ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 21(9): 3140-3146 (2025)

BIOINFORMATION
Discovery at the interface of physical and biological sciences
BIOMEDICAL INFORMATICS
www.bioinformation.net
Volume 21(9)
Research Article

Received September 1, 2025; Revised September 30, 2025; Accepted September 30, 2025, Published September 30, 2025
DOI: 10.6026/973206300213140

SJIF 2025 (Scientific Journal Impact Factor for 2025) = 8.478
2022 Impact Factor (2023 Clarivate Inc. release) is 1.9

Edited by P Kangueane
Citation: Yadav *et al.* Bioinformation 21(9): 3140-3146 (2025)

Analysis of outer membrane protein from *Orientia tsutsugamushi* for epitope mining

Jogender Yadav[1], Sudheer Gupta[2], Ankur Joshi[3], Abhijit P Pakhare[3], Priyal Gupta[1], Santenna Chenchula[4] & Shashank Purwar[1,*]

[1]Department of Microbiology, All India Institute of Medical Sciences Bhopal, Madhya Pradesh, India; [2]NGS & Bioinformatics Division, 3B BlackBio DX Ltd., Bhopal, Madhya Pradesh, India; [3]Department of Community and Family Medicine, All India Institute of Medical Sciences Bhopal, Madhya Pradesh, India; [4]Department of Pharmacology, All India Institute of Medical Sciences Bhopal, Madhya Pradesh, India; *Corresponding author

Affiliation URL:
https://www.aiimsbhopal.edu.in/
https://3bblackbio.com

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 21(9): 3140-3146 (2025)

Author contacts:
Jogender Yadav - E-mail: jogender.phd2022@aiimsbhopal.edu.in
Sudheer Gupta - E-mail: Sudheer.bioinfo@gmail.com
Ankur Joshi - E-mail: ankur.cfm@aiimsbhopal.edu.in
Abhijit P Pakhare - E-mail: Abhijit.cfm@aiimsbhopal.edu.in
Priyal Gupta - E-mail: priyal.phd2019@aiimsbhopal.edu.in
Santenna Chenchula - E-mail: csanten7@gmail.com
Shashank Purwar - E-mail: shashank.microbiology@aiimsbhopal.edu.in

Abstract:
Scrub typhus, caused by *Orientia tsutsugamushi*, is an emerging health threat that can lead to severe complications if untreated. Therefore, it is of interest to analyse 130 sequences of the 56 kDa outer membrane protein, a key immunodominant antigen. SNP profiling identified 300 conserved and 8 highly variable residues. Epitope prediction revealed 73 conserved and 6 variable B cell epitopes, along with 4 conserved and 1 variable $CD8^+$ T cell epitopes and 3 conserved and 1 variable $CD4^+$ T cell epitopes. Notably, five immunogenic regions were predicted to activate B cells and both $CD8^+$ and $CD4^+$ T cells simultaneously. Thus, we show valuable insight into immune recognition and evasion strategies of *O. tsutsugamushi*. Subject to experimental validation, the identified epitopes represent promising candidates for rational vaccine development, improved serodiagnostics and novel therapeutic interventions against scrub typhus.

Keywords: Orientia tsutsugamushi, Immune epitope, B-cell Epitope, T-cell Epitope, 56 kDa gene

Background:
Scrub typhus, caused by *Orientia tsutsugamushi* and transmitted by the larval stage of trombiculid mites, is a re-emerging infectious disease predominantly affecting rural regions across the Asia-Pacific. It is a major cause of acute febrile illness (AFI), often presenting with nonspecific symptoms and, in some cases, an eschar. If untreated, it can progress to severe complications, including multi-organ failure, with mortality rates ranging from 7–30% **[1, 2].** Globally, approximately one billion people are at risk and an estimated one million cases occur annually **[2–4]**. Multiple genes in *O. tsutsugamushi* have been investigated for their roles in pathogenicity and immune recognition, including the 22 kDa, 47 kDa, 56 kDa, 58 kDa, 110 kDa and Sca family proteins (ScaA–ScaE). Among these, the 56 kDa outer membrane protein is particularly notable for its high antigenicity and sequence variability **[5, 6]**. This gene, with an open reading frame of about 1600 bp encoding a 516–541 amino acid protein, is implicated in host cell invasion via fibronectin binding. It also contains four hypervariable regions (I–IV), which significantly contribute to its antigenic diversity **[7, 8]**. To date, more than 40 genotypes of *O. tsutsugamushi* have been identified based on variation in the 56 kDa gene **[9]**. Analysing both conserved and variable regions of the 56 kDa gene is crucial in epitope prediction studies for vaccine development, understanding immune response and designing diagnostic tools **[5, 6]**. Therefore, it is of interest to analyse the outer membrane protein from Orientia tsutsugamushi for epitope mining.

