigen/adhesin polymorphic membrane proteins seem to be under positive selection along specific serovars, which sustains an essential biological role of this extra-large paralogue family in chlamydial pathobiology. This study provides insight into how evolutionary inferences illuminate ecological processes such as adaptation to different niches, pathogenicity, or ecological success driven by arms races.

Genomic changes of microbial pathogens are directly linked to the evolutionary “arms race” that takes place between microbe and host during the infectious process, as a result of the antagonistic interaction, and they are a consequence of polymorphisms accumulated after selective pressure from the host’s inflammatory or immune response (32). However, the majority of coding genes present a higher number of synonymous rather than nonsynonymous substitutions, which indicates that purifying selection is operating to preserve the current function and structure of the protein, and only a small fraction of the genes are expected to be positively selected where diversification is favored through increased fitness (11). In order to understand the evolutionary forces that act on gene variation, major challenges are to identify loci that might have been under selection and to determine the type of natural selection that has influenced their evolutionary history (59). In the field of infectious diseases, site-specific inferences regarding positive selection on loci involved in drug resistance (19) or in the interaction with the host immune system have been proposed as complementary approaches for the development of vaccines against HIV and other viruses (33) and also to predict the evolution of virulent strains of the influenza virus (12). Also, it has been shown that core genes are equally subjected to positive selection as pathogen specific accessory genes (4), suggesting that blind genomic-scale analysis should be performed.

For a species such as *Chlamydia trachomatis* with a wide range of tissue tropism and ecological success, but presenting a nearly identical core and pan genome, and a DNA sequence similarity of >98% (39), the few existing polymorphisms are expected to be extremely informative of the adaptive evolution process. However, an excess of nonsynonymous substitutions alone is not sufficient to invoke positive selection, since it requires an increase in fitness caused by the corresponding amino acid replacement. Otherwise, it may represent the accumulation of slightly deleterious

mutations (not severely deleterious as these will not become fixed because they render their bearers nonviable) to the pathogen on behalf of Muller’s ratchet theory (27, 60). This is predicted to operate in intracellular replicating bacteria (as *C. trachomatis*, which replicates within a host vacuole named inclusion) that are subject to recurrent bottlenecks and replicate in small populations, with little opportunity for recombination and few back or compensatory mutations (2). Although it was recently shown (39) that recombination events affect many more chromosome regions than previously suspected in *C. trachomatis*, the frequency and the relative weight of recombination and mutation calculated for this pathogen ($p/\theta < 0.07$ and $r/m < 0.71$, respectively) (28, 49) indicates the point mutation events as the major evolutionary driving force.

In the present study, we used comparative genomics over 59 *C. trachomatis* strains (comprising all serovars) to clarify the mutational dynamics underlying both the separation of *C. trachomatis* as species, and the pathoadaptation driven by this arms race. We identified positively selected genes and codons that are hypothetically involved in the evolutionary adaptation of *C. trachomatis* serovars to different cell types: mucosal cells from the eye conjunctiva (responsible for trachoma) (serovars A to C), mucosal cells from the genitalia (primarily yielding cervicitis) (serovars D to K), and mononuclear phagocytes (yielding invasive diseases such as

Received 19 July 2012 Accepted 1 September 2012

Published ahead of print 7 September 2012

Address correspondence to João P. Gomes, j.paulo.gomes@insa.min-saude.pt.

Supplemental material for this article may be found at <http://jb.asm.org/>.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.01291-12

hemorrhagic proctitis and suppurative lymphadenitis) (serovars L1 to L3). Finally, we also detected positive selection events likely driving pathogenic and ecological success dissimilarities.

MATERIALS AND METHODS

***C. trachomatis* strains, cell culture, and DNA extraction.** The present study encompasses data from 59 *C. trachomatis* strains and the *Chlamydia muridarum* Nigg strain (also called the mouse pneumonitis strain [MoPn]) (see Table S1 in the supplemental material). These include *in silico* data from recently analyzed fully sequenced *C. trachomatis* strains (39) and eight historical prototype strains (Ba/Apache-2, C/TW3, F/IC-Cal3, G/UW57, H/UW43, I/UW12, J/UW36, and K/UW31) in order to enroll all 15 major serovars. The additional eight strains were propagated in HeLa 229 cell monolayers and, at 48 to 72 h postinfection, the cells were harvested, and a bacterium-enriched pellet was obtained and resuspended in 200 μ l of phosphate-buffered saline, as previously described (9). DNA was extracted using a QIAamp DNA minikit (Qiagen) according to the manufacturer's instructions and stored at -80°C until use.

Selection of loci and sequencing. Based on available *in silico* full-genome sequence data, we searched for polymorphic genes among *C. trachomatis* strains through the progressiveMauve algorithm (23) of the Mauve software v2.3.1. A detailed evaluation of the polymorphism of each locus was further performed using Lasergene 9.0 (DNASTAR, Madison, WI) and MEGA5 (76). Chromosome loci revealing an extremely low polymorphism were discarded from the present analysis since their use would hamper the accurate application of likelihood tests. We ended up with 75 top-ranked polymorphic genes (see Table S2 in the supplemental material). These were categorized according to their functional role, involving 20 housekeeping genes (HKs), 14 genes encoding well-known cell envelope proteins (CEPs), 31 genes coding for secreted proteins (SECs), and 10 genes coding for proteins with unknown function or for which the biological role is not consensual. The SEC category involves proteins secreted (either by the type III secretion system [T3SS], a machinery used by many bacterial pathogens to manipulate eukaryotic host cells by injecting virulence proteins, or by an undefined mechanism) into the cytosol of the host cells or to the inclusion membrane. For analyses enrolling divergence versus polymorphism, the corresponding orthologous genes of the *Chlamydia muridarum* Nigg strain were identified (by NCBI-BLAST search), and sequences were collected from the full-genome annotated in the GenBank database (accession number NC_002620) (65). For the strains that we needed to propagate because no *in silico* data were available, 75 genes were amplified and sequenced using standard procedures (36). The sequences and location of primers, as well as the amplicon sizes, are listed in the Table S3 in the supplemental material. Automated sequencing was achieved using BigDye Terminator v1.1 cycle sequencing chemistry, according to the manufacturer's instructions (Applied Biosystems), in an Applied Biosystems 3130xl genetic analyzer. Sequence reads were assembled using SeqBuilder software (DNASTAR), and alignments were generated using the CLUSTAL W algorithm implemented in both the MegAlign software (DNASTAR) and MEGA5. A concatenated alignment of the 75 genes was also constructed for all *C. trachomatis* and *C. muridarum* strains. Since the CLUSTAL W program generates alignment artifacts in the presence of insertion/deletion (indel) events by disrupting codons, we edited "by hand" the amino acid alignments rather than only automate the process before editing the corresponding nucleotide sequences. When strain-exclusive single nucleotide polymorphisms (SNPs), indel events and pseudogenes were identified, resequencing was performed from a newly extracted DNA, and new sequences reads were generated for comparative purposes.

