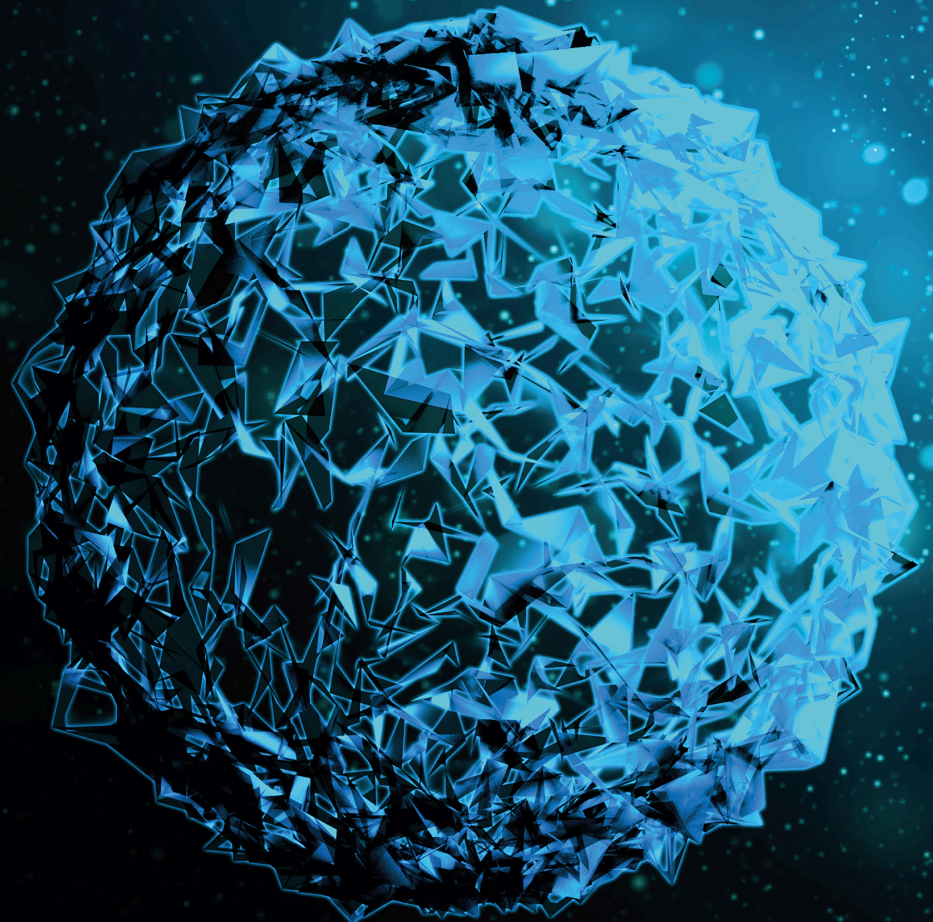


Adaptive Evolution of Autoimmune Proteins in Animals

Lead Guest Editor: Hafiz Ishfaq Ahmad

Guest Editors: Jinping Chen and Borhan Shokrollahi





Adaptive Evolution of Autoimmune Proteins in Animals

BioMed Research International

Adaptive Evolution of Autoimmune Proteins in Animals

Lead Guest Editor: Hafiz Ishfaq Ahmad
Guest Editors: Jinping Chen and Borhan
Shokrollahi



Copyright © 2024 Hindawi Limited. All rights reserved.

This is a special issue published in "BioMed Research International." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Section Editors

Penny A. Asbell, USA
David Bernardo , Spain
Gerald Brandacher, USA
Kim Bridle , Australia
Laura Chronopoulou , Italy
Gerald A. Colvin , USA
Aaron S. Dumont, USA
Pierfrancesco Franco , Italy
Raj P. Kandpal , USA
Fabrizio Montecucco , Italy
Mangesh S. Pednekar , India
Letterio S. Politi , USA
Jinsong Ren , China
William B. Rodgers, USA
Harry W. Schroeder , USA
Andrea Scribante , Italy
Germán Vicente-Rodríguez , Spain
Momiao Xiong , USA
Hui Zhang , China

Academic Editors

Immunology

Contents

Retracted: Status of Bioinformatics Education in South Asia: Past and Present

BioMed Research International

Retraction (1 page), Article ID 9842042, Volume 2024 (2024)

Retracted: Exploring the Potential of Interferon Gamma Gene as Major Immune Responder for Bovine Tuberculosis in River Buffalo

BioMed Research International

Retraction (1 page), Article ID 9840801, Volume 2024 (2024)

Retracted: Epi-Gene: An R-Package for Easy Pan-Genome Analysis

BioMed Research International

Retraction (1 page), Article ID 9830450, Volume 2024 (2024)

Retracted: *In Silico* Characterization of Growth Differentiation Factors as Inhibitors of TNF-Alpha and IL-6 in Immune-Mediated Inflammatory Disease Rheumatoid Arthritis

BioMed Research International

Retraction (1 page), Article ID 9824581, Volume 2024 (2024)

Retracted: Computer-Aided Prediction and Identification of Phytochemicals as Potential Drug Candidates against MERS-CoV

BioMed Research International

Retraction (1 page), Article ID 9815474, Volume 2024 (2024)

Retracted: Identification of Hearing Loss-Associated Variants of *PTPRQ*, *MYO15A*, and *SERPINB6* in Pakistani Families

BioMed Research International




Retraction (1 page), Article ID 9784048, Volume 2024 (2024)

Retracted: Identification of Peptides as Novel Inhibitors to Target IFN- γ , IL-3, and TNF- α in Systemic Lupus Erythematosus

BioMed Research International

Retraction (1 page), Article ID 9756587, Volume 2024 (2024)



[Retracted] Identification of Peptides as Novel Inhibitors to Target IFN- γ , IL-3, and TNF- α in Systemic Lupus Erythematosus

Ghulam Mustafa , Hafiza Salaha Mahrosh , Mahwish Salman, Sumaira Sharif , Raheela Jabeen, Tanveer Majeed, and Hafsah Tahir

Research Article (11 pages), Article ID 1124055, Volume 2021 (2021)






[Retracted] Epi-Gene: An R-Package for Easy Pan-Genome Analysis

Furqan Awan , Muhammad Muddassir Ali, Muhammad Hamid , Muhammad Huzair Awan,



Muhammad Hassan Mushtaq, Saeeda Kalsoom, Muhammad Ijaz, Khalid Mehmood , and Yongjie Liu 

Research Article (8 pages), Article ID 5585586, Volume 2021 (2021)



Structural and Evolutionary Adaptation of NOD-Like Receptors in Birds

Xueting Ma , Baohong Liu , Zhenxing Gong , Xinmao Yu , and Jianping Cai 
Research Article (11 pages), Article ID 5546170, Volume 2021 (2021)



[Retracted] Identification of Hearing Loss-Associated Variants of *PTPRQ*, *MYO15A*, and *SERPIN6* in Pakistani Families

Umair Mahmood, Shazia A. Bukhari , Muhammad Ali, Zubair M. Ahmed, and Saima Riazuddin 
Research Article (6 pages), Article ID 5584788, Volume 2021 (2021)




[Retracted] Status of Bioinformatics Education in South Asia: Past and Present

Muhammad Muddassir Ali, Muhammad Hamid , Muhammad Saleem, Saadia Malik, Natash Ali Mian, Muhammad Ahmed Ihsan , Nadia Tabassum, Khalid Mehmood, and Furqan Awan
Review Article (9 pages), Article ID 5568262, Volume 2021 (2021)





[Retracted] Computer-Aided Prediction and Identification of Phytochemicals as Potential Drug Candidates against MERS-CoV

Hafiza Salaha Mahrosh , Muhammad Tanveer , Rawaba Arif , and Ghulam Mustafa 
Research Article (7 pages), Article ID 5578689, Volume 2021 (2021)




Sequence and Structural Characterization of Toll-Like Receptor 6 from Human and Related Species

Ghulam Mustafa , Hafiza Salaha Mahrosh , and Rawaba Arif 
Research Article (9 pages), Article ID 5545183, Volume 2021 (2021)


Molecular Characterization of MHC Class I Genes in Four Species of the *Turdidae* Family to Assess Genetic Diversity and Selection

Muhammad Usman Ghani , Li Bo , An Buyang , Xu Yanchun , Shakeel Hussain, and Muhammad Yasir 
Research Article (14 pages), Article ID 5585687, Volume 2021 (2021)

[Retracted] Exploring the Potential of Interferon Gamma Gene as Major Immune Responder for Bovine Tuberculosis in River Buffalo

Maryam Javed , Syed Ahmed Raza , Asif Nadeem, Muhammad Muddassir Ali, Wasim Shehzad, and Khalid Mehmood 
Research Article (7 pages), Article ID 5532864, Volume 2021 (2021)

[Retracted] *In Silico* Characterization of Growth Differentiation Factors as Inhibitors of TNF-Alpha and IL-6 in Immune-Mediated Inflammatory Disease Rheumatoid Arthritis

Ghulam Mustafa , Hafiza Salaha Mahrosh, and Rawaba Arif
Research Article (11 pages), Article ID 5538535, Volume 2021 (2021)

Retraction

Retracted: Status of Bioinformatics Education in South Asia: Past and Present

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] M. M. Ali, M. Hamid, M. Saleem et al., "Status of Bioinformatics Education in South Asia: Past and Present," *BioMed Research International*, vol. 2021, Article ID 5568262, 9 pages, 2021.

Retraction

Retracted: Exploring the Potential of Interferon Gamma Gene as Major Immune Responder for Bovine Tuberculosis in River Buffalo

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named

external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] M. Javed, S. A. Raza, A. Nadeem, M. M. Ali, W. Shehzad, and K. Mehmood, "Exploring the Potential of Interferon Gamma Gene as Major Immune Responder for Bovine Tuberculosis in River Buffalo," *BioMed Research International*, vol. 2021, Article ID 5532864, 7 pages, 2021.

Retraction

Retracted: Epi-Gene: An R-Package for Easy Pan-Genome Analysis

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] F. Awan, M. M. Ali, M. Hamid et al., "Epi-Gene: An R-Package for Easy Pan-Genome Analysis," *BioMed Research International*, vol. 2021, Article ID 5585586, 8 pages, 2021.

Retraction

Retracted: *In Silico* Characterization of Growth Differentiation Factors as Inhibitors of TNF-Alpha and IL-6 in Immune-Mediated Inflammatory Disease Rheumatoid Arthritis

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named

external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] G. Mustafa, H. S. Mahrosh, and R. Arif, "*In Silico* Characterization of Growth Differentiation Factors as Inhibitors of TNF-Alpha and IL-6 in Immune-Mediated Inflammatory Disease Rheumatoid Arthritis," *BioMed Research International*, vol. 2021, Article ID 5538535, 11 pages, 2021.

Retraction

Retracted: Computer-Aided Prediction and Identification of Phytochemicals as Potential Drug Candidates against MERS-CoV

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] H. S. Mahrosh, M. Tanveer, R. Arif, and G. Mustafa, "Computer-Aided Prediction and Identification of Phytochemicals as Potential Drug Candidates against MERS-CoV," *BioMed Research International*, vol. 2021, Article ID 5578689, 7 pages, 2021.

Retraction

Retracted: Identification of Hearing Loss-Associated Variants of *PTPRQ*, *MYO15A*, and *SERPINB6* in Pakistani Families

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] U. Mahmood, S. A. Bukhari, M. Ali, Z. M. Ahmed, and S. Riazuddin, "Identification of Hearing Loss-Associated Variants of *PTPRQ*, *MYO15A*, and *SERPINB6* in Pakistani Families," *BioMed Research International*, vol. 2021, Article ID 5584788, 6 pages, 2021.

Retraction

Retracted: Identification of Peptides as Novel Inhibitors to Target IFN- γ , IL-3, and TNF- α in Systemic Lupus Erythematosus

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] G. Mustafa, H. S. Mahrosh, M. Salman et al., "Identification of Peptides as Novel Inhibitors to Target IFN- γ , IL-3, and TNF- α in Systemic Lupus Erythematosus," *BioMed Research International*, vol. 2021, Article ID 1124055, 11 pages, 2021.

Retraction

Retracted: Identification of Peptides as Novel Inhibitors to Target IFN- γ , IL-3, and TNF- α in Systemic Lupus Erythematosus

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] G. Mustafa, H. S. Mahrosh, M. Salman et al., "Identification of Peptides as Novel Inhibitors to Target IFN- γ , IL-3, and TNF- α in Systemic Lupus Erythematosus," *BioMed Research International*, vol. 2021, Article ID 1124055, 11 pages, 2021.

Research Article

Identification of Peptides as Novel Inhibitors to Target IFN- γ , IL-3, and TNF- α in Systemic Lupus Erythematosus

Ghulam Mustafa ¹, Hafiza Salaha Mahrosh ¹, Mahwish Salman,¹ Sumaira Sharif ², Raheela Jabeen,³ Tanveer Majeed,⁴ and Hafsah Tahir⁵

¹Department of Biochemistry, Government College University, Faisalabad, 38000, Pakistan

²Institute of Molecular Biology and Biotechnology, The University of Lahore, Pakistan

³Department of Biochemistry and Biotechnology, The Women University Multan, Pakistan

⁴Department of Biotechnology, Kinnaird College for Women, Lahore, Pakistan

⁵Department of Environmental Sciences, Quaid-i-Azam University, Islamabad, Pakistan

Correspondence should be addressed to Ghulam Mustafa; gmustafa_uaf@yahoo.com

Received 28 May 2021; Accepted 29 October 2021; Published 13 November 2021

Academic Editor: Arif Siddiqui

Copyright © 2021 Ghulam Mustafa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Autoimmune disorder is a chronic immune imbalance which is developed through a series of pathways. The defect in B cells, T cells, and lack of self-tolerance has been greatly associated with the onset of many types of autoimmune complications including rheumatoid arthritis, systemic lupus erythematosus (SLE), multiple sclerosis, and chronic inflammatory demyelinating polyneuropathy. The SLE is an autoimmune disease with a common type of lupus that causes tissue and organ damage due to the wide spread of inflammation. In the current study, twenty anti-inflammatory peptides derived from plant and animal sources were docked as ligands or peptides counter to proinflammatory cytokines. Interferon gamma (IFN- γ), interleukin 3 (IL-3), and tumor necrosis factor alpha (TNF- α) were targeted in this study as these are involved in the pathogenesis of SLE in many clinical studies. Two docking approaches (i.e., protein-ligand docking and peptide-protein docking) were employed in this study using Molecular Operating Environment (MOE) software and HADDOCK web server, respectively. Amongst docked twenty peptides, the peptide DEDTQAMMPFR with *S*-score of -11.3018 and HADDOCK score of -10.3 ± 2.5 kcal/mol showed the best binding interactions and energy validation with active amino acids of IFN- γ protein in both docking approaches. Depending upon these results, this peptide could be used as a potential drug candidate to target IFN- γ , IL-3, and TNF- α proteins to control inflammatory events. Other peptides (i.e., QEPQESQQ and FRDEHKK) also revealed good binding affinity with IFN- γ with *S*-scores of -10.98 and -10.55, respectively. Similarly, the peptides KHDRGDEF, FRDEHKK, and QEPQESQQ showed best binding interactions with IL-3 with *S*-scores of -8.81, -8.64, and -8.17, respectively.

1. Introduction

The innate and adaptive immune system controls the defense organization mediated by multiple components and molecules in an organism [1]. Different organs, signaling pathways, and compounds collectively perform various tasks to protect the organism from external and internal damage. Autoimmunity or production of autoantigens mainly brings disaster to the immune system due to inadequate immune tolerance [2]. The defect in the synergistic relationship between innate immunity and adaptive immu-

nity causes severe consequences including autoimmune disorders, inflammations like systemic lupus erythematosus, rheumatoid arthritis, Alzheimer's disease, multiple sclerosis, and many other complications [3].

Systemic lupus erythematosus (SLE) is the most common chronic autoimmune inflammatory disorder characterized by the presence of autoantibodies directed against own cells or tissues of the body. It is intermediated by B cells which generate autoantibodies against nuclear antigens, a type III hypersensitivity reaction that causes chronic systemic inflammation and tissue damage in the joints, skin,

brain, lungs, kidneys, and blood vessels [4]. The incidence of SLE prevalence has been predominantly recorded in young middle-aged females. According to studies, the highest prevalence can be seen in certain ethnicities, reflected in prevalence rates of approx. 40/100,000 persons in Northern European cohorts with comparison rates of 200/100,000 patients of African-American descent [5]. SLE is multifactorial in its origin with a wide range of clinical and serological manifestations. There have been many efforts to elucidate the pathogenesis of SLE with current recognition of genetic susceptibility, environmental triggers, and disruption in both the innate and adaptive immune systems [6].

Adaptive immunity is an antigen-specific host defense that comprises of B and T lymphocytes and immunoglobulins. The defected immunity in SLE results in a myriad of complications such as decreased T cell signaling, stimulation of autoaggressive T effector cells, and production of autoantibodies. Apart from all these, any dysregulation in B cells that produces proinflammatory cytokines (i.e., IL-1, IL-3, IL-6, IL-23, and TNF- α) pushes the inflammatory events and drives the SLE disease. Thus, targeting these B and T cells could be proved as a therapeutic advantage to control SLE [7].

The role of cytokines regarding SLE has collected much interest of scientists. Type I interferon family along with many other cytokines such as interleukins (IL-3, IL-6, IL-10, and IL-17) and tumor necrosis factor (TNF) is seen to be involved in SLE. These interleukins play a significant role in diseases which are linked to inflammation and autoimmunity [4]. This disease has several variants which are genetically linked with pathogenic mechanisms. Genetics and epigenetics are the factors which contribute directly to cause alterations in the cells of both innate and adaptive immune responses [8].

Greater than 95% of SLE patients have detectable serum antinuclear antibody (ANA). Anti-ds-DNA antibodies are highly specific for SLE and present in 65–70% of the patients (versus 0.5% of the healthy population). Anti-ds-DNA antibodies, anti-Ro, anti-La, anti-C1q, and anti-Sm antibodies have been demonstrated histologically in renal biopsy specimens. A number of specific antibodies have been associated with a particular expression of SLE. Anti-Ro and antinucleosome antibodies are most strongly linked with cutaneous lupus [9]. Many combination drug therapies have been in practice for the control of clinical manifestations of SLE, but these have been recorded with severe side effects, and also, the patients of SLE still show a higher standardized mortality rate (i.e., of 4.6-fold) with respect to the general population [10]. Therefore, the current study was planned to reveal anti-inflammatory peptides from plant and animal sources using molecular docking approach. The study includes protein-ligand and peptide-protein docking of twenty anti-inflammatory peptides counter to IFN- γ , IL-3, and TNF- α receptor proteins as targets for the treatment of SLE. In the current study, we have investigated the potential of these peptides as drug candidates to attenuate the inflammatory response and tissue destruction due to activation of proinflammatory cytokines, B cells, and T cells which lead towards autoantibody production.

2. Materials and Methods

The study includes the protein-ligand and peptide-protein docking of twenty anti-inflammatory bioactive peptides against three main receptor proteins (i.e., IL-3, TNF- α , and IFN- γ) that play leading roles in the pathogenesis of SLE. Molecular Operating Environment v.2015.10 (Chemical Computing Group ULC, Montreal, QC, Canada) [11] was employed for ligand-based docking, and HADDOCK v.2.4 [12] an online server was used for peptide-protein docking.

2.1. Ligand Database Preparation. An extensive literature survey was performed to explore plant and animal derived bioactive anti-inflammatory peptides. The chemical structures of these ligands were prepared using ACD/ChemSketch v.C40E41 (Advanced Chemistry Development, Inc., Toronto, Ontario, Canada) [13] and saved in MOL format in MOE database as ready-to-dock compounds after energy minimization.

2.2. Refinement of Receptor Proteins. The three-dimensional structures of IL-3 (PDB ID: 5UV8), IFN- γ (PDB ID: 6E3K), and TNF- α (PDB ID: 6OP0) were retrieved from protein data bank (<https://www.rcsb.org/>). Removal of solvent, addition of hydrogen atoms, energy minimization, and 3D protonation were performed using MOE with default parameters, and the minimized structure of each protein was saved to use as receptor protein for docking studies.

2.3. Ligand-Protein Docking. The active binding pocket of each receptor protein was selected using the site finder tool of MOE. The prepared ready-to-dock library of twenty anti-inflammatory peptides was docked counter to IL-3, IFN- γ , and TNF- α , and the top three ligands from each protein-ligand docking were selected on the basis of their interactions and S-scores. The algorithm of MOE provides top conformations on the basis of binding patterns of ligands with active amino acids of the receptor protein, minimum energy structure, and maximum occupancy of the binding pocket.

2.4. Peptide-Protein Docking. IL-3, IFN- γ , and TNF- α have been extensively studied due to their roles in the pathophysiology of many autoimmune disorders. The dysfunctioning of the signaling pathways of B and T cells leads towards the production of autoantibodies against the body's own tissues and cells. In the literature, many anti-inflammatory compounds have been reported for autoimmune disease; however, in this study, for the first time, peptide-protein docking was employed to explore the inhibitory effects of peptides counter to selected receptor proteins for the treatment of SLE. The most reported twenty anti-inflammatory peptides were used for docking studies against receptor proteins. For peptide-protein docking, the top peptides against each receptor protein were selected, which were obtained from the results of protein-ligand docking and used for further analysis of peptide-protein docking. The sequences of these peptides were retrieved from the literature and subjected to BLASTp to find their homologs. To predict three-

dimensional (3D) structure of each peptide, the PDB database was used during BLAST analyses to find the best templates of each peptide using the homology modeling approach. Modeller v.9.21 (Ben Webb, UCSF, CA, USA) [14] was used to predict 3D structures of selected peptides (i.e., DEDTQAMMPFR and KHDRGDEF).

HADDOCK v.2.4 was used to carry out the docking analysis between the best selected peptides and selected target proteins. Educational version of PYMOL Molecular Graphics System v2.0 (Schrödinger, LLC) was used to visualize the docked complexes and draw figures [15].

2.4.1. Molecular Dynamic Simulation. For 120 nanoseconds, Desmond (Schrödinger LLC) was used to model molecular dynamics in triplicates [16]. The earliest phase of protein-ligand complex for molecular dynamics simulation was used in the docking experiments. Molecular docking studies can predict ligand binding state in static situations. Docking is useful because it provides a static view of a molecule's binding pose at the active site of a protein [17]. By integrating Newton's classical equation of motion, MD simulations typically compute atom movements over time. Simulations were used to predict the ligand binding status in the physiological environment [18, 19].

The protein-ligand complex was preprocessed using Protein Preparation Wizard or Maestro, which included complex optimization and minimization. All of the systems were prepared using the System Builder tool. TIP3P (transferable intermolecular interaction potential 3 points), a solvent model with an orthorhombic box was chosen. In the simulation, the OPLS 2005 force field was used [20]. To make models neutral, counter ions were introduced. To mimic physiological conditions, 0.15 M NaCl was added. The NPT ensemble with 300 K temperature and 1 atm pressure was chosen for the entire simulation. The models were relaxed before the simulation. The trajectories were saved for examination after every 100 ps, and the simulation's stability was verified by comparing the root mean square deviation (RMSD) values of protein and ligand over time.

3. Results

3.1. Protein-Ligand Docking. Molecular docking predicts the intermolecular framework and binding interactions between ligand molecules and proteins. Different docking approaches play a leading role in drug discovery. This study includes the protein-ligand docking and peptide-protein docking of twenty anti-inflammatory peptides counter to IFN- γ , IL-3, and TNF-alpha. The top three hit peptides from each analysis were selected on the basis of energy structure and interactions with active amino acids of the respective receptor protein (Table 1).

3.2. Interaction Analyses. Amongst twenty peptides as ligands, the peptide DEDTQAMMPFR showed the best interactions counter to IFN- γ and TNF- α . The peptide with S-score of -11.31 showed interactions with Lys34, Arg42, Gln46, and Tyr53 amino acids of the binding pocket of IFN- γ (Figure 1). The second peptide, QEPQESQQ with S

-score of -10.98, exhibited interactions with Val22 and Glu39 residues of IFN- γ as a receptor protein (Figure S1). The third peptide FRDEHKK with S-score of -10.55 showed interactions with amino acids Val5, Glu39, and Tyr53 of IFN- γ (Figure S2). IFN- γ is a dimerized cytokine and known for its critical role in adaptive and innate immunity against variety of pathogens [21]. In addition to immunity response, it also activates a proinflammatory program in macrophages. The elevated level of IFN- γ has been observed in autoimmune complications including SLE [22].

For the IL-3, the peptide KHDRGDEF with S-score of -8.81 showed interactions with Cys16, Cys84, Pro86, and Leu87 amino acids of the binding pocket of IL-3 (Figure 2). The peptide FRDEHKK with S-score of -8.64 showed interactions with Asn15, Asp21, and Glu119 (Figure S3), and the peptide QEPQESQQ with S-score of -8.17 interacted with Pro83, Ala121, and Glu119 residues of IL-3 (Figure S4). Interleukin-3 is a multispecific hemopoietin, a glycoprotein cytokine that is synthesized by T cells in response to antigen. IL-3 has been involved in the proliferation and differentiation of many immune cells and both primitive multipotent and committed myeloid progenitors' cells [23, 24]. Elevated levels and increased IL-3-responsive progenitor cells have been reported in SLE patients [4].

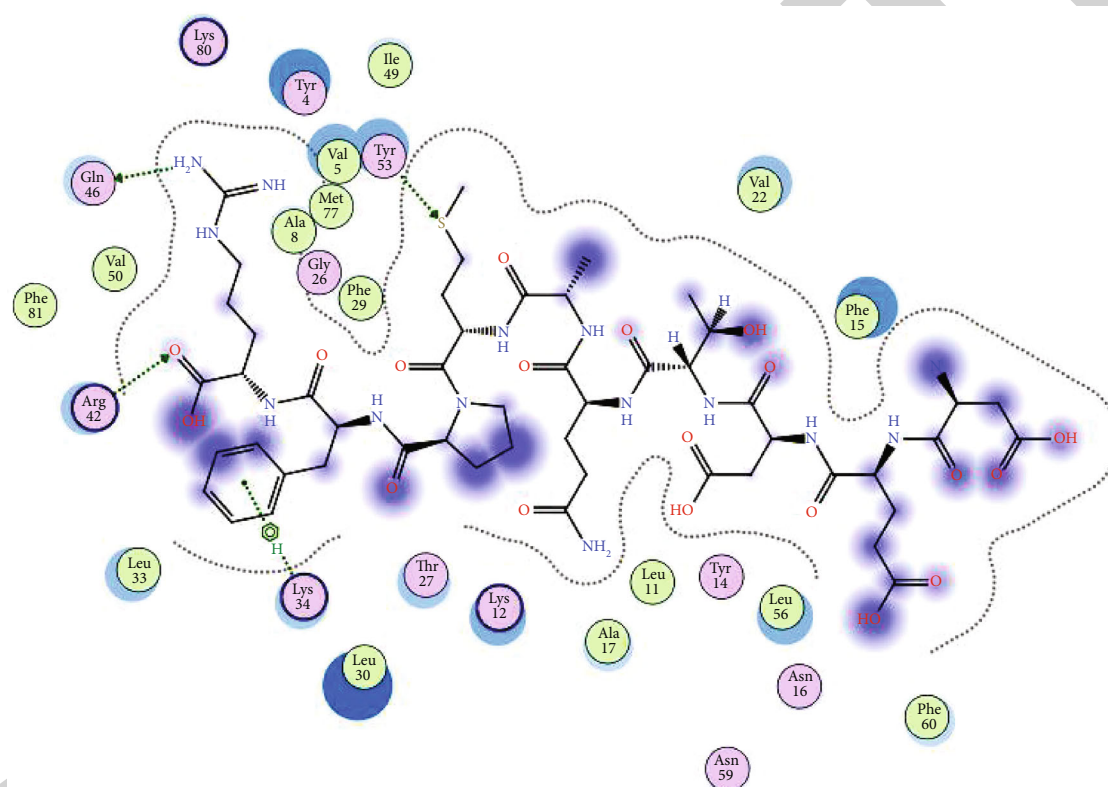
The other receptor protein TNF- α is an inflammatory cytokine and important for resistance to cancer and infection. The dysregulation and elevation of this protein has been associated with many autoimmune disorders. In our study, the peptide DEDTQAMMPFR with S-score of -8.20 showed interactions with Arg32 and Leu142 amino acids of the binding pocket of TNF- α (Figure 3). Other peptides (i.e., FRDEHKK and QEPQESQQ) with S-scores of -7.32 and -7.25 interacted with Asp143, Pro20 and Glu23 (Figure S5), and Arg32 amino acids (Figure S6) in the binding pocket of TNF- α , respectively. TNF- α is a proinflammatory cytokine, belongs to the super family of tumor necrosis factor, and secreted by macrophages in a defense mechanism to protect from damage by inducing inflammation against pathogenic stimuli. The elevated level or mutation in TNF- α signaling leads towards deleterious consequences including many autoimmune disorders [25].

Collectively, with IFN- γ , IL-3, and TNF- α , many other cytokines and different factors have been studied to understand the autoimmunity events. Thus, there is an immediate need for the present age to understand the autoimmunity to control excessive inflammatory responses and balance cytokine signaling.

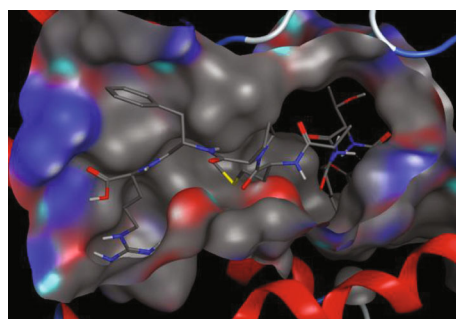
3.3. Peptide-Protein Docking. Peptide-protein interactions play a crucial role in a variety of regulatory and signaling pathways of the cell. The peptide-protein complex helps to open the key to elucidate important biological processes and to understand the underlying peptide-protein interactions. In this study, we used the top three ligands obtained from protein-ligand docking analysis for peptide-protein docking studies. The BLASTp was used to find suitable homologs and templates to build 3D structures of selected

TABLE 1: Top interactions of anti-inflammatory peptides with IFN- γ , IL-3, and TNF- α as receptor proteins.

Sr. No.	Peptide	Receptor	S-score	Interactions
1	DEDTQAMMPFR	IFN- γ	-11.30	Lys34, Arg42, Gln46, Tyr53
2	QEPQESQQ	IFN- γ	-10.98	Val22, Glu39
3	FRDEHKK	IFN- γ	-10.55	Val5, Glu39, Tyr53
4	KHDRGDEF	IL-3	-8.81	Cys16, Cys84, Pro86, Leu87
5	FRDEHKK	IL-3	-8.64	Asn15, Asp21, Glu119
6	QEPQESQQ	IL-3	-8.17	Pro83, Glu119, Ala121
7	DEDTQAMMPFR	TNF- α	-8.20	Arg32, Leu142
8	FRDEHKK	TNF- α	-7.32	Pro20, Glu23, Asp143
9	QEPQESQQ	TNF- α	-7.25	Arg32



(a)



(b)

FIGURE 1: Binding interactions of peptide DEDTQAMMPFR with receptor IFN- γ revealed through protein-ligand approach. (a) Interactions of peptide with the receptor. The Gln46 and Tyr53 are polar amino acids and acting as sidechain acceptor and donor, respectively. Arg42 and Lys34 are basic amino acids and acting as sidechain donor and acceptor, respectively. (b) Binding patterns of peptide with the receptor protein.

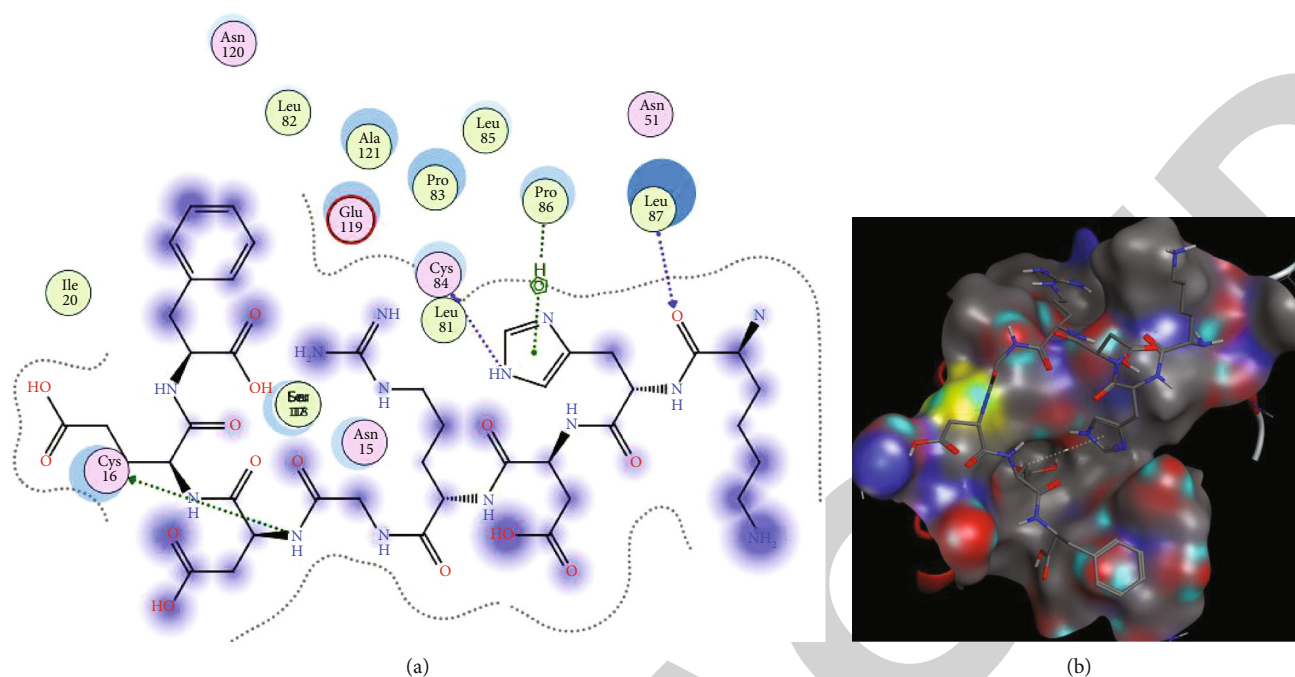


FIGURE 2: Binding interactions of peptide KHDRGDEF with receptor IL-3 revealed through protein-ligand approach. (a) Interactions of peptide with the receptor. The Cys16 Cys84 are polar amino acids and acting as sidechain and backbone acceptors, respectively. Pro86 and Leu87 are greasy amino acids and acting as sidechain acceptor and backbone donor, respectively. (b) Binding patterns of peptide with the receptor protein.

peptides in pdb format. Modeller 9.21 was employed for 3D structure predictions, and accurate models were selected on the basis of their DOPE values and GA341 scores. The HADDOCK server was used for docking of respective peptides counter to IFN- γ , IL-3, and TNF- α receptor proteins. Amongst selected peptides, only one peptide (i.e., DEDT-QAMMPFR) with the HADDOCK score of -10.3 ± 2.5 kcal/mol showed binding with IFN- γ (Figure 4). The educational version of PyMOL was used to visualize and draw the predicted cluster of peptide protein. The remaining two peptides counter to IL-3 and TNF- α are not discussed here due to their poor HADDOCK scores and energy structures.

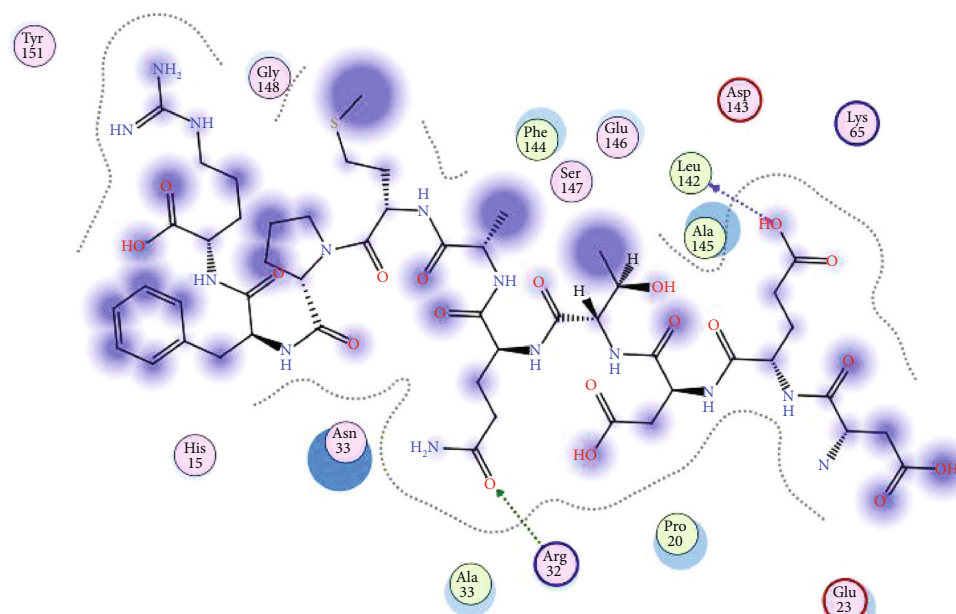
3.3.1. MD Simulation. The evolution of RMSD values for the C-alpha atoms of protein-ligand complex over time has been shown in Figure 5(a). The plot showed that the complex reached stability at 10 ns. This was increase in RMSD of peptide at 50 ns. After that, for the length of the simulation, fluctuations in RMSD values for target remained within 1.5 Angstrom which is absolutely acceptable [26]. The ligand fit to protein RMSD values fluctuated within 1.5 Angstrom after they have been equilibrated. These findings indicate that the peptide stayed firmly connected to the receptor throughout the simulation period.

On the RMSF graphic (Figure 5(b)), peaks represented the portions of the protein that fluctuated the most during the simulation. Protein tails (both N- and C-terminal) typically changed more than any other part of the protein. Alpha helices and beta strands, for example, are usually stiffer than the unstructured section of the protein and fluctuate less than loop portions. According to MD trajectories, the resi-

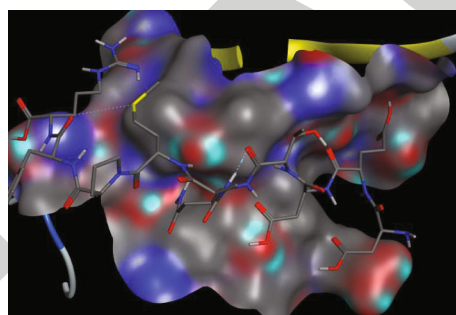
dues with greater peaks belonged to loop areas or N and C-terminal zones (Figure S7). Low RMSF values of binding site residues indicated that ligand binding to the protein is stable.

Alpha-helices and beta-strands are monitored as secondary structure elements during the simulation (SSE). The graph above depicts the distribution of SSE by residue index across the protein structure. Throughout the simulation, the left graphic showed the SSE composition for each trajectory frame while the right plot monitors each residue's SSE assignment through time.

Protein interactions with the ligand can be detected throughout the simulation. These interactions were categorized and summarized by types (Figure S7). The four types of protein-ligand interactions (or "contacts") include hydrogen bonds, hydrophobic interactions, ionic interactions, and water bridges. The "Simulation Interactions Diagram" panel in Maestro was used to study the subtypes of each interaction type. Over the course of the trajectory, the stacked bar charts were standardized: for example, a value of 0.7 indicated that the specific interaction was maintained for 70% of the simulation duration. Because some protein residues may make several interactions of the same subtype with the ligand, values above 1.0 are feasible. The majority of the significant ligand-protein interactions discovered by MD were hydrogen bonds and hydrophobic interactions (Figure 6). B:THR_27, D:GLY_75, and D:ASN_136 are the most important in terms of H-bond, and D:MET_25, D:PRO_77, and D:PRO_146 are the most important in terms of hydrophobic interactions.



(a)



(b)

FIGURE 3: Binding interactions of peptide DEDTQAMMPFR with receptor TNF- α revealed through protein-ligand approach. (a) Interactions of peptide with the receptor. Arg32 is a basic amino acid and acting as a sidechain donor while Leu142 is a greasy amino acid and acting as a backbone acceptor. (b) Binding patterns of peptide with the receptor protein.

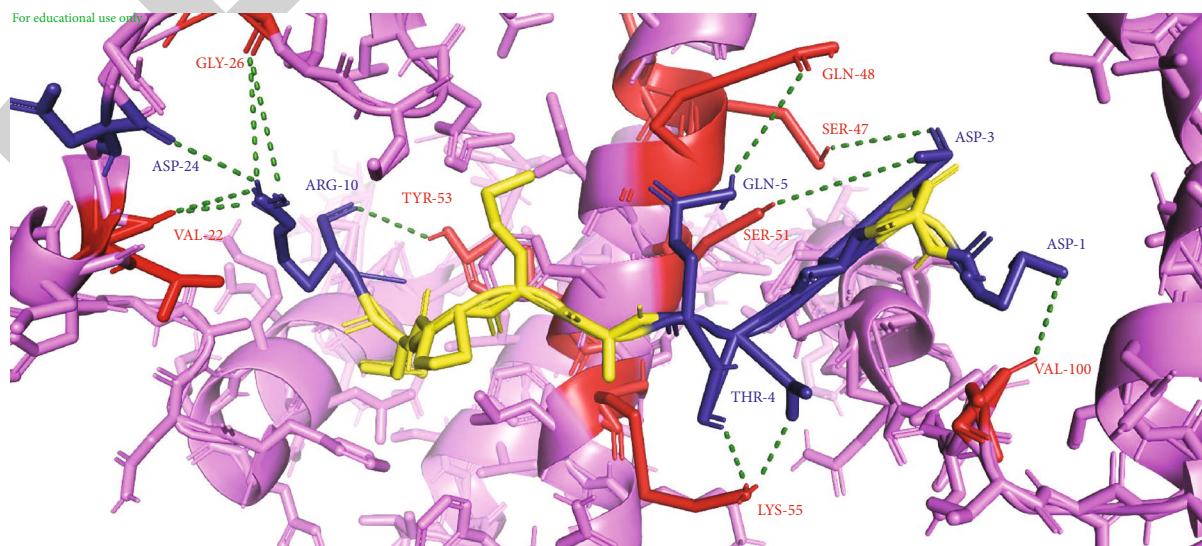


FIGURE 4: Peptide-protein interactions between DEDTQAMMPFR and IFN- γ . The peptide DEDTQAMMPFR has been represented in yellow color with deep blue-colored interacting residues, and IFN- γ is shown in violet color with red interacting residues.

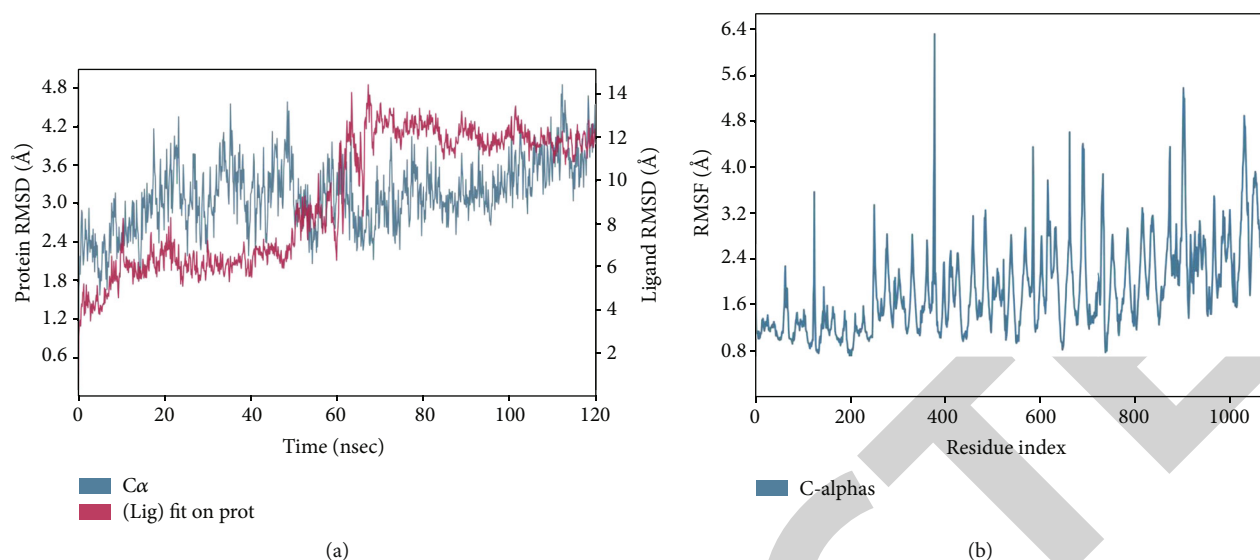


FIGURE 5: Root mean square deviation (RMSD) and residue wise root mean square fluctuation (RMSF). (a) RMSD of the C-alpha atoms of protein and ligand with time (receptor-peptide complex). The left Y-axis shows the variation of protein RMSD through time. The right Y-axis shows the variation of ligand RMSD through time. (b) RMSF of the protein.

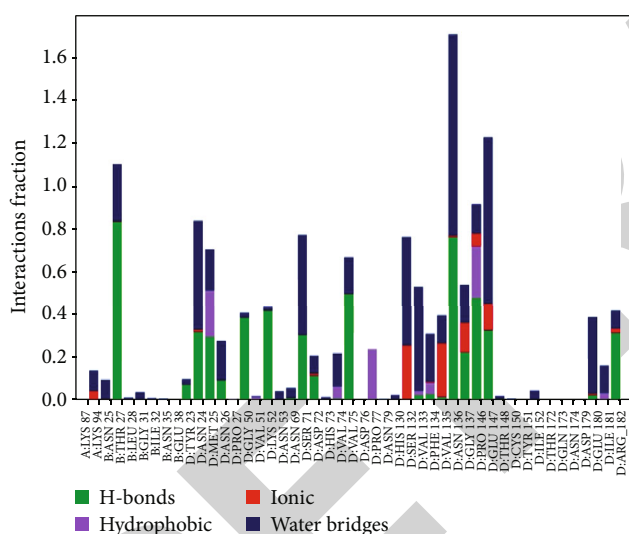


FIGURE 6: Protein-ligand contact histogram (H-bonds, hydrophobic, ionic, water bridges).

A timeline has exhibited the interactions and contacts (H-bonds, hydrophobic, ionic and water bridges) as described above. In Figure 7, the top panel displayed the total number of specific connections the protein made with the ligand over the duration of the journey. The bottom panel of each trajectory frame showed which residues interacted with the ligand. Some residues made many particular connections with the ligand which has been indicated by a deeper shade of orange color according to the scale to the right of the plot.

Over the course of the trajectory, the stacked bar charts were standardized: for example, a value of 1.0 signified that the exact interaction was maintained for 100% of the simulation duration. Values exceeding 1.0 are possible because some protein residues may make several interactions of the

same subtype with the ligand. The interactions of individual ligand atom with protein residues are showed in Figure 8. Interactions that last more than 30% of the simulation period in the selected trajectory (0.00 to 100.0 nsec) are shown.

The MMGBSA.py script from the Desmond module of the Schrodinger suite 2019-4 was used to perform the MM-GBSA analysis. Every frame was collected from each MD trajectory for binding free energy estimates of the receptor in combination with the peptide (Figure 9). Total energy was ranged from -996.226 to 5.056. The mean and median were -356.755 and -343.800, respectively, indicating good energy. Total binding free energy (kcal/mol) was calculated using the law of additivity in which individual energy modules such as coulombic, covalent, hydrogen bond, van der Waals, self-contact, lipophilic, solvation, and π - π stackings of the ligand and the protein were added together [27].

4. Discussion

Molecular docking is an elaborative approach to foresee the interactions between ligand and targeted amino acids in the binding pocket of the receptor protein [28]. Computational approaches including molecular docking help scientists to predict the binding capacities of different small molecules and peptides as drug candidates against different receptor proteins [29]. In the current study, we have used some plant- and animal-derived anti-inflammatory peptides ranging from 3 to 15 amino acid residues as ligands/peptides counter to proteins from different bacteria which are the leading cause of many autoimmune disorders.

The autoimmune disease SLE is a disorder of connective tissues with a wide range of clinical manifestations. The autoreactive B (bone marrow- or bursa-derived cells) and T cells (thymus cells) of adaptive and innate immunity play a leading role in the production of autoantibodies and lead

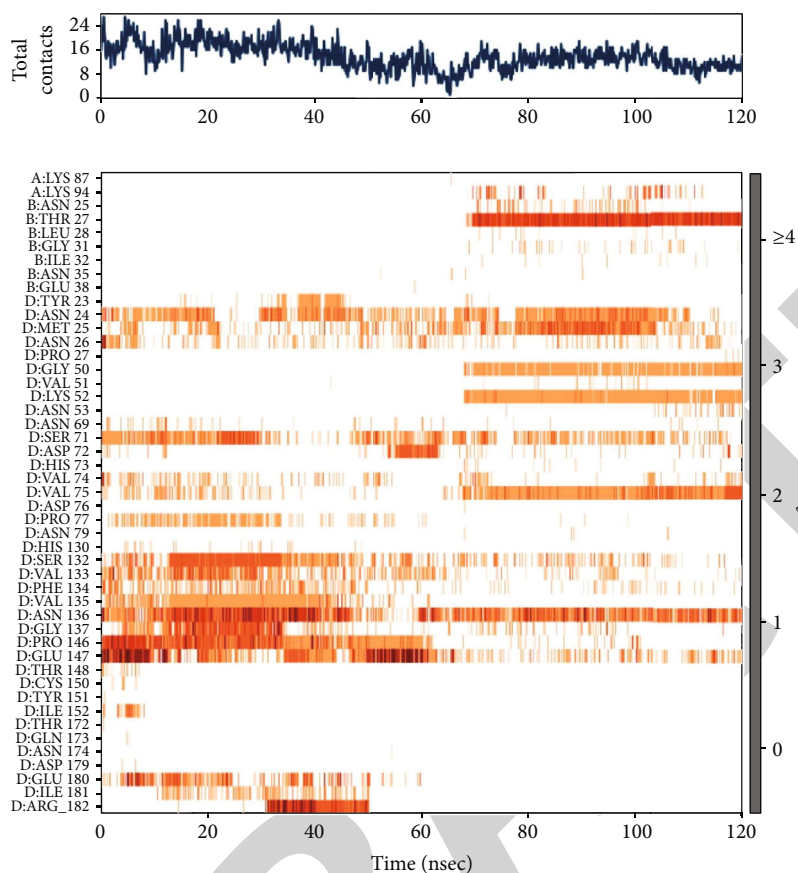


FIGURE 7: A timeline representation of the interactions and contacts (H-bonds, hydrophobic, ionic, and water bridges).

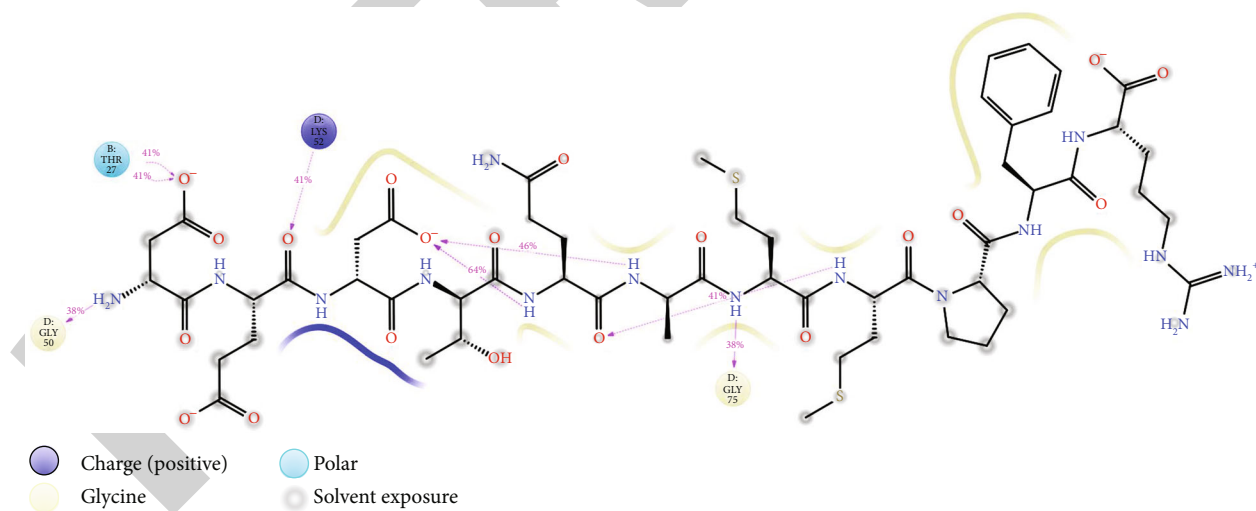


FIGURE 8: Ligand atom interactions with the protein residues.

towards autoimmune disorders including SLE [6]. Different factors such as IL-3, IL23, M-CSF, IFN- α , IFN- γ , and IL-6 have been reported in different studies to be involved in the production of autoantibodies but the main trigger for all these autoimmune complexities is still unknown [4, 30]. The elevated levels of proinflammatory cytokines and TNF- α have been reported in many studies as associated

mainly with autoreactive B cells and T cells to produce autoantibodies [4, 6].

IFN- γ or type II interferon is a pleiotropic cytokine that coordinates with a diverse array of cellular immunity processes [31]. IFN- γ is a homodimer and formed by noncovalent interactions of 17 kDa polypeptide dimers and crucially known for early control of pathogen spreading. IFN- γ has

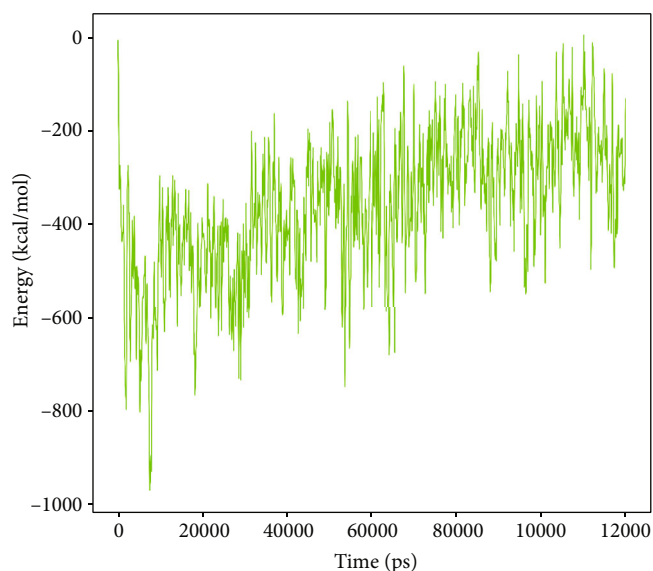


FIGURE 9: Estimation of binding free energy of the receptor in combination with the peptide using MM-GBSA.

been secreted mainly by CD4 T helper cells, CD8 cytotoxic T cells, and to a less extent by antigen-presenting cells (APCs) and natural killer cells [32]. The peripheral blood mononuclear cells (PBMCs) of SLE patients showed high level of IFN- γ transcript, and the T cells of SLE patients produce much more IFN- γ as compared to normal cells [33]. In the current study, strong interactions of three peptides (i.e., DEDTQAMMPFR, QEPQESQQ, and FRDEHKK) were found with the active amino acids present in the binding pocket of IFN- γ which could be used as potential inhibitors of IFN- γ to treat SLE.

IL-3 is a monomeric glycoprotein which is predominantly produced by activated T cells in response to stimuli. It serves as a bridge between the immune system (T-lymphocytes) and the hemopoietic system that in response to foreign stimuli generates cellular elements for cellular defense [23]. The dysregulation of IL-3 has been associated with various autoimmune diseases including arthritis and SLE. In SLE patients, elevated IL-3 responsive progenitor cells have been observed in spleen, which show an association between IL-3 and autoreactive cells [4]. The molecular docking approach used in this study exhibited strong interactions of three peptides (i.e., KHDRGDEF, FRDEHKK, and QEPQESQQ) with the active amino acids present in the binding pocket of IL-3. These peptides could be potential inhibitors of IL-3.

