Review

The Evolution of Rag Gene Enhancers and Transcription Factor E and Id Proteins in the Adaptive Immune System

Genki Yoshikawa [1], Kazuko Miyazaki [2], Hiroyuki Ogata [1],* and Masaki Miyazaki [2],*

1   Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji 611-0011, Japan; yoshikawagenki@gmail.com
2   Laboratory of Immunology, Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto 606-8507, Japan; kamiyazaki@infront.kyoto-u.ac.jp
*   Correspondence: ogata@kuicr.kyoto-u.ac.jp (H.O.); mmiyazaki@infront.kyoto-u.ac.jp (M.M.)

Abstract: Adaptive immunity relies on the V(D)J DNA recombination of immunoglobulin (*Ig*) and T cell receptor (*TCR*) genes, which enables the recognition of highly diverse antigens and the elicitation of antigen-specific immune responses. This process is mediated by recombination-activating gene (Rag) 1 and Rag2 (Rag1/2), whose expression is strictly controlled in a cell type-specific manner; the expression of *Rag1/2* genes represents a hallmark of lymphoid lineage commitment. Although *Rag* genes are known to be evolutionarily conserved among jawed vertebrates, how *Rag* genes are regulated by lineage-specific transcription factors (TFs) and how their regulatory system evolved among vertebrates have not been fully elucidated. Here, we reviewed the current body of knowledge concerning the cis-regulatory elements (CREs) of *Rag* genes and the evolution of the basic helix-loop-helix TF E protein regulating *Rag* gene CREs, as well as the evolution of the antagonist of this protein, the Id protein. This may help to understand how the adaptive immune system develops along with the evolution of responsible TFs and enhancers.

1. Introduction

Our body is protected from invading pathogens by immune responses, which are primarily mediated by two distinct types of cells, the adaptive and innate immune cells. These cells cooperatively function to induce inflammatory responses to eliminate pathogens from the body. Adaptive immune cells, such as T and B cells, elicit pathogen-specific immune responses through the recognition of specific antigens, while innate immune cells, including macrophages, neutrophils, dendritic cells, and histiocytes, are activated by pattern recognition receptors (PRRs), which recognize distinct microbial components. The adaptive immune system (AIS) relies on the assembly of T cell receptor (*TCR*) and immunoglobulin (*Ig*) genes from arrays of variable (V), diversity (D), and joining (J) gene segments. These assembled antigen receptors are able to recognize highly diverse antigens and elicit antigen-specific immune responses [1,2]. This V(D)J recombination of *TCR* and *Ig* genes is mediated by recombination-activating gene 1 (Rag1) and Rag2 protein complex, and Rag1/Rag2 (Rag1/2) complex recognizes and cleaves recombination signal sequences (RSSs) flanking the TCR and Ig V, D, and J gene segments [3]. Rag1 primarily binds and cleaves DNA, while Rag2 enhances Rag1 binding and is a vital co-factor for DNA cleavage [3]. Because the *Rag1/2* genes are exclusively expressed in T cell and B cell progenitor/precursor stages, their expression implies adaptive lymphoid lineage commitment [4]. *Rag* genes are known to be evolutionarily conserved among jawed vertebrates. The current model suggests that *Rag1/2* genes evolved from the ancestral RAG transposase genes [5,6]. The discovery of ProtoRag in the cephalochordate amphioxus strongly support this model [7], and a recent report presenting cryo-electron microscopy structures of RAG and ProtoRAG transposase

shows the mechanism underlying the properties of RAGs with appropriate DNA cleavage and transposition activities [8]. They identified two adaptations specific to jawed vertebrates in RAG1 and RAG2 that suppress RAG-mediated transposition [8].

After T cell lineage commitment in the thymus, TCRβ V(D)J rearrangement is initiated at immature CD4⁻CD8⁻ (double negative; DN) cells (T progenitor cell; pro-T cell), and DN cells are further divided into DN1-4 populations. Following the selection of TCRβ in DN3a cells, DN3a cells start proliferation and differentiate into DN3b-DN4 stage and further CD4⁺CD8⁺ (double positive; DP) cells (T precursor cell; pre-T), in which TCRα VJ rearrangement is performed [9]. As well as T cell, upon B cell lineage commitment from common lymphoid progenitors (CLPs) in the bone marrow, V(D)J recombination of Ig heavy chain (*Igh*) and light chain (*Igκ* and *λ*) occurs in B cell progenitor (pro-B) and precursor (pre-B) cells, respectively [10,11]. CLPs give rise to adaptive lymphoid cells (T and B cell), innate lymphoid cells (ILCs), and plasmacytoid dendritic cells (pDCs). T and B cell lineage commitments are instructed by a set of lineage-specific transcription factors' (TFs') expression: E2A, Ebf1, Foxo1, and Pax5 for B cell and E2A/HEB, Gata3, Tcf1, Bcl11b, Runx, Ikaros, and Pu.1 for T cell [12–16]. Notably, ILCs and T cells show functional similarities in cytokine production, and they commonly express Bcl11b, Tcf1, Gata3, and Runx during their development and activation [17]. What TFs drive adaptive lymphoid lineages? Importantly, E2A and HEB act in synergy to establish T cell identity. In addition, they simultaneously suppress the aberrant ILC development in thymus [18]. Similarly, Ebf1 and Pax5 determine the B cell lineage by repressing genes leading to the T cell and ILC lineages [19]. These adaptive lymphoid lineage-specific TFs were also expected to regulate *Rag1/2* gene expression to drive the difference between adaptive and innate immune cells. Lineage-specific TFs alter gene expression patterns by binding to specific DNA sequences within cis-regulatory elements (CREs), and TF bindings modulate the enhancer–promoter interactions. Many previous studies have attempted to identify the CREs associated with *Rag1/2* gene expression [4,14]. Although both developing T and B cells express the *Rag1/2* genes, the *Rag* gene enhancers differ in T and B cells. In T cells, an anti-silencer element (ASE), which is located 73 kb upstream of the *Rag2* gene and is 8 kb in length, is essential for *Rag1/2* gene expression in developing T cells, but not in developing B cells [20]. Recently, a T cell-specific *Rag* enhancer region (*R-TEn*; 1 kb) was identified within the ASE region, which is regulated by E2A [21]. In contrast, two B cell-specific enhancers (*R1B* and *R2B* (overlapping with *Erag*)) are critically required for *Rag1/2* expression in pro-B cells, and the deletion of both regions results in developmental arrest at the pro-B stage (Figure 1) [21,22]. These results raised questions regarding what lineage-specific TFs regulate these CREs to determine the adaptive lymphoid lineage and how these TFs and CREs have evolutionarily developed during the evolution of AIS. In this report, we reviewed the role of adaptive lymphocyte-specific TFs, especially the transcriptional balance between the E and Id proteins (E-Id axis) [23], and the mechanism by which E2A regulates *Rag1/2* expression, and the evolution of E-Id proteins and *Rag1/2* enhancers among species.
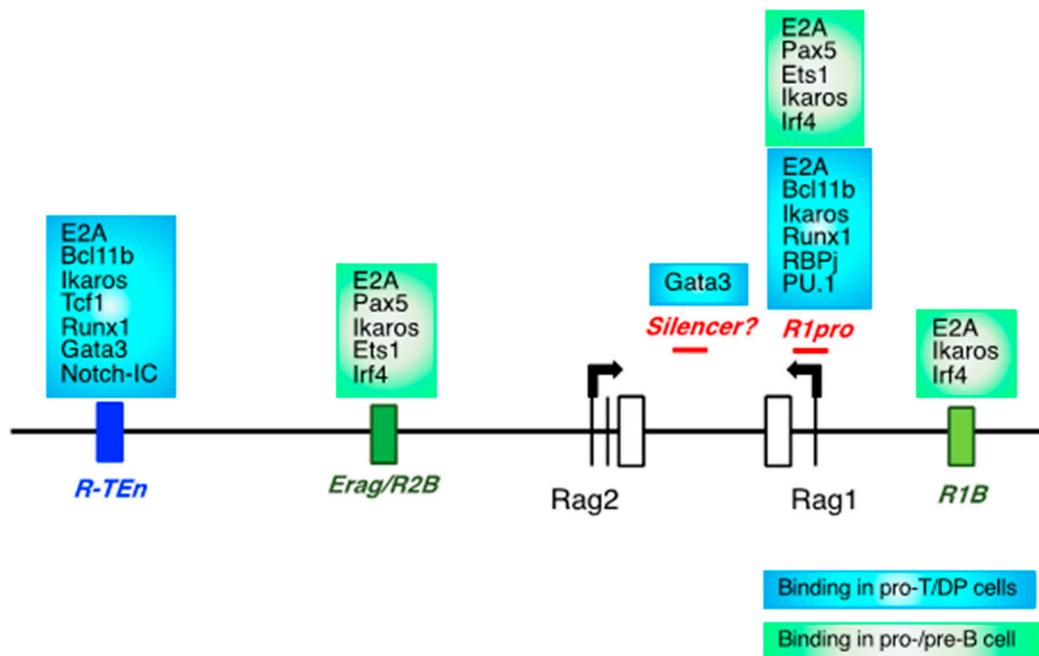
Figure 1. *Rag* gene enhancers in developing T and B cells. A schematic diagram of *Rag* gene locus and enhancers in developing T and B cells. The blue box indicates T cell-specific enhancer (R-*TEn*) and TF binding in pro-T and DP cells. The green boxes indicate B cell-specific enhancers (*R1B* and *R2B*) and TF binding in pro-/pre-B cells. The red lines indicate the *Rag1*-promoter region (*R1pro*) and silencer (*Silencer*) of *Rag* gene, respectively.

