Molecular features underlying the higher ecological success of C. trachomatis E and F genotypes

Alexandra Nunes[1*], Rita Ferreira[1], Vítor Borges[1], Maria J. Borrego[1], and João P. Gomes[1**]

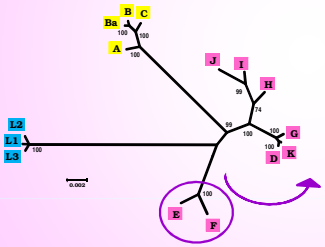
[1] Department of Infectious Diseases, National Institute of Health, Av. Padre Cruz, 1649-016 Lisbon, Portugal
*E-mail: alexandra.nunes@insa.min-saude.pt;
** Email: j.paulo.gomes@insa.min-saude.pt

Background

In the light of the >98% genomic similarity among the fully-sequenced *C. trachomatis* strains, the higher worldwide ecological success of E and F serovars is enigmatic. Cumulative data have been providing some clues about the secret underlying serovar's ecological success. We intend to provide a quick overview of the molecular aspects that distinguish E and F from the remaining serovars.

¤ Evolutionary history of C. trachomatis genotypes

Based on a high-scale concatenation-based phylogenomic study [1], using ~33% of all chromosome SNPs, E and F exhibit an independent evolutionary co-segregation, for which the polymorphism of some membrane proteins, housekeeping genes, and regulatory regions may be important for promoting: *i*) the formation of exclusive host-interacting regions (as already reported for some Pmps [2] and OmcB [3]); and *ii*) specificities on metabolic pathways (such as temporal protein synthesis, bacterial replication or energy metabolism – see table). These may confer E and F strains some functional and/or structural advantage in terms of infection and transmission.



IGRs	HKs	HPs	CEPs
CT144/CT145	*araD*	CT683	*incD*
rpoB/rl7	*gapA*		*incE*
pmpB/pmpC	*karG*		*pmpB*
rs2/ompA	*tsf*		*pmpC*
ompA/pbpB	*rs2*		*omcB*
CT683/CT684	CT686		*pbpB*
yfh0_fparB	*yfh0_1*		*poB*
thdF/pgsD	*parB*		*pmpD*
glyQ/pgsA	*dppF*		*pmpH*
	dppD		*pmpI*
	tiSS/RNA		

Figure 1 – Phylogeny based on ~50,000 bp/taxa. The chart shows the loci contributing to E/F segregation. IGRs=intergenomic regions; HKs=housekeeping genes; HPs=hypothetical protein genes; CEPs=cell envelope protein genes. **Figure taken from ref [1].**

¤ Analysis of specific intra-loci domains

Cumulative evidences [1,2] revealed the existence of E/F-specific mutational patterns for some loci, where mutations exclusive of E and F are clustered in specific domains with divergences of up to 45% to the remaining genital serovars, which may yield unique E/F conformational motifs for interaction with the host.

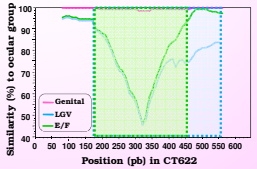


Figure 2 – Example of a SimPlot graph showing the nucleotide similarity between ocular, E/F, remaining genital, and LGV serovars. Intra-loci domains that are specific of a particular group of serovars are bordered by colored boxes. **Figure taken from ref [1].**

¤ Evaluation of recombination on C. trachomatis population

Data from an ongoing study, using a sampling of multiple recent isolates that reflects the worldwide distribution of each genotype, seem to evidence a clonal genomic structure for E and F strains, where a predominant favorable clone may be strongly maintained *in vivo*. Preliminary data show that the likelihood of E and F strains to undergo recombination is about 12-fold lower than that of the other genotypes ($P < 10^{-2}$).

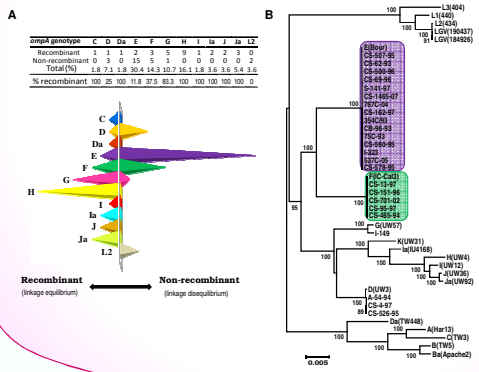


Figure 3 – (A) Impact of recombination on recent *C. trachomatis* isolates (n=56). To detect mosaic structures, 14 loci involving two statistically-confirmed recombination hotspots [4] and representing five well-separated regions of the *C. trachomatis* chromosome were used. Bar lengths are proportional to the absolute number of recombinant and non-recombinant strains from each genotype. (B) Phylogenetic tree (concatenation of 14 loci) based on ~15,000 bp/taxa, showing that all non-recombinant E and F isolates are genetically identical to the respective prototype strain.

¤ Worldwide analysis of ompA variability

Based on data from a worldwide survey [5], MOMP of E and F strain which together represent 42.3% of all analyzed specimens (>500 exhibit the lowest mutation rate (22.3-fold lower than that of the oth genotypes, $P < 10^{-20}$).

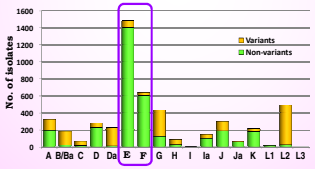


Figure 4 – Distribution *p* genotype of all MOM sequences reported worldwid Overall, the data pertain to total of 5,026 strains from distinct geographic regions five continents. **Figure taken from ref [5].**

¤ Evaluation of chlamydial infectious load

A previous quantitative study using >170 urine samples [6], revea similar infectious load among all genital strains, suggesting that, up entry, E and F strains do not seem to present a higher multiplica rate *in vivo*. Thus, the higher ecological success of E and F may defined at the adhesion/entry stage.

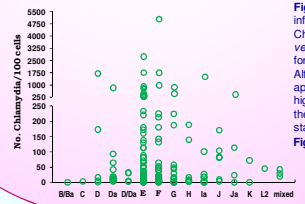


Figure 5 – *C. trachomatis* infectious load (No. of Chlamydia/100 cells) *versus ompA* genotype, for all urine samples. Although E and F strains apparently present the highest infectious loads, the distribution was not statistically significant. **Figure taken from ref [6].**

Conclusions

Full genomic data from multiple and diverse recent isolates will be essential to decipher the secret behind the higher ecological success of E and F strains. However, this overview suggests that a noticeable lack of chromosomal mosaicism together with a strikingly low mutational rate of the dominant antigen, the existence of exclusive host-interacting regions and specificities on metabolic pathways may be critical factors. Their apparent unique genomic make-up suggests the emergence of successful clones well-adapted to face the 'arms race' with the host.

References

[1] Nunes A, Nogueira PJ, Borrego MJ, and Gomes JP. (2008) *Genome Biol* 9:R153.
[2] Gomes JP, Nunes A, Bruno WJ, Borrego MJ, Florindo C, and Dean D. (2006) *J Bacteriol* 188:275-86.
[3] Moelleken K, and Hegemann JH. (2008) *Mol Microbiol* 67:403-19.
[4] Gomes JP, Bruno WJ, Nunes A, Santos N, Florindo C, Borrego MJ, and Dean D. (2007) *Genome Res* 17:50-6
[5] Nunes A, Nogueira PJ, Borrego MJ, and Gomes JP. (2010) *PLoS ONE* 5:e13171.
[6] Gomes JP, Borrego MJ, Atik B, Santo I, Azevedo J, Brito de Sá A, Nogueira P, and Dean D. (2006) *Microbes In*