TNF- α is a pleiotropic proinflammatory cytokine and contributes importantly to the development of B and T cells. TNF- α is a potent inflammatory mediator of chronic and acute inflammation and secreted by macrophages, T cells, and neutrophils [34]. The involvement of TNF- α in the pathogenesis of SLE has been observed in many clinical studies due to overexpression and elevated levels of TNF- α in SLE patients [35]. In current study, using molecular docking approach, three peptides (i.e., DEDTQAMMPFR, FRDEHKK, and QEPQESQQ) showed strong interactions

with the active amino acids present in the binding pocket of TNF- α and could be used in the treatment of SLE.

The overexpression and elevated levels of interferons, interleukins, and tumor necrosis factor point towards the relationship of these factors to the pathogenesis of SLE and many other autoimmune disorders. The blockage of IFN- γ , IL-3, and TNF- α could be proved as an effective strategy to control the tissue and organ damage in SLE. In clinical trials, there are many ongoing therapies to control signaling pathways and targeted autoantibodies. Till now, many drugs have been approved and marketed such as rituximab, epratuzumab, abetimus, sodium, obinutuzumab, lulizumab pegol, abatacept, and blisibimod for the treatment of autoimmune disorders. The drug blisibimod is a fusion protein which is an antagonist of BAFF with little encouraging outcomes obtained after a phase III trial for the treatment of SLE [30]. In spite of many drugs and combinational therapies, these drugs have been associated with severe after effects; so, there is a need of such types of drugs with maximum potency and minimum side effects.

The proinflammatory cytokines, interleukins, and interferons make up the defense system of the cell and play a crucial role in generating different molecules in response to external pathogenic stimuli. Any mutation and overexpression of any of these molecules lead towards the production of autoantibodies against body's own cells. The peptides reported in this study could be used as potential drug candidates counter to IFN- γ , IL-3, and TNF- α as receptor proteins. Further elaborative study is still needed to explore much more potential of these anti-inflammatory peptides.

5. Conclusion

The docking analysis and S-scores of selected peptides have revealed the potential of selected peptides as drug candidates counter to inflammatory autoreactive proteins to control autoimmunity. In the current study, we used two types of docking analysis (protein-ligand and peptide-protein docking) to check the configuration and orientation of ligands/peptides counter to selected receptor proteins. In our first approach, the peptide DEDTQAMMPFR showed strong interactions with active amino acids of IFN- γ (S-score -11.31) and TNF- α (S-score 8.20) receptor proteins. The conformations showed the occupancy of the maximum binding pocket by ligand molecules. In our second approach (peptide-protein docking), the same peptide also showed strong bonding with the active amino acids of IFN- γ via HADDOCK server. The IFN- γ -DEDTQAMMPFR complex with HADDOCK score of -10.3 ± 2.5 kcal/mol showed strong interactions amongst active amino acid residues of both peptide and receptor protein. Further, the MD simulation analysis also confirmed that the peptide stayed firmly connected to the receptor throughout the simulation period (i.e., 120 nanoseconds). The results of the current study have explored the potential of peptides of plant and animal sources as drug molecules to control autoimmunity. The study could be proved as an initial step for further use of these peptides after some required modifications as drug candidates against autoimmune disorders.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Authors' Contributions

GM and HSM conceived and planned the experiments. GM, SS, and HSM carried out the protein-ligand and protein-protein docking experiments. HT, MS, and TM drafted the manuscript. RJ and GM helped conducting molecular dynamics and simulation study. GM supervised the project and proofread the article. All authors discussed the results and commented on the manuscript.

Acknowledgments

The authors would like to gratefully acknowledge Department of Biochemistry, Government College University Faisalabad, for providing space and facilities to accomplish this study.

Supplementary Materials

Figure S1: interactions (a) and binding pattern (b) of QEP-QESQQ peptide with IFN- γ . Figure S2: interactions (a) and binding pattern (b) of FRDEHKK peptide with IFN- γ . Figure S3: interactions (a) and binding pattern (b) of FRDEHKK peptide with IL-3. Figure S4: interactions (a) and binding pattern (b) of QEPQESQQ peptide with IL-3. Figure S5: interactions (a) and binding pattern (b) of FRDEHKK peptide with TNF- α . Figure S6: interactions (a) and binding pattern (b) of QEPQESQQ peptide with TNF- α . Figure S7: protein secondary structure element distribution by residue index throughout the protein structure (Supplementary Materials). (*Supplementary Materials*)

References

- [1] L. B. Nicholson, "The immune system," *Essays in Biochemistry*, vol. 60, no. 3, pp. 275–301, 2016.
- [2] Y. Juarranz, "Molecular and cellular basis of autoimmune diseases," *Cells*, vol. 10, p. 474, 2021.
- [3] G. Mustafa, H. S. Mahrosh, and R. Arif, "Sequence and structural characterization of toll-like receptor 6 from human and related species," *BioMed Research International*, vol. 2021, Article ID 5545183, 9 pages, 2021.
- [4] T. A. Gottschalk, E. Tsantikos, and M. L. Hibbs, "Pathogenic inflammation and its therapeutic targeting in systemic lupus erythematosus," *Frontiers in Immunology*, vol. 6, p. 550, 2015.
- [5] G. Murphy and D. Isenberg, "Systemic lupus erythematosus," in *Encyclopedia of Immunobiology*, Academic Press, 2016.
- [6] D. Zucchi, E. Elefante, E. Calabresi, V. Signorini, A. Bortoluzzi, and C. Tani, "One year in review 2019: systemic lupus erythematosus," *Clinical and Experimental Rheumatology*, vol. 37, no. 5, pp. 715–722, 2019.
- [7] L. M. Mathias and W. Stohl, "Systemic lupus erythematosus (SLE): emerging therapeutic targets," *Expert Opinion on Therapeutic Targets*, vol. 24, no. 12, pp. 1283–1302, 2020.
- [8] G. C. Tsokos, M. S. Lo, P. C. Reis, and K. E. Sullivan, "New insights into the immunopathogenesis of systemic lupus erythematosus," *Nature Reviews Rheumatology*, vol. 12, no. 12, pp. 716–730, 2016.
- [9] C.-L. Murphy, C.-S. Yee, C. Gordon, and D. Isenberg, "From BILAG to BILAG-based combined lupus assessment—30 years on," *Rheumatology*, vol. 55, no. 8, pp. 1357–1363, 2016.
- [10] A. Mohamed, Y. Chen, H. Wu, J. Liao, B. Cheng, and Q. Lu, "Therapeutic advances in the treatment of SLE," *International Immunopharmacology*, vol. 72, pp. 218–223, 2019.
- [11] ULC C, *Molecular Operating Environment (MOE)*, 2013.08, , vol. 10, 1010 Sherbooke St. West, Suite# 910, Montreal, QC, Canada, H3A 2R7, 2018.
- [12] G. C. P. van Zundert, J. Rodrigues, M. Trellet et al., "The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes," *Journal of Molecular Biology*, vol. 428, no. 4, pp. 720–725, 2016.
- [13] Z. Li, H. Wan, Y. Shi, and P. Ouyang, "Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/draw, and ChemSketch," *Journal of Chemical Information and Computer Sciences*, vol. 44, no. 5, pp. 1886–1890, 2004.
- [14] B. Webb and A. Sali, "Comparative protein structure modeling using MODELLER," *Current Protocols in Bioinformatics*, vol. 54, no. 1, pp. 5.6. 1–5.6. 37, 2016.
- [15] W. L. DeLano, "Pymol: an open-source molecular graphics tool," *CCP4 Newsletter on protein crystallography*, vol. 40, no. 1, pp. 82–92, 2002.
- [16] K. J. Bowers, D. E. Chow, H. Xu et al., "Scalable algorithms for molecular dynamics simulations on commodity clusters," in *SC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, p. 43, Tampa, FL, USA, 2006.
- [17] L. G. Ferreira, R. dos Santos, G. Oliva, and A. D. Andricopulo, "Molecular docking and structure-based drug design strategies," *Molecules*, vol. 20, no. 7, pp. 13384–13421, 2015.
- [18] P. W. Hildebrand, A. S. Rose, and J. K. Tiemann, "Bringing molecular dynamics simulation data into view," *Trends in Biochemical Sciences*, vol. 44, no. 11, pp. 902–913, 2019.
- [19] M. A. Rasheed, M. N. Iqbal, S. Saddick et al., "Identification of lead compounds against Scm (fms10) in *Enterococcus faecium* using computer aided drug designing," *Life*, vol. 11, no. 2, p. 77, 2021.
- [20] D. Shivakumar, J. Williams, Y. Wu, W. Damm, J. Shelley, and W. Sherman, "Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field," *Journal of Chemical Theory and Computation*, vol. 6, no. 5, pp. 1509–1519, 2010.
- [21] S. Payne, *Chapter 6-Immunity and Resistance to Viruses, Viruses*. Academic Press, Cambridge, MA, USA, 2017.
- [22] S. Najafi, E. Rajaei, R. Moallemian, and F. Nokhostin, "The potential similarities of COVID-19 and autoimmune disease pathogenesis and therapeutic options: new insights approach," *Clinical Rheumatology*, vol. 39, no. 11, pp. 3223–3235, 2020.
- [23] F. J. Pixley and E. R. Stanley, "Cytokines and cytokine receptors regulating cell survival, proliferation, and differentiation in hematopoiesis," in *Handbook of Cell Signaling*, Elsevier, 2010.

Retraction

Retracted: Epi-Gene: An R-Package for Easy Pan-Genome Analysis

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.





The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] F. Awan, M. M. Ali, M. Hamid et al., "Epi-Gene: An R-Package for Easy Pan-Genome Analysis," *BioMed Research International*, vol. 2021, Article ID 5585586, 8 pages, 2021.

Research Article

Epi-Gene: An R-Package for Easy Pan-Genome Analysis

Furqan Awan ^{1,2}, Muhammad Muddassir Ali,³ Muhammad Hamid ⁴,
Muhammad Huzair Awan,⁵ Muhammad Hassan Mushtaq,² Saeeda Kalsoom,⁶
Muhammad Ijaz,⁷ Khalid Mehmood ⁸ and Yongjie Liu ¹

¹Joint International Research Laboratory of Animal Health and Food Safety, College of Veterinary Medicine, Nanjing Agricultural University, Nanjing 210095, China

²Department of Epidemiology and Public Health, University of Veterinary and Animal Sciences, Lahore 54000, Pakistan

³Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences, Lahore 54000, Pakistan

⁴Department of Statistics and Computer Sciences, University of Veterinary and Animal Sciences, Lahore 54000, Pakistan

⁵Computer Foundation Department, Cyber Brain Educational Institute, Lahore 54000, Pakistan

⁶Department of Biotechnology, Virtual University of Pakistan, Lahore 54000, Pakistan

⁷Department of Veterinary Medicine, University of Veterinary and Animal Sciences, Lahore 54000, Pakistan

⁸Faculty of Veterinary and Animal Sciences, The Islamia University of Bahawalpur, 63100, Pakistan

Correspondence should be addressed to Khalid Mehmood; khalid.mehmood@iub.edu.pk and Yongjie Liu; liyongjie@njau.edu.cn

Received 29 January 2021; Accepted 28 August 2021; Published 21 September 2021

Academic Editor: Harry Schroeder Jr

Copyright © 2021 Furqan Awan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The main aim of this study was to develop a set of functions that can analyze the genomic data with less time consumption and memory. Epi-gene is presented as a solution to large sequence file handling and computational time problems. It uses less time and less programming skills in order to work with a large number of genomes. In the current study, some features of the Epi-gene R-package were described and illustrated by using a dataset of the 14 *Aeromonas hydrophila* genomes. The joining, relabeling, and conversion functions were also included in this package to handle the FASTA formatted sequences. To calculate the subsets of core genes, accessory genes, and unique genes, various Epi-gene functions have been used. Heat maps and phylogenetic genome trees were also constructed. This whole procedure was completed in less than 30 minutes. This package can only work on Windows operating systems. Different functions from other packages such as dplyr and ggtree were also used that were available in R computing environment.

1. Introduction

In the last few years, sequencing technologies have made whole-genome sequencing easier and inexpensive [1, 2]. Consequently, this leads to a rise in prokaryotic genome sequences in a short time and at a small cost. This bloom did not limit it at the genus or species level [3, 4]. It expanded to sequence the strains of the same species in order to study the physiological diversity [5]. Moreover, these prokaryotic genome sequences helped us to investigate the outbreaks and their associated risk factors [6, 7].

Prokaryotes have more diverse and vibrant genomes as compared to eukaryotes [8, 9]. The main reason for this diversity, especially in bacterial pathogens, is frequent expo-

sure to a variety of stresses in their natural environment and in their host systems. This may lead to accumulation of unique genes for structural and regulatory mechanisms via gene transfers and mutations [10]. Contrarily, eukaryotes are more complex and multicellular organisms that have stringent stress management with minimal chance of introduction of unique genes [11]. This diversity among microbes could be a possible hurdle against their correct classification and identification as pathogenic and nonpathogenic strains [12].

Pan-genomic studies are found to be fruitful in the correct classification of the strains and identification of pathogenic genes related to the pathogenicity of that particular strain [13]. Such studies cluster all the genes and classify

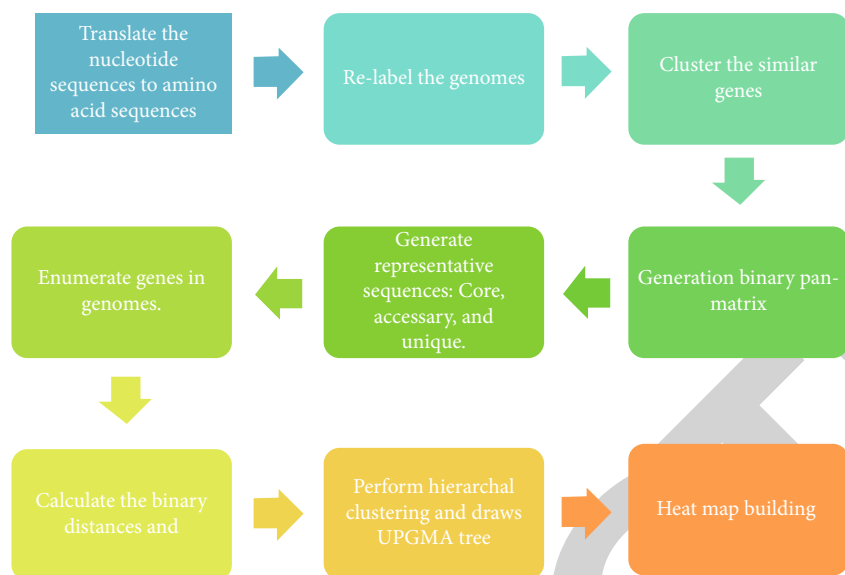


FIGURE 1: Steps involved in the work flow of Epi-gene package.

them into classes based on their presence in genomes [3]. Commonly, bacterial pan-genomes are comprised of conserved or core genes (shared by all) and dispensable genes (shared by some) [14]. Core gene clusters could be helpful in phylogenetic analysis, while the dispensable gene clusters are helpful in identifying the unique characters, especially antibiotic resistance and virulence factors [4]. These gene clusters serve as the backbone of pan-genomic studies, but this computation needs immense and ample time.

The main aim of this study was to develop a package that can statistically analyze the genomic data with less time consumption and require beginner-level programming skills. It was also intended to develop various functions that can perform data wrangling with the FASTA formatted sequences in R-language environment. In the current study, some features of the Epi-gene R-package are described and illustrated by using a dataset of the 14 *Aeromonas hydrophila* genomes. *A. hydrophila* is a well-known Gram-negative bacteria with diverse genetic architecture [10]. Therefore, Epi-gene was employed to investigate the pan-genome studies of highly diverse strains of *A. hydrophila*.

2. Methods

A case study has been described in this package, with R-code, which can serve as a template or guideline for the users to implement this study. Here, an overview of the package implementation and some steps for the analysis are provided (Figure 1).

2.1. R-Statistical Language. The R-statistical language is a free tool. Unlike other programming software, only beginner-level programming skills are enough for basic analyses [15]. It has a huge collection of packages and possible solutions for data handling, statistical calculations, and graphical representations. In the beginning, it was used to develop functions for purely statistical problems, but now,

TABLE 1: *A. hydrophila* genomes included in this study with the summary of calculated datasets.

Bacterial strains	ID	Total number of genes	Number of accessory genes	Number of unique genes
4AK4	org1	3928	323	445
Ah10	org2	4178	847	171
AHNIH1	org3	4176	854	162
AL0606	org4	4252	922	170
AL0971	org5	4319	1158	1
ATCC7966	org6	4076	812	104
D4	org7	4371	1201	10
GYK1	org8	4226	1039	27
J1	org9	4307	1141	6
JBN2301	org10	4404	1237	7
ML09-119	org11	4320	1159	1
NJ35	org12	4512	1199	153
PC104A	org13	4322	1161	1
YL17	org14	4099	694	245

it is being used for statistical calculations of huge genomic data [16–19]. The Epi-gene package focuses the microbial pan-genomics and offers various functions in this regard. It also uses the other packages in R for different calculations.

2.2. External Software Packages. External software such as Usearch was employed for the typical computation of gene clustering. Usearch is free for any user and can be downloaded easily after registration. It offers gene clustering computation in a very short time as compared to the Basic Local Alignment Search Tool (BLAST) [20]. Epi-gene directs Usearch for clustering and other functions from within R-language.

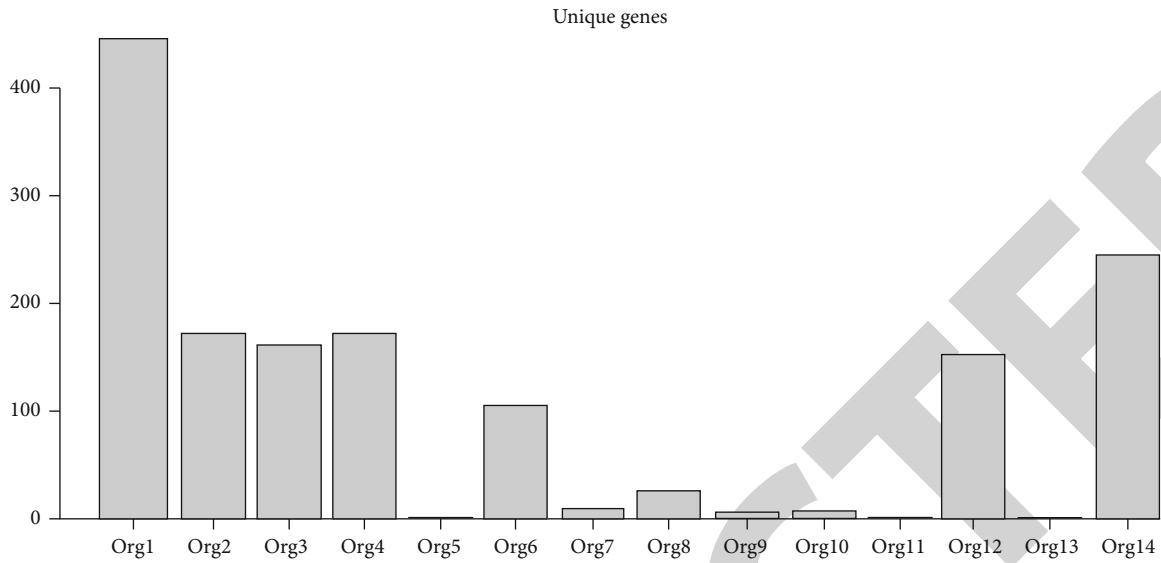


FIGURE 2: Graphical representation of unique genes across the included *A. hydrophila* strains.

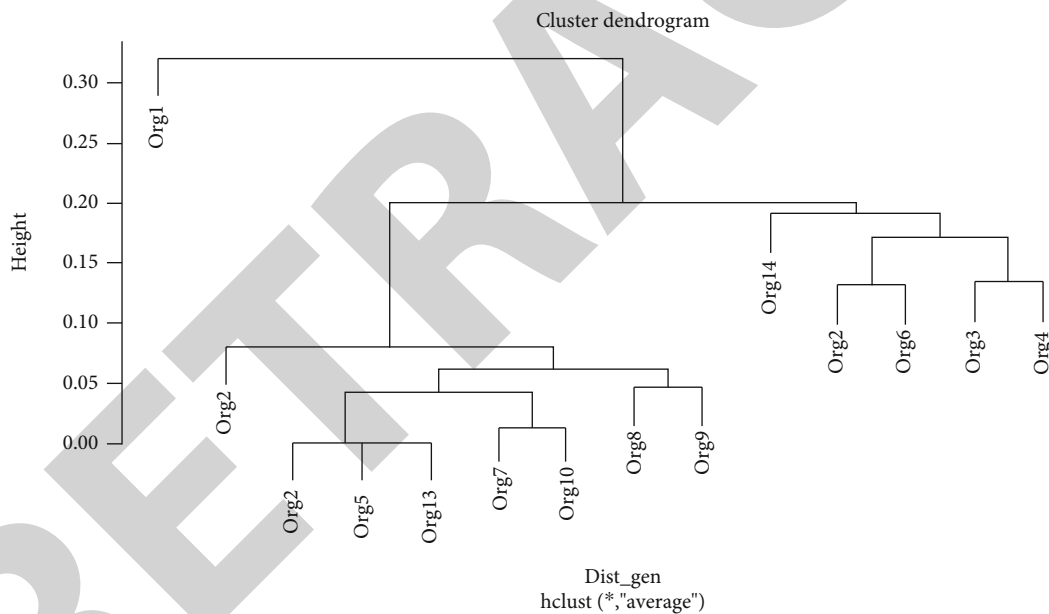


FIGURE 3: Dendrogram showing the phylogenetic relation of *A. hydrophila* genomes.

2.3. FASTA Format-Related Functions. FASTA format is a commonly used file extension format to store nucleotide and amino acid sequences. But handling a large number of files of this format is sometimes difficult. In this package, multiple functions are developed that will be utilized during this study but can also be utilized on individual needs. These functions include relabeling, joining multiple FASTA files, and conversion of FASTA format files to text delimited formats. These functions can be utilized with the commands of relabel, convert, and joining. Another useful function is developed to concatenate all the contigs or scaffolds in order to develop a single line genome sequence for user needs.

2.4. Binary Pan-Matrix. A pan-genome analysis is usually based on a pan-matrix. To compute this pan-matrix, there are two steps: the first step involves the heavy computations followed by the analyses that take pan-matrix as the input. A large number of amino acid sequences are compared which is the main constriction faced during a pan-genome study. To solve this computational problem, UCLUST is chosen. This is invoked from R by the function clustering in the Epi-gene package. UCLUST is 1000 times faster than BLAST whereas results are highly accurate as mentioned in previous studies [20, 21]. Based on this clustering, all the sequences are clustered into gene clusters that would represent classical gene families.

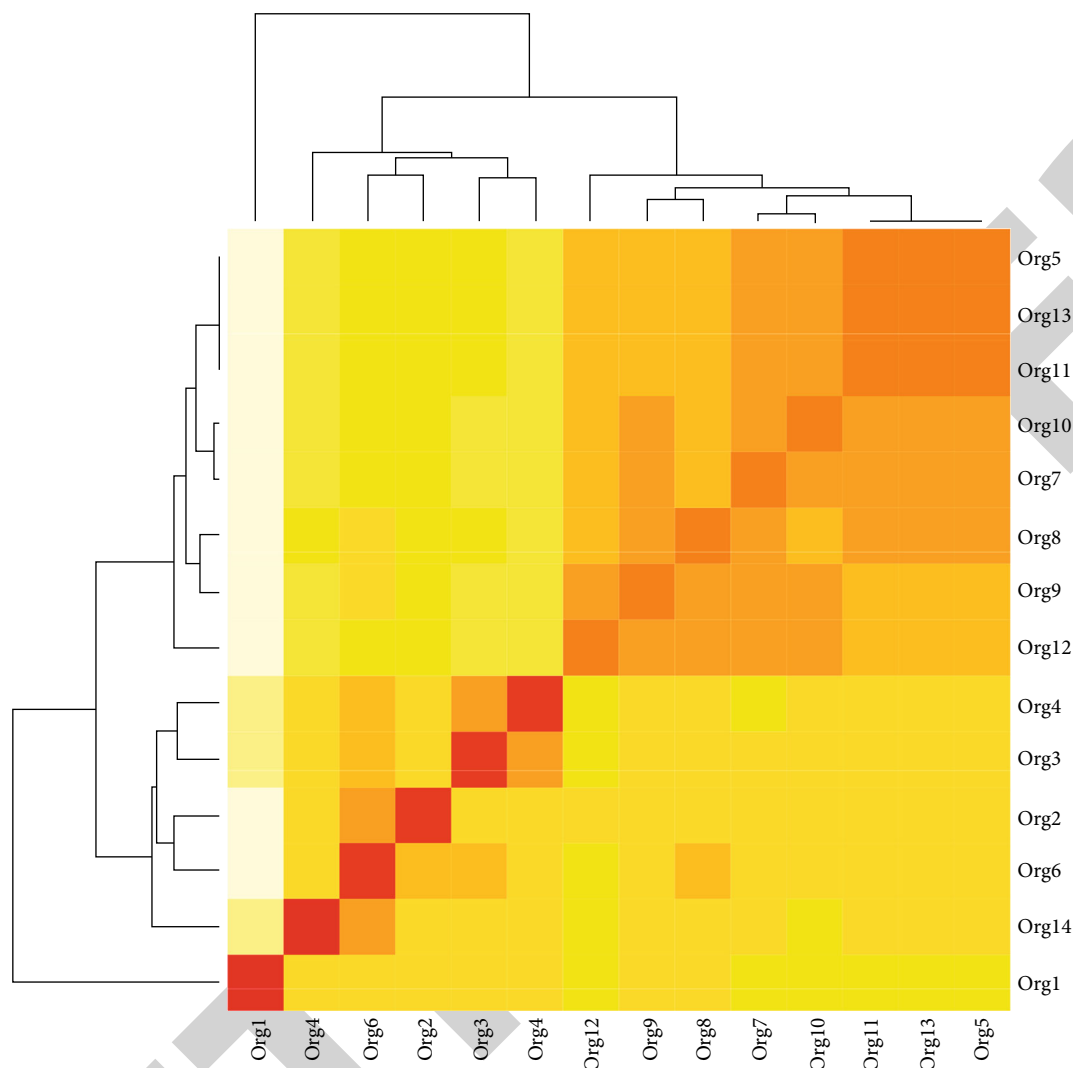


FIGURE 4: Heat map showing the graphical representation of phylogenetic relation of *A. hydrophila* genomes.

2.5. Analysis of Core, Accessory, and Unique Genes. The analysis of the core, accessory, and unique genes can be performed based on the previously calculated binary pan-matrix data. Core genes are defined as genes shared by all the genomes while the dispensable genes either present in two or more strains (accessory genes) or present in only one strain (unique genes) can also be identified. These three classes of genes can be enumerated and graphically represented according to individual need.

2.6. Phylogenetic Analyses. As the pan-matrix is based on the presence or absence of gene families, binary distances between genomes can be computed under the `distGen` function. This function can transform the pan-matrix values into continuous variables that can define the genome. Based on this function, it is possible to perform the hierarchical clustering of the genomes and can be displayed as pan-genome trees. This pan-genome tree can be illustrated by using the `Gentree` function.

2.7. Graphical Representation. Graphical representation is more illustrative than long and heavy tables. In the Epi-gene package, it is also possible to illustrate a heat map along with the pan-genome phylogenetic tree. A heat map is generated with the different possible user-defined pallets and colors.

2.8. Pan-Matrix Based on Sequence Identity. Another pan-matrix was also developed based on the sequence identity of the genes with each other in a cluster. Based on this pan-matrix, quantification of data is possible that can lead to further downstream statistical analyses. Possible statistical analyses involve the principal component analyses (PCA). This pan-matrix can be performed by the function of `id-matrix`. For further calculations of continuous data, other statistical packages can be utilized.

3. Implementation

To demonstrate some aspects of the Epi-gene package, the publicly available data for the 14 complete sequences of *A.*

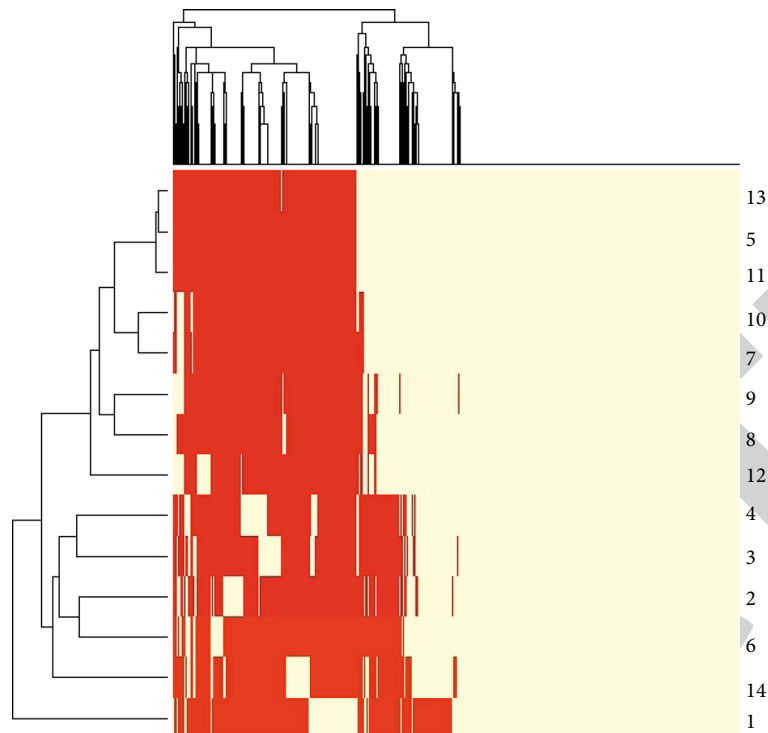


FIGURE 5: Heat map showing the phylogenetic relation of *A. hydrophila* genomes along with the presence or absence of clusters.

hydrophila were used. Within the Epi-gene, a case study document has been included that demonstrates all computations as a guideline for users.

First, genome sequences for 14 *A. hydrophila* genomes were downloaded from NCBI. Next, FASTA sequences were relabeled and joined together to form a multiple sequence file. Optionally, according to the user's need, these FASTA formatted sequences can be converted to txt format and single line sequence. The pan-genome based on 14 genomes was having a median of 4279.5 genes with a range of 3928 to 4512 and a total of 59490 sequences (Table 1). After clustering, pan-matrix was constructed from the homogenous gene clusters. All *A. hydrophila* genomes contain almost half of the core genes. There are 6394 gene clusters present in all 14 genomes. The core number of genes was found to be 3160 genes (Table 1). There was a high number of accessory genes present in this pan-genome ranging from 323 to 1237. A total of 1503 unique gene clusters were found in the pan-genome (Figure 2).

Clustering the genes also enabled us to analyze the phylogenetics of the organisms under study. Followed by clustering, a binary distance matrix was calculated that assigns the different values to different strains or organisms. The dendrogram showed more relevant organisms together via the neighbor joining clustering method (Figure 3).

The graphical heat map is an interactive tool that can express data in a more good way. Epi-gene has two types of heat map-related functions. The first function can generate a heat map with binary matrix assigned values. It is a short heat map with more relation to phylogenetics (Figure 4). The second function can generate a heat map with all the genes present or absent in a genome. The second

function could take more time because of the handling of large genomes (Figures 4 and 5).

The pan-matrix based on sequence identity can be utilized for multiple possible statistical analyses. In this study, we have performed principal component analyses (PCA) to understand more variation and dimension reduction. The scree plot based on eigenvalues could be seen in Figure 6(a). Moreover, based on the PCA, similar genomes were clustered close to each other (Figure 6(b)) as they were clustered in binary matrix-based clustering. Furthermore, a biplot was also drawn that was including the gene clusters as variables and genomes as individuals (Figure 7). These calculations could be further modified and used to select highly variable gene clusters.

4. Discussion

An increasing trend of genome-level research has opened many ways to focus on microbes. But handling a large number of genomes in a single analysis is a bottleneck [4]. In the current study, the package Epi-gene has addressed this issue by utilizing the UCLUST algorithm of the Usearch software package. It is already known that Usearch is 1000 times faster than BLAST [20]. The case study performed in this research took five minutes to perform clustering of all genes. This algorithm was also adopted in BPGA software [22]. But that software lacks technical support with restriction of options for further downstream analyses. Moreover, there are serious concerns over the source code of BPGA. But the Epi-gene is freely available and can be understood easily.

Handling a large number of FASTA formatted files in Windows and other operating systems is sometimes difficult.

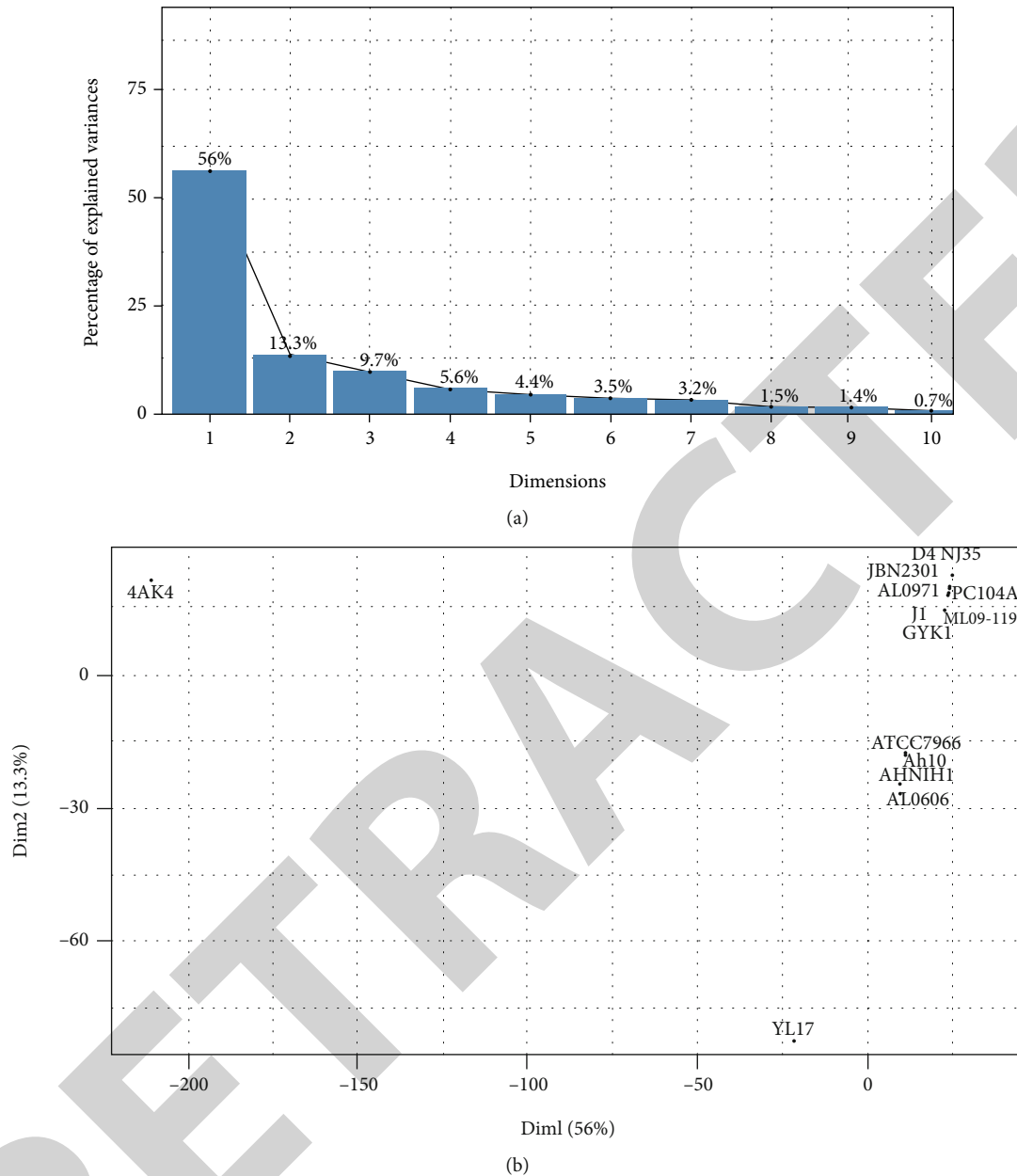


FIGURE 6: PCA-based results performed by utilizing the pan-matrix based on the sequence identity. (a) Scree plot describes the reduced dimensions and eigenvalues. (b) Individuals included in PCA show the clustering that is quite similar to the bin-matrix-based clustering.

Specifically, joining and relabeling the multiple FASTA formatted sequences are cumbersome and not easy. Furthermore, to perform these basic tasks, a user must be good at computer and programming skills. The Epi-gene can perform these FASTA format-related files in no time and require little time. In the case of Epi-gene, such joining and relabeling can be performed easily even if the user does not require advanced knowledge of programming. The Epi-gene package can calculate all the information related to pan-genomes, for instance, summary of pan-genome, median number of genes, set of core, and accessory and unique genes. The basic key to this calculation is the absence- or presence-based matrix. In other R-packages, up to author knowledge, only micropan is the package that

can construct a pan-matrix. The micropan is a fine approach towards pan-genomic study, but it uses BLAST which is slow and requires a long time [16].

Based on the binary pan-matrix, a pan-genome tree can also be constructed to estimate the phylogenetic relationship. This kind of tree demonstrates the difference in the number of gene clusters between genomes. There could be a variation between software regarding the tree construction as the distance calculation methods or clustering methods change. But overall results remain the same. In Epi-gene, no further functions were developed in the current version for pan-matrix based on sequence identity, as there are multiple packages already present that can handle this quantitative continuous data in a better way. For the present study,

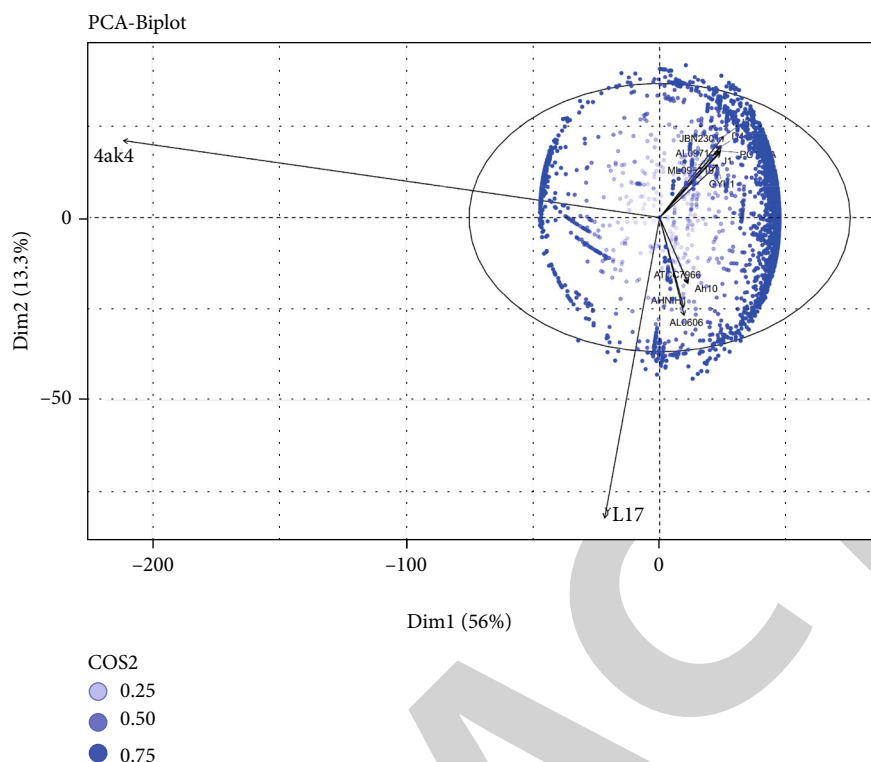


FIGURE 7: PCA-based biplot describing the genomes and homogenous gene clusters (colors filled based on cos2 character of variables).

the FactoMineR package was utilized. This package is solely meant for PCA calculations and graphical representations of the data [23]. Therefore, users are free to analyze this kind of data with multiple solutions.

Currently, Epi-gene is fully functional in Windows operating systems. Some functions in this package utilize the system commands to direct the Usearch for clustering functions. But, in the future, it is intended to design some more functions that will enable this package to work completely on LINUX operating systems.

5. Conclusion

Epi-gene is a promising functional package in R-statistical language with less time consumption and multiple graphical features. Furthermore, FASTA format handling functions will be helpful in studying sequences in R-language. A graphically clustered dendrogram showed more detailed information regarding genome relatedness. In the future, a recent version of this package will be updated according to future demands.

Data Availability

This package is freely available at the github repository (<http://furqan915.github.io/Epi-gene/>). The datasets generated during the current case study are available from the corresponding authors on reasonable request.

Conflicts of Interest

The authors declare that they have no competing interests.

Acknowledgments



This study was funded by the National Nature Science Foundation of China (31372454), the Jiangsu fishery science and technology project (D2017-3-1), the Independent Innovation Fund for Agricultural Science and Technology of Jiangsu Province (CX(17)2027), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

- [1] B. J. Traynor, "The era of genomic epidemiology," *Neuroepidemiology*, vol. 33, no. 3, pp. 276–279, 2009.
- [2] M. M. Ali, M. Hamid, M. Saleem et al., "Status of bioinformatics education in South Asia: past and present," *Bio Med Research International*, vol. 2021, article 5568262, pp. 1–9, 2021.
- [3] E. N. Gordienko, M. D. Kazanov, and M. S. Gelfand, "Evolution of pan-genomes of *Escherichia coli*, *Shigella* spp., and *Salmonella enterica*," *Journal of Bacteriology*, vol. 195, no. 12, pp. 2786–2792, 2013.
- [4] O. Lukjancenko, T. M. Wassenaar, and D. W. Ussery, "Comparison of 61 sequenced *Escherichia coli* genomes," *Microbial Ecology*, vol. 60, no. 4, pp. 708–720, 2010.
- [5] L. D. Alcaraz, *Pan-genomics: unmasking the gene diversity hidden in the bacteria species*, Peer J Pre Prints, 2014.

Research Article

Structural and Evolutionary Adaptation of NOD-Like Receptors in Birds

Xueting Ma ^{1,2}, Baohong Liu ^{1,2}, Zhenxing Gong ^{1,2}, Xinmao Yu ^{1,2},
and Jianping Cai ^{1,2}

¹State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Xujiaping 1, Lanzhou, Gansu Province 730046, China

²Jiangsu Co-Innovation Center for Prevention and Control of Animal Infectious Diseases and Zoonoses, Yangzhou, Jiangsu Province 225009, China

Correspondence should be addressed to Jianping Cai; caijianping@caas.cn

Received 25 February 2021; Revised 7 April 2021; Accepted 20 April 2021; Published 30 April 2021

Academic Editor: Hafiz Ishfaq Ahmad

Copyright © 2021 Xueting Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

NOD-like receptors (NLRs) are intracellular sensors of the innate immune system that recognize intracellular pathogen-associated molecular patterns (PAMPs) and danger-associated molecular patterns (DAMPs). Little information exists regarding the incidence of positive selection in the evolution of NLRs of birds or the structural differences between bird and mammal NLRs. Evidence of positive selection was identified in four avian NLRs (NOD1, NLRC3, NLRC5, and NLRP3) using the maximum likelihood approach. These NLRs are under different selection pressures which is indicative of different evolution patterns. Analysis of these NLRs showed a lower percentage of codons under positive selection in the LRR domain than seen in the studies of Toll-like receptors (TLRs), suggesting that the LRR domain evolves differently between NLRs and TLRs. Modeling of human, chicken, mammalian, and avian ancestral NLRs revealed the existence of variable evolution patterns in protein structure that may be adaptively driven.

1. Introduction

Nod-like receptors (NLRs) are pivotal sensor proteins of the innate immune system with diverse functions. They detect pathogen-associated molecular patterns (PAMPs) of invading microbes and danger-associated molecular patterns (DAMPs), thus initiating an innate immune response.

Structurally, members of the NLR family share a similar tripartite domain organization: a variable N-terminal domain, a central nucleotide-binding and oligomerization domain (NACHT), and a C-terminal domain [1].

NLRs are grouped into subfamilies based on their specific N-terminal protein-protein interaction domain: the caspase recruitment domain (CARD), the pyrin domain (PYD), the baculovirus inhibitor of apoptosis protein repeat (BIR) domain, or an acidic transactivation domain [1, 2]. These N-terminal domains are involved in recruiting downstream

effector molecules [2] and signal transduction. Moreover, the N-terminal region of an additional subfamily, NLRX1, contains a mitochondrial targeting sequence that shares no homology to any other protein [3, 4]. The central NACHT domains are similar to the STAND (signal-transduction ATPases with numerous domains) subclade of the AAA+ ATPases superfamily [5–7]. The central NACHT domain is required for protein oligomerization [1, 8].

NLRs detect PAMPs via a C-terminal leucine-rich repeats (LRR), which is a 20–29 residue repeated sequence with conserved and characteristically spaced hydrophobic residues [9, 10]. Recently, about 23 NLRs and NLR-like proteins have been reported in humans [11]. NLRs, including NLRP1, NLRP3, NLRP6, NLRP7, NLRC4, NLRP12, and NAIP, have been observed to activate the assembly of inflammasomes and mediate caspase-1 activation, leading to inflammation and cell pyroptosis [12, 13]. Other NLRs, such as NOD1, NOD2,

NLRC3, NLRC5, NLRP10, NLRX1, and CIITA, have been reported to induce the activation of nuclear factor- κ B (NF- κ B) and the mitogen activated protein kinase (MAPK) signaling pathway, or act as transcriptional regulators [12, 13].

Most studies on NLRs have focused on NLRs of mammals and fish. Recent studies on NLRs of early diverging organisms have suggested adaptive evolution in NLRs. In *Hydra*, which have large and complex NLR repertoires, NLRs recruit downstream adaptor molecules after activation. These NLRs, with the associated adaptor proteins, may induce apoptosis or activate putative NF- κ B/JNK transcription factors, thus regulating downstream cell responses and the expression of antimicrobial peptide [14]. It has been suggested that NLRs are ancient genes with putative cytoplasmic defense functions in basal metazoans and a common metazoan ancestor [14]. Gene expansion and domain gain, loss, and shuffling have occurred in the NLRs of *Hydra* and many other animals, indicating that NLRs have evolved in a species-specific manner to adapt to various ecological niches [14, 15].

In recent years, several important structures and functions of NLRs have been discovered in mammals. However, due to the different environmental conditions and ecological niches occupied by mammals and birds, conceivably their NLRs have evolved differently, generating unique structural and functional properties. Although evolution of the NOD subfamily of NLRs is known to be conserved in mammals, NOD2 is missing in chickens. In mammals, NOD2 senses specific bacterial muramyl dipeptides (MDP). In contrast to NOD2 in mammals, MDP are recognized by NLRP3 in chickens, suggesting that chicken NLRP3 replaces the role of mammalian NOD2 [16]. In mammals, NLRC5 is an important regulator of the MHC class-I antigen presentation pathway [17, 18]; however, one report showed no direct relationship between NLRC5 knockdown and MHC-I expression in chickens [19]. Moreover, the NLRP3 PYD domain, which is missing in *Xenopus* and zebrafish, has been identified in chickens [20]. Collectively, these observations suggest a pattern of molecular evolution of NLRs in vertebrates that is different from birds. Thus, the innate immune systems in all organisms, from *Hydra* and corals, to fish, birds, and mammals, have evolved in response to environmental conditions.

The aim of this work was, therefore, to identify evidence of positive selection in avian NLRs, to examine the structural and functional evolution of NLRs in birds, and to further elaborate the structural and functional diversity of NLRs between mammals and birds. Using Maximum Likelihood method, evidences of long-term selective pressure in the NLR genes have been found. To better understand how the CARD-binding properties of NLRs structurally and functionally diversified, ancestral NLR proteins were reconstructed and the structures of ancestral CARDS compared. Data from this study may provide more evidence for the role of positive selection in the evolution of NLRs.

2. Results

Our study identified some positive selection sites in the pathogen recognition domains of NLRs that were examined. To evaluate the functional significance of the inferred positively

selected sites, the location for each putatively positive site has been summarized in supplementary materials (Table 1). Our results provide evidence of positive selection in the NACHT domain of the avian NOD1, NLRC3, NLRP3, and NLRC5 (supplementary materials Table 1 and Figure 1). The NACHT domain plays a crucial role in the NLRs. It contains several characteristic motifs, namely, Walker A, Walker B, Sensor 1, and WH motif [21].

It has been reported that the sequence variations in the NACHT domain in human NLRs, especially in the vicinity of conserved regions or motifs, may influence the cycle of nucleotide-binding, -hydrolysis, -release, and/or conformational changes induced by NTP-hydrolysis, thus leading to inflammatory disorders [22, 23]. To investigate the effects of positive selection on the NACHT domain in four molecules, the positions of conserved motifs and residues under selective pressure were analyzed (Figure 1). None of observed positively selected codons were located within or close to the conserved motifs required for ATPase activity, activation, and oligomerization. As shown in the multiple sequence alignment between the *Gallus* and human sequences (Figure 1), motifs involved in NACHT functions are conserved among NLR proteins.

2.1. NLRC5. The quantity and distribution of positively selected codons varied among NLRs. The highest accumulation of positively selected sites occurred in NLRC5. Many sites under positive selection were located in N-terminal and C-terminal domains of the avian NLRC5. Remarkably, NLRC5 lacked the canonical N-terminal domain present in other defined domains of NLRs [24]. The N-terminal domain of NLRC5 is considered an atypical CARD. NLRC5 has a bipartite nuclear localization signal (NLS) which is required for nuclear localization of NLRC5 [17]. The NLS, the transport signal for nuclear protein import [25], is highly conserved in mammals and contains critical residues, i.e., Lys121, Arg122, Arg132, Arg133, and Lys134 in humans [26, 27]. Surprisingly, our sequence alignment indicates that the NLS (amino acids 121 to 134) was mostly absent in the *Gallus* and the avian ancestor NLRC5 sequences (Figure 2).

The highest accumulation of positively selected sites was observed in the LRR region of NLRC5 (Figure 3 and supplementary materials Table 1). NLRC5 contains a large LRR region that is different from other NLRs, with varying number of LRRs [26, 28]. To gain more insight into the functional significance of the putatively selected sites in LRRs, LRRSearch (<http://www.lrrsearch.com>) [29] was used to analyze the number and the location of the LRR of the chicken NLRC5. In general, the LRR domain contains 20–29 amino acid residues. It has a highly conserved segment (HCS) that consists of a consensus sequence and a variable segment (VS), which is located before the HCS of the next LRR. The HCS contains the consensus sequence LxxLxLxxN/CxL [10, 30, 31]. Of the 39 codons identified under positive selection in the LRRs of NLRC5, only 5 were localized in the HCS.

2.2. NOD1. The percentage of NOD1 sites under positive selection that were located in the known functional domains

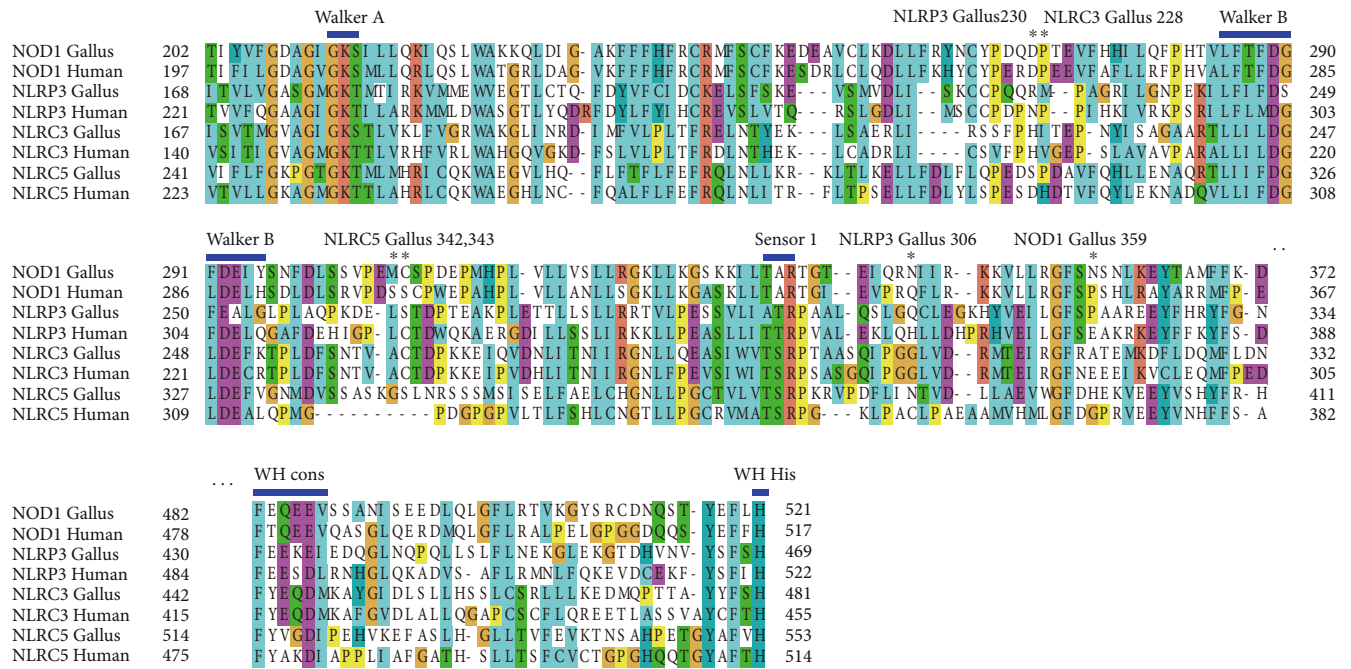


FIGURE 1: Adaptive substitutions in avian NLRs NACHT domain. We have aligned the NACHT domain of NLRs (human and Gallus shown). Adaptive amino acid substitutions and important motif of NACHT are indicated along the top of the alignment. Adaptive amino acid substitutions in NACHT are represented with star (*); motifs involved in ATP binding and hydrolysis of NACHT are indicated along the top of the alignment.

was low. In this study, only two positively selected sites (93 and 359) have been found within the functional domains. Amino acids 93 and 359 were located within the CARD domain and NACHT domain of avian NOD1, respectively. In the avian NOD1, Cys93 was replaced by Tyr, His, Arg, Leu, Phe, and Ser. Because the crystal structure of the human NOD1 CARD has been resolved and residues involved in the downstream signaling and interaction with RIP2 are recognized [32], we aligned the sequences of NOD1 CARD domain and plotted the CARD domain structure and electrostatic potential for human and *Gallus* NOD1 as well as for reconstructed ancestral NOD1 proteins (Figure 4). The results obtained henceforth suggested that the positively selected site 93 is not a key residue in the NOD1-RIP2 interaction. Since a similar degree of positive selection in interacting molecules is expected, we analyzed the positively selected codons in RIP2 (supplementary materials Table 1). Amino acids 324, 371, and 412 were identified to be under positive selection but were not located within the CARD domain of RIP2.

2.3. *NLR3 and NLRP3*. It has been reported that NLR3, together with NOD1, originated from a gene duplication event that occurred before the divergence of birds and mammals [33]. A minimum number of adaptive substitutions were present within the avian NLR3s (supplementary materials Table 1 and Figure 3), showing a stabilized selection pattern of evolution. Although the functions of many substitutions are unknown, they would support the hypothesis that species-specific adaptations occur because of different environmental conditions and pathogens. NLR3 participates in the assembly of inflammasome complexes.

Mutations in NLRP3 are associated with many inflammatory diseases in humans [34]. Our analysis of avian NLRP3s detected many amino acid sites that are under positive selection (supplementary materials Table 1 and Figure 3).

3. Discussion

Evidence of positive selection in the bird NLRs is reported in this study with variable quantity and distribution of positively selected residues.