2. Regulation of *Rag1/2* Gene by T or B Cell-Specific Enhancers

2.1. CREs for the Rag Gene, and Lineage-Specific Transcription Factors

Rag1/2 gene expression is stringently controlled in a cell type-specific manner. The first wave of RAG expression is required for the recombination of *TCRβ* and *Igh* genes in pro-T and pro-B cells, respectively. After the β-selection of pro-T cells and pre-BCR selection of pro-B cells, *Rag* expression is transiently downregulated during the developmental transition toward precursor stages (DP and pre-B cells). In precursor cells, *Rag1/2* genes are upregulated for the VJ recombination of *TCRα*, *Igκ*, and *λ* genes. Following successful *TCR* and *Ig* gene recombination, *Rag1/2* gene expression is completely suppressed in mature T and B cells [10,14,24,25]. Both in mouse and human, impairment or loss of *Rag* gene expression and functions results in severe combined immunodeficiency, resulting from developmental arrest at pro-T and pro-B cell stages [26,27]. Furthermore, persistent *Rag* expression led to profound immunodeficiency in mouse [28]. Thus, *Rag1/2* genes are stringently regulated during T and B lymphopoiesis. There have been in vivo and in vitro studies that attempted to identify the CREs associated with *Rag1/2* expression [4]. Deletion of *Erag*, which is 23 kb upstream of the *Rag2* gene, caused impaired *Rag1/2* expression in pro-B cells and a moderate developmental block at the pro-B stage but did not affect the *Rag* gene expression in T cell development [22]. Interestingly, deletion of two B cell-specific enhancers, *R1B* and *R2B* (partially overlapping with *Erag*), resulted in a developmental arrest at the pro-B stage, indicating the enhancer redundancy of *Rag* gene in B cells [21]. On the other hand, the T cell-specific *Rag* gene enhancer R-*TEn*, which is included in ASE, is critically important for *Rag1/2* expression and TCR recombination during thymocyte development, and its deletion led to a developmental block at the DN3 cells and DP cells in fetal and adult thymus without affecting B cell development [20,21,29]. Taken together, *Rag* gene expression is tightly regulated in a cell type-specific manner, and T and B cells use distinct enhancer regions for *Rag* gene expression.

Previous reports and ChIP-seq showed that many T cell- or B cell-specific TFs bind to these enhancer regions (T cells: Tcf1, Bcl11b, Gata3, Runx1, Satb1, and Ikaros; B cells: Pax5, Ebf1, Foxo1, Ets1, Irf4, and Ikaros) [21,29–31]. What TFs regulate these cell type-specific

enhancers for *Rag* genes? Notably, mutating the E-box motifs in the *R-TEn* enhancer (*R-TEn-E-box-mutant*) in mice, in which the binding of E2A is blocked, resulted in an impaired *Rag1/2* expression and blocked thymocyte development, as well as deletion of the entire *R-TEn* region. Of note, the *R-TEn-E-box-mutant* abolished chromatin accessibility throughout the entire *Rag* gene locus. These results indicated that E-protein binding to the T cell-specific *Rag* gene enhancer is required for T cell-specific spatial interactions to enhance *Rag1/2* expression [14,21]. Notably, blocking E2A binding to the *Rag1* gene promoter region (*R1pro*) by generating E-box motif mutations alone resulted in the complete loss of *Rag1* expression without affecting *Rag2* expression in both developing T and B cells, leading to developmental arrest at the pro-T and pro-B cell stages [21]. Taken together, these results strongly suggest that the activities of T cell-specific enhancer and *Rag1* promoter depend on the binding of E2A to these regions and that E2A is a core TF that specifies the adaptive lymphoid cell identity through the regulation of *Rag* gene expression.

