

RELATIONSHIP OF CD4 PROTEINS ACROSS FISHES AND OTHER VERTEBRATES THROUGH PHYLOGENETIC TREE ANALYSIS AND IMMUNOGLOBULIN DOMAIN PREDICTION

MUHAMAD DAFA FAUZI



DEPARTMENT OF BIOLOGY
FACULTY OF MATHEMATICS AND NATURAL SCIENCE
INSTITUT PERTANIAN BOGOR
BOGOR
2026

@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



STATEMENT REGARDING UNDERGRADUATE THESIS AND INFORMATION SOURCE AND COPYRIGHT TRANSFER

By this writing, I clarify that the undergraduate thesis “Relationship of CD4 Proteins Across Fishes and Other Vertebrates Through Phylogenetic Tree Analysis and Immunoglobulin Domain Prediction” is my own work under the supervision of the advising committee and has not been proposed for any institution. The copied information sources of published and unpublished writings of authors have been mentioned in the text and incorporated in the references in the last part of the thesis.

By this writing, I hand over the copyright of my thesis to IPB University.

Bogor, July 2026

Muhamad Dafa Fauzi
G3401221061



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRAK

MUHAMAD DAFA FAUZI. Hubungan Protein CD4 pada Ikan dan Vertebrata Lainnya Melalui Analisis Pohon Filogenetik dan Prediksi Domain Immunoglobulin. Dibimbing oleh TRI HERU WIDARTO, KANTHI ARUM WIDAYATI, dan WORAPONG SINGCHAT.

Sistem imun adaptif vertebrata menghasilkan respons spesifik terhadap patogen dan membentuk memori imunologis. Sistem ini melibatkan glikoprotein yang disebut *cluster of differentiation 4* (CD4), yang berperan dalam membantu sel T mengatur respons imun dan mengenali antigen dari *major histocompatibility complex* kelas II (MHC-II). Penelitian ini mengeksplorasi hubungan pada tingkat protein di antara 254 sekuen asam amino homolog CD4 pada vertebrata, yang terdiri atas 77 sekuen CD4 tetrapoda, 104 sekuen CD4-1 ikan, dan 73 sekuen CD4-2 ikan, dengan menganalisis hubungan filogenetik menggunakan metode Neighbour-Joining serta memprediksi struktur domain immunoglobulin (domain IG) menggunakan analisis SMART. Penelitian ini juga menggunakan protein LAG3 sebagai homolog eksternal sebagai akar pohon filogenetik. Hasil analisis filogenetik menunjukkan bahwa kebanyakan sekuen CD4 tetrapoda membentuk kelompok saudara (*sister group*) dengan kebanyakan sekuen CD4-1 dan CD4-2 ikan. Baik CD4-1 maupun CD4-2 ikan menunjukkan empat subklaster. Pada CD4-1, subklaster tersebut terdiri dari coelacanth, teleost basal, otophysa, dan euteleostei, sedangkan pada CD4-2, subklaster tersebut terdiri dari hiu, teleost basal, otophysa, dan euteleostei. Pola domain IG yang diprediksi pada CD4-1 ikan lebih menyerupai pola pada CD4 tetrapoda, yaitu memiliki dua hingga empat domain IG, dibandingkan dengan CD4-2 ikan yang diprediksi memiliki dua hingga enam domain IG. Temuan ini mendukung studi imunologi vertebrata.

Kata kunci: duplikasi protein, ikan teleost, imunitas adaptif



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRACT

MUHAMAD DAFA FAUZI. Relationship of CD4 Proteins Across Fishes and Other Vertebrates Through Phylogenetic Tree Analysis and Immunoglobulin Domain Prediction. Supervised by TRI HERU WIDARTO, KANTHI ARUM WIDAYATI, and WORAPONG SINGCHAT.

The adaptive immune system in vertebrates produces specific responses against pathogens and forms immunological memory. In this system, a specific glycoprotein named the cluster of differentiation 4 (CD4) supports T cells in regulating immune responses and recognizing antigens from the major histocompatibility complex class II (MHC-II). This study explored the protein-based relationships among 254 amino acid sequences of CD4 homologs across vertebrates, including 77 tetrapod CD4 sequences, 104 fish CD4-1 sequences, and 73 fish CD4-2 sequences, by analyzing the Neighbour-Joining phylogenetic relationships and predicting immunoglobulin domain (IG domain) structure using SMART analysis. Additionally, this study uses LAG3 protein as an external homolog as an outgroup. The phylogeny shows that most tetrapod CD4 sequences form a sister group to most fish CD4-1 and CD4-2 sequences. Both fish CD4-1 and CD4-2 show four sub-clusters. The Cd4-1 sub-clusters are represented by coelacanth, basal teleost, otophysa, and euteleostei, whereas CD4-2 sub-clusters are represented by shark, basal teleost, otophysa, and euteleostei. The predicted IG domain pattern of fish CD4-1 is more similar to the tetrapod CD4, which shows two to four predicted IG domains, than to that of fish CD4-2, which shows two to six predicted IG domains. These findings support vertebrate immunology studies.