3.1. *NOD1*. NOD1 is a crucial molecule in innate immunity. It activates the NF- κ B pathway by recruiting the protein receptor-interacting protein 2 (RIP2) through the interaction of CARDs. Two positively selected codons have been found in the functional domain of the *Gallus* NOD1, amino acids 93 and 359. Interactions between NOD1 and RIP2 mainly depend on electrostatic interactions in the CARD domain [32]. The CARD domain, as an important effector domain, belongs to the death domain superfamily. Members of the death domain superfamily always promote homotypic and/or heterotypic interactions with other effector domain containing proteins.

In NOD1, the CARD domain is indispensable for downstream signaling, involved in many different cellular processes, such as apoptosis and inflammation. Because of functional constraints, residues with critical functions may be under slow evolutionary rate [35]. It has been reported that the core residues which constitute the polar surface and the hydrophobic core of CARD are conserved [36]. As expected, the CARD domain of NOD1 had a lower percentage of

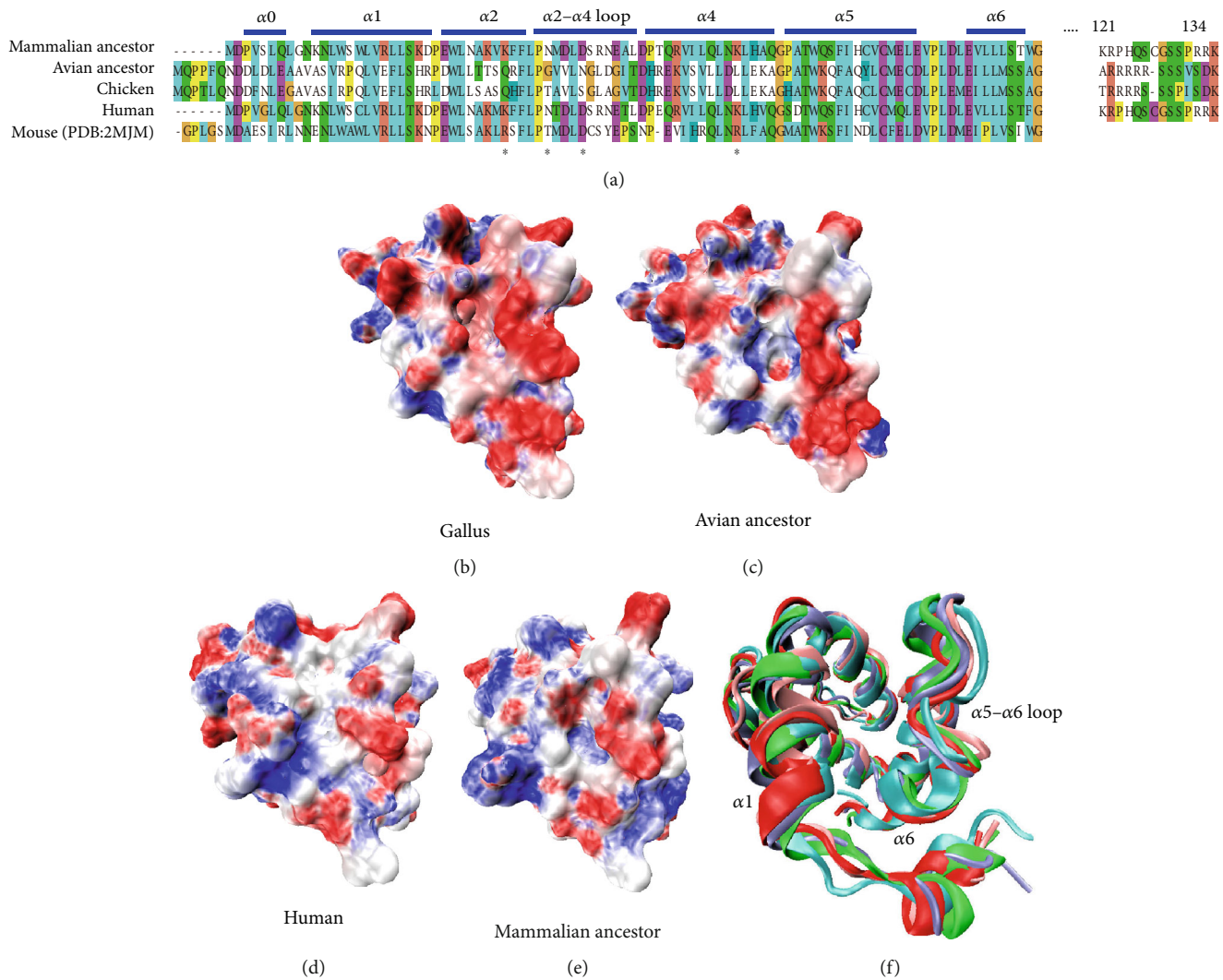


FIGURE 2: Sequence and primary functional differences in NLR5 CARD are established (human, chicken, mouse, mammalian ancestor, and avian ancestor). (a) Sequence alignments of NLR5 CARD domain. Adaptive amino acid substitutions in birds are represented with star (*). Secondary structures of protein are shown along the top of the alignment. NLS (amino acids 121 to 134 in human) is shown. (b–e) The atypical CARD of NLR5 consists of five α -helices ($\alpha 1$, $\alpha 2$, $\alpha 4$, $\alpha 5$, and $\alpha 6$) that are packed around a hydrophobic core. The CARD shape and electrostatic potential for human, chicken, mouse, mammalian ancestor, and avian ancestor are plotted. The surfaces are color-coded according to electrostatic surface potential: red, -10 kT; white, 0 kT; and blue, $+10$ kT. (f) Structural alignment of human, chicken, mammalian ancestor, and avian ancestor NLR5 atypical CARD.

positively selected codons than other NLRs, and it showed greater selective constraints.

To have further insight of the functional evolution of the NOD1 CARD and to understand if the positively selected residue (residue 93, located in $\alpha 5$) has the potential to influence the function of the CARD, we predicted the ancestral amino acid state of the CARD for mammals and birds and analyzed its 3D structure. The CARD domain of human NOD1 consists of six antiparallel α -helices packed around a hydrophobic core with residues E53, D54, and E56 of $\alpha 3$ involved in the interaction with RIP2 [32]. It has been reported that the surface shape of CARD and electrostatic interaction of oppositely charged residues plays a significant role in CARD-CARD interactions [32, 36, 37]. Mutations in L44, I57, and V41 greatly reduced activation of NOD1

signaling which demonstrated the significance of hydrophobic residues [32].

Our alignment results (Figure 4) showed that the ancestral CARD of mammals and avian species are conserved. Not only are the residues buried in the hydrophobic core and involved in interactions of the avian ancestor relatively well conserved but the shape and electrostatic potential of the binding area of the avian ancestor are very similar to human and chicken NOD1 (Figure 4). In human and mammalian ancestor, the NOD1 CARD domain residue 93 was replaced by Tyr, but the electrostatic distribution in this region was not altered (Figure 4). We examined the location of amino acid 93 and found that it was buried, leaving the electrostatic distribution on the surface of CARD unchanged. Similarly, analysis of avian RIP2 showed that the residues

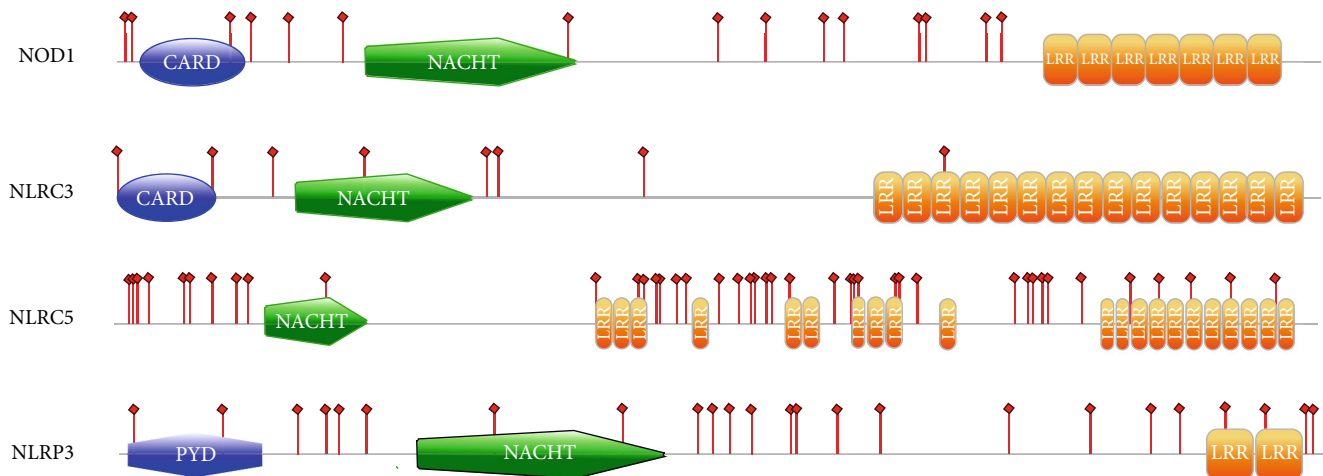


FIGURE 3: Location of positively selected sites of four NLR molecules in birds. Protein architectures were generated with PROSITE MyDomains image creator tool (<http://www.expasy.org/tools/mydomains/>). Red bars indicate adaptive amino acid substitutions. Colored areas represent different conserved domains (blue, CARD or PYD domain; green, NACHT domain; orange, LRR domain).

(R486, R525, and R530) of the RIP2 CARD domain involved in the interaction with NOD1 CARD were not under positive selection (supplementary materials Table 1), and these residues were conserved between avian and human (supplementary materials S1). This is consistent with its important biological function.

3.2. NLRC5. The N-terminal domain of NLRC5, which contains repeated α -helices and is strikingly distinct from other CARDS, is referred to as an atypical CARD. To investigate the structural differences of atypical CARDS, we compared the human, chicken, mammalian, and avian ancestral structures using homology modeling on an atypical mouse CARD, whose conformation is known [38]. Atypical CARD of NLRC5 consists of five α -helices that are packed around a hydrophobic core, with the difference to other CARDS (such as NOD1 CARD) being the absence of α_3 , which was replaced with an α_2 - α_4 loop. Interestingly, other studies of NOD1 CARD have suggested conserved α_2 and α_3 domains which form a putative interaction surface with the CARD of binding partner [36]. In our study, no positive selection is identified in α_2 and α_3 of CARD domains in NOD1 (positive selection 93 codon is in CARD α_5 of NOD1). NLRC5, however, also contacts with the binding partner by CARD-CARD interaction and contains four positively selected sites that fall in the α_2 , α_4 , and α_2 - α_4 loop of atypical CARD domain (supplementary materials Table 1 and Figure 2).

NOD1 interacts with its binding partner RIP2 by CARD-CARD interaction mainly through complementary surface shape and charge [38]. The hydrophobic interactions, however, make contributions to CARD-CARD interactions between NLRC5 and its partner RIG-I [38]. Moreover, the atypical CARD of NLRC5 interacts with the binding partner's CARD by its hydrophobic surface which is composed of α_1 , α_6 , and the α_5 - α_6 loop but not the α_2 and α_2 - α_4 loop [38]. Our alignment (Figure 2) showed that, except residues belonging to α_0 , α_2 , α_4 , and the α_2 - α_4 loop, most residues of the atypical CARD were conserved in chicken, human,

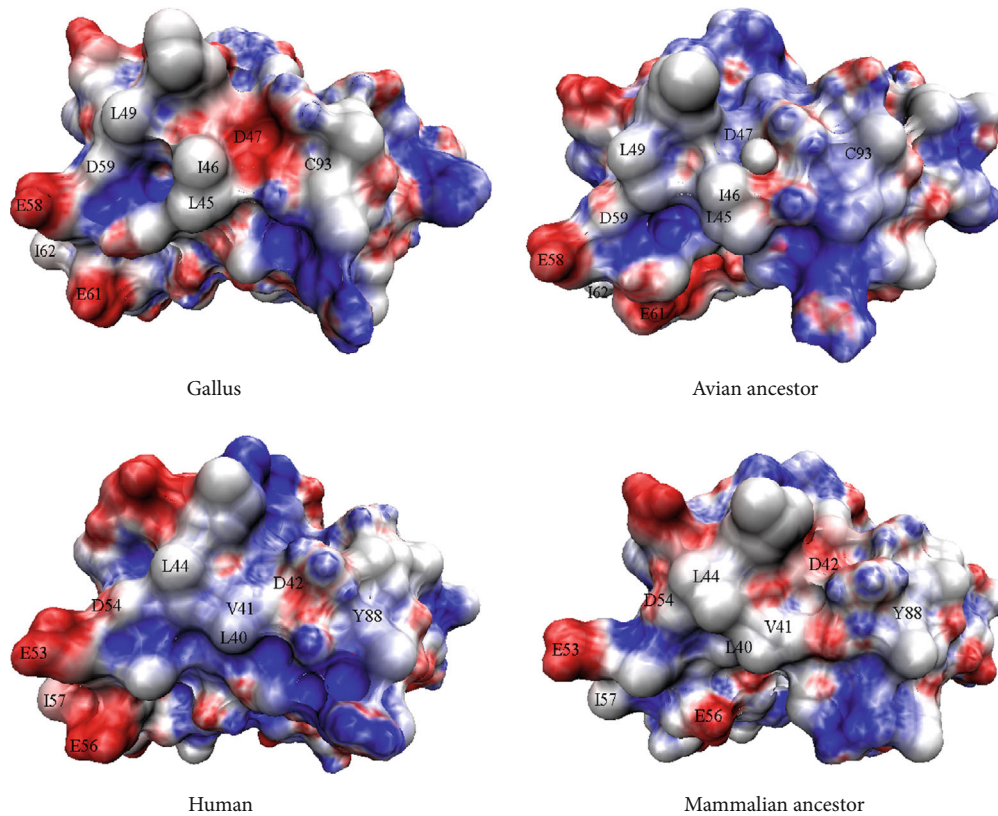
mouse, and avian and mammalian ancestors. For avian, positively selected sites mainly fall in the α_2 , α_4 , and α_2 - α_4 loop of NLRC5 (Figure 2); the α_1 , α_6 , and the α_5 - α_6 loop of NLRC5 that encode the RIG-I-binding region were more conserved than the α_2 and the α_2 - α_4 loop. A possible explanation may be that the residues in the RIG-I-binding region are directly involved in downstream signal transduction. The 3D structures of atypical CARDS were generally similar in the human, chicken, mouse, mammalian, and avian ancestors (Figure 2). But the electrostatic potentials of the pocket formed by α_1 , α_6 , and the α_5 - α_6 loop in the chicken and the avian ancestor are different from both the human and mammalian ancestral atypical CARDS which indicates a different evolution pattern of NLRC5 between birds and mammals.

Our additional alignment studies showed that the NLS (amino acids 121 to 134) was missing in the CARDS of bird NLRC5s (Figure 2). Sequence analysis of the chicken NLRC5 performed by cNLS Mapper [39] revealed a putative monopartite NLS at amino acids 1012-1025, which do not belong to CARD domain. In the avian ancestor, the NLS has been found absent using the cNLS Mapper prediction algorithm. The NLS is the transport signal for nuclear protein import [25] and is well conserved in mammals.

NLRC5 is an IFN- γ -inducible nuclear protein recognized as an important molecule for the MHC class-I antigen presentation pathway in humans [17, 18, 24]. Within the NLS sequence context, NLRC5 associates with promoters of MHC class-I genes and activates them, inducing expression of MHC class-I [17, 18]. Mutations within the NLS may prevent nuclear localization [17] and thereby reduce induction of MHC class-I expression [27]. It has, however, been reported that in chicken macrophages, there is no difference in MHC class-I gene expression between NLRC5 knockdown cells and controls, suggesting that NLRC5 may be a dispensable regulator of MHC class-I in chicken macrophages [19]. In zebrafish, overexpression of NLRC5 induces the expression of MHC class-II genes but not MHC class-I genes [40].

Gallus	26	ALLKVVYRELLVSKI	RHTQCLI	DNLI	NNEYFS	TEDA	EVVQ	PTQADKVRKI	LDLVQS	KGEEVS	EYFI	CVL	QKVT	DAYYEL	QPWLD	110
Avian ancestor	26	ALLKVVYRELLVSR	RNTQCLI	DNLI	KNDYFS	TEDA	EVVQ	PTQADKVRKI	LDLVQS	KGEEVS	EYFI	CVL	QKVT	DAYYEL	QPWLD	110
Human	21	QLLKS	NRELLVTHI	RNTQCLVDNLL	KNDYFS	AEDA	EVVQ	PTQADKVRKI	LDLVQS	KGEEVS	EFFLYLL	QQLAD	AYVDLR	PWLL	105	
Mammalian ancestor	21	KLLKI	NRELLVTHI	RNTQCLVDNLL	KNDYFS	AEDA	EVVQ	PTQADKVRKI	LDLVQS	KGEEVS	EFFLYVL	QQLAD	AYVDLR	PWLS	105	

(a)



(b)

FIGURE 4: Sequence and primary functional differences in NOD1 CARD. (a) Alignment of NOD1 CARD domain sequences. Star (*) indicates adaptive amino acid substitutions of NOD1 CARD in birds. (b) The NOD1 CARD shape and electrostatic potential for human, chicken, mouse, mammalian ancestor, and avian ancestor are plotted; the surfaces are color-coded according to electrostatic surface potential: red, -10 kT; white, 0 kT; and blue, $+10$ kT. Residues reported to be important for NOD1/RIP2 interaction, and NF- κ B activation in human is labeled.

These findings indicate the functional divergence of NLRC5 between mammals, birds, and fishes.

3.3. NLRC3. A minimal number of adaptive substitutions have been found within NLRC3 (supplementary materials Table 1 and Figure 3), suggesting a stabilized selection pattern in evolution. It is difficult to speculate about the impact of functional effects caused by substitutions in NLRC3 without an available crystal structure. Recent studies have suggested that NLRC3 functions to regulate the strength and time of the inflammatory response, suppress the activation of various innate immune signaling cascades, and prevent excessive immune responses [41, 42]. NLRC3 can regulate the STING signaling pathway and the TRAF6 signaling pathway. Also, NLRC3 negatively modulates STING activation via the response to cytosolic DNA, cyclic di-GMP, and DNA viruses [41]. Moreover, NLRC3 negatively regulates the activation of NF- κ B through

interaction with TRAF6 [42]. It has been demonstrated that an interaction between the NACHT domain of NLRC3 and STING is mediated by the LRR domain [41]. Thus, NLRC3 may decrease STING dependent innate immune activation [41]. The NACHT domain of NLRC3 also affects the NF- κ B pathway by binding with TRAF6 [42]. NLRC3 binds to TRAF6 via the TRAF-binding motif in the NACHT domain [42]. NLRC3 has two TRAF-binding motifs that are conserved among various species [42]. Our analysis found conserved TRAF2-binding sites in birds (Figure 5). NLRC3 has been reported to be an inhibitor of the PI3K-AKT-mTOR pathways [43]. Moreover, an NLRC3-like protein of zebrafish has a negative regulatory function on macrophage activation and inflammation. It has been reported that both the PYD and the NACHT domains are essential for function of the NLRC3-like protein in zebrafish, and loss of function mutations in the zebrafish NLRC3-like protein may result in systemic inflammation [44]. This provides support for cross-



FIGURE 5: WebLogo of the putative TRAF2-binding sites constructed from alignment of 41 birds NLRC3 sequences. The residue numbering corresponds to the residues from chicken NLRC3. TRAF-binding motif mainly contains four residues with (P/S/A/T)-X-(Q/E)-E in the NACHT domain; “P/S/A/T” represents Pro, Ser, Ala, or Thr; “X” indicates any amino acid; “Q/E” represents Gln or Glu. Major TRAF2-binding motifs in birds NLRC3 are at the region of 483-486 and 608-611. The letter size is proportional to the degree of amino acid conservation. The weblogo was made using the web-based application WebLogo (<http://weblogo.berkeley.edu>).

species functionality of NLRC3 [43]. Hence, it is possible that mutations in NLRC3 are likely to influence inflammatory signaling pathways and even induce inflammatory disorders. These studies, collectively, provide reasonable evidence that NLRC3 evolved under strong stabilizing selection.

3.4. NLRP3. NLRP3 is a member of the NLR family and belongs to the NLRP subclass characterized by the presence of a PYD. Human NLRP3 is activated in response to a broad range of stimuli including bacteria, fungi, yeast, viruses, parasites, pore-forming toxins, crystals, TLR ligands, bacterial RNA, and DAMPs, such as ATP and hyaluronan [45–48]. NLRP3 participates in the assembly of inflammasome complexes, which in most cases, help the host eliminate invading pathogens. Aberrant accumulation of inflammasome signals may result in disease in humans. Mutations in and around the NACHT of NLRP3 cause three auto-inflammatory diseases: FCAS (familial cold auto-inflammatory syndrome), MWS (Muckle-Wells Syndrome), and CINCA/NOMID (chronic infantile neurological cutaneous and articular syndrome/neonatal onset multisystemic auto-inflammatory disease) [34]. Several amino acids showed evidence of positive selection that were detected in or around the NACHT domain of NLRP3 in the chicken and avian ancestor sequences we examined (Figures 1 and 3).

A relatively recent study has demonstrated that the mutations may affect the stabilization of NLRP3 by increasing its half-life [49], showing that the NLRP3 protein, stable at lower temperatures, was degraded at 37°C [49]. In avian species, high metabolic rates may lead to high body temperatures—the mean body temperatures for all birds has been reported to be 38.54°C (± 0.96) for resting birds, 41.02°C (± 1.29) for birds in the active phase, and 43.85°C (± 0.94)

for highly active birds [50]. These findings suggest that positive selection found in or around the NACHT domain of bird NLRP3 may be associated with adaptation to a higher body temperature in birds. In the chicken, NOD2 was replaced with NLRP3 as another potential PRR to sense MDP [16]. Recent studies have reported that the NLRP3 gene is variable between mammalian and avian species [51]. The PYD domain of NLRP3 was not identified in *Xenopus* or the zebrafish [20]. Another inflammation-related molecule, IL-1 β , has also been identified to be highly variable between mammalian and avian species, suggesting potentially different mechanisms in the host inflammatory responses between these classes [51].

3.5. LRR Domains of NLRs. Previous research has shown a high number of positively selected codons located in LRR domain of TLRs [52]. Our results, however, indicate a different level of positive selection acting in LRR domain of NLRs. NLRC5, NOD1, NLRC3, and NLRP3 showed lower percentage of codons under positive selection in the LRR domain than TLRs. The TLRs are transmembrane proteins which recognize PAMPs through the extracellular domain of LRRs. The NLRs are intracellular, cytoplasmic sensors. Recent studies have emphasized that interactions between NLRs and PAMPs may be indirect and involve intermediates of host cells, which like R proteins indirectly sense PAMPs in plants [1, 53, 54]. This may explain the different patterns of evolution observed for LRRs in TLRs and NLRs. NLRC5 contains a large LRR region that is different from other NLRs, with the exact number of LRRs being uncertain [26, 28]. LRRs of the human NLRC5 are involved in the interaction with IKK α and IKK β , blocking phosphorylation and degradation of the inhibitory I κ B proteins, thus, inhibiting NF- κ B activation

[55]. In addition, the LRR region of human NLRC5 is responsible for nuclear export of NLRC5, as well as transcriptional activation [56]. Of the 39 codons identified under positive selection in the LRRs of the chicken and avian ancestral NLRC5, only 5 amino acids were localized in the HCS. Although the number of LRRs in NLRC5 is still controversial, the C-terminal LRR repeat of human NLRC5 has been reported to be of 36 amino acid residues in length [26] while others have suggested that the C-terminal capping motifs may serve to protect the protein [57]. Interestingly, recent studies have also described an NES (nuclear export signal) in the C-terminal region of LRRs that mediates nuclear export in humans [17].

It is reasonable to predict conserved evolution for this region. Our analyses mapped one codon under positive selection (residue 1800), which falls in the region closest to the C-terminal LRRs. Because of the large size of the LRR region and unusual structure of NLRC5, it has been speculated that NLRC5 might sense different stimuli, compared to other NLRs [28]. Some of the positively selected amino acid sites in LRRs may, thus, be involved in the recognition of these diverse stimuli. The detailed molecular basis of the role of NLRC5 in host defense and immune signaling, however, is largely still unknown.

4. Conclusion

The Nod-like receptors of the innate immunity represent the first line of defense against the pathogens. These are involved in pathogen recognition and, thus, need to evolve rapidly in a dynamic arms race with pathogens. Evidence of positive selection of NLRs in birds has been observed. Adaptive selection in avian NLRs was different than either avian TLRs or mammalian NLRs. NLRs have shown adaptive divergence in structure and function throughout the avian evolution. This work provides a basic understanding of structural and functional evolution in avian NLRs. By looking for structural differences between human, chicken, mammalian ancestral, and avian ancestral NLRs, we propose that the different environmental conditions encountered by mammals and birds might have induced structural and functional differentiation among members of the NLR family. Different NLRs with varied functions might have experienced unique pressures which might have induced the evolutionary change.

5. Materials and Methods

5.1. Data Collection. The coding regions of three chicken (Lingnanhuang: LNH) NOD-like-receptor genes, NOD1, NLRC3, and NLRP3, were obtained by polymerase chain reaction (PCR) amplification using gene-specific primers (primer details in supplementary materials S2). Chicken NLRC5 and all avian NLR sequences (accession number displayed in supplementary materials S3) were collected from NCBI (<http://www.ncbi.nlm.nih.gov>). NOD-like-receptor amino acid sequences of chicken, avian ancestors, and mammalian ancestors are displayed in supplementary materials S2.

5.2. Sequence Alignment. Sequence alignments were produced using PROBCONS version 1.12 [58]. The alignments used for phylogenetic analyses were processed using Gblocks v0.91b [59] to detect and filter potentially unreliable and misaligned regions. To detect positive selection in individual codons of the NLR sequences, a phylogenetic tree was reconstructed by MrBayesv3.2.2 (<http://mrbayes.csit.fsu.edu/>) [60]. All analyses were run with 2000,000 generation in MrBayes. Convergence was considered to have been achieved with split frequency values of <0.01. If the split frequency did not drop below 0.01, the analysis continued with additional 2000,000 generations. The first 25% of the topologies were discarded as burn-in. Sequence analysis of the chicken NLRC5 was performed by cNLS Mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi) [39].

5.3. Tests of Selection. All sequences for testing positive selections have been displayed in supplementary materials S3. To test positive selections at NLRs, the ratios of nonsynonymous (dN) to synonymous (dS) substitutions per site were compared in a maximum likelihood (ML) framework. A ratio of dN/dS > 1 is interpreted as evidence of positive selection. In order to improve the accuracy of the positive selection results, two ML frameworks were selected, the codeml program of PAML [61] and the HyPhy package of the Data Monkey Web Server (<http://www.datamonkey.org>) [62]. For codeml, two alternative models (M7 and M8) were chosen and compared with twice the difference of log-likelihood value ($2\Delta\ln L$) with 2 degrees of freedom to investigate whether sites were under positive selection in each NLRs. For codeml, the Bayes Empirical Bayes (BEB) posterior probability method was calculated in conjunction with site models to identify individual codons as adaptive [63]. In HyPhy, three distinct models, SLAC, FEL, and REL, were conducted; the level of statistical significance was set at a p value = 0.1 for SLAC and FEL and Bayes Factor = 50 for REL analysis, sites with a p value < 0.1 for SLAC and FEL, and Bayes Factor > 50 for REL was accepted to identify candidates for selection.

Protein-coding substitutions identified as more than or equal to two ML methods were considered adaptive. In order to identify robust sites under positive selection, only sites with evidence of selection in at least two of the ML methods were considered. In addition, the PAL2NAL program [64] was used to convert the amino acid alignment into a codon-based DNA alignment for PAML codeml test.

5.4. Ancestral Sequence Reconstruction. Ancestral protein sequences were reconstructed using MEGA 6 [65] with the maximum likelihood algorithm. All nucleotide sequences were converted to corresponding protein sequences were aligned with PROBCONS. For each alignment, the best fitting nucleotide substitution model for each codon position was evaluated using the AIC criterion in MEGA with “find best DNA/Protein models.” An evolutionary tree was reconstructed with the ML method. For each internal node of the tree, MEGA exported a file including information of most probable ancestral sequences, and program ExtAncSeqMEGA [66] could then extract the ancestral DNA and protein

sequences from the file. The accuracy scores of ancestral DNA and protein sequences were estimated. All ancestral sequences have been displayed in supplementary materials S2.

5.5. Homology Modeling and Structural Analysis. To obtain a more precise idea of the functional evolution of avian NLRs, structural homology models are built with MODELLER v9.5 [67]. Then, the structures were processed using the PDB2PQR server (http://nbcrc-222.ucsd.edu/pdb2pqr_2.0.0/) [68] to add hydrogen atoms and force field parameters. Electrostatic surface potentials were estimated with APBS web solver (Adaptive Poisson Boltzmann Solver). The visualization of structures and electrostatic potential maps was generated with VMD1.9.1 [69].

Data Availability

GenBank accession numbers of chicken (Lingnanhuang) NLR nucleotide sequences (generated in the course of the study) are as follows: MT385526, MT385527, and MT385528. The chicken sequences (third-party data) and ancestral sequences are deposited in Supplementary File S2. The accession number of avian nucleotide sequences analyzed in the article is deposited in Supplementary File S3.

Additional Points

Significance Statement. Despite several studies on the evolution of TLRs in birds, a clear picture of the evolution of birds NLRs has not yet emerged. We have found evidence of positive selection of NLRs within the birds, reconstructed avian, and mammalian ancestral sequence of NLRs and compared the structural and functional differences of NLRs between mammals and avian. Different evolution patterns of the molecular binding region were specifically investigated in our study. Our analyses of avian NLR family provide more evidence for the role of positive selection in the evolution of NLRs.

Conflicts of Interest

The authors declare no commercial or financial conflict of interest.

Acknowledgments

We thank Dr. Patricia Wilkins of Parasitology Services, USA, for editorial assistance. We are also grateful to Prof. Xiaoqing Xia at Institute of Hydrobiology, Chinese Academy of Sciences, for his academic help. Furthermore, we also appreciate the contributors who deposited the sequences used in this analysis in GenBank. This work was supported by the Innovative Special Project of Agricultural Sci-Tech (Grant No. CAASASTIP-2014-LVRI-09 to JPC).

Supplementary Materials

S1: RIP2 alignment. S2: primers and chicken sequences and ancestral sequences. S3: avian sequences accession number. Supplementary Table 1: sites found to be under positive selection. (*Supplementary Materials*)

References

- [1] G. Chen, M. H. Shaw, Y.-G. Kim, and G. Nuñez, "NOD-like receptors: role in innate immunity and inflammatory disease," *Annual Review of Pathology: Mechanisms of Disease*, vol. 4, no. 1, pp. 365–398, 2009.
- [2] Z. Ye and J. P.-Y. Ting, "NLR, the nucleotide-binding domain leucine-rich repeat containing gene family," *Current Opinion in Immunology*, vol. 20, no. 1, pp. 3–9, 2008.
- [3] C. B. Moore, D. T. Bergstralh, J. A. Duncan et al., "NLRX1 is a regulator of mitochondrial antiviral immunity," *Nature*, vol. 451, no. 7178, pp. 573–577, 2008.
- [4] D. Arnoult, F. Soares, I. Tattoli, C. Castanier, D. J. Philpott, and S. E. Girardin, "An N-terminal addressing sequence targets NLRX1 to the mitochondrial matrix," *Journal of Cell Science*, vol. 122, no. 17, pp. 3161–3168, 2009.
- [5] T. A. Kufer and P. J. Sansonetti, "NLR functions beyond pathogen recognition," *Nature Immunology*, vol. 12, no. 2, pp. 121–128, 2011.
- [6] T. Maekawa, T. A. Kufer, and P. Schulze-Lefert, "NLR functions in plant and animal immune systems: so far and yet so close," *Nature Immunology*, vol. 12, no. 9, pp. 817–826, 2011.
- [7] Y. Kadota, K. Shirasu, and R. Guerois, "NLR sensors meet at the SGT1-HSP90 crossroad," *Trends in Biochemical Sciences*, vol. 35, no. 4, pp. 199–207, 2010.
- [8] T. P. Monie, C. E. Bryant, and N. J. Gay, "Activating immunity: lessons from the TLRs and NLRs," *Trends in Biochemical Sciences*, vol. 34, no. 11, pp. 553–561, 2009.
- [9] P. Rosenstiel, G. Jacobs, A. Till, and S. Schreiber, "NOD-like receptors: ancient sentinels of the innate immune system," *Cellular and Molecular Life Sciences*, vol. 65, no. 9, pp. 1361–1377, 2008.
- [10] T. Wei, J. Gong, F. Jamitzky, W. M. Heckl, R. W. Stark, and S. C. Rössle, "LRRML: a conformational database and an XML description of leucine-rich repeats (LRRs)," *BMC Structural Biology*, vol. 8, no. 1, p. 47, 2008.
- [11] I. C. Allen, "Non-inflammasome forming NLRs in inflammation and tumorigenesis," *Frontiers in Immunology*, vol. 5, p. 169, 2014.
- [12] F. Barbé, T. Douglas, and M. Saleh, "Advances in Nod-like receptors (NLR) biology," *Cytokine & Growth Factor Reviews*, vol. 25, no. 6, pp. 681–697, 2014.
- [13] H. Zhu and X. Cao, "NLR members in inflammation-associated carcinogenesis," *Cellular & Molecular Immunology*, vol. 14, no. 5, pp. 403–405, 2017.
- [14] C. Lange, G. Hemmrich, U. C. Klostermeier et al., "Defining the origins of the NOD-like receptor system at the base of animal evolution," *Molecular Biology and Evolution*, vol. 28, no. 5, pp. 1687–1702, 2011.
- [15] M. Hamada, E. Shoguchi, C. Shinzato, T. Kawashima, D. J. Miller, and N. Satoh, "The complex NOD-like receptor repertoire of the coral *Acropora digitifera* includes novel domain combinations," *Molecular Biology and Evolution*, vol. 30, no. 1, pp. 167–176, 2013.
- [16] F. Martinon, L. Agostini, E. Meylan, and J. Tschopp, "Identification of bacterial muramyl dipeptide as activator of the NALP3/cryopyrin inflammasome," *Current Biology*, vol. 14, no. 21, pp. 1929–1934, 2004.
- [17] T. B. Meissner, A. Li, A. Biswas et al., "NLR family member NLRC5 is a transcriptional regulator of MHC class I genes,"

- Proc. Natl. Acad. Sci. USA*, vol. 107, no. 31, pp. 13794–13799, 2010.
- [18] Y. Yao, Y. Wang, F. Chen et al., “NLRC5 regulates MHC class I antigen presentation in host defense against intracellular pathogens,” *Cell Research*, vol. 22, no. 5, pp. 836–847, 2012.
- [19] L. Lian, C. Ciraci, G. Chang, J. Hu, and S. J. Lamont, “NLRC5 knockdown in chicken macrophages alters response to LPS and poly (I:C) stimulation,” *BMC Veterinary Research*, vol. 8, no. 1, p. 23, 2012.
- [20] K. J. Laing, M. K. Purcell, J. R. Winton, and J. D. Hansen, “A genomic view of the NOD-like receptor family in teleost fish: identification of a novel NLR subfamily in zebrafish,” *BMC Evolutionary Biology*, vol. 8, no. 1, p. 42, 2008.
- [21] M. Proell, S. J. Riedl, J. H. Fritz, A. M. Rojas, and R. Schwarzenbacher, “The Nod-like receptor (NLR) family: a tale of similarities and differences,” *PLoS One*, vol. 3, no. 4, article e2119, 2008.
- [22] P. Rosenstiel, A. Till, and S. Schreiber, “NOD-like receptors and human diseases,” *Microbes and Infection*, vol. 9, no. 5, pp. 648–657, 2007.
- [23] M. Albrecht, F. S. Domingues, S. Schreiber, and T. Lengauer, “Structural localization of disease-associated sequence variations in the NACHT and LRR domains of PYPAF1 and NOD2,” *FEBS Letters*, vol. 554, no. 3, pp. 520–528, 2003.
- [24] S. Benko, J. G. Magalhaes, D. J. Philpott, and S. E. Girardin, “NLRC5 limits the activation of inflammatory pathways,” *The Journal of Immunology*, vol. 185, no. 3, pp. 1681–1691, 2010.
- [25] A. Lange, R. E. Mills, C. J. Lange, M. Stewart, S. E. Devine, and A. H. Corbett, “Classical nuclear localization signals: definition, function, and interaction with importin α ,” *The Journal of Biological Chemistry*, vol. 282, no. 8, pp. 5101–5105, 2007.
- [26] J. A. Mótyán, P. Bagossi, S. Benkő, and J. Tózsér, “A molecular model of the full-length human NOD-like receptor family CARD domain containing 5 (NLRC5) protein,” *BMC Bioinformatics*, vol. 14, no. 1, p. 275, 2013.
- [27] T. B. Meissner, A. Li, Y.-J. Liu, E. Gagnon, and K. S. Kobayashi, “The nucleotide-binding domain of NLRC5 is critical for nuclear import and transactivation activity,” *Biochemical and Biophysical Research Communications*, vol. 418, no. 4, pp. 786–791, 2012.
- [28] A. Neerinx, K. Lautz, M. Menning et al., “A role for the human nucleotide-binding domain, leucine-rich repeat-containing family member NLRC5 in antiviral responses,” *The Journal of Biological Chemistry*, vol. 285, no. 34, pp. 26223–26232, 2010.
- [29] A. Bej, B. R. Sahoo, B. Swain, M. Basu, P. Jayasankar, and M. Samanta, “LRRsearch: an asynchronous server-based application for the prediction of leucine-rich repeat motifs and an integrative database of NOD-like receptors,” *Computers in Biology and Medicine*, vol. 53, pp. 164–170, 2014.
- [30] B. Kobe and A. V. Kajava, “The leucine-rich repeat as a protein recognition motif,” *Current Opinion in Structural Biology*, vol. 11, no. 6, pp. 725–732, 2001.
- [31] N. Matsushima, H. Miyashita, T. Mikami, and Y. Kuroki, “A nested leucine rich repeat (LRR) domain: the precursor of LRRs is a ten or eleven residue motif,” *BMC Microbiology*, vol. 10, no. 1, p. 235, 2010.
- [32] F. Manon, A. Favier, G. Núñez, J.-P. Simorre, and S. Cusack, “Solution structure of NOD1 CARD and mutational analysis of its interaction with the CARD of downstream kinase RICK,” *Journal of Molecular Biology*, vol. 365, no. 1, pp. 160–174, 2007.
- [33] A. L. Hughes, “Evolutionary relationships of vertebrate NACHT domain-containing proteins,” *Immunogenetics*, vol. 58, no. 10, pp. 785–791, 2006.
- [34] M. Lamkanfi and T.-D. Kanneganti, “Nlrp3: an immune sensor of cellular stress and infection,” *The International Journal of Biochemistry & Cell Biology*, vol. 42, no. 6, pp. 792–795, 2010.
- [35] B. Knudsen and M. M. Miyamoto, “A likelihood ratio test for evolutionary rate shifts and functional divergence among proteins,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 25, pp. 14512–14517, 2001.
- [36] J. J. Chou, H. Matsuo, H. Duan, and G. Wagner, “Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment,” *Cell*, vol. 94, no. 2, pp. 171–180, 1998.
- [37] H. Qin, S. M. Srinivasula, G. Wu, T. Fernandes-Alnemri, E. S. Alnemri, and Y. Shi, “Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1,” *Nature*, vol. 399, no. 6736, pp. 549–557, 1999.
- [38] P. G. M. Gutte, S. Jurt, M. G. Grütter, and O. Zerbe, “Unusual structural features revealed by the solution NMR structure of the NLRC5 caspase recruitment domain,” *Biochemistry*, vol. 53, no. 19, pp. 3106–3117, 2014.
- [39] S. Kosugi, M. Hasebe, M. Tomita, and H. Yanagawa, “Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 25, pp. 10171–10176, 2009.
- [40] X. M. Wu, Y. W. Hu, N. N. Xue et al., “Role of zebrafish NLRC5 in antiviral response and transcriptional regulation of MHC related genes,” *Developmental and Comparative Immunology*, vol. 68, pp. 58–68, 2017.
- [41] L. Zhang, J. Mo, K. V. Swanson et al., “NLRC3, a member of the NLR family of proteins, is a negative regulator of innate immune signaling induced by the DNA sensor STING,” *Immunity*, vol. 40, no. 3, pp. 329–341, 2014.
- [42] M. Schneider, A. G. Zimmermann, R. A. Roberts et al., “The innate immune sensor NLRC3 attenuates Toll-like receptor signaling via modification of the signaling adaptor TRAF6 and transcription factor NF- κ B,” *Nature Immunology*, vol. 13, no. 9, pp. 823–831, 2012.
- [43] R. Karki, S. M. Man, R. K. S. Malireddi et al., “NLRC3 is an inhibitory sensor of PI3K-mTOR pathways in cancer,” *Nature*, vol. 540, no. 7634, pp. 583–587, 2016.
- [44] C. E. Shiao, K. R. Monk, W. Joo, and W. S. Talbot, “An anti-inflammatory NOD-like receptor is required for microglia development,” *Cell Reports*, vol. 5, no. 5, pp. 1342–1352, 2013.
- [45] T.-D. Kanneganti, N. Özören, M. Body-Malapel et al., “Bacterial RNA and small antiviral compounds activate caspase-1 through cryopyrin/Nalp3,” *Nature*, vol. 440, no. 7081, pp. 233–236, 2006.
- [46] S. Mariathasan, D. S. Weiss, K. Newton et al., “Cryopyrin activates the inflammasome in response to toxins and ATP,” *Nature*, vol. 440, no. 7081, pp. 228–232, 2006.
- [47] F. Martinon, V. Pétrilli, A. Mayor, A. Tardivel, and J. Tschopp, “Gout-associated uric acid crystals activate the NALP3 inflammasome,” *Nature*, vol. 440, no. 7081, pp. 237–241, 2006.

- [48] S. B. Willingham, D. T. Bergstralh, W. O'Connor et al., "Microbial pathogen-induced necrotic cell death mediated by the inflammasome components CIAS1/cryopyrin/NLRP3 and ASC," *Cell Host & Microbe*, vol. 2, no. 3, pp. 147–159, 2007.
- [49] S. D. Brydges, J. L. Mueller, M. D. McGeough et al., "Inflammasome-mediated disease animal models reveal roles for innate but not adaptive immunity," *Immunity*, vol. 30, no. 6, pp. 875–887, 2009.
- [50] R. Prinzinger, A. Preßmar, and E. Schleucher, "Body temperature in birds," *Comparative Biochemistry and Physiology Part A: Physiology*, vol. 99, no. 4, pp. 499–506, 1991.
- [51] J. Ye, M. Yu, K. Zhang et al., "Tissue-specific expression pattern and histological distribution of NLRP3 in Chinese yellow chicken," *Veterinary Research Communications*, vol. 39, no. 3, pp. 171–177, 2015.
- [52] H. Areal, J. Abrantes, and P. J. Esteves, "Signatures of positive selection in Toll-like receptor (TLR) genes in mammals," *BMC Evolutionary Biology*, vol. 11, no. 1, p. 368, 2011.
- [53] T. Kawai and S. Akira, "The roles of TLRs, RLRs and NLRs in pathogen recognition," *International Immunology*, vol. 21, no. 4, pp. 317–337, 2009.
- [54] J. D. Jones and J. L. Dangl, "The plant immune system," *Nature*, vol. 444, no. 7117, pp. 323–329, 2006.
- [55] J. Cui, L. Zhu, X. Xia et al., "NLR5 negatively regulates the NF- κ B and type I interferon signaling pathways," *Cell*, vol. 141, no. 3, pp. 483–496, 2010.
- [56] A. Neerinx, G. M. Rodriguez, V. Steimle, and T. A. Kufer, "NLR5 controls basal MHC class I gene expression in an MHC enhanceosome-dependent manner," *The Journal of Immunology*, vol. 188, no. 10, pp. 4940–4950, 2012.
- [57] J. Bella, K. L. Hindle, P. A. McEwan, and S. C. Lovell, "The leucine-rich repeat structure," *Cellular and Molecular Life Sciences*, vol. 65, no. 15, pp. 2307–2333, 2008.
- [58] C. B. Do, M. S. Mahabhashyam, M. Brudno, and S. Batzoglou, "ProbCons: probabilistic consistency-based multiple sequence alignment," *Genome Research*, vol. 15, no. 2, pp. 330–340, 2005.
- [59] J. Castresana, "Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis," *Molecular Biology and Evolution*, vol. 17, no. 4, pp. 540–552, 2000.
- [60] J. P. Huelsenbeck, F. Ronquist, R. Nielsen, and J. P. Bollback, "Bayesian inference of phylogeny and its impact on evolutionary biology," *Science*, vol. 294, no. 5550, pp. 2310–2314, 2001.
- [61] Z. Yang, "PAML 4: phylogenetic analysis by maximum likelihood," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1586–1591, 2007.
- [62] S. L. Pond and S. D. Frost, "Datamonkey: rapid detection of selective pressure on individual sites of codon alignments," *Bioinformatics*, vol. 21, no. 10, pp. 2531–2533, 2005.
- [63] Z. Yang, W. S. Wong, and R. Nielsen, "Bayes empirical Bayes inference of amino acid sites under positive selection," *Molecular Biology and Evolution*, vol. 22, no. 4, pp. 1107–1118, 2005.
- [64] M. Suyama, D. Torrents, and P. Bork, "PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments," *Nucleic Acids Research*, vol. 34, no. Web Server, pp. W609–W612, 2006.
- [65] K. Tamura, G. Stecher, D. Peterson, A. Filipski, and S. Kumar, "MEGA6: molecular evolutionary genetics analysis version 6.0," *Molecular Biology and Evolution*, vol. 30, no. 12, pp. 2725–2729, 2013.
- [66] B. G. Hall, *Working with Various Computer Platforms Phylogenetic Trees Made Easy: A How-to Manual 5Edn*, Sinauer Associates Inc., U.S., New York, 2011.
- [67] A. Sali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints," *Journal of Molecular Biology*, vol. 234, no. 3, pp. 779–815, 1993.
- [68] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker, "PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations," *Nucleic Acids Research*, vol. 32, no. Web Server, pp. W665–W667, 2004.
- [69] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33–38, 1996, 27–38.

Retraction

Retracted: Identification of Hearing Loss-Associated Variants of *PTPRQ*, *MYO15A*, and *SERPINB6* in Pakistani Families

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] U. Mahmood, S. A. Bukhari, M. Ali, Z. M. Ahmed, and S. Riazuddin, "Identification of Hearing Loss-Associated Variants of *PTPRQ*, *MYO15A*, and *SERPINB6* in Pakistani Families," *BioMed Research International*, vol. 2021, Article ID 5584788, 6 pages, 2021.

Research Article

Identification of Hearing Loss-Associated Variants of *PTPRQ*, *MYO15A*, and *SERPINB6* in Pakistani Families

Umair Mahmood,¹ Shazia A. Bukhari ,¹ Muhammad Ali,² Zubair M. Ahmed,³ and Saima Riazuddin ³

¹Department of Biochemistry, Government College University, Faisalabad 38000, Pakistan

²Department of Animal Sciences, Quaid Azam University, Islamabad 46000, Pakistan

³Department of Otorhinolaryngology Head and Neck Surgery, University of Maryland School of Medicine, Baltimore, MD 21201, USA

Correspondence should be addressed to Shazia A. Bukhari; shaziabukhari@gcuf.edu.pk and Saima Riazuddin; sriazuddin@som.umaryland.edu

Received 17 February 2021; Revised 12 March 2021; Accepted 30 March 2021; Published 28 April 2021

Academic Editor: Hafiz Ishfaq Ahmad

Copyright © 2021 Umair Mahmood et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The inner ear is an essential part of a well-developed and well-coordinated hearing system. However, hearing loss can make communication and interaction more difficult. Inherited hearing loss (HL) can occur from pathogenic genetic variants that negatively alter the intricate inner ear sensory mechanism. Recessively inherited forms of HL are highly heterogeneous and account for a majority of prelingual deafness. The current study is designed to investigate genetic causes of HL in three consanguineous Pakistani families. After IRB approval, the clinical history and pure tone audiometric data was obtained for the clinical diagnosis of HL segregating in these three Pakistani families. We performed whole exome sequencing (WES) followed by Sanger sequencing in order to identify and validate the HL-associated pathogenic variants, respectively. The 3-D molecular modeling and the Ramachandran analysis of the identified missense variants were compiled to evaluate the impact of the variants on the encoded proteins. Clinical evaluation revealed prelingual severe to profound sensorineural HL segregating among the affected individuals in all three families. Genetic analysis revealed segregation of several novel variants associated with HL, including a canonical splice-site variant (c.55-2A>G) of *PTPRQ* in family GCFHL-01, a missense variant [c.1079G>A; p.(Arg360Gln)] of *SERPINB6* in family LUHL-01, and an insertion variant (c.10208-10211insCCACCAGCCCCGTGCCTC) within *MYO15A* in family LUHL-011. All the identified variants had very low frequencies in the control databases. The molecular modeling of p.Arg360Gln missense variant also predicted impaired folding of *SERPINB6* protein. This study reports the identification of novel disease-causing variants in three known deafness genes and further highlights the genetic heterogeneity of HL in Pakistani population.

1. Introduction

A significant portion of our genome, comprised of ~30,000 genes [1, 2], is associated with the development and function of hearing. Pathogenic variants in these genes account for around 50% of hearing loss (HL) cases. Approximately 5% of the world's population is affected with various kinds of HL [3]. HL is the most recurrent sensory disability in humans

with a frequency rate of 1–2:1000 babies. Worldwide, this targets around 360 million people of different ages [4]. More than 70% of hearing impairments are nonsyndromic with an inheritance of 70-90% autosomal recessive, 10-20% autosomal dominant, and 1-2% X-linked and mitochondrial inheritance [5]. Nonsyndromic recessive hearing loss (NSRHL) genes encode proteins widely spread in different tissues. However, variants in NSRHL genes specifically hinder the

intricate inner ear sensory mechanism [6, 7]. For nonsyndromic SNHL, 76 genes have been identified out of the 126 distinct autosomal genetic HL loci [8].

Among populations that have high rates of consanguineous marriages and are isolated based upon geographically, religiously, cultural, and social factors, recessive disorders are found in abundance [9]. In Pakistan, the ratio of consanguineous marriages is extremely high. Because of this, congenital severe hearing impairment accounts for 70% of the total HL cases in Pakistan [10]. Considering these facts, the present study was designed to determine the previously unknown nucleotide basis that was responsible for hearing damage in three large consanguineous Pakistani families affected with HL.

In the hearing process, several hundreds or even thousands of genes are involved in the proper development and functioning of the inner ear neurosensory epithelia. Diverse genes and their expressed protein families (e.g., solute carrier proteins, gap junction, and motor proteins) play orchestrated roles in the various molecular functions of the inner ear, including neurotransmitter release, maintenance of ionic homeostasis, control of adhesion in hair cells, intracellular transport, and protection of hair cell cytoskeleton. Together, all this makes it possible for us to hear sounds [3].

Over the last two decades, we witnessed rapid identification of genes and their pathogenic variants associated with hearing loss in humans. Studies in inbred families were an especially dominant part of this field. In the current study, whole exome sequencing (WES) was used for the identification of causative genes in large consanguineous Pakistani families with the HL phenotype. The current study reports novel variants of three known deafness genes and further highlights the genetic heterogeneity of HL in the Pakistani population.

2. Materials and Methods

2.1. Subjects and Clinical Evaluation. The study was compiled following the tenets of the Declaration of Helsinki for human subjects, and all the procedures that were followed were pre-approved by the Institutional Review Board Committees (HP-00061036) at the University of Maryland School of Medicine, Baltimore, MD, USA, and Government College University, Faisalabad, Pakistan. All the individuals consented in written local combination format for voluntarily inclusion in this project. The families were selected on the basis of (i) inheritance of disease phenotype and (ii) the number of affected individuals (≥ 3 affected). The family members were interviewed in detail to develop the family pedigree, associated disorders, and their follow-up. The physical examination, medical history, and pure tone audiometry data were assembled to highlight the clinical phenotype.

2.2. Whole-Exome Sequencing and Bioinformatics Analyses. WES was used to analyze the variants in the DNA sample of the affected individuals from each family, and the Agilent SureSelect Human Expanded All Exon V5 kit was used to recover genomic libraries and sequenced with an average of 100x coverage on an Illumina HiSeq4000 (Illumina, San

Diego, CA, USA). Reads were aligned with the Illumina Chastity Filter with the Burrows-Wheeler Aligner [11]. The GATK UnifiedUnityper module was used to call the variant sites, and the variant quality score recalibration method was used to filter single nucleotide variants [12]. The filtration of candidate variants, DNA sequencing, and PCR amplification was also performed [13]. Primer 3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) was used to design the primers used for PCR.

2.3. Molecular Modeling. The three-dimensional (3-D) structures of wild-type and mutant proteins SERPINB6 were generated by Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) by using an intense mode option. The Chimera online tool was used to visualize PDB structure of protein that was subsequently further analyzed. Mol-Probity was used to generate Ramachandran plots for both the wild-type and mutant protein PDB structures. Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) was used to align the sequence of the closely related species.

3. Results

3.1. Clinical Data. As part of our ongoing efforts to ascertain and clinically and genetically characterize Pakistani families with hearing loss [14, 15], three new large consanguineous families were enrolled from the Punjab province of Pakistan after approval from the Institutional Review Board (IRB) of University of Maryland School of Medicine, Baltimore, MD, USA, and the Government College University, Faisalabad, Pakistan. In all affected individuals, HL was observed from birth, except for the affected individuals of family GCFHL-01 who had progressive HL and were diagnosed after 4-5 years of age. The audiometric profile of affected members of family LUHL-01 revealed moderate to severe hearing damage, and the affected members of family LUHL-011 showed severe to profound HL. Previous medical history and clinical diagnosis did not reveal any apparent comorbidity with HL. Romberg and Tandem gait tests revealed normal vestibular system in affected individuals of all families. The peripheral vision loss, cornea opacity, or night blindness was not observed during ophthalmoscopic examination in any individuals from the included families.