2.2. Evolution of Rag Gene Enhancer

Enhancer regions play a crucial role in precise pattern and amounts of gene expression during development, and divergence of the DNA sequence within enhancer region is considered to be related to the phenotypic variations among species [32]. This suggests that the phylogenetic conservation of DNA sequences within *Rag* gene enhancers reflect the evolution of *Rag* gene regulation. Although *Rag1/2* genes are well known to be conserved among jawed vertebrates, the conservation of *Rag* gene enhancers had not been investigated. Thus, we investigated the conservation of *R-TEn*, *R1B,* and *R2B* regions and E-box motifs in these regions [21]. We found that DNA sequence similarities in *R-TEn* and *R2B* are readily observed among mammals, most birds, and reptiles; however, sequence similarities of these enhancers are not noticeable in the corresponding genomic regions of amphibians and fishes (Figure 2) [21]. Furthermore, we observed significant conservation of E-box motifs in conserved *R-TEn*, *R2B,* and *R1B* regions (Figure 2) [21]. These results show the discordance in the conservation of *Rag* genes and their enhancers among jawed vertebrates and the possibility of divergent cis-regulatory modules of *Rag* genes in terrestrial animals, aquatic animals, and amphibians. Thus, we proposed that terrestrial animals evolutionarily acquired the E protein-mediated regulatory mechanisms as enhancers to increase the *Rag* gene expression, which induce higher expression of *Rag genes* and enable a diverse range of *TCR* and *Ig* gene recombination to protect our bodies from a wide range of pathogens.

Regarding the evolution of AIS among vertebrates, cytidine deaminases CDA1 and CDA2 in jawless vertebrates are counterparts of *Rag1* and *Rag2* in jawed vertebrates and evolutionarily developed AIS as genome editors [33–35]. Furthermore, the recombination of *Ig* and *TCR* in fish seems to be more diverse than that in mammals, for example, the plasticity of T/B cells and the repertoire usage of *TCR* and *Ig* [36]. Given that the locations of B cell development among birds, reptiles, amphibians, and fish are different, it is reasonable that the variation in enhancer regions among species produces diversification of *Rag1/2* gene regulation, such as timing. Considering this, it is surprising that both enhancer and promoter activities are critically controlled by E protein binding.
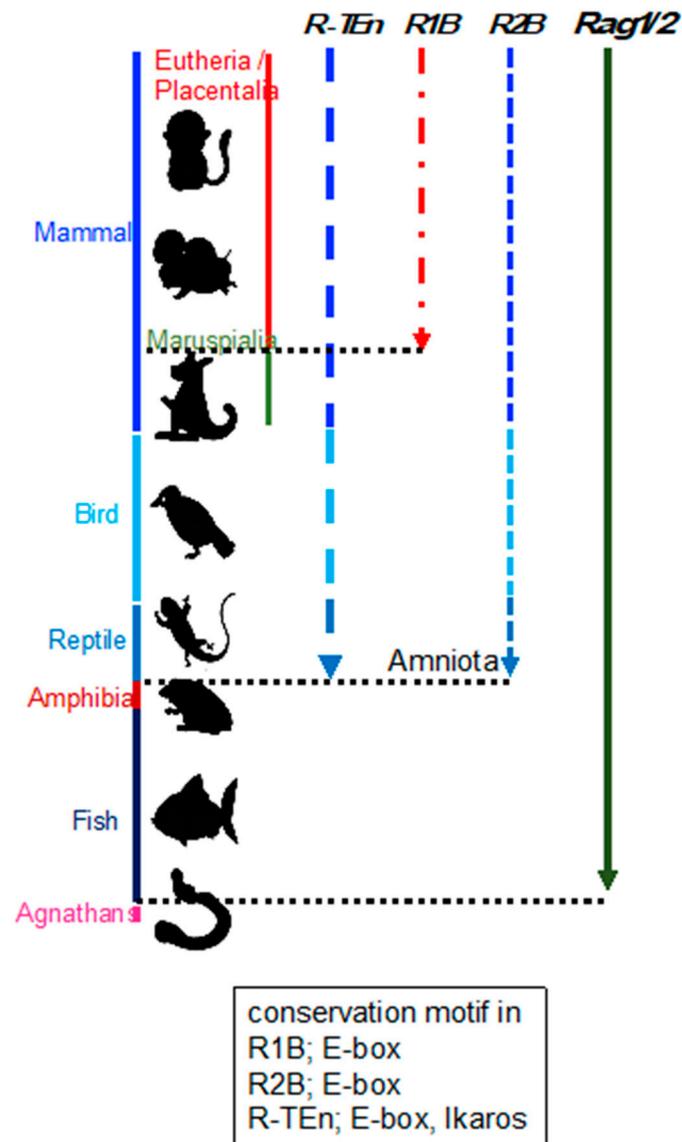
Figure 2. Schematic summary of the conservation of *R-TEn*, *R1B*, and *R2B* among vertebrates. Black, dotted lines indicate the border between placentaria and maruspialia, reptile and amphibia, and fish and agnathans. The conserved motifs in each enhancer region are shown in the box [21].

3. E Proteins and Id Proteins in Adaptive Lymphocyte Development

E proteins are basic helix-loop-helix (bHLH) transcription factors involved in multiple developmental processes. E proteins include E12, E47, E2-2 (TCF4), HEB (TCF12), daughterless (Da), and HLH-2 [23]. E proteins bind as homodimers or heterodimers to the E-box motif (CANNTG) within enhancer regions of their target genes. The mammalian E protein family plays important roles in hematopoiesis. However, the *Drosophila* gene product (da) and *Caenorhabditis elegans* gene product HLH-2 are involved in other developmental pathways. Da is essential for both neurogenesis and sex determination in *D. melanogaster* embryonic development [37]. HLH-2 is required for the development and function of the regulatory cells of the *C. elegans* somatic gonad [38]. Id proteins contain an HLH domain missing the basic region that is essential for specific DNA binding and form heterodimers with bHLH proteins such as E proteins [39]. When the Id protein forms heterodimers with the E protein, the Id protein antagonizes the DNA binding of E proteins and functions as a negative regulator [40]. Id proteins include Id1-4 and the *D. melanogaster* gene product *extramacrochaete* (*emc*) [23].

It is well established that a majority of adaptive lymphocyte development trajectories require regulation by E and Id proteins [23,39,41]. *E2A* (*Tcf3*) is critically required for B cell lineage commitment [42,43] and the E2A gene encodes E12 and E47 proteins, which are generated by differential splicing [44,45]. In lymphoid progenitor cells, E2A orchestrates the B cell fate, along with Ebf1, Foxo1, and other TFs [12,46]. Upon T cell lineage commitment, E2A and HEB act in synergy to establish T cell identity and to suppress ILC development [18]. Likewise, HEB plays a role in iNKT cell development [47], and E2A and HEB also play important roles in the positive selection of DP thymocytes [48]. Interferon-producing plasmacytoid dendritic cell (pDC) development is controlled by E2-2, whereas antigen-presenting classical dendritic cells (cDCs) are orchestrated by Id2 through neutralizing E2-2 activity [49–51].

