Characterization of the Human and Mouse T Cell Receptor Gamma Loci

by

Qun Zhou

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biology

We accept this thesis as conforming to the required standard

[redacted]

Dr. B. F. Koop, Supervisor (Department of Biology)

[redacted]

Dr. N. Sherwood, Departmental Member (Department of Biology)

[redacted]

Dr. W.E. Hintz, Departmental Member (Department of Biology)

[redacted]

Dr. T.W. Pearson, Outside Member (Department of Biochemistry & Microbiology)

[redacted]

Dr. C. Upton, External Examiner (Department of Biochemistry & Microbiology)

University of Victoria

ABSTRACT

T cell receptors (TCRs) are membrane proteins on T lymphocytes that recognize antigens presented by major histocompatibility complexes (MHC). The genes encoding TCRs are located at three loci, α/δ, β and γ. As multigenic families, TCR loci are rich in structural, gene regulatory, and evolutionary features. Previous studies predominantly focused on mRNA and partial germline sequences, which served to provide a limited understanding of such features. As a part of the Human Genome Project, the sequencing of both the human and mouse TCR γ loci has been completed in our lab during the past eight years, which has contributed to a more complete understanding of the features of the TCR γ loci.

The completely sequenced human TCR γ locus is 140 kb. The identified known coding elements include fourteen V (variable) gene fragments located at the 5' end, as well as five J (joining) and two C (constant) gene fragments at the 3' end. In addition, two V relics were identified by a sequence similarity comparison in the middle of Vγ9 and Vγ10, and Vγ11 and JP1, respectively. The human locus possesses a high percentage of short interspersed elements (SINEs) (13.3%) and a low percentage of long interspersed elements (LINEs) (6.35%). The phylogenetic analysis indicates that the nine tandemly repetitive V genes in the GV1 family are generated by six unequal crossing-over events and one gene conversion, and the two J-C clusters are also duplicated by unequal crossing-over events. The increased number of V, J and C fragments greatly enlarges recombinatory diversity. Further analysis of different functional domains of the members of the GV1 family suggests that the diversification of the duplicated V genes is generated by positive natural selection on the antigen binding domains. Thus, the above organizational and evolutionary features reveal that this locus evolved under a birth-and-death process.

In the 205 kb mouse TCR γ locus, 167 kb sequences covering the 5' and 3' end regions were determined, and structural, gene regulatory, and evolutionary features were characterized. The known coding gene fragments distributed in the four V-J-C clusters, the recombination and splicing signals, and enhancers were identified and located on the

genomic sequence. Unlike the human locus, this locus is structurally characterized by many LINEs (30.9%), and few SINEs (2.0%). Five large LINEs were generated by both retroposition and duplication combined in their respective clusters. In addition, by comparing the genomic sequence with the sequence of the known enhancer γE1, a new putative enhancer, γE4, was found 3.7 kb downstream of the 3' end of the Cγ4 gene fragment. This locus is also characterized by a few large locus-specific repeats. The phylogenetic analysis of these repeats suggests that clusters 2 and 3 are recent duplications, while clusters 1 and 4 are ancient clusters. Among the four clusters, cluster 2 reversed its transcriptional orientation, while cluster 3 became non-functional due to the pseudogene Cγ3. These organizational features, together with the clusteral organization, imply that the natural selection pressure on the mouse TCR γ chain decreased, with the result that cluster 2 had to preserve its genetic memory by changing transcriptional orientation to avoid deletion by recombination.

The characterization of the DNA fragments in cluster 4 of the mouse TCR γ locus at the γδ T cell level revealed that the putative enhancer identified by sequence comparison, γE4, is functional. The minimal enhancer is confined within a 294 bp region that contained NFγ2, NFγ3 and NFγ4, the three nuclear factor binding sequences identified in enhancer γE1. The maximum enhancer is a 651 bp fragment that encompasses the six NFγs. In addition, a 900 bp silencer was identified at the 5' end of the enhancer. The maximum enhancer exhibited maximal activity in the γδ T cell line PEER and slightly higher activities than that conferred by the control construct in the γδ T cell line Molt-13 and αβ T cell EL-4. The maximum enhancer did not display activity in nonlymphocyte HeLa cells. Therefore, similar to other TCR enhancers, γE4 is also T cell-specific, but not γδ lineage-specific. However, the fragment containing both the silencer and enhancer showed an increased activity in comparison with the enhancer in Molt-13, while a decreased activity relative to the enhancer in PEER. Since both Molt-13 and PEER are human γδ T cell lines expressing different Vγ genes, these different patterns of activity suggest that the combination of positive and negative elements, in addition to other regulatory elements in the TCR γ locus, may control the specific subtype of the γδ T cells

selection and may work together to control the spatial and temporal expression of γδ T cells.

Examiners:

[redacted]

Dr. B. F. Koop, Supervisor (Department of Biology)

[redacted]

Dr. N. Sherwood, Departmental Member (Department of Biology)

[redacted]

Dr. W.E. Hintz, Departmental Member (Department of Biology)

[redacted]

Dr. T.W. Pearson, Outside Member (Department of Biochemistry & Microbiology)

[redacted]

Dr. C. Upton, External Examiner (Department of Biochemistry & Microbiology)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I. Introduction

1. The immune system:

1.1) Overview of the immune system:

The environments that biological organisms live in are full of harmful agents, such as viruses, bacteria, parasites, and fungi. In the course of evolution, the immune system appeared to protect these living organisms from attack by invasive pathogens. As a defense, immune system functions become increasingly specific and selective (Elgert, 1996; Segel & Cohen, 2001). Immunity can be innate or acquired. Innate immunity is a kind of "non-specific protection", such as one's skin barrier which prevents the entrance of pathogens into the body and the process of phagocytosis which uses "nonspecific" immune cells. Such a defense exists in all animals and usually does not require prior contact with specific antigens. However, recent findings show that the innate immune system possesses a higher degree of specificity than was previously thought (Medzhitov et al., 1997). In contrast, acquired immunity, present in jawed vertebrate species, needs prior contact with antigens to activate the so-called specific immune system through two principle actors, B and T lymphocytes (Litman, 1999). Thus, one's specific immune system is able to recognize a particular pathogen and eliminate it mainly in a second invasion (Mak & Ferrick, 1998; Hughes & Yeager, 1997). When activated by specific antigens, B cells differentiate into plasma cells that secrete immunoglobulins (Igs) into biological fluid. The extracellular Igs can bind directly to pathogens to neutralize toxins and viruses or to opsonize bacteria in order to facilitate easier phagocytosis. T cells, however, recognize the antigen and MHC complex presented by antigen presenting cells (APCs) via T cell receptors (TCRs) on their cell membranes, and primarily perform immune regulatory and effector functions. The regulatory functions are mediated by helper T cells. Of the two types of helper T cells, the Th-1 cells secrete IL-2 and INF-γ to activate cytotoxic T cells and macrophages respectively, whereas the Th-2 cells secrete IL-4 and IL-5 to stimulate B cells which differentiate into plasma cells. The effector functions are carried out by cytotoxic T cells, which kill virus-infected, tumor and allograft cells either by releasing porphorins to destroy membranes or by inducing apoptosis (Mosman & Coffman, 1989; Kaufmann, 1996).

1.2) Origin of immune cells:

Lymphocytes are derived from the common haematopoietic stem cells in bone marrow (Figure 1). Precursors of B cells mature elsewhere. For example, they differentiate and mature in bone marrow in humans; but in birds, are processed in the bursa of Fabricius (Dunon & Imhof, 1996). However, T cell precursors have to pass through the thymus to become immunocompetent. Within the thymus, T cells experience two important processes termed thymic education: 1) the T cells that can bind self-MHC effectively are chosen by positive selection in the cortex and then continue to differentiate. Cells which do not react with self-MHC are killed via apoptosis; and 2) following positive selection, the cells selected will further experience a negative selection in both the cortex and medulla, wherein cells that can recognize the self-antigen-MHC complex are additionally killed. Thus, through these two processes, mature T cells which have the ability to react with both foreign antigens and self-MHCs are produced. Once lymphocytes become fully mature, they leave the thymus to reside in the peripheral immune organs such as the lymph nodes, spleen and mucosa-associated lymphoid tissues. Immune responses to antigens take place in the peripheral immune organs (Fehling et al., 1999).

1.3) Important cell surface molecules involved in the activation of T cells:

T cells initiate their immune responses when TCRs recognize and bind to specific antigens. This process is intimately involved with a series of protein molecules expressed on the T cell and APC surface (Jorgensen et al., 1992). Such membrane proteins include TCR, MHC, CD3 (cluster of determination), and CD4 or CD8, which form a complex with antigens to initiate the immune response. CD3 is composed of five subunits, δ, ε, γ, η, and ζ (Carson et al., 1991). They link to TCR by non-covalent bonds to form a TCR complex. These subunits can stabilize the structure of TCRs and transduce signals into the cytoplasm to activate T cells once TCRs interact with antigens. CD4 and CD8 are expressed on helper and cytotoxic T cells respectively. They bind to the constant region of the MHCs on the APCs, which stabilizes the interaction between the two cells (Davis et al., 1998; Konig, 2002). In addition, a non-antigen co-stimulation is also required because the efficient activation of T cells requires signals from both the TCR and an additional co-stimulatory receptor. For example, the CD28 protein is one of the most

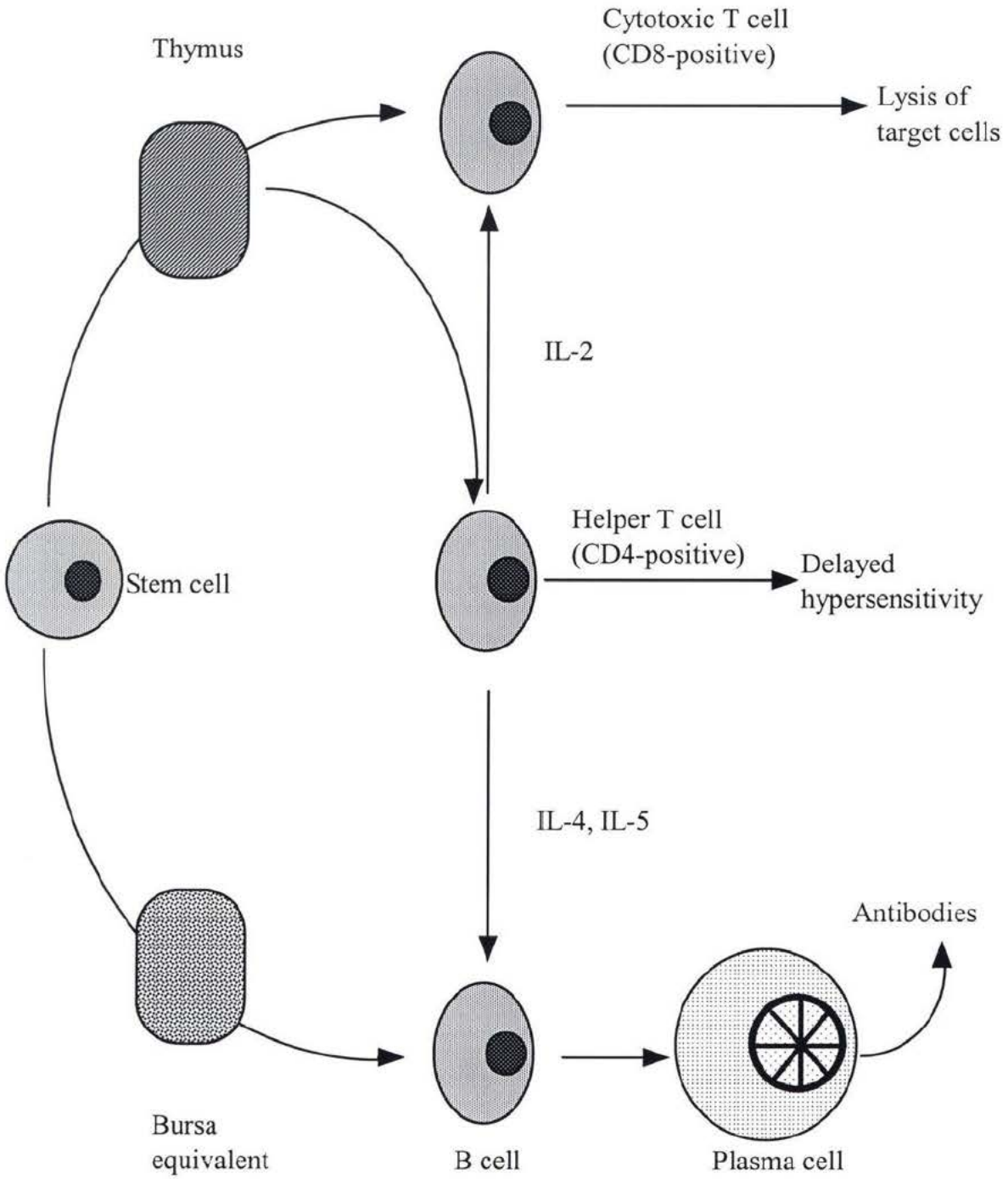


Figure 1. Origin of T and B lymphocytes. Both B and T cells are derived from the bone marrow stem cells. T cells mature in the thymus while B cells mature in the Bursa equivalent (e.g. bone marrow in Mammalia) (adapted from Brooks G.F. et al: Medical Microbiology, 19th ed. Appleton & Lange, 1991).

potent and well-characterized co-stimulatory molecules on T cells, which binds to CD80 proteins on APCs (Sansom, 2000; Salomon & Bluestone, 2001). Without the co-stimulatory receptors, the responsiveness of T cells to antigens can be inhibited or substantially decreased (Levinson et al., 1996; Lenschow et al., 1996).

Of the surface molecules previously mentioned, TCRs possess the maximum diversity in specific immune recognition. TCRs are divided into αβ and γδ subtypes, which are composed of two of the four peptide chains, α, β, γ and δ. Similar to Igs, each chain contains a variable region (V) and a constant region (C) (Davis, 1990). In the V regions, there are three complementary determination regions (CDRs) exhibiting high diversity and four flanking framework regions (FRs) with low diversity. The CDRs congregate on the surface of the TCR molecules to make contact with the antigen-MHC complexes and they contribute recognition specificity, while the FRs provide the framework structure of three-dimensional conformations (Hong et al., 1992). Among the three CDRs, CDR1 and CDR2 contact the helices of MHC to provide the basic affinity of the TCR/Ag/MHC complex. In addition, CDR3 is highly diverse because it corresponds to the junctional region of rearranged gene fragments of V-J and V-D-J and it tends to contact with the specific bound peptide fragments. Thus, CDR3 is the key determinant of immune specificity in antigen recognition (Fremont et al., 1996; Chien et al., 1996; Xu et al., 2000).