Methodology:
Sequences retrieval:
A systematic search was conducted in the National Centre for Biotechnology Information (NCBI) **[10]** database to retrieve 130 full-length amino acid sequences derived from the 56 kDa gene of *Orientia tsutsugamushi*. The search was performed using the keyword "*Orientia tsutsugamushi* 56 kDa gene "complete" in FASTA format. These sequences were exclusively sourced from human specimens collected between 1993 and 2024. The Gene Bank sequence AP008981.1 was used as the reference sequence.

MSA and SNP analysis:
Multiple sequence alignment (MSA) and single-nucleotide polymorphism (SNP) analysis were performed using the resources available at the Bacterial and Viral Bioinformatics Resource Centre (BV-BRC). For aligning the sequences, MAFFT was used as the alignment algorithm **[11-15]**. All 130 sequences were examined for MSA and SNP analysis to determine the entropy score, which facilitated the identification of genetic variations within the 56 kDa gene of *Orientia tsutsugamushi*. Based on the MSA and SNP analysis, the entropy score was calculated using the SNP analysis tool. The entropy score represents the distance between the amino acid residues in different sets of homologous protein sequences. Based on the range of entropy scores, they were subdivided into bins: 0-50, 51-100, 101-150, 151-200 and above 200. Residues with an entropy score ranging between 0 and 50 were considered conserved, while residues with an entropy score greater than 200 were considered highly variable. On the other hand, the variable category (51-200) was divided into three groups: Category 1 (51-100), Category 2 (101-150) and Category 3 (151-200).

B-Cell epitope prediction:
B-cell epitope (BCE) prediction was performed using the Immune Epitope Database and Tools (IEDB) **[16]**. Predictions were generated through four distinct methods available in the IEDB: Parker Hydrophilicity Prediction, Kolaskar & Tingaonkar Antigenicity, Karplus & Schulz Flexibility Prediction and Chou & Fasman Beta-Turn Prediction. The cut-off values were 1.704, 1.031, 0.997 and 0.971, respectively. These cut-off values were the default settings, as they were optimised during the standardisation and validation of the tool. A site was considered

a BCE if the score was above the cut-off value. A consensus result was established by considering a region as a B-cell epitope site if it elicited positive predictions from three or more of the aforementioned methods.

MHC-I epitope prediction:
Using the Immune Epitope Database and Tools (IEDB), MHC class-I epitope prediction was conducted. Predictions were generated through four distinct methods, namely NetMHCpan 4.1 BA **[17-21]**, ANN **[22-27]**, SMM **[28]** and SMMPMBEC **[29]**, which are available in the IEDB. The predictions were carried out for the 27 alleles recommended by the IEDB **[16]**.

MHC-II epitope prediction:
The prediction of MHC class-II epitopes was also carried out using the IEDB platform. Four different prediction methods, namely NetMHCIIPan 4.1 BA, NetMHCIIPan 4.3 BA, NetMHCIIPan 4.2 BA **[17, 30-34]** and NN-align 2.3 **[33, 35]**, were employed. The prediction was conducted for the 27 alleles suggested by the IEDB. Both MHC class I and MHC class II results were analysed based on the IC50 (half-minimal inhibitory concentration) value. A lower IC50 value indicates stronger binding efficiency of the peptide to the HLA molecule, while a higher IC50 value indicates weaker binding efficiency. According to the method's recommendations for both MHC class I and II peptides, if the IC50 value of any peptide is equal to or less than 50, it is considered a strong binder. Peptides with IC50 values between 51 and 500 are considered intermediate binders, while peptides with IC50 values greater than 500 are considered weak/non-binders **[16]**.

Table 1: List of strong HLA class I binders with their entropy score

S. No.	Peptide	Allele	Entropy Score
1	YLRNISAEV	HLA-A*02:01	172
2	KLYADLFTT	HLA-A*02:01	152
3	MLIASAMSA	HLA-A*02:03	**0**
4	LIASAMSAL	HLA-A*02:03	**0**
5	TLAAGMTIA	HLA-A*02:03	**31**
6	YLRNISAEV	HLA-A*02:03	172
7	VLSDKITKI	HLA-A*02:03	165
9	MVASGALGV	HLA-A*02:03	58
10	SINPLMASV	HLA-A*02:03	134
11	MLIASAMSA	HLA-A*02:06	**0**
12	NQIQLNFQI	HLA-A*02:06	143
13	KLYADLFTT	HLA-A*02:06	152
14	YTSGKIDGV	HLA-A*02:06	160
15	MVASGALGV	HLA-A*02:06	58
16	GSYMYSFSK	HLA-A*11:01	99
17	AVRRIAWLK	HLA-A*30:01	170
18	LQRHAGIRK	HLA-A*30:01	113
19	AINAAEGVY	HLA-A*30:02	97
20	AMSALSLPF	HLA-A*32:01	**13**
21	KIYSDIRQF	HLA-A*32:01	135
22	RQFAKVANI	HLA-A*32:01	155
23	MITGVESTR	HLA-A*68:01	138
24	EAAAAAAVR	HLA-A*68:01	112
25	YSFSKIEEK	HLA-A*68:01	105
26	MTIAPGFRA	HLA-A*68:02	86
27	DYAGIDYMV	HLA-A*68:02	112
28	ESFSIYAGL	HLA-A*68:02	138
29	YTSGKIDGV	HLA-A*68:02	160
30	MVASGALGV	HLA-A*68:02	58
31	RPRQDLNIL	HLA-B*07:02	**205**
32	APLPNSASV	HLA-B*07:02	162
33	AMSALSLPF	HLA-B*15:01	**13**
34	LMASVGVRY	HLA-B*15:01	134