Id2 and *Id3* are involved in both T and B cell development to modulate E protein DNA binding activity. *Id2* is particularly important for ILC, NK, and LTi cell development through the suppression of E protein activity [13,52]. In B cell development, Id3 is induced in response to TGFβ signaling for survival during early B cell development [53]. Id3 is highly expressed in naïve mature B cells and downregulated in activated germinal center B (GCB) cells, while E2A protein abundance is low in naïve B cells but high in GCB cells to induce AID expression in cooperation with E2-2 [54,55]. In T cell development, *Id3* is first upregulated by pre-TCR signaling in DN3 cells and further upregulated upon positive selection of TCR signaling in DP cells [56,57]. Similarly, γδ TCR-mediated signaling induces high levels of *Id3* abundance and E2A-Id3-Tcf1 transcription axis control γδ T cell development and effector function [58,59]. Furthermore, *Id3* plays a key role in follicular helper T (T_{FH}) and follicular cytotoxic T(T_{FC}) cell development through the regulation of CXCR5 expression [60–62]. Notably, *Id2* and *Id3* have distinct roles in the differentiation of CD8 cytotoxic T cells toward effector and memory T cells [63,64]. In addition, regulatory T (Treg) cells also critically require *Id2* and *Id3* expression to suppress systemic T_{H2} inflammation and function as a gatekeeper for follicular regulatory T (T_{FR}) cell [65]. Taken together, the E–Id protein axis controls the adaptive lymphocyte development and activation to maintain immunological homeostasis.

4. Evolution of E and Id Proteins

In this section, we address the question of how the E–Id axis was evolutionally developed. Emc is a negative feedback regulator that prevents runaway self-stimulation of *Da* gene expression in *Drosophila*. Coupled transcriptional feedback loops maintain the widespread Emc expression that restrains Da activity to induce neurons [66], suggesting that the transcriptional regulation system by E and Id proteins is conserved from the common ancestor of mammals and *Drosophila*.

Three E protein homologs and two Id protein homologs were found in the lamprey (*Petromyzon marinus*) (Figure 3). A reconstructed maximum likelihood phylogenetic tree of E protein homologs indicates that homologs of jawed vertebrates form three clades for E2A, E2-2, and HEB. Lamprey E protein homologs are located outside these three clades of E proteins of jawed vertebrates, although their positions are not well resolved (Figure 3A, Figure S1). A reconstructed maximum likelihood phylogenetic tree for Id protein homologs indicates that homologs of jawed vertebrates form four clades corresponding to Id1 to Id4 (Figure 3B, Figure S1). Lamprey Id protein homologs form a clade with Id2 of jawed vertebrates, although this position is not statistically supported. These multiple clades of E and Id proteins conserved in jawed vertebrates strongly suggest that these paralogs were generated through the widely recognized two rounds of whole genome duplication (WGD) in vertebrates [67]. It is plausible that ancestral jawed vertebrates probably had four paralogs for each of the E and Id proteins, and one of the four E protein paralogs was lost early in evolution prior to the divergence of jawed vertebrates. Recent research proposed that all extant vertebrates share the first duplication, which occurred in the Cambrian, and the second duplication is found only in jawed vertebrates and occurred in the Ordovician [67]. Given the unstable phylogenetic positions of lamprey E and Id

homologs, it is unclear from the current data whether this evolutionary scenario could also explain the existence of paralogs in jawless vertebrates.
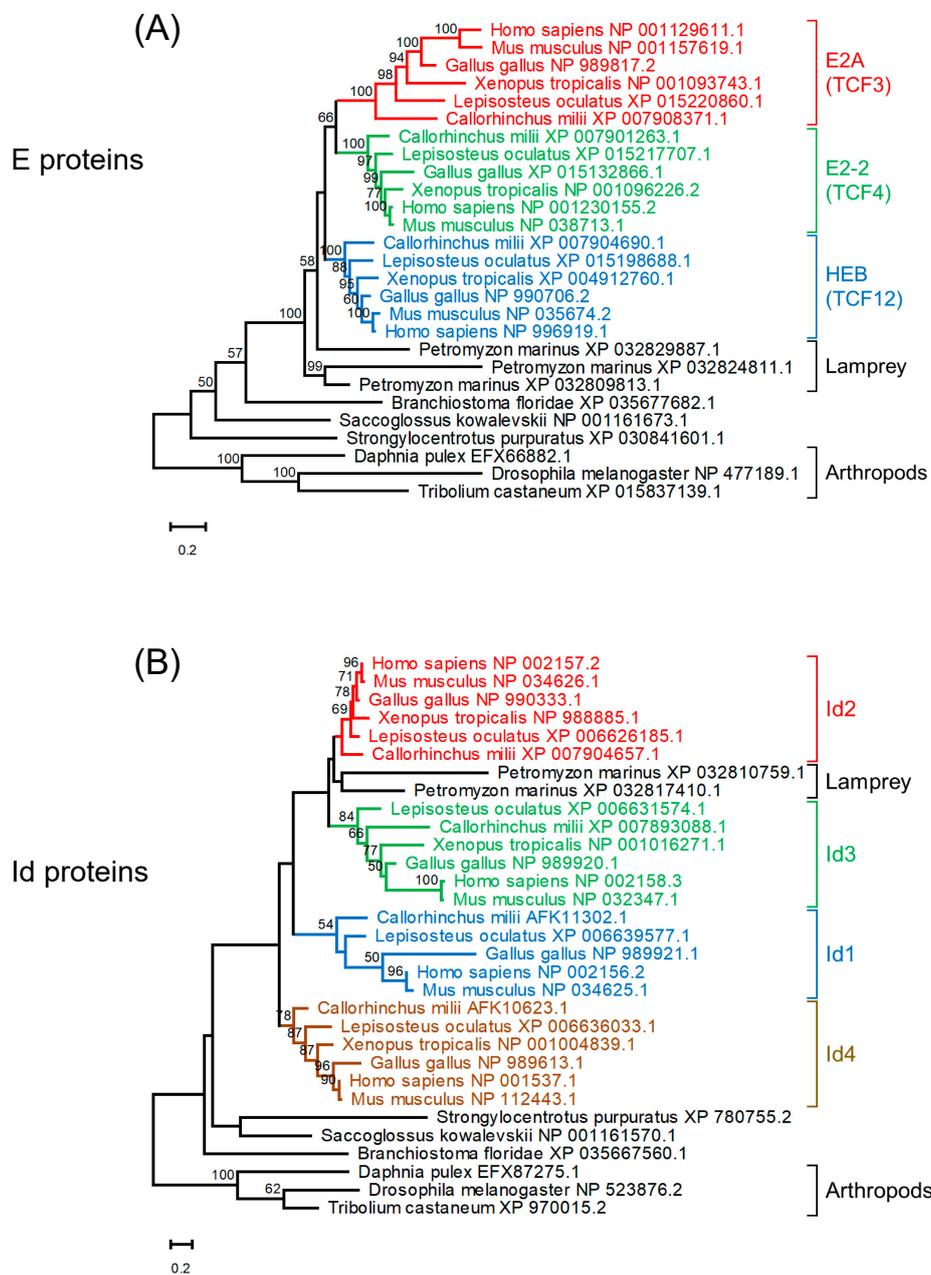

Figure 3. Maximum likelihood phylogenetic trees of homologs of E proteins (**A**) and Id proteins (**B**). Sequences were aligned using MAFFT (v7.453) [68] with default parameters. Tree reconstruction was performed using RAxML (version 8.2.12) [69] with the JTT + F substitution model and PROTGAMMA parameter with 100 bootstrap replicates. Phylogenetic trees were visualized using MEGA-X (version 10.2.4) [70]. Bootstrap values are given along the branches.