In addition to TCRs, MHCs are another type of key membrane molecules that possess specific immune recognition, which exhibit a high polymorphism between individuals. This MHC polymorphism is characterized by a large number of alleles at a given functional locus and a large number of nucleotide differences between these alleles (Ono et al., 1992). Two prime classical kinds of MHCs exist, namely MHC-I and MHC-II. MHC-1 molecules consist of an α chain and a β2-microglobulin, while MHC-II molecules consist of noncovalently associated α and β chains (Klein, 1986). The encoding genes for MHC-I are located in the H-2 locus in mice and HLA-A, B and C loci in humans, whereas the genes for MHC-II are in Ia in mice and HLA-DP, DQ and DR in humans. Since an individual gains one strand of DNA from each parent, most people are

endowed with two distinct variants of A, two of B and two of C, for a total of six distinct MHC-I genes. Likewise, most people possess two variants of each of the MHC-II genes, producing a total of six genes. The locus which contains these genes is split into three smaller loci named DP, DQ and DR. Since each locus contains one gene for an α unit and one for a β unit, most people have a maximum of twelve MHC II molecules. The MHC-I molecules are ubiquitously distributed and preferentially present antigens to cytotoxic T cells. The MHC-II molecules are predominantly expressed on lymphocytes, and preferentially present antigens to the helper T cells (Bjorkman et al., 1987). In the course of T cell activation, antigens are taken up into APCs, are processed and the epitopes bound to the cleft between the two α-helices of MHC-I and the equivalent site of MHC II on the outer surface of the newly synthesized MHC glycoproteins in endosomes. The MHC/peptide complexes are then transported to the cell membrane where they can bind TCRs and the CD3 complex (Watts, 1997). In general, MHC-I presents endogenously synthesized antigens, such as viral proteins, whereas MHC II presents antigens of extracellular microorganisms that have been phagocytized, such as bacterial proteins. Thus, with MHC restriction, the cytotoxic T cells can only attack the infected cells bearing MHC. In addition to the classical MHC-I and MHC-II, there also exist so-called MHC-like molecules. MHC class I-like molecules can present peptides with specific sequences on the cell surface. For example, the mouse MHC class I-like molecule T23 protein can bind to peptides of Hsp65. However, MHC class II-like molecules are not expressed on the cell surface. In endosomes, they have the ability to transfer the processed antigens to MHC class II molecules (Steele et al., 2000; McDevitt, 2000).

1.4) Somatic recombination:

Among the numerous immune molecules, the Ig and TCR protein chains are encoded by combinatory genes which are produced by assembling separated gene segments including variable (V), diversity (D), joining (J) and constant (C) fragments. TCR gene fragments are distributed on different gene loci on chromosomes. During the maturation of T cells, these gene fragments undergo somatic rearrangement in the thymus, which brings V, D and J together by looping out or inverting and then deleting the intervening DNA depending on the orientation of the sequences being joined (Figure 2) (Weaver,

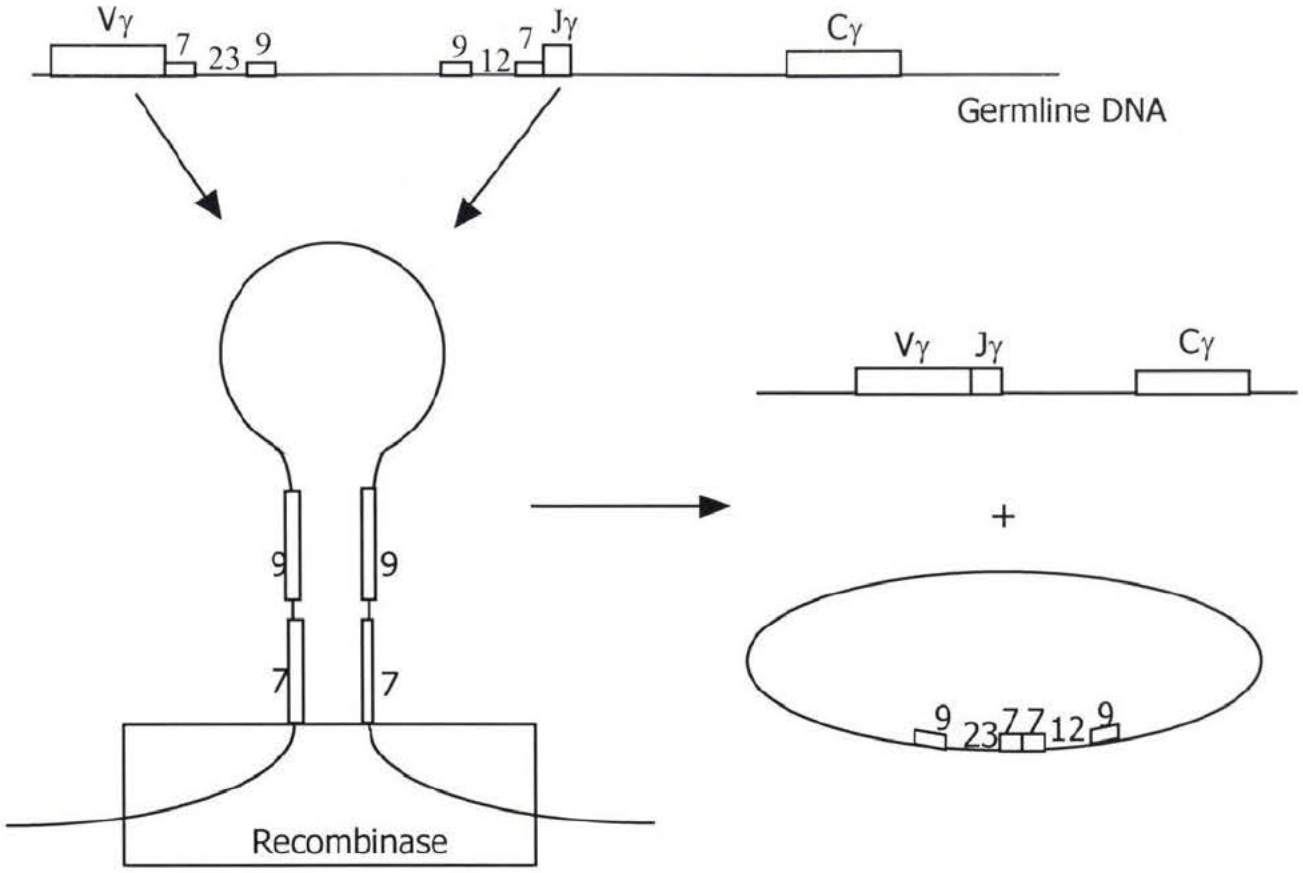


Figure 2. DNA rearrangement of V and J gene fragments. Each V, (D), and J gene element is flanked by an recombination signal sequence (RSS) which consists of a highly conserved heptamer separated by a spacer which in turn is separated by a nonamer. Recombinase brings the RSSs together and incises the stem structure and then ligates the broken ends (modified from Weaver, 1995).

1995). Recombination signal sequences (RSSs) are located at the 3' end of the V gene fragments and the 5' end of the J gene fragments. The RSS is composed of a seven bp conserved heptamer sequence CACAGTG, a spacer, and a nine bp conserved nonaner sequence ACAAAAACC. RSS-12 contains a 12 bp spacer, while RSS-23 contains a 23 bp spacer. Efficient recombination, which occurs almost exclusively between RSSs with different length spacers is also known as 12/23 restriction (Sakano et al., 1980; Eastman et al., 1996). A single enzyme complex known as the V(D)J recombinase is required for the rearrangement of the entire seven complex genetic loci encoding antigen receptor chains. The essential components of the V(D)J recombinase machinery include the lymphoid-specific and developmentally regulated proteins RAG-1 (recombination associated gene) and RAG-2, in addition to the DNA-dependent protein kinase (DNAPK), Ku70, Ku80, and the kinase catalytic subunit DNAPKcs, and the XRCC4 and DNA ligase IV protein of immature lymphocytes (Oettinger et al., 1990; Gilfillan, et al., 1993; Fugmann et al., 2000). In the first stage of recombination, the RSSs are recognized by RAG-1 and RAG-2 proteins and form a loop-like complex with RSSs. The RAG-1 and/or RAG-2 proceed to introduce nicks at the heptamer borders to generate double stranded breaks at both ends of the loop causing the DNA to be cleaved precisely between the RSSs and their flanking coding elements (Jackson et al., 1995; McBlane et al., 1995). In the second stage, the two coding ends are processed, which results in a loss and/or an addition of a small number of nucleotides, and joined to form a coding joint. As a result, this imprecise joining of the coding ends provides an important source of receptor diversity (Weis-Garcia et al., 1997; Hesslein & Schatz, 2001).

1.5) Immune diversity:

Confronting the huge array of different antigens in nature, the acquired immune system must generate diversity in order to provide organisms with an effective and specific defense. The large lymphocyte repertoire, expressing Igs and TCRs molecules, constitutes this important role of specific immunity (Davis et al., 1988). The diversity of the specific immune molecules is primarily generated during the rearrangement of different DNA fragments. A few molecular mechanisms are responsible for this process (Tonegawa, 1983). The sources of Ig diversity stem from the random recombination of

any single V, D and J distributed in each gene cluster, the random combination of different heavy and light protein chains, junctional diversity created by variable splicing by recombinases in addition to random insertions of N-nucleotides and the template addition of P nucleotides at junctional regions, and somatic mutation on V gene fragments (Hesslein & Schatz, 2001). In TCR molecules, however, only the first three mechanisms contribute to the formation of immune diversity, whereas somatic mutation on the V region does not (Roitt, 1995). Table 1 presents the theoretical possibility of generating immune diversity in V regions of the Igs and TCRs through the above mechanisms (Abbas et al., 1994). Based on these mechanisms, it is estimated that the mammalian immune system can discriminate at least 10^9 distinct antigenic determinants (Abbas et al., 1994).

2. Discovery and nomenclatures of TCR γ genes:

The mouse TCR γ gene was first discovered in Southern blots in 1984 by Saito et al. The probe they used was the insert of T cell specific clone pHDS58, which was derived from the cDNA in cytotoxic lymphocyte clones and similar to α and β TCR genes. Later, using both of the monoclonal antibodies to the human CD3 and the αβ TCR, a group of T cells were found to be CD3 positive and αβ TCR negative, and were defined as γδ T cells (Lew et al., 1986; Brenner et al., 1986). The γ chain was identified as a part of the γδ heterodimer associated with CD3 on the surface of T lymphocytes (Bank et al., 1986; Weiss et al., 1986b; Moingeon et al., 1986). Although the TCR γ genes were discovered at almost the same time as αβ TCRs (Acuto et al., 1983; Yanagi et al., 1984), knowledge of γδ T cells is relatively limited in comparison with that of αβ T cells, probably because almost 95% of the T cells in the peripheral blood of humans and mice are αβ T cells, while only 5% are γδ T cells (Hass et al., 1993; Kronenberg, 1994). In addition, because the αβ T cells perform the primary functions of the cellular immune system, more emphasis has been focused on the study of the αβ T cells. From 1984 to 1990, research on γδ T cells was mainly focused on the characterization of the molecular structure, gene rearrangements, tissue expression and development. Since the 1990s, most of the effort has been on the precise characterization of physiological and pathological functions, on

Table 1. Calculations of the possible human V gene diversity. The minimum number of specificities is generated by different mechanisms (adapted from Roitt, 1995).

	αβ TCR		γδ TCR		Ig		
	α	β	γ	δ	H	Lκ	Lλ
V gene fragment	50	57	8	3	~100	~70	25
D gene fragment	-	1,1	-	3	~4	-	-
J gene fragment	70	6,7	3,2	3	6	5	8
random recombination	VxJ 50x70	VxDxJ 57(13+7)	VxJ 8x5	VxDxJ 3x3x3	VxDxJ 100x4x6	VxJ 70x5	VxJ 25x8
Total	3500	1140	40	27	2400	350	200
Interchain recombination	3500x1140		40x27		2400x350		2400x200
Total	$4x10^6$		10^3		$9x10^6$		$5x10^6$
Other mechanisms: D in 3 reading frames, junctional diversity, N insertion	$4x10^9$		10^6		$9x10^9$		$5x10^9$
Somatic mutation	-		-		+++		+++

gene regulation and on evolution. These studies showed that γδ T cells share many common features with αβ T cells, but that they also possess numerous unique properties, making γδ T cells a distinctly different cell type (Cheng et al., 1991; Born et al., 1999). These unique features suggest that the γδ T cell population is not just an insignificant functional redundancy of αβ T cells, but rather that it coordinates the interplay between innate immunity and acquired (Mak & Ferrick, 1998).

Although γ and δ gene fragments have been found to be present in many species, from Chondrichthyes cartilaginous fish to primate chimpanzees, most of the studies have been devoted to humans and mice (Marchalonis et al., 1998; Miccoli et al, 2001). In the course of the discovery of TCR γ gene fragments, different nomenclatures have been proposed (Heilig et al., 1987; Garman et al., 1986; Asarnow et al., 1988; Lefranc et al., 1986a; Arden et al., 1995a & b). Table 2 summarizes the three most commonly used nomenclature systems for the V gene fragments of the mouse and human TCR γ. In this thesis, the nomenclature for mouse TCR by Heilig et al. (1986) and the nomenclature of human TCR by Lefranc et al. (1986a) are adopted.

3. The protein structure of TCR γ and δ chains:

The overall structure of γδ T cells resembles that of αβ TCR and Igs. The γ and δ chains that constitute γδ TCR are linked by disulphide bonds. In the manner of αβ TCRs, they possess monovalent binding capacity, which differs from the bivalent binding capacity of Ig molecules. The molecular mass of γδ TCR is between 77 and 90 KD. The γ chain mass is approximately 31 to 47 KD depending on the specific constant gene used, whereas the δ chain mass is between 45 and 48 KD (Weiss et al., 1986a; Cron et al., 1990).

The V domains of both the γ and δ chains form the antigen binding structure, which is similar to αβ TCRs and Igs. In these regions, there are three CDRs and four flanking framework regions (FR) (Hein, 1994). The CDRs are highly variable, whereas the FRs are relatively constant. CDR1 and CDR2 are encoded within the V gene, while CDR3 is

Table 2. Nomenclatures of the mouse and human Vγ genes.

Mouse			Human	
Garman et al. (1986)	Heilig and Tonegawa (1986)	WHO-IUIS (1995)	LeFranc et al. (1986)	WHO-IUIS (1995)
V1.1	V1	GV1S1	V1	GV1S1P
			V2	GV1S2
V1.2	V2	GV2S1	V3	GV1S3
			V4	GV1S4
V1.3	V3	GV3S1	V5	GV1S5
			V5P	GV1S5P
V2	V4	GV4S!	V6	GV1S6P
			V7	GV1S7P
V3	V5	GV5S1	V8	GV1S8
			VA	GV5S1P
V4	V6	GV5S2	V9	GV2S1
			V10	GV3S1P
V5	V7	GV5S3	VB	GV6S1P
			V11	GV4S1P

encoded by the V-J or V-D-J junctional region. Among the three CDRs, CDR3 is the most diverse (Davis et al., 1988; Schiffer et al., 1986). Structural comparison indicates that there is extensive similarity among the three types of specific immune molecules, αβ TCRs, γδ TCRs and Igs. For example, the length of CDR3 from the δ chain is significantly longer than that of the γ chain, similar to IgH chains which are longer than IgL chains (Hein, 1994). However, there is also sequence similarity between Vγ and Vβ and between Vδ and Vα, which clearly indicates a relationship to the two TCRs (Hayday, 2000). Therefore, it is difficult to determine the evolutionary relationship among these three types of molecules simply using structural similarity as a gauge. In humans, the majority of the peripheral γδ T cells are Vγ9Vδ2 positive and tend to recognize phosphoantigens (Miyagawa et al., 2001). Recently, the crystallographic structure of the human Vγ9Vδ2 TCR has been characterized. CDRs of the V domains reveal that the CDR1δ, CDR1γ, CDR2γ and CDR3 loops form a pocket between them. The chemical environment in the pocket formed by the unique composition of amino acid residues exhibits a chemically reasonable binding site for phosphorylated antigens, providing a possible explanation for the canonical usage of the Vγ9 and Vδ2 gene segments by phosphoantigen-reactive receptors (Allison et al., 2001).