3.2. Mutation Detection and Molecular Modeling. WES of three large consanguineous Pakistani families was performed to analyze the pathogenic variants for NSHL. The novel splicing variant c.55-2A>G of *PTPRQ* was detected as cosegregating in the affected individuals of families GCFHL-01, while an insertion variant c.10208-10211insCCACCAGGCCCGTGC CTC of *MYO15A* segregating with HL was found in family LUHL-011 (Figures 1(a) and 2(a)). Finally, a novel predicted missense variant c.1079G>A (p.(Arg360Gln)) of *SERPINB6* was identified in family LUHL-01 (Figures 1(a) and 2(a)). All the identified variants were predicted pathogenic by various in silico algorithms, were either absent or had low allele frequency in gnomAD database, and were classified as pathogenic or likely pathogenic according to ACMG classification (Table 1). Furthermore, the p.(Arg360Gln) predicted missense variant replaces an evolutionary conserved residue in the

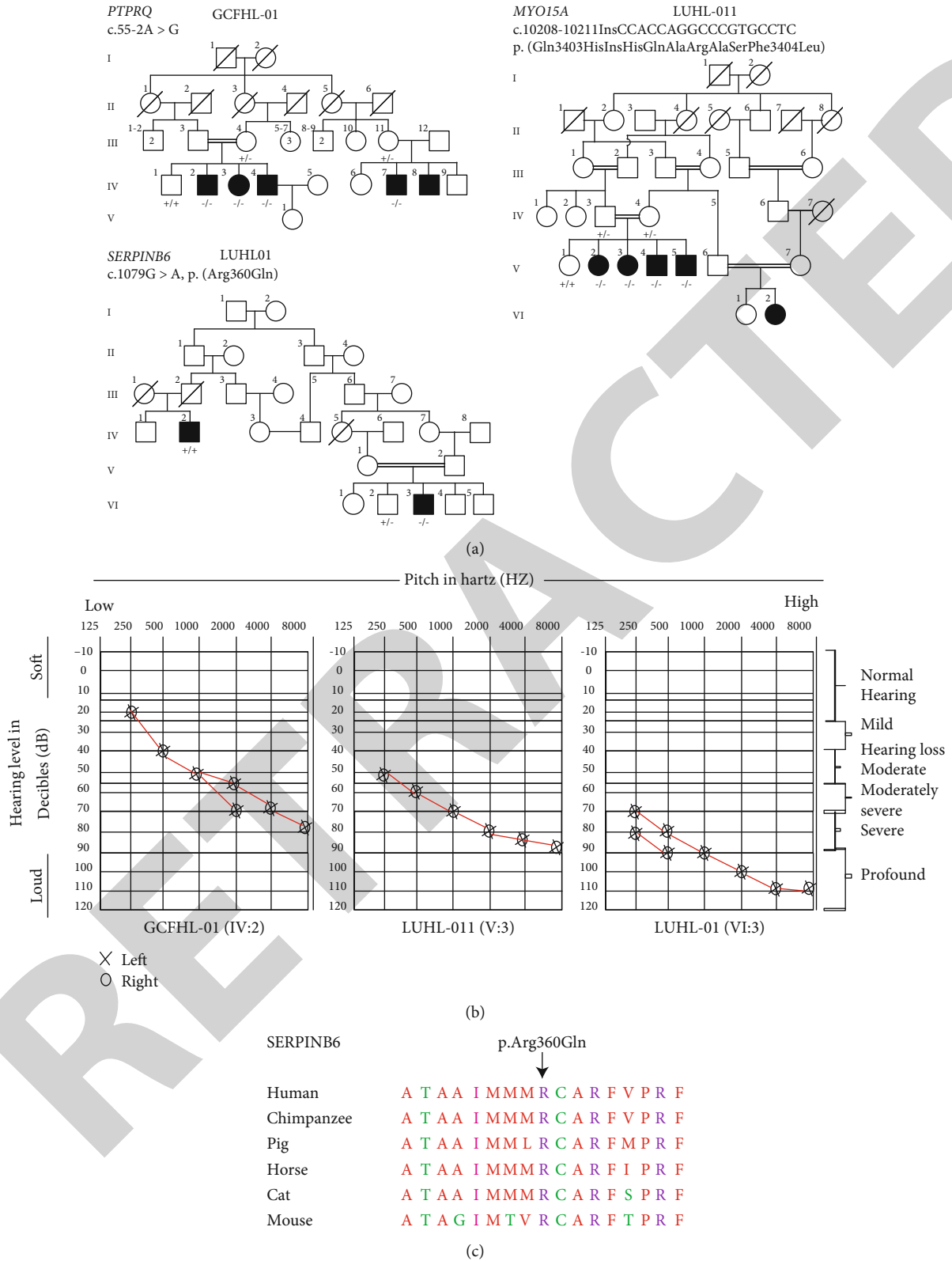


FIGURE 1: Family pedigrees, hearing loss (HL) phenotype and causative alleles. (a) Segregation of HL causing alleles in three Pakistani families. Double lines indicate consanguineous families, empty symbols represent unaffected individuals, and filled symbols affected individuals. The genotypes of the identified variants are also shown for each of the participating family members. All families had autosomal recessive mode of inheritance for HL. (b) Audiometric air conduction thresholds from the proband of each Pakistani family revealed varying degree of HL. (c) ClustalW multiple amino acid sequence alignment shows evolutionary conservation of arginine at position 360 of SERPINB6.

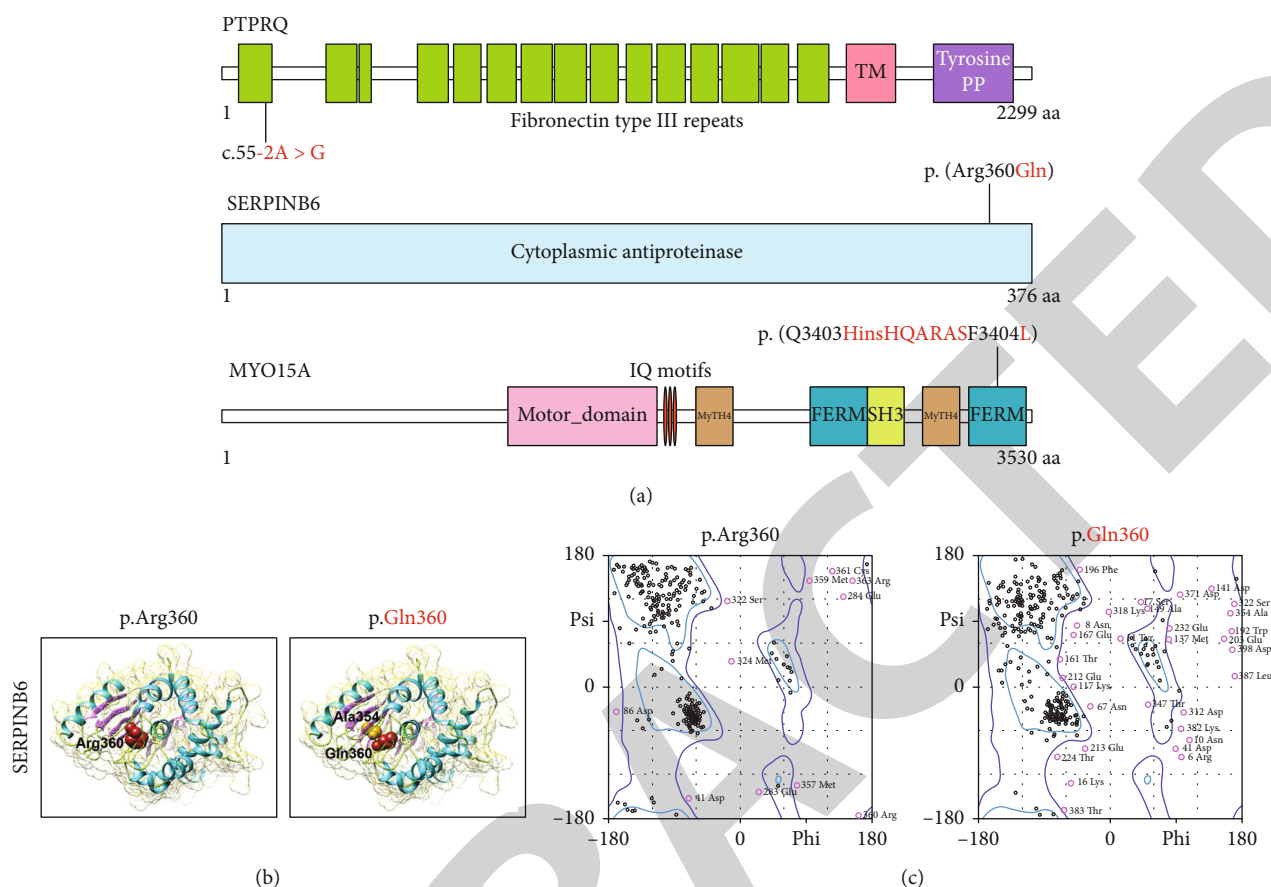


FIGURE 2: Protein structure, molecular modeling, and Ramachandran plots. (a) Schematic representations of PTPRQ, SERPINB6, and MYO15A proteins along with HL-causing variants were identified in Pakistani families. (b) 3-dimensional molecular modeling of SERPINB6. Protein secondary structures generated by Phyre2 are shown in respective colors: helix, cyan; strand, purple; and coils, green. The protein surface is displayed in a meshwork. Concerned residue is shown in firebrick color, while the aberrant hydrogen bonding due to p.Arg360Gln variant is displayed in golden yellow color. (c) Ramachandran plots of both wild and mutant protein PDB structures; the wild structure shows 88% of the residues existing in the favored region, and 96% of the residues are located in the allowed region, but 68% of the residues in mutant structure are found in the favored region and 86% are in the allowed region, respectively.

encoded protein (Figure 1(c)). All three variants are located in the functional domains of encoded proteins (Figure 2(a)).

3.3. Molecular Modeling. To determine the impact of p.Arg360Gln missense variant on the encoded SERPINB6 protein, the 3-D structure of wild-type and mutant proteins were generated using the online bioinformatics tool Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) and were visualized through Chimera software. The p.Arg360Gln replacement resulted in a smaller and negatively charged residue (Gln) as compared to the large and positively charged wild-type residue (Arg), which is predicted to cause alterations in the ionic interactions. This accounts for the aberrant interactions with neighboring residue alanine at position 354 (Figure 2(b)) and also might further disturb the secondary protein structure.

We also used the MolProbity tool to generate Ramachandran plots for both wild-type and mutant protein PDB structures. In the Ramachandran plot, the wild SERPINB6 protein shows that 88% of the residues reside within a favorable region and 96% residues are in allowed region, while 14 were outliers. In contrast, for the p.Arg360Gln mutant protein,

68% of the residues are found in the favored region, 86% are found in the allowed region, and 53 were outliers (Figure 2(c)). Overall, there is a significant difference in amino acid distribution of wild type versus mutant type (Figure 2(c)).

4. Discussion

Our study further expands the genetic landscape of inherited variants of HL-associated genes in the Pakistani population and revealed three novel variants in three known nonsyndromic deafness genes, *MYO15A*, *PTPRQ*, and *SERPINB6*. Although variants in *MYO15A* are a commonly known cause of HL worldwide (including the Pakistani population) [16], variants of *PTPRQ* and *SERPINB6* are relatively rare. In fact, to our knowledge, the p.(Arg360Gln) variant found in *SERPINB6* is only the fourth ever reported allele that is associated with HL in humans [17, 18]. *SERPINB6* is located on the chromosome at position 6p25.2 and encodes serpin (serine protease inhibitor) superfamily and subfamily ovalbumin-serpin B member 6, cytoplasmic anti-proteinase (CAP) protein. In the inner ear of the mouse, *SERPINB6* is highly

TABLE 1: Genes, identified variants, and their American College of Medical Genetics and Genomics (ACMG) classification.

Family	Gene	cDNA change	Protein change	CADD	GnomAD	Mutation taster	Polyphen2	ACMG classification
GCFHL01	<i>PTPRQ</i>	c.55-2A>G	N/A	N/A	N/A	Disease causing	N/A	Pathogenic (PVS1, PM2, PP3, and PP5)
LUHL011	<i>MYO15A</i>	c.10208-10211insCCACCAGGCCCGTGCCTC	N/A	N/A	N/A	Disease causing	N/A	Pathogenic (PVS1, PM2, PP3, and PP5)
LUHL-01	<i>SERPINB6</i>	c.1079G>A	p.(Arg360Gln)	N/A	0.001	Disease causing	Benign	Likely pathogenic (PS3, PP2, PP3, and BP4)

CADD: Combined Annotation Dependent Depletion (<https://cadd.gs.washington.edu/>); GnomAD: <https://gnomad.broadinstitute.org>. PVS1: pathogenic very strong [null variant (nonsense, frameshift, canonical ± 1 or 2 splice sites, initiation codon, single, or multiexon deletion) in a gene where LOF is a known mechanism of disease]; PM2: pathogenic moderate 2 [absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium]; PP3: pathogenic supporting 3 [multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.)]; PP5: pathogenic supporting 5 [reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation]; BP4: benign supporting 4 [benign computational verdict because of 1 benign prediction from GERP vs. no pathogenic predictions].

enriched in the organ of Corti sensory cells; however, expression is also found in the stria vascularis and spiral limbus region [19, 20]. Mice lacking *SERPINB6A* exhibit progressive degeneration of cochlear sensory cells and HL [19, 20]. We observed moderate to severe HL in family LUHL-01. However, it is not possible to determine if the hearing loss observed is progressive in nature or not, considering we only have data from a single audiometric report. Previous reports in human subjects with *SERPINB6* variants also lack the longitudinal analysis of hearing loss phenotype [17, 18]. We plan to follow up the evaluation of hearing phenotype in family LUHL-01 in future after the COVID-19 pandemic restriction lessen.

PTPRQ, located at chromosome 21q21.31, encodes the protein tyrosine phosphatase receptor Q (EC 3.1.1.48), a member of the protein tyrosine phosphatase receptor family type III. Previously, variants in *PTPRQ* have been reported for having dominant (DFNA73) [21], as well as recessively inherited (DFNB84) forms of nonsyndromic HL in human [22, 23]. We identified a novel splice variant (c.55-2A>G) of *PTPRQ* inherited in a recessive manner in family GCFHL-01. We did not observe HL among carriers of c.55-2A>G variant in family GCFHL-01. *In silico* analysis revealed cryptic splicing due to c.55-2A>G variants, which is predicted to cause exon skipping and premature truncation of the encoded protein, likely indicating a loss-of-function disease mechanism in the affected individuals of family GCFHL-01.

An inframe insertion variant in the carboxy tail FERM domain (Figure 2(a)) of *MYO15A* also was observed co-occurring with HL in family LUHL-011. Pathogenic variants of *MYO15A* have been extensively reported in families sensorineural nonsyndromic severe to profound HL [16]. Previously, a different insertion deletion (c. c.10208_10209delAGinsAC-CAGGCCCGTGCAGCTC) variant at the same nucleotide position, mutated in family LUHL-011, was documented in

another large Pakistani family [16], which could be coincidental or a mutation hot spot. In conclusion, our study further expands the genetic landscape of HL-associated variants of known deafness genes and provides information that can improve molecular diagnostics and genetic counseling for families segregating prelingual HL.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

The funders had no role in the experimental strategy; in the collection, analysis, and clarification of data; in the writing of the manuscript; or in the verdict to publish the data.

Conflicts of Interest

The writers announce no conflict of interest.

Authors' Contributions

S.A.B., M.A., and S.R. designed the experiment. U.M. is assigned to the methodology. S.A.B., Z.M.A., and S.R. are responsible for the software. Z.M.A. is involved in the confirmation. S.A.B. and Z.M.A. did the prescribed analysis. S.A.B., U.M., and Z.M.A. are responsible for the resources. Z.M.A. and S.R. organized the data. U.M., S.A.B., M.A., and S.R. wrote and prepared the original draft. Z.M.A., S.A.B., and S.R. wrote, reviewed, and edited the paper. S.A.B., M.A., and S.R. did the supervision. Z.M.A. and S.R. acquired funding.

Retraction

Retracted: Status of Bioinformatics Education in South Asia: Past and Present

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] M. M. Ali, M. Hamid, M. Saleem et al., "Status of Bioinformatics Education in South Asia: Past and Present," *BioMed Research International*, vol. 2021, Article ID 5568262, 9 pages, 2021.

Review Article

Status of Bioinformatics Education in South Asia: Past and Present

Muhammad Muddassir Ali,¹ Muhammad Hamid ,² Muhammad Saleem,³ Saadia Malik,⁴ Natash Ali Mian,⁵ Muhammad Ahmed Ihsan ,¹ Nadia Tabassum,⁶ Khalid Mehmood,⁷ and Furqan Awan⁸

¹*Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences, Lahore 54000, Pakistan*

²*Department of Statistics and Computer Science, University of Veterinary and Animal Sciences, Lahore 54000, Pakistan*

³*Department of Industrial Engineering, Faculty of Engineering-Rabigh, King Abdulaziz University, Jeddah 21589, Saudi Arabia*

⁴*Department of Information Systems, Faculty of Computing and Information Technology-Rabigh, King Abdulaziz University, Jeddah 21589, Saudi Arabia*

⁵*School of Computer and IT, Beaconhouse National University, Lahore 54000, Pakistan*

⁶*Faculty of CS and IT, Virtual University of Pakistan, Lahore 54000, Pakistan*

⁷*Faculty of Veterinary and Animal Sciences, The Islamia University of Bahawalpur, 63100, Pakistan*

⁸*College of Veterinary Medicine, South China Agricultural University, Guangzhou 510642, China*

Correspondence should be addressed to Muhammad Hamid; muhammad.hamid@uvas.edu.pk

Received 25 January 2021; Revised 10 March 2021; Accepted 26 March 2021; Published 27 April 2021

Academic Editor: Borhan Shokrollahi

Copyright © 2021 Muhammad Muddassir Ali et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Bioinformatics education has been a hot topic in South Asia, and the interest in this education peaks with the start of the 21st century. The governments of South Asian countries had a systematic effort for bioinformatics. They developed the infrastructures to provide maximum facility to the scientific community to gain maximum output in this field. This article renders bioinformatics, measures, and its importance of implementation in South Asia with proper ways of improving bioinformatics education flaws. It also addresses the problems faced in South Asia and proposes some recommendations regarding bioinformatics education. The information regarding bioinformatics education and institutes was collected from different existing research papers, databases, and surveys. The information was then confirmed by visiting each institution's website, while problems and solutions displayed in the article are mostly in line with South Asian bioinformatics conferences and institutions' objectives. Among South Asian countries, India and Pakistan have developed infrastructure and education regarding bioinformatics rapidly as compared to other countries, whereas Bangladesh, Sri Lanka, and Nepal are still in a progressing phase in this field. To advance in a different sector, the bioinformatics industry has to be revolutionized, and it will contribute to strengthening the pharmaceutical, agricultural, and molecular sectors in South Asia. To advance in bioinformatics, universities' infrastructure needs to be on a par with the current international standards, which will produce well-trained professionals with skills in multiple fields like biotechnology, mathematics, statistics, and computer science. The bioinformatics industry has revolutionized and strengthened the pharmaceutical, agricultural, and molecular sectors in South Asia, and it will serve as the standard of education increases in the South Asian countries. A framework for developing a centralized database is suggested after the literature review to collect and store the information on the current status of South Asian bioinformatics education. This will be named as the South Asian Bioinformatics Education Database (SABE). This will provide comprehensive information regarding the bioinformatics in South Asian countries by the country name, the experts of this field, and the university name to explore the top-ranked outputs relevant to queries.

1. Introduction

Bioinformatics is an emerging discipline, and with its advancements, there will be more opportunities for scientists in applied and pure research [1–5]. Nowadays, bioinformatics or computational biology has extreme prospect in genomic analysis and complete protein studies. Bioinformatics created a new community by maintaining networks among different areas and allow the researcher to work peacefully with the support of the whole community on its back. The cost of running the bioinformatics laboratory equipment and its teaching is far less than running molecular techniques and equipment. In bioinformatics, we mostly deal with software, algorithms, tools, and databases to perform the analysis of a desired sample.

European countries, USA, and UK are taking the lead in the field of bioinformatics [6–10]. South Asian countries are developing countries with immense pace; however, these bioinformatics fields are not much developed. Bioinformatics is a complex multidisciplinary field that analyzes and develops tools to manage and store the continuous growing biological data. This led to increase the importance and the efficiency of bioinformatics education in performing research [7, 11–14].

So, it is pertinent to take an overview of bioinformatics education in South Asia keeping in mind the diverse geographical and geopolitical situation of these countries. The current review would provide the current status of bioinformatics education, challenges, and development for the bioinformatics in South Asian countries. This would help to formulate policies for future development in the field of bioinformatics with more precision and insight.

The rest of the paper is organized as follows: in Section 2, history of bioinformatics in South Asia is discussed. The universities offering bioinformatics are described in Section 3. In Section 4, the bioinformatics curriculum in South Asia is presented. In Section 5, the bioinformatics conferences and workshop in South Asia are described in detail while Section 6 highlights the opportunities for bioinformatics in South Asia. In Section 7, the challenges the bioinformatics is facing in South Asia are defined while Section 8 describes the comprehensive platform framework for the development of South Asian Bioinformatics Education Database. In Section 9, suggestions for the improvement of bioinformatics education in South Asia is presented while Section 10 concludes the research and provides directions for future work.

2. History of Bioinformatics in South Asia

India is the first country among all the South Asian countries to introduce a formal study in bioinformatics. With the establishment of a nationwide network of distributed information centers (DICs) under the support of the Indian government and biotechnology information system (BTIS), India started the formal bioinformatics activities in the early 1980s [15, 16]. During the early years, generalized and short-term training programs were conducted in different areas of bioinformatics. In the 1980s, the main purpose of training programs was to build awareness of bioinformatics among IT professionals, biologists, and mathematicians [14, 16–

18]. In the coming years, these training programs will be evolved and will give more theme-based training on topics like biological database, database searches, algorithms and their applications, and structural bioinformatics and simulations. From the beginning of the new millennium, these DICs expanded their bioinformatics domain and started working on different bioinformatics aspects like genomics, proteomics, structural bioinformatics, and other domains [19–21].

COMSATS Institute of Information Technology (CIIT) and Muhammad Ali Jinnah University are the pioneers to introduce a bioinformatics study in Pakistan in 2003 [22]. Later, many other universities introduced bachelor's degree programs in this discipline, such as International Islamic University, Government College University, and University of Veterinary and Animal Sciences. From 2003, many workshops and conferences were conducted by the Higher Education Commission (HEC) of Pakistan on bioinformatics in institutions like Centre for Excellence in Molecular Biology (CEMB) and CIIT. In 2010, Regional Science Group Pakistan (RSG-P) was established, with a motto to change the scientific environment in Pakistan via knowledge sharing in computational science and in bioinformatics. RSG-P provided a platform for a bioinformatics to share their thoughts to enhance the prospect of bioinformatics in Pakistan [22].

Bangladesh is a young and prospective country on biological research especially on bioinformatics. As for now, there is not a single university in Bangladesh that has well-developed multidisciplinary bioinformatics course that incorporates most public-domain databases, research tools, peer-review journals, and research organizations. However, there are universities present in Bangladesh which offer bioinformatics study which helped a lot in pharmaceutical and biotechnological industry. The continuous progress of information and communication technology (ICT) had already laid the basis of bioinformatics studies and biological data analysis in Bangladesh. Bangladesh's pharmaceutical industry now meets 96 percent of the country's pharmaceutical needs [23], mainly because of extensive governmental funding in the pharmaceutical industry. The government of Bangladesh started offering a course combining pharmaceuticals with bioinformatics to promote and create awareness of bioinformatics within the country [23]. This helps to develop and stabilize companies for bioinformatics and pharmaceutical side by side.

In Nepal, two institutes teach bioinformatics to students. In 2003, the Department of Biotechnology at Kathmandu University started introductory courses on bioinformatics [24]. In 2012, Nepal acquired its first research database [25]. The database lacks expressed sequence tag (EST) records and genome survey sequenced record (GSS).

In Sri Lanka, bioinformatics came in between 2002, when the Department of Plant Sciences at the University of Colombo introduced a special course by the name of "Introduction of Bioinformatics." In 2010, four Sri Lankan universities started teaching bioinformatics. Currently, there is no government-made database of bioinformatics; however, the government organized many conferences and diplomas to increase the awareness of bioinformatics [26].

3. Universities Offering Bioinformatics

In India, more than a hundred universities are offering bioinformatics degree at different graduation and postgraduation levels [16, 26–28]. Details of some universities are provided in Table 1 (see Supplementary File) which includes a complete description about the universities' web link and the programs being offered.

In Pakistan, around thirty universities are offering bioinformatics as a professional degree [29]. A list of some universities along with their details of the courses is presented in Table 2 (see Supplementary File).

The study of bioinformatics is quite expensive in Sri Lanka. There are few institutes in Sri Lanka where bioinformatics is taught at the undergraduate level while only one university started offering bioinformatics at the Ph.D. doctorate level. Sri Lanka has currently more than 16 persons holding a Ph.D. degree in bioinformatics. Table 3 (see Supplementary File) furnishes details of the university which is presently offering bioinformatics degrees and courses in Sri Lanka [26].

In Nepal, bioinformatics was introduced in 2003. Nepal has only two institutes that are offering study in bioinformatics. A major factor for the late start of bioinformatics in Nepal is her landlocked geographical location on the world's map, and besides that, Ph.D. professors of bioinformatics are not in a collaboration with each other.

In Bangladesh, currently, three institutions are offering degree in bioinformatics, but progress is being made by the government of Bangladesh to make more institutes for biological sciences especially biotechnology and bioinformatics [25]. Table 4 (see Supplementary File) presents details of universities, which are offering bioinformatics in Nepal and Bangladesh.

The other countries of South Asia like Afghanistan, Bhutan, and Maldives are currently having no institution regarding the study of bioinformatics.

4. Bioinformatics Curriculum in South Asia

4.1. Diploma. The first diploma in bioinformatics was started in India during the year 2000. At first, biotechnology information system started an advanced diploma in bioinformatics in Jawaharlal Nehru University, University of Pune, University of Calcutta, Pondicherry University, and Madurai Kamaraj University [30]. Like India, Pakistan also started a one-year diploma in 2002 at the Pakistan Institute of Modern Sciences and Virtual University (VU) of Pakistan. In Nepal, the Department of Biotechnology started a one-year diploma in 2003 at the University of Kathmandu [24, 31].

The scheme of studies for a diploma is quite smooth and specially designed for those students who completed their bachelor in either basic science or medicine. The aim of these diploma programs is to build the awareness of bioinformatics among the student of natural science and to level the field for the introduction of undergraduate and postgraduate studies in South Asia especially.

4.2. Bachelors. Based on educational and industrial needs, every South Asian country varies in the curriculum of bioinformatics. In India, a bachelor's degree has various names, like B.Tech, BSc, M.Sc., and BS. The BSc and M.Sc. combine have worth equal to the four-year bachelor, while BS and B.Tech itself have a four-year program [32].

In Pakistan, Bangladesh, Nepal, and Sri Lanka, there is no concept of BSc; all universities offer only BS Hons program. Institute of Bioinformatics in India, COMSATS CIIT in Pakistan, the University of Kathmandu in Nepal, and the University of Colombo in Sri Lanka are some of the best institutes for bioinformatics study [29, 32, 33].

4.3. Masters. The University of Pune was the first university to start a master's degree program in bioinformatics in South Asia, with the financial support of the Department of Biotechnology (DBT) in 2002. During the early years, research or thesis was optional for masters in bioinformatics but now, it has become compulsory. In India alone, more than 29 universities are offering a master's degree in bioinformatics and the present curriculum has evolved from only theoretical bases to the research-oriented one. In Pakistan, 9 universities are offering masters in bioinformatics. Nepal, Sri-Lanka, and Bangladesh have only 1 university offering a master's degree in bioinformatics. Most South Asian universities are offering 90 credit hours for master's degree in bioinformatics. Research and thesis are now compulsory to get a degree in most South Asian institutes [19]. The master's level courses of bioinformatics in South Asia are specially designed to meet the local industry demands. The master's degree is more focused on the specific domain of bioinformatics like algorithms for the database and molecular biology, protein structure analysis, and microarray data.

4.4. Ph.D. In 1997, the University of Pune was the first institute throughout South Asia to start a Ph.D. in bioinformatics. The University of Pune is also the first university in South Asia to award a Ph.D. in bioinformatics in 2003. Currently, there are many public and private universities in India which are offering Ph.D. in bioinformatics like Indian Institute of Technology (IIT), National Centre for Biological Science (NCBS), Center for Cellular and Molecular Biology (CCMB), and National Institute of Pharmaceutical Education and Research (NIPER). These Ph.D. programs are funded by multiple institutes like DBT, Department of Science and Technology (DST), Ministry of Communication and Informational Technology (MCIT), and University Grant Commission (UGC) [34, 35].

In Pakistan, only 3 universities are offering a doctoral-level study in bioinformatics. Quaid-e-Azam University (QAU) was the first university in Pakistan which started offering a Ph.D. in bioinformatics while the other two universities are Hazara University and Capital University of Science and Technology. Ph.D. in bioinformatics was started in 2008 in Pakistan. There are many organizations that grant funding and support to bioinformatics' students. HEC of Pakistan is the biggest supporting organization; others include National Center for Bioinformatics (NCBI), Bioinformatics Research

Lab (BRL), and Institute of Molecular Science and Bioinformatics (ISMB) [31, 36].

Asian University for Women is the only university in Bangladesh which started offering a Ph.D. in bioinformatics in 2009. Bangladesh Bioinformatics and Computational Biology Association (BBCBA), Molecular Modelling and Drug Design Laboratory (MMDDL), and Bangladesh Council of Scientific and Industrial Research Laboratories (BCSIR) are the few institutes that deal with bioinformatics in Bangladesh.

Department of Biotechnology of Kathmandu University is the only department in Nepal which is offering Ph.D. in bioinformatics with the collaboration of some foreign organizations.

In Sri Lanka, Ph.D. in bioinformatics was started in 2013. The University of Colombo and the University of Peradeniya are two universities which are offering a Ph.D. degree in bioinformatics.

5. Bioinformatics Conferences and Workshop in South Asia

There are many steps taken by the HEC, and many teaching activities are in process since 2003. From 2003, the HEC supported financially on organizing conferences, committees, and conferences on bioinformatics. These workshops and conferences are organized by highly ranked institutes such as CIIT/COMSAT Islamabad, CEMB, and Punjab University Lahore. The major importance of these conferences and workshops is the interaction of local researchers with foreign or other local researchers and the transfer of new innovative ideas. Table 5 (see Supplementary File) provides information about workshops, symposiums, and conferences that were held in South Asia. Almost more than 33 workshops and conferences have been organized by the government of Pakistan and many universities to promote the awareness of bioinformatics and its applications among people in society. India is leading the race of bioinformatics in South Asia. Many conferences and workshops have been organized to enhance the promotion and knowledge of bioinformatics among Indians. Almost more than 100 conferences and workshops in the last 5 years had been organized in India just to increase the interaction of local researchers with the international researcher to discuss new topics and transfer of knowledge. The major aim of the Indian government is to comply with or attract foreign investors in India for the improvement of the native bioinformatics industry and market. Sri Lanka, Bangladesh, and Nepal had few conferences and workshops in the past years, but their motives and ideas had been reforming; therefore, almost 3 conferences on bioinformatics were organized by Bangladesh in this year [37]. The way of bioinformatics is soothing in these developing countries, and thus, it helps to advance their native technology regarding natural sciences. Few conferences and their descriptions are given below; the rest of the details: titles, organizers, and dates about conferences and workshops, are furnished in Table 5 (see Supplementary File).

5.1. Pre-18th FAOBMB Symposium Satellite Workshop on Bioinformatics, Pakistan. This symposium was held in the CEMB, Lahore, Pakistan, between 14th November and 19th November 2005. This conference is sponsored by different organizations like CEMB, Federation of Asian and Oceanian Biochemists and Molecular Biologists, APBioNet, and Progeniq Pte Ltd. This symposium is organized with the aim of bringing Pakistan's bioinformatics community together, understanding new bioinformatics concepts, and providing platforms for researchers, professors, and students to discuss certain issues. Five professors and scientists from Sweden and Singapore participated in this symposium.

5.2. Advances in Bioinformatics in the Postgenomic Era Workshop, Pakistan. The workshop was conducted in the National Center for Bioinformatics (NCBI) of QAU [34]. This workshop was sponsored by QAU, and the major aim of this workshop was to introduce new technology and strategies about bioinformatics in genomics and proteomics to train young scholars for the techniques which emerge in the bioinformatics field. All scholars and lecturers were from Pakistani universities.

5.3. Bioinformatics Workshop Hands-On Training on Analysis of Biological Data Using R, Pakistan. This workshop was held at VU on 5th January 2016. The workshop was sponsored by the VU and HEC [38]. The workshop had an aim to relate two different fields like bioinformatics and statistics to produce more perfect results, introducing R techniques which can be utilized in bioinformatics and building fundamental skills. All scholars, professors, and scientists in this workshop were from Pakistani universities like VU, COMSAT, CEMB, and QAU.

5.4. Computational Biology Workshop, Pakistan. This workshop was conducted by Habib University from 31st July to 4th August 2017. This workshop was sponsored by many private companies along with the university itself. The aim of the workshop was to provide information about computational techniques, how these techniques will communicate with biology, and the principle of biology and current affairs of research conducted in this field. Scholars and professor participated in this workshop were from America, Australia, UK, and Pakistan [38].

5.5. Indian Conference on Bioinformatics 2017 (Inbix'). This conference was organized by <http://Bioclues.org/> in association with Birla Institute of Scientific Research (BISR) in Jaipur, Rajasthan, India. BISR hosted this conference from 07 to 09 November 2017. This international conference was organized with the hope to foster the interactions and collaborations among young researchers and to connect them with the top researchers in the bioinformatics and computational biology. Temple-Smith from Boston University and Thomas Sicheritz from Technical University of Denmark were the keynote speakers at this conference [34].

5.6. Bifx India Virtual Conference 2010. Bifx India Virtual Conference was the first of its kind in India. It was organized by <http://Bioclues.org/> with the support of Bioinformatics

Organization of India, Asia Pacific Bioinformatics Network (APBioNet), and International Society for Computational Biology. The date of this conference was the 12th and 13th of February 2010. Dr. Søren Brunak (Denmark Technical University), Dr. Thomas Knudsen (National Center for Computational Toxicology, EPA, United States), Dr. Shoba Ranganathan, (Macquarie University, Australia), Dr. David Reif (National Center for Computational Toxicology, EPA, United States), Dr. Russel Thomas, (The Hamner Institute, United States), and Dr. Tin Wee Tan (National University of Singapore, Singapore) were the keynote speakers of this conference.

5.7. The Eighth Asia Pacific Bioinformatics Conference, India. The 8th Asia Pacific Bioinformatics Conference was held in the Indian Institute of Science, Bangalore, from January 18 to 21, 2010. The conference focused on the following topics which are sequence analysis, motif search/analysis, RNA analysis, physical and genetic maps, evolution and phylogeny, protein structure analysis, transcription, gene expression, proteomics, and population genetics/SNP/haplotyping.

5.8. International Conference on Bioinformatics (ICBINF-18), Sri Lanka. Research League of Sri Lanka has organized an international conference on bioinformatics in Kandy, on the 12th of October 2018. The aim of this conference was to provide a platform to the researchers and practitioners from both academia and industry to share cutting-edge development in the field of bioinformatics.

6. Opportunities for Bioinformatics in South Asia

Bioinformatics gained importance due to its advancements; there will be more opportunities for scientists in applied and pure research. The cost of running the bioinformatics laboratory equipment and its teaching is far less than running molecular techniques and equipment. Most of the bioinformatics software is freely available on the internet for academic purposes. The main perspective of bioinformatics is to store lots of biological data in the form of a database. The development of new standards and software coding rules would ease our way to conduct research by helping us to analyze and match the preexisting genomic data. The availability of good services for data deposition, good programming, and data analysis would gain much importance in bioinformatics to remove false data from databases.

India has a well-developed system of biotechnology as compared to other South Asian countries based on business to business (B2B) structure where it provides services to business. With the help of biotechnology, pharmaceutical and agriculture industries are changing the lives of local communities. The government of India has taken serious efforts to introduce bioinformatics in its universities which really helps India to increase its revenue in the global bioinformatics market. The advances in the field of bioinformatics and the pharmaceutical industry would be a huge bonus for India. The government of India has systematic effort in bioinformatics industry and research if it can properly capitalize on

its resources. IT companies like Infosys, Cognizant Technology, and Tata consultancy services (TCS) are now investing in computational biology along with other IT services. The Indian government and bioinformatics companies can look forward to bioinformatics services like deoxyribonucleic acid (DNA) mapping, DNA mining, DNA sequencing, and functional gene analysis. Biotechnology information system (BTIS) along with DBT is promoting bioinformatics and recently developed a Bio-IT Park. Bioinformatics in India has gained significant growth by information technology and institutional education. Biotechnology information system network (Bitnet) has developed a countrywide network that provides valuable information to students and experts in biotechnology and bioinformatics. These databases offer the students an opportunity to learn about processing, efficient organization, formatting of data, gene mapping, gene identification, new genes, and protein analysis. These networks and industries provide enormous job vacancy for bioinformaticians and also provide new innovative ideas to students to perform an industry-oriented and profitable research. Being the largest English-speaking scientific workforce in the world, we are hopeful that a lot of opportunities would come to India and Pakistan in the coming years [36].

Although the development capability and jobs of bioinformatics in Pakistan are less than in those India, still, it has a large potential for bioinformatics to prosper. Most of the Pakistani students graduated in bioinformatics are working either in academia or in pharmaceutical industries. Many types of bioinformatics jobs are available in Pakistan for the computational specialist as either implementation of system and program for management or analysis and storage of a vast amount of sequence DNA. The laboratory information management systems (LIMS) major objective is to improve the accuracy of results, to ease the access to databases, and to increase the utilization of data. RSG-P provides a forum for local bioinformaticians to collaborate with other students, experts, and academicians. It would provide an opportunity to uplift research in bioinformatics. Every lab requires qualified bioinformaticians to perform dry work and run their confirmatory tests. Development of EMBnet and e-book helped to develop a database for local students to add their research for the promotion of bioinformatics in Pakistan. There is an increasing opportunity for bioinformatics students in the pharmaceutical industry in drug development and disease control. As Pakistan is an agricultural country and most of the research work is done on plants and agriculture by agriculturists and plant biotechnologists, bioinformatics assists in this particular area to increase the workflow and this will promote the bioinformatics and cause vacancy of jobs in this particular area.

Due to its unique fauna and flora, Nepal is one of the most prospective places for biologists. There is a huge possibility that bioinformatics can get interesting information from the species of Nepal as most of them are biologically undiscovered. Every country needs computational data to conserve its natural resources, and bioinformaticians are the only data analysts present to analyze data with biological knowledge. In short, we can say that bioinformatics has a huge scope in agriculture and medical health in Nepal.

Bangladesh is still new in bioinformatics research. Bangladesh has no infrastructure for bioinformatics, and no university is well developed. However, recently, a Bangladeshi research group had developed a global network and introduced important research areas to talented researchers and students. Bangladeshi bioinformaticians would allow the pharmaceutical companies to gain new achievements and increase efficiency rate by the introduction of new tools and designs and new screening test; this provides more jobs and more learning for students. A large number of vacancies are empty for bioinformaticians in the pharmaceutical industry to organize and analyze biological data. However, unfortunately, there is no one to fill them.

In Sri Lanka, the opportunity for bioinformaticians is increasing day by day as the government has started an awareness program by combining forces with the University of Colombo and added one seat of bioinformatician in every hospital and research lab which collects and analyze the biological data. Sri Lanka also has a journal to publish local bioinformatics work with the name of *Sri-Lanka Journal of Bio-Medical Informatics* [36].

7. Challenges That Bioinformatics is Facing in South Asia

Being an integrated science, bioinformatics faces many challenges and problems in South Asia. Biology experts have no expertise in computer functioning, programming, and computer theories. Another major problem in South Asia is the lack of experts in bioinformatics. Governments sent people to foreign countries to acquire knowledge for the country, but most of them never returned [29, 33]. Some of the challenges that South Asian countries are facing to improve the status of bioinformatics are as follows.

7.1. Education Quality. The bioinformatician must have an understanding of biology, statistics, computer science, IT, and mathematics along with core bioinformatics knowledge. Only a handful of teachers with expertise enlisted above are available for teaching and research in South Asia. Although new universities are opening every year, still, academia is unable to produce well-trained bioinformaticians in South Asia [27, 29]. In India, an average of 35 persons are getting their Ph.D. degrees in bioinformatics every year, while in Pakistan, an average of 12 persons get their Ph.D. degree in bioinformatics. In Sri Lanka, there are only 16 Ph.D. scholars so far and most of them migrated to foreign countries. This brain drain is the main reason behind the shortage of teachers and experts in bioinformatics.

7.2. The Gap between Biological and Computer Sciences. Biology and computer sciences are two different fields, and both have different perspectives and scope. The biologist generally does not know about the deep understanding of computer science, and computer experts lack knowledge in biology. Understanding the concepts of both fields of study is a tiresome process, and sometimes, biologists face failure just because they are not able to communicate their ideas to IT experts or vice versa. There is not even a single algorithm

made by experts from Pakistan, India, Nepal, or Bangladesh. In 2014, HEC, BTIS, and BBCBA planned a scheme to add bioinformatics as a compulsory subject in all biology and natural science degrees aimed at improving the student extinct toward bioinformatics.

7.3. Learning Method and Education. In India, more than 90 institutes are teaching bioinformatics at bachelor's, master's, or Ph.D. levels, but unfortunately, a number of application-oriented students are far less than the desired amount. In Pakistan, the major problem is the curriculum. From 2003 to 2014, HEC amended the curriculum of bioinformatics eight times. On the basis of the multidisciplinary nature of the subject, multiple methods were applied to achieve the desired outcome. Courses regarding the programming skills were also enhanced. Adaption to teach bioinformatics is also a problem because teaching bioinformatics is enormously different from teaching other natural science degrees [29, 30, 33].

7.4. Lack of Algorithm Development. India is considered an IT hub for producing an enormous amount of software and computer programs, but unfortunately, there is no significant Indian contribution in the development of the bioinformatics algorithm. In South Asian countries, the algorithm production is very rare and there is not even a single tool or application which was produced by any South Asian expert [30, 31].

8. Comprehensive Platform Framework for the Development of South Asian Bioinformatics Education Database

A comprehensive platform framework is suggested to make a database about the status of South Asian bioinformatics education. The procedure for the development of this system is to form a database, which will store information regarding the status and knowledge of South Asian education from a bioinformatics scenario (Figure 1). The database will be named as SABE (South Asian Bioinformatics Education Database). Users would be able to take search regarding the bioinformatics in South Asian countries and the country name, by the experts of this field and by the university name to explore the top-ranked outputs relevant to queries. This single platform would be available to end users for providing useful updated complete information of South Asian bioinformatics education.

9. Suggestions for the Improvement of Bioinformatics Education in South Asia

These are some of the suggestions to improve the prospect of bioinformatics in South Asia:

- (1) The study of bioinformatics should be made compulsory in all degrees of natural science and along theory; measures must be taken to make the students familiar with practical and dry lab work including molecular modeling, BLAST, gene prediction, gene

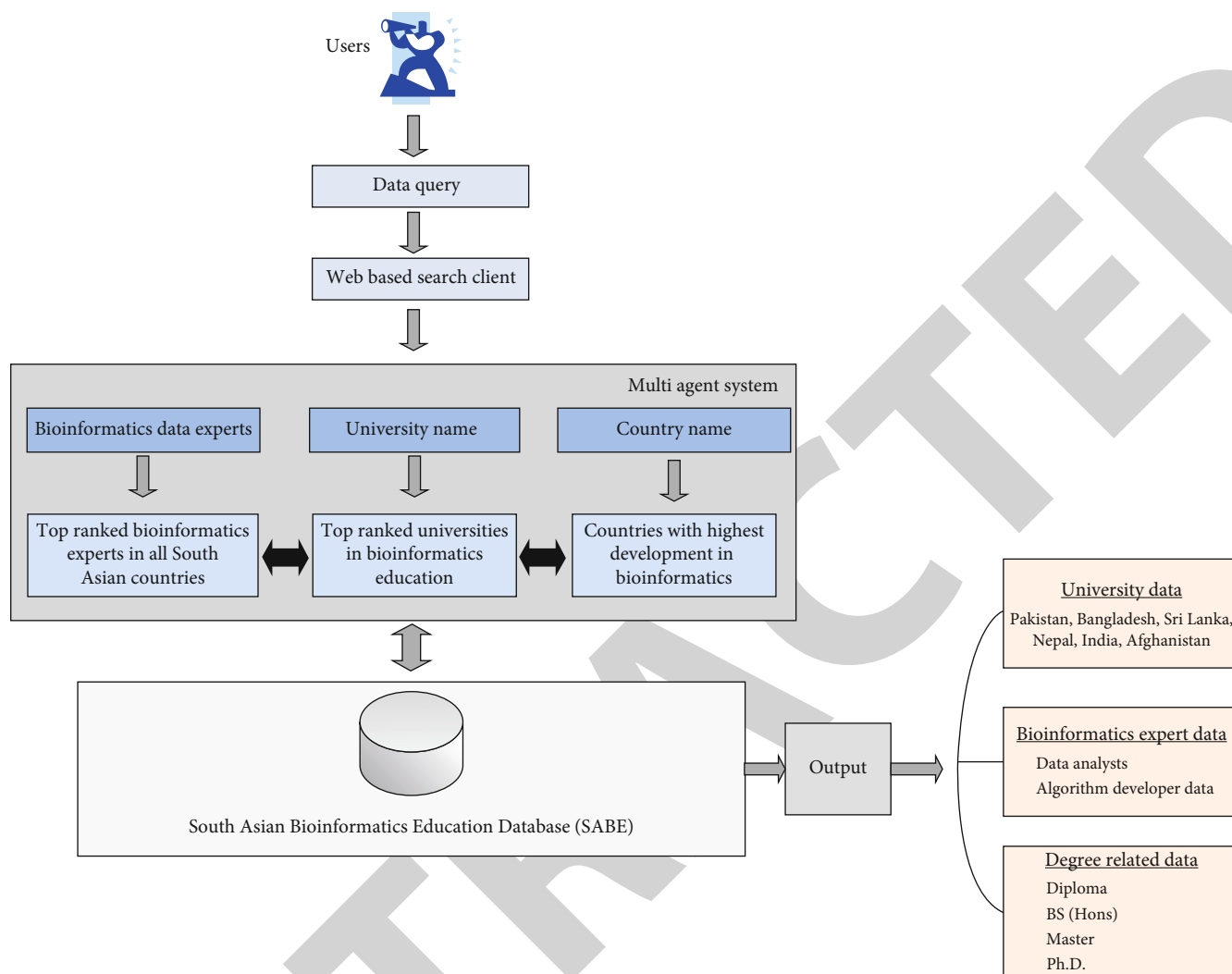


FIGURE 1: The schematic representation of SABE framework.

analysis, and complete protein studies. This will create more opportunities for bioinformatics graduates

- (2) Short courses and diplomas must be designed for imparting bioinformatics knowledge. These courses would help them to work effectively in a skillful and diverse workforce
- (3) Seminars must be arranged to create a diverse community having skills in mathematics, statistics, computer science, and biological sciences
- (4) Because most of the research was undertaken in universities, therefore, students and researchers in universities must be encouraged to attend training and conferences related to bioinformatics
- (5) Diploma courses are very necessary to make the bioinformatics workforce. In the past, universities like Allama Iqbal Open University (ALU), VU of Pakistan, and Bioinformatics Institute of India had started short courses and distance learning diplomas. But sadly, these courses are now aban-

doned in Pakistan. Pakistan must have to start their diploma programs again

- (6) Local universities must collaborate with foreign universities and research institutes. It would increase the worth of a local degree
- (7) To promote the prospect of bioinformatics, a persuasive campaign must be designed on social media to spread the message of bioinformatics throughout South Asia
- (8) The bioinformatics curriculum that is taught in South Asian universities should be recognized by the legal education bodies like HEC and BTIS in Pakistan and India, respectively

10. Conclusion

Bioinformatics is an interdisciplinary emerging field, and it leads a lot of expertise as it deals with many branches of science. The bioinformatics education started in South Asia

with less interest, but the interest is developing among students, governments, industry, and other stakeholders. Bioinformatics can fill the gap of the latest scientific research in South Asia if little extra and dedicated efforts are done by the government. The bioinformatics industry can be a source of income and a source of developing bioinformatics in South Asia. Despite many challenges and problems, bioinformatics is progressing in South Asia with great speed, but this pace of improvement can be accelerated with little more focused struggle by all stakeholders. At the end, based upon the literature review, a framework of the development of a centralized database (SABE) is suggested to collect and store the information on the current status of South Asian bioinformatics education.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Supplementary Materials

Table 1: some of the major universities in India offering different bioinformatics courses. Table 2: some of the Pakistani universities offering different courses and degrees in bioinformatics. Table 3: universities in Sri Lanka offering bioinformatics degree and courses. Table 4: universities offering bioinformatics courses in Nepal and Bangladesh. Table 5: major conferences and workshops of bioinformatics held in South Asian countries. (*Supplementary Materials*)

References

- [1] P. Lakshmi and D. Ramyachitra, "Review about bioinformatics, databases, sequence alignment, docking, and drug discovery," in *Statistical Modelling and Machine Learning Principles for Bioinformatics Techniques, Tools, and Applications*, pp. 11–23, Springer, 2020.
- [2] L. Liu, B. Song, J. Ma et al., "Bioinformatics approaches for deciphering the epitranscriptome: recent progress and emerging topics," *Computational and Structural Biotechnology Journal*, vol. 18, pp. 1587–1604, 2020.
- [3] A. J. Magana, M. Taleyarkhan, D. R. Alvarado, M. Kane, J. Springer, and K. Clase, "A survey of scholarly literature describing the field of bioinformatics education and bioinformatics educational research," *CBE—Life Sciences Education*, vol. 13, no. 4, pp. 607–623, 2014.
- [4] J. Guan and X. Gao, "Comparison and evaluation of Chinese research performance in the field of bioinformatics," *Scientometrics*, vol. 75, no. 2, pp. 357–379, 2008.
- [5] A. D. Baxevanis, G. D. Bader, and D. S. Wishart, *Bioinformatics*, John Wiley & Sons, 2020.
- [6] C. W. Van Gelder, R. W. Hooft, M. N. Van Rijswijk et al., "Bioinformatics in the Netherlands: the value of a nationwide community," *Briefings in Bioinformatics*, vol. 20, no. 2, pp. 375–383, 2019.
- [7] P. Papadopoulou, M. Lytras, and C. Marouli, "Bioinformatics as applied to medicine: challenges faced moving from big data to smart data to wise data," in *Biotechnology: Concepts, Methodologies, Tools, and Applications*, pp. 185–209, IGI Global, 2019.
- [8] V. Baillie Gerritsen, P. M. Palagi, and C. Durinx, "Bioinformatics on a national scale: an example from Switzerland," *Briefings in Bioinformatics*, vol. 20, no. 2, pp. 361–369, 2019.
- [9] A. Via, T. Blicher, E. Bongcam-Rudloff et al., "Best practices in bioinformatics training for life scientists," *Briefings in Bioinformatics*, vol. 14, no. 5, pp. 528–537, 2013.
- [10] Y. Machluf and A. Yarden, "Integrating bioinformatics into senior high school: design principles and implications," *Briefings in Bioinformatics*, vol. 14, no. 5, pp. 648–660, 2013.
- [11] A. Holzinger, M. Dehmer, and I. Jurisica, "Knowledge discovery and interactive data mining in bioinformatics-state-of-the-art, future challenges and research directions," *BMC Bioinformatics*, vol. 15, pp. 1–9, 2014.
- [12] R. Fang, S. Pouyanfar, Y. Yang, S.-C. Chen, and S. Iyengar, "Computational health informatics in the big data age: a survey," *ACM Computing Surveys (CSUR)*, vol. 49, pp. 1–36, 2016.
- [13] T. K. Karikari, "Bioinformatics in Africa: the rise of Ghana?," *PLoS Computational Biology*, vol. 11, no. 9, article e1004308, 2015.
- [14] T. K. Attwood, "Genomics: the babel of bioinformatics," *Science*, vol. 290, no. 5491, pp. 471–473, 2000.
- [15] J. Leipzig, "A review of bioinformatic pipeline frameworks," *Briefings in Bioinformatics*, vol. 18, pp. 530–536, 2017.
- [16] A. Som, P. Kumari, and A. Ghosh, "Advancing India's bioinformatics education and research: an assessment and outlook," *Journal of Proteins and Proteomics*, vol. 10, no. 3, pp. 257–267, 2019.
- [17] W. J. Ewens and G. R. Grant, *Statistical Methods in Bioinformatics: An Introduction*, Springer Science & Business Media, 2006.
- [18] R. Pereira, J. Oliveira, and M. Sousa, "Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics," *Journal of Clinical Medicine*, vol. 9, no. 1, p. 132, 2020.
- [19] L. da Fontoura Costa, "Bioinformatics: perspectives for the future," *Genetics and Molecular Research*, vol. 3, pp. 564–574, 2004.
- [20] G. Mboowa, I. Sserwadda, and D. Aruhomukama, "Genomics and bioinformatics capacity in Africa: no continent is left behind," *Genome*, vol. 999, pp. 1–11, 2020.
- [21] J. H. J. Tan, S. L. Kong, J. A. Tai et al., "Experimental and bioinformatics considerations in cancer application of single cell genomics," *Computational and Structural Biotechnology Journal*, vol. 19, pp. 343–354, 2021.
- [22] S. Manzoor, A. Niazi, and E. Bongcam-Rudloff, "A stepping stone to develop bioinformatics in Pakistan," *EMBnet. journal*, vol. 23, p. 891, 2017.
- [23] M. M. Islam, "Role of bioinformatics in developing country: Bangladesh," *Current Trends in Technology & Science*, vol. 2, pp. 160–165, 2013.
- [24] J. Guo, *Nepal Counts on Science to Turn Struggling Country around*, American Association for the Advancement of Science, 2008.
- [25] Y. Sapkota and S. Subedi, "Bioinformatics—an entry-level avenue for biomedical research in Nepal," *Frontiers in Genetics*, vol. 5, p. 42, 2014.
- [26] March 2020, <http://www.nrc.gov.lk/index.php/about-nrc/vision-mission.html>.
- [27] N. Mulder, R. Schwartz, M. D. Brazas et al., "The development and application of bioinformatics core competencies to

Retraction

Retracted: Computer-Aided Prediction and Identification of Phytochemicals as Potential Drug Candidates against MERS-CoV

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] H. S. Mahrosh, M. Tanveer, R. Arif, and G. Mustafa, "Computer-Aided Prediction and Identification of Phytochemicals as Potential Drug Candidates against MERS-CoV," *BioMed Research International*, vol. 2021, Article ID 5578689, 7 pages, 2021.

Research Article

Computer-Aided Prediction and Identification of Phytochemicals as Potential Drug Candidates against MERS-CoV

Hafiza Salaha Mahrosh ¹, Muhammad Tanveer ², Rawaba Arif ¹,
and Ghulam Mustafa ¹

¹Department of Biochemistry, Government College University, Faisalabad 38000, Pakistan

²Prince Sultan University, Riyadh, Saudi Arabia

Correspondence should be addressed to Ghulam Mustafa; gmustafa_uaf@yahoo.com

Received 13 February 2021; Accepted 31 March 2021; Published 12 April 2021

Academic Editor: Andrea Scribante

Copyright © 2021 Hafiza Salaha Mahrosh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Middle East respiratory syndrome coronavirus (MERS-CoV) is the major leading cause of respiratory infections listed as blueprint of diseases by the World Health Organization. It needs immediate research in the developing countries including Saudi Arabia, South Korea, and China. Still no vaccine has been developed against MERS-CoV; therefore, an effective strategy is required to overcome the devastating outcomes of MERS. Computer-aided drug design is the effective method to find out potency of natural phytochemicals as inhibitors of MERS-CoV. In the current study, the molecular docking approach was employed to target receptor binding of CoV. A total of 150 phytochemicals were docked as ligands in this study and found that some of the phytochemicals successfully inhibited the catalytic triad of MERS-CoV. The docking results brought novel scaffolds which showed strong ligand interactions with Arg178, Arg339, His311, His230, Lys146, and Arg139 residues of the viral domains. From the top ten ligands found in this study (i.e., rosavin, betaxanthin, quercetin, citromitin, pluviatilol, digitogenin, ichangin, methyl deacetylnomilinate, kobusinol A, and cyclocalamin) based on best *S*-score values, two phytochemicals (i.e., pluviatilol and kobusinol A) exhibited all drug-likeness properties following the pharmacokinetic parameters which are important for bioavailability of drug-like compounds, and hence, they can serve as potential drug candidates to stop the viral load. The study revealed that these phytochemicals would serve as strong potential inhibitors and a starting point for the development of vaccines and proteases against MERS-CoV. Further, *in vivo* studies are needed to confirm the efficacy of these potential drug candidates.

1. Introduction

Middle East respiratory syndrome (MERS) is a lethal respiratory syndrome caused by MERS coronavirus (MERS-CoV). In 2012, a 60-year-old man with lung failure in Jeddah, Saudi Arabia, was diagnosed with this syndrome. First time, in 2012, MERS-CoV was isolated from a 60-year-old Saudi male who died from severe respiratory failure [1]. MERS epidemic affected many countries with more severity (affected 1083 individuals in 23 countries) [2]. Viruses from the coronavirus family cause major respiratory and intestinal infections in animals and have been considered pathogenic to humans after the outbreak of SARS epidemic in 2002 and 2003 in China and MERS in 2012 in Saudi Arabia. SARS virus uses

an ACE2 enzyme as the receptor and affects the ciliated bronchial epithelial cells [3]. MERS-CoV targets dipeptidyl peptidase 4 (DPP4) and affects the unciliated bronchial epithelial cells [2].