5. Discussion

Jawed vertebrates (gnathostomes) possess AIS that can recognize and initiate a protective response against invading pathogens. AIS in jawed vertebrates is centered on T and B lymphocytes bearing TCR and B cell receptors (BCRs, Igs), which are generated through the V(D)J recombination mediated by Rag1 and Rag2 [3]. The TCRs recognize peptide fragments of antigens complexed with molecules encoded by the major histocompatibility complex (MHC) class I and class II genes. Therefore, *Ig*, *TCR*, *Rag1* and *Rag2*,

and *MHC* class I/II genes have an integral role in AIS in jawed vertebrates. Homologs of these genes have been identified in all extant classes of jawed vertebrates [71]. On the other hand, jawless vertebrates (agnathans) possess lymphocyte-like cells (LLCs) that morphologically resemble the T and B cells of jawed vertebrates [72–74]. Many of the genes encoding transcription factors involved in AIS in jawed vertebrates are also encoded in the genome of jawless vertebrates [75]. However, sequence and transcriptome analysis of jawless vertebrates provided no evidence for the presence of *Ig*, *TCR*, *Rag1/2*, or *MHC* genes [72,76,77]. Instead of *TCR* and *Ig*, different types of antigen receptors, which are known as variable lymphocyte receptors (VLRs), composed of highly diverse leucine-rich repeat (LRR) modules, have been identified in lampreys, and, like Rag-mediated recombination of *TCR* and *Ig* in jawed vertebrates, VLRs are generated by cytidine deaminase (CDA) 1 and 2 [33–35,73,78]. Interestingly, CDA2 is closely related to activation-induced cytidine deaminase (AID), which is essential for class-switch recombination (CSR) and somatic hypermutations (SHMs) in human and mice [34,79]. The similarities of multiple components of AIS between jawed and jawless vertebrates suggest that the roots of the system originated in the common ancestor of all vertebrates [80]. Additionally, a primitive AIS emerged already in the common ancestor of all vertebrates, although the possibility of independent acquisitions of the AIS in jawed and jawless vertebrates cannot be excluded. The *Rag1* and *Rag2* genes that are essential for AIS in jawed vertebrates are proposed to have arisen from a transposable element and transmitted vertically through chordate and jawed vertebrate evolution [6,7,81]. This scenario suggested that the acquisition of novel molecular capabilities was a crucial event in the evolution of AIS in jawed vertebrates. Of note, we have recently demonstrated that E2A regulates enhancer/promoter activity of *Rag* gene during T and B cell development [21] and Id2 is essential for innate lymphocytes, including dendritic cells, suggesting that the E–Id axis determines the cell fate of lymphocytes between adaptive and innate immune cells [18,23]. Therefore, it seems that the existence of paralogous E and Id genes at the time of the horizontal acquisition of transposable *Rag* genes contributed to the genesis of the AIS in jawed vertebrates. The two rounds of WGD are considered important for the acquired immunity of jawed vertebrates [82,83]. Previous research proposed that two rounds of WGD played a major role in the duplication of many signaling genes ancestrally used in nervous system development and function that were later co-opted for new functions during evolution of the AIS [84]. E-Id paralogous genes represent new evidence that paralogous genes arisen by WGDs could be co-opted for new functions in AIS in jawed vertebrates. In fact, functional redundancies among E proteins or Id proteins are observed in T cell lineage commitment and development (E2A and HEB, and Id2 and Id3), Treg cell function (Id2 and Id3), and germinal center B cell development (E2A and E2-2), suggesting that the duplication of E or Id genes could give rise to diversity and stability in AIS [48,60,65,85,86]. These observations also reinforce the hypothesis that not only the acquisition of novel molecular capabilities but also the co-option and redirection of preexisting systems are the major sources of innovation [87].