Phylogenomic analysis. Analyses of genetic diversity and phylogeny were conducted for each gene by using MEGA5. Briefly, we computed the overall mean distances (number of differences and p-distance) and matrices of pairwise comparisons at both nucleotide and amino acid levels. For phylogenetic analysis, individual trees were generated by using the neighbor-joining method with bootstrapping (67), and the evolutionary

distances were computed using the Kimura two-parameter method (52). For all of these analyses, the pairwise deletion option was selected since it excludes sites containing alignment gaps or missing data from the analysis only when necessary in the pairwise distance estimation. Truncated genes, which are expected to encode nonfunctional proteins, were excluded from the phylogenetic and evolutionary analyses, except for the strains with nondisrupted sequences, because their biological role may be phenotype specific.

Global analysis of molecular evolution. The nonsynonymous/synonymous substitution rate ratio (dN/dS) among related protein-coding DNA sequences, where dN refers to the number of nonsynonymous substitutions per nonsynonymous site and dS is defined as the number of synonymous substitutions per synonymous site, may be suggestive of the selective pressures driving the mutational trends (86). Initially, for a global analysis of these trends, we estimated the dN and dS values with MEGA5 by using the Kumar model (61). For each gene, the dN/dS was calculated over all *C. trachomatis* sequence pairs and between the sequences of the two species (*C. trachomatis* and *C. muridarum*). Moreover, in order to reinforce the comparison between the amount of evolutionary variation within the *C. trachomatis* species (polymorphism) and the variation between *C. trachomatis* and *C. muridarum* (divergence), we also applied the McDonald-Kreitman (MK) test (26, 55). However, since it has been assumed that the results from the MK test cannot directly discriminate the type of selection acting on genes (62), the subjacent MK test algorithm was only used to clarify the neutral and amino acid-altering mutational trends underlying the *C. trachomatis* speciation process. This kind of analysis is suitable for tracing the Muller's ratchet phenomenon, which is commonly observed in niche-restricted pathogens.

Evaluation of the directionality in *C. trachomatis* evolution. In order to search for genes on which positive selection putatively operates, two distinct approaches were applied. First, as a statistical support of the dN and dS estimations within *C. trachomatis* strains, the codon-based Z-test of selection was computed by MEGA5 using the Kumar method (61), where bootstrapping (1,000 replicates) was used for estimation of the variation in the statistic test. This test calculates the probability of rejecting the null hypothesis of strict neutrality ($dN = dS$) in favor of one of two alternative hypothesis: positive selection ($dN > dS$) or purifying selection ($dN < dS$). Results with P values of <0.05 were considered significant at the 5% level. On a second approach, the branch site test of positive selection (branch site test 2) (85, 88) was used using the codeml application from the phylogenetic analysis by maximum-likelihood (PAML) package (version 4.4d) (83). Alignments of nucleotide sequences from the 59 *C. trachomatis* strains and *C. muridarum* (built and corrected on MEGA5) were converted into the "interleaved" PHYLIP format using the BioEdit package (version 7.0.0; <http://www.mbio.ncsu.edu/bioedit/bioedit.html>), where stop codons were removed from sequences. The branch site test is a robust bioinformatic approach (84) that is recommended to infer positive selection in a lineage of interest (called foreground lineage) when several lineages in the phylogeny may have been subjected to distinct selective pressures (85, 88). The statistical significance of the presence of positive selection along the branch of interest was addressed by the likelihood-ratio test (LRT) (82). In branch site test 2, the LRT compares the twice of the log likelihood difference ($2\Delta\ell$) between two models (alternative and null models) with the chi-square distribution with one degree of freedom for P value calculation (88). The alternative model allows positive selection ($dN/dS \geq 1$) for the foreground branch, whereas the null model assumes the dN/dS ratios of ≤ 1 for all site classes in all branches in the phylogeny. When positive selection acting on a specific gene was suggested by a significant LRT ($P < 0.05$), a Bayes empirical Bayes analysis (87) was used to identify the specific positively selected sites within that gene along the foreground branches. Therefore, the branch site model requires an *a priori* definition and labeling of the foreground branches to be tested for positive selection, which should rely on well-defined biological hypotheses (85). Thus, based on the assumption that some genes might be involved in *C. trachomatis* phenotypic dissimilarities as a result

FIG 1 Chromosomal mapping of loci involved in the directional evolution of *C. trachomatis*. From the outside in, the first and second circles (light blue lines) refer to forward and reverse coding regions, respectively, according to the published genome of the *C. trachomatis* strain D/UW3. The 75 evolutionary informative genes evaluated in the present study are highlighted by dark blue lines. These loci encompass ca. 55% of the SNPs occurring within the chromosomal coding regions. The third (orange lines) and fourth circles (red lines) illustrate genes found to be under positive selection by the codon-based Z-test of positive selection (MEGA5) or the branch site test of positive selection (PAML), respectively. Circle 5 shows the GC skew plot. The origin of replication (ORI) and the termination region (TER) are also marked. The figure was built using DNAPlotter (16).

of targeted positive selective pressures, we created six comprehensible biological hypotheses (H1 to H6). The hypotheses evaluate the existence of genes under positive selection that may be involved in the following biological processes: specific cell appetite to columnar epithelial cells of ocular (H1) or genital (H2) mucosae and to mononuclear phagocytes (H3), pathogenic diversity among strains causing ocular disease (H4), genital disease (H5), or hemorrhagic proctitis and suppurative lymphadenitis (H6). Only genes for which the phylogeny supported any of these scientific hypotheses were tested.

Finally, since recombination may bias the results of positive selection, we used published data on recombination analysis enrolling all *C. trachomatis* genes (39, 41, 49) to inspect whether the genes selected for the present study showed evidence of recombination. Consequently, for the genes showing incongruent trees where unequivocal recombination was detected within a specific branch, the analysis of positive selection was excluded *a priori* for the corresponding biological hypothesis. On the other hand, genes yielding congruent trees but for which recombination had been previously detected (39, 41, 49) were still subjected to positive selection analysis and are properly identified in the present study.

Nucleotide sequence accession numbers. The nucleotide sequences determined in the present study were submitted to the GenBank database (<http://www.ncbi.nlm.nih.gov/GenBank/index.html>) and are currently available for consulting under accession numbers JQ066324 to JQ066722.

RESULTS

Polymorphism significance of the selected genes. The distribution of point mutations in *C. trachomatis* chromosome is highly heterogeneous. Although the selected genes (Fig. 1; see also Table S2 in the supplemental material) represent 11% of the coding region length, they encompass ca. 55% of all chromosomal SNPs occurring within coding regions, which corresponds to a total of 5,083 polymorphic sites among the 59 strains. In fact, we found that any given chromosomal SNP has a 10.0 times higher probability (odds ratio 95% confidence interval = 9.3 to 10.7) to belong to the pool of genes under evaluation than to show up in any other gene (Fisher exact test, $P < 10^{-7}$).

FIG 2 Evidence for Muller's ratchet phenomenon. These graphs show non-synonymous versus synonymous mutational dynamics on the *C. trachomatis*/*C. muridarum* separation process. The scattering plot depicts the results concerning the evaluation of nonsynonymous and synonymous substitutions within the *C. trachomatis* species (open circles) (reflecting polymorphism) and between the species *C. trachomatis* and *C. muridarum* (crosses) (reflecting divergence). The neutrality line is also shown. The box plots display the dispersion of the overall decays in the dN and dS values [i.e., $dN(\text{divergence})/dN(\text{polymorphism})$ and $dS(\text{divergence})/dS(\text{polymorphism})$, respectively]. Outliers and extreme values are indicated by open circles and asterisks, respectively. The considerable lower decay values for dN suggests that the accumulation of deleterious mutations among strains from *C. trachomatis* species results from genetic bottleneck due to niche restriction (Muller's ratchet effect).