MERS-CoV uses the machinery of the host cell and creates the copies of its genome in the cytoplasm of the host cell using its replication machinery. CoV completes its life cycle in four steps as S protein mediated cellular fusion and interaction with DPP4, expression of replicase for the replication, and transcription and release of virions [4]. DPP4 is directly linked with the human immune system, expressed on the cell surface and involved in various processes such as signal transduction, apoptosis, regulators for T-cell activation, regulation and cleavage of peptidase hormones, and

neuropeptides [5]. DPP4 is expressed on the epithelial and endothelial cell surface of human organs [6]. In addition to bat, recently, camel has also been found as an intermediate host of the MERS-CoV, and this discovery has attracted the attentions of scientists as the gene fragments and neutralizing antibodies have been reported from the camels in May 2015 [7]. From the recent studies, it has been found that the MERS-CoV receptor-binding domains interact easily with the DPP4 of the bat with no hindrance and restriction at the cellular level [8].

MERS-CoV virus affects the host cells by binding to the cellular receptor DPP4 via the S1 subunit and secondly by the formation of the fusion core by the S2 subunit [9]. The main mechanism of MERS-CoV is its binding to DPP4 on the cell surface that causes the mutations in human membranous glycoproteins and in mammals (i.e., camels). Human MERS-CoV virus differs from the camel MERS-CoV on the basis of their nucleotide sequences. MERS-CoV has some strategies to escape from antiviral responses as the virus contains many structural and nonstructural proteins that modulate the host immune and antiviral defence proteins. Combination therapy with ribavirin and beta-interferon has been used for early-stage patients, but in the future, it might be possible that the use of antiviral and combination therapy would be proven fruitful against MERS-CoV [10]. Recently, the implication of computational biology has opened a new door in the vicinity of drug designing. Molecular docking is a key technique to foresee the binding capacity and interactions of ligands to design drugs. The current study was therefore planned to target nonstructural protein (nsp13) of MERS-CoV and docked with plant phytochemicals to explore potential drug candidates against the virus.

2. Materials and Methods

A ready-to-dock library of 140 phytochemicals was prepared and docked against MERS-CoV nsp13 via the Molecular Operating Environment (MOE) software. MOE is a comprehensive suite specifically designed for protein/DNA modelling, protein structure analysis, drug designing, peptide modelling, molecular docking, and stimulations.

2.1. Ligand Database Preparation. An extensive literature study was performed to hunt for antiviral compounds from different medicinal plants with potential activity against different viral diseases. The chemical structures of all the phytochemicals were downloaded from the PubChem database (<http://www.ncbi.nlm.nih.gov/pccompound>) and saved in the MOE database after energy minimization.

2.2. Refinement of Receptor Protein. The three-dimensional (3D) structure of the MERS-CoV nsp13 was retrieved from the Protein Data Bank (PDB) using PDB ID 5WWP (<http://www.rcsb.org/pdb>) and optimized by removal of water molecules, addition of hydrogens, 3D protonation, and energy minimization using MOE.

2.3. Molecular Docking. The docking algorithm of MOE was used to dock a prepared ligand database with an active site of the MERS-CoV nsp13 receptor protein. A siteFINDER tool

was used to find the binding residues with default parameters such as rescoring 1: London dG, retain: 10, refinement: force field, rescoring 1: London dG, and retain: 10 to predict the interactions of ligands with active residues of the MERS-CoV nsp13. After docking, the *S*-score was the criteria to select the appropriate confirmations between ligands and receptor protein.

2.4. Drug Scan. A drug scan of selected phytochemicals was executed following the “Lipinski rule of five” (<http://www.scbio-iitd.res.in/software/drugdesign/lipinski.jsp>) in order to assure the appropriate molecular properties of the ligands as potential drug candidates. These rules predict the druggability of selected candidates on the bases of molecular mass (≤ 500 Dalton), hydrogen bond donors (< 5), hydrogen bond acceptors (< 10), molar refractive index (40-130), and $\log P$ (≤ 5) [11]. Evaluation of ADMET-based properties (i.e., absorption, distribution, metabolism, excretion, and toxicity) classifies the likeness and substructure recognition of lead compounds. Therefore, to further validate the potential and bioavailability of selected hits, all the compounds were subjected to the admetSAR server [12]. The admetSAR facilitates researchers to freely predict the ADMET properties of drug molecules from the perspective of medical chemistry.

3. Results

3.1. Molecular Docking. The library of 140 compounds was docked against nsp13 of MERS-CoV using MOE. Out of 140 docked ligands, only top 10 ligands with the best *S*-scores and binding patterns with interactive amino acid residues of the hotspot conserved regions of the binding pocket were selected for further analysis. MOE provides multiple conformations of each phytochemical, and on the basis of the minimum *S*-score, top 10 selected phytochemicals were rosavin, betaxanthin, quercetin, citromitin, pluviatilol, digitogenin, ichangin, methyl deacetylnomilinate, kobusinol A, and cyclocalamin. Details of each ligand including their *S*-scores, RMSD values, and interacting residues are shown in Table 1.

3.2. Interaction Analysis. Among the top 10 selected ligands, rosavin with the minimum *S*-score of -15.40 showed potential hydrophobic interactions with binding residues (i.e., Arg339, Asn361, Lys146, and Cys309) of the active site of the receptor protein. In this study, Arg339 was reported as the main interactive amino acid residue in all the interactions except for pluviatilol, ichangin, and cyclocalamin. Interactions between the ligands and the binding residues of the receptor protein are given in Table 1.

Rosavin showed interactions with four amino acid residues (i.e., Arg339, Asn361, Lys146, and Cys309) of the receptor protein (Figure S1). Rosavin is an O-acyl carbohydrate extracted from the plant *Rhodiola rosea* used as dietary herb for stress relief, CNS stimulation, and mental disorders. Betaxanthin has shown interactions with Arg339, His311, and Lys146 of the MERS-CoV nsp13 protein (Figure S2). Betalains are Tyr-derived plant pigments which provide red-violet (betacyanins) and yellow (betaxanthins) colours to

TABLE 1: Interaction of top ten phytochemicals with MERS-CoV nps13.

Sr. no.	PubChem ID	Compound name	S-score	RMSD value	Residues
1	9823887	Rosavin	-15.40	1.64	Arg339, Asn361, Lys146, Cys309
2	135926572	Betaxanthin	-14.64	2.45	Arg339, His311, Lys146
3	5280343	Quercetin	-14.24	1.13	Arg339, Arg139, Arg390, His230
4	12303287	Citromitin	-13.73	1.33	Arg339, Arg139, Thr382
5	70695727	Pluviatilol	-13.01	1.14	Arg560, Val533, His554, Pro514, Asn516
6	441886	Digitogenin	-13.15	0.83	Arg339, Arg139, Glu143, Lys146, Asn361
7	441801	Ichangin	-12.99	2.52	Arg202, Tyr205, Thr532, Asn177
8	272822441	Methyl deacetylnomilate	-12.45	2.17	Arg339, Arg139
9	274366689	Kobusinol A	-12.25	1.30	Arg339, Arg139, Asn361, His311
10	13857944	Cyclocalamin	-12.34	1.04	Arg139, Asn179, Lys146

TABLE 2: Pharmacokinetic parameters important for bioavailability of compounds' drug-likeness properties of selected phytochemicals.

Sr. no.	Ligand	Mass (<500 D)	HBD (≤ 5)	HBA (≤ 10)	Log P (<5)	Molar refractivity (40-130)	Violations
1	Rosavin	428	6	10	2.75	101.31	1
2	Betaxanthin	358	2	6	1.71	84.39	0
3	Quercetin	302	5	7	0.52	64.36	0
4	Citromitin	404	0	8	3.56	99.23	0
5	Pluviatilol	356	1	6	3.10	90.31	0
6	Ichangin	488	1	9	4.54	125.39	0
7	Digitogenin	448	3	5	6.06	140.22	2
8	Methyl deacetylnomilate	504	2	9	4.81	128.66	1
9	Kobusinol A	374	2	6	3.07	97.76	0
10	Cyclocalamin	502	1	9	4.90	130.08	1

HBD: hydrogen bond donors; HBA: hydrogen bond acceptors; log P : the logarithm of octanol/water partition coefficient.

fruits and vegetables. Betacyanins are heterocyclic plant pigments reported with antioxidant activity.

Quercetin is a polyphenolic flavonoid. It is a potential chemopreventive and anti-inflammatory agent and has been reported from various plants including *Allium cepa*, *Ginkgo biloba*, *Malus domestica*, and Buckwheat tea. Quercetin showed interactions with Arg339, Arg139, Arg390, and His230 residues of the viral protein (Figure S3). Similarly, citromitin interacted with Arg339, Arg139, and Thr382 residues of the receptor protein (Figure S4). Citromitin belongs to the family of organic compounds known as 8-o-methylated flavonoids. Citromitin is a hexamethoxyflavanone and has been reported as the major chemical constituent of *Citrus*.

Ichangin belongs to the sesquiterpenoid class of terpenes. It showed interactions with Arg202, Tyr205, Thr532, and Asn177 (Figure S5). Digitogenin is a steroidal saponin and also called as 2- α -hydroxy steroid, 3- β -hydroxy, or 15- β -hydroxy steroids. Digitogenin has been extracted from the plant *Digitalis purpurea* and reported with potential membrane-related applications such as antihyperlipidemia. Digitogenin exhibited interactions with Arg339, Arg139, Glu143, Lys146, and Asn361 residues of the viral protein (Figure S6). Methyl deacetylnomilate showed interactions with Arg339 and Arg139 of the receptor protein (Figure S7). Methyl deacetylnomilate belongs to the class of citrus limonoids present in lime, citrus, and grapefruit and has

been reported as an anti-inflammatory and anticancer agent. Cyclocalamin belongs to the class of organic compound known as steroid lactone. Cyclocalamin has been extracted from citrus and mandarin orange (tangerine). Cyclocalamin showed binding interactions with Arg139, Asn79, and Lys 146 of the receptor protein (Figure S8).

In the light of the current results, it is assumed that these natural phytochemicals would be proven as good antagonists against MERS-CoV (nsp13) protein to stop its proliferation in the future. Results also demonstrated the potency of these selected ligands as they can block the target site of the viral replicative nonstructural protein. The results from the interactions of ligands and target protein elucidate that these phytochemicals would be strong drug candidates against the MERS-CoV nsp13.

3.3. Drug Scan. The Lipinski rule of five (Ro5) illustrates the durability of potential drug candidates. According to this rule, the molecular mass of the molecule should be ≤ 500 Dalton, high lipophilicity should be expressed as $\log P < 5$, hydrogen bond donors should be ≤ 5 , hydrogen bond acceptors should be ≤ 10 , molar refractivity should be between 40 and 130. Only those that accomplished all five parameters could be proven as potential drug candidates. Among the top 10 ligands selected in this study, all ligands except methyl deacetylnomilate and cyclocalamin followed the Ro5. The

TABLE 3: ADMET profiling of the best selected phytochemicals.

	Rosavin	Betaxanthin	Quercetin	Citromitin	Pluviatilol	Ichangin	Digitogenin	MD	Kobusinol A	Cyclocalamin
Absorption										
BBB	-	+	-	-	+	+	+	+	+	+
HIA	-	+	+	+	+	+	+	+	+	+
Caco-2 permeability	-	-	-	+	+	-	-	-	+	+
PGS	NS	Substrate	NS	NS	NS	Substrate	NS	Substrate	NS	Substrate
PGI	NI	NI	NI	Inhibitor	Inhibitor	NI	NI	Inhibitor	Inhibitor	Inhibitor
ROCT	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
Metabolism										
CYP3A4 substrate	Substrate	Substrate	Substrate	Substrate	NS	Substrate	Substrate	Substrate	Substrate	Substrate
CYP2C9 substrate	Substrate	NS	NS	NS	NS	NS	NS	NS	NS	NS
CYP2D6 substrate	NS	NS	NS	Substrate	Substrate	NS	NS	NS	Substrate	NS
CYP3A4 inhibition	NI	NI	Inhibitor	Inhibitor	Inhibitor	Inhibitor	NI	Inhibitor	NI	Inhibitor
CYP2C9 inhibition	NI	NI	NI	NI	Inhibitor	NI	NI	NI	Inhibitor	NI
CYP2C19 inhibition	NI	NI	NI	Inhibitor	Inhibitor	NI	NI	NI	Inhibitor	NI
CYP2D6 inhibition	NI	NI	NI	NI	Inhibitor	NI	NI	NI	NI	NI
CYP1A2 inhibition	NI	NI	Inhibitor	Inhibitor	Inhibitor	NI	NI	NI	Inhibitor	NI
Toxicity										
AMES toxicity	NAT	NAT	NAT	NAT	NAT	NAT	NAT	NAT	NAT	NAT
Carcinogens	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC

MD: methyl deacetylnomilinate; BBB: blood-brain barrier; HIA: human intestinal absorption; PGS: P-glycoprotein substrate; PGI: P-glycoprotein inhibitor; ROCT: renal organic cation transporter; NS: nonsubstrate; NI: noninhibitor; NAT: non-Ames toxic; NC: noncarcinogenic.

drug assessment test also revealed fewer adverse effects of the selected ligands. Rosavin, methyl deacetylnomilinate, and cyclocalamin violated only one rule, so on the basis of overall drug profiling, these ligands could also be accepted as good drug candidates (Table 2).

ADMET properties of drug molecules were checked after the durability. All the selected top ten ligands accomplished the criteria of being good drug candidates as they are non-toxic and noncarcinogenic. Rosavin, betaxanthin, and citromitin showed a little variation, but the overall results could be accepted as they are nontoxic (Table 3).

The interactions and binding patterns of the best phytochemicals (i.e., pluviatilol and kobusinol A) with nsp13 of MERS-CoV are shown in Figures 1 and 2, respectively.

The fifth ligand pluviatilol is a principal lignan precursor from the class of polyphenol isolated from *Lindera obtusiloba*. Pluviatilol has been reported with antiallergic and anti-inflammatory activities. The binding patterns of pluviatilol showed interactions with Arg560, Val533, His554, Pro514, and Asn516 residues of the receptor protein.

4. Discussion

Middle East respiratory syndrome coronavirus (MERS-CoV) is a lethal zoonotic pathogen that causes severe respiratory illness which leads to respiratory and kidney failure. MERS-CoV was firstly reported in Saudi Arabia as a causative agent for respiratory and renal disease with approx. fatality rate of ~35%. In 2019, about 203 new cases of MERS were reported because there was no typical targeted vaccine for the MERS-CoV [13]. Most of the cases were associated with travellers, and predominately human-to-human associations and transmission were more common. Phylogenetic and epidemiological studies revealed bats as initial and camels as intermediate reservoirs for MER-CoV prevalence [14].

The phylogenetic study revealed a close relationship of MERS-CoV with the *Tylosycteris* bat coronavirus HKU4 and *Pipistrellus* bat coronavirus HKU5. This close genomic relationship suggests the zoonotic origin of MERS-CoV (from bat coronaviruses) [15]. Beta-coronavirus (β -CoVs) are positive-strand RNA viruses that belong to the family of Coronaviridae and order Nidovirales. The genome of a

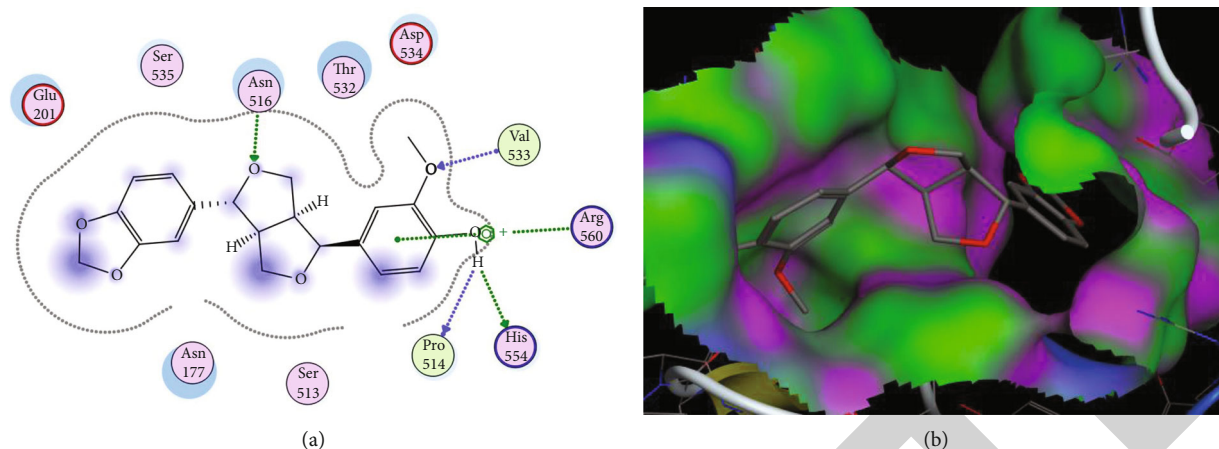


FIGURE 1: Interactions (a) and binding patterns (b) of pluviatilol with MERS-CoV nps13.

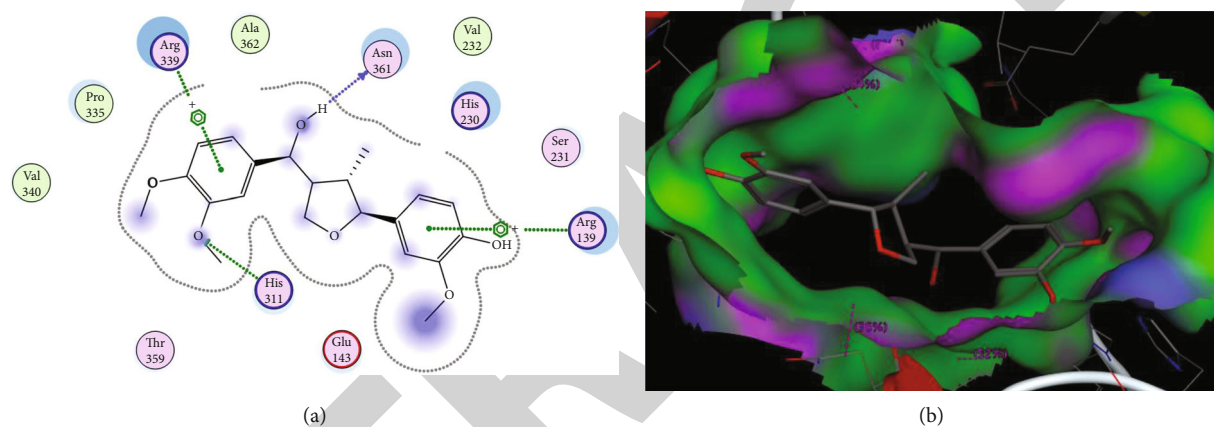


FIGURE 2: Interactions (a) and binding patterns (b) of kobusinol A with MERS-CoV nps13. Kobusinol A was the ninth ligand, and it showed interactions with Arg339, Arg139, Asn361, and His311 residues of the viral protein. Kobusinol A is a lignin and has been extracted from leaves of different plants, and it plays an important role as an antimicrobial agent against a variety of microbial diseases.

MERS-CoV encodes major structural and nonstructural proteins that mediate in viral replication and progression of the infection. The nsp13 used in this was also a nonstructural helicase of the MERS-CoV genome involved in multiple functioning. Spike glycoprotein displays a multidomain architecture and organized homodimer units that mediate the entry of the virus inside the host cell. The variety in sequence and structure of spike glycoprotein mediates virus to alter the mechanism of its fusion and attachment to the host receptors which leads towards the difficulty in targeting a specific site of the viral protein [14].

The domains in the spike protein are comprised of N-terminal and C-terminal. N-terminal subunits are comprised of folded domains that mediate viral attachment to DPP4 receptors in the host cell while C-terminal fuses the membranes of the virus with the host cell to provide spot for initiation of infection. Upon interaction with DPP4 on the epithelial cells, the virus infects the lungs of the host [16]. In the MERS-CoV genome, the 3-terminal genes encode proteins with structural and functional variability to make a complex which provides targets for antiviral agents. On the

other hand, the 5-terminal of the MERS-CoV genome is comprised of large open reading frames (i.e., ORF1a and ORF1b) which encode different nonstructural and structural proteins including spike, membrane, envelope, and nucleocapsid structural proteins [17].

Computational biology skills help scientists to evaluate the possibilities of binding patterns of different ligands before the experimental work in the laboratory [18]. Molecular docking is an elaborative method to discover the binding features of different small molecules as drugs against the target proteins [19–21]. The natural flora plays a vital role in drug screening and development [22–24]. The crude extracts of different medical plants have been reported with biologically active compounds with multiple clinical activities against many diseases [25, 26]. Biomedical or herbal decoction of many medicinal plants has been in practice since prehistoric times for management of multiple disorders [27]. For the past few years, medical plants have played an important role in the pharmaceuticals to treat many uncured diseases. Plants are good sources of phytochemicals that play role as inhibitors to the binding sites of the target moiety [28].

The purpose of this study includes the discovery of novel compounds as an antagonist of MERS-CoV nsp13. The MERS-CoV nsp13 is a nonstructural helicase that predominantly forms the core of membrane-bound replication-transcription complexes [29]. Extensive studies have been made to develop the vaccine against MERS-CoV, but the efforts have yet not proven successful. So, there is dire need to design new drugs against the virus.

The current study was based on the docking of different phytochemicals from different medicinal plants against the MERS-CoV nps13. The 3D structure of the target protein of Middle East respiratory CoV (human beta corona virus) was downloaded from PDB and saved in the MOE software after following the docking protocols. After docking, top 10 molecules were shortlisted on the bases of the best S-scores and interactions with amino acids at the active site of the viral protein using MOE software. Moreover, a detailed study for ADMET scanning of lead compounds was also carried out. All the selected compounds in this study showed no diverse effects in absorption in the human body. Distribution of a drug through the blood brain barrier is a key factor in the drug discovery as most of the drugs cannot cross the BBB due to the tight junctions around the brain [30]. Other types of models such as HIA (human intestine absorption), renal organic cation transporter, and P-glycoprotein substrate also help to evaluate the compounds as good drug candidates. Another model is comprised of the clusters of the isoenzyme cytochrome P50 involved in the metabolism of the drug, and 75% association of drug metabolism is linked with these enzyme clusters. From the results of the admetSAR screening, all the selected phytochemicals were found as nontoxic and noncarcinogenic.

From the results of interacting amino acid residues, Arg178, Arg339, His311, His230, Lys146, and Arg139 have been found as more common interacting residues in this study. In the chemical interactions, Arg339 and Arg139 as basic and directly interacting residues shared the common side chain position. Results demonstrated the potential of these ligands to bind and block the target site of the viral protein. It can be concluded that these 10 molecules can be used as drugs in the future with no severe side effects.

5. Conclusion

Beta coronavirus is the main causative agent of the Middle East respiratory syndrome, and instead of many contributions by scientists, no accurate drug against MERS-CoV has been discovered since its emergence. Through molecular docking, scientists find the potency of many natural compounds to act as drugs against disease causative agents. The present study discovered 10 strongly interacting phytochemicals (i.e., rosavin, betaxanthin, quercetin, citromitin, pluviatilol, digitogenin, ichangin, methyl deacetylnomilate, kobusinol A, and cyclocalamin) from different medicinal plants against MERS-CoV, and out of these two phytochemicals (i.e., pluviatilol and kobusinol A), all Ro5 and pharmacokinetic parameters were followed. On the bases of chemical interactions and drug scanning of these molecules, it was revealed that they would inhibit MERS-CoV replica-

tion. The aim of this study was to figure out the novel bioactive compounds from medicinal plants with the maximum potential to inhibit the target viral protein which could be used as efficient drug candidates against the MERS-CoV nps13.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

All authors of this article would like to thank Prince Sultan University for their financial and academic support to conduct this research and publish it in *BioMed Research International*.

Supplementary Materials

The supplementary file contains the interactions and binding patterns of rosavin, betaxanthin, quercetin, citromitin, ichangin, digitogenin, methyl deacetylnomilate, and cyclocalamin with nps13 of MERS-CoV (Figures S1–S8, respectively). (*Supplementary Materials*)

References

- [1] E. I. Azhar, D. S. C. Hui, Z. A. Memish, C. Drosten, and A. Zumla, "The Middle East respiratory syndrome (MERS)," *Infectious Disease Clinics*, vol. 33, no. 4, pp. 891–905, 2019.
- [2] B. T. Bradley and A. Bryan, *Emerging respiratory infections: the infectious disease pathology of SARS, MERS, pandemic influenza, and Legionella*, in *Seminars in diagnostic pathology*, Elsevier, 2019.
- [3] H. S. Mahrosh and G. Mustafa, "The COVID-19 puzzle: a global nightmare," *Environment, Development and Sustainability*, pp. 1–28, 2021.
- [4] F. Li and L. Du, *MERS coronavirus: an emerging zoonotic virus*, Multidisciplinary Digital Publishing Institute, 2019.
- [5] C. Durinx, A. M. Lambeir, E. Bosmans et al., "Molecular characterization of dipeptidyl peptidase activity in serum: soluble CD26/dipeptidyl peptidase IV is responsible for the release of X-Pro dipeptides," *European Journal of Biochemistry*, vol. 267, no. 17, pp. 5608–5613, 2000.
- [6] A. S. Omrani, J. A. Al-Tawfiq, and Z. A. Memish, "Middle East respiratory syndrome coronavirus (MERS-CoV): animal to human interaction," *Pathogens and global health*, vol. 109, no. 8, pp. 354–362, 2015.
- [7] G. Wong, W. Liu, Y. Liu, B. Zhou, Y. Bi, and G. F. Gao, "MERS, SARS, and Ebola: the role of super-spreaders in infectious disease," *Cell Host & Microbe*, vol. 18, no. 4, pp. 398–401, 2015.
- [8] W. Widagdo, N. M. Okba, W. Li et al., "Species-specific colocalization of Middle East respiratory syndrome coronavirus attachment and entry receptors," *Journal of virology*, vol. 93, no. 16, 2019.

Research Article

Sequence and Structural Characterization of Toll-Like Receptor 6 from Human and Related Species

Ghulam Mustafa , Hafiza Salaha Mahrosh , and Rawaba Arif 

Department of Biochemistry, Government College University, Faisalabad 38000, Pakistan

Correspondence should be addressed to Ghulam Mustafa; gmustafa_uaf@yahoo.com

Received 21 February 2021; Revised 17 March 2021; Accepted 26 March 2021; Published 12 April 2021

Academic Editor: Borhan Shokrollahi

Copyright © 2021 Ghulam Mustafa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Toll-like receptors (TLRs) play an important role in the innate immune response against various pathogens. They serve as expected targets of natural selection in those species which are adapted to habitats with contrasting pathogen burdens. Till date, sufficient literature about TLRs especially TLR6 is not available. The current study was therefore planned to show evolutionary patterns of human TLRs generally and TLR6 specifically along with their conservation and diversity. The study also deals with characteristic polymorphic patterns of TLR6 in humans which are involved in serious clinical consequences. The sequence analysis of TLR6 from different mammals revealed conserved regions in the protein sequence. With respect to TLR6 evolution, human showed a close evolutionary relationship with chimpanzee and orangutans, while monkeys were appeared in a separate clade showing a distant evolutionary relationship. Old World monkeys and New World monkeys made their separate clades but both have evolved from a common ancestor. The C-terminal of human TLRs (TLR1 to TLR10) exhibited more conservation as compared to other regions. The phylogram of human TLRs showed that TLR6 is closely related to TLR1 and both TLRs shared a common ancestor with TLR10. The domain analysis has revealed that TLR1 and TLR10 have least (i.e., 4) number of leucine-rich repeat (LRR) while TLR6 contains five LRRs. Three single nucleotide polymorphisms were found in TLR6 which were found to be associated with benign. Conclusively, the current comparative sequence analyses and phylogenetic analyses provided informative insights into the process of TLR evolution in mammals. Furthermore, the polymorphism analysis would serve as a useful marker in the early detection of susceptibility and resistance against cancers and other diseases in humans.

1. Introduction

The host immunity begins with first line of defence including skin, endothelium, lysozyme, saliva, gastric juice, and mucous membranes that prevent foreign infectious agents to reach into the cellular potential sites to cause infections. The second and third lines of defense are comprised of non-specific resistance (innate immunity) and specific resistance (acquired immunity), respectively. The former relies on inflammations and activation of phagocytic cells, and the latter is cell-mediated immunity linked with antibody production and activation of antigen-sensitized cytotoxic T cells. The professional T cells, B cells, leukocytes, phagocytes, and release of cytokines play a central role in the cellular immunity by neutralizing the threats [1].

Cell-mediated immunity plays a critical role in combating multiple types of infections and pathogenic attacks. Cell-mediated innate immunity is nonspecific to invaders, dependent on the phagocytes and proteins to recognize the conserved regions of pathogen-associated molecular patterns (PAMPs) and/or pathogenic endogenous damage-associated molecular patterns (DAMPs) to destroy them quickly [2]. The innate immunity is mainly involved the recognition of DAMPs and PAMPs via pattern recognition receptors (PRRs) [3]. Toll-like receptors (TLRs) are type I glycoproteins that belong to the family of PRRs and respond towards the special repertoire of pathogenic molecules including PAMPs and DAMPs. RLRs are pattern recognition receptors with 10-27 extracellular leucine-rich repeats, intracellular cytoplasmic Toll/interleukin-1 receptor (TIR) domain, and

a single transmembrane helix, and these are the structurally conserved regions of TLRs [4]. To date, 10 human TLRs (TLRs 1-10) and 13 murine TLRs (TLRs 1-9 and TLRs 11-13) have been reported. The TLR10 has been found to be nonfunctional in mouse [5].

Toll-like receptors are expressed on all innate immunity cells including macrophages, basophils, neutrophils, natural killer cells, and dendritic cells as well as on adaptive immune cell lymphocytes (T and B cells) [6]. The cellular repair mechanism is activated by activation of TLRs that stimulate the signalling cascade of host immune system which helps the release of immune modulators and cytokines [7]. The signalling cascade of TLRs depends on the nature of stimuli and is regulated by two distinct pathways (i.e., TIR-domain-containing adapter-inducing interferon- β - (TRIF-) dependent pathway and MyD88-dependent pathway). TLR3 and TLR4 utilize the TRIF-dependent pathway that leads towards the stimulation of type 1 interferons, and all TLRs except TLR3 utilize MyD88-dependent pathway associated with inflammatory cytokine production [6]. The overexpression of TLRs drastically alters the homeostasis of immune system by sustaining the levels of proinflammatory cytokines, and type 1 interferon contributes in the development and progressions of autoimmune diseases including lupus erythematosus, rheumatoid arthritis, Alzheimer's disease, and type I diabetes mellitus [6].

TLR6 forms a heterodimer with TLR2 to broaden the ligand capacity against different pathogens. The mutations in TLR1, TLR2, and TLR6 lead towards the progression of different autoimmune disorders. Different studies have supported the evidence of involvement of TLR6 in the progression of many autoimmune complications including sepsis, coronary artery disease, and intestinal inflammation [8–10]. Therefore, in current study, the evolutionary and genetic level study of TLR6 was conducted which will provide substantial knowledge about the variants and polymorphisms associated with different disorders. In this study, we have adopted computational biology mode of study to explore polymorphic residues and estimated evolutionary relationships of TLR6 among various mammals and conservation of human TLRs.

2. Materials and Methods

2.1. Retrieval of Human TLR6 Protein Sequence. The human TLR6 protein sequence was retrieved from NCBI's Entrez protein database with accession number: AAY88762.1 and analysed using PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool) [11] available on the NCBI website (<http://www.ncbi.nlm.nih.gov/>).

2.2. Phylogenetic Analysis. Along with human TLR6 protein sequence, forty most similar reported sequences of TLR6 from different mammals were also retrieved from protein database for phylogenetic systematics. The expected threshold value was set to 0.05 in PSI-BLAST, and only those sequences were selected which showed *E*-value better than threshold. All sequences were aligned using ClustalX and imported into the MEGA7 program [12] for manual align-

ment. Neighbor-joining (NJ) phylogenetic tree was reconstructed using MEGA7 with 100 bootstrap replicates [13].

2.3. Comparative Sequence Analysis and Domain Organization of Human Toll-Like Receptors. The protein sequences of all TLRs (i.e., TLRs 1-10) from humans were retrieved from NCBI's Entrez protein database to reveal their diversity and conservation at amino acid level. Multiple sequence alignment of protein sequences of all TLRs was performed through Geneious [14] for comparative study. An NJ phylogenetic tree of all human TLRs was also reconstructed using MEGA7 with 100 bootstrap replicates to explore their evolutionary relationships. To identify conserved protein domains in human TLRs, the SMART server (<http://smart.embl-heidelberg.de/>) was used with the default parameters [15].

2.4. Allelic Distribution and Polymorphism of Human TLR6 Gene. The database of single nucleotide polymorphism (dbSNP) of NCBI was explored to reveal allelic distribution and TLR6 polymorphism associated with diseases. The dbSNP contains single nucleotide variations, microsatellites, and small-scale insertions and deletions in humans. Only those SNPs of human TLR6 were selected in this study which exhibited clear clinical significance.

3. Results

BLAST has been the most popular algorithm for similarity search that can accommodate nucleotide or protein sequences. BLAST was used to identify local regions of similarity and statistical significance of TLR6 protein sequences from selected organisms. Geneious was employed to perform multiple sequence alignment (Figure 1). The truncated sequences were deleted, and longer sequences were shortened in the multiple sequence alignment to make them all equal in length. The consensus sequence is also shown above the alignment. To show identity among TLR6 sequences, bars are shown in the alignment. Higher the green bar in the alignment, higher is the identity among those regions. The comparative sequence analysis is showing greater diversity among TLR6 protein sequences among selected organisms in the regions of amino acids 1 to 100. High sequence conservation is revealed in the regions of 640 to 790.

3.1. Analysis of TLR6 Phylogenetic Tree. To establish the evolutionary relationships, a phylogenetic tree of TLR6 protein from human and forty members of mammals with maximum identity was reconstructed (Figure 2). The evolutionary history was inferred using the neighbor-joining method [16]. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches [17]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method [18] and are in the units of the number of amino acid substitutions per site. Overall mean distance was found to be 0.143 that shows estimates of average evolutionary divergence over all sequence pairs.

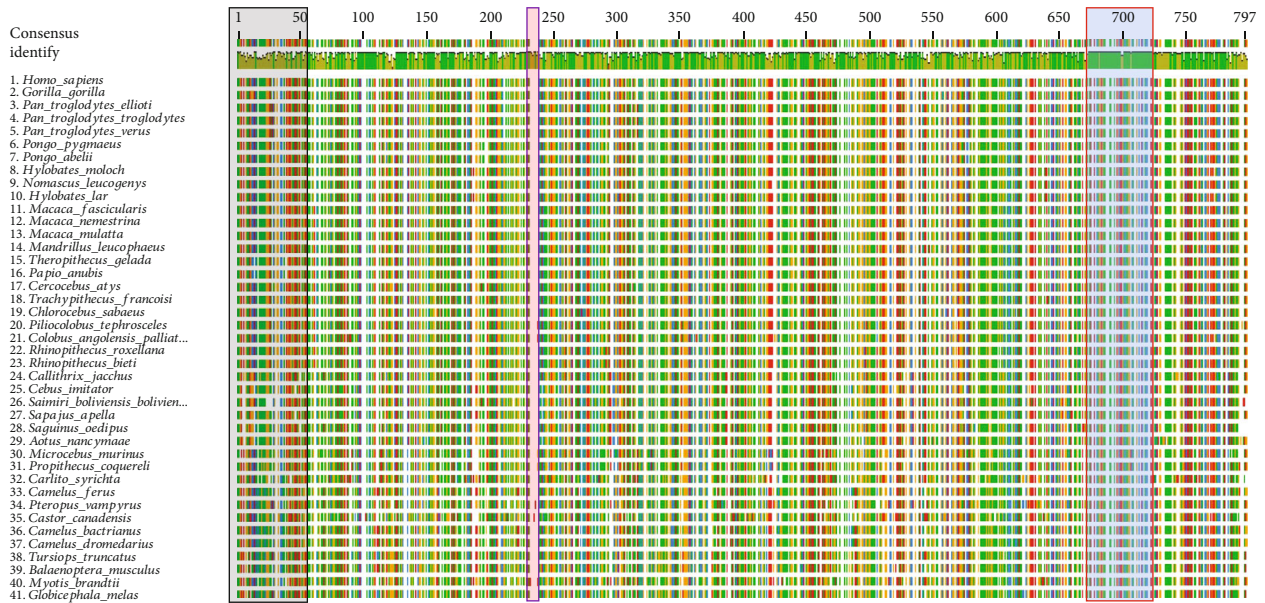


FIGURE 1: Multiple sequence alignment of TLR6 from selected organisms to show sequence similarities/differences. Similar color bars showing similar sequences. Red highlighted areas are showing regions present in the large flying fox (*Pteropus vampyrus*) and the North American beaver (*Castor canadensis*) but not in other mammals. Highly variable regions are depicted as black rectangle. The blue highlighted region is highly agreed with the consensus.

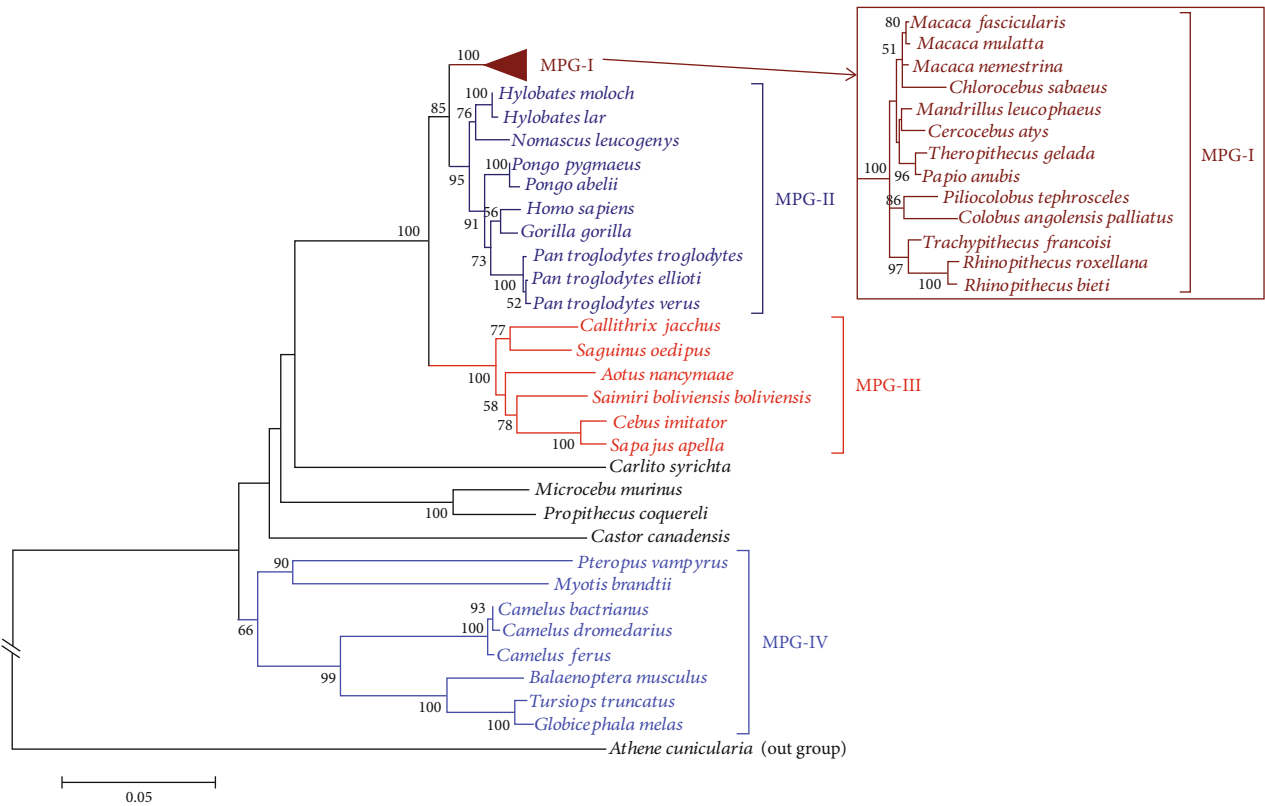


FIGURE 2: Phylogenetic relationships of TLR6 from human and other homologs. The phylogram is labelled on the basis of their monophyletic groups. The protein sequence of TLR1 from *Athene cunicularia* was used as an out-group.

The phylogram was fully resolved into different clades. The phylogram has been divided into four different monophyletic groups (i.e., MPG-I, MPG-II, MPG-III, and MPG-IV) based on cladding patterns of all taxa. *Homo sapiens* was appeared in MPG-II and showed close evolutionary relationships with gorilla and chimpanzee. Two orangutan species (i.e., *Pongo pygmaeus* and *Pongo abelii*) also appeared in this clade. The latter is one of the two species of Asian ape, while other orangutan species belong to rainforests of Borneo and Sumatra. Three gibbon species also joined the MPG-I (*Hylobates moloch*, *Hylobates lar*, and *Nomascus leucogenys*) showing their evolutionary closeness with other taxa in the group.

The MPG-I was found to be a clade of monkeys. Long-tailed or crab-eating macaque (*Macaca fascicularis*), rhesus macaque (*Macaca mulatta*), and southern pig-tailed macaque (*Macaca nemestrina*) were found to be closely related species, while all three distantly related to the sabaean monkey (*Chlorocebus sabaeanus*). The drill (*Mandrillus leucophaeus*) showed a close evolutionary relationship with the sooty mangabey (*Cercocebus atys*); whereas, gelada (*Theropithecus gelada*) and the olive baboon (*Papio anubis*) were found to be related species. Similarly, the Ashy red colobus monkey (*Piliocolobus tephrosceles*) and angolan colobus monkeys (*Colobus angolensis palliatus*) showed close evolutionary relationship. The leaf monkey or langur (*Trachypithecus francoisi*) was distantly related to two related species which were golden monkey (*Rhinopithecus roxellana*) and the black snub-nosed monkey (*Rhinopithecus bieti*).

The MPG-III was consisted of New World monkeys which are small to mid-sized primates. A node with 100 bootstrap value is showing that the New World monkeys and Old World monkeys shared a common ancestor. The Philippine tarsier (*Carlito syrichta*) and the North American beaver (*Castor canadensis*) both did not appear in any clade and showing divergence from other monophyletic groups. Two lemur species, the gray mouse lemur (*Microcebus murinus*) and coquerel's sifaka (*Propithecus coquereli*), exhibited a strong evolutionary relationship.

The fourth clade (i.e., MPG-IV) was consisted of diverse group of taxa. The large flying fox (*Pteropus vampyrus*) and a Brandt's bat (*Myotis brandtii*) were found to be evolutionary related species. Three *Camelus* species also appeared in this clade. Then, some marine organisms such as blue whale (*Balaenoptera musculus*), the common bottlenose dolphin or Atlantic bottlenose dolphin (*Tursiops truncatus*), and the long-finned pilot whale (*Globicephala melas*) also joined this clade with strong evolutionary relationships to each other. The protein sequence of TLR1 from a burrowing owl (*Athene cunicularia*) was used to root the tree.

3.2. Conservation and Diversification of Human TLRs. The protein sequence of human TLR6 was compared with other human TLRs (TLRs 1, 2, 3, 4, 5, 7, 8, 9, and 10), and their multiple sequence alignment is shown in Figure 3. More sequence conservation has been depicted at C-terminal of the TLR proteins. Dark shaded areas are highly conserved, light shaded areas are somewhat conserved, while unshaded areas are showing diversification among TLR sequences.

The sequence logo of human TLRs is shown in Figure S1. In logo, the conserved sequences are represented by large single-letter codes of amino acids at specific positions. The sequences which are missing from other TLRs are highlighted by pink rectangles.

3.3. Evolutionary Relationships and Domain Organization Analysis of Human TLRs. The multiple sequence alignment of human TLRs was further used to reconstruct a phylogenetic tree (Figure 4). Interesting relationships among human TLRs were revealed from this phylogram. The TLR6 was found to be closely related to TLR1 as both have evolved from a common ancestor that is supported by 100 bootstrap value at their ancestral node. Both TLRs further shared a common ancestor with TLR10 and showing their evolutionary relationship with TLR10. TLR1 and TLR10 showed minimum number of leucine-rich repeat (LRR) domains and also both have signal peptides, while TLR6 contains five LRR domains and also does not have a signal peptide. TLR2 and TLR4 are showing distant evolutionary relationships with TLR1, TLR6, and TLR10. A high bootstrap value (82%) on the ancestral node of TLR3 and TLR5 is showing that both TLRs have evolved in a close evolutionary relationship. TLR7 and TLR8 shared a common ancestor and showing that both are closely related which is supported by a high bootstrap value of 98%, and both TLRs further joined the ancestral node of TLR9 showing an evolutionary relationship with TLR9. It has been depicted by a high bootstrap value (93%) that TLR3 and TLR5 have close evolutionary relationship and shared a common ancestor with TLR9, TLR7, and TLR8 which is showing that a distant evolutionary relationship is present between both groups of human TLRs. TIR domain was found in all TLRs, while TLR3 and TLR9 showed maximum number of domains (i.e., 20). LRR carboxyl-terminal (LRR_CT) domain was not found in TLR7. LRR amino-terminal (LRR_NT) domain was found only in TLR3 and TLR7.

3.4. Comparative Sequence Analysis of Human TLR6 and TLR1. Based on phylogenetic analysis, the human TLR6 showed a strong evolutionary relationship with TLR1. Therefore, a sequence logo was generated through comparative sequence analysis of both proteins to access similarity between both proteins (Figure 5). The sequence logo is shown above the aligned sequences. The amino acids in their single-letter codes are shown in the sequence logo. Higher the size of the single-letter code of amino acids, higher the conservation is at that position. A high similarity between both sequences has been labelled in red rectangles. It has been revealed that both sequences are not highly similar in the region of 1 to 145.

3.5. SNPs of Human TLR6 Gene with Clinical Significance. The single nucleotide polymorphisms (SNPs) of human TLR6 gene with serious clinical significance are given in Table 1. Only the SNPs with known clinical significances were added. A total of four SNPs have been found in dbSNP that are associated with benign. Three variants of single nucleotide variations and one variant of deletion were found

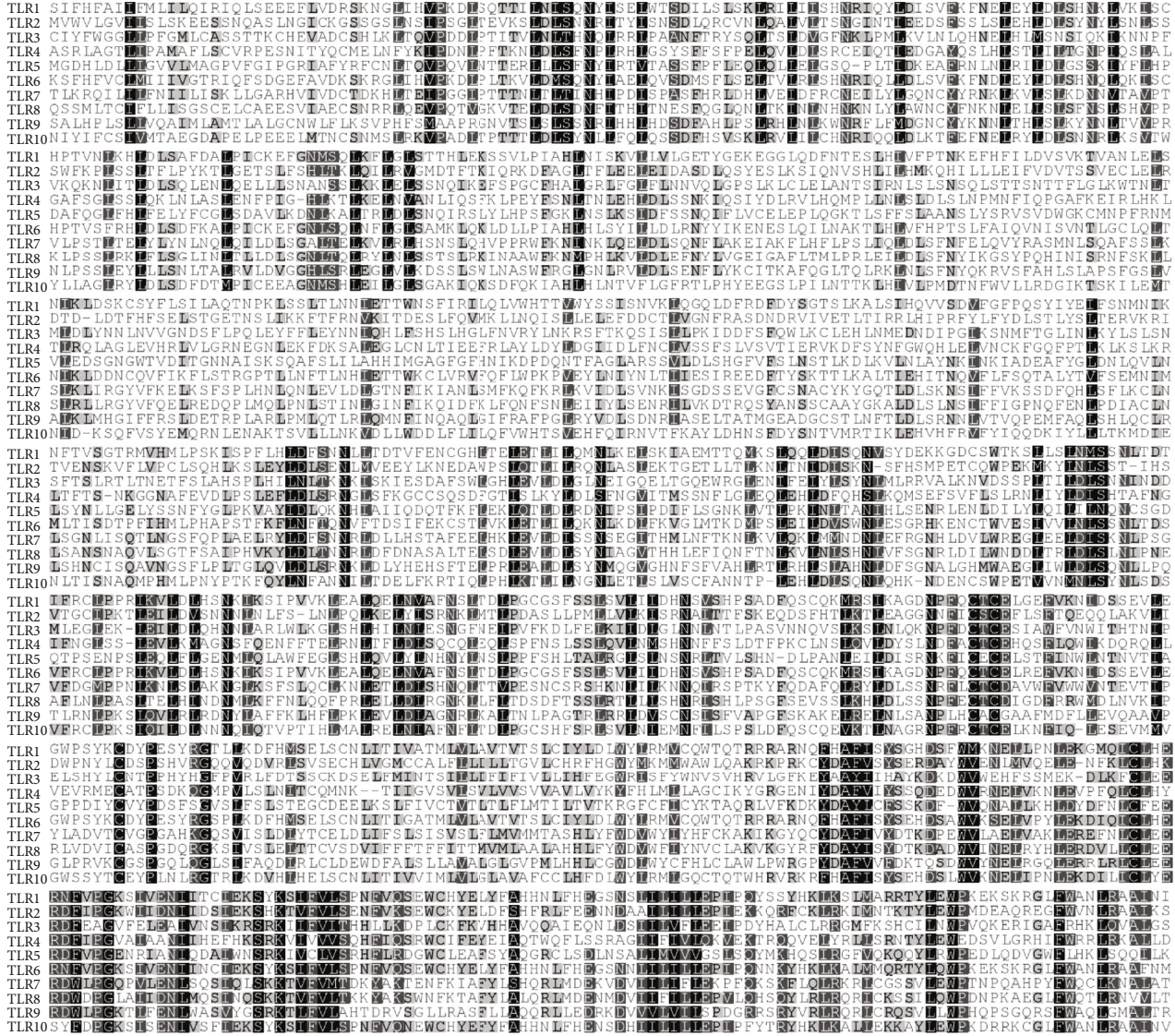


FIGURE 3: Multiple sequence alignment of human TLRs (TLR1 to TLR10). The dark shaded areas are showing conservation among various TLRs (more the intensity of the dark shaded area, higher is the conservation in that region). The conservation scale: black = conserved, gray = average, and white = variable.

in these SNPs. In first SNP, the thiamine (T) is replaced with cytosine (C) and the highest frequency of the alternate allele (C) has been found in African American. In second SNP, the cytosine (C) was the reference allele in TLR6 gene which is altered with thiamine (T) which causes a missense mutation. The highest frequency of the alternate allele (T) has been found in Europeans. Third SNP was also a SNV, and in this polymorphism, cytosine (C) has been altered with adenine (A) and causes a missense mutation with its highest frequency in South Asia. The last SNP was found to be a deletion in which adenine and thiamine (AT) were deleted and caused in frameshift mutation.

4. Discussion

The inflammatory responses have been closely associated with TLR6-mediated extracellular signal-regulated kinase (ERK) and p38 pathways in macrophages [19]. The onset of

single nucleotide polymorphisms (SNPs) in TLR6 encoding gene or protein leads towards alteration in the functioning of this PRR. The polymorphic residues on the amino acid sequence of TLRs result in SNPs that alter the inflammatory signalling cascade associated with different disorders [20]. In this study, we have demonstrated the phylogenetic relationships of TLRs specifically TLR6 reported in humans. In this phylogram, the TLR6 of human showed a strong evolutionary relatedness with gorilla. Woodman et al. [21] also conducted a phylogenetic analysis of TLR6 from different mammals including primates. In their phylogram, the human also showed a strong evolutionary relationship with chimpanzee and gorilla. In another study, Soni et al. [22] reconstructed a phylogenetic tree of DNA sequences of different TLRs (TLR1 to TLR10) of various vertebrates. They found that TLR2, TLR6, TLR9, and TLR10 were revealed to be conserved during the entire course of their evolution.

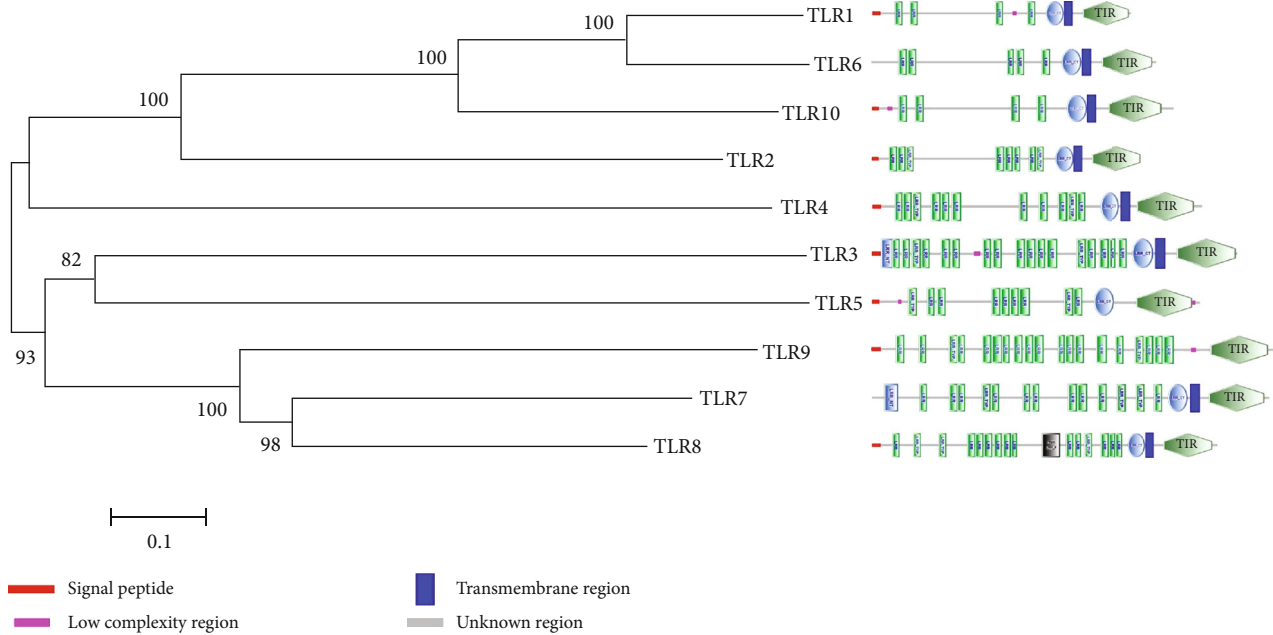


FIGURE 4: Phylogenetic tree of human TLRs. The evolutionary history was inferred using the NJ method. The bootstrap consensus tree inferred from 100 replicates is taken to represent the evolutionary history of the taxa analysed. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. Different colors and shapes are representing different domains and regions.