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **1983**, *302*, 575–581. [CrossRef]
2. Schatz, D.G.; Oettinger, M.A.; Baltimore, D. The V(D)J recombination activating gene, RAG-1. *Cell* **1989**, *59*, 1035–1048. [CrossRef]
3. Schatz, D.G.; Ji, Y. Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* **2011**, *11*, 251–263. [CrossRef]
4. Kuo, T.C.; Schlissel, M.S. Mechanisms controlling expression of the RAG locus during lymphocyte development. *Curr. Opin. Immunol.* **2009**, *21*, 173–178. [CrossRef]
5. Fugmann, S.D. The origins of the Rag genes—From transposition to V(D)J recombination. *Semin. Immunol.* **2010**, *22*, 10–16. [CrossRef] [PubMed]
6. Carmona, L.M.; Schatz, D.G. New insights into the evolutionary origins of the recombination-activating gene proteins and V(D)J recombination. *FEBS J.* **2017**, *284*, 1590–1605. [CrossRef]
7. Huang, S.; Tao, X.; Yuan, S.; Zhang, Y.; Li, P.; Beilinson, H.; Zhang, Y.; Yu, W.; Pontarotti, P.; Escriva, H.; et al. Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell* **2016**, *166*, 102–114. [CrossRef] [PubMed]
8. Zhang, Y.; Cheng, T.C.; Huang, G.; Lu, Q.; Surleac, M.D.; Mandell, J.D.; Pontarotti, P.; Petrescu, A.J.; Xu, A.; Xiong, Y.; et al. Transposon molecular domestication and the evolution of the RAG recombinase. *Nat. Cell Biol.* **2019**, *569*, 79–84. [CrossRef] [PubMed]
9. Rothenberg, E.V. Programming for T-lymphocyte fates: Modularity and mechanisms. *Genes Dev.* **2019**, *33*, 1117–1135. [CrossRef]
10. Muljo, S.A.; Schlissel, M.S. Pre-B and pre-T-cell receptors: Conservation of strategies in regulating early lymphocyte development. *Immunol. Rev.* **2000**, *175*, 80–93. [CrossRef]
11. Mandel, E.M.; Grosschedl, R. Transcription control of early B cell differentiation. *Curr. Opin. Immunol.* **2010**, *22*, 161–167. [CrossRef]
12. Lin, Y.C.; Jhunjhunwala, S.; Benner, C.; Heinz, S.; Welinder, E.; Mansson, R.; Sigvardsson, M.; Hagman, J.; Espinoza, C.; Dutkowski, J.; et al. A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat. Immunol.* **2010**, *11*, 635–643. [CrossRef]
13. Serafini, N.; Vosshenrich, C.A.J.; Di Santo, J.P. Transcriptional regulation of innate lymphoid cell fate. *Nat. Rev. Immunol.* **2015**, *15*, 415–428. [CrossRef] [PubMed]
14. Miyazaki, K.; Miyazaki, M. The interplay between chromatin architecture and lineage-specific transcription factors and the regulation of Rag gen expression. *Front. Immunol.* **2021**, *12*, 6597612021. [CrossRef] [PubMed]
15. Boller, S.; Li, R.; Grosschedl, R. Defining B Cell Chromatin: Lessons from EBF1. *Trends Genet.* **2018**, *34*, 257–269. [CrossRef] [PubMed]
16. Hosokawa, H.; Rothenberg, E.V. How transcription factors drive choice of the T cell fate. *Nat. Rev. Immunol.* **2021**, *21*, 162–176. [CrossRef] [PubMed]
17. Eberl, G.; Colonna, M.; Di Santo, J.P.; McKenzie, A.N.J. Innate lymphoid cells: A new paradigm in immunology. *Science* **2015**, *348*, aaa6566. [CrossRef]
18. Miyazaki, M.; Miyazaki, K.; Chen, K.; Jin, Y.; Turner, J.; Moore, A.J.; Saito, R.; Yoshida, K.; Ogawa, S.; Rodewald, H.-R.; et al. The E-Id Protein Axis Specifies Adaptive Lymphoid Cell Identity and Suppresses Thymic Innate Lymphoid Cell Development. *Immunity* **2017**, *46*, 818–834.e4. [CrossRef]
19. Nechanitzky, R.; Akbas, D.; Scherer, S.; Györy, I.; Hoyler, T.; Ramamoorthy, S.; Diefenbach, A.; Grosschedl, R. Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. *Nat. Immunol.* **2013**, *14*, 867–875. [CrossRef]
20. Yannoutsos, N.; Barreto, V.M.; Misulovin, Z.; Gazumyan, A.; Yu, W.; Rajewsky, N.; Peixoto, B.R.; Eisenreich, T.; Nussenzweig, M.C. A cis element in the recombination activating gene locus regulates gene expression by counteracting a distant silencer. *Nat. Immunol.* **2004**, *5*, 443–450. [CrossRef]
21. Miyazaki, K.; Watanabe, H.; Yoshikawa, G.; Chen, K.; Hidaka, R.; Aitani, Y.; Osawa, K.; Takeda, R.; Ochi, Y.; Tani-Ichi, S.; et al. The transcription factor E2A activates multiple enhancers that drive Rag expression in developing T and B cells. *Sci. Immunol.* **2020**, *5*, eabb1455. [CrossRef] [PubMed]
22. Hsu, L.-Y.; Lauring, J.; Liang, H.-E.; Greenbaum, S.; Cado, D.; Zhuang, Y.; Schlissel, M.S. A Conserved Transcriptional Enhancer Regulates RAG Gene Expression in Developing B Cells. *Immunity* **2003**, *19*, 105–117. [CrossRef]
23. Murre, C. Helix-loop-helix proteins and the advent of cellular diversity: 30 years of discovery. *Genes Dev.* **2019**, *33*, 6–25. [CrossRef]
24. Wilson, A.; Held, W.; Macdonald, H.R. Two waves of recombinase gene expression in developing thymocytes. *J. Exp. Med.* **1994**, *179*, 1355–1360. [CrossRef] [PubMed]
25. Krangel, M.S. Mechanics of T cell receptor gene rearrangement. *Curr. Opin. Immunol.* **2009**, *21*, 133–139. [CrossRef] [PubMed]
26. Mombaerts, P.; Iacomini, J.; Johnson, R.S.; Herrup, K.; Tonegawa, S.; Papaioannou, V.E. RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* **1992**, *68*, 869–877. [CrossRef]

27. Shinkai, Y.; Rathbun, G.; Lam, K.P.; Oltz, E.M.; Stewart, V.; Mendelsohn, M.; Charron, J.; Datta, M.; Young, F.; Stall, A.M.; et al. RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **1992**, *68*, 855–867. [CrossRef]

28. Wayne, J.; Suh, H.; A Sokol, K.; Petrie, H.T.; Witmer-Pack, M.; Edelhoff, S.; Disteche, C.M.; Nussenzweig, M.C. TCR selection and allelic exclusion in RAG transgenic mice that exhibit abnormal T cell localization in lymph nodes and lymphatics. *J. Immunol.* **1994**, *153*, 5491–5502. [PubMed]

29. Hao, B.; Naik, A.; Watanabe, A.; Tanaka, H.; Chen, L.; Richards, H.W.; Kondo, M.; Taniuchi, I.; Kohwi, Y.; Kohwi-Shigematsu, T.; et al. An anti-silencer– and SATB1-dependent chromatin hub regulates Rag1 and Rag2 gene expression during thymocyte development. *J. Exp. Med.* **2015**, *212*, 809–824. [CrossRef] [PubMed]

30. Naik, A.; Byrd, A.T.; Lucander, A.C.; Krangel, M.S. Hierarchical assembly and disassembly of a transcriptionally active RAG locus in CD4+CD8+ thymocytes. *J. Exp. Med.* **2019**, *216*, 231–243. [CrossRef]

31. Amin, R.H.; Schlissel, M.S. Foxo1 directly regulates the transcription of recombination-activating genes during B cell development. *Nat. Immunol.* **2008**, *9*, 613–622. [CrossRef] [PubMed]

32. Long, H.K.; Prescott, S.L.; Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **2016**, *167*, 1170–1187. [CrossRef]

33. Rogozin, I.B.; Iyer, L.M.; Liang, L.; Glazko, G.V.; Liston, V.G.; I Pavlov, Y.; Aravind, L.; Pancer, Z. Evolution and diversification of lamprey antigen receptors: Evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat. Immunol.* **2007**, *8*, 647–656. [CrossRef]

34. Trancoso, I.; Morimoto, R.; Boehm, T. Co-evolution of mutagenic genome editors and vertebrate adaptive immunity. *Curr. Opin. Immunol.* **2020**, *65*, 32–41. [CrossRef]

35. Morimoto, R.; O'Meara, C.P.; Holland, S.J.; Trancoso, I.; Souissi, A.; Schorpp, M.; Vassaux, D.; Iwanami, N.; Giorgetti, O.B.; Evanno, G.; et al. Cytidine deaminase 2 is required for VLRB antibody gene assembly in lampreys. *Sci. Immunol.* **2020**, *5*, eaba0925. [CrossRef]

36. Hsu, E.; Pulham, N.; Rumfelt, L.L.; Flajnik, M.F. The plasticity of immunoglobulin gene systems in evolution. *Immunol. Rev.* **2006**, *210*, 8–26. [CrossRef]