The constant region of TCR γ chains is composed of an immunoglobulin-like region, a hinge region, a transmembrane (TM) region, and a short cytoplasmic region (Marchalonis et al., 1998). The hinge region is the most variable and has a different length. There is a cysteine residue in the hinge which forms a disulfide bond with the TCR δ chain (Hein, 1994). In the mammalian γδ TCRs identified so far, Cγ chains differ greatly in their length of hinge, their number of cysteine residues and their sites for N-glycosylation. For example, the mouse Cγ1 is glycosylated, whereas the Cγ2 chain is not glycosylated. The hinge of Cγ4, which is encoded by two exons, is 18 amino acid residues longer than those of the other three Cγ chains (Vernooij et al., 1993). In addition, the crystallographic structure of γδ TCR reveals that the C domains are markedly different from those of αβ TCRs although they are similar in overall structure at the V domains. For example, Cδ is composed of a regular immunoglobulin-like domain with a regular three-stranded β-sheet

as its outer face, while Cα contains an unusual secondary structure instead of a regular outer β-sheet. All of these structural differences may enable γδ TCRs to form recognition/signaling complexes that are different from αβ TCRs (Allison et al., 2001).

4. Development of γδ T cells:

Similar to αβ T cells, γδ T cells mature predominantly in the thymus. However, it was found that a portion of γδ T cells also develops in extrathymic regions (Lambolez et al., 2002). In the thymus, the development of T cells mainly involves two processes. First, each stem cell differentiates into either the αβ or γδ T cell, in a process termed T cell lineage selection. Second, those T cells which become γδ subtypes experience thymic selection, during which the precursor cells gradually become mature T cells that are capable of responding to foreign antigens (Robey & Fowlkes, 1994; Bluestone et al., 1995).

4.1) αβ and γδ lineage determination:

Precursor T cells have not yet expressed the important differentiation surface markers such as CD4 and CD8, when entering the thymus. Under the effects of thymic hormones and other factors, these cells continue to differentiate to a point from which they are able to diverge into either the αβ or γδ T cell lineage (Fehling et al., 1999). First, experiments revealed that the common precursor cell at the divergent point did not possess CD3, CD4 or CD8 markers because both the αβ and γδ T cell types could differentiate from the developmentally advanced TN (triple negative, CD3-CD4-CD8-) population, which has lost the ability to develop into B, NK (natural killing) cells and thymic dendrite cells. Further study showed that when a series of developmentally consecutive immature TN thymocytes, such as CD25-CD44+TN, CD25+CD44+TN, CD25+CD44-/lowTN (-/low is weakly positive) and CD25-CD44-/lowTN, were injected intravenously and intrathymicly, only the first three subsets generated both αβ and γδ T cells. Therefore, the divergent point might reside within the fourth pre-T-cell line CD25+CD44-/lowTN (Shortman et al., 1991; Suda & Zlotnik, 1993). The fact that the CD25+CD44-/lowTN is the first subset in which the rearrangements of α, β and γ were detected significantly,

further supports the above point (Dudley et al., 1995). However, more tangible evidence is still needed in order to define the precise precursor cell at the branch point.

The mechanism underlying the lineage separation between αβ and γδ T cells in the thymus is less well understood although there are a few models which have attempted an explanation (Fehling et al., 1999). The "sequential rearrangement model" proposes that the αβ and γδ lineage determination may be influenced by the rearrangement of TCR genes. The common precursor cells may first attempt to generate functional γ and δ rearrangements. If the rearrangements on both chromosomes are successful, the precursor cells will be forced to develop into the γδ lineage. However, those cells with unsuccessful rearrangements of γ and δ genes will generate αβ TCR (Pardoll et al., 1987; Dudley et al., 1995). This model uses the fact that γ, δ and β genes are rearranged significantly before α in the fetal thymus (Fowlkes et al., 1989) and that the first γδ T cells appear around 14 days, but that the first αβ T cells appear in the thymus between 17 and 18 days (Hedrick et al., 1993). In the "competitive rearrangement model", the γ, δ and β genes are believed to compete with one another to form the receptors. If the γ and δ genes are initially arranged in a productive fashion, the cells will proceed to the γδ lineage; but if a functional β gene is assembled first, the cell will follow the αβ pathway. The support stems from findings showing that the rearrangements of γ, β and δ take place almost at the same developmental stage, but without a temporal sequence. Therefore, lineage selection may be determined by competition between the three types of genes (Godfrey et al., 1994; Petrie et al., 1995). Further support for this hypothesis emanates from the somatic cell hybrids generated by the fusion of γδ+ and αβ+ thymocytes, in which the TCR γ gene is extinguished, suggesting the existence of competition among genes. The third model, the "separate lineage model," proposes that the outcome of TCR gene rearrangement is completely irrelevant to the lineage decision. Thus, the commitment to the αβ and γδ pathways is manifested by some unknown mechanism. This model is not satisfying because it does not explain how a lineage decision is achieved (Winoto & Baltimore, 1989a). The last model, "maintenance of the lineage decision," emphasizes how a determined lineage maintains its lineage status. One proposal states that T cells

committed to the αβ lineage activate the silencers of TCR γ to inhibit the expression of γδ TCRs (Ishida et al., 1990). A second proposal states that in cells which follow the αβ pathway, a programmed excision event is triggered to delete the Dδ, Jδ and Cδ regions, thereby permanently preventing the formation of the δ chain (de Villartay & Cohen, 1990).

4.2) Thymic selection:

Once the potent progenitor cells have chosen the γδ T cell route, these γδ T cells will undergo a series of screening steps in the cortex of the thymus based on their reactivity to self-antigens and/or MHCs. The vast majority of thymocytes that are not selected will die *in situ* via apoptosis (Rodewald et al., 1999).

It is clear that γδ T cells undergo negative selection in the thymus during their maturation (Dent et al., 1990). Experiments suggest that potentially self-reactive γδ T cells are negatively selected in the thymus. When transgenic mice containing γδ TCR specific to T10b, a MHC-I-like molecule expressed in the H-2 region, are bred to mice with H-2b MHC, the self-reactive transgenic γδ T cells are largely deleted from the thymus and spleen. Further histological evidence has shown that these potential self-reactive γδ thymocytes in H-2b/d transgenic mice become large and down-regulate the T cell receptor before being clonally deleted via programmed cell death (Dent et al., 1990, 1993).

The model for positive selection for MHC, however, has not been generally accepted (Robey et al., 1998; Hayday, 2000). This is partly because it has not been clearly defined whether γδ T cells command a similar mechanism for recognizing the ligand/MHC complex as αβ T cells. In mice lacking β2-microglobulin (β2-M), CD8+ αβ T cells fail to develop because they are not positively selected without MHC class I expression. However, the same mice show a normal distribution of γδ T cells in the thymus, peripheral lymphoid tissues, and intraepithelial locations (Zijlstra et al., 1990). Likewise, normal γδ T cells were found in double mutant mice which lacked classical MHC I and II

expression (Grusby et al., 1993). However, in mice expressing transgenic γδ TCRs with specificity for nonclassical class I molecules Ib, encoded in the Tla region, β2-M expression was required for the normal development of transgenic-expressing γδ T cells (Wells et al., 1991). Such discoveries suggest that the majority of γδ T cell are able to develop normally without MHC class I/II dependent selection, but that a minority require the nonclassical MHC for their development (Schweighoffer et al., 1996). A possible explanation is that the positive selection for γδ T cells could be exceptional, only occurring among those γδ T cells whose TCRs possess αβ T cell-like specificities.

In addition to the selection of MHC-reactive cells, the thymus has been suggested to be involved in the selection of some specific γδ T cell phenotypes. An important feature of the mouse γδ T cells is that a few subtypes express invariant TCRs which equally recognize monomorphic ligands, probably self-proteins, which suggests that the maintenance of a particular junctional sequence may be driven by a thymic selection mechanism. Supporting evidence emerged in early experiments examining the invariant junctional sequences of the mice Vγ5/Vδ1 and Vγ6/Vδ1 subsets. For example, it was found that normal development of the Vγ5+ subset required a fetal thymic environment, which is consistent with thymic selection (Ikuta et al. 1990). However, in mice lacking TCR δ, thus deleting the possibility of selection for certain surface-expressed protein products by the thymus, the junctions of a significant portion of Vγ5-Jγ1 and Vγ6-Jγ1 rearrangements are still invariant. This finding reveals that the recombinase may also play a role in the determination of invariant junctions (Itohara et al., 1993). The above seemingly contradictory findings can be easily reconciled by assuming that the invariant TCR repertoire is possibly the result of two processes: a) the recombination machinery generating a series of biased gene rearrangements; and b) the thymus selecting the cells bearing TCRs with the respective invariant idiotype containing the particular conformation for the monomorphic ligand.

4.3) Extrathymic differentiation:

The γδ T cells in the intestine are thought to develop locally because a large number of intestinal γδ T cells have been identified in the intestinal epithelia of athymic nude mice in addition to thymectomized mice reconstituted with bone marrow (Racha et al., 1994; Suzuki et al., 2000). This hypothesis is confirmed by studies showing that the stem cells of the bone marrow home directly to the gut cryptopatches, by-passing the thymus, and developing into a population of precursor cells that give rise to γδ T cells (Saito et al., 1998). Consistent with the notion of the extrathymic development of T cells, expression of RAG and V(D)J rearrangement was identified in these precursor cells (Lambolez et al., 2002). The intestinal epithelium may play a similar role in the development of intestinal γδ T cells similar to the role the thymic epithelium plays in intrathymic T cell development (Guy-Grand et al., 1991). The intestinal differentiation is IL-7 dependent because gut cryptopatches from IL-7-/- mice did not support γδ T cell development (Kanamori et al., 1996).

5. Tissue distribution of γδ T cells:

Unlike αβ T cells, which predominately circulate in the peripheral blood, some subsets of γδ T cells are distributed in distinct anatomical locations. The disproportionate abundance of these T cells in specific tissues is correlated to their specific functions. This property is one of the essential factors that render γδ T cells a unique component in the immune system (Mak & Ferrick, 1998).

5.1) γδ T cell tissue distribution in mice:

Similar to αβ T cells, the majority of γδ T cells leave the thymus and colonize in distinct tissues once they become mature (Table 3A). The γδ T cells can be detected in the thymus around day 13 of gestation, three days before αβ T cells are found. During their maturation in the thymus, three subtypes of γδ T cells, Vγ5+, Vγ6+ and Vγ4+ cells, exhibit a waved pattern of expression (Figure 3) (Haas et al., 1993). The first wave is the Vγ5+ T cell population which appears in the fetal thymus around day 14 and then declines at a rate making it difficult to detect by day 18. Around day 16, the second wave, Vγ6 bearing T cells, are present. Immediately following the Vγ6 T cells, Vγ4 bearing T

Table 3A. Characteristics of the mouse γδ T cells.

Subset	Location	TCR usage	Diversity
Vγ5	Skin	V5J1C1-γ V1D2J1C-δ	None
Vγ6	Vagina, uterus, tongue	V6J1C1-γ V1D2J1C-δ	None
Vγ1	Spleen, intestine, skin	V1J4C4-γ V6,4,5,7-δ	High
Vγ4	Blood, lymph nodes, spleen	V4J1C1-γ V5,4,6,7-δ	High
	Intestine	V7J1C1-γ V4,5,6,7-δ	High
	Lung	Vγ4 Vδ6	High
	Liver	V1,2-γ Vδ6	High
	Mammary gland	V4, 5-γ Vδ4	High

Table 3B. Characteristics of the human γδ T cells.

Subset	Location	TCR usage	Diversity
Vδ1	Thymus	V1C2-γ Vδ1 Most non S-S	High
Vδ2	Blood	V2C1-γ Vδ3 Most S-S	High

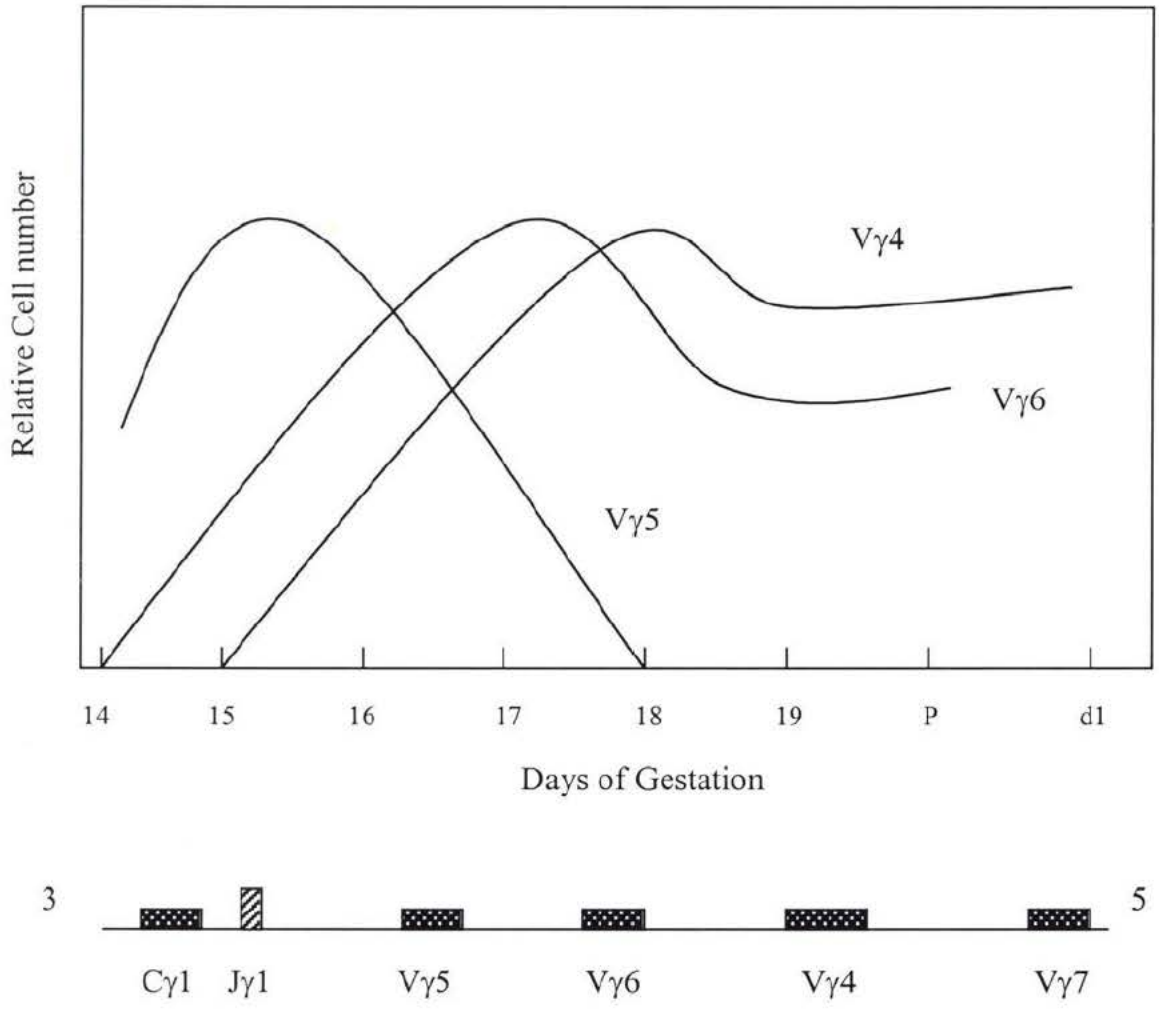


Figure 3. The order of expression of the mouse Vγ genes in the fetal thymus parallels the gene order in the γ locus. T cells bearing the mouse V genes in the cluster 1 show a waved pattern of expression in the fetal thymus. This sequential expression parallels the gene order on the chromosome (adapted from Vernooij, 1992).

