scientific reports

Check for updates

OPEN

Identifying HLA *DRB1-DQB1* alleles associated with *Chlamydia trachomatis* infection and in silico prediction of potentially-related peptides

Leidy Pedraza[1,2,6], Milena Camargo[1,3,6], Darwin A. Moreno-Pérez[1,3,6], Ricardo Sánchez[1,4], Luisa Del Río-Ospina[1], Indira M. Báez-Murcia[1], Manuel E. Patarroyo[1,4,5] & Manuel A. Patarroyo[1,4,5]✉

HLA class II (HLA-II) genes' polymorphism influences the immune response to *Chlamydia trachomatis* (*Ct*), it is considered a sexually transmitted infection. However, associations between HLA-II alleles and *Ct*-infection have been little explored in humans; this study was thus aimed at determining HLA-*DRB1-DQB1* alleles/haplotypes' effect on *Ct*-infection outcome in a cohort of Colombian women. Cervical sample DNA was used as template for detecting *Ct* by PCR and typing HLA-*DRB1-DQB1* alleles/haplotypes by Illumina MiSeq sequencing. Survival models were adjusted for identifying the alleles/haplotypes' effect on *Ct*-outcome; bioinformatics tools were used for predicting secreted bacterial protein T- and B-cell epitopes. Sixteen HLA-*DRB1* alleles having a significant effect on *Ct*-outcome were identified in the 262 women analysed. *DRB108:02:01G and *DRB1**12:01:01G were related to infection-promoting events. Only the *DQB1**05:03:01G allele related to clearance/persistence events was found for HLA-*DQB1*. HLA-*DRB1* allele homozygous women were associated with events having a lower probability of clearance and/or early occurrence of persistence. Twenty-seven peptides predicted in silico were associated with protective immunity against *Ct*; outer membrane and polymorphic membrane protein-derived peptides had regions having dual potential for being T- or B-cell epitopes. This article describes HLA-*DRB1-DQB1* alleles/haplotypes related to *Ct*-infection resolution and the peptides predicted in silico which might probably be involved in host immune response. The data provides base information for developing future studies leading to the development of effective prevention measures against *Ct*-infection.**

Chlamydia trachomatis (*Ct*) is the commonest bacteria-related, sexually-transmitted infection (STI) worldwide[1]; the WHO's Report on global sexually transmitted infection surveillance estimated that there are 127 million new cases annually[2]. Most *Ct* infections have an asymptomatic clinical course, but some might lead to severe complications, such as pelvic inflammatory disease (PID) and recurrent abortions[1].

Ct mainly affects 18–20 year-old women and 20–24 year-old men, asymptomatic cases mainly occurring in women (close to 90%)[3]; however, symptomatic infections usually occur weeks or months after exposure. Abundant secretion, dysuria and postcoital bleeding are amongst the commonest symptoms[3].

Previous studies have revealed that late detection of *Ct* asymptomatic infection (together with other agents such HPV) could lead to conditions such as squamosas cell carcinoma and cervical cancer (CC)[4]. It has been

[1]Molecular Biology and Immunology Department, Fundación Instituto de Inmunología de Colombia (FIDIC), 111321 Bogotá D.C., Colombia. [2]MSc Programme in Microbiology, Universidad Nacional de Colombia, 111321 Bogotá D.C., Colombia. [3]Animal Science Faculty, Universidad de Ciencias Aplicadas y Ambientales (U.D.C.A), 111166 Bogotá D.C., Colombia. [4]Faculty of Medicine, Universidad Nacional de Colombia, 111321 Bogotá D.C., Colombia. [5]Health Sciences Division, Main Campus, Universidad Santo Tomás, 110231 Bogotá D.C., Colombia. [6]These authors contributed equally: Leidy Pedraza, Milena Camargo and Darwin A. Moreno-Pérez. ✉email: mapatarr.fidic@gmail.com

Variable	Years	
Mean age (SD)	41.7 (23.1)	
Median age at first intercourse (IQR)	18.0 (4.0)	
Years of active sex life (SD)	23.1 (10.6)	
	n	(%)
Ethnicity—not mestizo[a]	7	(2.6)
More than three lifetime sexual partners	55	(21.8)
Marital status—married/cohabiting	231	(88.1)
Having had more than two pregnancies	37	(14.5)
A history of abortions	93	(35.5)
A history of sexually-transmitted diseases	60	(22.9)
Contraceptive method—hormonal	29	(11.9)
Educational level—illiterate/primary	122	(47.8)
Living outside Bogotá	180	(68.7)
Average monthly income—less than minimum wage[b]	232	(89.6)
HPV positive	218	(83.2)
Cytological findings—abnormal	23	(8.9)

Table 1. Base line demographic characteristics and risk factors for the 262 women included in the study. [a]This category included Indigenous and Afro Colombian ethnicities. [b]The Colombian minimum average monthly income would be roughly US$ 250.

shown that a fourth of *Ct* infections become spontaneously resolved and without treatment; factors, such as host immune response, thus significantly contribute towards eliminating genital tract infection[5].

The major histocompatibility complex (MHC) is a mechanism greatly determining infections' clinical course; establishing its role in *Ct* infection dynamics would thus explain their outcome[5]. Some studies have suggested the relationship between class II HLA alleles, *Ct* and tubal factor infertility (TFI)[6] as DQA**03:01*, DQA**05:01* and DQB**04:02* have been described as being associated with *Ct* and DRB1**75:03* and DRB5**01:01* with TFI[6].

An efficient immune response against intracellular pathogens like *Ct* requires cell-mediated immunity, this being stimulated by bacterial peptides presented to T-cells by the MHC[7]. Predicting these molecules and establishing their potential role in the natural history of *Ct* infection could contribute towards understanding its dynamics[8].

Studies have searched for prophylactic interventions; most have evaluated the immune response in mice experimentally infected by *Ct* or *C. muridarum*[9,10]. Murine model studies have demonstrated the importance of an HLA class II molecule-mediated immune response regarding bacterial infection resolution[11]. However, few studies have comprehensively evaluated HLA alleles' role regarding human *Ct* infection outcome and the peptides probably related to immune response[10,12]. This study was thus aimed at determining HLA-*DRB1-DQB1* alleles/haplotypes' effect on *Ct* infection, persistence, clearance and redetection in a cohort of Colombian women. The results provided information for developing suitable prevention measures for managing and controlling *Ct* infection.