Divergence versus polymorphism: detection of Muller's ratchet. Considering that *C. trachomatis* and *C. muridarum* species evolved from a last common ancestor (65), the comparison between the amount of evolutionary variation within the *C. trachomatis* species (polymorphism) and between *C. trachomatis* and *C. muridarum* (divergence) may shed some light on the evolutionary mutational dynamics that drove the *C. trachomatis* speciation. Accordingly, the divergence of the two species was evaluated through estimation of the dN/dS ratio between orthologous genes. All genes revealed dN/dS values lower than 1, where the mean value was 0.21 (standard deviation [SD] = 0.14) (Fig. 2). This observation suggests an unequivocal higher weight of synonymous than nonsynonymous changes on the species divergence, in agreement with the neutral theory of molecular evolution (53), which postulates that the fixation of selectively neutral mutations by random genetic drift is the major factor responsible for species divergence.

Subsequently, we compared these results with the dN/dS values obtained solely within *C. trachomatis* species (i.e., among the 59 strains). We observed that there was a high, but dissimilar, decrease in both dS and dN values after the separation of the two species from a common ancestor (Fig. 2). In fact, the observed mean of the decay of dS values ($147.2 [SD \pm 243.7]$) was 5.5 times higher than the one observed for dN ($26.9 [SD \pm 19.9]$). This was a consistent trend, as the MK test algorithm yielded a similar decays ratio of 4.7. Globally, this suggests that, since *C. trachomatis* was established as the species, the nonsynonymous changes increased their relative weight to synonymous changes in contrast with the evolutionary process that originated the separation of the two species. This observation is consistent with the Muller's ratchet theory (27, 60), which assumes an accumulation of slightly

FIG 3 dN versus dS by gene functional category. This graph represents dN and dS values estimated for each gene for all 59 *C. trachomatis* strains. Housekeeping genes (HKs), genes encoding well-known cell envelope proteins (CEPs), genes coding for proteins secreted into the cytosol of the host cells or to the inclusion membrane (SEC), and “other genes” (see Materials and Methods and Table S2 in the supplemental material for details) are represented by squares, circles, triangles, and crosses, respectively. The neutrality line is also shown.

deleterious nonsilent mutations on microbial populations repeatedly subjected to genetic bottlenecks. In fact, the obligate intracellular lifestyle of *C. trachomatis*, which is characterized by niche-restricted and low-size populations, and expected low frequency of recombination relative to mutation events (28, 49) may lead to less effective elimination and consequent accumulation of nonsilent mutations (2).

Evolutionary trends within the *C. trachomatis* species. We further focused on studying the accumulation of mutations after *C. trachomatis* speciation, with special emphasis on protein-altering changes that have contributed to phenotype divergences, which may help to clarify the *C. trachomatis* niche-specific adaptation. For each gene, the dN/dS ratio was evaluated over all *C. trachomatis* sequence pairs (Fig. 3), where genes exhibiting dN/dS ratios greater than 1 and a significant P value (<0.05) in the codon-based Z-test of positive selection ($dN/dS > 1$) were considered as putative targets of positive selection. Twenty-seven genes exhibited overall dN/dS values higher than 1, in which 15 (including 14 SECs) revealed a significant Z-test P value. As predicted, all housekeeping genes presented dN/dS ratios of less than 1, which indicates that the genes involved in regulatory/metabolic functions are less likely targeted by diversifying selection (47). On the other hand, 22 of the 31 SECs support an opposite scenario, which is relevant since these proteins contact directly with the host and thus are more prone to be involved in pathoadaptation. We also investigated whether the types of mutation are dependent on the degree of genetic variability by evaluating the relationship between dN and dS values and the nucleotide polymorphism (i.e., the p-distance) among *C. trachomatis* strains (Fig. 4). Globally, as observed above for species divergence, we found that the increment in polymorphism is essentially driven by fixation of silent mutations, which presented an increase rate ~ 4 -fold higher than for the nonsilent changes.

Distribution of dN and dS versus disease outcomes and ecological success. We also investigated whether the general distribution of both silent and nonsilent mutations among *C. trachomatis* corresponds to strains clustering by disease outcomes. We used

FIG 4 Genetic variability versus type of mutation. Distribution of dN (triangles) and dS (circles) values according to the nucleotide polymorphism (mean genetic p-distance) of the 75 genes under evaluation among *C. trachomatis* strains. The slope values of the trend lines indicate a nearly 4-fold-higher increase of dS with p-distance than of dN .

the concatenated sequences encompassing all genes under evaluation to calculate the dN and dS distances between each strain and the different groups of strains (i.e., three disease groups) (Fig. 5). Our results sustain a nonrandom accumulation of mutations where strains with the same cell appetite are unequivocally clustered either by silent mutations or protein-changing alterations. Indeed, the genetic distances between strains with dissimilar tropism are 1.8- to 10.5-fold (for dS) and 1.7- to 6.7-fold (for dN) higher than the ones between strains with similar cell appetite. In addition, the highly ecological succeeded strains causing non-invasive genital infections (mostly from serovars E and F) are slightly separated from the remainder genital strains (Fig. 5). This analysis clearly supports that our approach for detecting positively selected genes (and codons) relying on rationally established biological hypotheses (see Materials and Methods) may be a useful step for understanding the molecular basis underlying *C. trachomatis* phenotypic differences.

FIG 5 Nonrandom distribution of both nonsynonymous and synonymous mutations according to tropism and ecological success. The three-dimensional scatter plot shows the genetic distances between each of the 59 strains and the three disease groups by both dN (gray) and dS (black) estimations. Values were estimated by using the concatenated sequences enclosing all genes under evaluation. Strains infecting mononuclear phagocytes (shown in the bottom of the cube), the columnar epithelial cells of ocular (shown in the left side of the cube), or genital mucosae compose three major clusters for both dS and dN . Within the non-invasive genital strains, the more clinically prevalent strains (labeled with an ellipse) are clustered apart.

FIG 6 Positive selection driving the directional *C. trachomatis* evolution toward niche-specific adaptation. The figure represents a model tree encompassing all 59 *C. trachomatis* strains that was created to facilitate a proper visualization of all biological hypotheses. These evaluate the existence of genes under positive selection (through the branch site test of positive selection) that may be involved in the following: specific cell appetite to columnar epithelial cells of ocular (H1) (serovars A to C) or genital (H2) (serovars D to K) mucosae and to mononuclear phagocytes (H3) (serovars L1 to L3) and pathogenic diversity among strains causing ocular disease (H4), genital disease (H5), or hemorrhagic proctitis and suppurative lymphadenitis (H6). This test was applied to each individual gene tree in an independent manner for each hypothesis, depending on the tree topology. The figure boxes show the genes found to be positively selected for each biological hypothesis. The likelihood-ratio test (LRT) was used to infer the statistical significance (P values) of positive selection in the foreground branches. ** and *, significance of $P < 0.01$ and $P < 0.05$, respectively; §, genes presenting congruent trees, but for which a specific biological hypothesis may be affected by recombination (49). See Table 1 for details on the positively selected codons.