Results and Discussion:
In this study, the reference sequence of the 56 kDa antigen of *Orientia tsutsugamushi* comprised 516 amino acids. Among these, 300 residues exhibited low entropy scores (<50), indicating high conservation across different strains. In contrast, 8 residues showed high entropy scores (>200), signifying considerable variability. Complete entropy data are presented in **Figures 1–3** and Supplementary File 1. Using four different prediction methods, we identified 179 B-cell epitope (BCE) residues that were predicted by at least three methods. Of these, 73 were located in conserved regions, while 6 were within highly variable regions. These findings are illustrated in **Figure 4** and detailed in Supplementary File 2. In **Figure 4**, BCEs in conserved regions are marked in yellow (with green highlighting zero-entropy residues), while red indicates BCEs in variable regions. T-cell epitope prediction was also performed. For MHC class I, 508 peptides were identified across 27 common alleles, with 33 classified as strong binders. Among these, four peptides were located in conserved regions: MLIASAMSA (HLA-A02:03, HLA-A02:06), LIASAMSAL (HLA-A02:03), TLAAGMTIA (HLA-A02:03) and AMSALSLPF (HLA-A32:01, HLA-B15:01). One strong binder, RPRQDLNIL (HLA-B*07:02), was identified in a highly variable region (**Table 1** and Supplementary File 3). For MHC class II, 502 peptides were predicted, of which 142 were strong binders. Three strong binders were located in conserved regions: MKKIMLIASAMSALS (HLA-DRB108:02), KIMLIASAMSALSLP (HLA-DRB109:01) and LIASAMSALSLPFLA (HLA-DQA105:01/DQB103:01). One peptide, ADGKKHLSLITGIPF (HLA-DRB1*07:01), was found in a highly variable region (**Table 2** and Supplementary File 4).

We further analysed regions with overlapping B-cell and T-cell epitopes. These combined antigenic sites are promising vaccine candidates as they could stimulate both humoral and cellular immune responses. Five strong MHC class I peptides also overlapped with BCEs and strong MHC class II binders: VLSDKITKI (HLA-A02:03), YTSGKIDGV (HLA-A02:06, HLA-A68:02), KIYSDIRQF (HLA-A32:01), MITGVESTR (HLA-A68:01) and LMASVGVRY (HLA-B15:01) (**Table 3**). Identifying conserved epitopes is crucial for designing vaccines and diagnostics with broad strain coverage. Conversely, studying variable regions helps us understand immune escape mechanisms that may contribute to disease severity or persistence. The TSA56 protein, a dominant surface antigen constituting ~20% of the bacterial proteome, remains a major focus for vaccine and serological assay development **[5, 36]**. The 300 conserved residues identified in our study offer a strong foundation for further exploration of immune targets. Meanwhile, the 8 highly variable residues could contribute to strain-specific immune responses or immune evasion strategies. B-cell epitopes play a central role in antibody-mediated immunity, which is vital in the defence against *O.*

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 21(9): 3140-3146 (2025)

tsutsugamushi **[37, 38]**. Previous studies have demonstrated the importance of B-cell responses in acquired immunity **[6, 39, 40]**. For instance, Namsa *et al*. **[6]** proposed several B-cell epitopes, including QQQGQVTAQEA, FSKIEEKYSINP, PISIADRDFGIDIPN and PFGGTLAAGMTI. We found conserved residues within these sequences. However, two peptides reported by them—GGNAFANQIQLNF and TDSEGKKHLSLTTG—were located in highly variable regions and were not predicted as B-cell epitopes in our analysis. Similarly, Dolley *et al*. **[36]** identified a linear B-cell epitope (VGETKADSVGGKDAPI) that binds a Karp strain-specific monoclonal antibody. Our findings showed this region contains variable residues, indicating that some strains may evade recognition by such antibodies. T-cell immunity is also crucial for protection **[5, 6, 36, 41–43]**. N-Y Ha *et al*. **[5]** identified two MHC class I peptides—YLRNISAEV and QLYKDLVKL—that triggered robust IFN-γ responses in recovered patients. We found YLRNISAEV to be a strong MHC class II binder, though it was located in a variable region, while QLYKDLVKL was not predicted as a strong binder in our study. Notably, the same study reported MLIASAMSA (HLA-A*0201) as a strong MHC class I binder, which we also identified in a conserved region. Ramaiah *et al*. **[8]** reported CD4+ T-cell epitopes from Indian and South Korean isolates, identifying 68 and 14 epitopes, respectively. We found overlaps with 23 of the 68 Indian epitopes and 13 of the 14 South Korean epitopes, although none were located in conserved regions. To our knowledge, this is the first comprehensive study analyzing variability within the 56 kDa antigen and identifying both conserved and variable regions capable of inducing B-cell, CD4+ and CD8+ T-cell responses. These epitopes hold promise as vaccine candidates and diagnostic markers.

