

GUTAIID: a curated database linking gut microbial antigens to autoimmune mechanisms

Laibah Hashmi¹, Shoaib Ur Rehman¹, Farzana Jabeen², Masood Ur Rehman Kayani^{1,*}

¹Metagenomics Discovery Lab, Department of Sciences, School of Interdisciplinary Engineering & Sciences (SINES), National University of Sciences & Technology (NUST), Srinagar Highway, Sector H-12, Islamabad 44000, Pakistan

²School of Electrical Engineering and Computer Science (SEECS), National University of Sciences & Technology (NUST), Srinagar Highway, Sector H-12, Islamabad 44000, Pakistan

*Corresponding author. Masood Ur Rehman Kayani, Metagenomics Discovery Lab, Department of Sciences, School of Interdisciplinary Engineering & Sciences (SINES), National University of Sciences & Technology (NUST), Srinagar Highway, Sector H-12, Islamabad 44000, Pakistan. E-mail: masood.kayani@gmail.com

Citation details: Hashmi, Laibah, Rehman, Shoaib Ur, *et al.* GUTAIID: a curated database linking gut microbial antigens to autoimmune mechanisms. Database (2026) Vol. 2026: article ID baag029; DOI: <https://doi.org/10.1093/database/baag029>

Abstract

Gut dysbiosis is widely recognized as a contributor to autoimmune diseases, as it can lead to the expression of microbial antigens that disrupt immune regulation through specific molecular mechanisms. However, existing resources do not systematically link gut microbial antigen sequences to the specific autoimmune mechanisms through which they act. Here, we present GUTAIID (Gut Microbes in Autoimmune Disorders), a literature-curated database of gut microbial antigens annotated with experimentally supported autoimmune mechanisms. Peer-reviewed studies published from October 1970 to September 2024 were manually screened, yielding 73 potential antigens that operate through nine molecular mechanisms, including protein citrullination, epitope spreading, molecular mimicry, and immune modulation, amongst others. The corresponding protein sequences were retrieved from UniProtKB, and redundancy was removed with MMseqs2. For the database implementation, data were delivered through a lightweight LAMP (Linux–Apache–MySQL/MariaDB–PHP) stack with server-side HTML/Bootstrap rendering, MySQL indexing, and HTTPS-secured downloads. Users can browse, keyword-search, or bulk-download sequence archives via a five-tab interface (Home, Downloads, Search, Team, and About). GUTAIID thus enables mechanism-oriented exploration of gut microbial antigens and supports downstream biomarker and therapeutic discovery in autoimmune research. **Database URL:** <https://gutaid.mgdiscoverylab.com/>

Introduction

Autoimmune diseases are chronic disorders in which the immune system starts attacking the body's own tissues, leading to persistent inflammation and organ damage. Collectively, these conditions affect approximately 3%–8% of the global population and have a higher prevalence in females. They also impose a substantial clinical and economic burden [1]. Clinical manifestations may involve a single organ, as in autoimmune hepatitis or vitiligo, or present as systemic diseases such as multiple sclerosis and type 1 diabetes. This diversity highlights the wide range of tissues that may be targeted [2]. In recent years, growing evidence has identified disruptions in the gut microbiome as a common environmental factor capable of disturbing immune homeostasis and contributing to the development of autoimmunity [3].

Recent studies have shown that gut–microbiome dysbiosis is a pathogenic cofactor in rheumatoid arthritis (RA). These studies report consistent shifts in microbial community structure and reduced taxonomic diversity in patients compared with healthy controls [4–6]. Mechanistically, dysbiosis promotes autoimmunity in several ways. First, impaired intestinal barrier function allows luminal antigens to enter the bloodstream, allowing access to mu-

cosal immune cells and triggering systemic autoreactivity [7, 8]. Second, increased antigenic overlap between microbial and host proteins leads to molecular mimicry and loss of immune tolerance [9, 10]. Third, an enrichment of microbes and viruses that amplify pro-inflammatory cytokine networks linked to cartilage and bone destruction [11, 12]. Together, these studies indicate that the gut microbiome acts as a modulatory layer that influences autoimmune risk factors, thereby influencing the onset and progression of RA and related autoimmune disorders.