37. Caudy, M.; Vässin, H.; Brand, M.; Tuma, R.; Jah, L.Y.; Jan, Y.N. daughterless, a Drosophila gene essential for both neurogenesis and sex determination, has sequence similarities to myc and the achaete-scute complex. *Cell* **1988**, *55*, 1061–1067. [CrossRef]

38. Sallee, M.D.; Littleford, H.E.; Greenwald, I. A bHLH Code for Sexually Dimorphic Form and Function of the C. elegans Somatic Gonad. *Curr. Biol.* **2017**, *27*, 1853–1860.e5. [CrossRef]

39. Kee, B.L. E and ID proteins branch out. *Nat. Rev. Immunol.* **2009**, *9*, 175–184. [CrossRef]

40. Benezra, R.; Davis, R.L.; Lockshon, D.; Turner, D.L.; Weintraub, H. The protein Id: A negative regulator of helix-loop-helix DNA binding proteins. *Cell* **1990**, *61*, 49–59. [CrossRef]

41. Belle, I.; Zhuang, Y. E Proteins in Lymphocyte Development and Lymphoid Diseases. *Curr. Top. Dev. Biol.* **2014**, *110*, 153–187. [CrossRef]

42. Bain, G.; Maandag, E.C.; Izon, D.J.; Amsen, D.; Kruisbeek, A.M.; Weintraub, B.C.; Krop, I.; Schlissel, M.S.; Feeney, A.J.; Van Roon, M.; et al. E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell* **1994**, *79*, 885–892. [CrossRef]

43. Zhuang, Y.; Soriano, P.; Weintraub, H. The helix-loop-helix gene E2A is required for B cell formation. *Cell* **1994**, *79*, 875–884. [CrossRef]

44. Murre, C.; McCaw, P.S.; Baltimore, D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* **1989**, *56*, 777–783. [CrossRef]

45. Yamazaki, T.; Liu, L.; Lazarev, D.; Al-Zain, A.; Fomin, V.; Yeung, P.L.; Chambers, S.M.; Lu, C.-W.; Studer, L.; Manley, J.L. TCF3 alternative splicing controlled by hnRNP H/F regulates E-cadherin expression and hESC pluripotency. *Genes Dev.* **2018**, *32*, 1161–1174. [CrossRef]

46. Lin, Y.C.; Benner, C.; Mansson, R.; Heinz, S.; Miyazaki, K.; Miyazaki, M.; Chandra, V.; Bossen, C.; Glass, C.K.; Murre, C. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat. Immunol.* **2012**, *13*, 1196–1204. [CrossRef] [PubMed]

47. D'cruz, L.M.; Knell, J.; Fujimoto, J.K.; Goldrath, A.W. An essential role for the transcription factor HEB in thymocyte survival, Tcra rearrangement and the development of natural killer T cells. *Nat. Immunol.* **2010**, *11*, 240–249. [CrossRef]

48. Jones, M.E.; Zhuang, Y. Acquisition of a Functional T Cell Receptor during T Lymphocyte Development Is Enforced by HEB and E2A Transcription Factors. *Immunity* **2007**, *27*, 860–870. [CrossRef]

49. Cisse, B.; Caton, M.L.; Lehner, M.; Maeda, T.; Scheu, S.; Locksley, R.; Holmberg, D.; Zweier, C.; Hollander, N.S.D.; Kant, S.G.; et al. Transcription Factor E2-2 Is an Essential and Specific Regulator of Plasmacytoid Dendritic Cell Development. *Cell* **2008**, *135*, 37–48. [CrossRef]

50. Ghosh, H.S.; Cisse, B.; Bunin, A.; Lewis, K.L.; Reizis, B. Continuous Expression of the Transcription Factor E2-2 Maintains the Cell Fate of Mature Plasmacytoid Dendritic Cells. *Immunity* **2010**, *33*, 905–916. [CrossRef]

51. Grajkowska, L.T.; Ceribelli, M.; Lau, C.M.; Warren, M.E.; Tiniakou, I.; Higa, S.N.; Bunin, A.; Haecker, H.; Mirny, L.A.; Staudt, L.M.; et al. Isoform-Specific Expression and Feedback Regulation of E Protein TCF4 Control Dendritic Cell Lineage Specification. *Immunity* **2017**, *46*, 65–77. [CrossRef]