Results

HLA-*DRB1-DQB1* allele effects on *Ct* outcome. The demographic characteristics of 262 women complying with the retrospective study's inclusion criteria were analysed; mean age was 41.7 years-old (23.1 SD) and median age at onset of sexual life 18.0 years-old (4.0 IQR) (Table 1). *Ct* was detected retrospectively; study results gave the highest rate for persistence events (26.0 per 100 women/month), followed by clearance (16.1 per 100 women/month) and redetection events (15.6 per 100 women/month) (Supplementary Fig. S1). Survival data was estimated (using Kaplan Meier survival functions) for each event; Fig. 1 shows the probability of Ct infection, clearance, persistence and redetection risk throughout the follow-up period.

Forty-seven HLA-*DRB1* alleles were found using Illumina MiSeq sequencer whose frequency distribution has been previously published[13]. The *DRB1**04:07:01G allele had the greatest prevalence in the target population (Supplementary Table S1); this allele has been categorised as a common allele and has been reported in all continents' populations[14]. Multivariate models (parametric and semiparametric) were constructed for identifying data associated with *Ct* outcomes; they were adjusted to fit the covariables associated with outcomes in univariate analysis (Supplementary Table S2).

Sixteen *DRB1* alleles had a statistically significant *p* value, indicating their possible effect regarding a particular infection event (greater probability (GP), earlier occurrence (EO), later occurrence (LO) and lower probability (LP) (Table 2 and Supplementary Tables S3 and S4). Interestingly, three alleles were related to more than one event; *DRB1**08:02:01G and *DRB1**12:01:01G had concordant effects (GP for persistence and EO for redetection and LO for infection and GP for clearance) while *DRB1**14:02:01G had an opposite effect (LO for infection and GP for persistence) (Table 2). Fourteen HLA-*DQB1* alleles were identified in the target population (Supplementary Table S1), of which only *DQB1**05:03:01G had a statistically significant *p* value related to clearance and persistence events, having a GP effect for both (Table 2 and Supplementary Tables S5 and S6). It was also

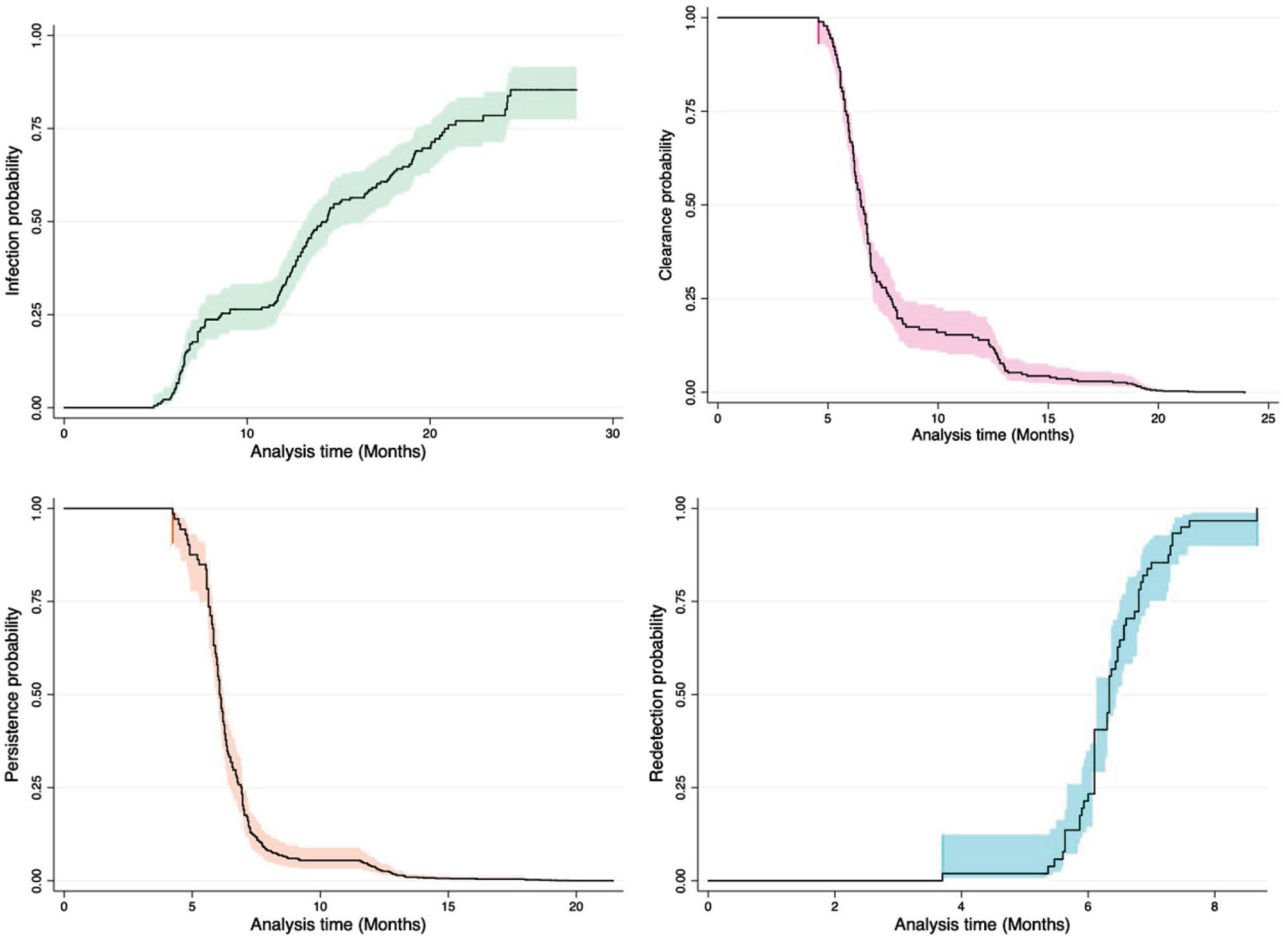

Figure 1. Kaplan–Meier estimator for evaluating Ct infection, clearance, persistence and redetection.

found that HLA-*DRB1* homozygous women had a LP effect on clearance and an EO effect related to persistence events, suggesting that this characteristic could represent a genetic disadvantage for its carriers as it makes them more susceptible to *Ct* infection. However, more experimental evidence is required to confirm such hypothesis.