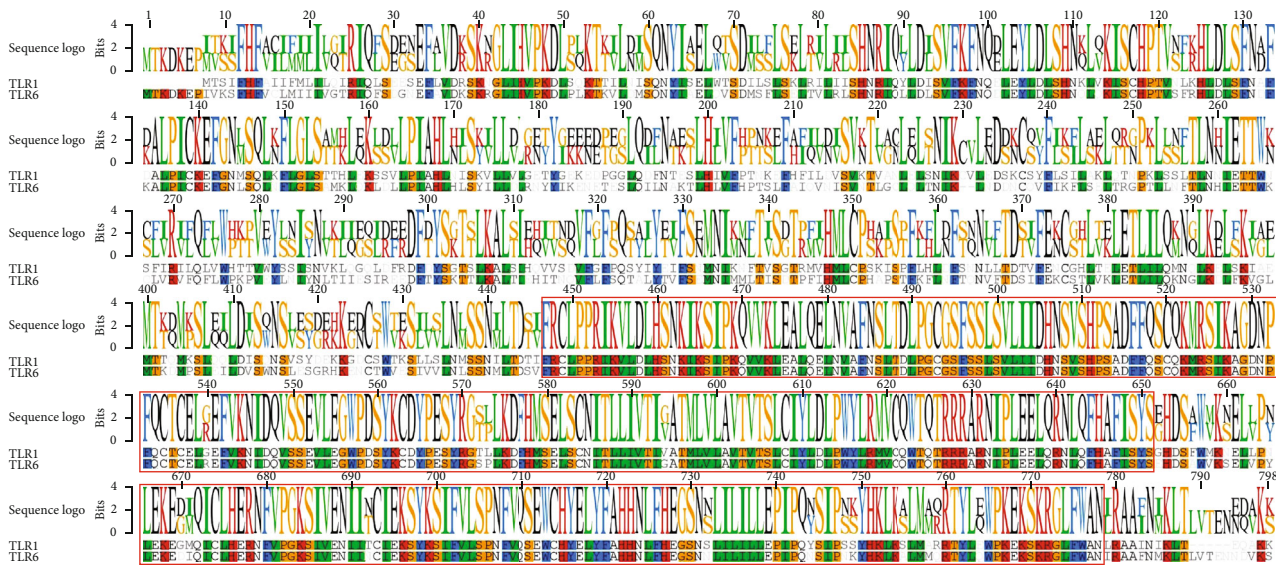


FIGURE 5: Comparison of human TLR6 and TLR1 protein sequences. Red rectangles are showing conservation between both sequences. Sequence logo has been generated on the basis of sequence conservation and shown above the alignment.

Phylogeny is an effective tool for determining the evolutionary relationships among various species in terms of protein and its functionality [23, 24]. To further elucidate the evolutionary relationships of TLRs, a phylogram among all human TLR proteins was also reconstructed. It was revealed that TLR6 is closely related to TLR1, and both TLRs were emerged from the ancestral node of TLR10. Similar findings were reported by Kumari and Singh [25] who conducted

phylogenetic analysis of TLR6 gene fragment of different animals. The TLR6 was found nearest to TLR1 and TLR10 mRNA of the *Sus scrofa*. The results of Banerjee et al. [26] are also in accordance to these findings. They have reported that genes of TLR1, TLR6, and TLR10 might have arisen from a gene duplication event which could be evidently proved by the presence of all three genes on the same chromosome (BTA6) and also all three genes belong to the same

TABLE 1: Distribution, polymorphism, and clinical significance of alleles of human TLR6 gene.

Sr. no.	Reference SNP	Variant type	Alleles	Functional consequence	Clinical significance	Allele frequency (global) [†]		Population with highest Alt allele frequency
						Ref allele	Alt allele	
1	rs5743812	SNV	T>C	Coding sequence variant, synonymous variant	Benign	T = 0.99731	C = 0.00269	African American (C = 0.0364)
2	rs5743816	SNV	C>T	Missense variant, coding sequence variant	Benign	C = 0.87967	T = 0.12033	European (T = 0.14193)
3	rs75244616	SNV	C>A	Missense variant, coding sequence variant	Benign	C = 0.99933	A = 0.00067	South Asian (A = 0.05)
4	rs863223364	DEL	AT>-	Frameshift variant, coding sequence variant	Benign	NA	NA	NA

SNV: single nucleotide variation; DEL: deletion; NA: not available. [†]The ALFA project provides aggregate allele frequency from database of Genotypes and Phenotypes (dbGaP).

family of TLR2. TLR2 and TLR4 appeared in the same clade in this study. Tania et al. [27] also reported that both genes are highly conserved across all mammalian species.

In mammals, the pattern recognition receptors have distinct classes including RIG-I-like receptors (RLRs), Toll-like receptors (TLRs), NOD-like receptors (NLRs), AIM2-like receptors (ALRs), and C-type lectin receptors (CLRs) which trigger the immediate host defence that leads to the induction of innate immunity responses [5, 28]. Among all the discovered PRRs, TLRs extensively initiate the survey of recognition of self- and non-self-antigens [29]. Toll-like receptors are generally expressed on cellular surface or in endosome in the dimer form, preferentially as homodimers or heterodimers. Among the reported Toll-like receptors, TLRs 1-2 and TLRs 4-6 are expressed on cell membrane, while TLR3, TLR7, TLR8, and TLR9 are localized in endosomes [30]. The TLRs are mainly associated with invader recognition features via eternal PAMPs of pathogens. These ligands can be characterized into three types, i.e., proteins are recognized by TLR5, TLR3, TLR7, TLR8, and TLR9; nucleic acids are recognized by TLR5; and lipids are recognized by TLR2/TLR6, TLR2/TLR1, and TLR4, respectively [28].

Each innate immune cell contains special TLRs that mediate in cellular immunity by recognizing the specific PAMP/DAMP and inducing proinflammatory cytokines. The intracellular Toll/interleukin-1 (IL-1) receptor (TIR) domains of TLRs play a leading role in the regulation of downstream signalling cascade [4]. TLRs reinvigorate adapter proteins that act as a platform for activations of IL-1R-associated protein kinases (IRAK) 1, 2, and 4 and TNF receptor-associated factor 6 (TRAF6) which translocate the NF- κ B, proinflammatory transcription factor, IRF3, and AP-1. Each transcriptional factor transcribes specific genes to encode different proteins such as proinflammatory cytokines, chemokines, antimicrobial peptides, and type 1 interferon.

The differences in producing immune molecules or polymorphisms in TLRs not only exhibit substantial influence on responses to a wide range of pathogens but are also associated with susceptibility and resistance to different diseases. In this study, a total of four SNPs were found in human TLR6 gene with serious clinical significance and result in the develop-

ment of benign. In a study, Elmaghraby et al. [31] have reported two novel nonsynonymous SNPs in TLR6 which were resulted from transversions. In another study, Mariotti et al. [32] found 855G>A SNPs and 2315T>C SNPs in the TLR6. In a study of prostate cancer, nine TLR6 SNPs were reported by Sun et al. [33]. Similarly, Noreen and Arshad [34] have reported that a C>T SNP (C745T) is associated with asthma and colitis, and an A>G SNP (A1401G) is associated with an increased risk of prostate cancer. Another SNP A>C, G,T (rs5743810) affects the expression and function of TLR6. The change in amino acid from Ser to Pro reduces the function of TLR6 and weakens the regulation of innate immune system in predisposed individuals. This mutation can potentially develop cancer [35]. The polymorphisms in the coding regions are believed to be involved in cancer development because of their roles in gene expression regulation. It has been found that TLR6 functionally interacts with TLR2 for mediating the cellular response against bacterial lipoproteins and causes the activation of NF- κ B pathway and inflammatory events. This activation further contributes towards tumor development and progression [36]. In spite of its enhanced expression and activation, the function and role of TLR6 in cancer are still not clearly understood. Identification of single nucleotide polymorphism (SNP) in genes involved with the innate immune response can be a useful marker in early detection of resistance or susceptibility in humans.

5. Conclusion

The genes encoding TLRs play an important role in the early defense against pathogens. The evolutionary history of TLR6 and evolutionary relationships of human TLRs were studied through phylogenetic analyses. Direct sequence comparisons and standard evolutionary approaches were employed to determine sequence conservation and diversity in human TLRs and TLR6 between human and other closely related species. In addition, the comparison of patterning of human TLR domains was performed. Single nucleotide polymorphisms involved in the development of benign were also revealed from human TLR6. These findings will facilitate the further exploration of TLRs especially TLR6 roles in regulating the immune system of human.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors would like to gratefully acknowledge the Department of Biochemistry, Government College University Faisalabad, for providing space and facilities to accomplish this study.

Supplementary Materials

Figure S1: sequence logo of human TLRs. In logo, the conserved sequences are represented by large single-letter codes of amino acids at specific positions. The sequences which are missing from other TLRs are highlighted by pink rectangles. (*Supplementary Materials*)

References

- [1] K. M. Hedayat and J.-C. Lapraz, *The Theory of Endobiogeny: Volume 3: Advanced Concepts for the Treatment of Complex Clinical Conditions*, Academic Press, 2019.
- [2] J. Zindel and P. Kubers, "DAMPs, PAMPs, and LAMPs in immunity and sterile inflammation," *Annual Review of Pathology: Mechanisms of Disease*, vol. 15, no. 1, pp. 493–518, 2020.
- [3] L. Nie, S.-Y. Cai, J.-Z. Shao, and J. Chen, "Toll-like receptors, associated biological roles, and signaling networks in non-mammals," *Frontiers in Immunology*, vol. 9, 2018.
- [4] W. Gao, Y. Xiong, Q. Li, and H. Yang, "Inhibition of Toll-like receptor signaling as a promising therapy for inflammatory diseases: a journey from molecular to nano therapeutics," *Frontiers in Physiology*, vol. 8, 2017.
- [5] T. Kawasaki and T. Kawai, "Toll-like receptor signaling pathways," *Frontiers in Immunology*, vol. 5, 2014.
- [6] S. R. El-Zayat, H. Sibaii, and F. A. Mannaa, "Toll-like receptors activation, signaling, and targeting: an overview," *Bulletin of the National Research Centre*, vol. 43, no. 1, 2019.
- [7] Y. Wang, E. Song, B. Bai, and P. M. Vanhoutte, "Toll-like receptors mediating vascular malfunction: lessons from receptor subtypes," *Pharmacology & Therapeutics*, vol. 158, pp. 91–100, 2016.
- [8] L. Choteau, H. Vancaeyneste, D. Le Roy et al., "Role of TLR1, TLR2 and TLR6 in the modulation of intestinal inflammation and *Candida albicans* elimination," *Gut Pathogens*, vol. 9, no. 1, 2017.
- [9] H. Wang, Y. Li, Y. Wang, H. Li, and L. Dou, "MicroRNA-494-3p alleviates inflammatory response in sepsis by targeting TLR6," *European Review for Medical and Pharmacological Sciences*, vol. 23, no. 7, pp. 2971–2977, 2019.
- [10] C. Wang, Q. Li, H. Yang et al., "MMP9, CXCR1, TLR6, and MPO participant in the progression of coronary artery disease," *Journal of Cellular Physiology*, vol. 235, no. 11, pp. 8283–8292, 2020.
- [11] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.
- [12] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets," *Molecular Biology and Evolution*, vol. 33, no. 7, pp. 1870–1874, 2016.
- [13] G. Afzal, G. Mustafa, S. Mushtaq, and A. Jamil, "DNA barcodes of Southeast Asian spiders of wheat agro-ecosystem," *Pakistan Journal of Zoology*, vol. 52, no. 4, 2020.
- [14] M. Kearse, R. Moir, A. Wilson et al., "Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data," *Bioinformatics*, vol. 28, no. 12, pp. 1647–1649, 2012.
- [15] G. Mustafa, R. Arif, S. A. Bukhari, M. Ali, S. Sharif, and A. Atta, "Structural and functional annotation of citrate synthase from *Aspergillus niger* ANJ-120," *Pakistan Journal of Pharmaceutical Sciences*, vol. 31, 2018.
- [16] N. Saitou and M. Nei, "The neighbor-joining method: a new method for reconstructing phylogenetic trees," *Molecular Biology and Evolution*, vol. 4, no. 4, pp. 406–425, 1987.
- [17] J. Felsenstein, "Confidence limits on phylogenies: an approach using the bootstrap," *Evolution*, vol. 39, no. 4, pp. 783–791, 1985.
- [18] E. Zuckerkandl and L. Pauling, "Evolutionary divergence and convergence in proteins," in *Evolving Genes and Proteins*, pp. 97–166, Elsevier, 1965.
- [19] R. Roy, S. K. Singh, M. Das, A. Tripathi, and P. D. Dwivedi, "Toll-like receptor 6 mediated inflammatory and functional responses of zinc oxide nanoparticles primed macrophages," *Immunology*, vol. 142, no. 3, pp. 453–464, 2014.
- [20] S. Mukherjee, S. Huda, and S. P. Sinha Babu, "Toll-like receptor polymorphism in host immune response to infectious diseases: a review," *Scandinavian Journal of Immunology*, vol. 90, no. 1, p. e12771, 2019.
- [21] S. Woodman, A. J. Gibson, A. R. Garcia et al., "Structural characterisation of Toll-like receptor 1 (TLR1) and Toll-like receptor 6 (TLR6) in elephant and harbor seals," *Veterinary Immunology and Immunopathology*, vol. 169, pp. 10–14, 2016.
- [22] B. Soni, B. Saha, and S. Singh, "Systems cues governing IL6 signaling in leishmaniasis," *Cytokine*, vol. 106, pp. 169–175, 2018.
- [23] R. Arif, S. Ahmed, and G. Mustafa, "In silico study to reveal annotation and significant interactions of human defensin with its isoforms and their phylogeny," *Indian Journal of Pharmaceutical Sciences*, vol. 82, no. 3, 2020.
- [24] U. Hameed, S. Khalid, S. Javed, and G. Mustafa, "Importance of cytochrome oxidase subunit 1 in phylogeny reconstruction of various classes of subphylum vertebrata Usman Hameed," *Pure and Applied Biology*, vol. 9, no. 3, 2020.
- [25] N. Kumari and L. Singh, "TLR6 gene polymorphism, sequence analysis, and phylogenetic tree in *Sus scrofa*," *The Indian Journal of Veterinary Sciences and Biotechnology*, vol. 12, no. 4, 2017.
- [26] P. Banerjee, S. K. Gahlawat, J. Joshi, U. Sharma, M. S. Tandia, and R. K. Viji, "Sequencing, characterization and phylogenetic analysis of TLR genes of *Bubalus bubalis*," *DHR-IJBLS*, vol. 3, pp. 137–158, 2012.
- [27] M. Tandia, B. Mishra, P. Banerjee, J. Joshi, S. Upasna, and R. Viji, "Phylogenetic and sequence analysis of Toll like receptor genes (TLR-2 and TLR-4) in buffaloes," *Indian Journal of Animal Sciences*, vol. 82, no. 8, 2012.

- [28] M. Farrugia and B. Baron, "The role of Toll-like receptors in autoimmune diseases through failure of the self-recognition mechanism," *International Journal of Inflammation*, vol. 2017, 12 pages, 2017.
- [29] M. K. Vidya, V. G. Kumar, V. Sejian, M. Bagath, G. Krishnan, and R. Bhatta, "Toll-like receptors: significance, ligands, signaling pathways, and functions in mammals," *International Reviews of Immunology*, vol. 37, no. 1, pp. 20–36, 2018.
- [30] T. Kawai and S. Akira, "TLR signaling," in *Seminars in Immunology*, pp. 24–32, Elsevier, 2007.
- [31] M. Elmaghraby, A. El-Nahas, M. Fathala, F. Sahwan, and M. T. El-Dien, "Association of Toll-like receptors 2 and 6 polymorphism with clinical mastitis and production traits in Holstein cattle," *Iranian Journal of Veterinary Research*, vol. 19, no. 3, 2018.
- [32] M. Mariotti, J. L. Williams, S. Dunner, A. Valentini, and L. Pariset, "Polymorphisms within the Toll-like receptor (TLR)-2, -4, and -6 genes in cattle," *Diversity*, vol. 1, no. 1, pp. 7–18, 2009.
- [33] J. Sun, F. Wiklund, F.-C. Hsu et al., "Interactions of sequence variants in interleukin-1 receptor-associated kinase4 and the toll-like receptor 6-1-10 gene cluster increase prostate cancer risk," *Cancer Epidemiology Biomarkers & Prevention*, vol. 15, no. 3, pp. 480–485, 2006.
- [34] M. Noreen and M. Arshad, "Association of TLR1, TLR2, TLR4, TLR6, and TIRAP polymorphisms with disease susceptibility," *Immunologic Research*, vol. 62, no. 2, pp. 234–252, 2015.
- [35] A. Semaili, M. Almutairi, M. Rouabhia et al., "Novel sequence variants in the TLR6 gene associated with advanced breast cancer risk in the Saudi Arabian population," *PloS One*, vol. 13, no. 11, p. e0203376, 2018.
- [36] E. J. Hennessy, A. E. Parker, and L. A. O'neill, "Targeting Toll-like receptors: emerging therapeutics?," *Nature Reviews Drug Discovery*, vol. 9, no. 4, pp. 293–307, 2010.

Research Article

Molecular Characterization of MHC Class I Genes in Four Species of the *Turdidae* Family to Assess Genetic Diversity and Selection

Muhammad Usman Ghani ¹, Li Bo ¹, An Buyang ², Xu Yanchun ¹, Shakeel Hussain,¹ and Muhammad Yasir ³

¹College of Wildlife Resources and Protected Area, Northeast Forestry University, Harbin 150040, China

²Department of Stem Cell Biology and Medicine, Graduate School of Medical Science, Kyushu University, Fukuoka 810-0000, Japan

³Department of Life Science and Technology, Huazhong Agricultural University, Wuhan, China

Correspondence should be addressed to Muhammad Usman Ghani; drusmanghani466@gmail.com and Li Bo; libo_770206@126.com

Received 9 February 2021; Revised 9 March 2021; Accepted 19 March 2021; Published 12 April 2021

Academic Editor: Hafiz Ishfaq Ahmad

Copyright © 2021 Muhammad Usman Ghani et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In vertebrate animals, the molecules encoded by major histocompatibility complex (MHC) genes play an essential role in the adaptive immunity. MHC class I deals with intracellular pathogens (virus) in birds. MHC class I diversity depends on the consequence of local and global environment selective pressure and gene flow. Here, we evaluated the MHC class I gene in four species of the *Turdidae* family from a broad geographical area of northeast China. We isolated 77 MHC class I sequences, including 47 putatively functional sequences and 30 pseudosequences from 80 individuals. Using the method based on analysis of cloned amplicons ($n = 25$) for each species, we found two and seven MHC I sequences per individual indicating more than one MHC I locus identified in all sampled species. Results revealed an overall elevated genetic diversity at MHC class I, evidence of different selection patterns among the domains of PBR and non-PBR. Alleles are found to be divergent with overall polymorphic sites per species ranging between 58 and 70 (out of 291 sites). Moreover, transspecies alleles were evident due to convergent evolution or recent speciation for the genus. Phylogenetic relationships among MHC I show an intermingling of alleles clustering among the *Turdidae* family rather than between other passerines. Pronounced MHC I gene diversity is essential for the existence of species. Our study signifies a valuable tool for the characterization of evolutionary relevant difference across a population of birds with high conservational concerns.

1. Introduction

The major histocompatibility complex (MHC) is a group of molecules encoded by certain genes that are most polymorphic to have been described in vertebrates' genomes [1]. Two types of MHC gene families, class I and class II, are useful to cell surface glycoproteins that regulate the immune response. MHC class II molecules are heterodimers consisting of an α chain and a β chain; both contribute to presenting peptides from the processing of extracellular pathogens such as bacteria to the CD4+ T-helper cells [2]. Heterodimer mol-

ecules of MHC class I are made up of an α chain and a non-MHC molecule, the $\beta 2$ microglobulin. The α chain constitutes a cytoplasmic tail, a transmembrane domain, and three extracellular domains named $\alpha 1$, $\alpha 2$, and $\alpha 3$ [3] that are encoded by exons 2, 3, and 4, respectively. The MHC class I molecules are expressed in almost all somatic cells and trigger an adaptive immune response by presenting endogenously derived peptides of viral protein and an individual's own body cells to CD8+ cytotoxic T-cells [4]. Polymorphism is largely confined within the region encoding the ABS (antigen-binding site) of MHC class I [5]. Maintenance of

surprising diversity is supposed to take place by two types of selection: heterozygote advantage and frequency-dependent selection. Heterozygotes could recognize a broader range of antigens from multiple pathogens and therefore have more fitness than either individual having a homozygote [6]. Other is frequency-dependent selection, in which rare alleles deliver a selective advantage where pathogens have found a means to escape against common immune defensive alleles in the population. Thus, alteration in the pathogen community with time and locality results in MHC variation in the host population. Generally, in an individual possessing huge numbers and diverse MHC alleles; more pathogens can be recognized [1].

Structural diversity and immune response have been explored in numerous research, including genomics [7, 8], ailment [9–11], and mate choice [12–14]. Sequence similarity at PBR-based assignment to the locus is frequently hampered by various evolutionary indicators due to current recombination, duplication, and/or concerted evolution as well as positive selection mediated by a variety of pathogens [15]. Thus, numerous studies emphasized MHC genes as important markers to evaluate the adaptive potential and evolutionary status of a threatened population [16].

The emerging scenario inspires researchers to collect statistics from a group of wild taxa to enlarge our understanding of the evolution of the MHC gene [17]. Despite significant efforts, protocols for locus-specific MHC genotyping in avian are still difficult to achieve and remarkably rare [18]. MHC studies in population of wild birds remain neglected possibly due to complications in amplifying gene sequences from bird species not closely related to systematically studied chicken [19, 20].

A significant decline in habitats and fragmentation of available habitats are predisposing factors for dramatic deterioration in population sizes [21]. The avian genus *Turdus* is one of the broadly distributed passerine genera, with 65 documented extant species. The genus is listed wild territorial birds that are beneficial to china having economic and research value. Birds of this genus are strongly migratory thus experiencing a variety of environments. Up to the present, there are no studies on MHC class I genes in *Turdidae* species, which is the first step towards exploring the role of selection mediated by pathogens in the maintenance of MHC class I diversity. Precisely, this study aims to (1) Measured locus-specific variation in MHC I exon 3 genes across the *Turdidae* family to evaluate the mode of evolution by which such variation comes about. To achieve this, we have measured the diversity and selection at MHC I genes to make available the variations that exist across the *Turdidae* family. (2) We investigate the numbers of alleles possessed by each species and the general features of alleles in terms of functional genetic diversity. (3) Phylogenetic analyses to assess evolutionary relationships and processes driving avian MHC I diversity among four species of the *Turdidae* family and other avian species.

2. Material and Methods

2.1. Study Population. The study population was non-sympatrically distributed 80 individuals of four species of

genus *Turdus* of the *Turdidae* family. Samples include two to three contour feathers, tissue from breast and liver of birds accidentally injured or died during migratory season of 2017–19 in autumn and deposit in State key laboratory of wildlife detection center in northeast forestry university, stored at 4°C. The geographical location of sample material is presented (Figure 1).

2.2. Extraction of Genomic DNA. Region of calamus to the rachis of contour feathers was excised, tissues from skeletal muscles were minced, placed into a 1.5 ml Eppendorf tube containing TNE buffer (10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM EDTA, 1% SDS). Total genomic DNA was extracted with AxyPrep Multisource Genomic DNA Miniprep Kit (AXYGEN, China) according to the manufacturer's instructions. The DNA concentration was measured with Nanopore Spectrophotometer at 260 nm absorbance. Samples above 100 ng/μl concentration were used for further analysis.

2.3. PCR, Cloning, and Sequencing. Polymerase chain reaction was conducted using motif specific primers designed for the amplification of MHC class I genes in great reed warbler. The forward primers HN36 5'-TCCCCACAGGTCTC CACACAGT-3' and HN46 reverse 5'-ATCCCAAATTC CCACCCACCTT-3' correspond to exon 3 region in the flanking introns, the region coding most of the peptide-binding site in MHC molecules (subunit α2) [22–24]. The primers were used due to their successful amplification in many passerine species. Amplification was performed in the reaction mixture containing 20 ng DNA template, 0.2 μM of each primer, 25 μl 2× EasyTaq® PCR SuperMix (+dye) (Trans, China), and water (deionized) to reach 50 μl as final volume. Thermal cycling for MHC class I amplification began with one cycle at 94°C for 5 min, followed by 30 cycles of denaturation consisting of sequential steps of 94°C for 30s, 52°C for 30s, and 72°C for 30s, ending with a single extension step at 72°C for 5 min. Purification was carried out with AxyPrep™ DNA Gel Extraction Kit in accordance with the manufacturer's protocol. Purified PCR product was cloned using pEASY®-T5 Zero Cloning Kit containing Trans1-T1 Phage resistant chemically competent cells (Transgen Biotech). PCRs were performed for positive clones using M13 forward and reverse primers. Several colonies (20–25) per individual were selected and used as a template for sequencing directionally on an automatic sequencer (ABI PRISM 3730; Invitrogen Biotechnology Co. Ltd.).

2.4. Definition of Allele. Since few artifacts introduced during the recombination of PCR products in cloning [25, 26]. Amplification, cloning and sequencing were performed twice. Sequences were verified and referred to as an Allele; either minimum of three sequences have the same nucleotide composition or repeated in both events. The sequences which showed any deletion, insertion, or premature stop codons within exons were identified as presumed pseudogene sequence, and others were considered as putative functional allele (PFA) [27]. All sequences appropriate to our criteria

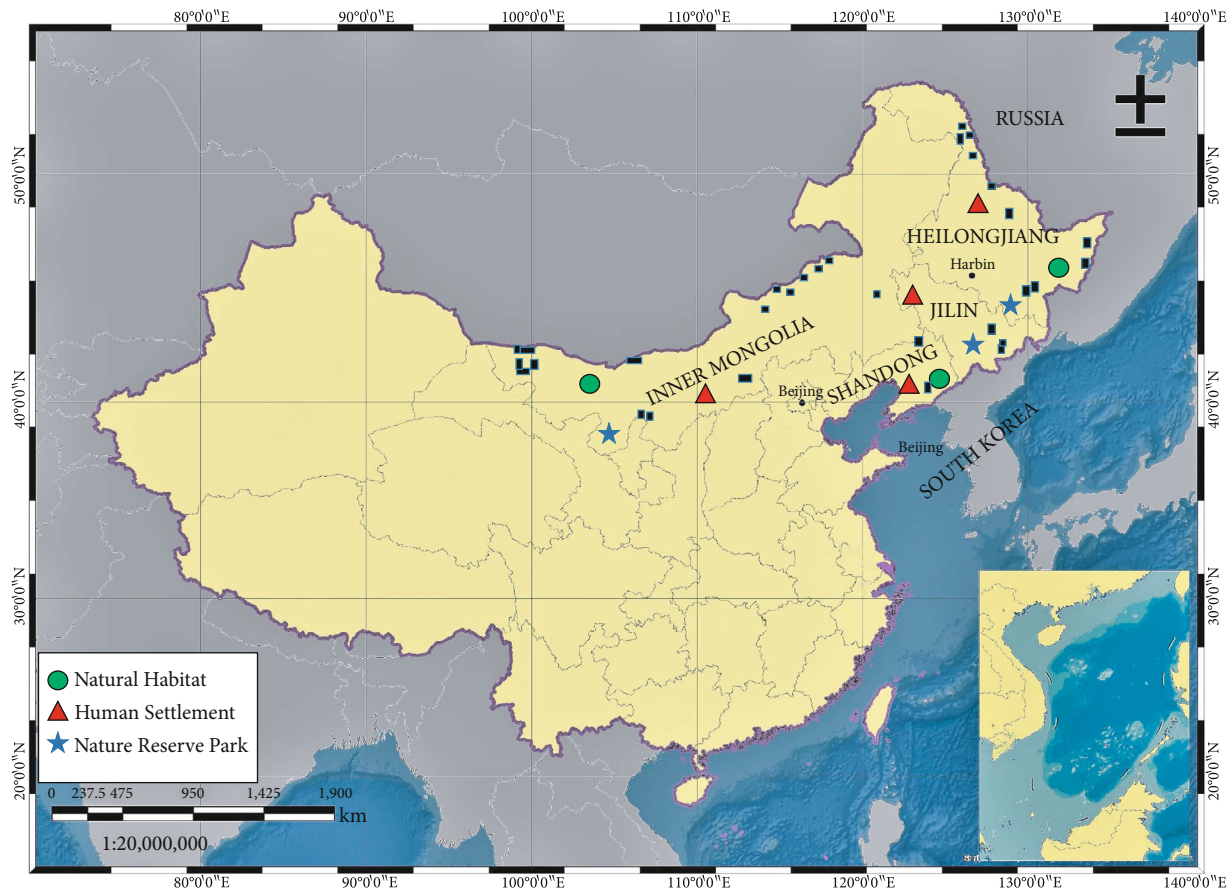


FIGURE 1: Geographical locations of samples included in our study. Square represents the actual site of the sample, and size of the square represents the approximate diameter of the sample's geographical range.

have been deposited into the GenBank (Accession No: MN849308-54).

2.5. Data Analysis

2.5.1. Sequence Analysis. Chromatogram signals of all sequencing were examined with chromas 2.2.6. Sequences without ambiguous signals were selected. Vector sequence from the MHC class I gene was removed using seqMan in the DNASTar7.1 package. Sequence editing and organization were done with BioEdit [28]. Sequences were aligned individually and then altogether four sampled species using CLUSTAL X [29]. The unique alleles were named according to the nomenclature for MHC in non-human species [30]. NCBI BLAST [31] was used for sequences confirmation representing close identity to passerine species previously published MHC class I exon 3 sequences. Sequences having at least one stop codon (shift in the reading frame due to indels or nonsense sequences) were classified as pseudogenes. Based upon sequences found to be translatable, a minimum number of functional loci MHC class I was estimated using a conservational approach that all Loci from samples species' individual were in heterozygote state.

The average pairwise nucleotide distances (Kimura 2-parameter model - K2P), and the Poisson-corrected amino acid distances were calculated using MEGA7.0. Standard

errors were obtained through 1000 bootstrap replicates. Haplotypes identification (N_a), the average number of nucleotide differences (K), polymorphic sites (S) and nucleotide diversity (π) were measured by DnaSP 5.10 [32].

2.6. Inference of Recombination. Recombination can influence the outcomes of selection, we first tested recombination. Analyses were implemented for the nucleotide alignment of exon 3 in the Recombination Detection Program version 4 (RDP4). Several method, including RDP [33], GENECONV [34], Chimaera [35], MaxChi [36], BootScan [33], SiScan [37], and 3Seq [38], were used to detect recombination events. In addition, the online GARD tool, provided by the Datamonkey webserver (<http://www.datamonkey.org/>), was used for recombination signals assessment [39].

2.7. Tests for Selection. For selection, we conduct a priori classification of peptide binding region (PBR) and non-peptide region upon inferred passerine PBR sequences [40, 41] homology sites with chicken MHC [42, 43] and human HLA [44]. The identification of sites subjected to selection in MHC class I Exon 3 was performed using various methods. The first standard selection test (Tajima's D , Fu and Li's F^* , and Fu and Li's D^*) were calculated using DnaSP 5.0 [32]. Second method was the calculation of parameter (ω) for functional alleles. It was carried out an

overall estimation of d_N/d_S of MHC class I Exon 3 and the other was codons comprising only PBR and non-PBR, which was calculated with MEGA 7.0 according to the Nei-Gojobori method [45] with the Jukes and Cantor correction. Standard error estimates were derived from 1000 bootstrap replicates. Z test of historical positive selection [46] was calculated in MEGA 7.0. Third, the Maximum likelihood implemented in codeml in PAML 4.9 was used for identification of sites involved in the positive selection, which are indicated where the ratio ω (d_N/d_S) larger than 1 [47]. Two different models corresponding ω were tested: M7 (beta), M8 (β and ω). To find whether the alternative model (M8) provided better fitter than the M7, we performed Likelihood ratio tests to compare twice the difference of the log-likelihood ratios ($2\Delta\ln L$) using a distribution χ^2 . PSSs in the M8 model was identified by PP more than 95% using the Bayes empirical Bayes procedure. Positively selected sites were verified at each codon site separately using many complementary approaches implemented in Datamonkey (<http://www.datamonkey.org/>) [48] in addition to afore mention methods. Specifically, we used MEME [49], FEL, SALC [50], and FUBER [51].

2.8. Phylogenetic Analysis. To assess the phylogenetic relationship, we construct two phylogenies (One for sampled species and other representing MHC class I sequences of related passerines plus sampled species) using Bayesian inference. We find the GTR+T nucleotide substitution model [52] that fits our data using MrModeltest [35] through the Akaike Information Criterion (AICc) [53]. Bayesian Markov chain Monte Carlo (MCMC) was run for two million generations and sampling every 1,000 generations to ascertain when log Likelihood reached stationary phase. The phylogenetic tree was summarized in MrBayes v3.1.2 [54] and the first 25% of the tree as burn-in was removed. Fig tree was used for visualization of the consensus tree. Exploration of relation between sampled species and related avian species, we conducted a maximum likelihood (ML) analysis with MEGA 7.0 [55]. The data were analyzed with the T92+G model. We conducted 1000 bootstrap replicates to estimate the support. Values greater than 75% were indicated in the ML phylogenetic trees. The species covered are mainly from *Passeridae*, *Acrocephalidae*, *Paridae*, *Motacillidae*, *Muscicapidae*, *Hirundinidae*, *Phylloscopidae*, *Fringillidae*, *Cardinalidae*, and *Sturnidae*. To further identify allelic lineages among sampled species and related avian species, we conducted the Neighbor-Net algorithm in SplitsTree 4.14.8. Neighbor-Net networks were based on uncorrected *P*-distances and carried out 1000 bootstrap replicates to estimate nodal support. Nodal support values (>75%) were displayed.

3. Results

3.1. Characterization of Alleles. We successfully and selectively amplified MHC class I exon 3 genes across 80 individuals from four species of the *Turdidae* family using HN36 and HN46 primers. An average of 22.7 clones per individual was sequenced. Sequences varied between 459 and 579 base pairs. The multiple sequence alignments of all sampled species were 411 base pair long. The final aligned MHC class I

TABLE 1: Amplification success and genetic diversity within each of the four species of the *Turdidae* family investigated. MHC class I exon 3 gene size (*L*), the overall number of polymorphic sites per allele repertoire (*S*), and the average number of nucleotide differences (*K*). Nucleotide diversity π at all sites: PBR and non-PBR.

Species	<i>L</i>	<i>S</i>	<i>K</i>	π
<i>Turdus naumanni</i>	285	64	33.7	0.118
				0.211
				0.091
				0.113
<i>Turdus eunomus</i>	285	58	32.32	0.183
				0.079
<i>Turdus ruficollis</i>	291	65	35.28	0.121
				0.247
				0.087
<i>Turdus atrogularis</i>	291	70	43.95	0.151
				0.309
				0.093

dataset included 285-291 bp (Primers not include). Analysis of gDNA alignment revealed a total of 77 distinct Haplotypes/alleles including 47 PFA. Each sequence was confirmed to exhibit similarity (81%-93%) with earlier reported passerine MHC class I sequences based upon BLAST search. The numbers of PFA sequences found in a single individual ranged from one to five, indicating that one to three loci exist in three of the four species of the *Turdidae* family. However, the number of putative functional alleles found in a single individual ranged from two to seven in *Turdus atrogularis* exhibiting two to four loci. Number of the individual tested, number of PFA and pseudogene retrieved, the minimum number of functional loci estimated is given in Table 1. Three alleles (*Tuna-MHCI * PFA05 = Tuen-MHCI * PFA09*, *Tuna-MHCI * PFA07 = Tuen-MHCI * PFA02* and *Tuen-MHCI * PFA05 = Tuna-MHCI * PFA015*) were shared among *Turdus naumanni* and *Turdus eunomus*. Two alleles (*Turu-MHCI * PFA05 = Tuat-MHCI * PFA02* and *Turu-MHCI * PFA09 = Tuat-MHCI * PFA08*) were also detected among individuals of *Turdus ruficollis* and *Turdus atrogularis*. Interestingly, genotypes comprising of one allele were by far the most repeated (26.67%, 8/30), followed by genotypes comprising two (16.67%, 5/30) and four alleles (13.3%, 4/30) in the population of *Turdus naumanni*. Almost pattern was consistent in population of *Turdus eunomus* and *Turdus rufficollis*. Genotypes constituting one allele (23.3%, 7/30) were the most repeated followed by three (16.67%, 5/30) in *Turdus eunomus*. Genotypes comprising one allele (33.33%, 5/15) were repeated in the population of *Turdus rufficollis*. Allelic repetition was absent in population of *Turdus atrogularis*.

Of the 77 sequences, 30 were non-translatable due to indels or the presence of stop codons resulted changes in the reading frame. Sequences were thus presumed to be pseudogenes. The number of identified pseudogenes within the four species ranged between three and five in most individuals of study population, and six of the thirteen pseudogene

TABLE 2: Recombinants detected in *Turdidae* family MHC class I alleles, parent sequences and breakpoints detected by the recombination detection program (RDP) and the genetic algorithm for recombination detection (GARD), and the RDP analyses.

	Recombination event 1	Recombination event 2
Recombinant	<i>Tuna_MHCI * PFA06</i>	<i>Tuna_MHCI * PFA02</i>
Maj P	<i>Tuna_MHCI * PFA02</i>	<i>Tuna_MHCI * PFA09</i>
Min P	<i>Tuna_MHCI * PFA011</i>	Unknown (<i>Tuna_MHCI * PFA01</i>)
BP 1 location	148 (148)	254 (253); $P < 0.001$
BP 2 location	Absent	Absent
RDP methods		
RDP	NS	<0.05
GENECONV	<0.001	<0.001
BootScan	<0.01	<0.05
MaxChi	<0.001	<0.01
Chimaera	<0.01	<0.01
SiScan	<0.001	<0.001
3Seq	<0.001	<0.01
BP from GARD	Absent	

Note: NS indicates not significant. Maj P and Min P represent major and minor parents, respectively. BP denotes breakpoint. The numbers in parentheses are BP locations in the recombinant nucleotide sequences without gaps. The values after the semicolon are Max chi values for those BPs. ** indicates $P < 0.01$.

sequences were found to be identical in three individuals from the population of sampled species. We cannot ignore the likelihood that some of the identified pseudogene sequences may be due to PCR or sequencing artifacts, as such events would more often result in nonfunctional sequences. The nucleotide deletion result in loss of 3 amino acids was obvious in *Tuna-MHCI * PS07-9* and *Tueu-MHCI * PS01-04* and *Tueu-MHCI * PS08*. Both nucleotide deletion, frame shift mutation and premature stop codons were detected in *Turu-MHCI * PS01,03* and *MHCI * PS09* at amino acid 33 encoding Exon 3. Loss of 3 amino acids was at position 78 was detected in *Tuat-MHCI * PS05* and *Tuat-MHCI*PS06*.

3.2. Analysis of Genetic Diversity. Overall we find an elevated genetic diversity (π) within exon 3 alleles repertoire among individuals of *Turdus atrogularis* was (0.151) than *Turdus eunomus* (0.113). The average number of nucleotides difference (K) varied between 43.95 in *Turdus atrogularis* and 32.32 in *Turdus eunomus*.

3.3. Analysis of Recombination. The recombination detection program not only analyzes brake points but also identify parent sequences. We ran the test of recombination by pooling all putative functional alleles recovered from four species of the *Turdidae* family. We only find one potential recombination event in *Tuna-MHCI * PFA06* in *Turdus naumanni* at two recombinant breakpoints at position 148 and 253. *Tuna-MHCI * PFA02* as major and *Tuna-MHCI * PFA011* minor parent. Likewise, a single recombination was significant in *Tueu-MHCI * PFA07*. We detected no recombination among other alleles. However, these recombinations were only significant in two out of seven tests and not consistent with recombination breakpoint identified by GARD, hence

the results represent that overall recombination is not likely to have any prominent effects on tests for positively selected sites (Table 2). The recombination breakpoints identified by these two programs are often inconsistent, probably because they use different computational methods.

3.4. Analysis of Selection. Considering that the evolutionary history of each domain might have been different, we tested each domain separately for evidence of positive selection. Selection statistics by traditional methods did not disclose any statistical significant signal of selection that deviate from neutral expectations for *Turdus eunomus* (Tajima's D : -0.87309, $P > 0.10$; Fu and Li's D^* test statistic: 0.36, $P > 0.10$; Fu and Li's F^* test statistic: 0.03, $P > 0.10$) and *Turdus atrogularis* (Tajima's D : -0.86107, $P > 0.10$ Fu and Li's D^* test statistic: 0.19, $P > 0.10$; Fu and Li's F^* test statistic: -0.077, $P > 0.10$). Still, overall d_N value was significantly higher statistically than d_S in *Turdus atrogularis* (1.687) and ratio d_N/d_S was more pronounced at codons presumably coding PBR (1.994) than codons not involved in such activity (0.884) is presented in (Table 3).

Application of Likelihood models represents that the model M8 allows for positive selection provides a better than the neutral evolution models M7. Sites being positively selected were recognized, are given in (Table 4). In total, we find 12 codons under positive selection in sampled species, of which three sites (25%) match homologues codons found positively selected in other avian species and one (8.3%) matched human peptide binding region (Table 4).

Usually consistent with the above finding, every substitute test (MEME, SALC, FEL, and FUBAR) for positive selection implemented in online adoptive evolutionary server Datamonkey (Weaver et al., 2018) identify numerous codons under positive selection (Figure 2) and (Figure 3).

TABLE 3: The average rates of nonsynonymous (d_N) and synonymous (d_S) substitutions and the result of Z-test and the average nucleotide distances (d_{nt}) and amino acid distances (d_{aa}) for PBR and non-PBR and all sites in MHC class I of the *Turdidae* family.

Species	Domain	$d_N \pm SE$	$d_S \pm SE$	Z	P	ω	$d_{nt} \pm SE$	$d_{aa} \pm SE$
<i>Turdus naumanni</i>	All sites	0.142 ± 0.026	0.104 ± 0.013	0.610	0.543	1.365	0.051 ± 0.08	0.246 ± 0.035
	PBR	0.281 ± 0.059	0.153 ± 0.032	1.848	0.034	1.835	0.041 ± 0.045	0.372 ± 0.082
	Non-PBR	0.051 ± 0.062	0.067 ± 0.012	1.356	0.476	0.761	0.034 ± 0.09	0.126 ± 0.029
<i>Turdus eunomus</i>	All sites	0.134 ± 0.021	0.107 ± 0.019	0.729	0.457	1.251	0.043 ± 0.010	0.202 ± 0.031
	PBR	0.179 ± 0.023	0.102 ± 0.032	1.442	0.383	1.175	0.057 ± 0.031	0.366 ± 0.011
	Non-PBR	0.056 ± 0.041	0.049 ± 0.012	1.792	1.000	1.142	0.051 ± 0.011	0.134 ± 0.029
<i>Turdus ruficollis</i>	All sites	0.146 ± 0.048	0.101 ± 0.025	0.911	0.771	1.445	0.059 ± 0.017	0.191 ± 0.035
	PBR	0.264 ± 0.037	0.138 ± 0.011	1.643	0.002	1.912	0.147 ± 0.045	0.453 ± 0.078
	Non-PBR	0.070 ± 0.018	0.068 ± 0.028	0.061	0.476	1.029	0.034 ± 0.049	0.126 ± 0.052
<i>Turdus atrogularis</i>	All sites	0.189 ± 0.091	0.112 ± 0.025	1.040	0.150	1.687	0.063 ± 0.053	0.211 ± 0.039
	PBR	0.321 ± 0.013	0.161 ± 0.069	1.012	0.435	1.994	0.207 ± 0.045	0.572 ± 0.162
	Non-PBR	0.069 ± 0.062	0.078 ± 0.012	1.813	0.476	0.884	0.078 ± 0.091	0.206 ± 0.041

*The errors were attained through 1000 bootstrap replicates which are in parentheses. Bold represents significant results.

TABLE 4: Estimation of d_N and d_S substitution rates for sites positively selected and their ratio for codons chosen a priori (PBR and non-PBR).

Species	Comparison	Model	lnL value	Parameter estimates	PSSs	LRT	TS value
<i>Turdus naumanni</i>	Tuna 1-30	M7 (beta)	-703.53	$P = 0.19, q = 0.138, P_0 = 0.351$	Not allowed	M7 vs. M8	5.47
		M8 (beta and omega)	-722.49	$P = 0.87, q = 0.117, P_1 = 0.09, \omega = 3.11$	39F,41L 88T		
<i>Turdus eunomus</i>	Tuna 1-30	M7 (beta)	-692.13	$P = 0.13, q = 0.97, P_0 = 0.347$	Not allowed	M7 vs. M8	4.56
		M8 (beta and omega)	-714.11	$P = 0.91, q = 0.158, P_1 = 0.11, \omega = 3.76$	41L, 52P,88T		
<i>Turdus ruficollis</i>	Turu1-15	M7 (beta)	-811.27	$P = 0.49, q = 0.187, P_0 = 0.411$	Not allowed	M7 vs. M8	6.21
		M8 (beta and omega)	-834.88	$P = 0.131, q = 0.232, P_1 = 0.13, \omega = 3.94$	29Y, 78F , 88G		
<i>Turdus atrogularis</i>	Tuat1-5	M7 (beta)	-847.53	$P = 0.83, q = 0.211, P_0 = 0.585$	Not allowed	M7 vs. M8	7.10
		M8 (beta and omega)	-849.78	$P = 0.173, q = 0.279, P_1 = 0.17, \omega = 4.11$	39H, 78F, 91Q		

*The log likelihood values and parameters estimated were computed using codeML implemented in PAML 4.9. PSSs were inferred in model M8 by BEB with posterior probabilities > 95%.

Across all tests for positive selection, four codons (9, 29, 65, and 88) were frequently identified by all methods as having under positive selection. Of these, codons (42, 59) were corresponding to PBR in human and codons 9, 29, 64, and 88 also match homology to PBR, known as positively selected among passerine in general [56] (Figure 4). The ten most frequent MHC class I alleles retrieved from sampled species displayed 87%-91% sequence similarity to 18 sequences from five other passerine families (*Acrocephalidae*, *Passeridae*, *Muscicapidae*, *Paridae*, *Passerellidae*). None of the 77 alleles studied had 100% sequence similarity to other published sequences to GenBank; thus, it establishes no allelic pair in

the study population that was 100% sequence likeness shared by another species.

3.5. Phylogenetic Analysis. In phylogenetic analysis, we observed that sampled species form a well-supported monophyletic clade with *Erithacus rubecula* members of the *Turdidae* family in maximum likelihood analysis. Bayesian analysis represents that most of the alleles shared among *Turdus atrogularis* and *Turdus ruficollis*. This pattern was almost consistent among *Turdus naumanni* and *Turdus eunomus* presented in Figure 4. The Net network of putative functional and pseudogene MHC class I exon 3 sequences in

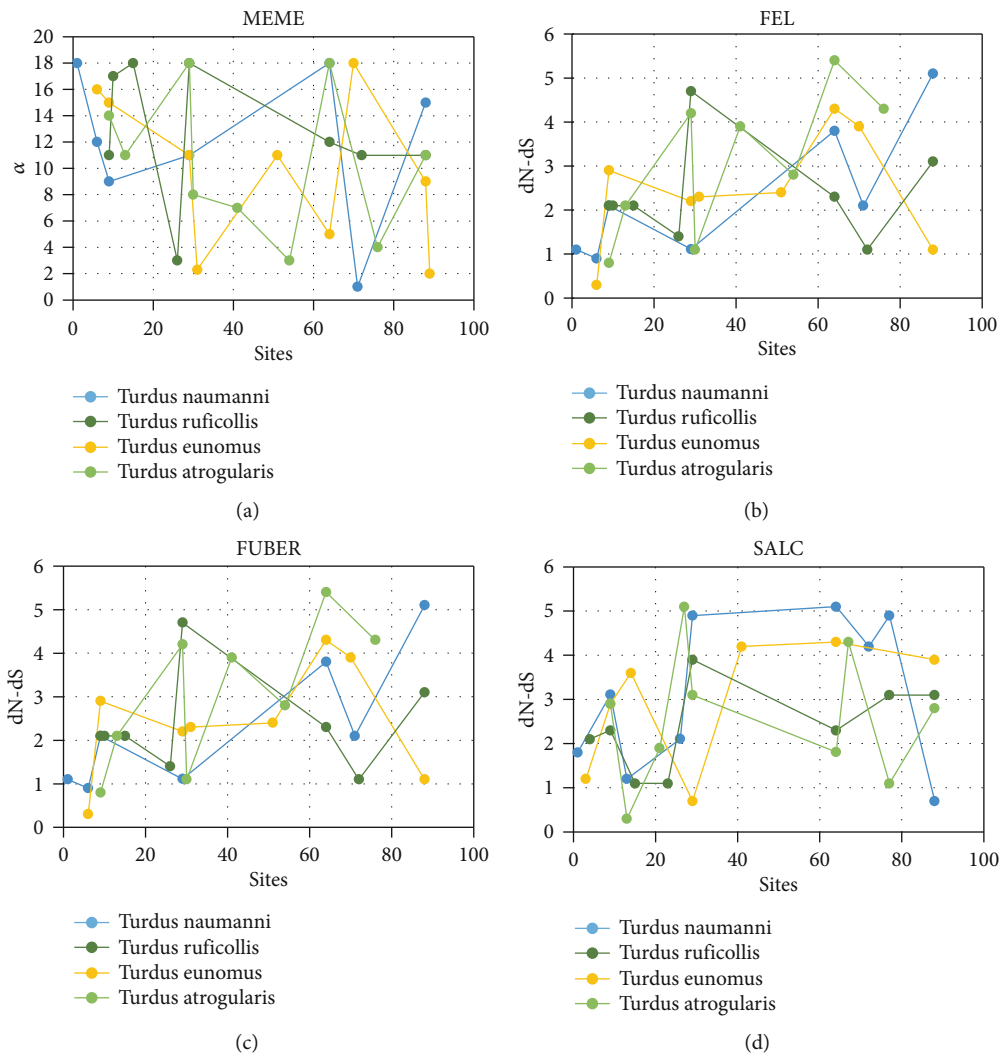


FIGURE 2: Positively selected sites using the online adoptive evolutionary server Datamonkey with (a) MEME, (b) FEL, (c) FUBAR, and (d) SALC. Substitution tests identify numerous codons showing the signature of positive selection.

the *Turdidae* family with other passerines indicate that allelic distribution among them is almost congruent with limited divergence. For instance, *Tueu-MHCI * PFA02* and *Tuna-MHCI * PFA07* networks formed a monophyletic clade in the phylogenetic network of exon 3. Three alleles were shared among *Turdus naumanni* and *Turdus eunomus* two among *Turdus ruficollis* and *Turdus atrogularis*. The clustering of the sequences among species could be due to transspecies polymorphism or orthology [57].

4. Discussion

In this study, we have for the first time characterize MHC Class I gene in four species of the *Turdidae* family in the order Passeriformes from the wide geographical area of Northeast china. Analysis of MHC class I sequences revealed a total of 77 distinct Haplotypes/alleles including 47 putative functional alleles ever reported in passerine species, a group which is reported to have surprising MHC diversity [58, 59]. According to our findings based on MHC class I sequences, the functional loci in an individual ranged from

one to three in three of the four species, which was consistent with findings from other passerine species studied till now [60]. In addition, we detected a large number of presumed pseudogene sequences in the sampled population as it retains important information about the evolution of MHC. This is not surprising, as it is consistent with the expectation of evolution by birth-and-death [61]. We made a significant effort to characterize the variation in regions of MHC class I exon 3 in our study population, we find that the primers would make some unlikely bias in allelic variations among individuals. Hence, MHC class I alleles variations per individual should, largely be due to copy number of genes variation among individuals, which has been confirmed in other birds [62]. Few MHC class I alleles were shared between *Turdus naumanni* and *Turdus eunomus* as well as among individuals of *Turdus ruficollis* and *Turdus atrogularis* is indicating allelic sharing due to common ancestors or challenging common pathogens, as this event is frequent in numerous avian species such as owls, ardeid birds, penguins and passerines [63–65].

Generally, abundant variation in genetic material in a species is an indicator of the capacity to adapt to numerous

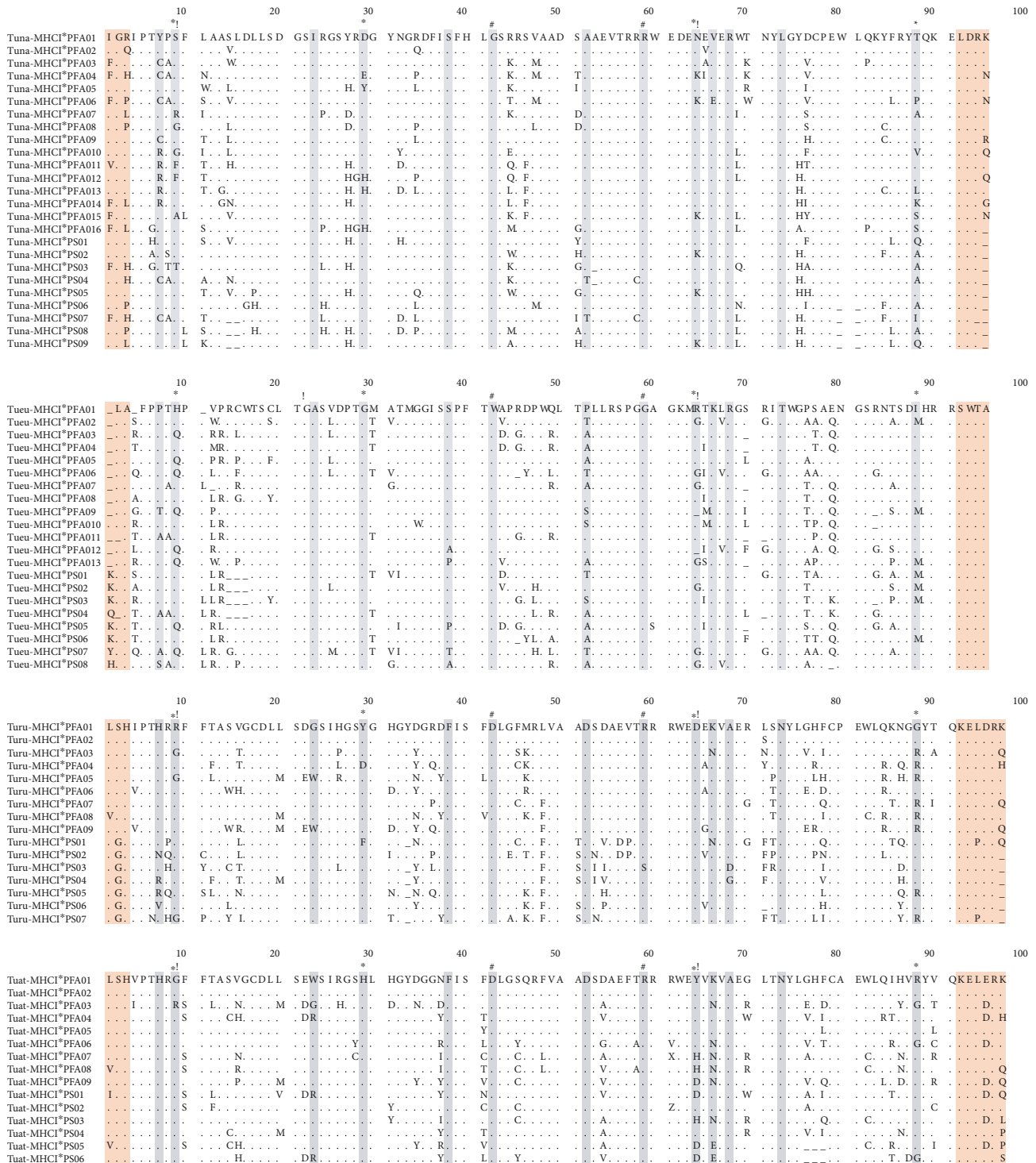


FIGURE 3: Alignment of deduced amino acid sequences of all alleles retrieved from four species of the *Turdidae* family. Dots (.) indicate identity with the reference sequence. Light gray color represents codons presumably coding for peptide binding regions upon alignment with human and other avian species. The light brown region indicates flanking introns. - indicates missing nucleotides. * represents codons positively selected in almost all of the tests performed for selection analysis. ! indicates the homologue region also positively selected in most of avian species. # inferred homology with human.

environmental changes by that species. Rapidly evolving environmental pathogens would cause MHC genes to exhibit enlarged genetic diversity in species [66, 67]. Collectively, in our study, we find elevated genetic diversity among func-

tional sequences and significant divergence, whereas pseudogene has low genetic variation and limited divergence. Similar results also have been described in other passerine species, including common yellow throat [68], great reed

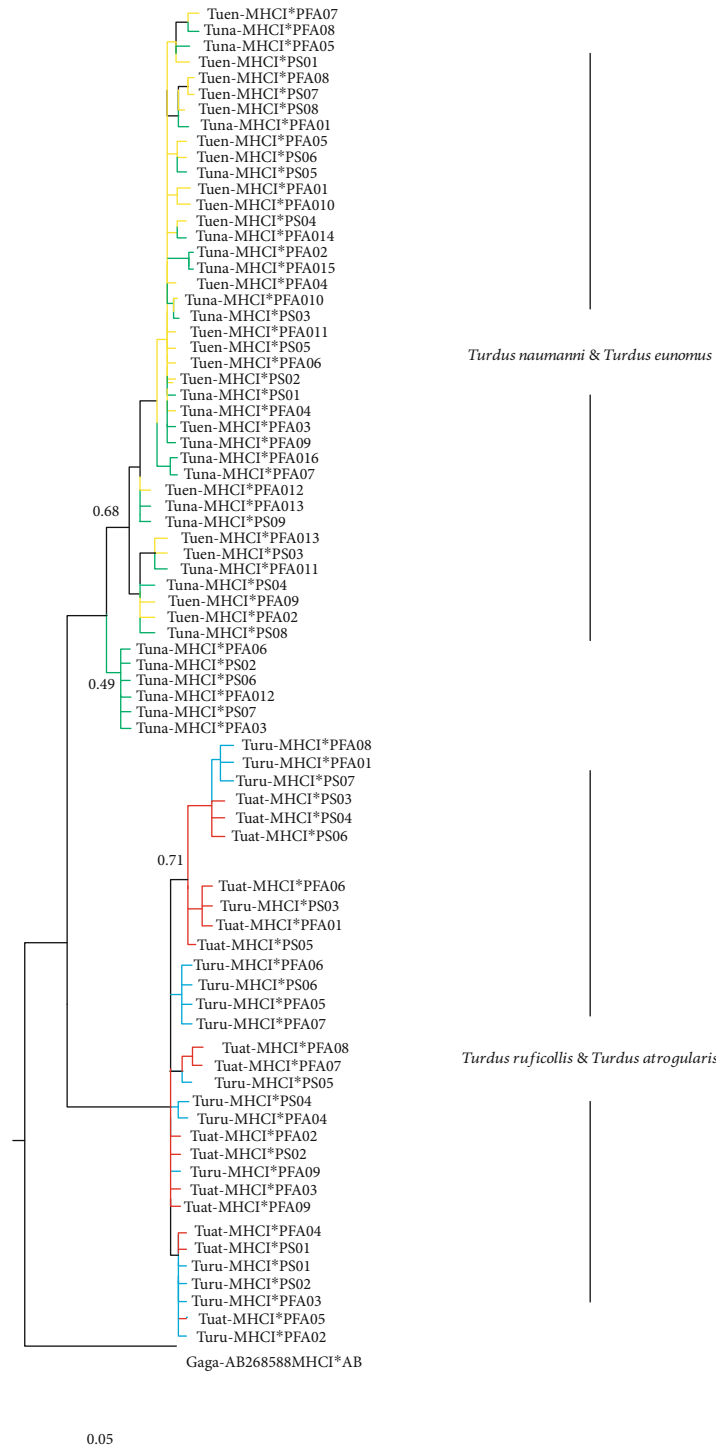


FIGURE 4: Bayesian phylogenetic reconstruction of MHC class I exon 3 of four species of the *Turdidae* family. All the nodes are well supported (PP > 0.90%) unless indicated otherwise. AB268885 *Gallus gallus* was used as an outgroup.

warbler and the great tit [69]. The allelic variation described in our study could be due to increased immunological defense against the internal pathogen since these are highly unlikely to adapt to novel, infrequent variant [15].