52. Verykokakis, M.; Zook, E.C.; Kee, B.L. ID'ing innate and innate-like lymphoid cells. *Immunol. Rev.* **2014**, *261*, 177–197. [CrossRef]
53. Kee, B.L.; Rivera, R.R.; Murre, C. Id3 inhibits B lymphocyte progenitor growth and survival in response to TGF-beta. *Nat. Immunol.* **2001**, *2*, 242–247. [CrossRef] [PubMed]
54. Chen, S.; Miyazaki, M.; Chandra, V.; Fisch, K.M.; Chang, A.N.; Murre, C. Id3 Orchestrates Germinal Center B Cell Development. *Mol. Cell. Biol.* **2016**, *36*, 2543–2552. [CrossRef] [PubMed]
55. Gloury, R.; Zotos, D.; Zuidscherwoude, M.; Masson, F.; Liao, Y.; Hasbold, J.; Corcoran, L.M.; Hodgkin, P.D.; Belz, G.T.; Shi, W.; et al. Dynamic changes in Id3 and E-protein activity orchestrate germinal center and plasma cell development. *J. Exp. Med.* **2016**, *213*, 1095–1111. [CrossRef]
56. Miyazaki, M.; Rivera, R.R.; Miyazaki, K.; Lin, Y.C.; Agata, Y.; Murre, C. The opposing roles of the transcription factor E2A and its antagonist Id3 that orchestrate and enforce the naive fate of T cells. *Nat. Immunol.* **2011**, *12*, 992–1001. [CrossRef]
57. Engel, I.; Murre, C. E2A proteins enforce a proliferation checkpoint in developing thymocytes. *EMBO J.* **2003**, *23*, 202–211. [CrossRef] [PubMed]
58. Lauritsen, J.P.; Wong, G.W.; Lee, S.Y.; Lefebvre, J.M.; Ciofani, M.; Rhodes, M.; Kappes, D.J.; Zúñiga-Pflücker, J.C.; Wiest, D.L. Marked induction of the helix-loop-helix protein Id3 promotes the gammadelta T cell fate and renders their functional maturation Notch independent. *Immunity* **2009**, *31*, 565–575. [CrossRef]
59. Fahl, S.P.; Contreras, A.V.; Verma, A.; Qiu, X.; Harly, C.; Radtke, F.; Zúñiga-Pflücker, J.C.; Murre, C.; Xue, H.H.; Sen, J.M.; et al. The E protein-TCF1 axis controls γδ T cell development and effector fate. *Cell Rep.* **2021**, *34*, 108716. [CrossRef]
60. Miyazaki, M.; Miyazaki, K.; Chen, S.; Chandra, V.; Wagatsuma, K.; Agata, Y.; Rodewald, H.-R.; Saito, R.; Chang, A.N.; Varki, N. The E–Id protein axis modulates the activities of the PI3K–AKT–mTORC1–Hif1a and c-myc/p19Arf pathways to suppress innate variant TFH cell development, thymocyte expansion, and lymphomagenesis. *Genes Dev.* **2015**, *29*, 409–425. [CrossRef]
61. Liu, X.; Chen, X.; Zhong, B.; Wang, A.; Wang, X.; Chu, F.; Nurieva, R.I.; Yan, X.; Chen, P.; Van Der Flier, L.G.; et al. Transcription factor achaete-scute homologue 2 initiates follicular T-helper-cell development. *Nat. Cell Biol.* **2014**, *507*, 513–518. [CrossRef]
62. Leong, Y.A.; Chen, Y.; Ong, H.S.; Wu, D.; Man, K.; Deleage, C.; Minnich, M.; Meckiff, B.J.; Wei, Y.; Hou, Z.; et al. CXCR5+ follicular cytotoxic T cells control viral infection in B cell follicles. *Nat. Immunol.* **2016**, *17*, 1187–1196. [CrossRef]
63. Yang, C.Y.; Best, J.A.; Knell, J.; Yang, E.; Sheridan, A.D.; Jesionek, A.K.; Li, H.S.; Rivera, R.R.; Lind, K.C.; D'Cruz, L.M.; et al. The transcriptional regulators Id2 and Id3 control the formation of distinct memory CD8+ T cell subsets. *Nat. Immunol.* **2011**, *12*, 1221–1229. [CrossRef]
64. Omilusik, K.D.; Shaw, L.A.; Goldrath, A.W. Remembering one's ID/E-ntity: E/ID protein regulation of T cell memory. *Curr. Opin. Immunol.* **2013**, *25*, 660–666. [CrossRef]
65. Miyazaki, M.; Miyazaki, K.; Chen, S.; Itoi, M.; Miller, M.; Lu, L.-F.; Varki, N.; Chang, A.N.; Broide, D.H.; Murre, C. Id2 and Id3 maintain the regulatory T cell pool to suppress inflammatory disease. *Nat. Immunol.* **2014**, *15*, 767–776. [CrossRef]
66. Bhattacharya, A.; Baker, N.E. A Network of Broadly Expressed HLH Genes Regulates Tissue-Specific Cell Fates. *Cell* **2011**, *147*, 881–892. [CrossRef] [PubMed]
67. Simakov, O.; Marlétaz, F.; Yue, J.-X.; O'Connell, B.; Jenkins, J.; Brandt, A.; Calef, R.; Tung, C.-H.; Huang, T.-K.; Schmutz, J.; et al. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat. Ecol. Evol.* **2020**, *4*, 820–830. [CrossRef] [PubMed]
68. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]
69. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef]
70. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
71. Cannon, J.P.; Haire, R.N.; Rast, J.P.; Litman, G.W. The phylogenetic origins of the antigen-binding receptors and somatic diversification mechanisms. *Immunol. Rev.* **2004**, *200*, 12–22. [CrossRef] [PubMed]
72. Mayer, W.E.; Uinuk-Ool, T.; Tichy, H.; Gartland, L.A.; Klein, J.; Cooper, M.D. Isolation and characterization of lymphocyte-like cells from a lamprey. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14350–14355. [CrossRef] [PubMed]
73. Guo, P.; Hirano, M.; Herrin, B.R.; Li, J.; Yu, C.; Sadlonova, A.; Cooper, M.D. Dual nature of the adaptive immune system in lampreys. *Nat. Cell Biol.* **2009**, *459*, 796–801. [CrossRef]
74. Boehm, T.; Hirano, M.; Holland, S.J.; Das, S.; Schorpp, M.; Cooper, M.D. Evolution of Alternative Adaptive Immune Systems in Vertebrates. *Annu. Rev. Immunol.* **2018**, *36*, 19–42. [CrossRef]
75. Rothenberg, E.V.; Pant, R. Origins of lymphocyte developmental programs: Transcription factor evidence. *Semin. Immunol.* **2004**, *16*, 227–238. [CrossRef]
76. Cooper, M.D.; Alder, M.N. The Evolution of Adaptive Immune Systems. *Cell* **2006**, *124*, 815–822. [CrossRef]
77. Suzuki, T.; Shin, I.T.; Kohara, Y.; Kasahara, M. Transcriptome analysis of hagfish leukocytes: A framework for understanding the immune system of jawless fishes. *Dev. Comp. Immunol.* **2004**, *28*, 993–1003. [CrossRef]
78. Pancer, Z.; Amemiya, C.T.; Ehrhardt, G.R.A.; Ceitlin, J.; Gartland, G.L.; Cooper, M.D. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nat. Cell Biol.* **2004**, *430*, 174–180. [CrossRef]
79. Muramatsu, M.; Kinoshita, K.; Fagarasan, S.; Yamada, S.; Shinkai, Y.; Honjo, T. Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. *Cell* **2000**, *102*, 553–563. [CrossRef]

80. Müller, V.; De Boer, R.J.; Bonhoeffer, S.; Szathmáry, E. An evolutionary perspective on the systems of adaptive immunity. *Biol. Rev.* **2018**, *93*, 505–528. [CrossRef]

81. Kapitonov, V.V.; Koonin, E.V. Evolution of the RAG1-RAG2 locus: Both proteins came from the same transposon. *Biol. Direct* **2015**, *10*, 1–8. [CrossRef]

82. Kasahara, M. What do the paralogous regions in the genome tell us about the origin of the adaptive immune system? *Immunol. Rev.* **1998**, *166*, 159–175. [CrossRef]

83. Kasahara, M.; Suzuki, T.; Du Pasquier, L. On the origins of the adaptive immune system: Novel insights from invertebrates and cold-blooded vertebrates. *Trends Immunol.* **2004**, *25*, 105–111. [CrossRef]

84. Okada, K.; Asai, K. Expansion of signaling genes for adaptive immune system evolution in early vertebrates. *BMC Genom.* **2008**, *9*, 218. [CrossRef]

85. Jones-Mason, M.E.; Zhao, X.; Kappes, D.; Lasorella, A.; Iavarone, A.; Zhuang, Y. E Protein Transcription Factors Are Required for the Development of CD4+ Lineage T Cells. *Immunity* **2012**, *36*, 348–361. [CrossRef]

86. Wöhner, M.; Tagoh, H.; Bilic, I.; Jaritz, M.; Poliakova, D.K.; Fischer, M.; Busslinger, M. Molecular functions of the transcription factors E2A and E2-2 in controlling germinal center B cell and plasma cell development. *J. Exp. Med.* **2016**, *213*, 1201–1221. [CrossRef]

87. Litman, G.W.; Rast, J.P.; Fugmann, S.D. The origins of vertebrate adaptive immunity. *Nat. Rev. Immunol.* **2010**, *10*, 543–553. [CrossRef]