HLA-*DRB1-DQB1* haplotype effect on *Ct* outcome. Forty-seven of the 142 *DRB1-DQB1* haplotypes had statistically significant values when analysing the effect of each haplotype on *Ct* outcome (Table 3). Twenty-seven associations were found regarding infection events; 16 related to the event (GP/EO) (e.g. *DRB1**01:02:01G-*DQB1**03:03:02G) whilst 11 associations did not (LO) (e.g. *DRB1**12:01:01G-*DQB1**03:02:01G). *DRB1**09:01:02G-*DQB1**03:01:01G and *DRB1**12:01:01G-*DQB1**03:02:01G had an EO effect on clearance whilst *DRB1**04:05:01-*DQB1**03:01:01G was associated with LP (Table 3 and Supplementary Table S7).

Eighteen associations were found regarding persistence; sixteen of them related to the event (GP/EO), *DRB1**04:05:01-*DQB1**03:01:01G and *DRB1**11:01:01G-*DQB1**03:02:01G having the greatest effect. By contrast, *DRB1**10:01:01G-*DQB1**05:01:01G and *DRB1**14:02:01G-*DQB1**05:01:01G reduced persistence (LP/LO). Eight haplotypes were related to redetection events (Table 3 and Supplementary Table S8).

Nine haplotypes were related to more than one event when analysed; 6 were related to some infection events, such as EO for DRB1*04:05:01-DQB1*02:01:01G on infection and redetection, LO for DRB1*12:01:01G-DQB1*03:02:01G on infection and GP on clearance;

whilst others had an opposite effect on events, LO of *DRB1**01:01:01G-*DQB1**03:01:01G infection and EO on persistence (Table 3).

In silico predicted peptides associated with *Ct* events. Peptides derived from proteins predicted as being secreted (Supplementary Table S9) were predicted in silico since it has been shown that they could be associated with protective immunity or susceptibility to *Ct* infection[7,10]. Fifteen out of 24 proteins had peptides having T-cell epitopes binding strongly to HLA-II-*DRB1* but not to -*DQB1* molecules (Supplementary Tables S10 to S24). Some peptides might have been related to increased susceptibility against *Ct* as they were associated with GP and EO effects related to persistence or redetection events. Twenty-seven peptides were related to protection (of which 7 were OMP-derived and 17 from PMP) since they were associated with effects related to *Ct* elimination (LO of infection, GP of clearance and LP of persistence) (Supplementary Tables S10 to S24). Interestingly, 11

Locus	Allele	Event	Coeff.	95%CI	Pc[d]	Effect
DRB1	**01:02:01G**	Infection	**8.36**	**1.71–4.08**	**0.000**	**Greater probability**
	01:03:01	Redetection	**0.01**[c]	**0.00–0.05**	**0.000**	**Earlier occurrence**
	03:02:01	Infection	**− 0.98**[b]	**− 1.29–(− 0.67)**	**0.000**	**Earlier occurrence**
	03:02:02	Infection	**− 0.55**[b]	**− 0.86–(− 0.25)**	**0.000**	**Earlier occurrence**
	04:10:01G	Infection	**2.91**[b]	**2.49–3.36**	**0.000**	**Later occurrence**
	08:01:01G	Infection	**− 0.87**[b]	**− 1.20–(− 0.54)**	**0.000**	**Earlier occurrence**
	08:02:01G	Persistence	**12.71**	**4.35–37.09**	**0.000**	**Greater probability**
		Redetection	**− 0.32**[a]	**− 0.51–(− 0.13)**	**0.030**	**Earlier occurrence**
	09:01:02G	Clearance	**5.53**	**2.91–10.49**	**0.000**	**Greater probability**
	10:01:01G	Persistence	**0.33**	**0.19–0.57**	**0.000**	**Lower probability**
	11:01:01G	Persistence	**22.20**	**5.70–86.49**	**0.000**	**Greater probability**
	11:01:02	Infection	**3.00**[b]	**2.48–3.51**	**0.000**	**Later occurrence**
	12:01:01G	Infection	**3.23**[b]	**2.73–3.73**	**0.000**	**Later occurrence**
		Clearance	**5.53**	**2.91–10.49**	**0.000**	**Greater probability**
	13:05:01	Persistence	**3.67**	**1.74–7.75**	**0.032**	**Greater probability**
	14:01:01G	Persistence	**3.92**	**1.86–8.25**	**0.000**	**Greater probability**
	14:02:01G	Infection	**2.40**[b]	**1.87–3.10**	**0.000**	**Later occurrence**
		Persistence	**27.79**	**5.30–145.59**	**0.000**	**Greater probability**
	16:01:01	Redetection	**− 0.17**[a]	**− 0.27–(− 0.08)**	**0.000**	**Earlier occurrence**
DQB1	**05:03:01G**	Clearance	**4.97**	**1.79–13.74**	**0.024**	**Greater probability**
		Persistence	**3.92**	**1.86–8.25**	**0.000**	**Greater probability**

Table 2. HLA *DRB1-DQB1* alleles associated with *Ct* infection, clearance, persistence and redetection. Alleles affecting an event are shown as green associations [i.e. greater probability (GP) or earlier occurrence (EO)] and those hindering them [lower probability (LP) or later occurrence (LO)] as red associations. *Pc* corrected *p* value, *95%CI* 95% confidence interval, *Coeff* regression coefficient, *DRB1* DR beta 1, *DQB1* DQ beta 1. Values in bold indicate statistical significance based on 95%CI, $p < 0.05$. [a]The Cox proportional hazards model did not fulfil the assumption of proportionality; a logistic parametric model was constructed. [b]The Cox proportional hazards model did not fulfil the assumption of proportionality; a lognormal parametric model was constructed. [c]The Cox proportional hazards model did not fulfil the assumption of proportionality; a Gompertz parametric model was constructed. [d]The Bonferroni method was used for correcting all the models' *p*-values. Analysis was adjusted for age, onset of sexual life, lifetime amount of sexual partners, planning method, abortions, history of other STI and HPV infections.

of them had potential B-cell epitope regions, thus highlighting their possible role as cell and/or humoral immune response mediators (Table 4).