cells with more TCR diversity than the first two subtypes start to appear in the thymus. The mechanism underlying this sequential expression is still poorly understood. However, it has been suggested that it is controlled by regulating the V(D)J recombination of the TCR γ locus and is predetermined at the level of the hematopoietic stem cell (Ye et al., 2001).

a) Vγ5:

The Vγ5 bearing T cells reside exclusively in the epidermis. The fact that both the Vγ5+ fetal thymocytes and the epidermal T cells have the same invariant junctional sequence of γ and δ chains suggests that Vγ5 bearing T cells are originated from the thymus rather than from extrathymic locations (Lafaille et al., 1989; Asarnow et al., 1988). The TCRs on these T cells exclusively employ the Vγ5Jγ1Cγ1 and Vδ1Dδ2Jδ2Cδ chains (Havran et al., 1988; Ito et al., 1989). The T cells express Thy and CD3 markers on their cell surface, just as do the mature T cells in the peripheral blood, but they do not contain CD4/CD8 molecules. Vγ5+ T cells exhibit dendritic morphology (dendritic epithelial cell-DEC). This cell population is the predominant type of lymphocytes in the outer layer of the skin and distributes around the keratinocytes (Elbe et al., 1989). The invariant TCRs and specific localization suggest that the Vγ5+ T cells play an important role in skin wound healing. In the skin, when the keratinocytes are damaged by infection or other environmental factors, the Vγ5+ T cells express keratinocyte growth factors (KGFs) in addition to expressing cytokines such as IL-2 and IFN-γ by recognizing antigens on these cells with invariant TCRs. Meanwhile, the CD8+ T cells are also recruited. The Vγ5+ T cells will then destroy the infected keratinocyes. KGFs and cytokines activate keratinocytes to proliferate and reepithelize, thus the wounded tissue is repaired (Havran et al., 1991; Tigelaar et al., 1995; Jameson et al., 2002).

b) Vγ6:

Vγ6 bearing γδ T cells appear in the thymus as another invariant subtype. The Vδ gene used is Vδ1, which joins together with Dδ2 and Jδ2. There are no CD4, CD8 and Thy-1 markers expressed on the cell surface (Ito et al., 1989). Normally, these γδ T cells home

to and reside in the mucosal epithelia of the tongue, vagina, uterus and lung (Hayes et al., 1996; Sim et al., 1994). Almost half of the T cells in the mucosal epithelia of these organs are γδ T cells, while Vγ6 positive T cells form the majority. Different from Vγ5+ T cells, Vγ6+ T cells are not confined to the tissues they colonize during ontogeny. In pregnancy, they proliferate in large numbers and are often activated in the placenta (Heyborne et al., 1992). In diseases associated with strong inflammatory responses, Vγ6+ T cells appear at the regions of inflammation, including the liver and testis (Mukasa et al., 1997; Roark et al., 1996) where they are normally infrequent or even absent.

c) Vγ4:

On the newly generated Vγ4+ cells, the TCRs are invariant in the first two weeks, but are gradually replaced by TCRs with limited junctional diversity (Sim et al., 1994). These cells circulate and appear in various lymphoid tissues such as the spleen and the lymph nodes, and have been detected in the lung and the lactating gland (Augustin et al., 1989; Reardon et al., 1990). Those Vγ4+ cells expressed in the epithelia of the lung tend to use Vδ5, while the cells expressed in lymphoid tissues pair with Vδ4, 5, and 7. In the lung, the primary γδ T cells in the early fetus are Vγ6+, but as one's age increases, Vγ4+ cells replace them. In addition, the number of Vγ4/Vδ4 T cells differs widely in a variety of inbred mice (Sperling et al., 1992). For example, the percentage of splenic Vγ4 cells in the B6 strains of mice is quite high (over 50%), whereas in other strains of mice such as the DBA/2, the percentage half of that, suggests Vγ4 subtypes in these mice strains are selected differently, by either the thymus or other mechanisms (Sim et al., 1990).

d) Vγ1:

Varying from the above three subtypes, the junctions of the Vγ1 TCRs are highly diverse. The Vδ chains that Vγ1 pairs with include Vδ2, 3, 5, 7 and 8. There are two portions of Vγ1+ cells, the thymus dependent and the thymus independent (Kaufmann, 1996). The thymus dependent Vγ1+ cells arise during later stages of fetus growth and continue to mature in the adult thymus (O'Brien et al., 1989 & 1992). The cells distribute in a broad range of tissues such as that of a newborn mouse's thymus and spleen, adult

spleen and liver, placenta, and as rare clones in skin and the lactating mammary gland (Happ et al., 1989; Nagler-Anderson et al., 1992; O'Brien et al., 1989, 1992; Reardon et al., 1990, Roberts et al., 1991). The extrathymically developed T cells preferentially localize in the intestinal epithelia (Pereira et al., 1995). Hybridomas expressing Vγ1 TCRs can be stimulated by a variety of ligands such as bacteria and bacterial extracts (Born et al., 1999), which suggests that Vγ1 bearing T cells play an important function against microbial infections. In fact, Vγ1+ T cells were found to provide protection against a variety of diseases such as viral and bacterial infections, as well as tumors (Fu et al., 1994a; Penninger et al., 1995; Nakamura et al., 1999). For example, it has been reported that γδ T cells play a protective role in the early phase of murine cytomegalovirus infection, as 90% of the γδ T cells are Vγ1-bearing T cells (Ninomiya et al., 2000).

e) Vγ7:

Vγ7 pairs with Vδ4, 5, 6 or 7, while the overall diversity of these TCRs is very high. Vγ7 bearing T cells primarily colonize the epithelia of the intestines and they are a predominant phenotype in the i-IEL (intestine intraepithelial lymphocytes) along with Vγ1+ cells (Arstila et al., 2000). In addition, a portion of T cells of this subtype are also found in the liver (Asarnow et al., 1989; Itohara et al., 1990; Tsuji et al., 1996). The Vγ7+ T cells differ from most other γδ T cells by their expression of CD8αα homodimers and rather variable levels of surface Thy-1 markers (Goodman et al., 1988; Lefrancois et al., 1989). Some experimental evidence shows that Vγ7+ cells can develop independently of the thymus (Bandeira et al., 1991; Pereira et al., 1995). However, other experiments indicate that at least a portion of these cells are normally derived from thymic emigrants (Kaufmann, 1996). Therefore, like Vγ1 bearing cells, Vγ7+ T cells also contain both thymus dependent and independent subtypes.

5.2) Tissue distribution of γδ T cells in humans:

Although the tissue distribution of γδ T cells in humans has been studied more than that of mice, limited experiments have shown that tissue specific distribution is also

evident in humans (Table 3B) (Porcelli et al., 1991; Machugh et al., 1997). However, at variance with the prominent invariant junctions of γδ TCRs in mice, the majority of V(D)J junctions of the γ and δ genes in human γδ TCRs show considerable diversity (Ullrich et al., 1990; Jarry et al., 1990; Deusch et al., 1991). In peripheral blood, humans display a close percentage of γδ T cells to that of mice (Porcelli et al., 1991). It was found that the relative percentage of the two predominant subsets, Vγ9/Vδ2 and Vγ9/Vδ1, changed with the development of the individual. Vγ9/Vδ2 occupies 25% in umbilical cord blood and increases to 70% in adult peripheral blood (Parker et al., 1990). On the contrary, Vγ9/Vδ1 is more plentiful in the early stage of development but subsequently decreases. In the intestine, although a small portion of i-IEL is γδ T cells and the γδ:αβ ratio is 1:5, primarily Vδ1 and Vδ3 and GV1 family members are utilized, which suggests some kind of preferential localization (Ullrich et al., 1990; Deusch et al., 1991; Soderstrom et al., 1996; De Libero, 2000). Different subtypes of γδ T cells are evenly distributed in the spleen and the lymph nodes except that the γδ:αβ ratio in the lymph nodes is ~1:50 (Groh et al., 1989). In human skin and mucosa, γδ T cells are only a small percentage whereas αβ T cells are the predominant lymphocytes. Even so, the percentage of epithelial γδ T cells in the total T cells is still larger than that found in peripheral blood and lymphoid organs (Hass et al., 1993). Very few γδ T cells are found in organs such as the liver, kidney, salivary gland and lung (Falini et al., 1989; Richmond et al., 1993).

5.3) γδ T cell distribution in other vertebrates:

The tissue specific expression of γδ T cell subsets is also evident in other species but the patterns of distribution are diverse. In ruminants such as sheep and cattle, the portion of γδ T cells in the peripheral blood consists of 60% of the total T lymphocytes, which contrasts with the only 5 to 10% attributed to humans and mice. Similar to mice, there are a large number of γδ T cells in the epithelia of the gut and tongue region. However, there is only a small population of γδ T cells in the peripheral lymphoid tissues (Hein et al., 1991; Wilson et al., 1999). In birds such as chickens, there are a large number of γδ T cells. The γδ T cells in the blood and spleen occupy 20% to 50% of the total lymphocyte population, while such cells are also predominant in the epithelia of the intestines (Bucy

et al., 1988; Cooper et al., 1989). Studies on the identification of the biological significance of the differential tissue distribution in different species are only beginning.

6. Recognition and activation of γδ T cells:

6.1) Antigens for γδ T cells:

One critical setback for integrating γδ T cells into the specific immune system is definition of what ligands the γδ T cells can recognize, particularly as their chemical nature and immune responsive features are poorly understood (Born et al., 1999; Richards & Nelson, 2000). A variety of antigens have been tested to screen the antigen specific responses *in vitro*. Both peptidic and nonpeptidic antigens were used to stimulate γδ T cells and hybridomas randomly. A multitude of responses transpired. Among these antigens, some have been molecularly defined whereas others have only been partially defined. For example, the molecularly defined soluble antigens, tetanus toxoid, mycobacterial 60-kDa heat shock protein (Hsp-60), Hsp-60-derived peptides, staphyloccocal enterotoxin A, listeriolysin O and lipopolysaccharides were found to stimulate γδ T cell responses *in vitro* (Hass et al., 1993; Born et al., 1999). In the case of partially defined antigens, it has been found that heat-killed bacteria, bacterial extracts, mycobacterial purified protein derivative, low molecular weight protease-resistant components of mycobacterial extracts, and poly (GT) can stimulate responses of γδ T cells *in vitro* in both mice and humans (Dembic et al., 1990; Holoshitz et al., 1989; Kabelitz et al., 1990; O Brien et al., 1989; Panchamoorthy et al., 1991; Pfeffer et al., 1990 & 1992). Such tests prove that the antigens for human γδ T cells are different from those of mice although they are both MHC independent. Human peripheral blood γδ T cells recognize low molecular weight non-peptidic phosphoantigens and the intraepithelial cells recognize MHC class I related antigens. However, the mouse blood γδ T cells directly recognize large protein antigens such as non-classical MHC class I molecules rather than non-peptidic antigens (Morita et al., 1995; Hayday, 2000).

Among the known ligands of γδ T cells, phospholigands have attracted extensive attention and are the most clearly defined antigens recognized by human γδ T cells

(Hayday, 2000). When isolating and characterizing the antigen from mycobacteria, phosphoantigens were found to be the antigens responsible for γδ T cell expansion. The cell subtype which responds to the antigens is mainly the human Vγ9Vδ2 T cells, which expand 2 to 10 fold (8%-60% of all circulating T cells) during microbial infections (Bukowski et al., 1999). Recently, *E. coli* was also found to activate human Vγ9Vδ2 T cells (Feurle et al., 2002). Some phosphate-containing antigens can elicit proliferative and cytokine responses of γδ T cells rapidly at very low concentrations. These immune responses are stimulated directly because accessory cells such as APCs are not strictly required (Tanaka et al., 1995). Site mutagenesis of the junctional regions of the TCR γ chain shows that γδ T cell reaction to phosphate antigens is dependent upon the junctional region of the TCR γ chains and upon pairing of Vγ9 and Vδ2 TCR chains (Bukowski et al., 1998). The positively charged lysine residues in the TCR-γ chain CDR3 region encoded by the germline J1 gene play a crucial role in the response of Vγ9Vδ2 T cells to the diverse small molecular mass nonpeptide antigens (Miyagawa et al., 2001). Further analysis of the chemical structure of the nonpeptide components extracted from mycobacteria suggests that such cells possess low molecular mass and contain phosphate groups but other structures of these non-peptide components are diverse, ranging from nucleotide derivatives to isoprenyls and sugars (Schoel et al., 1994). All of these antigens elicit γδ T cell subset specific responses. For example, human Vγ9Vδ2 γδ T cells recognize isopentenyl pyrophosphate in mycobacteria, while the Vγ1Vδ1 bearing T cells do not (Tanaka et al., 1995). The cleavage of the phosphate group completely abolishes the stimulatory activity on the human Vγ9Vδ2 γδ T cells. Therefore, the unique susceptibility of γδ T cells to phosphate antigens could be important to the selective induction of γδ T cells, which may be pertinent to their prominent immunoregulatory function by determining the flavor of their regulatory activities (Burk et al., 1997). Since the phosphoantigens are widely distributed in the biosphere as well as among bacteria and more advanced organisms, the reactivites of γδ T cells toward these molecules are likely very wide, perhaps extending to the recognition of Gram-positive and Gram-negative bacteria as well as certain parasites (Hayday, 2000).

6.2) Antigen recognition features of γδ T cells:

The trimolecular complex of TCR, foreign peptide and self-MHC has been viewed as one of the central laws of cellular immunology (Kaufmann et al., 1996). However, although αβ and γδ TCRs are similar in overall structure, there are some fundamental differences in their antigen recognition. First, the crystallographic analysis of γδ TCR has suggested that the recognition of antigens by γδ TCR resembles that of Ig rather than αβ TCR (Li et al., 1998; Allison et al., 2001). Second, αβ and γδ T cells possess both the same and different surface differentiation markers. For example, like αβ T cells, γδ T cells are also associated with the CD3 complex, and the activation of γδ T cells requires both TCR and CD3 (Hass et al., 1993). However, the majority of γδ T cells do not express CD4 and CD8, which are required for binding αβ TCR to MHC molecules (Hass et al., 1993; Schild et al., 1994). Only a small portion of γδ T cells in the intestine express CD4 or CD8 (Goodman et al., 1988). Third, the most striking difference between αβ and γδ T cells is the role of MHC in the recognition of antigens. Until now, most studies showed that antigens recognized by γδ T cells were not dependent on MHC and were not processed by APCs (Hayday, 2000). For instance, a herpes virus protein was found to be recognized directly by a human Vγ9+ γδ T cell clone (Sciammas et al., 1994). Therefore, the antigen/γδ TCR recognition is similar to the interactions between antigens and antibodies. However, a few studies have revealed that γδ T cells are under the control of MHC-like proteins such as Qa-1, which is located in the H-2T locus in mice (Porcelli, 1995). Unlike the small, processed peptide presented to αβ T cells, the antigens presented by Qa-1 are not processed and therefore the MHC-like proteins seem to present large polypeptides. For example, it has been found that Hsp60 preferentially binds to the Qa-1 molecule. A third pattern of recognition has also been proposed, wherein some γδ T cells were suspected to perform antigen recognition in combination with other molecules other than MHC or MHC-like proteins. One of these candidates is the NRAMP (natural resistance-associated macrophage proteins) gene product that is believed to transport intracellular antigens to the cell surface. It has been found that γδ T cell surveillance of hematpoietic tumors is NRAMP-dependent instead of MHC-dependent (Penninger et al., 1995).