It is also important to note how these altered microbial communities translate into immune dysfunction. In the literature curated for GUTAIID (Gut Microbes in Autoimmune Disorders), gut microbial antigens were linked to nine recurring autoimmune mechanism categories. These include molecular mimicry, superantigen activation, immune activation, epitope spreading, chronic infection, protein citrullination, immune modulation, inflammation-induced response, and immune system disruption. Molecular mimicry arises when microbial epitopes resemble host proteins, activating cross-reactive T- and B-cell clones that mistake self-antigens as targets. Persistent inflammation facilitates epitope spreading, in which newly exposed self-determinants become secondary immune targets [13, 14]. Superantigen activation can

Received: 25 November 2025. Revised: 27 April 2026. Accepted: 7 May 2026

© The Author(s) 2026. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

bypass normal peptide specificity by crosslinking MHC class II molecules with T-cell receptors, resulting in broad T-cell activation and abnormal helper signals that may promote autoreactivity [15, 16]. More broadly, sustained immune activation, chronic infection, and inflammation-induced responses can maintain prolonged antigen exposure, cytokine release, and tissue damage. This lowers the threshold for autoreactive lymphocyte activation and perpetuates autoimmune pathology [17–19]. Protein citrullination contributes by generating neoepitopes that can be recognized as non-self. Altered immune modulation or immune system disruption can weaken regulatory pathways that normally preserve immune homeostasis and self-tolerance [20–22]. Understanding how these microbial antigens exploit these mechanisms could guide biomarker discovery and inspire microbiome-targeted interventions for autoimmune diseases. This would pave the way for enhanced disease management.

To achieve this goal, researchers need a unified resource that connects gut antigens with their specific modes of immune disruption. Existing databases only partially meet this need. The Immune Epitope Database (IEDB) compiles pathogen, allergen, and self-epitopes but rarely specifies the microbial source or the mechanism underlying autoimmune involvement [23]. Microbe disease catalogues such as HMDAD and gutMDisorder provide taxon-level associations without sequence-level data [24, 25], whereas vaccine-oriented repositories (e.g. VIOLIN) prioritize protective antigens rather than those that break tolerance [26]. Prediction tools like VaxiJen offer *in silico* antigenicity scores with no experimental confirmation or mechanistic annotation [27]. Consequently, information about microbial antigens that act through molecular mimicry, epitope spreading, or bystander activation remains scattered across individual studies. GUTAIID addresses this gap by consolidating experimentally verified gut antigens and explicitly tagging each sequence with its reported autoimmune mechanism. It thereby provides a single, curated platform for mechanism-oriented research.

GUTAIID is the first literature-curated database dedicated to gut microbial antigenic sequences associated with well-defined autoimmune mechanisms. IEDB, HMDAD, and gutMDisorder focus on epitopes or taxon–disease associations rather than per-antigen mechanistic mapping. GUTAIID complements these resources by explicitly connecting each curated antigen to a defined autoimmune mechanism. By integrating mechanistic annotation with sequence-level data, GUTAIID offers three practical advantages. (i) It eliminates the laborious task of manually screening scattered literature sources to collect relevant information. (ii) It supplies a set of non-redundant, sequence-verified antigens that are ready for downstream bioinformatics and experimental workflows. (iii) It allows one-click FASTA downloads for immediate use without any need for editing. Collectively, these capabilities establish GUTAIID as a foundational reference for aiding mechanistic research on autoimmune disorders.

Materials and methods

Data collection

Peer-reviewed studies published from October 1970 to September 2024 were searched in PubMed and Google Scholar with the keywords gut microbiome, intestinal microbiota, autoimmune disease (including disease-specific names), and autoimmune mechanisms

(molecular mimicry, superantigen activation, immune activation, epitope spreading, chronic infection, protein citrullination, immune modulation, inflammation-induced response, immune system disruption). The search strategy prioritized studies describing autoimmune disease mechanisms in human subjects; however, studies employing validated animal models were also considered when they provided mechanistic insight relevant to autoimmune pathogenesis. All retrieved publications were curated manually to confirm that they (i) identified a bacterial antigen and (ii) provided evidence of its involvement in autoimmunity through a defined molecular mechanism. In addition to direct experimental validation, studies presenting biologically plausible indirect mechanistic evidence, such as molecular mimicry or immune activation pathways, were retained when supported by peer-reviewed literature. No conflicting evidence was identified among the selected studies, and multiple independent publications were used where available to support each curated antigen–mechanism relationship.