Discussion

Ct is the commonest sexually-transmitted bacterial pathogen worldwide; it can provoke serious consequences regarding reproductive sexual health once it becomes a chronic infection. Despite significant advances having been made regarding its control, clear and effective tools for reducing its impact on public health are still not

Haplotype *DRB1 DQB1*		Event	Coeff.	95%CI	Pc[d]	Effect
01:01:01G	03:01:01G	Infection	2.73[b]	2.21–3.25	0.000	Later occurrence
		Persistence	0.48[a]	0.31–0.66	0.000	Earlier occurrence
01:01:01G	03:02:01G	Persistence	4.76	2.03–11.14	0.000	Greater probability
01:01:01G	03:03:02G	Infection	3.10[b]	2.61–3.59	0.000	Later occurrence
01:02:01G	03:03:02G	Infection	14.82	4.54–48.31	0.000	Greater probability
01:03:01	02:01:01G	Persistence	9.52	2.77–32.71	0.000	Greater probability
		Redetection	0.01[c]	0.00–0.05	0.000	Earlier occurrence
01:03:01	03:01:01G	Infection	− 1.27[b]	− 1.64–(− 0.90)	0.000	Earlier occurrence
01:03:01	03:02:01G	Infection	2.97[b]	2.44–3.49	0.000	Later occurrence
01:03:01	05:01:01G	Persistence	9.52	2.77–32.71	0.000	Greater probability
		Redetection	0.01[c]	0.00–0.05	0.000	Earlier occurrence
03:01:01G	05:01:01G	Infection	14.08	5.36–36.96	0.000	Greater probability
03:02:01	04:02:01G	Infection	− 0.98[b]	− 1.29–(− 0.67)	0.000	Earlier occurrence
03:02:02	03:01:01G	Infection	− 0.55[b]	− 0.86–(− 0.25)	0.000	Earlier occurrence
04:02:01	03:01:01G	Infection	0.83[b]	0.41–1.26	0.000	Earlier occurrence
04:03:01G	02:01:01G	Persistence	5.25	2.10–13.15	0.000	Greater probability
04:03:01G	03:01:01G	Infection	5.32	2.58–10.95	0.000	Greater probability
04:03:01G	03:02:01G	Persistence	6.89	2.82–16.83	0.000	Greater probability
04:05:01	02:01:01G	Infection	− 0.80[b]	− 1.30–(− 0.48)	0.000	Earlier occurrence
		Redetection	− 0.47[a]	− 0.61–(− 0.34)	0.000	Earlier occurrence
04:05:01	03:01:01G	Clearance	0.23	0.11–0.47	0.000	Lower probability
		Persistence	51.72	10.70–249.94	0.000	Greater probability
04:05:04	03:02:01G	Infection	3.08[b]	2.57–3.59	0.000	Later occurrence
04:08:01	02:01:01G	Infection	3.10[b]	2.60–3.61	0.000	Later occurrence
04:08:01	03:01:01G	Infection	3.10[b]	2.60–3.61	0.000	Later occurrence
07:01:01G	02:01:01G	Redetection	0.01[c]	0.00–0.05	0.000	Earlier occurrence

Table 3. (continued)

available. Therefore, comprehensively evaluating a *Ct* infection dynamics-related immune response represents an alternative approach for developing effective control tools[15,16].

This study was focused on comprehensively investigating (the first time) the effect of HLA-*DRB1-DQB1* alleles on *Ct* infection outcome, given that it has been shown that HLA molecules could be related to *Ct*-induced diseases, such as trachoma[17], PID[6,18] and infertility[19,20], or be associated with infection prevalence and bacterial reinfection[8,21,22]. It was found that HLA-*DRB1* alleles were associated with cervical-related *Ct* infection outcome (Tables 2 and 3). Some alleles were less common for this locus and occurred at lower frequency in the target population. MHC-pathogen coevolution models indicate that less commonly occurring alleles provide greater protection against pathogens than more commonly occurring ones to which pathogens may have become adapted[23,24].

*DQB1*05:03:01G was only associated with *Ct* clearance and persistence events (Table 2); however, previous studies have reported HLA-*DQB1* (*DQB1*06* and *DQB1*04:02*) alleles' association with *Ct* infection and reinfection and increased bacterial persistence marker cHSP60[8,21,22]. Such discrepancy could be explained by the genetic background of the particular population being studied (African compared to South-American in this study) thereby contributing to modulating an immune response to bacterial infection[6,25]. However, these alleles only had similar associations to those reported in previous studies when configured as haplotypes, i.e. when they have been combined with a DRB1 allele (Table 3).

07:01:01G	**03:02:01G**	Persistence	**5.25**	**2.10–13.15**	**0.000**	**Greater probability**
07:01:01G	**05:01:01G**	Persistence	**9.52**	**2.77–32.71**	**0.000**	**Greater probability**
		Redetection	**0.01**[c]	**0.00–0.05**	**0.000**	**Earlier occurrence**
07:11	**04:02:01G**	Infection	**2.63**[b]	**2.14–3.12**	**0.000**	**Later occurrence**
08:01:01G	**03:01:01G**	Infection	**− 0.87**[b]	**− 1.20–(− 0.54)**	**0.000**	**Earlier occurrence**
08:02:01G	**03:01:01G**	Persistence	**31.40**	**5.96–165.42**	**0.000**	**Greater probability**
09:01:02G	**03:01:01G**	Clearance	**5.53**	**2.91–10.49**	**0.000**	**Greater probability**
10:01:01G	**05:01:01G**	Persistence	**0.36**	**0.21–0.62**	**0.000**	**Lower probability**
11:01:01G	**03:01:01G**	Persistence	**21.83**	**5.62–84.75**	**0.000**	**Greater probability**
11:01:01G	**03:02:01G**	Persistence	**51.72**	**10.70–249.94**	**0.000**	**Greater probability**
11:04:01G	**02:01:01G**	Persistence	**10.34**	**2.87–37.24**	**0.000**	**Greater probability**
11:04:01G	**03:02:01G**	Infection	**0.65**[b]	**0.38–0.92**	**0.000**	**Earlier occurrence**
11:04:01G	**05:01:01G**	Infection	**− 1.27b**	**− 1.64–(− 0.90)**	**0.000**	**Earlier occurrence**
12:01:01G	**03:02:01G**	Infection	**3.23**[b]	**2.73–3.73**	**0.000**	**Later occurrence**
		Clearance	**5.53**	**2.91–10.49**	**0.000**	**Greater probability**
13:01:01G	**05:02:01G**	Redetection	**− 0.17**[a]	**− 0.27–(− 0.08)**	**0.000**	**Earlier occurrence**
13:01:01G	**06:02:01G**	Infection	**9.58**	**4.44–20.64**	**0.000**	**Greater probability**
13:03:01G	**04:02:01G**	Infection	**− 1.07**[b]	**− 1.44–(− 0.71)**	**0.000**	**Earlier occurrence**
13:05:01	**03:01:01G**	Infection	**2.50**[b]	**1.87–3.13**	**0.000**	**Later occurrence**
		Persistence	**4.20**	**1.97–8.95**	**0.000**	**Greater probability**
13:05:01	**03:02:01G**	Infection	**2.50**[b]	**1.87–3.13**	**0.000**	**Later occurrence**
		Persistence	**4.20**	**1.97–8.95**	**0.000**	**Greater probability**
14:02:01G	**03:02:01G**	Infection	**0.69**[b]	**0.41–0.97**	**0.000**	**Earlier occurrence**
14:02:01G	**04:02:01G**	Persistence	**31.42**	**5.96–165.42**	**0.000**	**Greater probability**
14:02:01G	**05:01:01G**	Persistence	**0.48**[a]	**0.31–0.66**	**0.000**	**Earlier occurrence**
15:01:01G	**05:02:01G**	Infection	**− 0.98**[b]	**− 1.29–(− 0.67)**	**0.000**	**Earlier occurrence**
15:01:01G	**06:03:01G**	Infection	**9.58**	**4.44–20.64**	**0.000**	**Greater probability**
16:01:01	**05:02:01G**	Redetection	**− 0.17**[a]	**− 0.27–(− 0.08)**	**0.000**	**Earlier occurrence**
16:01:01	**06:03:01G**	Redetection	**− 0.17**[a]	**− 0.27–(− 0.08)**	**0.000**	**Earlier occurrence**
16:02:01G	**05:01:01G**	Infection	**2.73**[b]	**2.21–3.25**	**0.000**	**Later occurrence**