Positive selection driving bacterial specific appetite to different human cell types. The phylogenetic analysis revealed genes whose trees cluster all strains that preferentially infect the same human cell type in a single branch. Thus, genes (and codons) targeted by positive selection along those branches may be involved on specific host-cell interactions. To evaluate this, we conducted the branch site test of positive selection under the biological hypotheses H1 to H3 (see Materials and Methods). All of the genes and the inferred positively selected codons found to be putatively involved in specific adaptive processes are described in Fig. 6 and Table 1. Five genes were found to be under positive selection in the evolutionary process that drove the segregation of ocular strains (branch H1). These include genes encoding two polymorphic membrane proteins (Pmps) (CT869/*pmpE* and CT870/*pmpF*), one Pmp-like protein (CT050), one inclusion membrane protein (Inc) (CT115/*incD*), and the translocated actin recruiting phosphoprotein (CT456/*tarp*). The Pmps and Incs are among the most promising research targets for which there is cumulative evidence of their involvement in biological mechanisms, such as adhesion, immune response elicitation, or subversion of intracellular trafficking (see Table 1 for details) (21, 25, 35, 38, 66, 68, 78). For example, PmpF was predicted *in silico* to contain T-cell epitopes that bind HLA class I and II alleles (15). On the other hand, Tarp is a chlamydial effector of the T3SS associated with the chlamydial invasion of the host cells (44) by mediating host actin polymerization and inclusion development (20, 42).

For the biological hypothesis H2, we found significant evidence supporting a fixation of adaptive mutations driving a better appetite to the columnar epithelial cells of genital mucosa (branch H2) solely for the gene CT105. Although its function is unknown, a previous study of heterologous expression in yeast

(71) suggested that CT105 may be involved in modulation of host cellular functions. Nevertheless, it is worth noting that CT105 is a pseudogene for ocular strains; thus, we may be facing a scenario of a gene strictly needed for tropism functions other than those involving the ocular conjunctiva. The frequent tree incongruence involving genital strains hampered the evaluation of several genes for this specific biological hypothesis (Table 1).

Regarding the branch that clusters all strains infecting the mononuclear phagocytes (branch H3), we detected 18 genes likely under positive selection. This set includes one HK and genes encoding 11 SECs (seven Incs), four CEPs (including three Pmps), and two proteins with unknown function (Table 1). Besides the general relevance of Incs and Pmps (explained above), we would highlight the SEC CT223, an Inc suggested to play a role in the subversion of host cell functions, either by containing SNARE-like (eukaryotic soluble *N*-ethylmaleimide-sensitive attachment protein receptors) motifs (impact in intracellular trafficking) (25) or by blocking host cell cytokinesis (1). The SEC CT622 is an antigen putatively secreted by the T3SS (21, 30, 37), whereas CT867 and CT868 are proteases that possess deubiquitinating and deneddylating activities (57), which may suggest a role in virulence. The invasive infection patterns of L1 to L3 strains, along with the expectation that these strains were the first to diverge from a common *C. trachomatis* ancestor (73), may justify the high number of genes detected under diversifying selection on branch H3 rather than on branches H1 and H2. These genes may play a role in the specificity of mononuclear phagocytes-bacteria interactions that yield invasive infections with L1 to L3 strains.

Positive selection driving pathogenic diversity among strains infecting the same human cell type. The detection of positive selection acting on specific genes along branches of strains

TABLE 1 Positively selected genes and inferred codons putatively involved in specific adaptive evolution based on the branch site test of positive selection by PAML

Biological hypothesis with positive selection (H1 to H6) ^a			Specific codons under positive selection ^d		Putative function/expt evidence ^c		Specific biological hypothesis excluded due to putative recombination ^f
ORF ^c (gene)	<i>P</i> (LRT test) ^c						
CT050	H1	<10 ⁻¹⁰	ND		Pmp-like protein identified in the inclusion Lumen (48, 71)		H2, H5
	H3	0.0352	51K*, 80K*, 366K*, 523S*, 546K*				
	H6	<10 ⁻⁷	183N*, 184S*, 223D**				
CT051	H3	0.0004	563V*		Pmp-like protein (48, 71)		H2, H5
	H6	0.0002	435R**				
CT105	H2	0.0004	ND		Function unknown		–
CT115 (<i>incD</i>)	H1	<10 ⁻⁴	11D**, 12G*		Inc (68), T3SS effector (75)		H2, H5
	H3	<10 ⁻⁹	150L*, 151E*, 153S*, 155S**				
CT147	H3	<10 ⁻⁵	1407I*		Inc; human early endosomal antigen 1 (EEA1) homologue (8), immunodominant antigen (69), involvement in pathogenic differences <i>in vivo</i> (51)		H1, H2, H4, H5
	H6	0.0003	ND		Inc associated with host kinases in microdomains that interact with the host centrosomes (58)		–
CT222	H3	<10 ⁻⁷	124I*, 125S**, 126V**		Inc (6), T3SS effector (75)		–
CT223	H3	0.0051	127G*, 206R*		Subversion of intracellular trafficking (25), host cell cytokinesis blockage (1)		–
	H5	0.0082	99R*, 152S*				
CT233 (<i>incC</i>)	H6	0.0432	5 M*, 6S**, 7D**, 8I*, 11K*, 14I*		Inc (7), T3SS effector (29, 75)		–
	H3	0.0004	8Y*, 24N*, 80T**, 89I**		Inc (45), T3SS effector (24)		–
CT249	H3	0.0104	ND		Inc (6), T3SS effector (75)		H1, H2, H4, H5
CT288	H3	0.0015	ND		Adhesin (38), antigen (21, 35, 78)		H2, H5
	H4	0.0198	58A*, 75T**, 235A*, 820A*, 947E*, 998A*, 1061N**, 1171T**				
CT414 (<i>pmpC</i>)	H6	0.0017	ND		Adhesin (38), antigen (21, 35, 78)		H2, H5
	H6	<10 ⁻⁶	145P*, 544I**, 598V**		Inc (6), T3SS effector (75), antigen (72)		–
CT442 (<i>crpA</i>)	H6	0.0154	24A*, 29K**, 48I*, 104I**, 133D*, 137V**		Translocated actin-recruiting phosphoprotein/early T3SS effector (20), contribution for the pathogen phagocytosis (44), involvement in pathogenic differences <i>in vivo</i> (51), antigen (30, 81)		–
CT456 (<i>tarp</i>)	H1	0.0094	447H*, 978H*				
CT529	H3	<10 ⁻²⁴	ND				
	H4	0.0049	ND				
	H5	<10 ⁻⁸	189A**, 237G*, 407S*, 481A**				
	H6	<10 ⁻²⁸	134N*, 139I*, 252G*, 301D*, 351D**, 399G*, 404D**, 493D**, 494D*, 530K*, 577R*, 603A*, 888G*, 891D**, 909K*				
CT622	H3	0.0009	3A*		Inc (31), T3SS effector (75), antigen (31, 81)		–
	H3	0.0012	ND		Involvement in pathogenic differences <i>in vivo</i> (51), T3SS effector (37), antigen (21, 30)		H1, H2, H4, H5
CT694	H7 ^g	0.0190	ND		Immunodominant antigen (21, 69), early T3SS effector (40), modulation of host cell processes (10)		–
CT852 (<i>yhgN</i>)	H3	<10 ⁻⁴	202M**, 203L*		Putative integral membrane protein (79)		H2, H5
CT859 (<i>ispH</i>)	H3	0.0135	ND		4-Hydroxy-3-methylbut-2-enyl diphosphate reductase (74)		H2, H5
CT867	H3	0.0141	ND		Deubiquitinase and deneddylase (57)		H1, H2, H4, H5
CT868	H3	0.0064	ND		Deubiquitinase and deneddylase (57)		H1, H2, H4, H5
	H6	<10 ⁻⁵	9S*, 17R**, 26S**, 30R**, 105T**, 136D**, 138I**, 157Q*, 158T**, 222P**, 238R**, 244V**, 256S**, 287R**, 300P**, 303N**, 307E**, 311F**, 322Y**, 323D**, 324S**, 325K**, 339R**, 340G**, 343S**, 354H**, 358K**, 361L**		Involvement in pathogenic differences <i>in vivo</i> (51), inhibition of a crucial pathway for host inflammatory responses (54)		
CT869 (<i>pmpE</i>)	H1	0.0348	59N*, 139I*, 469A*		Adhesin (38), antigen (35, 78)		H2, H5
CT870 (<i>pmpF</i>)	H1	<10 ⁻⁴	ND		Adhesin (38), antigen (15)		–
	H3	0.0020	ND				