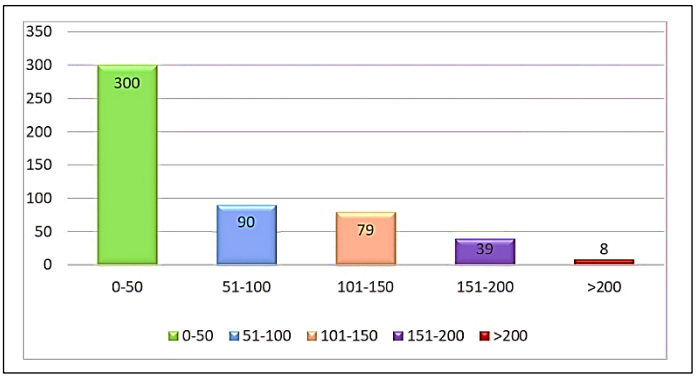

Figure 1: Overview of the entropy scores of all residues

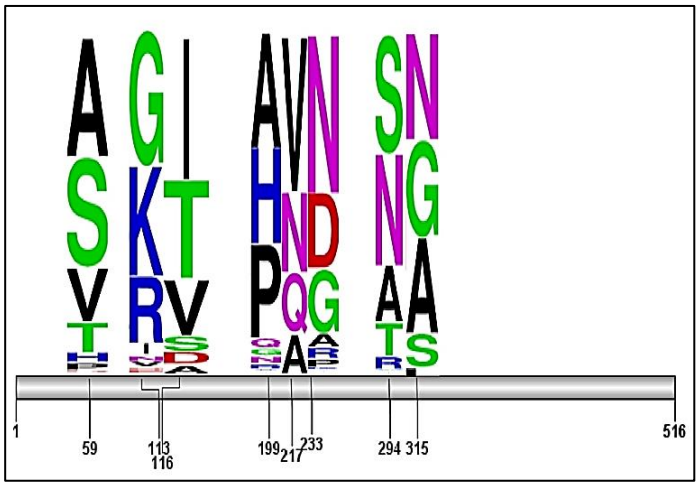

Figure 3: Highly variable sites with the amino acid substitutions on various positions.

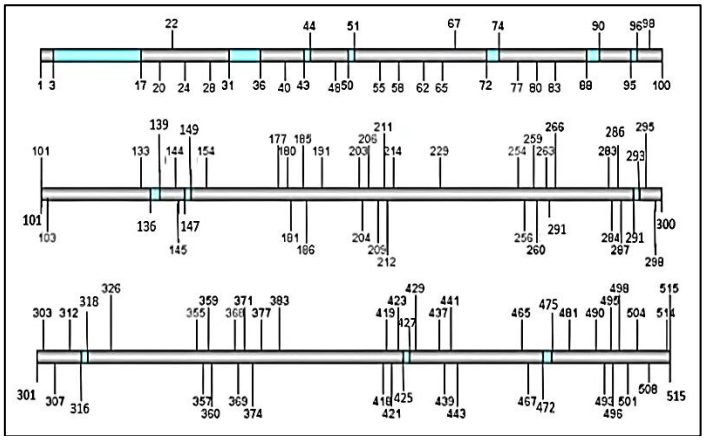

Figure 2: Conserved sites having zero entropy score (highlighted in blue region indicate the region in the sequence with zero entropy score)

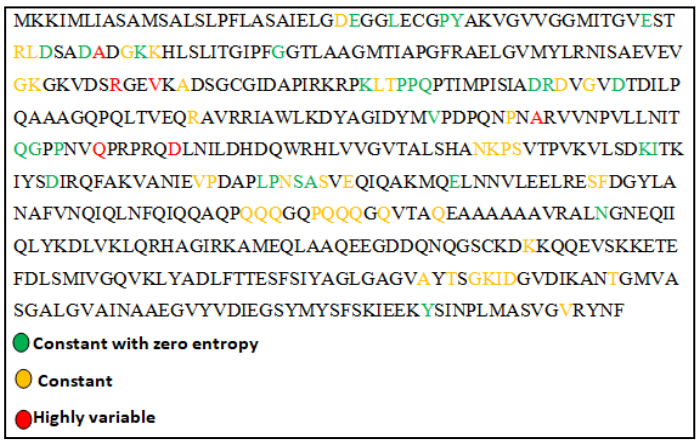

Figure 4: Presenting B cell residues with constant and variable regions

Table 2: List of HLA class II strong binders with their entropy score

S. No.	Peptide	Allele	Entropy Score	S. No.	Peptide	Allele	Entropy Score
1	TLAAGMTIAPGFRAE	HLA-DRB1*01:01	86	72	VMYLRNISAEVEVGK	HLA-DRB3*02:02	180
2	AAGMTIAPGFRAELG	HLA-DRB1*01:01	86	73	MYLRNISAEVEVGKG	HLA-DRB3*02:02	180
3	MTIAPGFRAELGVMY	HLA-DRB1*01:01	86	74	IEEKYSINPLMASVG	HLA-DRB3*02:02	134