Recombination has been considered an important mechanism that influences allelic diversity and driving evolution of the MHC gene [70, 71] We only find one potential recom-

bination event in *Tuna-MHCI * PFA06* at two recombinant breakpoints at position 148 and 253 identified with recombination detection program. Similarly, single recombination was significant in *Tueu-MHCI * PFA07*. Recombination pattern was also restricted two out of seven tests; hence our finding indicate recombination is unlikely to have any significant influence on tests for PSs. Though we could not find any

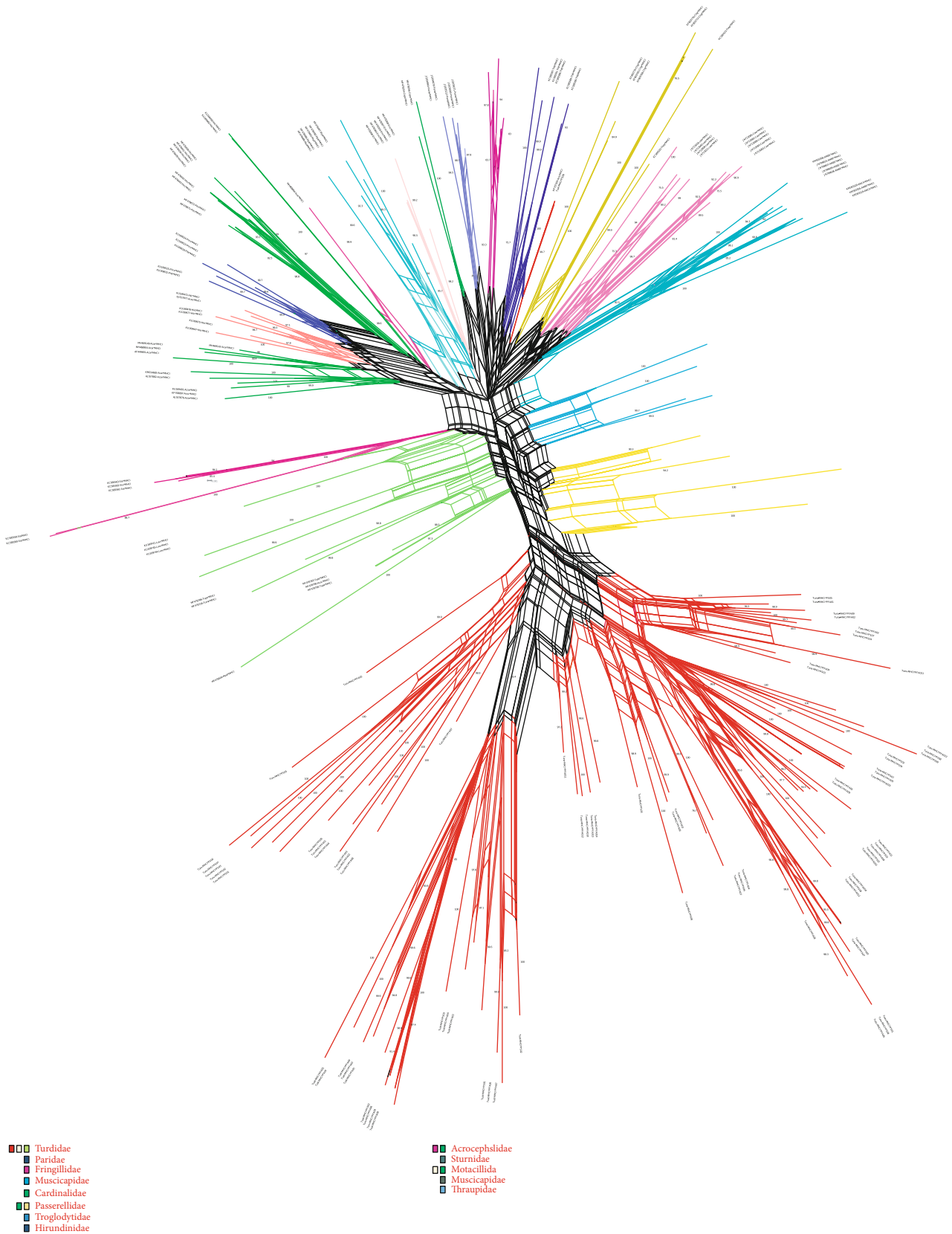


FIGURE 5: The phylogenetic networks of MHC class I of four species of the *Turdidae* family along with homologue sequences from passerine species. GenBank accession numbers are provided. Species names are mentioned in the lower right side with branch colors. Neighbor-Net networks based on uncorrected *P*-distances and carried out in 1000 bootstrap replicates to estimate nodal support. Nodal support values (>75%) were displayed.

substantial recombination among other alleles, qualitatively our result suggests a role for recombination during the evolution of MHC class I in our species studied. Our finding is consistent with, that micro-recombination is frequently observed in MHC genes [57]. Further study of recombinant function in the future will contribute to a detailed understanding of its role in the evolution of the MHC gene.

Positive selection is the maintainer of alleles having the advantageous mutation that maintain fitness of an individual. In our study, the classical test of selection Tajima's D , Fu and Li's D^* and Fu and Li's F^* showed no deviation from neutral selection or balance selection. Considering the level of variation, conventional methods used to find selection are not influential [72]. As sites positively selected are likely to accumulate more non-synonymous than synonymous substitutions, influencing amino acid variation to result in functional modifications in proteins [73]. Our study revealed differential expression of selection pattern in functional sequences on regions related with PBR and non-PBR of the MHC class I gene. Codons involved in peptide binding region revealed more non-synonymous substitution than synonymous ($d_N/d_S = 1.99$) in *Turdus atrogularis* as compared to non-peptide binding region ($d_N/d_S = 0.884$), pattern was consistent among all species tested, which might be enlightened that stronger selection pressure from intracellular pathogens than extracellular pathogens [74]. Evidence of positive selection at PBR of MHC has been reported in the house sparrow (PBR $d_N/d_S = 1.55$ vs. non-PBR $d_N/d_S = 0.51$) [75] and golden pheasant (PBR $d_N/d_S = 1.45$ vs. non-PBR $d_N/d_S = 0.91$) [76]. Of the 12 codons in total among species tested exhibit positive selection with Likelihood methods using PAML, 9, 29, 64, and 88 match homologues codons found positively selected in other passerine species.

It should be noted that the pooling of all alleles across loci will mostly reduce selection detection tests, so the outcomes might be conservative, but will be less prone to false positives [77, 78]. Therefore, attention should be given while inferring about the detected diversity in MHC and the possible effects of selection on individual loci. Our results suggested that $\alpha 2$ domain of MHC class I exon 3 of all species are under positive selection pressure. Pronounced positive selection at antigen-binding sites permits a species or population to present a larger repertoire of peptides (antigens), thus increase the defensive ability against parasitic and pathogenic infections.

Finally, phylogenetic clustering of MHC class I data set of sampled species when pooled with other passerine species produces a contrasting pattern. In general, the MHC class I sequence of the *Turdidae* family clustered together with sequences from congeneric species. We found increased sequences similarities between the same species rather than within species (trans specific likenesses), is usually described with trans species polymorphism (TSP), which occurs due to alleles passage from ancestral to the decedent via partial arrangement of lineages [79]. Although trans specific similarities can be described with convergent evolution due to the results of similar environmental selective pressure. Studies indicate that TSP is a primary mechanism responsible for clustering of alleles at avian MHC class I [80] (Figure 5).

5. Conclusion

Our study shows that species of the *Turdidae* family has retained significant MHC class I diversity, which supports high conservational value and contributes to the evolution of MHC class I genes. Importantly, we specifically amplify the exon 3 locus and provide an opportunity to avoid chimera formation during molecular characterization of hyper-variable genes of immunity. At the same time, our study is the first to validate contrasting patterns of allelic diversity and positive selection upon inferred PBR and non-PBR codons which supported the hypothesis that different mechanisms can shape evolutionary paths of MHC class I.

Abbreviations

MHC:	Major histocompatibility complex
PBR:	Peptide-binding region
CDs:	Cluster of differentiation
CDRs:	Complementary-determining regions
ABS:	Antigen-binding site
TCR:	T-cell receptors
SDS:	Sodium dodecyl sulfate
EDTA:	Ethylene diamine tetra acetic acid
PCR:	Polymerase chain reaction
HLA:	Human leukocyte antigen
MEGA:	Molecular Evolutionary Genetics Analysis
GTR:	General time-reversible model
PFA:	Putative function alleles
GARD:	Genetic algorithm for recombination detection
RDP:	Recombination detection program
PP:	Posterior probability
BP:	Breakpoint
d_N :	Nonsynonymous substitution
d_S :	Synonymous substitution
d_{nt} :	Nucleotide distance
d_{aa} :	Amino acid distance
SE:	Standard error
PSSs:	Positively selected sites
BEB:	Bayes empirical Bayes
MEME:	Mixed effects model of evolution
SALC:	Single likelihood ancestor counting FEL, fixed effect likelihood
FUBAR:	Fast unconstrained Bayesian approximation
PAML:	Phylogenetic analysis using maximum likelihood
TSP:	Transspecies polymorphism.

Data Availability

The data of this study will be available openly to readers, and they can access the data supporting the conclusions of the study.

Disclosure

The manuscript has been presented in "pre-print" at <https://www.researchgate.net/publication/346148804>.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

MUG, AY, and LB designed the study. MUG carried out the experiment and drafted the experiment. LB supervised the whole study, provided recommendations for, and revised the MS.YCX provided valuable suggestion for the MS. All authors contributed to and approved the current manuscript draft.

Acknowledgments

This study was supported by the Fundamental Research Funds for Central Universities (grant no. 2572018BE04). Authors are indebted to Kang Hui and Wang Dong for their technical assistance in experiments and willingly providing guidance during silico data analysis. The authors also thank Jacob Njaramba Ngatia and Dr. Mehboob Ahmad for their valued suggestions.

References

- [1] K. Murphy, P. Travers, and M. Walport, *Janeway's Immunobiology*, Garland science, New York, 2008.
- [2] K. K. Jensen, M. Andreatta, P. Marcatili et al., "Improved methods for predicting peptide binding affinity to MHC class II molecules," *Immunology*, vol. 154, no. 3, pp. 394–406, 2018.
- [3] P. J. Bjorkman and P. Parham, "Structure, function, and diversity of class I major histocompatibility complex molecules," *Annual Review of Biochemistry*, vol. 59, no. 1, pp. 253–288, 1990.
- [4] X.-H. Zhang, Z. X. Dai, G. H. Zhang, J. B. Han, and Y. T. Zheng, "Molecular characterization, balancing selection, and genomic organization of the tree shrew (*Tupaia belangeri*) MHC class I gene," *Gene*, vol. 522, no. 2, pp. 147–155, 2013.
- [5] M. Yeager and A. L. Hughes, "Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution," *Immunological Reviews*, vol. 167, no. 1, pp. 45–58, 1999.
- [6] S. Piertney and M. Oliver, "The evolutionary ecology of the major histocompatibility complex," *Heredity*, vol. 96, no. 1, pp. 7–21, 2006.
- [7] J. K. Kulski, T. Shiina, T. Anzai, S. Kohara, and H. Inoko, "Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man," *Immunological Reviews*, vol. 190, no. 1, pp. 95–122, 2002.
- [8] J. Kelley, L. Walter, and J. Trowsdale, "Comparative genomics of major histocompatibility complexes," *Immunogenetics*, vol. 56, no. 10, pp. 683–695, 2005.
- [9] Å. Langefors, J. Lohm, M. Grahn, Ø. Andersen, and T. . Schantz, "Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 268, no. 1466, pp. 479–485, 2001.
- [10] P. W. Hedrick, "Pathogen resistance and genetic variation at MHC loci," *Evolution*, vol. 56, no. 10, pp. 1902–1908, 2002.
- [11] H. Westerdahl, "Passerine MHC: genetic variation and disease resistance in the wild," *Journal of Ornithology*, vol. 148, no. S2, pp. 469–477, 2007.
- [12] S. Paterson and J. M. Pemberton, "No evidence for major histocompatibility complex-dependent mating patterns in a free-living ruminant population," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 264, no. 1389, pp. 1813–1819, 1997.
- [13] C. Landry, D. Garant, P. Duchesne, and L. Bernatchez, "Good genes as heterozygosity: the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*)," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 268, no. 1473, pp. 1279–1285, 2001.
- [14] G. J. Knafler, J. A. Clark, P. D. Boersma, and J. L. Bouzat, "MHC diversity and mate choice in the Magellanic penguin, *Spheniscus magellanicus*," *Journal of Heredity*, vol. 103, no. 6, pp. 759–768, 2012.
- [15] J. A. Borghans, J. B. Beltman, and R. J. De Boer, "MHC polymorphism under host-pathogen coevolution," *Immunogenetics*, vol. 55, no. 11, pp. 732–739, 2004.
- [16] A. Jepson, W. Banya, F. Sisay-Joof et al., "Quantification of the relative contribution of major histocompatibility complex (MHC) and non-MHC genes to human immune responses to foreign antigens," *Infection and Immunity*, vol. 65, no. 3, pp. 872–876, 1997.
- [17] S. V. Edwards, J. Gasper, D. Garrigan, D. Martindale, and B. F. Koop, "A 39-kb sequence around a blackbird Mhc class II gene: ghost of selection past and songbird genome architecture," *Molecular Biology and Evolution*, vol. 17, no. 9, pp. 1384–1395, 2000.
- [18] D. Canal, M. Alcaide, J. A. Anmarkrud, and J. Potti, "Towards the simplification of MHC typing protocols: targeting classical MHC class II genes in a passerine, the pied flycatcher *Ficedula hypoleuca*," *BMC Research Notes*, vol. 3, no. 1, p. 236, 2010.
- [19] G. Kroemer, A. Bernot, G. Behar et al., "Molecular genetics of the chicken MHC: current status and evolutionary aspects," *Immunological Reviews*, vol. 113, no. 1, pp. 119–145, 1990.
- [20] R. Zoorob, A. Bernot, D. M. Renoir, F. Choukri, and C. Auffray, "Chicken major histocompatibility complex class II B genes: analysis of interallelic and inter-locus sequence variance," *European Journal of Immunology*, vol. 23, no. 5, pp. 1139–1145, 1993.
- [21] M. M. Peacock and A. T. Smith, "The effect of habitat fragmentation on dispersal patterns, mating behavior, and genetic variation in a pika (*Ochotona princeps*) metapopulation," *Oecologia*, vol. 112, no. 4, pp. 524–533, 1997.
- [22] C. Bonneaud, G. Sorci, V. Morin, H. Westerdahl, R. Zoorob, and H. Wittzell, "Diversity of Mhc class I and IIB genes in house sparrows (*Passer domesticus*)," *Immunogenetics*, vol. 55, no. 12, pp. 855–865, 2004.
- [23] H. Westerdahl, "No evidence of an MHC-based female mating preference in great reed warblers," *Molecular Ecology*, vol. 13, no. 8, pp. 2465–2470, 2004.
- [24] M. Promerová, T. Albrecht, and J. Bryja, "Extremely high MHC class I variation in a population of a long-distance migrant, the scarlet rosefinch (*Carpodacus erythrinus*)," *Immunogenetics*, vol. 61, no. 6, pp. 451–461, 2009.
- [25] W. BABIK, P. TABERLET, M. J. EJSMOND, and J. RADWAN, "New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system," *Molecular Ecology Resources*, vol. 9, no. 3, pp. 713–719, 2009.

- [26] T. Kanagawa, "Bias and artifacts in multitemplate polymerase chain reactions (PCR)," *Journal of Bioscience and Bioengineering*, vol. 96, no. 4, pp. 317–323, 2003.
- [27] S. Abduriyim, Y. Nishita, P. A. Kosintsev et al., "Evolution of MHC class I genes in Eurasian badgers, genus *Meles* (Carnivora, Mustelidae)," *Heredity*, vol. 122, no. 2, pp. 205–218, 2019.
- [28] T. A. Hall, *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*, Nucleic acids symposium series, Information Retrieval Ltd., London, 1999.
- [29] M. A. Larkin, G. Blackshields, N. P. Brown et al., "Clustal W and Clustal X version 2.0," *Bioinformatics*, vol. 23, no. 21, pp. 2947–2948, 2007.
- [30] J. Klein, R. E. Bontrop, R. L. Dawkins et al., *Nomenclature for the major histocompatibility complexes of different species: a proposal*, The HLA system in clinical transplantation, Springer, 1993.
- [31] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.
- [32] P. Librado and J. Rozas, "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data," *Bioinformatics*, vol. 25, no. 11, pp. 1451–1452, 2009.
- [33] D. Martin, D. Posada, K. A. Crandall, and C. Williamson, "A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints," *AIDS Research & Human Retroviruses*, vol. 21, no. 1, pp. 98–102, 2005.
- [34] M. Padidam, S. Sawyer, and C. M. Fauquet, "Possible emergence of new geminiviruses by frequent recombination," *Virology*, vol. 265, no. 2, pp. 218–225, 1999.
- [35] D. Posada, "jModelTest: phylogenetic model averaging," *Molecular Biology and Evolution*, vol. 25, no. 7, pp. 1253–1256, 2008.
- [36] J. M. Smith, "Analyzing the mosaic structure of genes," *Journal of Molecular Evolution*, vol. 34, no. 2, pp. 126–129, 1992.
- [37] M. J. Gibbs, J. S. Armstrong, and A. J. Gibbs, "Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences," *Bioinformatics*, vol. 16, no. 7, pp. 573–582, 2000.
- [38] M. F. Boni, D. Posada, and M. W. Feldman, "An exact non-parametric method for inferring mosaic structure in sequence triplets," *Genetics*, vol. 176, no. 2, pp. 1035–1047, 2007.
- [39] S. L. Kosakovsky Pond, D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. W. Frost, "GARD: a genetic algorithm for recombination detection," *Bioinformatics*, vol. 22, no. 24, pp. 3096–3098, 2006.
- [40] C. N. Balakrishnan, R. Ekblom, M. Völker et al., "Gene duplication and fragmentation in the zebra finch major histocompatibility complex," *BMC Biology*, vol. 8, no. 1, p. 29, 2010.
- [41] M. Alcaide, J. Muñoz, J. Martínez-de la Puente, R. Soriguer, and J. Figuerola, "Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian coot *Fulica atra* (Aves: Rallidae)," *Ecology and Evolution*, vol. 4, no. 6, pp. 688–698, 2014.
- [42] J. Kaufman, J. Salomonsen, and M. Flajnik, "Evolutionary conservation of MHC class I and class II molecules—different yet the same," in *Seminars in immunology*, Elsevier, 1994.
- [43] C. S. Hee, S. Gao, B. Loll et al., "Structure of a classical MHC class I molecule that binds "non-classical" ligands," *PLoS Biology*, vol. 8, no. 12, article e1000557, 2010.
- [44] P. Bjorkman, M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger, and D. C. Wiley, "The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens," *Nature*, vol. 329, no. 6139, pp. 512–518, 1987.
- [45] M. Nei and T. Gojobori, "Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions," *Molecular Biology and Evolution*, vol. 3, no. 5, pp. 418–426, 1986.
- [46] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods," *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731–2739, 2011.
- [47] Z. Yang, "PAML 4: phylogenetic analysis by maximum likelihood," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1586–1591, 2007.
- [48] S. L. K. Pond and S. D. Frost, "Datamonkey: rapid detection of selective pressure on individual sites of codon alignments," *Bioinformatics*, vol. 21, no. 10, pp. 2531–2533, 2005.
- [49] B. Murrell, J. O. Wertheim, S. Moola, T. Weighill, K. Scheffler, and S. L. Kosakovsky Pond, "Detecting individual sites subject to episodic diversifying selection," *PLoS Genetics*, vol. 8, no. 7, article e1002764, 2012.
- [50] S. L. Kosakovsky Pond and S. D. Frost, "Not so different after all: a comparison of methods for detecting amino acid sites under selection," *Molecular Biology and Evolution*, vol. 22, no. 5, pp. 1208–1222, 2005.
- [51] B. Murrell, S. Moola, A. Mabona et al., "FUBAR: a fast, unconstrained Bayesian approximation for inferring selection," *Molecular Biology and Evolution*, vol. 30, no. 5, pp. 1196–1205, 2013.
- [52] T. Lecocq, S. Dellicour, D. Michez et al., "Scent of a break-up: phylogeography and reproductive trait divergences in the red-tailed bumblebee (*Bombus lapidarius*)," *BMC Evolutionary Biology*, vol. 13, no. 1, p. 263, 2013.
- [53] H. Bozdogan, "Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, 1987.
- [54] F. Ronquist and J. P. Huelsenbeck, "MrBayes 3: Bayesian phylogenetic inference under mixed models," *Bioinformatics*, vol. 19, no. 12, pp. 1572–1574, 2003.
- [55] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets," *Molecular Biology and Evolution*, vol. 33, no. 7, pp. 1870–1874, 2016.
- [56] P. Minias, E. Pikus, L. A. Whittingham, and P. O. Dunn, "A global analysis of selection at the avian MHC," *Evolution*, vol. 72, no. 6, pp. 1278–1293, 2018.
- [57] D. H. Bos and B. Waldman, "Evolution by recombination and transspecies polymorphism in the MHC class I gene of *Xenopus laevis*," *Molecular Biology and Evolution*, vol. 23, no. 1, pp. 137–143, 2006.
- [58] C. Loiseau, M. Richard, S. Garnier et al., "Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*)," *Molecular Ecology*, vol. 18, no. 7, pp. 1331–1340, 2009.
- [59] C. R. Freeman-Gallant, E. M. Johnson, F. Saponara, and M. Stanger, "Variation at the major histocompatibility complex in Savannah sparrows," *Molecular Ecology*, vol. 11, no. 6, pp. 1125–1130, 2002.
- [60] M. Alcaide, M. Liu, and S. V. Edwards, "Major histocompatibility complex class I evolution in songbirds: universal

- primers, rapid evolution and base compositional shifts in exon 3," *PeerJ*, vol. 1, article e86, 2013.
- [61] M. Nei and A. P. Rooney, "Concerted and birth-and-death evolution of multigene families," *Annual Review of Genetics*, vol. 39, no. 1, pp. 121–152, 2005.
- [62] H. C. Miller and D. M. Lambert, "Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petrioidae)," *Immunogenetics*, vol. 56, no. 3, pp. 178–191, 2004.
- [63] R. Burri, H. N. Hirzel, N. Salamin, A. Roulin, and L. Fumagalli, "Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution," *Molecular Biology and Evolution*, vol. 25, no. 6, pp. 1180–1191, 2008.
- [64] E. F. Kikkawa, T. T. Tsuda, D. Sumiyama et al., "Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus *Spheniscus*)," *Immunogenetics*, vol. 61, no. 5, pp. 341–352, 2009.
- [65] J. A. Eimes, S. I. Lee, A. K. Townsend, P. Jablonski, I. Nishiumi, and Y. Satta, "Early duplication of a single MHC IIB locus prior to the passerine radiations," *PLoS One*, vol. 11, no. 9, article e0163456, 2016.
- [66] A. L. Hughes, T. Ota, and M. Nei, "Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules," *Molecular Biology and Evolution*, vol. 7, no. 6, pp. 515–524, 1990.
- [67] D. J. Penn, K. Damjanovich, and W. K. Potts, "MHC heterozygosity confers a selective advantage against multiple-strain infections," *Proceedings of the National Academy of Sciences*, vol. 99, no. 17, pp. 11260–11264, 2002.
- [68] J. L. Bollmer, F. H. Vargas, and P. G. Parker, "Low MHC variation in the endangered Galápagos penguin (*Spheniscus mendiculus*)," *Immunogenetics*, vol. 59, no. 7, pp. 593–602, 2007.
- [69] I. Sepil, S. Lachish, and B. C. Sheldon, "Mhc-linked survival and lifetime reproductive success in a wild population of great tits," *Molecular Ecology*, vol. 22, no. 2, pp. 384–396, 2013.
- [70] H. Schaschl, F. Suchentrunk, S. Hammer, and S. J. Goodman, "Recombination and the origin of sequence diversity in the DRB MHC class II locus in chamois (*Rupicapra* spp.)," *Immunogenetics*, vol. 57, no. 1-2, pp. 108–115, 2005.
- [71] P. Minias, Z. W. Bateson, L. A. Whittingham, J. A. Johnson, S. Oyler-McCance, and P. O. Dunn, "Contrasting evolutionary histories of MHC class I and class II loci in grouse—effects of selection and gene conversion," *Heredity*, vol. 116, no. 5, pp. 466–476, 2016.
- [72] J. A. Anmarkrud, A. Johnsen, L. Bachmann, and J. T. Lifjeld, "Ancestral polymorphism in exon 2 of bluethroat (*Luscinia svecica*) MHC class II B genes," *Journal of Evolutionary Biology*, vol. 23, no. 6, pp. 1206–1217, 2010.
- [73] Q.-Q. Zeng, K. He, D. D. Sun et al., "Balancing selection and recombination as evolutionary forces caused population genetic variations in golden pheasant MHC class I genes," *BMC Evolutionary Biology*, vol. 16, no. 1, p. 42, 2016.
- [74] J. W. Wynne, M. T. Cook, B. F. Nowak, and N. G. Elliott, "Major histocompatibility polymorphism associated with resistance towards amoebic gill disease in Atlantic salmon (*Salmo salar* L.)," *Fish & Shellfish Immunology*, vol. 22, no. 6, pp. 707–717, 2007.
- [75] Å. A. Borg, S. A. Pedersen, H. Jensen, and H. Westerdahl, "Variation in MHC genotypes in two populations of house sparrow (*Passer domesticus*) with different population histories," *Ecology and Evolution*, vol. 1, no. 2, pp. 145–159, 2011.
- [76] Q. Ye, K. He, S. Y. Wu, and Q. H. Wan, "Isolation of a 97-kb minimal essential MHC B locus from a new reverse-4D BAC library of the golden pheasant," *PLoS One*, vol. 7, no. 3, article e32154, 2012.
- [77] M. A. Gillingham, A. Courtiol, M. Teixeira, M. Galan, A. Bechet, and F. Cezilly, "Evidence of gene orthology and trans-species polymorphism, but not of parallel evolution, despite high levels of concerted evolution in the major histocompatibility complex of flamingo species," *Journal of Evolutionary Biology*, vol. 29, no. 2, pp. 438–454, 2016.
- [78] E. Marmesat, K. Schmidt, A. P. Saveljev, I. V. Seryodkin, and J. A. Godoy, "Retention of functional variation despite extreme genomic erosion: MHC allelic repertoires in the Lynx genus," *BMC Evolutionary Biology*, vol. 17, no. 1, p. 158, 2017.
- [79] W. Jaratlerdsiri, S. R. Isberg, D. P. Higgins, L. G. Miles, and J. Gongora, "Selection and trans-species polymorphism of major histocompatibility complex class II genes in the order Crocodylia," *PLoS One*, vol. 9, no. 2, article e87534, 2014.
- [80] K. T. Ballingall, M. S. Rocchi, D. J. McKeever, and F. Wright, "Trans-species polymorphism and selection in the MHC class II DRA genes of domestic sheep," *PLoS One*, vol. 5, no. 6, article e11402, 2010.

Retraction

Retracted: Exploring the Potential of Interferon Gamma Gene as Major Immune Responder for Bovine Tuberculosis in River Buffalo

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named

external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] M. Javed, S. A. Raza, A. Nadeem, M. M. Ali, W. Shehzad, and K. Mehmood, "Exploring the Potential of Interferon Gamma Gene as Major Immune Responder for Bovine Tuberculosis in River Buffalo," *BioMed Research International*, vol. 2021, Article ID 5532864, 7 pages, 2021.

Research Article

Exploring the Potential of Interferon Gamma Gene as Major Immune Responder for Bovine Tuberculosis in River Buffalo

Maryam Javed ¹, Syed Ahmed Raza ¹, Asif Nadeem,² Muhammad Muddassir Ali,¹ Wasim Shehzad,¹ and Khalid Mehmood ³

¹Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences, Lahore, Pakistan

²Department of Biotechnology, Virtual University of Pakistan, Lahore, Pakistan

³Faculty of Veterinary and Animal Sciences, The Islamia University of Bahawalpur, 63100, Pakistan

Correspondence should be addressed to Maryam Javed; maryam.javed@uvas.edu.pk

Received 19 February 2021; Revised 8 March 2021; Accepted 23 March 2021; Published 7 April 2021

Academic Editor: Borhan Shokrollahi

Copyright © 2021 Maryam Javed et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Bovine tuberculosis (bTB) is a widespread zoonotic infection targeting the livestock sector, especially in developing countries, and posing a risk to humans and animal populations. Its recent prevalence in river buffaloes has been estimated as higher as 33.7%. In emergent countries like Pakistan, there is likeliness of human-livestock interfaces extensively and lacking of effective preventive measures that illustrate the risk of spreading the infection at a remarkable rate. The river buffalo (*Bubalus bubalis*) is an upkeep host of *Mycobacterium bovis* and is responsible for disease transmission among buffaloes and other livestock species. In this study, potential molecular biomarkers in the Interferon-gamma gene (IFNg) were identified after genomic screening of river buffaloes. Unique genomic loci in river buffalo proved the novelty of the genomic structure of this phenomenal animal but also highlighted its significance in natural immunity against the *Mycobacterium*. A total of eight single nucleotide polymorphisms were identified in the coding region of IFNg. The SNPs in the exonic region were all transitions, i.e., the conversion of purines to pyrimidines. These SNPs were analyzed for Hardy Weinberg Equilibrium, χ^2 test, gene diversity, and protein structural conformation. Pathway analysis in tuberculosis revealed that IFNg inhibits the antigen-presenting cells (APC) through JAK and STAT pathways. Network analysis of IFNg proteins in both species showed strong associations among the immunity-related proteins (interleukins, tissue necrosis factors) and receptors of interferons. The identified polymorphic sites might be novel-potential markers for the selection of animals with superior immune response against bTB and can be exploited as promising genomic sites for breeding the resistant animal herds to combat *Mycobacterium* infection in a long run.

1. Introduction

Bovine tuberculosis is a widespread infectious disorder affecting animals as well as humans. A major cause of this infection is a slow-growing, obligatory bacterium, which exists inside the cells, *Mycobacterium bovis*. This microbe is referred to as a well-adapted and “successful” pathogen. Its distribution is global, and several countries have been affected by its devastation in terms of economic losses and reduced herd health at livestock farms [1]. Many advanced countries have opted for the use of pasteurization in order to reduce the spread of infection, but the disease continues to cause economic/production losses when poorly controlled. The Office International des Epizooties categorized bTB as a

List-B disease, a disease which has been declared as major public health concern over the last decade and affects the socioeconomic status of local farmers and animal breeders. It also bears a significant impact on the animal trading and export of animal goods [2, 3]. In underdeveloped countries like Africa, *M. bovis* infection has been reported in a wide range of animal species [4]. Bovine TB is also one of the major zoonotic diseases in Pakistan. The prevalence of bovine TB in buffaloes in the country has been reported from 0.51% to 12.72% [5]. In Pakistan and India, the incidence of bTB in cattle and buffalo was found in about 2.25% of the population in 1969 and 10% in 1975. In some countries, human disease caused by *M. bovis* is merely reported as TB to avoid inquiries from disease control departments, which

TABLE 1

(a) Description of condition scores for Positive samples

Score	Condition	Features (antemortem)	(Postmortem)	PPD*
1	L-	Marked emaciation	Caseous necrosis	+ve
2	L	Transverse processes project prominently, neural spines appear sharply.	Mineralization	+ve
3	L+	Individual dorsal spines are pointed to the touch; hips, pins, tail-head, and ribs are prominent. Transverse processes visible, usually individually	Cavitation	+ve

*Purified protein derivative.

(b) Description of condition scores for negative samples

Score	Condition	Features (antemortem)	(Postmortem)	PPD
1	F-	Animal smooth and well covered, but fat deposits are not marked. Dorsal spines can be felt with firm pressure, but feel rounded rather than sharp.		-ve
2	F	Fat cover in critical areas can be easily seen and felt; transverse processes cannot be seen or felt.	No lesion	-ve
3	F+	Heavy deposits of fat clearly visible on tail-head, brisket, and cod; dorsal spines, ribs, hooks, and pins fully covered and cannot be felt even with firm pressure.		-ve

might produce problems of patient privacy [6]. Transmission of *Mycobacterium bovis* in animal herds is usually via inhalation and ingestion [7]. Resistance to infection could be mediated either by the innate or adaptive immune system, and, therefore, the genomic regions related to either of these processes may be considered as candidate genes for bTB susceptibility loci [7]. IFN γ is clearly an imperative cytokine in the control of the infection largely due to its potent role in macrophage activation [8–11] and is one of the major responders in *Mycobacterium* infections.

The aim of the study provides evidence in the progression of bovine tuberculosis to clinical stages is associated with reduced expression of IFN γ at the site of infection. Bovine IFN γ is the major gene involved in the production of IFN γ , a cytokine involved in delayed type of hypersensitivity response in buffalo. In the present study, candidate polymorphisms in IFN γ gene that play a part in the immunologic response were identified. Hardy Weinberg equilibrium and χ^2 testing depicted the significance of each locus in a population. Association testing was performed to evaluate the significance of each locus with bTB susceptibility. Network and pathway analyses were also performed. Finally, protein structural configuration was studied to evaluate the genomic closeness with other bovine species. Results of this study illustrate the uniqueness of the genomic architecture of IFN γ in Pakistani river buffalo, and identified polymorphisms provide a better understanding of superior animal selection that confers immunity against bTB.

2. Materials and Methods

This study was conducted to explore the novel SNPs in the exonic regions of IFN γ , which affect the immunologic status of the buffalos that ultimately leads to the resistance or susceptibility of bovine TB in buffalo. This research work was conducted at the Postgraduate Genomics lab, Institute of Biochemistry and Biotechnology, University of Veterinary

TABLE 2: IFN γ gene primers.

Sr. no.	Primer names	5'-3' sequences
1	IFNG1 (F)	5-ccagcaccacaaaggagacac-3
	IFNG1 (R)	5-gaagactagagatgagagccca-3
2	IFNG2 (F)	5-gtgccagcatccaagttaa-3
	IFNG2 (R)	5-agcaacaggaatcagccaa-3
3	IFNG3 (F)	5-tcctaagtactcataggcctga-3
	IFNG3 (R)	5-tgtttcatttaaccagcccc-3

and Animal Sciences (UVAS), Lahore. This study has been approved by the University of Veterinary and Animal Sciences Research Ethics Committee and has been performed in accordance with the ethical standards.

2.1. Animals Selection Criteria. A clutch of 50 animals (buffalos) with specific phenotypic features for the Nili-Ravi buffalo breed were selected from different relative herds with their Bovine TB status records from UVAS Pattoki campus, Research Farm B, and Buffalo Research Institute (BRI) Pattoki. These 50 animals were bifurcated into two heads for the purpose of blood sampling, each group consisting of 25 animals. The two groups were as follows:

- (1) Nili Ravi buffalo showing positive skin tuberculin test and Physical features
- (2) Nili Ravi buffalo with negative skin tuberculin test and no previous history of lung infection

The characterization of the Positive bTB animals was done on the basis of certain phenotypic parameters as emaciation, prominent transverse process, sharpened neural spines, dorsal spines individually pointed to the touch and hips, pins, tail-head, and ribs were eminent. Scoring system proposed

TABLE 3: SNPs identified in IFNg in Nili Ravi buffalo.

SNP	Location	Wild	Mutant	Transition/ transversion	Chi ² (<0.05)	HWE
<i>p.264G > C</i>	45830264	G	C	Transversion	0.040418	S*
<i>p.274G > A</i>	45830274	G	A	Transition	0.033603	S
<i>p.294 T > C</i>	45830294	T	C	Transition	0.243009	NS*
<i>p.336G > A</i>	45830336	G	A	Transition	0.000649	S
<i>p.342 T > C</i>	45830342	T	C	Transition	0.000262	S
<i>p.433 T > A</i>	45830433	T	A	Transversion	1.000000	NS
<i>p.1441A > C</i>	45831441	A	C	Transversion	0.015112	S
<i>p.1545G > C</i>	45831545	G	C	Transversion	0.000111	S

*S: significant. *NS: nonsignificant.

for gross pathologic condition was used as proposed by [12]. This scoring system was based on the lesions in viscera specially lungs. Lesions were scored for caseous necrosis, mineralization, cavitation, granuloma formation, etc. In this system, varying degrees of lesions from firm or hard white, grey, or yellow nodule with a yellow, caseous, necrotic center, which was dry and solid to thin-walled suppurative abscesses were categorized as postmortem positive Elizabeth et al.1996. Detail is mentioned in Tables 1(a) and 1(b).

2.2. Sampling Strategy. The blood sampling of mentioned animal groups (based on the phenotypic expression of anatomical signs antemortem and postmortem scoring) was carried out. Animals group ($n = 25$) without these lesions was classified as postmortem negative.

2.3. Blood Sampling. Observing standard operating procedures blood samples were drawn from the jugular vein into a 50 ml falcon tube containing 200 μ l of anticoagulant, i.e., EDTA (ethylene diamine tetra—acetic acid).

2.4. Genomic DNA Extraction/Quantification. DNA extraction was done by using standard organic method using phenol-chloroform, isoamyl alcohol (Sambrook and Russell, 2001) [13], and quantification was done by using NanoDrope Spectrophotometer and 0.8% agarose gel electrophoresis.

2.5. Genetic Characterization of IFNg Gene. Interferon-gamma in bovine species is located on chromosome 5. It has 4 exons with a total length of 2743 bp. Genomic characterization was done by designing specific primers to amplify the various regions of the gene by using Primer3 software (<http://frodo.wi.mit.edu/>) and using the sequence reported in the NCBI database (<https://www.nlm.nih.gov>) (Table 2).

2.6. Amplification and Sequencing the PCR Products. Primers were optimized and amplified on specific annealing temperature and PCR reaction mixture. PCR products were precipitated by using ethanol and sequenced using the ABI Genetic analyzer 3130 XL (Applied Biosystems, USA).

2.7. Bioinformatics Analysis. Sequences were aligned by using ClustalW, Multiple Alignment Tool (<https://www.genome.jp/tools-bin/clustalw>). Protein structural configuration of

TABLE 4: Allele frequency of multiple loci identified in IFNg in Nili Ravi buffalo.

Allele/locus	Allele A	Allele B
<i>p.264G > C</i>	0.3030	0.6970
<i>p.274G > A</i>	0.7523	0.2477
<i>p.294 T > C</i>	0.3733	0.6267
<i>p.336G > A</i>	0.3939	0.6061
<i>p.342 T > C</i>	0.6371	0.3629
<i>p.433 T > A</i>	0.2097	0.7903
<i>p.1441A > C</i>	0.5432	0.4568
<i>p.1545G > C</i>	0.9137	0.0863

TABLE 5: Shannon index of genomic variations in IFNg gene depicting diversity.

Locus	Sample size	na*	ne*	I
<i>p.264G > C</i>	50	2.0000	1.9231	0.6595
<i>p.274G > A</i>	50	2.0000	2.0000	0.6743
<i>p.294 T > C</i>	50	2.0000	1.6000	0.6288
<i>p.336G > A</i>	50	2.0000	2.0000	0.5710
<i>p.342 T > C</i>	50	2.0000	1.7241	0.1937
<i>p.433 T > A</i>	50	2.0000	1.1050	0.6402
<i>p.1441A > C</i>	50	2.0000	1.6000	0.6743
<i>p.1545G > C</i>	50	2.0000	1.0000	0.5823

IFNg in *Bos taurus* and Nili-Ravi buffalo was created by using Phyre2 software (<http://www.sbg.bio.ic.ac.uk/>). Further pathway analysis was of IFNg in tuberculosis disease pathway was performed through KEGG (<https://www.genome.jp/kegg/pathway.html>), and network analysis of IFNg in *Bos taurus* and Nili-Ravi buffalo was carried out using STRING (ver. 11.0) knowledgebase resource (<https://string-db.org/>).

2.8. Statistical Analysis. Evaluation of the difference between allelic and genotypic frequency of the gene under study was

TABLE 6: Summary of heterozygosity statistics of IFNg in Nili Ravi Buffalo.

Locus	Obs_Hom	Obs_Het	Exp_Hom*	Exp_Het*	Nei**	Ave_Het
<i>p.264G > C</i>	0.8000	0.2000	0.4947	0.5053	0.4800	0.4800
<i>p.274G > A</i>	0.8000	0.2000	0.4737	0.5263	0.5000	0.5000
<i>p.294 T > C</i>	0.5000	0.5000	0.6053	0.3947	0.3750	0.3750
<i>p.336G > A</i>	1.0000	0.0000	0.4737	0.5263	0.5000	0.5000
<i>p.342 T > C</i>	1.0000	0.0000	0.5579	0.4421	0.4200	0.4200
<i>p.433 T > A</i>	0.9000	0.1000	0.9000	0.1000	0.0950	0.0950
<i>p.1441A > C</i>	0.9000	0.1000	0.6053	0.3947	0.3750	0.3750
<i>p.1545G > C</i>	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Mean	0.8900	0.1100	0.6584	0.3416	0.3245	0.3245
St. dev	0.1595	0.1595	0.2203	0.2203	0.2093	0.2093

done by using chi-square test and association analysis through calculating the odds ratio using SHEsis (<http://analysis.bio-x.cn>).

3. Results

The purpose of this study was to identify the effect of single nucleotide polymorphisms in the interferon-gamma gene in Nili Ravi buffalo. Different statistical and bioinformatics tools were used to analyze the genetic variations.

3.1. Statistical Significance of Data. A total of eight SNPs (single nucleotide polymorphisms) were identified in the coding region of IFNg. Among all the eight SNPs identified, four were transitions (*p.274G > A*, *p.294 T > C*, *p.336G > A*, and *p.342 T > C*) and remaining were transversions (*p.264G > C*, *p.433 T > A*, *p.1441A > C*, and *p.1545G > C*) (Table 3). *p.294 T > C* and *p.433 T > A* were nonsignificant ($P < 0.05$) and were obeying Hardy Weinberg Equilibrium (HWE). *p.294 T > C* was observed having the highest value for chi²; 0.243009 (>0.05). This variation is a potentiated marker for single locus association analysis and further selection in the breeding program. The remaining all were significantly deviating from HWE. The lowest chi² score was observed for *p.1545G > C* (0.000111 <0.05) (Table 3).

Allelic frequencies were also calculated for all loci (Table 4). The frequency range for allele-A was from 0.2097 to 0.9137, while for allele-B was from 0.0863 to 0.7903. Shannon index was also calculated for all variations depicting the gene diversity for all loci. The range of the *I*-score was 0.1937 to 0.6743, indicating the lower to medium diversity of this gene in river buffalo (Table 5).

Heterozygosity statistics were also calculated indicating the expected and observed heterozygosity of each locus. The average score of gene heterozygosity was 0.3245 with a standard deviation of 0.2093 (Table 6). Haplotypes were constructed by blocking the genomic variations. The odds ratio was calculated with a confidence interval (CI) of 95%, and maximum score observed was 2.667 with chi² value 0.007 (Table 7). This indicated the significance of these variations to move together into the next generation.

TABLE 7: Association analysis of haplotypes by calculating odds ratio in IFNg gene.

Haplotype	Chi ²	Pearson's <i>P</i>	Odds ratio (95% CI)
A B A A B B B A	0.916	0.338615	—
A B B A B B B A	0.007	0.931562	1.133
A B B B B B B A	0.036	0.931562	1.133
B B A A B B B B	0.094	0.848878	1.167
B B B A B B B A	0.007	0.289049	2.667
B B B B A B B A	0.916	0.122060	—

3.2. Bioinformatics Significance of the Data. Bioinformatics analysis was done to construct a three-dimensional protein structural model and secondary structure prediction (Figure 1). It was observed that the numbers of residues were the same in Nili-Ravi buffalo and *Bos taurus* protein of IFNg. The structural model of both species did not show significant variation. Both models were carrying α -Helix and β -sheets. A significant variation in structural configuration was observed in Nili-Ravi buffalo in the form of a transmembrane helix between residues 66-81 directing from extracellular matrix to cytoplasmic medium (Figure 2). This helix could not be predicted in the 3-D structure of IFNg protein in *Bos taurus*. Pathway analysis in tuberculosis revealed that IFNg inhibits the antigen-presenting cells (APC) through JAK and STAT pathways (Figure 3). No difference in pathways of IFNg mediated inhibition of APC was found in *B. Taurus* and *B. Bubalis*. Network analysis of IFNg proteins in both species showed strong associations among the immunity-related proteins (interleukins, tissue necrosis factors) and receptors of interferons (Figure 4).

4. Discussion

The bacterium, *Mycobacterium bovis*, causes chronic respiratory disease and has a major impact on the cattle and buffalo industry worldwide as well as posing a risk to humans and other animal populations. In Punjab (Pakistan), the prevalence of bovine tuberculosis varies in buffaloes from 12.48% to 33.72% [14]. Despite over sixty years of costly

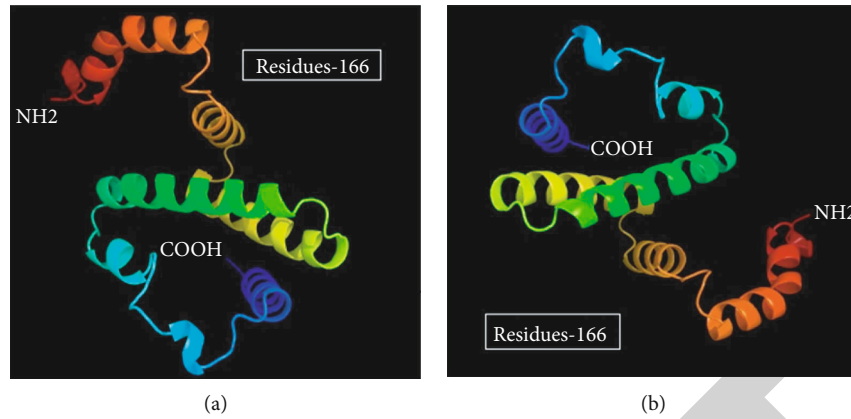


FIGURE 1: 3-D protein structural configuration of IFNg protein in Nili Ravi Buffalo and *Bos taurus*.

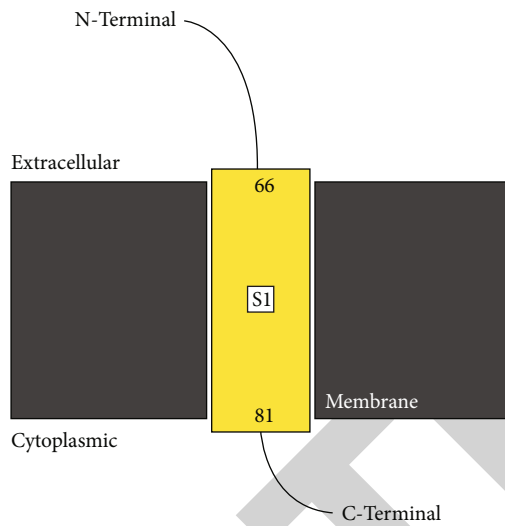


FIGURE 2: Transmembrane loop identified in IFNg protein of Nili Ravi buffalo.

eradication programs, including the vaccination, regular monitoring at farms, and periodic slaughter of infected animals, the number of cases continues to rise. According to Health Department, Government of Punjab, Pakistan is ranked 8th among the 22 high tuberculosis burdened countries in the world and about 300,000 new cases are added each year. There are also issues like drug residues in animal products and drug resistance by the pathogens that demand new control strategies for this infection. Previous studies have suggested that animals differ genetically in their risk of bovine tuberculosis. This has opened up the possibility of breeding animals that have a lower risk of becoming infected with bovine tuberculosis. Colin [15] and his team provided the first hand report about the genetic basis to natural resistance to infections such as bovine tuberculosis, with an estimated heritability of 0.48 (standard error, 0.096; $P < 0.01$). The breeding of selection lines of resistant and susceptible animals provided an ideal strategy for reducing the number and severity of outbreaks of Tb in farmed animals. Identify-

ing candidate genes and novel polymorphisms are crucial to economically important infectious diseases, which ultimately can provide a method for the selection of resistant animals in a more targeted and efficient way. But this solution is a long-term one and needs extensive research. Previous studies have suggested that animals differ genetically in their risk of bovine tuberculosis [16–18]. This has opened up the possibility of breeding animals, which have a lower risk of becoming infected with bovine tuberculosis.

In this context, this study was planned for genetic screening of the Interferon-gamma gene in the Nili Ravi buffalo of Pakistan. A total of eight SNPs were identified in the coding region of the IFNg gene. The one SNP found in the current research is in compliance with the research on IFNg, hence, seven SNPs found in the current research are novel in Nili Ravi buffalo. SNPs in the IFNg gene have also been reported by [19] and considered as candidate markers for controlling the Bovine Tuberculosis in Chinese Holstein cattle.

For the analysis of population, genetics at all the loci POPGENE 32 software was used. By using this software, overall allele frequency, heterozygosity, probability using Chi-square test, and Likelihood ratio test Hardy-Weinberg equilibrium, gene diversity of all SNP position was calculated.

The results of the study follow the postulates of the Hardy-Weinberg equilibrium demonstrating that the two of the identified alleles were randomly distributed throughout the population, no migration had occurred, no bottlenecks happened, and the population remained large in numbers. These nonsignificant variations were obeying HWE and can be potential markers for genetic selection. Polymorphisms with a probability value below 0.05 indicated that the population at these polymorphic sites was not obeying Hardy-Weinberg equilibrium. This indicated that at these positions, alleles were not equally distributed in population.

The mean value of effective number of alleles (n_e) was 1.59522, and the mean value score for the observed number of alleles (n_a) was 2.0000, which is higher than (n_e) demonstrating that Mutant alleles are more in the animal population, which could not pass on to the next generation as a whole. The mean value of observed homozygosity and observed heterozygosity was 0.8900 and 0.1100. Similarly,

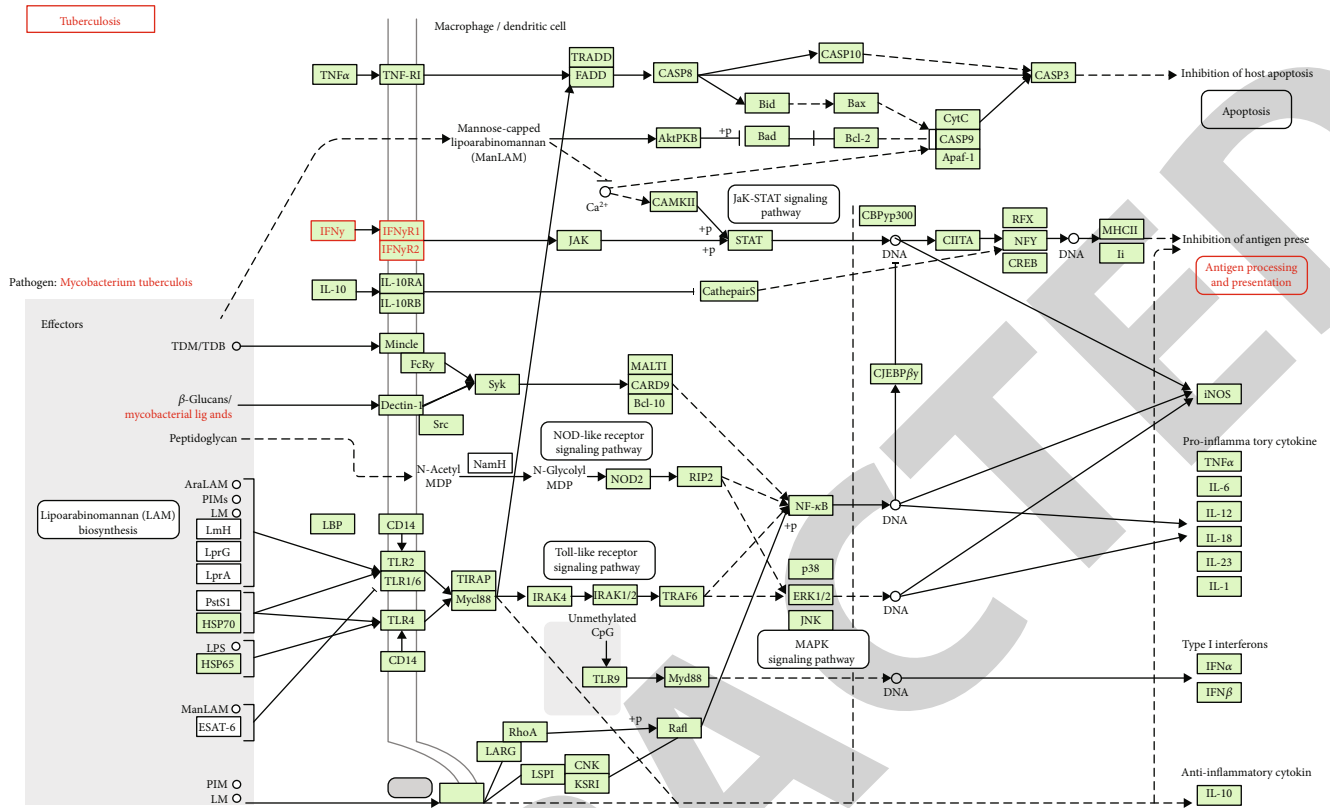


FIGURE 3: KEGG Pathway analysis of IFN γ protein in tuberculosis disease. The role of IFN γ is represented in the inhibition of antigen-presenting cells through the JAK and STAT pathways.