CT871 (<i>pmpG</i>)	H3	0.0082	ND	Adhesin (38), antigen (21, 35, 78)	H2, H5
	H6	<10 ⁻⁴	258G*, 320K*, 324S*, 812Q**		
CT872 (<i>pmpH</i>)	H5	0.0303	ND	Adhesin (38), antigen (21, 78)	–
	H6	0.0302	ND		

^a Open reading frame (ORF) numbers are based on genome annotation of the strain D/UW3 (GenBank accession no. NC_000117).

^b The hypotheses were created to evaluate the existence of genes under positive selection involved in particular biological processes: specific cell adherence to columnar epithelial cells of ocular (H1) or genital (H2) mucosae and to mononuclear phagocytes (H3) and pathogenic diversity among strains causing ocular disease (H4), genital disease (H5), or hemorrhagic proctitis and suppurative lymphadenitis (H6).

^c The likelihood-ratio test (LRT) was used to infer the statistical significance of positive selection in the foreground branches (P). The degree of freedom is 1 for the comparisons of alternative hypothesis versus the null hypothesis in the branch site test 2.

^d The posterior probabilities that each site belongs to the site class of positive selection on the foreground lineages are inferred by the Bayes empirical Bayes analysis. Positively selected sites are indicated by asterisks (**, $P > 0.99$; *, $P > 0.95$). ND, not discriminated (an excess of positively selected codons hampered their discrimination by PAML, or the identified codons revealed $P < 0.95$). For simplification purposes, amino acid positions for biological hypotheses H1 and H4 are based on the protein sequence annotation for the strain A/Har13, whereas those for H2 and H5 are based on the annotation for strain D/UW3, and those for H3 and H6 refer to strain L2/434. Within CT456/*tarp* and CT868, the positions of the codons labeled with a “S” symbol refer to strain L1/1322/p2 since L2/434 is deleted in this region.

^e Where appropriate, references are indicated in parentheses.

^f Recombination was detected in previous studies (39, 41, 49). –, no hypothesis excluded.

^g This gene was detected to be under positive selection specifically for the most clinically prevalent genital strains (mostly from E and F serovars).

causing similar disease outcomes may be useful for understanding adaptive alterations underlying niche-specific pathogenic dissimilarities. Among the ocular strains, we detected two genes under positive selection (branches H4) (Fig. 6): one *pmp* (CT413/*pmpB*) and CT456/*tarp* (Table 1). CT456/*tarp* had been already indicated as potentially involved in distinct pathogenic patterns displayed by two ocular strains (both of serovar A) when infecting cynomolgus monkeys (51).

Regarding the detection of positive selection driving pathogenic differences among strains causing noninvasive urogenital disease (branches H5), we found the above-described virulence factors CT223 and CT456/*tarp*, as well as one *pmp* (CT872/*pmpH*). Once more, the analysis of positive selection underlying pathogenicity among noninvasive genital strains was impaired by the described recombination events involving these strains (Table 1) (39, 41, 49). Of notice was the detection of positive selection events governing the evolutionary segregation of the most succeeded genital strains (mostly from E and F serovars) for the gene coding CT694. This protein is an immunodominant antigen (69), and also demonstrated was its secretion into the host cytoplasm by the T3SS at early time points after infection (as CT456/*Tarp*), where it localizes to host cell membranes and interacts with eukaryotic AHNK, an actin-binding protein. It is believed that CT694 may act by regulating membrane fluidity or by remodeling actin filaments during invasion or early stages of *C. trachomatis* development (10, 40). Thus, differences in immune evasion strategies or in host-cell manipulation during invasion may be crucial biological processes underlying ecological success.

Despite the remarkable genomic homogeneity of *C. trachomatis* strains that infect the mononuclear phagocytes (Fig. 5), we found 11 positively selected genes along the branches embodying SNPs that distinguish these strains (branches H6). These include four *pmp*'s (CT413/*pmpB*, CT414/*pmpC*, CT871/*pmpG*, and CT872/*pmpH*), two genes encoding Pmp-like proteins (CT050 and CT051), three genes coding for Inc's (CT147, CT233/*incC*, and CT442/*crpA*), CT456/*tarp*, and CT868. Besides what was generally described above for these proteins, it is worth noting that the Inc CrpA (6) is a T3SS substrate (75) that may play a role in immune evasion since it was found to be targeted by CD8⁺ T cells in response to infection in mice (72). Also notable is that all adaptive codons inferred for the CT233/*incC* (Table 1) correspond to the IncC N-terminal domain and, more specifically, within the first 15 residues, where it is known to reside the secretion signals recognized by T3SSs (75).

Evolutionary inferences and associated bias. The branch site test may generate some erroneous detected positively selected genes. Although this was not assessed in the present study, recent robust evaluations estimated a range of 0 to 5% of false positives on this test (84). Nevertheless, this is more problematic when performing interspecies analyses since it involves highly divergent sequences (5), which is not the case of the intraspecies analysis performed here. Furthermore, to prevent violations of model assumptions, we applied very conservative criteria to calculate P values in the LRT by using χ^2_1 as the null distribution (5, 88). Recombination is another critical factor that may bias the estimation of positive selection. Since previous data based on full-genome sequences (39, 41, 49) detected recombination for some of the genes enrolled in the present study, some specific biological hypothesis could not be subjected to the branch site test of positive selection. These specific exclusions are indicated in Table 1. The

remaining biological hypotheses were tested since recombination is not observed in the corresponding branches. For example, for the CT147 tree, where some ocular strains are shown within a genital branch (hampering the analysis of the hypotheses H1, H2, H4, and H5), the hypotheses H3 and H6 could still be validated. In another scenario, when recombination is known to occur in genes presenting strong congruent trees (as for CT870/*pmpF*), all hypothesis were evaluated. For these specific cases (indicated in Fig. 6), the results should be eyed with caution.