Table 3. HLA *DRB1-DQB1* haplotypes associated with *Ct* infection, clearance, persistence and redetection. Analysis was adjusted for age, age at onset of sexual life, lifetime amount of sexual partners, planning method, abortions, history of other STI and HPV infections. Alleles affecting an event are shown as green associations [i.e. greater probability (GP) or earlier occurrence (EO)] and those hindering them [lower probability (LP) or later occurrence (LO)] as red associations. Values in bold indicate statistical significance based on 95%CI, $p < 0.05$. *Pc* corrected p value, *95%CI* 95% confidence interval, *Coeff* regression coefficient, *DRB1* DR beta 1, *DQB1* DQ beta 1. [a]The Cox proportional hazards model did not fulfil the assumption of proportionality, the logistic parametric model was constructed. [b]The Cox proportional hazards model did not fulfil the assumption of proportionality, the lognormal parametric model was constructed. [c]The Cox proportional hazards model did not fulfil the assumption of proportionality, the Gompertz parametric model was constructed. [d]The Bonferroni method was used for correcting all models' *p* values.

Protein name	Event	Locus	Allele	Effect	Peptide
3-OXO-RED	Clearance	*DRB1*	*09:01	Major probability	RGSPGQTNYAAAKAGIIGFS
OMP-C	Infection		*04:10	Later occurrence	MAESLSTNVISLADTKAKDN
	Clearance		*09:01	Major probability	SYVCKPVEYVISVSNPGDLV
					VENPVPDGYAHSSGQRVLTF
PMP-A	Infection		*04:10	Later occurrence	ANKKGGAIYAQYVNLEQNQD
			*12:01		SGDFAGSRILFLNNQQITFE
					EDVANRTSIFNQPVHLYNGT[a]
	Clearance		*09:01	Major probability	PFFSPKGMIVFSGANLLDDA[a]
PMP-B	Infection		*04:10	Later occurrence	TCENSHRLQFLKNSSDKQGG
					SITNITGIIEIANNKATDVG
					VSFENITSLKVQTNGAEEKG
	Persistence		*10:01	Lower probability	RNNPVCKYRVLSSNEAGQVI
					ITLENGSFIFERNQANKRGA
					YFHKGSEYSYQASVYGGKFL
PMP-D	Infection		*04:10	Later occurrence	GAIAAQEIVSIQNNQAGISF
					NCELVDNGYVLFRDNRGRVY
PMP-F			*04:10		RGKAEQPILSIETTNDGQLG
			*12:01		GAIRSGGPIRFLNNQDVLFY[a]
PMP-G			*04:10		VAGVRGGGIAAVQDGQQGVS
			*12:01		EIGVGLPIVITPSKLYLNEL[a]
					GTIYSKTDLLLLNNEKFSFY[a]
LSU	Clearance		*09:01	Major probability	IAVKGNEVFVTPAAHVVDRP
OMP-B					RHASCPIDYIIANSQANPEV[b]
OMP-H	Infection		*04:10	Later occurrence	VLLNEDIVLSIDSSADKTDA
					SSADKTDAVIKVLDDSFQNN
	Persistence		*10:01	Lower probability	AAELRKKFEDLSAEYNTAQG
VPP	Clearance		*09:01	Major probability	EPLKDQQIILGTTSTPVAAK

Table 4. Protein regions having the potential to be T- or B-cell epitopes. T-(green) or B-cell (red) epitope prediction is shown; regions having dual prediction are shown in purple. [a]Peptides which could be related to GP of clearance when they are presented by the same allele. [b]Peptides related to LP of persistence when presented by *DRB1**10:01 allele.

LP of clearance and EO of persistence were found for homozygous HLA-*DRB1* (Supplementary Tables S3 and S4). It has been reported that homozygosity is related to susceptibility to infection whilst heterozygosity is associated with a higher probability of eliminating it, possibly due to a greater immune response, given the broader amount of HLA-II restricted epitopes that can be presented to T-cells[24,26,27]. It is worth noting that findings regarding alleles causing effects on events promoting or reducing infection (Table 2) are useful when designing *Ct* infection control strategies; for example, considering peptides presented by *DRB1**12:01 could represent a good strategy since it is associated with events related to infection resolution whilst peptides presented by DRB1*08:02 should be avoided as it is related to events associated with infection, such as persistence and redetection.