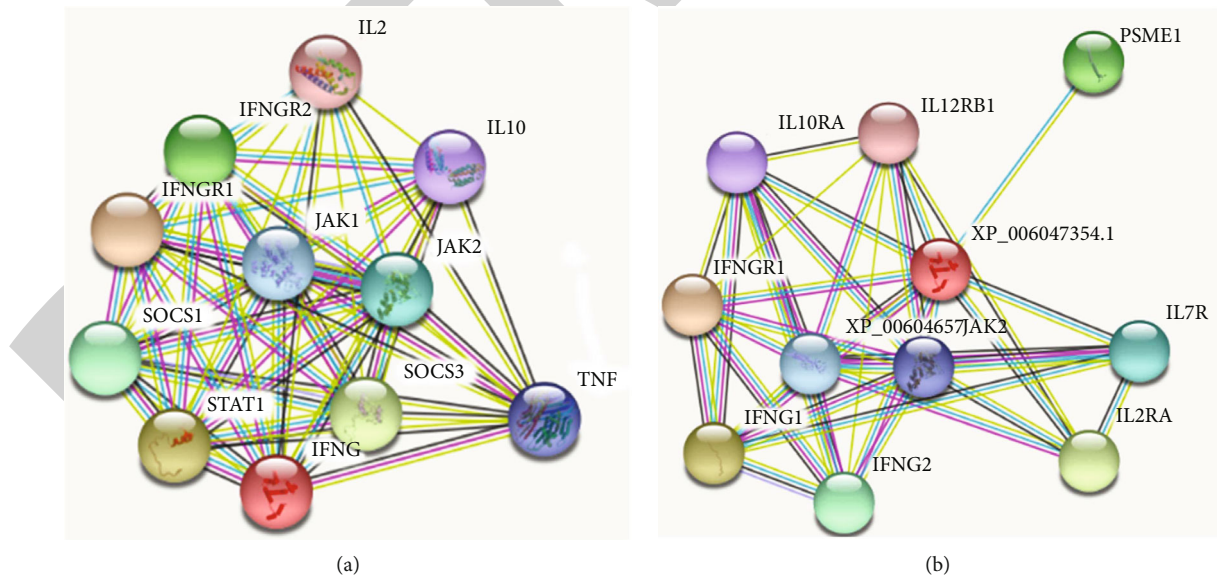


FIGURE 4: Network analysis of IFN γ protein of (a) *B. Taurus* and (b) *Nili Ravi buffalo* using STRING knowledgebase (11.0 Ver.).

expected homozygosity and heterozygosity mean were 0.6584 and 0.3416, respectively. The average heterozygosity mean value was 0.3245 ± 0.2093 (SD).

3-D protein structure of IFN γ protein in buffalo and cattle was compared, and unique secondary structures were

observed. A transmembrane helix was predicted in buffalo, which was missing in the cattle IFN γ protein of cattle. Moreover, it is also revealed that IFN γ inhibits the antigen-presenting cells (APC) through JAK and STAT pathways in tuberculosis disease [9, 11].

Retraction

Retracted: *In Silico* Characterization of Growth Differentiation Factors as Inhibitors of TNF-Alpha and IL-6 in Immune-Mediated Inflammatory Disease Rheumatoid Arthritis

BioMed Research International

Received 12 March 2024; Accepted 12 March 2024; Published 20 March 2024

Copyright © 2024 BioMed Research International. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named

external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] G. Mustafa, H. S. Mahrosh, and R. Arif, "*In Silico* Characterization of Growth Differentiation Factors as Inhibitors of TNF-Alpha and IL-6 in Immune-Mediated Inflammatory Disease Rheumatoid Arthritis," *BioMed Research International*, vol. 2021, Article ID 5538535, 11 pages, 2021.

Research Article

***In Silico* Characterization of Growth Differentiation Factors as Inhibitors of TNF-Alpha and IL-6 in Immune-Mediated Inflammatory Disease Rheumatoid Arthritis**

Ghulam Mustafa , Hafiza Salaha Mahrosh, and Rawaba Arif

Department of Biochemistry, Government College University, Faisalabad 38000, Pakistan

Correspondence should be addressed to Ghulam Mustafa; gmustafa_uaf@yahoo.com

Received 25 February 2021; Accepted 20 March 2021; Published 27 March 2021

Academic Editor: Borhan Shokrollahi

Copyright © 2021 Ghulam Mustafa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Tumor necrosis factor alpha (TNF- α) plays a critical role in the progression of inflammation and affects the cells of the synovial membrane. Another key factor in the progression of rheumatoid inflammation is interleukin-6 (IL-6). Both TNF- α and IL-6 promote the proliferation of synovial membrane cells thus stimulating the production of matrix metalloproteinases and other cytotoxins and leading towards bone erosion and destruction of the cartilage. Growth differentiation factor-11 (GDF11) and growth differentiation factor-8 (GDF8) which is also known as myostatin are members of the transforming growth factor- β family and could be used as antagonists to inflammatory responses which are associated with rheumatoid arthritis. In the current study, to elucidate the evolutionary relationships of GDF11 with its homologs from other closely related organisms, a comprehensive phylogenetic analysis was performed. From the phylogram, it was revealed that the clade of Primates that belong to superorder Euarchontoglires showed close evolutionary relationships with order Cetartiodactyla of the Laurasiatheria superorder. Fifty tetrapeptides were devised from conserved regions of GDF11 which served as ligands in protein-ligand docking against TNF- α and IL-6 followed by drug scanning and ADMET profiling of best selected ligands. The peptides SAGP showed strong interactions with IL-6, and peptides AFDP and AGPC showed strong interactions with TNF- α , and all three peptides fulfilled all the pharmacokinetic parameters which are important for bioavailability. The potential of GDF8 as an antagonist to TNF- α and IL-6 was also explored using a protein-protein docking approach. The binding patterns of GDF8 with TNF- α and IL-6 showed that GDF8 could be used as a potential inhibitor of TNF- α and IL-6 to treat rheumatoid arthritis.

1. Introduction

The immune system is a collection of chemicals, cells, and pathways that protect the cells from foreign invaders. Beyond chemical and physical barriers to pathogens, the immune system has two fundamental pillars in line of defense which include innate immunity and adaptive immunity [1]. Innate immunity is a nonspecific rapid immune response activated within minutes to hours with the lack of immunologic memory. On the other hand, adaptive immunity is an antigen-dependent response activated upon the subsequent exposure of host cells to antigens. There is synergy between the events in innate immunity and adaptive immunity, and any defect in either system provokes severe problems and diseases such

as autoimmune diseases, inflammations, and immunodeficiency disorders [2].

The immune system preserves homeostasis in immune integrity to protect the host cells from infectious agents. Furthermore, the immune system undergoes processed pathways to initiate self-tolerance to avoid the destruction of tissues or cells of the host immune system. The impairment of the self-tolerance mechanism leads towards the onset of severe autoimmune disorders [3]. Mounting evidence has clearly demonstrated the relation of innate immunity responses in initiation, progression, and maintenance of different autoimmune disorders including lupus erythematosus, rheumatoid arthritis, Alzheimer's disease, Addison's disease, and type I diabetes mellitus [2]. Autoimmunity involves the

self-destruction of own tissues by the production of abnormal responses due to self-reactive T-cells, cytokines, and autoantibodies. To design potential drugs and targets, these self-reactive species are the main focus to maintain different autoimmune disorders [4].

In this study, our main focus was on the systemic autoimmune disease rheumatoid arthritis (RA). Rheumatoid arthritis (RA) is an autoimmune disease characterized by synovial inflammation, decalcification, bone erosion, and destruction of cartilage which result in joint impairment. Currently, there are five main classes of drugs that have been in practice in RA therapy including analgesics, glucocorticoids, biologic disease-modifying antirheumatic drugs (bDMARDs), nonbiologic disease-modifying antirheumatic drugs, and nonsteroidal anti-inflammatory drugs (NSAIDs). The most common biological disease-modifying antirheumatic drugs are TNF- α inhibitors. Despite many drug therapies, the RA patients did not get any significant clinical benefits from these agents [5].

On the basis of sequence identity, shared inhibitors, and receptor utilization, the members of the transforming growth factor- β (TGF- β) family may be divided into three subclasses, i.e., TGF- β , bone morphogenetic protein (BMP)/GDF, and activin/inhibin [6, 7]. A similar protein fold has been observed in all proteins of the TGF- β family, and efforts have been made to understand the structural basis of these individual ligands for their differential signaling. Growth differentiation factor-11 (GDF11) also known as bone morphogenetic protein 11 (BMP11) is a member of the transforming growth factor- β (TGF- β) family [8]. From this family, GDF11 and GDF8 (also known as myostatin) have been found to be the most closely related members and emerged through a gene duplication event in an ancestral gene after the point when divergence of amphioxus occurred from vertebrates [9]. It has been observed that both GDF8 and GDF11 bind with type II receptors of activin which trigger signaling via the Smad2/3 pathway after subsequent phosphorylation of these receptors [10].

The aim of this study includes the investigation of potential tumor necrosis factor alpha (TNF- α) and interleukin-6 (IL-6) inhibitors which could be used in RA maintenance. Currently, the attenuation of inflammatory responses has attracted the attention of researchers in the field of many inflammatory disorders. Recently, the therapeutic effects of GDF11 in mouse models as an antagonist of inflammatory responses associated with RA have been explored [11]. In the current study, we therefore explored the inhibitory nature of GDF11 against the proinflammatory cytokines responsible for rheumatoid arthritis.

2. Materials and Methods

2.1. Phylogenetic Analysis. Human *GDF11* gene sequence was retrieved from the NCBI nucleotide database and analyzed using BLAST (Basic Local Alignment Search Tool) [12] available on the NCBI website (<http://www.ncbi.nlm.nih.gov/>). Along with *GDF11* gene sequence of *Homo sapiens*, seventy-four most similar reported sequences of *GDF11* from different mammals were also retrieved from GenBank for

phylogenetic systematics. All sequences were aligned using ClustalX and imported into the MEGA7 program [13] for manual alignment. A Neighbor-Joining (NJ) phylogenetic tree was reconstructed using MEGA7 with 100 bootstrap replicates.

2.2. Ligand-Based Molecular Docking. A ready-to-dock library of 50 ligands was prepared and docked against tumor necrosis factor alpha (TNF- α) and interleukin-6 (IL-6) with the help of Molecular Operating Environment (MOE) software.

2.3. Ligand Selection and Database Preparation. The protein sequence of GDF11 was retrieved from NCBI's Entrez Protein under accession No. O95390.1. MEME Suite was used to predict three motifs from nine selected homologs of GDF11 [14, 15]. The chemical structures of 50 tetrapeptides were drawn from the consensus sequences of predicted motifs using ACD/ChemSketch and saved in MOL format. All the tetrapeptides were minimized and saved into the MOE database in .mdb format for docking studies.

2.4. Retrieval and Optimization of Receptor Proteins. Three-dimensional structures of TNF- α (PDB ID: 6OP0) and IL-6 (PDB ID: 5FUC) were retrieved from Protein Data Bank (PDB) and optimized by removing water molecule, addition of hydrogen atoms, energy minimization, and 3D protonation using MOE software. The ready-to-dock database of 50 tetrapeptides was docked counter to TNF- α and IL-6 separately. The top five interaction conformations of each receptor protein were selected on the basis of binding patterns and energy conformations.

2.5. Protein-Ligand Docking. Active sites of receptor proteins were predicted by a site finder tool of MOE with default parameters. The prepared database of 50 tetrapeptides from the conserved sequences of GDF11 was docked counter to receptor proteins separately using MOE. The docking program of MOE provides the top conformations on the basis of maximum occupancy of binding pocket and minimum energy structure. The top five ligands from each protein-ligand docking study were selected on the basis of S-scores and kept aside for further verification through drug scan analysis.

2.6. Drug Scanning and ADMET Profiling. Lipinski's rule of five explores the therapeutic nature of selected drug molecules by distinguishing the drug-like and nondrug-like properties of drug candidates. The durability of hit compounds was evaluated using the online server SwissADME [16]. The ADMET (metabolism, distribution, excretion, absorption, and toxicity) attributes of target molecules were discovered using admetSAR server [17]. Only those ligands were considered to be the potential or lead drug candidates that accomplished all the ADMET models successfully.

2.7. Protein-Protein Docking. Appropriate interactions between the member of TGF- β and TNF- α and IL-6 are necessary to check binding residues between them and to inhibit the residues of receptor protein(s). In this study, the chain A

of myostatin (PDB ID: 6UMX) was selected as a template and downloaded from PDB to dock against receptor proteins TNF- α (PDB ID: 6OP0) and IL-6 (PDB ID: 5FUC). The inhibitors of TNF- α and IL-6 have been extensively studied and researched due to their role in reduction of inflammatory response in autoimmune disorders. The HADDOCK server was used to dock myostatin with TNF- α and IL-6 [18]. To visualize the interactions, the educational version of PYMOL was used [19]. Moreover, the online database PUBsum was used to demonstrate the interacting residues involved in protein-protein interactions [20].

3. Results and Discussion

Basic Local Alignment Search Tool (BLAST) is still the most popular search algorithm and can accommodate nucleotide or protein sequences. BLAST was used to identify local regions of similarity and statistical significance of *GDF11* nucleotide sequences from selected organisms. Multiple sequence alignment was also performed through Geneious [21] (Fig S1 of Supplementary file). The truncated sequences were deleted, and longer sequences were shortened in the multiple sequence alignment to make them all equal in length.

3.1. Analysis of *GDF11* Phylogenetic Tree. To establish the evolutionary relationships, a phylogenetic tree of the *GDF11* gene from seventy-six members of the Eutheria clade of mammals with maximum identity including *H. sapiens* was reconstructed (Figure 1). The evolutionary history was inferred using the Neighbor-Joining method [22, 23]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches [24, 25]. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method [26] and are in the units of the number of base substitutions per site. Overall mean distance was found to be 0.052. The tree is fully in accordance with the classification system, and all organisms appeared in their respective clades and subclades. The organisms divided themselves under three superorders, i.e., Euarchontoglires, Laurasiatheria, and Afrotheria, with two, four, and five orders covered, respectively. *Oryctolagus cuniculus* has appeared in Glires and *Homo sapiens* in Primates orders under the same superorder of Euarchontoglires. From the phylogeny of *GDF11*, it was inferred that *O. cuniculus* had the strongest evolutionary relationship with *Ochotona princeps* while *H. sapiens* with *Gorilla gorilla gorilla*. Some interesting observations were also found in the phylogeny of *GDF11*. The Primates order that belongs to superorder Euarchontoglires is showing more identity with order Cetartiodactyla of the Laurasiatheria superorder, inferring that there would be some evolutionary closeness between two orders. The organisms of order Chiroptera have not appeared in a single clade and are divided into two small clades, i.e., Chiroptera I and Chiroptera II. Chiroptera II is showing more closeness with the Perissodactyla order hypothesizing

that the organisms under both clades have strong evolutionary relationships.

A detailed phylogenetic analysis for *GDF11* sequences was not found previously, and therefore, in this study, we conducted a comprehensive phylogenetic analysis. In our phylogenetic analysis, the mammalian *GDF11* clades were divided into different clades on the basis of their suborders. Biga et al. [27] also conducted a phylogeny of *GDF11* using protein sequences. They found that *GDF11* from mammals clustered with that of zebrafish and also found that all myostatin and *GDF11* sequences were grouped together showing high identity between both proteins. In another study [28], *GDF11* and myostatin in vertebrates were found to be more closely related to each other than piscine myostatin when a phylogenetic analysis of *GDF11* and myostatin was performed.

More than 30 ligands have been reported from the TGF- β family, which are structurally related but functionally distinct. The family has 3 subclasses, i.e., TGF- β s, BMPs, and activin/myostatins. From the TGF- β family, *GDF11* and *GDF8* (myostatin) have 90% sequence identity within their mature signaling domain and both are members of the activin/myostatin subclass [29]. Talking about the function of *GDF11*, it is essential for the development of mammals and has been recognized for the regulation of aging of various tissues. *GDF8*, however, has been found as a strong negative regulator of muscle growth and modulates various metabolic processes [30].

Other differences in *GDF11* and *GDF8* are also studied and observed that *GDF11* mRNA is found largely in various tissues [31] but most abundant in the kidney and spleen [32], whereas *GDF8* mRNA is mainly found in cardiac and skeletal muscles. Both proteins are found in the bloodstream, and the functional effects of their circulation are still not clear, but it has been suggested that their presence may act as hormonal signals. It was suggested that a number of features and functions of both GDFs could overlap because of their high sequence similarity, but the latest studies have shown that both ligands have different functions [29]. Due to identical residues in the region, both proteins possibly bind type II receptors in the same fashion, but the residues have been found divergent in the prehelix loop and wrist helix between *GDF11* and *GDF8* conferring that the binding of type I receptors might be different in both ligands [33].

Because of high sequence similarity and conservation between *GDF11* and *GDF8*, the functions of both the ligands were assumed to be similar [34]. It is likely that much of this assumption reflects the fact that *GDF11* is a comparatively understudied ligand of the TGF- β family as compared to other members especially *GDF8*, and thus, the tools to study the biology of *GDF11* are continuously evolving. Therefore, we have studied *GDF11* in more detail and gave its comprehensive evolutionary relationships through its phylogeny.

3.2. Devising of Tetrapeptides as Ligands. Protein BLAST (BLASTp) was used to explore homologs of *GDF11* based on query converge. MEME Suite was then used to explore three motifs from the selected homologs. The most conserved 50 tetrapeptides were devised from the selected

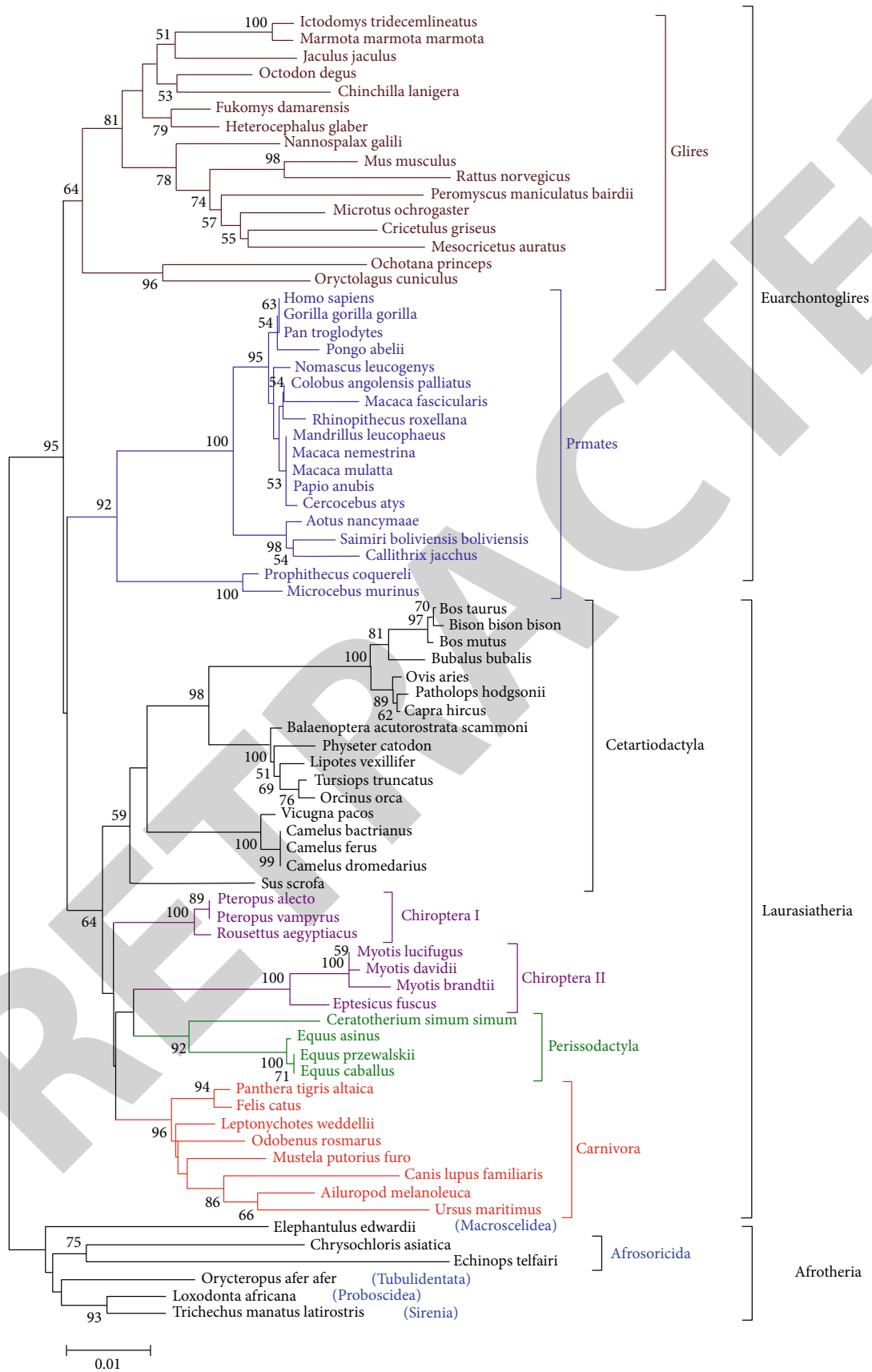


FIGURE 1: Phylogenetic relationships of the *GDF11* gene in selected organisms. The tree is labeled on the basis of orders and suborders of selected animals. The animals of the same suborder cladded together.

motifs as ligand molecules and saved in the MOE database.

3.3. Protein-Ligand Docking. Molecular docking is a computer-aided drug discovery used to predict the perfect orientation of ligand molecules. This study includes a ready-to-dock library of 50 tetrapeptides devised from GDF11 and docked against tumor necrosis factor alpha and interleukin-6. The top hit compounds were selected on the bases of *S*-scores and energy validations for further analyses.

3.4. Interaction Analyses. According to given commands, MOE gave different confirmations of all ligands. The top five ligands from each protein-ligand study were selected as hit compounds on the basis of *S*-scores and binding residues. For TNF- α , the peptide TETV exhibited the best *S*-score (i.e., -14.7233) and showed interactions with amino acids Ser147 and His15 of the binding pocket (Figure 2). Other peptides (i.e., ISMA, ANPR, AFDP, and AGPC with *S*-scores -14.4874, -14.4748, -14.1934, and -13.6792, respectively) also showed interactions with active amino acids of the binding site (Figs 2S to 5S, respectively, of Supplementary file). Table 1 is showing peptides with their docking scores and interacting residues of TNF- α .

Other receptor protein interleukin-6 is a pleiotropic anti-inflammatory myokine and proinflammatory cytokine encoded by the IL-6 gene in humans. In our study, the peptide AQET exhibited the best *S*-score (-11.4547) with interactive amino acid residues (i.e., Met67 and Glu172) of the binding pocket (Figure 3). Four peptides (i.e., DGSP, TETV, SAGP, and GSAG with *S*-scores of -10.9771, -10.2757, -10.0603, and -9.9370, respectively) showed excellent binding interactions with active amino acids (Figs 6S to 9S, respectively, of Supplementary file). Table 1 is showing peptides with their docking scores and interacting residues of IL-6.

The elevated level of plasma IL-6 is associated with insulin resistance in diabetes mellitus [35]. IL-6 binds to the receptor on the cellular surface and forms an IL-6/sIL-6R complex that leads towards activation of cells. IL-6 also promotes T-cell differentiation and at the same time collaborates with TNF- α and IL-1 to induce systemic inflammatory response [36]. TNF- α is also a proinflammatory cytokine used by the immune system in a defense mechanism to induce inflammation against pathogenic response. The deregulation or elevated signaling of TNF- α alters the balance of T-cells thus leading towards the inflammatory diseases such as inflammatory bowel disease, rheumatoid arthritis (RA), and ankylosing spondylitis (AS). Collectively with TNF- α , the role of IL-6 has been studied in humans and different animal models to understand the phenomena of autoimmunity [37]. Thus, there is a dire need for new drugs to regulate the cytokine signaling and reduce the inflammatory responses to treat autoimmune disorders.

3.5. Drug Scan and ADMET Screening. The Lipinski rule of five distinguishes the drug-like and nondrug-like molecules based on five parameters (i.e., molecular mass: ≤ 500 daltons, hydrogen bond donor: < 5 , hydrogen bond acceptor: < 10 ,

molar refractive index: 40-130, and partition coefficient ($\log P$): ≤ 5). The three peptides SAGP, AFDP, and AGPC showed no deviation and fully accomplished the criteria of potent drug candidates. The remaining peptides AQET, DGSP, TETV, SAGP, GSAG, ISMA, and ANPR showed only one violation of Lipinski's rule of five (Table 2). The ADMET-based drug profiling is necessary to estimate the bioavailability of drug candidates. The ADMET study evaluates the hit compounds using different threshold values to estimate the absorption of leading compounds. In the current study, all the selected compounds were found to be noncarcinogens and non-Ames toxic (Table 3). Based on overall drug profiling, these ligands could be used as potential inhibitors of TNF- α and IL-6.

3.6. Protein-Protein Docking. Protein-protein docking is the prediction of binding complexes between two proteins to demonstrate the protein-protein interference. The HADDOCK server was used to perform template-based docking and to predict the interacting residues through hybrid algorithms. In this study, the docking analysis was performed to dock myostatin (GDF8) protein (which showed $> 85\%$ sequence identity with GDF11) with TNF- α and IL-6. The results revealed good binding patterns and interactions between amino acids of two protein complexes. The HADDOCK binding score of GDF-TNF- α was found to be -62.2 kcal/mol and that of GDF-IL-6 was found to be -48.1 kcal/mol. The interactions between GDF8 and TNF- α are shown in Figure 4 while interactions between GDF8 and IL-6 are shown in Figure 5.

The hydrogen bonds between GDF8 and both target proteins are shown in red stick representation. In the case of GDF-TNF- α interactions, it was observed that the active amino acids of GDF8 were involved in six hydrogen bonds with TNF- α ; meanwhile, 16 hydrogen bonds were observed in GDF-IL-6 interactions. The online database PUBsum also demonstrates the hydrogen bonding along with disulfide bonds between the active amino acids of these protein complexes. The overall results and binding patterns of GDF8 with TNF- α and IL-6 indicated that GDF8 could be used as a potential inhibitor of TNF- α and IL-6 receptor proteins.

The molecular docking approach predicts the best orientation of ligands binding with particular amino acids of the binding pocket of receptor molecules [15]. Molecular docking is the major section of drug discovery as it predicts the biological interactions and bioavailability of hit compounds. Computational drug discovery is the most economic and effective route of *in silico* drug designing [38]. Based on the type of ligands, docking can be classified into many types like protein-small molecule (ligand) docking, protein-protein docking, and protein-nucleic acid docking [39]. In the current study, we used two approaches (i.e., ligand-based protein docking and protein-protein docking) of TGF- β members counter to two target proteins such as tumor necrosis factor alpha and interleukin-6.

Human TNF- α is 17 kDa in size, nonglycosylated, and encoded by genes on 6p23-6q12. TNF- α mainly involves in a variety of functions including phagocytosis of neutrophilic granulocytes, cytolysis and cytoablation of many tumor cells,

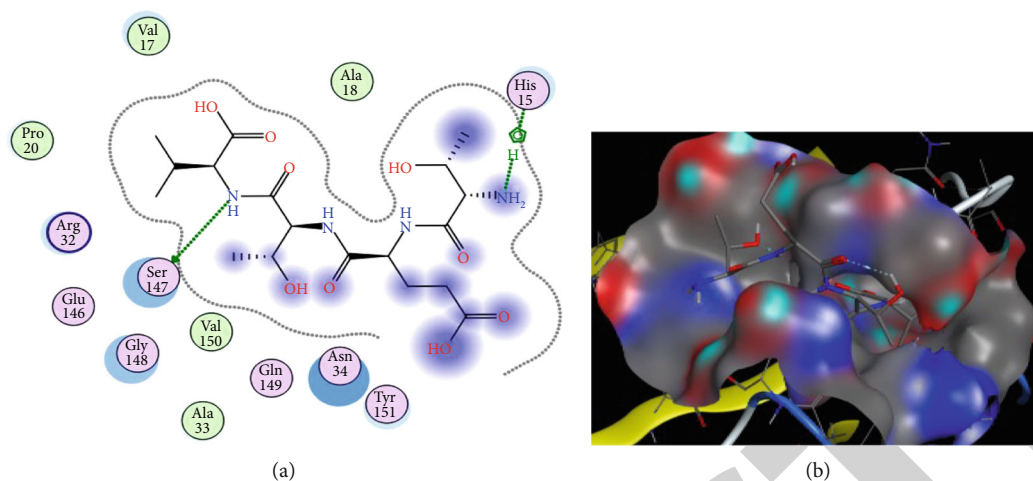


FIGURE 2: Interactions (a) and binding patterns (b) of TETV with TNF- α .

TABLE 1: Top interactions of GDF11 tetrapeptides with TNF- α and IL-6 as receptor proteins.

Sr. No.	Peptide	Receptor	S-score	Interacting residues
1	TETV	TNF- α	-14.7233	Ser147, His15
2	ISMA		Gly148	
3	ANPR		Arg32, Glu146, Val150	
4	AFDP		Ala33, Ser147	
5	AGPC		His15, Gln149, Val150, Ser147	
6	AQET		Met67, Glu172	
7	DGSP		Met67, Arg179	
8	TETV	IL-6	-10.2757	Glu172, Arg179, Arg168
9	SAGP		Glu172, Arg179, Ser169, Ser176	
10	GSAG		Glu172, Glu172, Arg179	

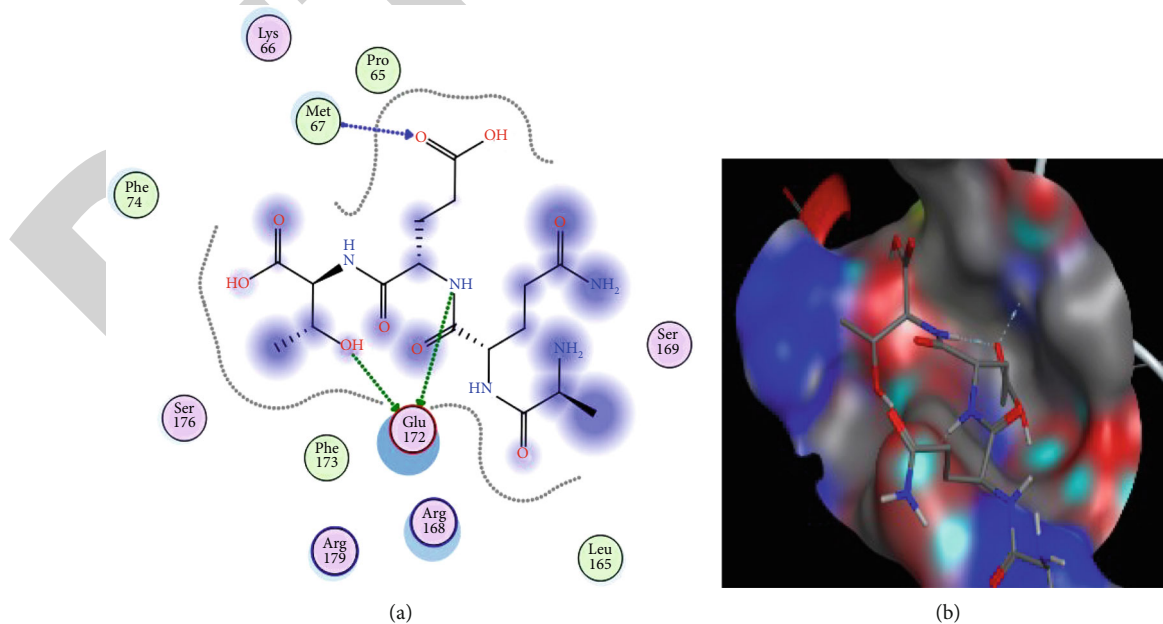


FIGURE 3: Interactions (a) and binding patterns (b) of AQET with IL-6.

TABLE 2: Pharmacokinetic parameters important for bioavailability of compounds drug-likeness properties of selected peptides.

Sr. No.	Peptides	Receptor	Molecular properties [†]					Violations
			Mass (<500 D)	HBD (≤ 5)	HBA (≤ 10)	Log <i>P</i> (<5)	Molar refractivity (40-130)	
1	SAGP	IL-6	330.34	5	7	-2.32	81.15	0
2	AFDP	TNF- α	448.47	5	8	-0.75	115.86	0
3	AGPC	TNF- α	346.40	4	6	-1.58	87.91	0
4	TETV	TNF- α	448.47	10	8	-2.25	106.22	1
5	AQET	IL-6	447.44	8	10	-2.93	103.16	1
6	GSAG	IL-6	290.27	6	7	-2.92	64.83	1
7	ISMA	TNF- α	420.52	6	7	-0.93	106.07	1
8	ANPR	TNF- α	456.50	8	8	-3.21	116.03	1
9	DGSP	IL-6	374.35	6	9	-2.95	87.73	1

HBD: hydrogen bond donors; HBA: hydrogen bond acceptors; log *P*: the logarithm of octanol/water partition coefficient.

TABLE 3: ADMET profiling of the best selected peptides.

	Peptides									
	AQET	DGSP	TETV	SAGP	GSAG	ISMA	ANPR	AFDP	AGPC	
Absorption										
BBB	+	+	+	+	+	+	+	-	+	
HIA	-	-	-	-	-	-	-	-	-	
Caco-2 permeability	-	-	-	-	-	-	-	-	-	
PGS	NS	Substrate	NS	NS	NS	Substrate	Substrate	Substrate	Substrate	
PGI	NI	NI	NI	NI	NI	NI	NI	NI	NI	
ROCT	NI	NI	NI	NI	NI	NI	NI	NI	NI	
Metabolism										
CYP3A4 substrate	NS	Substrate	NS	Substrate	NS	NS	Substrate	Substrate	Substrate	
CYP2C9 substrate	NS	NS	NS	NS	NS	Substrate	NS	NS	NS	
CYP2D6 substrate	NS	NS	NS	NS	NS	NS	NS	NS	NS	
CYP3A4 inhibition	NI	NS	NI	NS	NI	NI	NI	NI	NI	
CYP2C9 inhibition	NI	NI	NI	NI	NI	NI	NI	NI	NI	
CYP2C19 inhibition	NI	NI	NI	NI	NI	NI	NI	NI	NI	
CYP2D6 inhibition	NI	NI	NI	NI	NI	NI	NI	NI	NI	
CYP1A2 inhibition	NI	NI	NI	NI	NI	NI	NI	NI	NI	
Toxicity										
AMES toxicity	NAT	NAT	NAT	NAT	NAT	NAT	NAT	NAT	NAT	NAT
Carcinogens	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC

BBB: blood-brain barrier; HIA: human intestinal absorption; PGS: P-glycoprotein substrate; PGI: P-glycoprotein inhibitor; ROCT: renal organic cation transporter; NS: nonsubstrate; NI: noninhibitor; NAT: non-Ames toxic; NC: noncarcinogenic.

and modulations of many other proteins such as IL-1 and IL-6 [40]. TNF- α is a pleiotropic cytokine, a key mediator of chronic and acute inflammations secreted by macrophages, neutrophils, and T-cells [41]. With the combination of IL-6 and IL-1, TNF- α performs a variety of functions including endothelium alterations, inhibition of anticoagulatory mechanisms, and B-cell differentiation and proliferation. Although TNF- α is a necessary cytokine, its overexpression or deregulation is associated with different autoimmune diseases [40].

Interleukin-6 is a proinflammatory cytokine and an anti-inflammatory myokine transiently produced by the host immune system as the result of an infection. Human IL-6

protein is of ~20 kDa and contains 212 amino acids encoded by genes on chromosome 7p21 [42]. Interleukin-6 plays a critical role in host immunity, acute-phase reactions, and hematopoiesis. Although the expression and regulation of IL-6 have been completely controlled by posttranscriptional mechanisms, any mutation in the IL-6 regulatory mechanism leads towards chronic inflammation and autoimmunity [43]. Under the presumption that IL-6 and TNF- α cause autoimmunity, therefore, several drugs have been designed to treat autoimmune diseases. All the anti-TNF- α and IL-6 therapies have been totally designed to negate the TNF- α and IL-6 activity by blocking the TNF transcription and NF-mediated downstream signaling [40, 44].

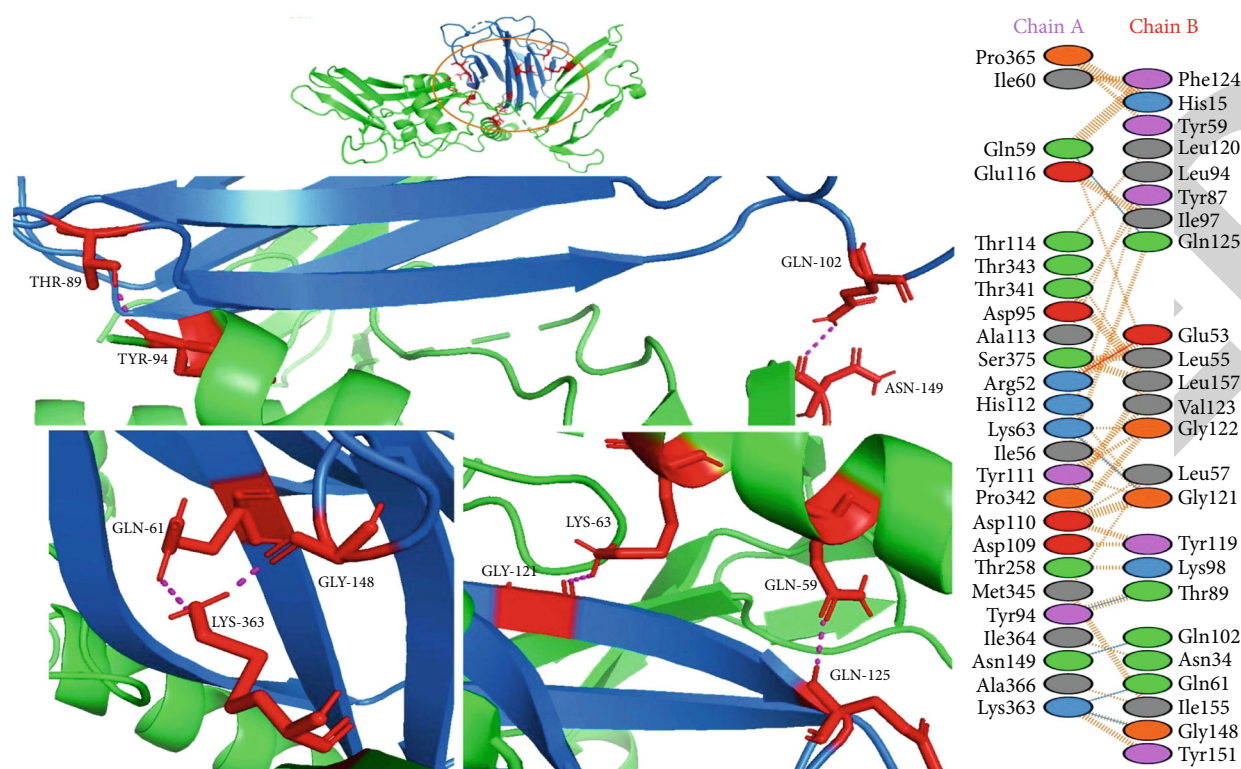


FIGURE 4: Protein-protein interactions between myostatin and TNF- α . (a) GDF8/TNF- α docked complex in cartoon representation is shown at the top. Myostatin is shown in green while TNF- α is shown in blue. Interacting residues of the GDF8/TNF- α complex are shown in red color stick representation. (b) All interacting residues of GDF8 and TNF- α . Blue color: H-bonds, red color: salt bridges, yellow color: disulfide bonds, orange color: other contacts. Hydrogen bonds are shown with blue lines. The colors of interacting residues are representing properties of amino acids (i.e., blue: positive, red: negative, green: aliphatic, grey: aliphatic, pink: aromatic, orange: Pro/Gly, and yellow: Cys).

The blockage of TNF- α and IL-6 has been expected as an effective strategy for the treatment of many autoimmune disorders. Many IL-6 antibodies have been developed to inhibit the IL-6 signaling and transduction pathways. Among all the designed antibodies, anti-IL-6Rab commonly known as tocilizumab showed efficient results in clinical trials for rheumatoid arthritis as anti-IL-6Rab. The tolerability profiles of anti-IL-6Rab monotherapy proved safe and potent and therefore could be used as an anti-RA antibody. Several clinical studies supported the inhibitory nature of tocilizumab as an anti-IL-6 receptor monoclonal antibody and showed astonishing efficacy in RA [42, 45]. In another study, suppression of proinflammatory cytokine responses by targeting the JAK/STAT pathway has been demonstrated as a therapeutic option in saphenous vein graft failure. The results of different clinical trials including suppression of cytokine signaling (SOCS3) showed that it appears as a major contributor to suppression of unwanted vascular inflammation and proliferation [46].

Wang et al. [47] demonstrated the dual nature of different ligands as potential inhibitors of TNF- α and IL-6. The results of their study showed that two molecules (i.e., ZINC19701771 and ZINC06576501) exhibited the best binding pattern and pharmacological properties and therefore could be served as potential candidates for the development of an anti-RA drug. In the current study, we used two

approaches of molecular docking to check the therapeutic nature of two members of the TGF- β family (i.e., GDF11 and GDF8/myostatin) as antagonists of TNF- α and IL-6. In the first approach, ligand-based docking was performed and 50 tetrapeptides were devised from conserved regions of GDF11 counter to two inflammatory receptor proteins such as TNF- α and IL-6. On the basis of confirmations and orientation of ligands, the top five peptide molecules were selected for further pharmacokinetic analysis. The peptide TETV with the best *S*-score (-14.7233) showed interactions with amino acids Ser147 and His15 present in the binding pocket of TNF- α . Similarly, the peptide AQET was found with the top *S*-score (-11.4547) and interacted with amino acids Met67 and Glu172 of the binding pocket of IL-6 receptor protein.

The drugability test of all selected peptides fulfilled the criteria of being potential drug candidates by following the Lipinski rule of five. Out of ten selected peptides, three peptides (i.e., SAGP, AFDP, and AGPC) followed all the five parameters of Lipinski's rule of five while other peptides showed only one violation. The ADMET profiling of all leading compounds was also found satisfactory as all hit compounds were revealed to be noncarcinogens and non-Ames toxic. Therefore, based on overall drug profiling, these ligands could be used as potential inhibitors of TNF- α and IL-6.

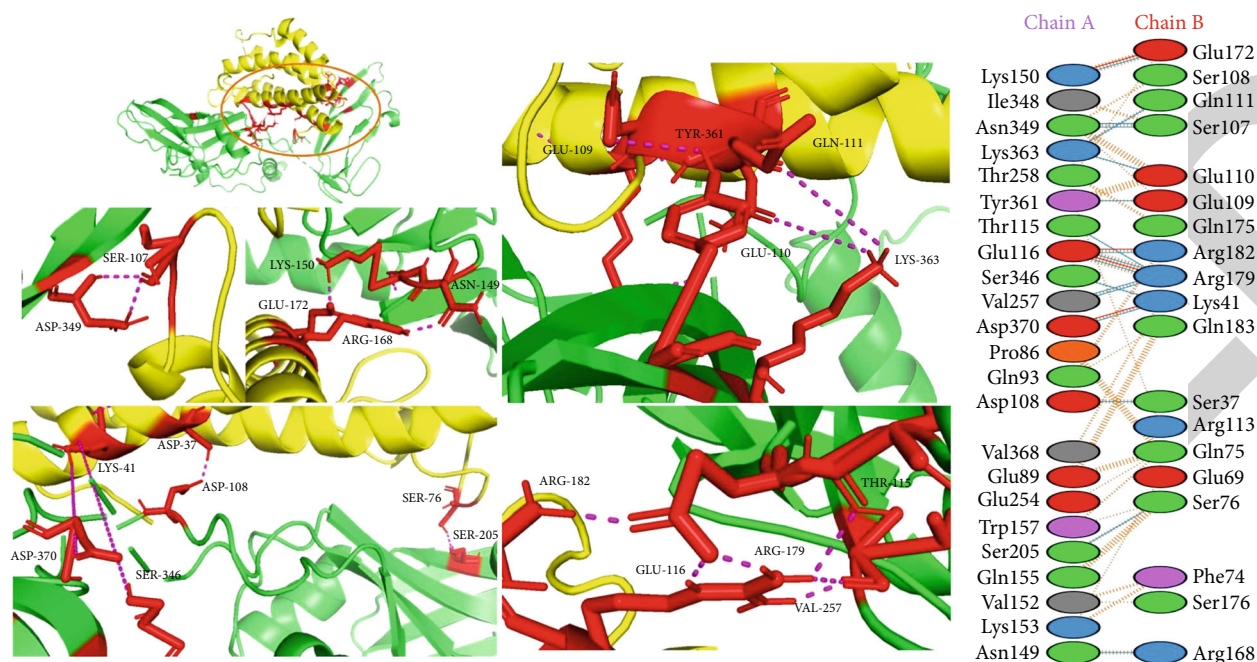


FIGURE 5: Protein-protein interactions between myostatin and IL-6. (a) GDF8/IL-6 docked complex in cartoon representation is shown at the top. Myostatin is shown in green, and IL-6 is shown in yellow. Interacting residues of the GDF8/IL-6 complex are shown in red color stick representation. (b) All interacting residues of GDF8 and IL-6. Blue color: H-bonds, red color: salt bridges, yellow color: disulfide bonds, and orange color: nonbonded contacts. Hydrogen bonds are shown with blue lines. The colors of interacting residues are representing properties of amino acids (i.e., blue: positive, red: negative, green: aliphatic, grey: aliphatic, pink: aromatic, orange: Pro/Gly, and yellow: Cys).

In the next approach, we used protein-protein docking via the HADDOCK server to check the binding interactions of GDF8 with TNF- α and IL-6. The binding score of GDF-TNF- α was found to be -62.2 kcal/mol with 6 hydrogen bonds between active residues, and GDF-IL-6 showed a binding score of -48.1 kcal/mol with 16 hydrogen bonds. Later, all these interactions were confirmed by comparing the results of HADDOCK with those of PUBsum. From results of this study, it can be concluded that both members of TGF- β can be used as therapeutic agents against two main inflammatory cytokines (i.e., TNF- α and IL-6) for the treatment of rheumatoid arthritis. This study can be used as a reference to explore more aspects and involvement of GDF8 and GDF11 in the treatment of autoimmune disorders.

4. Conclusion

From a comprehensive phylogenetic analysis of GDF11, protein-ligand and protein-protein docking, the potential of TGF- β members (GDF11 and GDF8) as drug inhibitors of TNF- α and IL-6 receptor proteins has been revealed. Three tetrapeptides (i.e., SAGP: against IL-6; AFDP and AGPC: against TNF- α) devised from GDF11 with good docking scores fulfilled all the criteria of being good drug candidates. TNF- α and IL-6 are the proinflammatory cytokines that regulate the inflammatory responses and signaling upon infections. In the case of deregulation of these cytokines, the circumstances lead towards the onset of autoimmunity and resistance in self-tolerance. In the current study, different approaches were employed to explore the inhibitory nature

of GDF11 and GDF8 counter to TNF- α and IL-6. The energy validations and scoring of molecular docking results showed the potential of GDF11 and GDF8 as drug candidates with no toxicity.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors would like to gratefully acknowledge the Department of Biochemistry, Government College University Faisalabad, for providing space and facilities to accomplish this study.

Supplementary Materials

Fig S1: multiple sequence alignment of GDF11 gene from selected organisms to show sequence similarities/differences. Fig 2S: interactions (a) and binding patterns (b) of ISMA with TNF- α . Fig 3S: interactions (a) and binding patterns (b) of ANPR with TNF- α . Fig 4S: interactions (a) and binding patterns (b) of AFDP with TNF- α . Fig 5S: interactions (a) and binding patterns (b) of AGPC with TNF- α . Fig 6S: interactions (a) and binding patterns (b) of DGSP with

IL-6. Fig 7S: interactions (a) and binding patterns (b) of TETV with IL-6. Fig 8S: interactions (a) and binding patterns (b) of SAGP with IL-6. Fig 9S: interactions (a) and binding patterns (b) of GSAG with IL-6. (*Supplementary Materials*)

References

- [1] M. F. Bachmann and M. Kopf, "On the role of the innate immunity in autoimmune disease," *The Journal of Experimental Medicine*, vol. 193, no. 12, pp. F47–F50, 2001.
- [2] J. S. Marshall, R. Warrington, W. Watson, and H. L. Kim, "An introduction to immunology and immunopathology," *Allergy, Asthma & Clinical Immunology*, vol. 14, no. S2, p. 49, 2018.
- [3] M. Zouali and A. La Cava, "Editorial: innate immunity pathways in autoimmune diseases," *Frontiers in Immunology*, vol. 10, p. 1245, 2019.
- [4] C. Castro and M. Gourley, "Diagnosis and treatment of inflammatory myopathy: issues and management," *Therapeutic advances in musculoskeletal disease*, vol. 4, no. 2, pp. 111–120, 2012.
- [5] G. George, G. Shyni, and K. Raghu, "Current and novel therapeutic targets in the treatment of rheumatoid arthritis," *Inflammopharmacology*, vol. 28, no. 6, pp. 1457–1476, 2020.
- [6] C. A. Innis, J. Shi, and T. L. Blundell, "Evolutionary trace analysis of TGF- β and related growth factors: implications for site-directed mutagenesis," *Protein Engineering*, vol. 13, no. 12, pp. 839–847, 2000.
- [7] G. Mustafa, M. J. Iqbal, M. Hassan, and A. Jamil, "Bioinformatics characterization of growth differentiation factor 11 of *Oryctolagus cuniculus*," *Journal of the Chemical Society of Pakistan*, vol. 39, pp. 1089–1095, 2017.
- [8] L. W. Gamer, K. A. Cox, C. Small, and V. Rosen, "Gdf11 is a negative regulator of chondrogenesis and myogenesis in the developing chick limb," *Developmental Biology*, vol. 229, no. 2, pp. 407–420, 2001.
- [9] F. Xing, X. Tan, P.-J. Zhang et al., "Characterization of amphioxus GDF8/11 gene, an archetype of vertebrate MSTN and GDF11," *Development Genes and Evolution*, vol. 217, no. 7, pp. 549–554, 2007.
- [10] T. Poggioli, A. Vujic, P. Yang et al., "Circulating growth differentiation factor 11/8 levels decline with age," *Circulation Research*, vol. 118, no. 1, pp. 29–37, 2016.
- [11] W. Li, W. Wang, L. Liu et al., "GDF11 antagonizes TNF- α -induced inflammation and protects against the development of inflammatory arthritis in mice," *The FASEB Journal*, vol. 33, no. 3, pp. 3317–3329, 2019.
- [12] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.
- [13] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets," *Molecular Biology and Evolution*, vol. 33, no. 7, pp. 1870–1874, 2016.
- [14] T. L. Bailey, M. Boden, F. A. Buske et al., "MEME SUITE: tools for motif discovery and searching," *Nucleic acids research*, vol. 37, no. Web Server, pp. W202–W208, 2009.
- [15] A. Mushtaq, G. Mustafa, T. M. Ansari, M. A. Shad, J. Cruz-Reyes, and A. Jamil, "Antiviral activity of hexapeptides derived from conserved regions of bacterial proteases against HCV NS3 protease," *Pakistan Journal of Pharmaceutical Sciences*, vol. 34, no. 1, pp. 215–223, 2021.
- [16] A. Daina, O. Michielin, and V. Zoete, "SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules," *Scientific Reports*, vol. 7, no. 1, article 42717, 2017.
- [17] H. Yang, C. Lou, L. Sun et al., "admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties," *Bioinformatics*, vol. 35, no. 6, pp. 1067–1069, 2019.
- [18] C. Dominguez, R. Boelens, and A. M. Bonvin, "HADDOCK: a protein–protein docking approach based on biochemical or biophysical information," *Journal of the American Chemical Society*, vol. 125, no. 7, pp. 1731–1737, 2003.
- [19] W. L. DeLano, "Pymol: an open-source molecular graphics tool," *CCP4 Newsletter on protein crystallography*, vol. 40, no. 1, pp. 82–92, 2002.
- [20] R. A. Laskowski, "PDBsum new things," *Nucleic Acids Research*, vol. 37, no. Database, pp. D355–D359, 2009.
- [21] M. Kearse, R. Moir, A. Wilson et al., "Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data," *Bioinformatics*, vol. 28, no. 12, pp. 1647–1649, 2012.
- [22] N. Saitou and M. Nei, "The neighbor-joining method: a new method for reconstructing phylogenetic trees," *Molecular Biology and Evolution*, vol. 4, no. 4, pp. 406–425, 1987.
- [23] G. Afzal, G. Mustafa, S. Mushtaq, and A. Jamil, "DNA barcodes of Southeast Asian spiders of wheat agro-ecosystem," *Pakistan Journal of Zoology*, vol. 52, no. 4, pp. 1433–1441, 2020.
- [24] J. Felsenstein, "Confidence limits on phylogenies: an approach using the bootstrap," *Evolution*, vol. 39, no. 4, pp. 783–791, 1985.
- [25] A. Mushtaq, T. M. Ansari, G. Mustafa, M. A. Shad, J. Cruz-Reyes, and A. Jamil, "Isolation and characterization of nprB, a novel protease from *Streptomyces thermovulgaris*," *Pakistan Journal of Pharmaceutical Sciences*, vol. 33, no. 5, pp. 2361–2369, 2020.
- [26] M. Kimura, "A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences," *Journal of Molecular Evolution*, vol. 16, no. 2, pp. 111–120, 1980.
- [27] P. R. Biga, S. B. Roberts, D. B. Iliev et al., "The isolation, characterization, and expression of a novel GDF11 gene and a second myostatin form in zebrafish, *Danio rerio*," *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, vol. 141, no. 2, pp. 218–230, 2005.
- [28] K. S. Kim, Y. J. Kim, J. M. Jeon et al., "Molecular characterization of myostatin-like genes expressed highly in the muscle tissue from Morotoge shrimp, *Pandalopsis japonica*," *Aquaculture Research*, vol. 41, no. 11, pp. e862–e871, 2010.
- [29] R. G. Walker, T. Poggioli, L. Katsimpardi et al., "Biochemistry and biology of GDF11 and myostatin: similarities, differences, and questions for future investigation," *Circulation Research*, vol. 118, no. 7, pp. 1125–1142, 2016.
- [30] A. C. McPherron, T. V. Huynh, and S.-J. Lee, "Redundancy of myostatin and growth/differentiation factor 11 function," *BMC Developmental Biology*, vol. 9, no. 1, p. 24, 2009.
- [31] M. Nakashima, T. Toyono, A. Akamine, and A. Joyner, "Expression of growth/differentiation factor 11, a new member of the BMP/TGF β superfamily during mouse embryogenesis," *Mechanisms of Development*, vol. 80, no. 2, pp. 185–189, 1999.