Manual curation produced a catalogue of 73 immunogenic antigens. For each antigen, details collected from the literature, such as the name, source organism, and reported mechanism of immune modulation, were recorded. Then the list of antigen names was provided to custom Perl 5.34 scripts that queried the UniProt REST API to download every associated reviewed and unreviewed protein entry for each antigen. These sequences constituted the core dataset for the downstream processing workflow and the database construction detailed in the following sections.

Data processing

During download, basic metadata—including entry count, minimum and maximum sequence length—were computed and stored. Sequence headers were then normalized to the convention <AntigenID>_<counter>, providing unique, traceable identifiers for downstream analyses. Redundant entries were collapsed with MMseqs2 v15.6f452 at a 100% sequence-identity threshold, yielding a non-redundant representative set that still preserved closely related antigen variants for downstream analyses. The curated FASTA files, per-antigen statistics, and cluster membership tables were consolidated into a comma-separated values (CSV) archive that underpins the GUTAIID web interface.

Database implementation

The GUTAIID web application was implemented on a conventional LAMP architecture (Linux, Apache, MySQL/MariaDB, PHP). Server-side functionality was developed in PHP 8.2 and deployed on a Linux host running Apache 2.4. Curated data were stored in a relational MariaDB 10.10.2 instance comprising two core tables, *mechanism* and *antigen*. Pages were rendered server-side with HTML5 and Bootstrap 3.0.2, while client-side interactivity used vanilla JavaScript and Chart.js for data visualization. Query performance was optimized with indexed fields and offset-based pagination (10 records per request), which adequately supported the current dataset without external caching or full-text search. Each antigen record was assigned a unique, auto-incremented identifier; all traffic was served over HTTPS. Administrative updates were executed via a secure CSV importer, and non-redundant FASTA archives were distributed from a dedicated/FASTA_files/directory. The codebase was maintained in a private Git repository and made

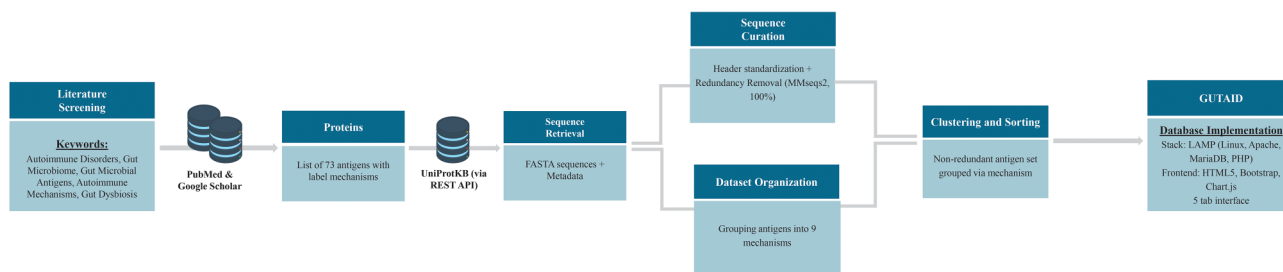


Figure 1 Framework for developing GUTAIID.

available for academic research on request. The framework of the database construction is illustrated in [Fig. 1](#) and [Supplementary Fig. 1](#).

Results

Database contents and usage

GUTAIID hosts 73 experimentally verified bacterial antigens, each assigned to one or more of nine literature-derived autoimmune mechanisms. [Table 1](#) presents the mechanism-wise distribution of these antigens and the corresponding sequence totals identified in this study. Every entry combines a non-redundant UniProt sequence cluster with mechanism tags and basic statistics, all accessible through a streamlined web interface. Users can browse, search, or download the full dataset via five dedicated tabs—Home, Downloads, Search, Team, and About.

The specific functions offered by each of these pages are detailed in the sections that follow.

Home

The Home page greets visitors with a brief mission banner—linking to the ‘About’ section for fuller context—followed by an icon grid of the nine key molecular mechanisms indexed in GUTAIID. Selecting any mechanism opens a dedicated view where antigens associated with that pathway are listed in a paginated table displaying their names, mechanism descriptors, sequence-length range, and cluster size, alongside a one-click FASTA download for each entry. A pie chart on the Home page summarizes the overall mechanism distribution, highlighting molecular mimicry as the most frequent category.