4	TIAPGFRAELGVMYL	HLA-DRB1*01:01	86	75	EEKYSINPLMASVGV	HLA-DRB3*02:02	134
5	IAPGFRAELGVMYLR	HLA-DRB1*01:01	130	76	EKYSINPLMASVGVR	HLA-DRB3*02:02	134
6	APGFRAELGVMYLRN	HLA-DRB1*01:01	130	77	KYSINPLMASVGVRY	HLA-DRB3*02:02	134
7	GFRAELGVMYLRNIS	HLA-DRB1*01:01	172	78	KDLVKLQRHAGIRKA	HLA-DRB4*01:01	113
8	HDQWRHLVVGVTALS	HLA-DRB1*01:01	184	79	YKDLVKLQRHAGIRK	HLA-DRB5*01:01	113
9	IQLNFQIQQAQPQQQ	HLA-DRB1*01:01	143	80	KDLVKLQRHAGIRKA	HLA-DRB5*01:01	113
10	LNFQIQQAQPQQQGQ	HLA-DRB1*01:01	143	81	VKLQRHAGIRKAMEQ	HLA-DRB5*01:01	113
11	TESFSIYAGLGAGVA	HLA-DRB1*01:01	138	82	ETEFDLSMIVGQVKL	HLA-DRB5*01:01	158
12	SFSIYAGLGAGVAYT	HLA-DRB1*01:01	112	83	TEFDLSMIVGQVKLY	HLA-DRB5*01:01	158
13	FSIYAGLGAGVAYTS	HLA-DRB1*01:01	112	84	EFDLSMIVGQVKLYA	HLA-DRB5*01:01	105
14	IYAGLGAGVAYTSGK	HLA-DRB1*01:01	138	85	FDLSMIVGQVKLYAD	HLA-DRB5*01:01	105
15	GVDIKANTGMVASGA	HLA-DRB1*01:01	161	86	LIASAMSALSLPFLA	HLA-DQA1*05:01/DQB1*03:01	**13**
16	NTGMVASGALGVAIN	HLA-DRB1*01:01	145	87	PYAKVGVVGGMITGV	HLA-DQA1*05:01/DQB1*03:01	93
17	EEKYSINPLMASVGV	HLA-DRB1*01:01	134	88	YAKVGVVGGMITGVE	HLA-DQA1*05:01/DQB1*03:01	93
18	EKYSINPLMASVGVR	HLA-DRB1*01:01	134	89	AKVGVVGGMITGVES	HLA-DQA1*05:01/DQB1*03:01	93
19	GQPQLTVEQRAVRRI	HLA-DRB1*03:01	186	90	KVGVVGGMITGVEST	HLA-DQA1*05:01/DQB1*03:01	138
20	QPQLTVEQRAVRRIA	HLA-DRB1*03:01	186	91	VGVVGGMITGVESTR	HLA-DQA1*05:01/DQB1*03:01	138
21	PQLTVEQRAVRRIAW	HLA-DRB1*03:01	186	92	GVVGGMITGVESTRL	HLA-DQA1*05:01/DQB1*03:01	138
22	QLTVEQRAVRRIAWL	HLA-DRB1*03:01	186	93	SLITGIPFGGTLAAG	HLA-DQA1*05:01/DQB1*03:01	140
23	SDKITKIYSDIRQFA	HLA-DRB1*03:01	165	94	LITGIPFGGTLAAGM	HLA-DQA1*05:01/DQB1*03:01	140
24	DKITKIYSDIRQFAK	HLA-DRB1*03:01	165	95	ITGIPFGGTLAAGMT	HLA-DQA1*05:01/DQB1*03:01	140
25	KITKIYSDIRQFAKV	HLA-DRB1*03:01	165	96	TGIPFGGTLAAGMTI	HLA-DQA1*05:01/DQB1*03:01	140
26	ITKIYSDIRQFAKVA	HLA-DRB1*03:01	165	97	GIPFGGTLAAGMTIA	HLA-DQA1*05:01/DQB1*03:01	140
27	TKIYSDIRQFAKVAN	HLA-DRB1*03:01	165	98	IPFGGTLAAGMTIAP	HLA-DQA1*05:01/DQB1*03:01	140
28	GGMITGVESTRLDSA	HLA-DRB1*07:01	138	99	PFGGTLAAGMTIAPG	HLA-DQA1*05:01/DQB1*03:01	86
29	ADGKKHLSLITGIPF	HLA-DRB1*07:01	**213**	100	GGTLAAGMTIAPGFR	HLA-DQA1*05:01/DQB1*03:01	86
30	DGKKHLSLITGIPFG	HLA-DRB1*07:01	175	101	TDILPQAAAGQPQLT	HLA-DQA1*05:01/DQB1*03:01	199
31	GKKHLSLITGIPFGG	HLA-DRB1*07:01	140	102	DILPQAAAGQPQLTV	HLA-DQA1*05:01/DQB1*03:01	199
32	YSDIRQFAKVANIEV	HLA-DRB1*07:01	175	103	ILPQAAAGQPQLTVE	HLA-DQA1*05:01/DQB1*03:01	199
33	SDIRQFAKVANIEVP	HLA-DRB1*07:01	175	104	LPQAAAGQPQLTVEQ	HLA-DQA1*05:01/DQB1*03:01	199
34	DIRQFAKVANIEVPD	HLA-DRB1*07:01	175	105	PQAAAGQPQLTVEQR	HLA-DQA1*05:01/DQB1*03:01	199
35	IRQFAKVANIEVPDA	HLA-DRB1*07:01	175	106	QAAAGQPQLTVEQRA	HLA-DQA1*05:01/DQB1*03:01	199
36	KANTGMVASGALGVA	HLA-DRB1*07:01	145	107	VTAQEAAAAAAVRAL	HLA-DQA1*05:01/DQB1*03:01	126
37	TGMVASGALGVAINA	HLA-DRB1*07:01	60	108	TAQEAAAAAAVRALN	HLA-DQA1*05:01/DQB1*03:01	126
38	SINPLMASVGVRYNF	HLA-DRB1*07:01	134	109	AQEAAAAAAVRALNG	HLA-DQA1*05:01/DQB1*03:01	126