Keywords: adaptive immunity, duplication protein, teleost fish



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Copyright belongs to IPB, 2026
All Rights Reserved

*Prohibited to cite part or all of this paper without citing or mention the source.
Citation is only for the education importance, research, writing scientific paper,
report writing, critical writing, or review a problem, and the citation is not harm
IPB importance.*

*Prohibited to announce and multiply part or all of this paper in any form
without IPB permissions.*

RELATIONSHIP OF CD4 PROTEINS ACROSS FISHES AND OTHER VERTEBRATES THROUGH PHYLOGENETIC TREE ANALYSIS AND IMMUNOGLOBULIN DOMAIN PREDICTION

MUHAMAD DAFA FAUZI

Undergraduate Thesis Intended to Acquire
Bachelor's Degree in Biology Study Program

**DEPARTMENT OF BIOLOGY
FACULTY OF MATHEMATICS AND NATURAL SCIENCE
INSTITUT PERTANIAN BOGOR
BOGOR
2026**

@Hak cipta milik IPB University

IPB University





@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Exeminer Team on the Undergraduate Thesis Examination:

- 1 Prof. Dr. Ir. Miftahudin, M.Si.

Research Title : Relationship of CD4 Proteins Across Fishes and Other Vertebrates
Through Phylogenetic Tree Analysis and Immunoglobulin Domain
Prediction

Name : Muhamad Dafa Fauzi

Student ID : G3401221061

Approved by

Supervisor 1:

Ir. Tri Heru Widarto, M.Sc.

Supervisor 2:

Dr. Kanthi Arum Widayati, S.Si., M.Si.

Supervisor 3:

Dr. Worapong Singchat



Acknowledge by

Head of The Study Program:

Dr. Ir. Iman Rusmana, M.Si

NIP 19650720 199103 1002

Examination Date:
August 06 2026

Passdate:



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



FOREWORDS

Praise and gratitude are offered to Allah SWT for all His blessings and grace, which have enabled the author to complete this scientific work entitled “Relationship of CD4 Protein Across Fishes and Other Vertebrates Through Phylogenetic Tree Analysis and Immunoglobulin Domain Prediction”. This research was conducted from August 2025 to December 2025. The preparation and writing of this thesis would not have been possible without the help, support, and prayers of many parties. Therefore, the author would like to express sincere gratitude and appreciation to:

1. Ir. Tri Heru Widarto, M.Sc., as the first supervisor, Dr. Kanthi Arum Widayati, S.Si., M.Si., as the second supervisor, and Dr. Worapong Singchat, as the third supervisor, for all the directions, suggestions, and guidance that were very meaningful during the process of compiling this undergraduate thesis.
2. Kornorn Srikulnath, Ph.D., as the team leader in the Animal Genomics and Bioresource Research Center (AGB), for the guidance, support, and motivation provided to the author.
3. Academic advisors, seminar moderators, and all educational staff in the Department of Biology for their assistance and service.
4. The examiners who have provided input, criticism, and constructive corrections to improve the quality of this thesis.
5. The author’s beloved family: Mr. Mulyono, Mrs. Kusmiati, Rizky Agustina, Nisrina Nuraini, Biandari Bunga Nafeeza, Agus Sugiat Tirta Setia, Muhammad Abyan Firjatullah, Gazala Butsaina Ulfah, and Sorrae for their endless love, prayers, and support.
6. All friends from Cheezone, AGB Lab, FH2 Boarding House, Kopken, Biology IPB 59 (Orchinus Aegis), Project Team 9 Biology, Abiotic 2023, Bioworld HIMABIO IPB 2024 and 2025, EnL KSE IPB 2024 and 2025, Krissana dormitory, Sesaat, and many others who have shared their support, laughter, and precious memories during the author’s study period.
7. All individuals and parties who have contributed to the completion of this research.

The author hopes that this scientific work will benefit those in need and contribute to the advancement of science, especially in the field of biology.