Downloads

The Downloads section provides direct access to the curated sequence sets. All nine molecular mechanisms are listed on the Downloads page; clicking any mechanism starts an instant download of a compressed FASTA file of all sequences associated with that particular pathway. A separate ‘Download All FASTA’ button packages the entire collection—covering every mechanism—into a single archive for users who need a complete export. This layout supports both focused and comprehensive data retrieval without requiring additional navigation.

Search

The Search page offers a complete, tabular view of every antigen in GUTAIID, displaying its mechanism label, sequence-length range, cluster size, and a direct FASTA download link. A case-insensitive query bar at the top of the search page supports keyword retrieval by antigen name, molecular mechanism, or a combination of both, instantly filtering the table to matching records. This layout enables rapid identification of any antigen in the database while preserving access to the same statistical details provided elsewhere in the database.

Team

The Team page lists the Principal Investigator, Research Scientist, and Database Developer from the Metagenomics Discovery Lab, National University of Sciences & Technology. Institutional affiliation and contact e-mail are provided for each team member, promoting transparency and enabling direct communication with the individuals responsible for GUTAIID’s conception, curation, and technical development.

About

The About page outlines the scientific rationale for GUTAIID, traces its origins in metagenomic research, and concisely explains the literature-driven and UniProt-based methodology used to curate gut microbial antigens. Collectively, these details furnish users with the context required to understand the breadth and scientific significance of the database.

The graphical user interface of GUTAIID is shown in [Fig. 2](#).

An example of usage

To demonstrate the usage of GUTAIID, we queried heat-shock protein 65 (Hsp65) from *Mycobacterium tuberculosis*, an antigen repeatedly implicated in autoimmune pathogenesis through molecular mimicry and cross-reactive anti-Hsp65 responses [28–30]. On the Search page, the user first selects Antigen as the query type and then enters any common synonym (‘HSP65’, ‘hsp65’, or ‘heat shock protein 65’). The case-insensitive filter instantly returns a single row summarizing the record: tagged mechanisms (molecular mimicry; immune response to bacterial heat-shock proteins), sequence-length span (63–541 aa), and a 401-member non-redundant cluster, together with a one-click Download FASTA link. This allows immediate retrieval of every curated Hsp65 sequence for downstream analysis, while the table itself provides a concise snapshot of the antigen’s immunological context.