39	MKKIMLIASAMSALS	HLA-DRB1*08:02	**20**	110	QEAAAAAAVRALNGN	HLA-DQA1*05:01/DQB1*03:01	126
40	KIMLIASAMSALSLP	HLA-DRB1*09:01	**0**	111	EAAAAAAVRALNGNE	HLA-DQA1*05:01/DQB1*03:01	142
41	KANTGMVASGALGVA	HLA-DRB1*09:01	145	112	AAAAAAVRALNGNEQ	HLA-DQA1*05:01/DQB1*03:01	142
42	TGMVASGALGVAINA	HLA-DRB1*09:01	60	113	TESFSIYAGLGAGVA	HLA-DQA1*05:01/DQB1*03:01	138
43	KYSINPLMASVGVRY	HLA-DRB1*09:01	134	114	ESFSIYAGLGAGVAY	HLA-DQA1*05:01/DQB1*03:01	138
44	IQLYKDLVKLQRHAG	HLA-DRB1*11:01	161	115	SFSIYAGLGAGVAYT	HLA-DQA1*05:01/DQB1*03:01	138
45	QLYKDLVKLQRHAGI	HLA-DRB1*11:01	56	116	FSIYAGLGAGVAYTS	HLA-DQA1*05:01/DQB1*03:01	138
46	LYKDLVKLQRHAGIR	HLA-DRB1*11:01	113	117	SIYAGLGAGVAYTSG	HLA-DQA1*05:01/DQB1*03:01	138
47	YKDLVKLQRHAGIRK	HLA-DRB1*11:01	113	118	IYAGLGAGVAYTSGK	HLA-DQA1*05:01/DQB1*03:01	138
48	KDLVKLQRHAGIRKA	HLA-DRB1*11:01	113	119	YAGLGAGVAYTSGKI	HLA-DQA1*05:01/DQB1*03:01	138
49	DLVKLQRHAGIRKAM	HLA-DRB1*11:01	113	120	AGLGAGVAYTSGKID	HLA-DQA1*05:01/DQB1*03:01	138
50	DGKKHLSLITGIPFG	HLA-DRB1*12:01	175	121	GLGAGVAYTSGKIDG	HLA-DQA1*05:01/DQB1*03:01	168
51	GKKHLSLITGIPFGG	HLA-DRB1*12:01	140	122	LGAGVAYTSGKIDGV	HLA-DQA1*05:01/DQB1*03:01	168
52	KKHLSLITGIPFGGT	HLA-DRB1*12:01	140	123	VDIKANTGMVASGAL	HLA-DQA1*05:01/DQB1*03:01	145
53	GVMYLRNISAEVEVG	HLA-DRB1*13:02	180	124	DIKANTGMVASGALG	HLA-DQA1*05:01/DQB1*03:01	145
54	KIDGVDIKANTGMVA	HLA-DRB1*13:02	160	125	KANTGMVASGALGVA	HLA-DQA1*05:01/DQB1*03:01	145
55	ANTGMVASGALGVAI	HLA-DRB1*13:02	145	126	ANTGMVASGALGVAI	HLA-DQA1*05:01/DQB1*03:01	145
56	NTGMVASGALGVAIN	HLA-DRB1*13:02	145	127	NTGMVASGALGVAIN	HLA-DQA1*05:01/DQB1*03:01	145
57	SGALGVAINAAEGVY	HLA-DRB1*13:02	97	128	TGMVASGALGVAINA	HLA-DQA1*05:01/DQB1*03:01	60
58	GALGVAINAAEGVYV	HLA-DRB1*13:02	97	129	GMVASGALGVAINAA	HLA-DQA1*05:01/DQB1*03:01	60
59	ALGVAINAAEGVYVD	HLA-DRB1*13:02	97	130	MVASGALGVAINAAE	HLA-DQA1*05:01/DQB1*03:01	62
60	SINPLMASVGVRYNF	HLA-DRB1*13:02	134	131	VASGALGVAINAAEG	HLA-DQA1*05:01/DQB1*03:01	62
61	GKKHLSLITGIPFGG	HLA-DRB1*15:01	140	132	GGTLAAGMTIAPGFR	HLA-DQA1*01:02/DQB1*06:02	86
62	RAELGVMYLRNISAE	HLA-DRB1*15:01	130	133	GTLAAGMTIAPGFRA	HLA-DQA1*01:02/DQB1*06:02	86
63	NEQIIQLYKDLVKLQ	HLA-DRB1*15:01	161	134	TLAAGMTIAPGFRAE	HLA-DQA1*01:02/DQB1*06:02	86
64	EQIIQLYKDLVKLQR	HLA-DRB1*15:01	161	135	VTAQEAAAAAAVRAL	HLA-DQA1*01:02/DQB1*06:02	126
65	TESFSIYAGLGAGVA	HLA-DRB1*15:01	138	136	TAQEAAAAAAVRALN	HLA-DQA1*01:02/DQB1*06:02	126
66	VTPVKVLSDKITKIY	HLA-DRB3*01:01	165	137	AQEAAAAAAVRALNG	HLA-DQA1*01:02/DQB1*06:02	126
67	PVKVLSDKITKIYSD	HLA-DRB3*01:01	165	138	QEAAAAAAVRALNGN	HLA-DQA1*01:02/DQB1*06:02	126
68	AELGVMYLRNISAEV	HLA-DRB3*02:02	172	139	IKANTGMVASGALGV	HLA-DQA1*01:02/DQB1*06:02	145
69	ELGVMYLRNISAEVE	HLA-DRB3*02:02	172	140	ASGALGVAINAAEGV	HLA-DQA1*01:02/DQB1*06:02	62
70	LGVMYLRNISAEVEV	HLA-DRB3*02:02	180	141	KLYADLFTTESFSIY	HLA-DPA1*01:03/DPB1*04:01	152
71	GVMYLRNISAEVEVG	HLA-DRB3*02:02	180	142	ADLFTTESFSIYAGL	HLA-DPA1*01:03/DPB1*04:01	152