DISCUSSION

A well-known metaphor in evolutionary biology is the adaptive landscape represented by a two-dimensional plot of all genotypes in a specific environment, with their fitness represented by the height of the landscape. For each new environment, in order to climb the fitness peak, bacteria will have to acquire new beneficial mutations, which will likely be differentially spread among different genotypes (80). Presumably, the radiation of *C. trachomatis* species into strains with different cell appetite may be explained by this scenario. Indeed, the different environments are represented by the dissimilar human tissues that strains preferentially infect (ocular, genital tract and lymph nodes), which present heterogeneous properties in terms of competing flora, immune response, and physiological characteristics (such as pH and hormonal concentration). Also, strains present dissimilar levels of fitness: serovars E and F together account for >40% of all genital infections worldwide (64), and serovars A and L2 clearly predominate in ocular (3) and lymphogranuloma (79) infections, respectively. On the other hand, *C. trachomatis* is an obligate intracellular bacterium with a low doubling time (9) and population size (34) and is thus subjected to transmission bottlenecks which make this pathogen a target for the accumulation of deleterious mutations on behalf of the Muller's ratchet theory. The validity of the Muller's ratchet has been evaluated either in RNA viruses (18), which present high mutation rates, are subjected to recurrent bottlenecks, and whose rate of compensatory back-mutations is low, or even in large free-living bacteria such as *Salmonella enterica* serovar Typhimurium, where these contributing factors are clearly attenuated (2). In the present study when the values of *dN* and *dS* are compared independently, it is noticeable that, after the *C. muridarum*/*C. trachomatis* separation from a common ancestor, the values of *dS* show a 147-fold decrease, whereas the values of *dN* only decreased 27-fold (Fig. 2), which suggests the existence of an accumulation of deleterious mutations due to genetic bottleneck, as postulated by Muller. We speculate that, in addition to this scenario, some nonsilent changes may reflect adaptive mutations to different niches rather than deleterious mutations specific of same-niche infecting strains. Thus, the detection of positive selection events acting on particular genomic regions may help to elucidate genotype-phenotype relationships. However, there are few cases where genotype and phenotype are unequivocally linked in *C. trachomatis*, because no straightforward tools to genetically manipulate this pathogen are available thus far. One of the few examples is illustrated by the mutational pattern of the *trpBA* operon (this operon encodes tryptophan synthase, which uses indole as a substrate), where genital strains possess an intact and active operon, whereas it is truncated by point mutations or small indels in strains infecting the ocular conjunctiva (where indole is rare) (13). Thus, mutations that are beneficial in one genetic background are not necessarily beneficial in another background. In

our study, a similar scenario may stand for CT105, which is a pseudogene for the ocular strains, whereas our results suggest that it may be involved in the strains' appetite to the genital epithelium (Fig. 6 and Table 1).

Our results showed that nonsilent changes differentiate strains with different cell appetite or pathogenesis (Fig. 5) and involve genes whose functions may underlie distinct phenotypes. In fact, among the 25 genes identified as positively selected along specific lineages (Table 1), we found genes encoding proteins implicated in immune response elicitation (such as CT147, CT442/*crpA*, CT529, CT694, and *pmp's*) (21, 30, 31, 69, 72, 78), proteolytic activity (such as CT867 and CT868) (57), and subversion of host-cell functions (such as CT223 and CT456/*tarp*) (25, 44). Some of these genes were also identified in a previous study (49), but no genotype-phenotype associations could be established because only six serovars were evaluated, in contrast to the present study, which constitutes a considerable scale-up in terms of genetic variability (enrolling all major 15 serovars represented by 59 strains). A detailed view of the positively selected *loci* that we have detected revealed 11 genes (CT050, CT051, CT115/*incD*, CT147, CT223, CT413/*pmpB*, CT456/*tarp*, CT868, CT870/*pmpF*, CT871/*pmpG*, and CT872/*pmpH*), supporting two or more biological hypotheses for adaptive changes (Fig. 6). Although this seems intriguing in terms of evolutionary directionality, experimental evidences suggest multiple functions for some of them (Table 1) or a distinct *modus operandi* for a specific function. The most striking example is illustrated by CT456, which codes for Tarp. Strong experimental evidence showed that this T3SS effector is associated with the recruitment of host-cell actin observed at early stages of invasion, involving a C-terminal actin-binding domain (ABD) and a proline-rich region (43). Whereas the invasive serovar L2 contains a single functional ABD and it is believed that the proline-rich domain also plays a role in actin nucleation, Tarp from strains with different cell appetite contain multiple ABD sites that are able to nucleate actin without the need of the respective proline-rich domain (44). These data suggest that strains may use Tarp distinctly for actin nucleation. Also, Tarp harbors an N-terminal tyrosine-rich repeat domain (the number of repeats are serovar dependent) that is tyrosine phosphorylated by host cell kinases (42). Curiously, some positively selected sites found to be associated with the infection of mononuclear phagocytes (Table 1) are located precisely within the tyrosine-rich repeat domain. Moreover, there seems to be a pattern of amino acid substitution, where positive selection is operating on exactly the same amino acid positions within the repeated regions, involving always the exchange between aspartic acid (D) and glycine (G) for seven of the positively selected sites (Table 1). In support of recently published data by Mehlitz et al. (56), it can be speculated that, as for the Tarp C-terminal region, a dissimilar *modus operandi* of the Tarp N-terminal region may underlie the distinct phenotype properties of *C. trachomatis* strains. Also of relevance is CT868, a deubiquitinating and deneddylating enzyme that likely interferes in multiple cellular processes. Indeed, a recent study demonstrated that CT868 can inhibit the host inflammatory responses by blocking the nuclear factor- κ B pathway, a known mechanism by which pathogenic microorganisms evade the host immune responses (54). In light of these data and considering the privileged representation of these 11 *loci* in Fig. 6, they should be considered as crucial research targets to improve our knowledge of the pathobiology of *C. trachomatis*.

It is also noteworthy that, among the nine-member Pmp paralogue family, six genes seem to be under positive selection for specific phenotypes. Cumulative evidences indicate that Pmps may function as fine-tune determinants of *C. trachomatis* pathobiology, either by antigenic variation (15, 21, 63, 78) or by host-cell adhesion (22, 36, 38), hypothetically through a shutoff mechanism at the inclusion level (77). It is posited that the accelerated evolution between paralogues is common and constitutes a mechanism for the generation of new genes and new biochemical functions (46).

Darwinian evolution generally relies on the existence of an adaptive pathway in which intermediate steps provide a gradual improvement of fitness. Thus, adaptive changes should not completely rule out synonymous mutations. In fact, the latter may alter the immediate protein adaptive landscape (by changing the proximal amino acids), providing the protein with new opportunities to evolve (14). Also, synonymous mutations can change RNA secondary structure and influence its stability (17), as well as create codons with different frequency usage (associated with tRNA abundance), which was already shown to affect the translation efficiency in several microorganisms (50, 70). We have previously shown synonymous changes to more favorable codons for the *C. trachomatis* major antigen (64), and it is reasonable to expect that several other loci present synonymous changes that, in a camouflaged way, become adaptive. Our results support a non-random accumulation of synonymous mutations in *C. trachomatis*. In fact, we found that strains infecting the same human cell type are clearly the most closely related through *dS* analysis (Fig. 5), which suggests a nonstochastic fixation of synonymous mutations.

In conclusion, our results support a directional evolution of *C. trachomatis* toward niche-specific adaptation in addition to a background of Muller's ratchet deleterious mutations. Whereas the molecular basis for organ/cell appetite is likely complex, these data suggest that population genetics and evolutionary inferences may be crucial to a comprehensive understanding of the resulting phenotypes, by guiding subsequent experimental procedures to specific targets.