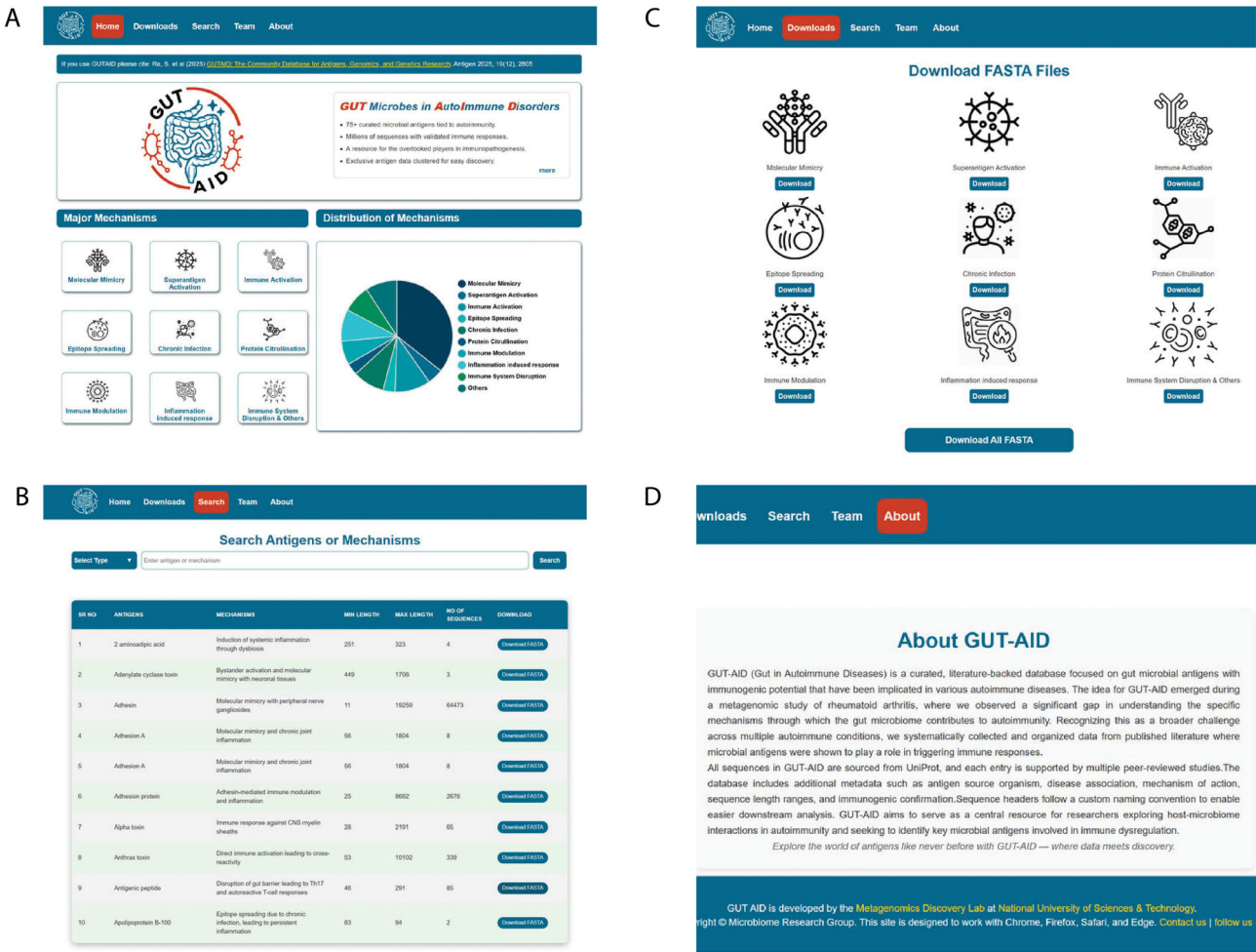


Figure 2 The graphical user interface of GUT-AID. (A) The Home page of GUT-AID. (B) The search page of GUT-AID. The user can search for antigens via antigen name or mechanism or both. (C) The Download page of GUT-AID allows the user to download antigen sequences by mechanism. (D) The About page of GUT-AID.

Table 1 Mechanism-wise summary of antigens and their corresponding total sequences across non-mutually exclusive, literature-derived autoimmune mechanism categories.^a

Major mechanism	Number of antigens	Percentage of total (73)	Total sequence
Molecular mimicry	34	46.58%	1 194 149
Superantigen activation	4	5.48%	3 652
Immune activation	25	34.25%	708 105
Epitope spreading	3	4.11%	4 052
Chronic infection	5	6.85%	782 480
Protein citrullination	3	4.11%	10 874
Immune modulation	11	15.07%	559 101
Inflammation induced response	15	20.55%	603 842
Immune system disruption and others	3	4.11%	287 807

^aSequence counts represent post-clustering totals at 100% sequence similarity

The same interface supports pathway-based exploration. Switching the query type to mechanism and entering ‘molecular mimicry’ repopulates the table with all antigens assigned to that category, each accompanied by its own statistics and individual download button. For users who prefer a bulk export, the Downloads page offers a single ‘molecu-

lar mimicry’ tile that delivers a compressed FASTA archive encompassing every mimicry-associated antigen in GUT-AID. Together, these two search options, single-antigen queries and mechanism-level downloads, show how GUT-AID simplifies access to curated sequences for research on microbe-driven autoimmunity.

Conclusion

GUTAIID is the first specialized repository that unites literature-verified gut microbial antigens with explicit autoimmune-mechanism tags and non-redundant UniProt sequences. Its lightweight LAMP infrastructure and streamlined download options give researchers rapid, mechanism-aware access to 73 antigens spanning nine molecular mechanisms, providing an immediate resource for dissecting host–microbiome interactions in autoimmunity.