Table 3: List of protective immune response peptides including B cell Residues

S. No.	MHC class I	MHC Class II
1	VLSDKITKI (HLA-A*02:03)	VTPVK*VLSDKITKI*Y (HLA-DRB3*01:01)

		PVKVLSDKITKIYSD (HLA-DRB3*01:01)
2	YTSGKIDGV (HLA-A*02:06)	LGAGVA*YTSGKIDGV* (HLA-DQA1*05:01/DQB1*03:01)
3	YTSGKIDGV (HLA-A*68:02)	
4	KIYSDIRQF (HLA-A*32:01)	SDKIT*KIYSDIRQF*A (HLA-DRB1*03:01)
		DKIT*KIYSDIRQF*AK (HLA-DRB1*03:01)
		KIT*KIYSDIRQF*AKV (HLA-DRB1*03:01)
		IT*KIYSDIRQF*AKVA (HLA-DRB1*03:01)
5	MITGVESTR (HLA-A*68:01)	GG*MITGVESTR*LDSA (HLA-DRB1*07:01)
		VGVVGG*MITGVESTR* (HLA-DQA1*05:01/DQB1*03:01)
		GVVGG*MITGVESTR*L (HLA-DQA1*05:01/DQB1*03:01)
6	LMASVGVRY (HLA-B*15:01)	SINP*LMASVGVRY*NF (HLA-DRB1*07:01)
		KYSINP*LMASVGVRY* (HLA-DRB1*09:01)
		SINP*LMASVGVRY*NF (HLA-DRB1*13:02)
		KYSINP*LMASVGVRY* (HLA-DRB3*02:02)

Limitations:
This study has some limitations. We focused only on the 56 kDa antigen, rather than the whole proteome. Because entropy scores apply to individual positions, we assigned each BCE residue its own entropy score, but for MHC peptides (9–15 residues), we used the highest entropy value within each peptide. Finally, our predictions covered only 27 common MHC alleles, so some relevant epitopes could have been missed.

Conclusion:
We identified 6 immunogenic regions in 56 kDa protein that represent potential protective antigens against *O. tsutsugamushi*. We also identified several conserved as well as highly variable immune epitopes which can induce B cell, CD8+T cell and CD4+ T cells. Subject to experimental validation, these proposed protective antigen epitopes could be valuable for development of vaccine and serology or point of care test against *Orientia tsutsugamushi* infection.