7. Functions of γδ T cells:

Until now, nearly all the functions of αβ T cells have been described in γδ T cells (Born et al., 1999). These include both the immunoregulatory and immunoprotective functions. However, experiments have shown that immune resistance to antigens is not prominent compared to the immune regulatory functions except for the characteristic immune surveillance of epithelium (Hayday, 2000). With γδ T cells challenged by pathogens, there is only a small resistance change. In addition, these experiments are usually carried out in immune compromised animals or in γδ T cells being challenged with a heavy load of pathogens. On the other hand, with the disturbance of γδ T cells, a dramatic change has transpired in the immune response involving both innate and acquired immunity. The small percentage of γδ T cells in the peripheral blood determines that they cannot play a major role in immune protection. This percentage is consistent with that of many regulatory T cells such as CD4+CD25+ and αβ TCR+CD4-CD8- T cells (Zhang et al., 2001), plus their unique anatomical distribution, suggests that γδ T cells are important in immunoregulation and epithelial repair rather than immunoprotection.

7.1) The immunoregulatory functions of γδ T cells:

T cells exert their immune regulatory functions on the innate, and the specific humoral immunity and cell mediated immunities (Mak & Ferrick, 1998). These effects are dependent on the effective communication and interaction among B cells, T cells, macrophages and NK cells, including the secretion of specific proliferation or differentiation lymphokines (Hayday, 2000).

Numerous evidences have shown that γδ T cells influence antibody production. When co-cultured with the surface-Ig positive mouse lymphoma CH12, a population of Thy-1+, CD4-CD8-, and B220+ γδ T cells can promote lymphoma Ig secretion (Sperling et al., 1989). Further, when the isolated γδ T cells from chicken ovalbumin-immunized mice were transferred into normal mice, the IgE production was inhibited very efficiently

(McMenamin et al., 1994). Some experiments have also indicated that γδ T cells can induce Ig isotype switching in B cells. Mice deficient with αβ T cells possess entire Ig isotypes with a relatively higher level of IgG1 and IgE. However, in the αβ and γδ T cell double deficient mouse, all the Ig isotypes are lacking (Wen et al., 1994). Such findings provide convincing evidence that γδ T cells play an important role in humoral immune response.

Studies suggest as well that γδ T cells can influence the development of αβ T cells and the immune responses of the peripheral αβ T effector cells. In mice injected with anti TCR-δ mAb, which results in the disappearance of γδ T cells in lymph nodes and spleen, the reactivity of αβ T cells increased (Kaufmann et al., 1993). The increased αβ T cell functions may suggest that γδ T cells have the ability to control αβ T cell response negatively under normal circumstances. In contrast to the primarily inhibitory effect of γδ T cells on αβ T cells, αβ T cells were found to stimulate response in the γδ T cell positively. It has been observed that peritoneal γδ T cells from *Listeria*-immune mice show an enhanced potential to expand when αβ T cells are activated *in vitro* (Skeen et al., 1993). The stimulation of γδ T cells by αβ T cells was abrogated by a protein-permeable semipermeable membrane, which excludes cytokines as the sole cause (Vila et al., 1995). Therefore, the influence between γδ and αβ T cells is probably through cell-to-cell interaction.

In addition to the regulation of B and T cells, γδ T cells also show some functional relationship with macrophages and natural killer (NK) cells. Macrophages from γδ T cell deficient mice stop secreting their major lymphokine TNF-α in response to lipopolysaccharides. This is not an intrinsic defect of macrophages because preincubating them with γδ T cells from normal mice can restore their capability to secrete TNF-α (Nishimura et al., 1995). The effect of γδ T cells on macrophages is partially due to the production of IFN-γ from γδ T cells because the IFN-γ mAb inhibits this effect. Although the direct interaction between γδ T cells and natural killer cells is not supported, functional relationships between them have been described in the observation of an

infectious disease. In a TCR δ-/- mouse model given the *Listeria monocytogenes* infection, it was found that INF-γ production by NK cells was impaired in comparison with normal mice (Ladel et al., 1996). This is probably a defect caused by the reduced production of TNF-α in the absence of γδ T cells. The response to listeria infection involves two phases, with macrophages and NK cells critical in the first phase and αβ T cells responsible for the second phase of clearance (King et al., 1999). The above finding may suggest that γδ T cells provide a link between the innate and the adaptive immune responses against murine listeriosis, regulating both early NK and late αβ T-cell reactivities (Born et al., 1999).

7.2) Resistance to pathogens:

In the immune system, the importance and precise role of γδ T cells to the resistance of pathogens remains to be determined. γδ T cells increased in a variety of infectious diseases, such as tuberculosis, salmonellosis, tularemia, brucellosis, malaria, toxoplasmosis, HIV, listeria, and Herpes simplex virus type 1, among others. Diverse methods including direct infection, blocking γδ T cells with monoclonal antibodies (mAbs), and knock-out studies have been carried out. Immune responses to pathogens have been identified in almost all the diseases tested, but the direct contribution of γδ T cells to host resistance is prominent only in a few immune compromised animal models. Based on these results, it was concluded that the major contribution of γδ T cells in the resistance to pathogens is through the primary immune response in collaboration with cells of the innate immune system. Some of the responses are essential to immune protection, whereas others are not essential and are functional redundancy of αβ T cells (Hass et al., 1993; Hayday, 2000).

One of the best-studied examples of resistance to pathogens by γδ T cells is the infection by the gram-positive rod *Listeria monocytogenes*, in which γδ T cells reveal a substantial protective role. In mice, either the TCR δ or TCR β gene deletion mutants were resistant to experimental listeriosis. Additional treatment of TCR β-deficient mutants with TCR γδ mABs rendered these mice susceptible to listeriosis (Mombaerts et

al., 1993). Further studies on mice with depleted Vγ1+ T cells show that the γδ T cells used are predominantly Vγ1Vδ6 positive and the response to *Listeria monocytogenes* is through IFN-γ (Harty & Bevan, 1995; Nakamura et al., 1999). These findings suggest that γδ T cells participate in protective antimicrobial immunity at an early phase of the primary infection before the αβ T cells appear. In addition, the critical protective role of γδ T cells is also identified in malaria infections (Tsuji et al., 1994). In αβ T cell deficient mice, the immunization of irradiation-inactivated sporozoites induces an immune response that significantly inhibits the development of the parasite's liver stages. However, the malaria is exacerbated by the antibody-mediated transient *in vivo* depletion of γδ T cells (van der Heyde et al., 1993). This finding is regarded as evidence that γδ T cells are involved in protective immunity against liver stages of malaria parasites without a need for αβ T cells.

In addition, responses to γδ T cells were also examined in viruses. Although direct evidence that γδ T cells react to virus infected cells has not been found, accumulation of γδ T cells at the sites where microbes replicate as well as selective γδ T cell expansion in the peripheral blood and lymphoid organs have been observed in numerous murine models (Hayday, 2000). For example, γδ T cells appear to have protective effects in the herpes simplex virus (HSV-1) infected mice (Sciammas et al., 1997). When stimulated by HSV-1, both human and mouse γδ T cells lyse the infected cells. But in mice lacking γδ T cells, the severity of HSV-induced epithelial lesions was much increased. Thus, by arresting viral replication, γδ T cells protect individuals from HSV-1 infection (Ninomiya et al., 2000).

7.3) Inflammatory reaction:

With infection and injury, inflammation helps living organisms restore their normal structures and functions. The cellular response of leukocytes plays an important role in this process. In healthy immunocompetent mice, after intravenous or peritoneal inoculation of *Listeria monocytogenes*, the infected intracellular bacterium is cleared within 7-10 days following a vigorous inflammatory response and subsequent

development of listeria-specific cytolytic αβ T cells. Some studies suggest that γδ T cells are involved in the inflammatory resolution. In the infected mice, there are small granulomatous lesions detected around the infected hepatocytes in the liver, while γδ T cells appear in the lesion about the same time as αβ T cells. However, if γδ T cells do not appear, the resolution of inflammation is severely impaired and this results in serious parenchymal damage. Following infection, αβ and γδ T cell-depleted mice each showed exacerbated liver lesions (Fu et al., 1994a; Mombaerts et al., 1993). However, lesions in the αβ T cell-depleted mice were rich in bacteria and almost devoid of neutrophils, whereas the lesions in γδ cell-depleted mice were abscessed and rich in polymorphonuclear leukocytes. Circles of necrotic hepatocytes infiltrated by neutrophils surrounded the abscesses, and there was extensive parenchymal damage. In the whole body, the γδ-deficient mice did not show excessive bacterial accumulation, which is consistent with the function of the γδ cells overall control of pathogen growth (O'Brien et al., 2000).

7.4) Other functions:

Apart from the major immune regulatory function and resistance to pathogens, γδ T cells were also found to be involved in a broad range of physiological and pathological processes, which include grafts of tissue, wound healing, epithelial homeostasis, autoimmunity and tumor surveillance (Hayday, 2000; Ferrarini et al., 2002). For example, experiments show that γδ T cells can delay allograph rejection. C3H/HeJ mice injected with multiple minor incompatible B10.BR cells exhibit delayed rejection of subsequent B10.BR skin grafts. This inhibition of C3H-anti-B10.BR immunity is mediated by CD4-8- γδ T cells (Gorczynski, 1994). In addition, studies of intestinal epithelia of γδ T cell deficient mice provide evidence that γδ T cells are involved in epithelial homeostasis (Komano et al., 1995). In TCR δ-/- mice, the absence of γδ T cells correlates with a reduction in epithelial cell turnover while epithelial differentiation was diminished. On the whole, the immune functions of γδ T cells are as important as B cells and αβ T cells. More studies are needed to establish the detailed relationships between γδ T cells and these immune components.

8. Gene regulatory features of TCR γ:

Immune cells are derived from the potent stem cells. Hematopoietic lineage determination and differentiation are controlled by multiple factors. The coordinated transcriptional regulation of the lineage-specific genes is one of the important molecular mechanisms for lineage determination (Rodewald & Fehling, 1998). For example, it has been proposed that both the lineage selection of αβ and γδ T cells and the ordered expression of the mouse Vγ gene subtypes in the thymus are controlled by transcriptional regulation of the accessibility of the recombinatory gene segments to the recombinases (Sleckman et al., 1996). In addition, as multigene families, Ig and TCR gene loci are good models for studying transcriptional regulation because they are rich in regulatory elements and complicated regulatory machinery (Staudt, 1991; Leiden, 1993). Studies of the transcriptional regulation of the TCR genes have led to many characteristic findings. For instance, a great number of ubiquitous and lymphoid-specific transcriptional factors and nuclear binding elements have been identified (Faisst et al., 1992; Leiden, 1993; Lee et al., 2001). In the TCR γ loci, diverse regulatory elements including enhancers, promoters, silencers and boundary elements have also been characterized. These elements have proved essential in determining both the temporal and spatial biological features of γδ T cells.

8.1) TCR γ promoters:

It has been suggested that the promoter of each V gene is a key element in controlling the specific subtype determination (Sunaga et al., 1997). Presently, only the promoters of murine Vγ1 and Vγ5 have been functionally characterized in both the human and mouse TCR γ loci (Ofir et al., 1995; Clausell & Tucker, 1994). Multiple positive and negative elements identified in these two promoters show interesting contrasting features: the Vγ1 promoter contains a TATA box, while the Vγ5 promoter lacks an obvious TATA box; the intrinsic activity of the Vγ1 promoter is very high; the Vγ5 has a very low activity (which needs to couple to an enhancer to demonstrate its activity); the proximal region of the Vγ1 promoter has negative activity, but the proximal region of the Vγ5 shows positive

effects; both promoters possess GC elements that bind Sp1-like proteins. Mutations in the element enhance the activity of Vγ1 promoter by 50% to 80%, while it reduces the activity of the Vγ5 promoter by 80%. Such opposing features have been proposed to explain the ordered fashion of the V genes expressed in the mouse TCR γ locus: conditions that enhance the activity of one promoter inhibit the activity of another promoter and thus lead to sequential expression (Ofir et al., 1995).

In the human TCR γ locus, the promoters have not been functionally studied, but characterized by sequence comparison. The alignment of the TCR γ V gene promoters reveals that they all contain a heptamer sequence (CTGCAGG). It is possible that a *cis*-acting element is involved in transcriptional regulatory mechanisms common to all human TCR γ genes, although the common functional importance remains to be determined (Hettmann et al., 1992). In the peripheral blood, Vγ9Vδ2 T cells are the dominant subtype, which play a major role in responding to infection (Bukowski et al., 1998). The sequencing characterization of the promoter of the Vγ9 shows that this promoter lacks TATA and CAAT boxes but has short tandem repeats. These repeats may regulate the activation of Vγ9 in a way similar to the TATA box. Moreover, there are characteristic decanucleotides (AGGTGGTTGAG) in the promoter regions of both the Vδ2 and Vγ9. Since Vγ9 is preferentially expressed in association with Vδ2 and they are both simultaneously expressed early in T cell differentiation, the common decanucleotide may possibly be responsible for this simultaneous expression (Dariavach & Lefranc, 1989).

8.2) TCR γ enhancers:

Enhancers can increase the transcriptional activity of genes independent of their transcriptional orientations and distance relative to the transcriptional starting sites. Presently, a mouse enhancer γE1 and a human enhancer HγE have been functionally characterized (Spencer et al., 1991; Kappes et al., 1991; Hettmann & Cohen, 1994). The murine TCR γ enhancer γE1 is located at the 3' end of the Cγ1 gene. In addition, by hybridization, using γE1 as the probe, two additional putative enhancers were located at

the 3' end of the Cγ3 and the 5' end of the Cγ2, respectively. This is not surprising because the J-C regions of clusters 2 and 3 are duplication products of cluster 1 (Vernooij et al., 1993). There has been only one human enhancer (HγE) detected, that of 6.5 kb 3' of Cγ2 (Lefranc et al., 1995). The sequence similarity between mouse γE1 and HγE is around 50%. The functional characterization of this enhancer suggests that it is also functionally similar to γE1 (Hettmann et al., 1994). Although both the human and mouse TCR γ enhancers show activity on transcription in γδ T cells, they also display activity in αβ T cells, which suggests that the enhancer is not the sole factor responsible for the sublineage selection. Among all the transcriptional regulatory elements, the role of enhancers in the regulation of V(D)J recombinations has been analyzed in the most detail. *In vivo* studies of the TCR α, β and δ enhancers show that they play a critical role in TCR gene rearrangement (Lauzurica & Krangel, 1994; Leduc et al., 2000). Deletion of the enhancers results in the inhibition of the V(D)J recombination. However, in the TCR γ locus, deletion of the enhancer γE1 only modestly reduces TCR γ gene rearrangement (Xiong et al., 2002). Therefore, the temporal and lineage-specific control of the TCR gene rearrangement may rely on the activities of multiple elements within the TCR γ locus.

8.3) TCR γ silencers:

In contrast to enhancers, silencers are a combination of short DNA sequence module elements that suppress the transcription of genes, independent of orientation or position relative to the transcriptional start site (Brand et al., 1985; Strachan & Read, 1996). Two inhibitory elements have been identified in the mouse TCR γ locus. One is between Vγ6 and Vγ5 in cluster 1 (Clausell & Tucker, 1994), the other may be at the 3' region of enhancer γE1 (Kappes et al., 1991). A transgenic study provided direct evidence for such a TCR γ silencer, which is probably located 3' of the TCR Cγ1 gene segment (Ishida et al., 1990). In the human TCR γ locus, two silencers have been identified at the 3' end of the human TCR Cγ2 (Lefranc et al., 1995). The human enhancer and silencer have equal enhancing and inhibitory activity in both the αβ T cell line Jurkat and the γδ T cell line PEER. However, when the enhancer is combined with the two silencers, they show

inhibitory activity in the αβ T cell lines Jurket and EL-4, but show an enhancing activity in γδ T cell line PEER (Lefranc et al., 1995). Therefore, the combination of human TCR γ enhancers and silencers provides one source of differential effect in the determination of the sublineage specificity.