The current version of GUTAIID should be considered the first release of an expanding reference resource. The present version reflects a deliberately conservative curation strategy based primarily on PubMed and Google Scholar literature and restricted to bacterial antigens with experimentally supported links to autoimmune disease. Accordingly, the current coverage may be influenced by literature availability and reporting bias, and fungal or viral antigens were not included in this initial release. All sequence sets are presently represented at a 100% sequence-identity clustering threshold, while future versions will also provide additional sequence collections at multiple clustering levels. Future work will expand both content and capacity. Scheduled literature searches and UniProt synchronizations will add newly reported antigens, while back-end upgrades—caching, full-text indexing, and a public MIT-licensed GitHub release—will improve query performance, scalability, and community involvement. These developments will ensure that GUTAIID remains a robust, up-to-date platform for advancing microbe-mediated autoimmunity research.

Acknowledgements

The authors are highly thankful to members of the Metagenomics Discovery Lab at the SINES, NUST, Islamabad, Pakistan, for their useful discussion and feedback on our manuscript.

Author contributions

L.H. performed the literature search, curation, and sequence retrieval, the construction of the sequence catalogue, and prepared the manuscript. S.U.R. performed the database construction and developed the GUI for the database. M.U.R.K. conceived the idea, supervised the research, and provided resources for the project. F.J. co-supervised the research. All authors have read and agreed to the submission of the manuscript.

Supplementary material

Supplementary material is available at [Database](#) online.

Conflicts of interest

The authors declare no competing interests.

Funding

No dedicated research funding was received for this project.

Data availability

The publicly retrieved gene sequences and relevant details are available from the following URL: <https://gutaidd.mgdiscoverylab.com/>. Supplementary material, including a complete list of bacterial antigens, their mechanism, and supporting references, and a list of bacteria reported in autoimmune diseases are shared as [Supplementary Tables 1 and 2](#), respectively. This data has also been shared on Zenodo: <https://doi.org/10.5281/zenodo.19661956>

References

- 1 Yamamoto K. Mechanisms of autoimmunity. *Autoimmune Dis* 2004;47:403–406.
- 2 Sudres M, Verdier J, Truffault F *et al*. Pathophysiological mechanisms of autoimmunity. *Ann NY Acad Sci* 2018;1413:59–68. <https://doi.org/10.1111/nyas.13560>
- 3 Proal AD, Marshall TG. Autoimmune disease in the era of the metagenome. *Autoimmun Rev* 2009;8:677–81. <https://doi.org/10.1016/j.autrev.2009.02.016>
- 4 Koh JH, Lee EH, Cha KH *et al*. Factors associated with the composition of the gut microbiome in patients with established rheumatoid arthritis and its value for predicting treatment responses. *Arthritis Res Ther* 2023;25:32. <https://doi.org/10.1186/s13075-023-03013-x>
- 5 Kishikawa T, Maeda Y, Nii T *et al*. Metagenome-wide association study of gut microbiome revealed novel aetiology of rheumatoid arthritis in the Japanese population. *Ann Rheum Dis* 2020;79:103–11. <https://doi.org/10.1136/annrheumdis-2019-215743>
- 6 Chu XJ, Cao NW, Zhou HY *et al*. The oral and gut microbiome in rheumatoid arthritis patients: a systematic review. *Rheumatology* 2021;60:1054–66. <https://doi.org/10.1093/rheumatology/keaa835>
- 7 Luca F, Shoenfeld Y. The microbiome in autoimmune diseases. *Clin Exp Immunol* 2019;195:74–85. <https://doi.org/10.1111/cei.13158>
- 8 Zhang Y, Li L, Guo C *et al*. The role of intestinal mucosal barrier in autoimmune diseases: a comprehensive review. *Front Immunol* 2022;13:871713.
- 9 Cusick MF, Libbey JE, Fujinami RS. Molecular mimicry as a mechanism of autoimmune disease. *Clin Rev Allergy Immunol* 2012;42:102–11. <https://doi.org/10.1007/s12016-011-8294-7>
- 10 Zhao T, Wei Y, Zhu Y *et al*. Gut microbiota and rheumatoid arthritis: from pathogenesis to novel therapeutic opportunities. *Front Immunol* 2022;13:1007165. <https://doi.org/10.3389/fimmu.2022.1007165>
- 11 Brennan FM, McInnes IB. Evidence that cytokines play a role in rheumatoid arthritis. *J Clin Invest* 2008;118:3537–45. <https://doi.org/10.1172/JCI36389>
- 12 Guo R, Li S, Zhang Y *et al*. Dysbiotic oral and gut viromes in untreated and treated rheumatoid arthritis patients. *Microbiol Spectr* 2022;10:00348–22. <https://doi.org/10.1128/spectrum.00348-22>
- 13 Oldstone MBA. Molecular mimicry as a mechanism for the cause and the prevention of autoimmune disease. *Curr Opin Microbiol* 2005;8:359–63.
- 14 Miller SD, Turley DM, Podojil JR. Antigen-specific tolerance strategies for the prevention and treatment of autoimmune