References:
[1] Stephen S *et al*. *J Clin Diagn Res*. 2018. [DOI: 10.7860/JCDR/2018/38157.12271]
[2] Musa TH *et al*. *BioMed Res Int*. 2020 **2020**:1 [PMID: 33123578].
[3] Sonthayanon P *et al*. *J Clin Microbiol*. 2009 **47**:430. [PMID: 19091812]
[4] Kim MH *et al. Infect Chemother*. 2019 **51**:161. [PMID: 31270995]
[5] Ha NY *et al*. *PLoS Negl Trop Dis*. 2015 **9**:e0003585 [PMID: 25768004]
[6] Namsa ND *et al. Preprint.* [DOI: 10.21203/rs.3.rs-1486762/v1]
[7] Varghese GM *et al*. *Emerg Infect Dis*. 2015 **21**:64. [PMID: 25530231]
[8] Ramaiah A *et al*. *PLOS ONE*. 2018 **13**:e0196240. [PMID: 29698425]
[9] Nallan K *et al*. *Trop Med Infect Dis*. 2023 **8**:174. [PMID: 36977175]
[10] Sayers EW *et al*. *Nucleic Acids Res*. 2024 **52**:D33 [PMID: 37994677]
[11] Katoh K *el al. Nucleic Acids Res.* 2002 **30**:3059. [PMID: 12136088]
[12] Katoh K *et al*. *Brief Bioinform*. 2019 **20**:1160. [PMID: 28968734]
[13] Edgar RC. *Nucleic Acids Res*. 2004 **32**:1792. [PMID: 15034147]
[14] Edgar RC. *BMC Bioinformatics*. 2004 **5**:113. [PMID: 15318951]
[15] Darling AE *et al*. *PLoS ONE*. 2010 **5**:e11147. [PMID: 20593022]
[16] Vita R *et al*. *Nucleic Acids Res*. 2019 **47**:D339 [PMID: 30357391]
[17] Reynisson B *et al*. *Nucleic Acids Res*. 2020 **48**:W449. [PMID: 32406916]
[18] Jurtz V *et al*. *J Immunol*. 2017 **199**:3360. [PMID: 28978689]
[19] Nielsen M & Andreatta M. *Genome Med*. 2016:33. [PMID: 27029192]
[20] Hoof I *et al*. *Immunogenetics*. 2009 **61**:1. [PMID: 19002680]
[21] Nielsen M *et al*. *PLoS ONE*. 2007 **2**:e796. [PMID: 17726526]
[22] Andreatta M & Nielsen M. *Bioinformatics*. 2016 **32**:511. [PMID: 26515819]
[23] Lundegaard C *et al*. *Nucleic Acids Res*. 2008 **36**:W509. [PMID: 18463140]
[24] Lundegaard C *et al*. *Trends Biotechnol*. 2006 **24**:537 [PMID: 17045685.]
[25] Lundegaard C *et al*. *Bioinformatics*. 2008 **24**:1397 [PMID: 18413329]
[26] Nielsen M *et al*. *Protein Sci*. 2003 **12**:1007. [PMID: 12717023]
[27] Buus S *et al*. *Tissue Antigens*. 2003 **62**:378. [PMID: 14617044]
[28] Peters B & Sette A. *BMC Bioinformatics*. 2005 **6**:132. [PMID: 15927070]
[29] Kim Y *et al*. *BMC Bioinformatics*. 2009 **10**:394. [PMID: 19948066]
[30] Nilsson JB *et al*. *Sci Adv*. 2023 **9**:eadj6367. [PMID: 38000035]
[31] Nilsson JB *et al*. *Commun Biol*. 2023 **6**:442. [PMID: 37085710]
[32] Kaabinejadian S *et al*. *Front Immunol*. 2022 **13**:835454. [PMID: 35154160]
[33] Jensen KK *et al*. *Immunology*. 2018 **154**:394. [PMID: 29315598]
[34] Andreatta M *et al*. *Immunogenetics*. 2015 **67**:641. [PMID: 26416257]
[35] Nielsen M & Lund O. *BMC Bioinformatics*. 2009 **10**:296. [PMID: 19765293]
[36] Dolley A *et al*. *Heliyon*. 2024 **10**:e23616. [PMID: 38187223]

[37] Chen HW *et al*. *Infect Immun*. 2009 **77**:2311. [PMID: 19289508]
[38] Sanchez-Trincado JL *et al*. *J Immunol Res*. 2017 **2017**:1. [PMID: 29445754]
[39] Seong SY *et al*. *Microbes Infect*. 2001. [PMID: 11226850]
[40] Sankar S *et al*. *J Cell Biochem*. 2019 **120**:5869. [PMID: 30320912]
[41] Walker DH & Mendell NL. *Npj Vaccines*. 2023 **8**:11. [PMID: 36759505]
[42] Valbuena G & Walker DH. *Front Cell Infect Microbiol*. 2013 **2**:170 [PMID: 23316486].
[43] Hauptmann M *et al*. *PLoS Negl Trop Dis*. 2016 **10**:e0004991. [PMID: 27606708]