It has been suggested that Ag presentation during adaptative response could be an important mechanism for controlling *Ct* infection[8,28]; identifying T-cell antigens able to stimulate protection-inducing immunity is thus the key for developing anti-*Ct* vaccines[29]. Analysing *Ct* molecules' T-cell epitopes (Supplementary Tables S10 to S24) whose role in protection-inducing immunity was experimentally evaluated[10] revealed that 3-oxoacyl-[acyl-carrier protein] reductase had a peptide related to a GP of *Ct* clearance (Table 4). This peptide had been found in an immunoproteomics study demonstrating that inoculating dendritic cells previously pulsed with a peptide mixture (including the peptide discussed here) triggered a response partially protecting mice from intranasal and genital tract *Chlamydia* infection[30]. It has been demonstrated that CPAF-derived peptides could be related to a protection-inducing effect in a HLA-*DR4* (HLA-*DRB1*04:01*) transgenic mouse model[31], however, no allele/peptide association was found in this study (Supplementary Table S12), possibly due to the allele's low frequency in the studied population (less than 0.2%), suggesting an allele-specific effect.

The OMPs and PMPs had various regions containing T-cell epitopes, the most important ones being related to events associated with *Ct* infection elimination (Table 4). Interestingly, there was discrepancy amongst several events which could have been explained by binding core mutations for some PMP-derived peptides (PMP-B and PMP-F); these enabled discriminating between invasive (L2 and LGV) and non-invasive variants (A, B, C, D, E, F, H, J, Ja and La) (Supplementary Table S16 and S18).

It has been reported recently that the *Ct* OMP (CTH522) protein being evaluated in phase 1 trials was able to trigger a more consistent cell-mediated immune response profile after its immunisation using CAF01 liposomes compared to the placebo group, thus highlighting its potential usefulness as a vaccine candidate[32]. Such result, added to this study's findings, supports the idea that T-cell epitopes derived from the antigens analysed here (mainly surface-derived molecules, particularly totally conserved ones) could be regions of interest for the future design of novel interventions aimed at controlling *Ct* infection.

In silico analysis suggested that some protection-related predicted peptides would specifically stimulate T-cells whilst others would stimulate both T- and B-cells (Table 4). Vaccination with various OMP serovars (D, E and F) has elicited an antibody (Ab) response neutralising bacteria in vitro[33]. Furthermore, PMPs can trigger an immune response against genital[34] and ocular[35] *Chlamydia* infection and a serological response in humans[36]. Mice vaccinated with DC/PMP-derived peptides (G, E and F) or with just immunogens in formulation have developed immunity against genital tract and pulmonary *Chlamydia* infection, significantly reducing bacteria in UFI assays[7,37]. Interestingly, phase 1 clinical trial vaccination using the *Ct*-OMP version showed accelerated seroconversion, increased IgG titres and enhanced mucosal profile, thus making CTH522 a promising candidate for further clinical development[32].

As CD4 T-cells are essential for resolving primary genital infection[38] and CD8 T-cells are important for eliminating *Ct*-infected cells by effector mechanisms[39], then peptides stimulating both types of effector cells (B- and T), as predicted here, could be considered most suitable for controlling *Ct* infection and therefore as promising candidates for future studies (Table 4).

Antigenic peptides recognised by both CD4 and CD8 T-cells could be promising diagnostic and therapeutic tool candidates since one of the main limitations for developing an effective vaccine lies in identifying *Ct* epitopes capable of being recognised by both cell types[40]; an Ab-mediated immune response would reduce bacterial load, thereby facilitating further elimination of infection via a cell-mediated immune response[33,40].

Considering this and given the in silico analysis performed here, it can thus be suggested that a universal anti-*Ct* vaccine should contain peptides having the following characteristics: they should induce immunity and be protein-derived, have a high degree of conservation, be associated with protection-related events (such as GP of clearance and LP of infection, redetection and persistence) and be able to stimulate T- and B-cell responses.

HLA-*DRB1-DQB1* alleles/haplotypes having an effect on *Ct* resolution have thus been reported here, along with in silico predicted epitopes derived from protection-related proteins targeting *Ct* infection. However, functional read-outs for demonstrating the effect of presentation regarding the predicted antigens (i.e. wet-laboratory assays) were not performed here, thus failing to obtain a complete panorama of anti-*Ct* immune responses constitutes a limitation of this study.

Addressing new prophylactic and therapeutic targets must become a high priority as the tools used to date for *Ct* control have not had a significant impact on reducing bacterial infection load. Future analysis should be aimed at validating predicted epitopes' immunogenic and immunological in vitro and in vivo properties and their safe and efficacy regarding humans. Such data will provide relevant knowledge for understanding the usefulness of peptides as a vaccine component and the influence of host factors on the clinical course of *Ct* infection.

Materials and methods

Study design and participants. A cohort was studied between 2007 and 2010; that previous study was aimed at determining the natural history of HPV infection in women from the Colombian cities of Bogotá, Girardot and Chaparral; all the women were attending hospital clinics as outpatients. The study's objective was explained to them and they voluntarily accepted participating in the study by signing an informed consent form, as described previously[13]. Retrospective analysis inclusion criteria consisted of having available cervical samples for typing HLA-*DRB1* and *DQB1* and *Ct* detection, women having attended at least four follow-up sessions (one base line and three visits) and 6-monthly periods between visits (±3 months).

The women filled in a survey form during each visit for compiling data regarding sociodemographic information and risk factors. Such information included data regarding whether they had received/used any type of treatment between visits; none of the women reported using antibiotics during follow-up. The women did not receive antibiotic treatment for *Ct* infections detected during the study, given the retrospective nature of *Ct* detection. The Universidad del Rosario's School of Medicine and Health Sciences Research Ethics Committee approved the study (CEI-ABN026-000135). All procedures were performed in accordance with Helsinki Declaration guidelines.

***Ct* detection and HLA-*DRB1-DQB1* typing.** Previously obtained genomic DNA (gDNA)[13,41] was used as template for detecting *Ct* by conventional PCR, amplifying a cryptic plasmid ORF2 segment from KL5/KL6 and KL1/KL2 primers[41]. An Illumina MiSeq (San Diego, CA, USA; Histogenetics, Ossining, NY, USA) sequencer was used for typing HLA alleles from *DRB1-DQB1* loci exons 2 and 3; the IPD-IMGT/HLA database (https://www.ebi.ac.uk/ipd/imgt/hla) published in January 2018 (3.31.0) was used for assigning alleles[13].

Statistical analysis. Qualitative variables were expressed as percentages. The Chi^2 or Fisher's exact tests were used for evaluating association/concordance amongst categorical variables. Continuous variables were expressed as means [with standard deviations (SD) for measure of dispersion] or medians [interquartile ranges (IQR)].