- disease. *Nat Rev Immunol* 2007;7:665–77. <https://doi.org/10.1038/nri2153>
- 15 Tuffs SW, Dufresne K, Rishi A *et al.* Novel insights into the immune response to bacterial T cell superantigens. *Nat Rev Immunol* 2024;24:417–34. <https://doi.org/10.1038/s41577-023-00979-2>
 - 16 Friedman SM, Posnett DN, Tumang JR *et al.* A potential role for microbial superantigens in the pathogenesis of systemic autoimmune disease. *Arthritis Rheum* 1991;34:468–80. <https://doi.org/10.1002/art.1780340412>
 - 17 Fujinami RS, Herrath MG, Christen U *et al.* Molecular mimicry, bystander activation, or viral persistence: infections and autoimmune disease. *Clin Microbiol Rev* 2006;19:80–94. <https://doi.org/10.1128/CMR.19.1.80-94.2006>
 - 18 Rivera-Correa J, Rodriguez A. Divergent roles of antiself antibodies during infection. *Trends Immunol* 2018;39:515–22. <https://doi.org/10.1016/j.it.2018.04.003>
 - 19 Xiang Y, Zhang M, Jiang D *et al.* The role of inflammation in autoimmune disease: a therapeutic target. *Front Immunol* 2023;14:1267091. <https://doi.org/10.3389/fimmu.2023.1267091>
 - 20 Sakaguchi S, Yamaguchi T, Nomura T *et al.* Regulatory T cells and immune tolerance. *Cell* 2008;133:775–87. <https://doi.org/10.1016/j.cell.2008.05.009>
 - 21 Ciesielski O, Biesiekierska M, Panthou B *et al.* Citrullination in the pathology of inflammatory and autoimmune disorders: recent advances and future perspectives. *Cell Mol Life Sci* 2022;79:94. <https://doi.org/10.1007/s00018-022-04126-3>
 - 22 Rocamora-Reverte L, Melzer FL, Würzner R *et al.* The complex role of regulatory T cells in immunity and aging. *Front Immunol* 2020;11:616949. <https://doi.org/10.3389/fimmu.2020.616949>
 - 23 Vita R, Mahajan S, Overton JA. The Immune Epitope Database (IEDB): 2019 update. *Nucleic Acids Res* 2019;47.
 - 24 Ma L, Chen H, Yu B. HMDAD: a human microbe–disease association database. *Briefings Bioinf* 2017;18:85–97. <https://doi.org/10.1093/bib/bbw005>
 - 25 Li J, Ding J, Li H. gutMDisorder v2.0: a comprehensive database for dysbiosis of gut microbiota in disorders and interventions. *Nucleic Acids Res* 2022;51:D717–D722. <https://doi.org/10.1093/nar/gkac871>
 - 26 He Y, Xiang Z, Mobley HLT. VIOLIN: vaccine Investigation and Online Information Network. *Nucleic Acids Res* 2008;36:D923–D928. <https://doi.org/10.1093/nar/gkm1039>
 - 27 Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf* 2007;8:doi:10.1186/1471-2105-8-4
 - 28 Dow CTM. Paratuberculosis heat shock protein 65 and human diseases: bridging infection and autoimmunity. *Autoimmune Dis* 2012;2012:150824.
 - 29 Singh I, Yadav AR, Mohanty KK *et al.* Molecular mimicry between HSP 65 of Mycobacterium leprae and cytokeratin 10 of the host keratin; role in pathogenesis of leprosy. *Cell Immunol* 2012;278:63–75. <https://doi.org/10.1016/j.cellimm.2012.06.011>
 - 30 Rajaiah R, Moudgil KD. Heat-shock proteins can promote as well as regulate autoimmunity. *Autoimmun Rev* 2009;8:388–93. <https://doi.org/10.1016/j.autrev.2008.12.004>