8.4) Locus control regions:

The locus control region (LCR) is a set of *cis*-acting elements that regulate chromatin accessibility of a gene locus (Strachan & Read, 1996). The open conformation of transcriptionally active chromatin domains makes them more accessible to cleavage by the enzyme DNase I, a hypersensitive site (HsA). LCR first acts at a stage in development that renders all genes in the locus competent, but transcriptionally inactive. Once they are in this competent state, each gene is further regulated by stage-specific enhancers or repressors (Reitman et al., 1993). An HsA has been located at 3' of the mouse α/δ locus (Diaz et al., 1994). It is possible that the differential regulation of TCR α versus δ genes during development is due to the relative distance between the gene and the LCR. In the mouse TCR γ locus, an HsA is identified between Vγ7 and Vγ4 (Baker et al., 1999). HsA supports transcription in a largely position-independent fashion in mature T cells. The enhancer γE1, without HsA, supports transcription in both mature and immature T cells with the restriction of distance and is necessary for the consistent rearrangement of transgenic recombinations. These two elements together, however, can regulate the transcription of the TCR γ gene without the restriction of distance and cell line. Therefore, the HsA provides chromatin-opening activity and together with the enhancer γE1, constitutes a T-cell-specific locus control region for the TCR γ locus.

8.5) Boundary elements:

Boundary elements or insulators are 0.5 to 3 kb DNA segments, which are assumed to be transcriptionally neutral but can block or insulate the spread of positive or negative regulatory elements (Blackwood et al., 1998; Bell et al., 2001). Figure 4 presents the schematic view of enhancers, promoters and boundary elements. In the TCR loci, only one enhancer-blocking element (BEAD-1) was identified between α and δ gene segments in the human TCR α/δ locus (Zhong et al., 1997). BEAD-1 is a 2 kb located DNA

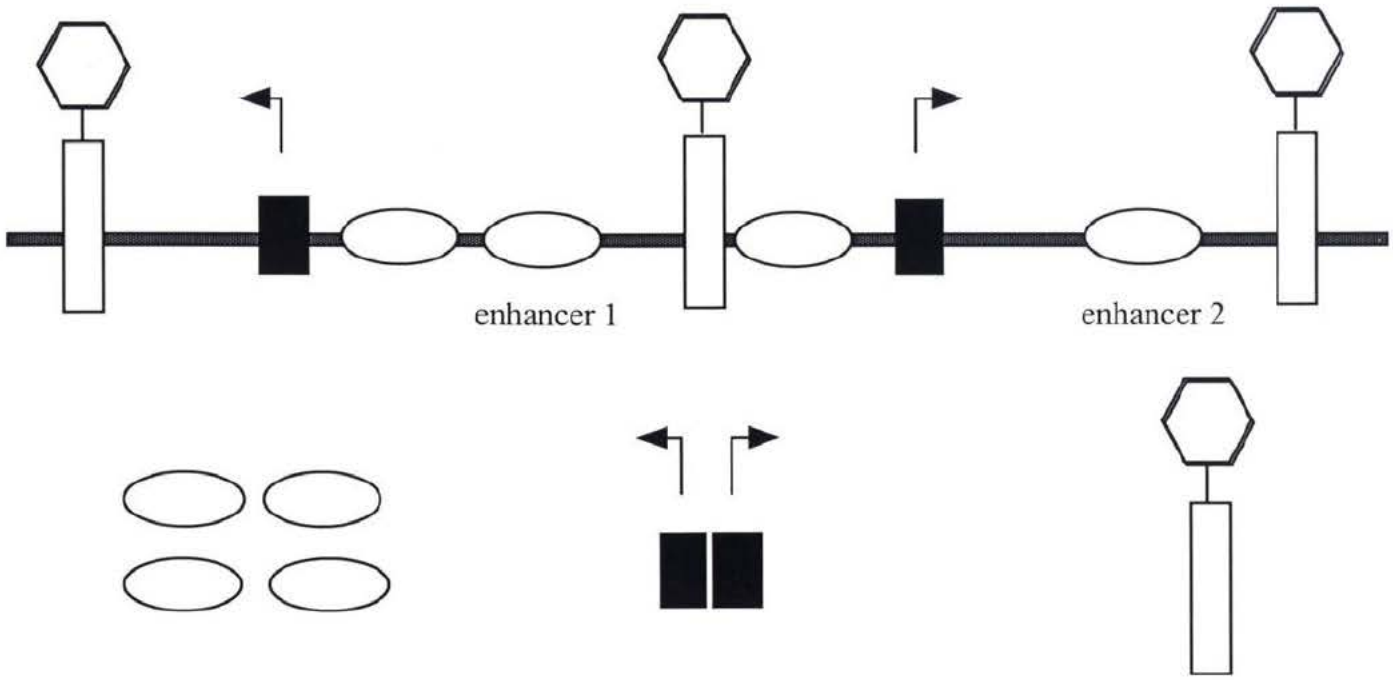


Figure 4. DNA elements that affect the transcriptional regulation by RNA polymerase. The horizontal line depicts a segment of the genome. Transcriptional enhancers contain arrays of recognition sites for sequence-specific DNA-binding factors. One enhancer may activate its cognate promoter and other enhancers which are physically close. When located between an enhancer (ovals) and a promoter (arrows), boundary elements (hexagons) impair the ability of the enhancer to activate transcription from the promoter. It is thought that boundary elements might function to demarcate regulatory domains (Blackwood et al., 1998).

between the 3' end of TCR δ gene segments and the 5' end of TCR α joining gene segments. When it was placed between a TCR γ enhancer and a promoter in a construct, BEAD-1 blocked the effect of enhancer on the promoter *in vitro*, by separating the TCR α/δ locus into distinct regulatory domains controlled by the TCR δ enhancer and the TCR α enhancer. Thus, BEAD-1 can prevent the TCR δ enhancer from opening the chromatin of the TCR α joining gene segments for VDJ recombination at an early stage of T cell development and ensure that the α and δ gene fragments can be rearranged independently. However, a recent *in vivo* study on mice with the human BEAD-1 analogous region deleted, revealed that the TCR α and δ gene expression and rearrangement are not completely perturbed (Sleckman et al., 2001). Therefore, although the role of BEAD-1 in human TCR α/δ locus cannot be ruled out, its contribution to TCR developmental rearrangement is not critical. Other mechanisms may be more important in the developmental control of the subtype of gene fragments in the TCR α/δ locus.

9. Evolution of the TCR genes:

The acquired immune system is believed to have originated about 550 million years ago. The four kinds of key molecules, RAG, MHC, Ig light and heavy chains, including TCR α, β, γ and δ chains, are presented in a wide variety of extant cartilaginous fish, birds and mammals, therefore these distinct gene families were present 450 MYA (Marchalonis et al., 1998; Litman et al., 1999). Such a conclusion is supported by failed experimental attempts to isolate homologous genes in the more primitive agnathan vertebrates, lampreys and hagfish (Rast et al., 1997; Matsunaga, 1998). Neither lamprey nor hagfish possess a thymus or spleen (Zapata & Cooper, 1990). In addition, comparative analysis of some of the representative DNA sequences of the TCR identified in jawed vertebrates, artiodactyl mammals (Hein, 1994), rodents and primates (Arden et al., 1995a,b), carnivores (Hein, 1994), birds (Gobel et al., 1994; Rast et al., 1994), elasmobranches (Rast et al., 1997) and teleost fish (Kamper & McKinney, 2002), also supports the above hypothesis. Figure 5 presents the major evolutionary events in which the different specific immune molecules appear in the major vertebrate species.

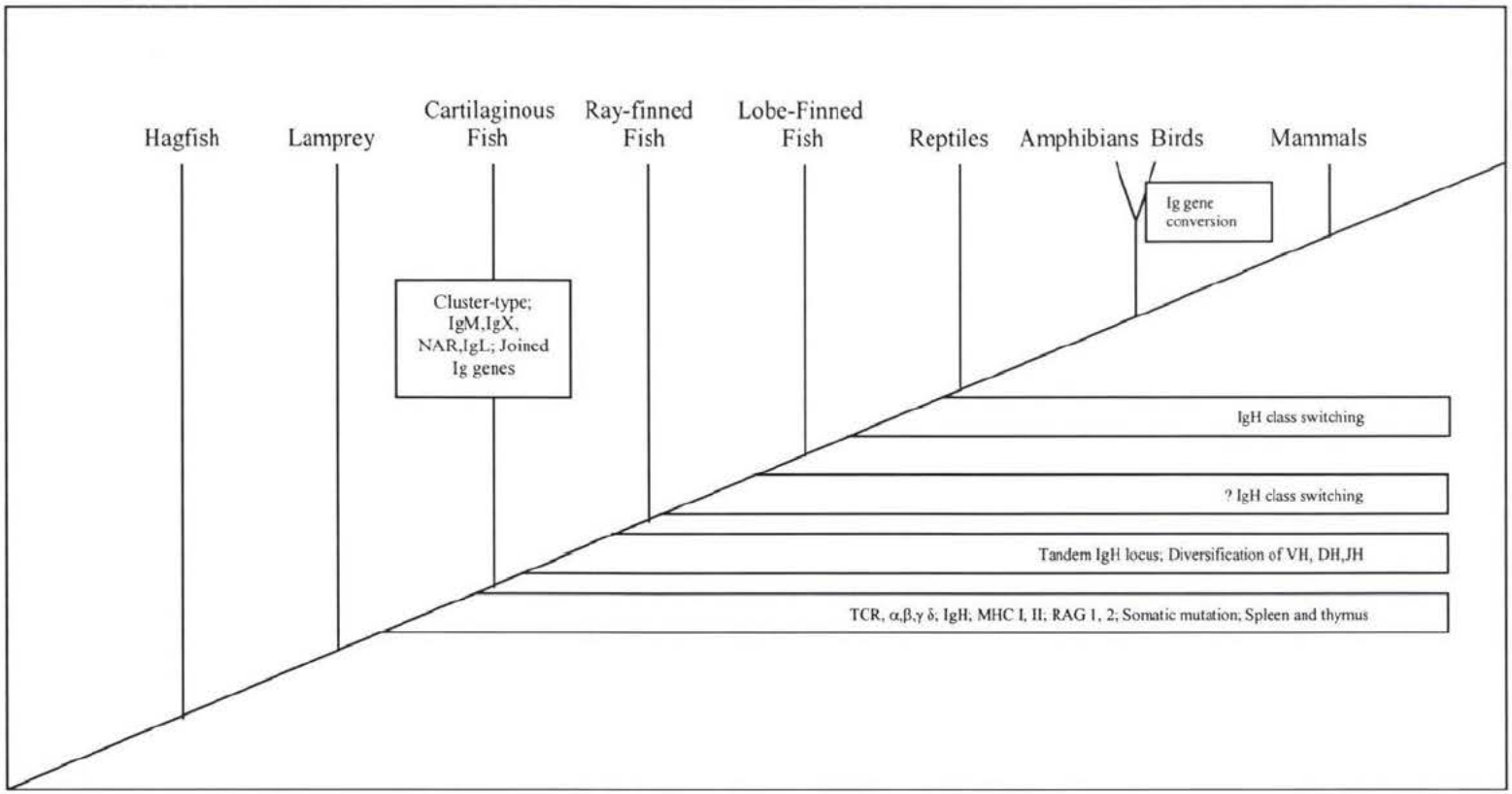


Figure 5. Major events in the evolution of rearranging antigen binding receptors superimposed on a phylogenetic tree of the major vertebrate groups. Ig, immunoglobulin; IgH, Ig heavy chain; IgL, Ig light chain; NAR, novel antigen receptor; DH, Ig heavy chain diversity region; JH, Ig heavy chain joining region; VH, Ig heavy chain variable region; RAG, recombination activating gene; TCR, T cell receptor; MHC, major histocompatibility complex.

In depicting a picture of the early evolutionary history of the adaptive immune system, the primitive receptor's structural form has raised controversy. The primitive receptor molecule was proposed to exist in invertebrate, protochordates and jawed vertebrates as a nonrearranged form (Litman et al., 1999). Actually, a few such ancestral-like candidate genes and gene products have been identified in jawed vertebrates, such as Southern pufferfish, Xenopus and chicken (Rast et al., 1995; Chretien et al., 1996). Such genes are characterized by non-rearrangement and possession of Ig/TCR-like characters. Among them, the prototypic gene in the Southern pufferfish contains a V domain and an Ig C2-type domain (C1 and C2 refer to Ig C1 and C2 type domains), as well as a TM and a Cyt region. The V domain exhibits a higher degree of similarity to TCR and to a lesser degree to Ig V regions, probably as the Ig genes underwent more marked changes than the TCR genes (Rast & Litman, 1998). Therefore, the general structure of V-C2-TM-Cyt may have been the earliest nonrearranging predecessor of those of Ig and TCR. Since IgC1-type domains are restricted in distribution to the vertebrate rearranging receptor, along with classes I and II MHC proteins, a protein of the form V-C2-TM-Cyt may be the predecessor of the nonrearranging V-C1-TM-Cyt-type molecule. The generation of segmental rearranging receptors was hypothesized to be due to a transpositional interruption event of DNA mediated by transposase function associated with RAG gene on the nonrearranging V-C1-TM-Cyt-type molecule (Agrawal et al., 1998).

A second aspect related to the origin of the immune system is the location of the earliest adaptive immunity. The extensive conserved localization of gut-associated immune tissues of all jawed vertebrates suggests that the adaptive immune system originated from the gut of ancient jawed fish (Matsunaga, 1998). Further, from the gastrointestinal lamina propria of cold-blooded vertebrates and birds to the cryptopatches of mouse intestine, the lymphopoietic cells serve the same function. In mice, some T cell subtypes can even develop and mature without the thymus (Saito et al., 1998). Actually, the above hypothesis is not that complicated in terms of the predatory behavior of the ancient jawed fishes. The lamprey, which is more ancient than the jawed fish, suck for food. But in jawed fish, the jaws enable a predatory manner of hunting food. Thus, physical injuries to the food canal by swallowed food, such as bone and scales, increase.

Consequently, injuries and concomitant infections in the gastrointestinal tract of primitive jawed fish may be the major selective force responsible for the acquisition of adaptive immunity (Anderson & Matsunaga, 1996). In the long history of the evolution of adaptive immunity, γδ T cells may be one of the ancient cells in the gut-associated immune tissues, earlier than B and αβ T cells. This is supported by phylogenetic analysis of the TCR V and C gene fragments across vertebrate taxa, which reveals that TCR γ and δ sequences are more ancient than β and Ig genes (Richards & Nelson, 2000). In addition, cellular features of γδ T cells possess the capability to recognize both self- and nonself-antigens, membrane-bounded, and respond to a variety of antigens quickly without the assistance of APCs (Born et al., 1999), all of which meet the criteria for primordial specific immune cells. The most primitive acquired immune cell likely originated from the γδ T cell-like cell in the gut of the cartilaginous fish.