ACKNOWLEDGMENTS

This study was supported by grant PTDC/SAU-MII/099623/2008 from the Fundação para a Ciência e a Tecnologia (FCT). V.B. and R.F. are recipients of Ph.D. fellowships (SFRH/BD/68527/2010 and SFRH/BD/68532/2010, respectively) from the FCT. A.N. is a recipient of a postdoctoral fellowship (SFRH/BPD/75295/2010) from the FCT.

REFERENCES

- Alzhanov DT, Weeks SK, Burnett JR, Rockey DD. 2009. Cytokinesis is blocked in mammalian cells transfected with *Chlamydia trachomatis* gene CT223. *BMC Microbiol.* 9:2. doi:10.1186/1471-2180-9-2.
- 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.
- Andreasen AA, et al. 2008. *Chlamydia trachomatis ompA* variants in trachoma: what do they tell us? *PLoS Negl. Trop. Dis.* 2:e306. doi:10.1371/journal.pntd.0000306.
- Anisimova M, Bielawski J, Dunn K, Yang Z. 2007. Phylogenomic analysis of natural selection pressure in *Streptococcus* genomes. *BMC Evol. Biol.* 7:154. doi:10.1186/1471-2148-7-154.
- Anisimova M, Yang Z. 2007. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Mol. Biol. Evol.* 24:1219–1228.
- Bannantine JP, Griffiths RS, Viratyosin W, Brown WJ, Rockey DD. 2000. A secondary structure motif predictive of protein localization to the chlamydial inclusion membrane. *Cell Microbiol.* 2:35–47.
- Bannantine JP, Rockey DD, Hackstadt T. 1998. Tandem genes of *Chlamydia psittaci* that encode proteins localized to the inclusion membrane. *Mol. Microbiol.* 28:1017–1026.
- Belland RJ, et al. 2003. Genomic transcriptional profiling of the developmental cycle of *Chlamydia trachomatis*. *Proc. Natl. Acad. Sci. U. S. A.* 100:8478–8483.
- Borges V, et al. 2010. Normalization strategies for real-time expression data in *Chlamydia trachomatis*. *J. Microbiol. Methods* 82:256–264.
- Bullock HD, Hower S, Fields KA. 2012. Domain analyses reveal that *Chlamydia trachomatis* CT694 protein belongs to the membrane-localized family of type III effector proteins. *J. Biol. Chem.* 287:28078–28086.
- Bush RM. 2001. Predicting adaptive evolution. *Nat. Rev. Genet.* 2:387–392.
- Bush RM, Bender CA, Subbarao K, Cox NJ, Fitch WM. 1999. Predicting the evolution of human influenza A. *Science* 286:1921–1925.
- Caldwell HD, et al. 2003. Polymorphisms in *Chlamydia trachomatis* tryptophan synthase genes differentiate between genital and ocular isolates. *J. Clin. Invest.* 111:1757–1769.
- Cambray G, Mazel D. 2008. Synonymous genes explore different evolutionary landscapes. *PLoS Genet.* 4:e1000256. doi:10.1371/journal.pgen.1000256.
- Carlson JH, Porcella SF, McClarty G, Caldwell HD. 2005. Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains. *Infect. Immun.* 73:6407–6418.
- Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25:119–120.
- Chamary JV, Hurst LD. 2005. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol.* 6:R75.
- Chao L. 1990. Fitness of RNA virus decreased by Muller's ratchet. *Nature* 348:454–455.
- Chen L, Perlina A, Lee CJ. 2004. Positive selection detection in 40,000 human immunodeficiency virus (HIV) type 1 sequences automatically identifies drug resistance and positive fitness mutations in HIV protease and reverse transcriptase. *J. Virol.* 78:3722–3732.
- Clifton DR, et al. 2004. A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin. *Proc. Natl. Acad. Sci. U. S. A.* 101:10166–10171.
- Coler RN, et al. 2009. Identification and characterization of novel recombinant vaccine antigens for immunization against genital *Chlamydia trachomatis*. *FEMS Immunol. Med. Microbiol.* 55:258–270.
- Crane DD, et al. 2006. *Chlamydia trachomatis* polymorphic membrane protein D is a species-common pan-neutralizing antigen. *Proc. Natl. Acad. Sci. U. S. A.* 103:1894–1899.
- Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi:10.1371/journal.pone.0011147.
- Dehoux P, Flores R, Dauga C, Zhong G, Subtil A. 2011. Multi-genome identification and characterization of chlamydia-specific type III secretion substrates: the Inc proteins. *BMC Genomics* 12:109. doi:10.1186/1471-2164-12-109.
- Delevoe C, et al. 2008. SNARE protein mimicry by an intracellular bacterium. *PLoS Pathog.* 4:e1000022. doi:10.1371/journal.ppat.1000022.
- Egea R, Casillas S, Barbadilla A. 2008. Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Res.* 36(web server issue)W157–62157–62162.
- Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* 78:737–756.
- Ferreira R, et al. 2012. Impact of loci nature on estimating recombination and mutation rates in *Chlamydia trachomatis*. *G3 (Bethesda)* 2:761–768.
- Fields KA, Mead DJ, Dooley CA, Hackstadt T. 2003. *Chlamydia trachomatis* type III secretion: evidence for a functional apparatus during early-cycle development. *Mol. Microbiol.* 48:671–683.
- Finco O, et al. 2011. Approach to discover T- and B-cell antigens of intracellular pathogens applied to the design of *Chlamydia trachomatis* vaccines. *Proc. Natl. Acad. Sci. U. S. A.* 108:9969–9974.
- Fling SP, et al. 2001. CD8⁺ T cells recognize an inclusion membrane-associated protein from the vacuolar pathogen *Chlamydia trachomatis*. *Proc. Natl. Acad. Sci. U. S. A.* 98:1160–1165.
- Fraser C, Hanage WP, Spratt BG. 2005. Neutral microepidemic evolution of bacterial pathogens. *Proc. Natl. Acad. Sci. U. S. A.* 102:1968–1973.