Bogor, July 2026

Muhamad Dafa Fauzi



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



TABLE OF CONTENTS

LIST OF FIGURES	xiii
LIST OF APPENDICES	xiii
I INTRODUCTION	1
1.1 Backgrounds	1
1.2 Objective	1
II METHOD	2
2.1 Time and Place	2
2.2 Materials and Equipment	2
2.3 Methodology	2
2.3.1 Data Collection of CD4, CD4-1 and CD4-2	2
2.3.2 Analysis Phyllogenetic of CD4, CD4-1 and CD4-2	2
2.3.3 Analysis of Predicted IG Domain Structures in CD4, CD4-1, and CD4-2	3
III RESULT AND DISCUSSION	4
3.1 Result	4
3.1.1 Copy Number Variations of the CD4, CD4-1, CD4-2 and LAG3 Protein	4
3.1.2 Phylogenetic Analysis of CD4, CD4-1, and CD4-2	4
3.1.3 Phylogenetic Analysis of CD4-1	5
3.1.4 Phylogenetic Analysis of CD4-2	6
3.1.5 Predicted IG Domain of CD4, CD4-1 and CD4-2 of fish and other vertebrates	7
3.2 Discussion	9
IV CONCLUSION AND RECOMMENDATION	12
4.1 Conclusion	12
4.2 Recommendation	12
REFERENCES	13
APPENDICES	15
CURRICULUM VITAE	56



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University

LIST OF FIGURES

- 1 Neighbor-joining tree of the 261 amino-acid sequences of CD4, CD4-1, and CD4-2 proteins. This tree included 77 species of tetrapods for CD4 (yellow), 104 species of fish for CD4-1 (green), 73 species of fish for CD4-2 (purple), and 7 amino acid sequences of LAG3 (black) as outgroups. Arrow: bootstrap value; solid grey square: aves; empty diamond: reptiles; empty triangle: mammals; empty circle: amphibian; solid black square: coelacanth (CD4-1); solid black triangle: shark (CD4-2); solid black diamond: *Hippoglossus* (CD4-2). The alignment use ClustalW in MEGA12, then trimmed with smart-gap (default) in ClipKIT. This tree was constructed using MEGA12 and beautified with iTOL and Canva. One LAG3 sequence from *Oryzias latipes* was used to re-root the tree. 5
- 2 Phylogenetic tree of the CD4-1 protein based on 111 amino acid sequences. This tree includes 104 sequences of fish CD4-1 and 7 sequences of LAG3 (black branch) as an outgroup. The cluster was represented by different colors, such as outgroup (black), Coelacanth (red), basal teleost (cream), otophysa (green), and euteleostei (purple). The arrow represented the bootstrap value. The alignment uses ClustalW in MEGA12, then trimmed with smart-gap (default) in ClipKIT. This tree was constructed using MEGA12 and beautified with iTOL and Canva. One LAG3 sequence from *Oryzias latipes* was used to re-root the tree. 6
- 3 Phylogenetic tree of CD4-2 protein based on 80 amino acid sequences. This tree uses 73 sequences of fish for CD4-2 and 7 sequences of LAG3 (black branch) as an outgroup. The cluster was represented by different colors, such as outgroup (black), shark (blue), basal teleost (cream), otophysa (green), and euteleostei (purple). The solid diamond represented the euteleostei species, which separated from that cluster (*Hippoglossus hippoglossus* and *Hippoglossus stennolepis*). The arrow represented the bootstrap value. The alignment uses ClustalW in MEGA12, then trimmed with smart-gap (default) in ClipKIT. This tree was constructed using MEGA12 and beautified with iTOL and Canva. One LAG3 sequence from *Oryzias latipes* was used to re-root the tree. 7
- 4 Comparison of the SMART-predicted IG domains in 77 CD4 sequences from tetrapods (A), 104 CD4-1 sequences from fish (B), and 73 CD4-2 sequences from fish (C). The silhouette represented the cluster, followed by the number of species. The green square represented the predicted IG domain in each cluster. 8

LIST OF APPENDICES

- 1 Information of LAG3 protein 16
- 2 Phylogenetic of CD4, CD4-1, and CD4-2 before coloring 17
- 3 Phylogenetic tree of CD4-1 before coloring 18
- 4 Phylogenetic tree of CD4-2 before coloring 19

5	Sequence information of CD4, CD4-1, and CD4-2 protein. Acc No: Accession Number; Seq Mod: Sequence Model; Chr: Chromosome; *: species that only have one copy (only CD4-1 or CD4-2).	20
6	Completeness, Assembly version, and access date of CD4, CD4-1, and CD4-2. *: species that only have one copy (only CD4-1 or CD4-2).	34
7	Sequence length and SMART analysis result of CD4, CD4-1, and CD4-2. SP: Signal Peptide; IG: Immunoglobulin Domain; TR: Transmembran Region; *: species that only have one copy (only CD4-1 or CD4-2); **: catfish species (Siluriformes).	46