Although the immune system of cartilaginous fish differs radically from that of mammals in many histological, molecular, and functional aspects, the generally similar TCR molecular structure among distantly related species can provide important evolutionary information (Rast et al., 1997). As one of the functional regions, V domains of TCR α, β, γ and δ chains have been phylogenetically analyzed extensively in diverse representative vertebrate species. In Figure 6, three major clusters, which correspond to the TCR Vγ set (I), the TCR Vβ set (II), and the TCR Vα/δ group (III), are formed. Therefore, the TCR V domains are consistent with their nominal identification, although a few pose difficulties in defining their group. Among them, a few V domains are found to be conserved throughout vertebrate evolution at the amino acid level. For example, in group I, there is an association between human Vγ4 and skate Vγ5 (30% identity), and similarity between chicken Vγ and skate Vγ2 (40% identity). Some residues which are typically found in most mammalian and avian Vγ sequences are conserved in skate, such as the IHWYR stretch around Trp35 and the TYYCAW stretch around Cys92 (Charlemagne et al., 1998). In addition to TCR Vγ, there is considerable conservation between long separated species in the TCR Vα and Vβ domains (Partula et al., 1995, 1996). Having arisen early in evolution, such conservation indicates that these TCR V

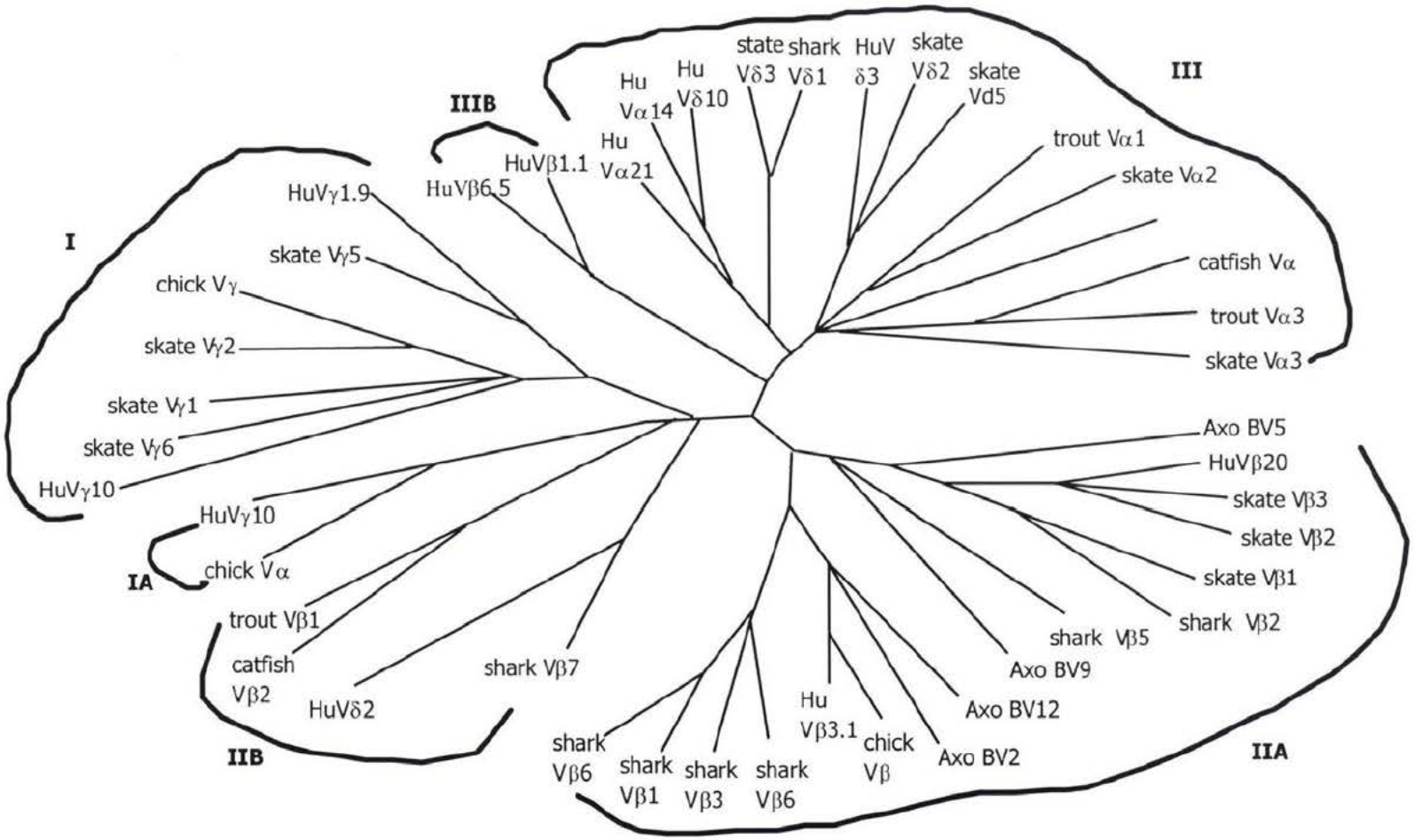


Figure 6. Unrooted tree of representative TCR V domains of diverse vertebrate species. The dendogram was calculated using the program CLUSTALW (Thompson et al., 1994) and displayed using the program Treeview (Page, 1996). The clusters are predominantly Vγ(I), Vβ(II), and Vα/δ (III), while mixed subsets are designated IA, IIA, IIB, and IIIB.

regions have been maintained for around 450 MYAs. One reason for maintaining the V domains is probably strong selection pressure. One force may be infection from a common type of bacteria or virus infections (Marchalonis et al., 1998). Another force may be from the molecular constraint of MHC. In comparison with Ig molecules, the structure of TCRs seems constrained. MHC I and II are also extraordinarily well conserved in all jawed vertebrates. Therefore, the strong functionally linked TCR and MHC molecules have been reciprocally frozen during evolution to maintain their structural conservation (Charlemagne et al., 1998).

The phylogenetic analysis of another functional region, the C domain, also generated a more accurate evolutionary inference because the overall evolution of C is slow relative to V gene fragments and therefore they are better conserved (Anderson et al., 1995; Marchalonis et al., 1998). Figure 7 reveals the phylogenetic analysis of the C regions of TCR, MHC I, MHC II, Ig H and Ig L chains, which constitute five independent clusters. This phylogenetic pattern of C gene fragments is consistent with the conclusion that these three groups of molecules, sequentially derived from the ancestral locus along with the appearance of vertebrates, became clearly distinct in gnathostome vertebrates. Among the sequences of TCRs, the Cγ chains are more variable and evolve faster than the C chains of other loci. For example, there is not much similarity between the Cγ gene of skate and that of mammals, but Cα and Cβ have 60% and 77-94% amino acid identity respectively throughout mammals (Charlemagne et al., 1998). This is mainly because of the fast sequence divergence that resulted from the generation of the multiple isotyes of Cγ chains by gene duplication or triplication (Hein, 1994). Multiple isotypes of Cγ genes were found in all species of primates (e.g. humans), rodents (e.g. mice) and artiodactyls (sheep, cows and pigs) (Ciccarese et al., 1997). The five Cγ genes in sheep are descendants of an ancestral pool that existed before the primate-rodent-artiodactyl diverged (Hein et al., 1993) while the two human and four mouse Cγ genes arose after the separation of primates and rodents. It has been estimated using the molecular clock that the distance between cattle and sheep Cγ gene fragments is 18 MYA, which coincides with the 12 MYA of the duplication of Cγ genes in humans and mice. This time period corresponds to the last Tertiary period of the Miocene (24.6-5.1 MYA), when the mammalian

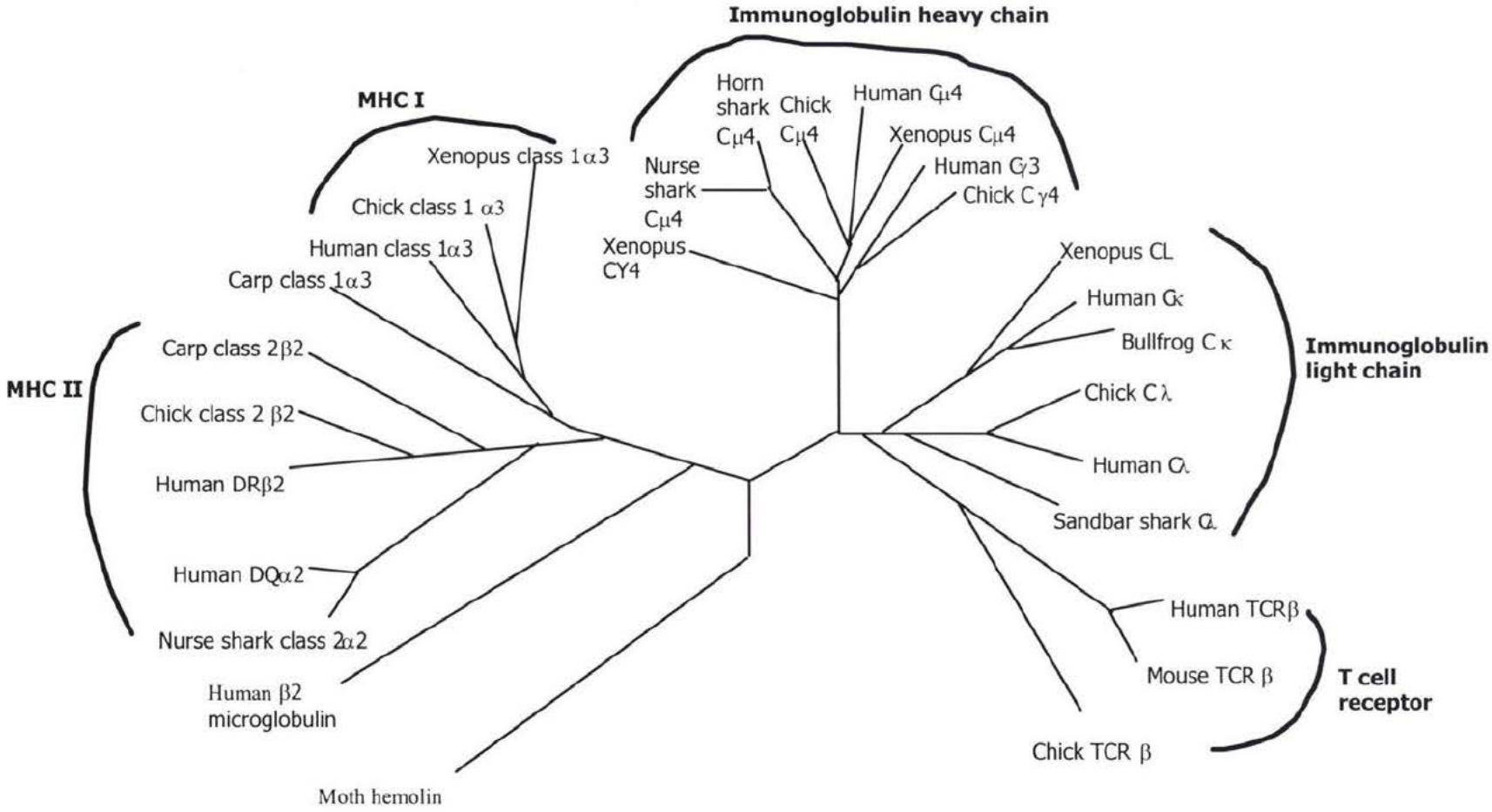


Figure 7. Rooted phylogenetic tree depicting the relationship among the immunoglobulin constant domain of light chains, heavy chains, T cell receptors, and MHC products of different species. Moth hemolin is used as the outgroup (Marchalonis et al., 1998).

radiation gave rise to a great number of the genera that persist today. This suggests that the rapid duplications of Cγ genes in mammals were subjected to similar evolutionary pressures (Ciccarese et al., 1997).

10. The Human Genome Project:

Since the discovery of the double helix structure of DNA in 1953, molecular biology and related techniques have been developing rapidly. These developments eventually make the complete deciphering of the whole human genomic sequence possible. The Human Genome Project (HGP) was officially started in the United States in 1990 and it was estimated to be completed in 15 years (Collins & Galas, 1993). The US Human Genome Project is the major contributor but similar projects were also established in the UK, France, Japan and other countries. In order to co-ordinate the different national efforts, the International Human Genome Organization (HUGO) was established to facilitate the exchange of research resources, encourage public debate and advise on the implications of human genome research (McKusick, 1989). Since 1998, some private genomic companies have joined the sequencing competition, adopting the whole-genome shotgun strategy as well as high-throughput sequencing machines, which have greatly sped up the project. In June 2000, the first working draft (a total of 2.297 Gbp of sequence) that has mapped 91% of the entire human genome and 9% of uncertain gaps (100,000 gaps) was announced by publicly funded genomic groups and the Celera Genomics Co-operation (Lander et al., 2001; Venter et al., 2001). In April 2003, the project was finished (Collins et al., 2003).

The initial goals of the HGP relate to three aspects: 1) to improve the present physical map of human chromosomes; 2) to develop supportive technologies and facilities for human genome research; and 3) to expand communication networks and computational and database capacities (Collins & Galas, 1993). So far, most of these goals have been accomplished. Firstly, around 24500 to 26383 human genes have been detected among a total of expected 32,000 genes (Lander et al., 2001; Venter et al., 2001). In addition to the human sequence, the genomes of a few medically and important models including the *Haemophilus influenzae*, *E. coli*, yeast *S. cerevisiae,* fruit fly *Drosophila melanoganster*,

Arabidopsis thaliana, nematode *Caenorhabditis elegans* and mouse genomes have been fully sequenced (Mural et al., 2002; Waterson et al., 2002). Meanwhile, some genome projects are targeting the screening of disease genes. Secondly, in the process of sequencing the HGP, DNA cloning and sequencing techniques have been revolutionized by the application of multidisciplinary knowledge. These mainly include the application of the shotgun sequencing strategy and automatic sequencers. Thirdly, numerous databases and bioinformatic techniques have been built. The large amount of information being generated in the genome project can be accessed from remote computer stations through the internet. Distant users can also input raw data that is processed by the database managers and can analyse data that is already stored in the databases. Some of the major sequence databases, such as the most popular databases GenBank and EST bank, are located at dedicated centers in the USA (National Center of Biotechnology Information), UK (European Bioinformatics Institute) and Japan (National Institute of Genetics) (Brown, 2000).

The so-called post-genomic era is characterized by the accumulation of sequence data applications. Proteomics focuses on the large-scale identification and quantification of proteins, including the determination of their localization, modifications, interactions, activities and ultimately, their function (Fields, 2001). One of the relatively immediately available applications of HGP is the rapid development of pharmacogenetics, which is based on single-nucleotide polymorphisms (SNPs) (Liggett, 2001). SNPs are the single nucleotide variations on genomes among different individuals and ethnic groups. As most drugs show significant interindividual variation in therapeutic efficacy, SNP based treatments may individualize the dosage of the therapeutical agent. By comparing random reads against the human genomic sequence, about 1.4 million have been identified (The International SNP Map Working Group, 2001). Other medical applications include the identification of genes associated with hereditary diseases, development of drugs to treat these diseases (Peltonen & McKusick, 2001) and rapid cytogenetic diagnosis (Antonarakis, 2001). In fact, a total of 1112 disease genes have been discovered and more than 30 genes among them were cloned using the draft genome sequence from public databases. Collectively, the knowledge, resources and technologies emanating from the

HGP are revolutionizing every branch of science, medicine and other related fields, such as biology, behavioral science, medical diagnosis, monitoring and treatment of diseases, pharmaceutical industry, and biotechniques, in an interdisciplinary way.

Within such a background of biological revolution, as a part of the HGP, characterizing the human and mouse T cell receptor gamma loci started in 1994 and finished in the middle of 1997. These loci are multigene families, and therefore their sequences provided excellent models for the understanding of genomic structure and organization. Meanwhile, characterization and comparison of these two sequences also provide deeper insight into the understanding of the immune system and molecular evolution.