Ct infection was defined for this study's purposes as PCR detection of bacterial DNA at any point during follow-up (2 years). Clearance was understood as the elimination of infection via a previous positive *Ct* result. A percentage of infections persisted before becoming eliminated; this event was evaluated in the study and was defined as *Ct* being detected during two or more consecutive follow-ups. Redetection was defined as bacterial detection after not having detected bacterial DNA during a previous follow-up.

Incidence rates for events were reported along with 95% confidence intervals (95%CI). The Kaplan–Meier estimator was used for estimating the probability of subjects continuing event-free. Cox proportional hazards models were constructed for evaluating outcome probability; such models' coefficients were expressed as hazard rate (HR) and used for identifying alleles/haplotypes related to the evens being evaluated. Schoenfeld residuals were plotted to test the proportional hazard assumption; the covariables considered for the plot were those

having $p < 0.200$ in univariate analysis (Supplementary Table S2). Variance inflation factor (VIF) and tolerance values were used for evaluating multicollinearity between covariables[13].

Different parametric survival models were constructed when proportional hazards assumptions were not met. Akaike (AIC) and Bayesian (BIC) information criteria were used for selecting the models having the best fit. The Bonferroni method was used for correcting p-values for each model[13]. All two-tailed hypothesis tests (except those involved in constructing the models) were run with 0.05 significance. STATA14 software was used for analysis.

T- and B-cell epitope prediction. *Ct* variant-derived protein amino acid (aa) sequences (Supplementary Table S25) were downloaded from the PATRIC 3.5.11 database (https://www.patricbrc.org) and analysed first using classical pathway secretion predictor (SignalP 5.0)[42] and those not assigned by this predictor were analysed by the non-classical one, SecretomeP 2.0[43].

The Technical University of Denmark's Systems Biology Department's Center for Biological Sequence Analysis' NetMHCIIpan 3.2 server was used for assessing peptides having high predicted HLA-*DRB1-DQB1* allele binding activity, i.e. proteins predicted as secreted[44]. Peptides having < 2.0% rank were considered to have strong binding. The BepiPred 2.0 tool (0.6 epitope threshold) was used for calculating B-cell epitopes derived from proteins whose peptides were associated with protection[45].

Data availability

The datasets produced and/or analysed during this study are available from the corresponding author on reasonable request.

Received: 18 February 2021; Accepted: 9 June 2021
Published online: 18 June 2021

References

1. Nguyen, N. D. N. T. *et al.* Parenteral vaccination protects against transcervical infection with *Chlamydia trachomatis* and generate tissue-resident T cells post-challenge. *NPJ Vaccines* **5**, 1–12 (2020).
2. WHO. *Report on Global Sexually Transmitted Infection Surveillance*. World Health Organization, Geneva, Switzerland. (2018).
3. Markle, W., Conti, T. & Kad, M. Sexually transmitted diseases. *Prim. Care Clin. Off. Pract.* **40**, 557–587 (2013).
4. Fogue, P. *et al.* Molecular characterization of lower vaginal swabs for Human papilloma virus in association with *Chlamydia trachomatis* infection in Cameroonian women. *J. Infect. Public Health* **11**, 314–320 (2018).
5. Geisler, W. M. *et al.* Epidemiological and genetic correlates of incident *Chlamydia trachomatis* infection in North American adolescents. *J. Infect. Dis.* **190**, 1723–1729 (2004).
6. Ness, R. B., Brunham, R. C., Shen, C. & Bass, D. C. Associations among human leukocyte antigen (HLA) class II DQ variants, bacterial sexually transmitted diseases, endometritis, and fertility among women with clinical pelvic inflammatory disease. *Sex. Transm. Dis.* **31**, 301–304 (2004).
7. Yu, H. *et al.* Chlamydia muridarum T cell antigens and adjuvants that induce protective immunity in mice. *Infect. Immun.* **80**, 1510–1518 (2012).
8. Geisler, W. M., Tang, J., Wang, C., Wilson, C. M. & Kaslow, R. A. Epidemiological and genetic correlates of incident *Chlamydia trachomatis* infection in North American adolescents. *J. Infect. Dis.* **190**, 1723–1729 (2004).
9. Hafner, L. M. & Timms, P. Development of a *Chlamydia trachomatis* vaccine for urogenital infections: novel tools and new strategies point to bright future prospects. *Expert Rev. Vaccines* **17**, 57–69 (2018).
10. de la Maza, L. M., Zhong, G. & Brunham, R. C. Update on *Chlamydia trachomatis* vaccinology. *Clin. Vaccine Immunol.* **24**, e00543 (2017).
11. Timms, P. & Hafner, L. Development of a vaccine for *Chlamydia trachomatis*: challenges and current progress. *Vaccine Dev. Ther.* **5**, 45–58 (2015).
12. Zhong, G. & Brunham, R. C. Update on *Chlamydia trachomatis* vaccinology. *Clin. Vaccine Immunol.* **24**, 1–25 (2017).
13. Del Rio-Ospina, L. *et al.* Identifying the HLA DRB1-DQB1 molecules and predicting epitopes associated with high-risk HPV infection clearance and redetection. *Sci. Rep.* **10**, 7306 (2020).
14. Hurley, C. K. *et al.* Common, intermediate and well-documented HLA alleles in world populations: CIWD version 3.0.0. *HLA* **95**, 516–531 (2020).
15. Unemo, M. *et al.* Sexually transmitted infections: challenges ahead. *Lancet Infect. Dis.* **17**, e235–e279 (2017).
16. Ziklo, N., Huston, W. M., Hocking, J. S. & Timms, P. *Chlamydia trachomatis* genital tract infections: when host immune response and the microbiome collide. *Trends Microbiol.* **24**, 750–765 (2016).
17. Conway, D. J. *et al.* HLA class I and II polymorphisms and trachomatous scarring in a *Chlamydia trachomatis*-endemic population. *J. Infect. Dis.* **174**, 643–646 (1996).
18. Kimani, J. *et al.* Risk factors for *Chlamydia trachomatis* pelvic inflammatory disease among sex workers in Nairobi, Kenya. *J. Infect. Dis.* **173**, 1437–1444 (1996).
19. Kinnunen, A. H. *et al.* HLA DQ alleles and interleukin-10 polymorphism associated with *Chlamydia trachomatis*-related tubal factor infertility: a case-control study. *Hum. Reprod.* **17**, 2073–2078 (2002).
20. Cohen, C. R. *et al.* Immunogenetic correlates for *Chlamydia trachomatis*-associated tubal infertility. *Obstet. Gynecol.* **101**, 438–444 (2003).
21. Gaur, L. K. *et al.* Association of *Chlamydia trachomatis* heat-shock protein 60 antibody and HLA class II DQ alleles. *J. Infect. Dis.* **180**, 234–237 (1999).
22. Olson, K. M., Tang, J., Brown, L., Press, C. G. & Geisler, W. M. HLA-DQB1*06 is a risk marker for chlamydia reinfection in African American women. *Genes Immun.* **20**, 69–73 (2019).
23. Borghans, J. A., Beltman, J. B. & De Boer, R. J. MHC polymorphism under host-pathogen coevolution. *Immunogenetics* **55**, 732–739 (2004).
24. Hedrick, P. W. Pathogen resistance and genetic variation at MHC loci. *Evolution* **56**, 1902–1908 (2002).
25. Mosaad, Y. M. Clinical role of human leukocyte antigen in health and disease. *Scand. J. Immunol.* **82**, 283–306 (2015).
26. Blackwell, J. M., Jamieson, S. E. & Burgner, D. HLA and infectious diseases. *Clin. Microbiol. Rev.* **22**, 370–385 (2009).
27. Burgner, D., Jamieson, S. E. & Blackwell, J. M. Genetic susceptibility to infectious diseases: Big is beautiful, but will bigger be even better?. *Lancet Infect. Dis.* **6**, 653–663 (2006).
28. Brunham, R. C. Perspective: my 37 year journey through Chlamydia research: *Chlamydia* antigen analysis using monoclonal antibodies and major histocompatibility complex molecules. *Pathog. Dis.* **75**, 1–9 (2017).