33. Gaschen B, et al. 2002. Diversity considerations in HIV-1 vaccine selection. *Science* 296:2354–2360.
34. Gomes JP, et al. 2006. Correlating *Chlamydia trachomatis* infectious load with urogenital ecological success and disease pathogenesis. *Microbes Infect.* 8:16–26.
35. Gomes JP, Hsia RC, Mead S, Borrego MJ, Dean D. 2005. Immunoreactivity and differential developmental expression of known and putative *Chlamydia trachomatis* membrane proteins for biologically variant serovars representing distinct disease groups. *Microbes Infect.* 7:410–420.
36. Gomes JP, et al. 2006. Polymorphisms in the nine polymorphic membrane proteins of *Chlamydia trachomatis* across all serovars: evidence for serovar Da recombination and correlation with tissue tropism. *J. Bacteriol.* 188:275–286.
37. Gong S, Lei L, Chang X, Belland R, Zhong G. 2011. *Chlamydia trachomatis* secretion of hypothetical protein CT622 into host cell cytoplasm via a secretion pathway that can be inhibited by the type III secretion system inhibitor compound 1. *Microbiology* 157:1134–1144.
38. Grimwood J, Stephens RS. 1999. Computational analysis of the polymorphic membrane protein superfamily of *Chlamydia trachomatis* and *Chlamydia pneumoniae*. *Microb. Comp. Genomics* 4:187–201.
39. Harris SR, et al. 2012. Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. *Nat. Genet.* 44:413–419.
40. Hower S, Wolf K, Fields KA. 2009. Evidence that CT694 is a novel *Chlamydia trachomatis* T3S substrate capable of functioning during invasion or early cycle development. *Mol. Microbiol.* 72:1423–1437.
41. Jeffrey BM, et al. 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.
42. Jewett TJ, Dooley CA, Mead DJ, Hackstadt T. 2008. *Chlamydia trachomatis* tarp is phosphorylated by src family tyrosine kinases. *Biochem. Biophys. Res. Commun.* 371:339–344.
43. Jewett TJ, Fischer ER, Mead DJ, Hackstadt T. 2006. Chlamydial TARP is a bacterial nucleator of actin. *Proc. Natl. Acad. Sci. U. S. A.* 103:15599–15604.
44. Jewett TJ, Miller NJ, Dooley CA, Hackstadt T. 2010. The conserved Tarp actin binding domain is important for chlamydial invasion. *PLoS Pathog.* 6:e1000997. doi:10.1371/journal.ppat.1000997.
45. Jia TJ, Liu DW, Luo JH, Zhong GM. 2007. Localization of the hypothetical protein CT249 in the *Chlamydia trachomatis* inclusion membrane. *Wei Sheng Wu Xue Bao* 47:645–648. (In Chinese.)
46. Jordan IK, et al. 2001. Constant relative rate of protein evolution and detection of functional diversification among bacterial, archaeal, and eukaryotic proteins. *Genome Biol.* 2:research0053.
47. Jordan IK, Rogozin IB, Wolf YI, Koonin EV. 2002. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res.* 12:962–968.
48. Jorgensen I, Valdivia RH. 2008. Pmp-like proteins Pls1 and Pls2 are secreted into the lumen of the *Chlamydia trachomatis* inclusion. *Infect. Immun.* 76:3940–3950.
49. 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.
50. Kanaya S, Yamada Y, Kudo Y, Ikemura T. 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* 238:143–155.
51. Kari L, et al. 2008. Pathogenic diversity among *Chlamydia trachomatis* ocular strains in nonhuman primates is affected by subtle genomic variations. *J. Infect. Dis.* 197:449–456.
52. Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111–120.
53. Kimura M. 1983. The neutral theory of molecular evolution. Cambridge University Press, London, England.
54. Le Negrate G, et al. 2008. ChlADub1 of *Chlamydia trachomatis* suppresses NF- κ B activation and inhibits I κ B α ubiquitination and degradation. *Cell Microbiol.* 10:1879–1892.
55. McDonald JH, Kreitman M. 1991. Adaptive evolution at the *adh* locus in *Drosophila*. *Nature* 351:652–654.
56. Mehlitz A, et al. 2010. Tarp regulates early *Chlamydia*-induced host cell survival through interactions with the human adaptor protein SHC1. *J. Cell Biol.* 190:143–157.
57. Misaghi S, et al. 2006. *Chlamydia trachomatis*-derived deubiquitinating enzymes in mammalian cells during infection. *Mol. Microbiol.* 61:142–150.
58. Mital J, Miller NJ, Fischer ER, Hackstadt T. 2010. Specific chlamydial inclusion membrane proteins associate with active Src family kinases in microdomains that interact with the host microtubule network. *Cell Microbiol.* 12:1235–1249.
59. Mitchell-Olds T, Willis JH, Goldstein DB. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nat. Rev. Gen.* 8:845–856.
60. Muller HJ. 1964. The relation of recombination to mutational advance. *Mutat. Res.* 106:2–9.
61. Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York, NY.
62. Nielsen R. 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity* 86:641–647.
63. Nunes A, et al. 2007. Comparative expression profiling of the *Chlamydia trachomatis* *pmp* gene family for clinical and reference strains. *PLoS One* 2:e878. doi:10.1371/journal.pone.0000878.
64. Nunes A, Nogueira PJ, Borrego MJ, Gomes JP. 2010. Adaptive evolution of the *Chlamydia trachomatis* dominant antigen reveals distinct evolutionary scenarios for B- and T-cell epitopes: worldwide survey. *PLoS One* 5:e13171. doi:10.1371/journal.pone.0013171.
65. Read TD, et al. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* 28:1397–1406.
66. Rockey DD, Scidmore MA, Bannantine JP, Brown WJ. 2002. Proteins in the chlamydial inclusion membrane. *Microbes Infect.* 4:333–340.
67. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–425.
68. Scidmore-Carlson MA, Shaw EI, Dooley CA, Fischer ER, Hackstadt T. 1999. Identification and characterization of a *Chlamydia trachomatis* early operon encoding four novel inclusion membrane proteins. *Mol. Microbiol.* 33:753–765.
69. Sharma J, et al. 2006. Profiling of human antibody responses to *Chlamydia trachomatis* urogenital tract infection using microplates arrayed with 156 chlamydial fusion proteins. *Infect. Immun.* 74:1490–1499.
70. Sharp PM, Tuohy TM, Mosurski KR. 1986. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 14:5125–5143.
71. Sisko JL, Spaeth K, Kumar Y, Valdivia RH. 2006. Multifunctional analysis of *Chlamydia*-specific genes in a yeast expression system. *Mol. Microbiol.* 60:51–66.
72. Starnbach MN, et al. 2003. An inclusion membrane protein from *Chlamydia trachomatis* enters the MHC class I pathway and stimulates a CD8⁺ T cell response. *J. Immunol.* 171:4742–4749.
73. Stephens RS. 2002. Chlamydiae and evolution: a billion years and counting, p 3–12. In Schachter J, et al. (ed), *Chlamydial infections: proceedings of the 10th International Symposium on Human Chlamydial Infections*. International Chlamydia Symposium, San Francisco, CA.
74. Stephens RS, et al. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* 282:754–759.
75. Subtil A, et al. 2005. A directed screen for chlamydial proteins secreted by a type III mechanism identifies a translocated protein and numerous other new candidates. *Mol. Microbiol.* 56:1636–1647.
76. Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum-parsimony methods. *Mol. Biol. Evol.* 28:2731–2739.
77. Tan C, et al. 2010. Variable expression of surface-exposed polymorphic membrane proteins in *in vitro*-grown *Chlamydia trachomatis*. *Cell Microbiol.* 12:174–187.
78. Tan C, et al. 2009. *Chlamydia trachomatis*-infected patients display variable antibody profiles against the nine-member polymorphic membrane protein family. *Infect. Immun.* 77:3218–3226.
79. Thomson NR, et al. 2008. *Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. *Genome Res.* 18:161–171.
80. Vos M. 2009. Why do bacteria engage in homologous recombination? *Trends Microbiol.* 17:226–232.
81. Wang J, et al. 2009. A chlamydial type III-secreted effector protein (Tarp) is predominantly recognized by antibodies from humans infected with *Chlamydia trachomatis* and induces protective immunity against upper genital tract pathologies in mice. *Vaccine* 27:2967–2980.

82. Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* 15:568–573.
83. Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586–1591.
84. Yang Z, dos Reis M. 2011. Statistical properties of the branch-site test of positive selection. *Mol. Biol. Evol.* 28:1217–1228.
85. Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19:908–917.
86. Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17:32–43.
87. Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22:1107–1118.
88. Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22:2472–2479.