29. Brunham, R. C. & Rey-Ladino, J. Immunology of Chlamydia infection: implications for a *Chlamydia trachomatis* vaccine. *Nat. Rev. Immunol.* **5**, 149–161 (2005).
30. Karunakaran, K. P. *et al.* Immunoproteomic discovery of novel T cell antigens from the obligate intracellular pathogen Chlamydia. *J. Immunol.* **180**, 2459–2465 (2008).
31. Li, W. *et al.* A T cell epitope-based vaccine protects against chlamydial infection in HLA-DR4 transgenic mice. *Vaccine* **31**, 5722–5728 (2013).
32. Abraham, S. *et al.* Safety and immunogenicity of the chlamydia vaccine candidate CTH522 adjuvanted with CAF01 liposomes or aluminium hydroxide: a first-in-human, randomised, double-blind, placebo-controlled, phase 1 trial. *Lancet Infect. Dis.* **19**, 1091–1100 (2019).
33. Olsen, A. W., Follmann, F., Erneholm, K., Rosenkrands, I. & Andersen, P. Protection against *Chlamydia trachomatis* infection and upper genital tract pathological changes by vaccine-promoted neutralizing antibodies directed to the VD4 of the major outer membrane protein. *J. Infect. Dis.* **212**, 978–989 (2015).
34. Karunakaran, K. P. *et al.* Outer membrane proteins preferentially load MHC class II peptides: implications for a *Chlamydia trachomatis* T cell vaccine. *Vaccine* **33**, 2159–2166 (2015).
35. Inic-Kanada, A. *et al.* Delivery of a chlamydial adhesin N-PmpC subunit vaccine to the ocular mucosa using particulate carriers. *PLoS ONE* **10**, e0144380 (2015).
36. Vasilevsky, S., Stojanov, M., Greub, G. & Baud, D. Chlamydial polymorphic membrane proteins: regulation, function and potential vaccine candidates. *Virulence* **7**, 11–22 (2016).
37. Yu, H., Jiang, X., Shen, C., Karunakaran, K. P. & Brunham, R. C. Novel *Chlamydia muridarum* T cell antigens induce protective immunity against lung and genital tract infection in murine models. *J. Immunol.* **182**, 1602–1608 (2009).
38. Farris, C. M. & Morrison, R. P. Vaccination against Chlamydia genital infection utilizing the murine *C. muridarum* model. *Infect. Immun.* **79**, 986–996 (2011).
39. Wizel, B., Nystrom-Asklin, J., Cortes, C. & Tvinnereim, A. Role of CD8(+)T cells in the host response to Chlamydia. *Microbes Infect.* **10**, 1420–1430 (2008).
40. Russi, R. C., Bourdin, E., Garcia, M. I. & Veaute, C. M. I. In silico prediction of T- and B-cell epitopes in PmpD: first step towards to the design of a *Chlamydia trachomatis* vaccine. *Biomed. J.* **41**, 109–117 (2018).
41. Quinónez-Calvache, E. M. *et al. Chlamydia trachomatis* frequency in a cohort of HPV-infected Colombian women. *PLoS ONE* **11**, 1–14 (2016).
42. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).
43. Bendtsen, J. D., Kiemer, L., Fausboll, A. & Brunak, S. Non-classical protein secretion in bacteria. *BMC Microbiol.* **5**, 58 (2005).
44. Nielsen, M. *et al.* Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS Comput. Biol.* **4**, e1000107 (2008).
45. Jespersen, M. C., Peters, B., Nielsen, M. & Marcatili, P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* **45**, W24–W29 (2017).

Acknowledgements

This work was supported by Colombia's General Royalties System, project BPIN-233, special agreement 021, entitled, "Developing a management plan for controlling cervical cancer in the Amazonas department," which was approved by the Amazonas Department Local Government and Colombia's Science, Technology and Innovation Department (COLCIENCIAS). Neither entity participated in developing the study. We would like to thank the Colombia's Science, Technology and Innovation Department (COLCIENCIAS) for financing MC as part of the Colombian Programme for Promoting Researcher Training—Call 617 (PhD study in Colombia). Sara Cecilia Soto De Leon provided technical support. We would like to thank Jason Garry for translating the manuscript.

Author contributions

L.P., M.C. and D.A.M.P.: designed the study, performed the experiments, analysed the data, wrote the first draft and revised the final draft of the manuscript. R.S. and L.D.R.O.: designed the study, contributed to data analysis and critically reviewed the first and final drafts of the manuscript. I.M.B.M.: contributed to the experiments and critically reviewed the first and final drafts of the manuscript. M.E.P: and M.A.P.: critically reviewed the first and final drafts of the manuscript. All authors approved the final draft.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-92294-w.

Correspondence and requests for materials should be addressed to M.A.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021