Chapter II. Characterization of the human TCR gamma locus

Summary:

The human TCR γ locus is a 140 kb region on chromosome 7. Characterization of the complete DNA sequence reveals that the coding regions occupy 4.55%, the locus-specific repeats 57%, and the genome-wide interspersed repeats 24%. The known genes including fourteen V, five J, and two C gene fragments were identified. The previously reported one promoter and two silencers, and the presently reported one enhancer were also identified on the sequence. Evolutionary analysis of the TCR γ locus suggests that this locus originated from a common primordial structure, which contains a single V gene in the GV1 subfamily and a single J-C cluster. Through a series of unequal crossing-overs and gene conversions, the present TCR γ structure emerged with a much-expanded GV1 subfamily and two repeating J-C clusters. Following the duplications, the V gene fragments were further diversified at the CDRs by positive natural selection. Thus, the V-J-C combinatorial diversity is greatly increased by these processes. Meanwhile, extensive gene duplication and dysfunction in the GV1 family also suggested that this locus has evolved through the birth-and-death process. In general, complete sequencing of the human TCR γ locus provides a good model to gain an insight into the organization, evolution, and functionality of multigene families. In particular, evolutionary analyses have proved to be a powerful tool for a structural and functional understanding of the genome.

Introduction:

Gene duplications are the key force of gene and genome evolution (Jeffreys & Harris, 1982). More than a third of a typical eukaryotic genome consists of duplicated genes and gene families (Wagner, 2001). Gene duplication and the subsequent diversification due to relaxed selective constraints is the simplest way to generate new genes with novel and complex biological functions (Ohta, 1991). The size of duplication is not only limited to a specific gene, it could be part of a chromosome, the entire chromosome or even the whole genome (Li, 1997). Three kinds of mechanisms, slipped-strand mispairing, unequal crossing-over and gene conversion, are responsible for the gene duplication (Li, 1997). Unequal crossing-over is widely believed to be the predominant biological mechanism for medium to large sized tandemly repeated sequences (Elemento et al., 2002; Graham, 1995). This process takes place in the M phase of meiosis, when the chromosomes line up in tetrad configuration, in which the homologous regions form a cross and then exchange genetic materials. In the resulting two alleles, one gains a piece of DNA from the other chromosome, whereas the other one loses some of the DNA (Li, 1997). Thus, gene duplications regulate the amount of gene products of the same function that are essential for the survival of an organism by an increased or decreased gene number. On the other hand, the increased gene members also provide raw materials for developing genes with novel functions.

Diverse evolutionary forces operate on the multigene family to shape the organization to cope with evolutionary needs. For the last two decades, it has been commonly believed that most multigene families are subject to concerted evolution (Li, 1997). This view was originated primarily from the studies of ribosomal RNA genes and histone genes (Arnheim, 1983; Matsuo & Yamazaki, 1989), in which unequal crossing-over and gene conversion are required to homogenize the family members to meet the need of large amounts of similar gene products (Scott et al., 1984). It was later extrapolated that most of the multigene families are under concerted evolution. In recent years, however, studies on the multigene families associated with immune response or disease resistance have shown that the gene members in these families diversify dramatically instead of being homogenized constantly after duplication by unequal crossing-overs (Nei, 1997). The

above gene families are believed to be subject to birth-and-death evolution rather than concerted evolution, in which some of the duplicated gene members are diversified to cope with different pathogens while others become dysfunctional (Piontkivska et al., 2002). In addition, the hypervariable regions are substantially higher than the frame regions of the genes in the above gene families. The evolutionary force behind the difference between the two types of regions is not clear. One hypothesis, based on the studies of the MHC, suggested that this was due to mini-gene conversion (Kappes & Strominger, 1988; Ohta, 1995). The study on disease resistant genes and immunoglobulin VH gene families, however, revealed that positive selection was responsible for the selection of advantageous amino acid diversity (Ota & Nei, 1994). With such unclear questions, the TCR γ locus, as another multigene family of the immune system with characteristic gene composition and organization, will provide a good model for studying the diverse evolutionary forces shaping the organization of multigene families and the different functional regions of gene members.

T lymphocytes play the central role in the specific immune system, destroying or eliminating pathogens such as viruses, bacteria, parasites and cancer cells with immune regulatory and cytolytic functions (Levinson et al., 1996). The features of αβ T cells have been clearly defined but the role of γδ T cells in the immune system remains uncertain (Kaufmann, 1996; Hayday, 2000). An increasing body of work indicates that the two types of T cells are different in numerous aspects, including function, ontogeny, and expression (Born et al., 1999). For example, recognition of antigens by α/β T cells takes place in the context of MHC molecules whereas γδ T cells do not require MHC for antigen recognition. The human γδ TCR is a heterodimer which consists of δ and γ polypeptide chains (Davis et al., 1988). Each polypeptide chain is coded by V, J and C gene fragments. The human TCR γ genes are located on chromosome 7 at band 7p14-p15 (Bensmana et al. 1991; Rabbitts et al., 1985; Lefranc et al., 1990). The organization of the human TCR γ locus has been characterized through cDNA and selected germline data analyses (Lefranc et al., 1990). The variable region is at the 5' end and it is followed by two clusters of joining-constant regions. The fourteen V gene segments are classified into six gene subfamilies. Each V gene segment has three hypervariable regions that

putatively correspond to antigen binding CDR regions (Huck et al., 1988; Hein et al., 1993). In the GV1 family, nine human TCR γ V1 genes share 83-86% similarity to each other and the J and C genes in the two J-C clusters also have a high degree of similarity, which suggests that they were generated by gene duplications. However, the mechanism underlying the duplications at these regions has not been determined. This research will document features with respect to the organization, structure, gene regulation and evolution of the complete sequencing of the entire human TCR γ multigene locus. In addition, the impact of the complete genomic structure and evolution on the functionality of TCR were also investigated.

Methods and materials:

The DNA of the human TCR γ locus was obtained from two genomic libraries constructed with the pWE15 cosmid and BAC vectors respectively (Lai et al., 1991). Five clones were screened out with a 300 bp probe, homologous to the GV1 gene exon 2, from the cosmid library. The BAC library was screened by 150-260 bp probes homologous to the Vγ10, Vγ11 and constant genes, and two clones were identified. These clones were mapped based on the results of restriction digestion and Southern hybridization. In addition, two lambda clones, λSH3 and λS9, which were originally isolated by Lefranc et al., (1986 a & b), were obtained from the American Type Culture Collection (Rockville, Maryland). The lambda DNA was extracted using the lambda DNA extraction kit (Qiagen). The DNA fragment contained a sequence between Vγ8 and VγA, that could not be cloned into stable cosmid, BAC, PAC, or lambda vectors. Using PCR primers designed from the exon sequences of Vγ8 and VγA, this piece of DNA was isolated from the human genomic DNA template. Cosmid and BAC DNA were extracted by the alkaline-lysis method (Maniatis et al., 1989) and PCR was conducted using Perkin Elmer's GeneAmp XL PCR kit. All the DNA of cosmids, BAC, lambda clones, and the PCR products were subjected to nebulization for random fragmentation (Povinelli et al., 1993). The fragmented DNA was then treated by Mung-bean nuclease, T4 polymerase, and Klenow polymerase sequentially. The subsequent blunt-ended DNA fragments were ligated with M13 vectors, which were previously linearized with *Hinc* II and dephosphorylated with calf intestinal alkaline phosphotase. Transformation of the M13

clones was conducted by electroporation. The templates were prepared using the Qiaprep kit (Qiagen, Santa Clarita, CA) and sequencing was conducted using ABI sequencers. An approximate seven-fold redundancy of sequences was assembled using Seqman (DNASTAR, Maddison, WI), PhrapPhred, and Consed (Phil Green et al., unpublished data, E-mail: phg@u.Washington.edu) software in combination.

The TCR γ genes and regulatory elements were located on the resulting genomic sequence by sequence comparison with published data (Lefranc & Rabbitts, 1985; Lefranc et al., 1986a & b; Font et al., 1988; Huck et al., 1988; Buresi et al. 1989). Synonymous and nonsynonymous nucleotide distances were calculated using the MEGA program (Kumar et al., 1993) based on the method of Nei and Gojobori (1986). Boundaries of the locus specific repeat were identified by dot-matrix comparisons using GeneAssist 5.5 (ABI, Perkin-Elmer Corp) and DOTTER (ftp://sanger.ac.uk/pub/dotter, Sonnhammer and Durbin, 1996). Multiple sequence alignments of GV1 repeats were made using the MAP program (match= 10, mismatch= -17, gap open = -30, and gap extension= -5). Genome wide repetitive elements were identified using RepeatMasker software (Smith and Green, unpublished). Evolutionary events of the GV1 family and J-C regions were searched by phylogenetic scanning along the multiple sequence alignments (Slightom et al. 1988; Fitch et al., 1991) using the intervals of 60 bp, 100 bp, 600 bp, 1000 bp and 3000 bp respectively. UPGMA trees were constructed using PAUP 4.00d55 program (Swofford, 1997) and were rooted by the most distant sequence. Adjacent trees along the multiple alignments that showed similar phylogenetic branching patterns were merged until two regions showed distinct patterns of phylogeny.

Results and discussion:

Figure 8 represents the genomic organization of the human TCR γ locus. The entire locus is 140691 bp (GenBank: AF159056), which is smaller than the previous estimations (160 kb by Lefranc et al., 1989; Strauss et al., 1987; 150 kb by Fox et al., 1989). In contrast to the V-J-C clusteral organization of the mouse TCR γ locus, the V region of the human locus is located at the 5' end whereas the region containing the two

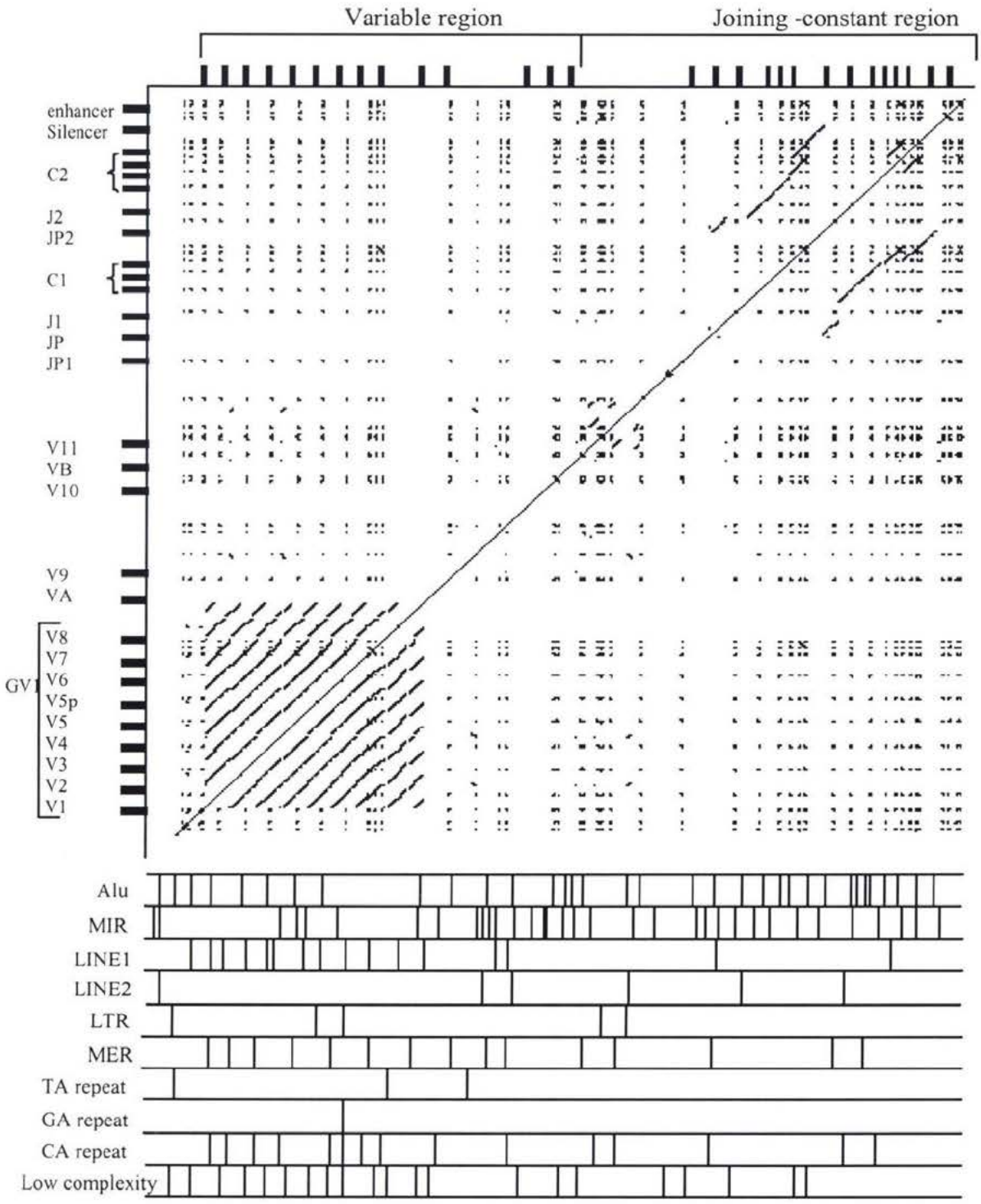


Figure 8. Features of the human TCR γ locus. The top graph is the dot-matix self-comparison of the 140691 bp of the human locus. The dots and lines indicate similar sequences between different regions. The comparison is made using Pipmaker software. The lines below the matrix are the major types of genomic wide repeats in this locus.

J-C clusters is at the 3' end. The length of the two regions is 95 kb and 45 kb respectively (Bensmana et al., 1991).

1. Coding elements:

1.1) Variable genes:

The coding gene segments occupy about 4.5% of the entire genomic sequence. Fourteen V gene fragments were identified in the complete genomic sequence. These genes have been classified into six subfamilies: GV1, GV2, GV3, GV4, GV5 and GV6 (Lefranc, 1990; Arden, 1995a & b). In the GV1 family, nine gene fragments exist, which are designated as Vγ1, Vγ2, Vγ3, Vγ4, Vγ5, Vγ5P, Vγ6, Vγ7, and Vγ8 from the 5' end to the 3' end sequentially. But in the other five subfamilies, each contains only a single gene. Each V gene consists of two exons, separated by an intron. Among the V genes, only six are functional while the remainders are pseudogenes. These pseudogenes became dysfunctional due to reasons involving deletion in the rearrangement signals (Vγ1), in-frame stop codons (Vγ5P and VγA), reading frame shift (Vγ7), or the combination of the above factors (Vγ6 and VγB). (Lefranc, 1986b; Forster, 1987). In addition, two V gene relics are revealed by dotmatrix comparison. One is between Vγ9 and Vγ10, in which the sequence is similar to the human Vγ10. The other V gene relic is between Vγ11 and JP1. Its sequence resembles exon 1 in the human Vγ10, the region between the 3' intron and 5' exon 2 of the human VγB and mouse Vγ4. The highly diverged sequence of the two relics from other V genes makes the identification of their boundaries and reading frames difficult. The three possible reading frames all lead to early stop codons. No DNA, EST, or protein sequences have been found to match the two relics. Therefore, the two V relics have been dysfunctional for a long time and have never regained their function.

1.2) J and C genes:

In the first J-C cluster, the three J genes, JP1, JP, and J1 are located 5' upstream of Cγ1. The second J-C cluster includes only JP2 and J2 in addition to the 5' end Cγ2. Cγ1 and Cγ2 each consist of three distinct kinds of exons, i.e. exon 1, exon 2 and exon 3. Exon 1 is an immunoglobulin-like domain of 330 bp. Exon 2 which contains only ten

amino acids, encodes to the hinge region, while the 545 bp exon 3 codes for the transmembrane domain and the region which extends into the cytoplasm (Saito et al., 1984). There is a single copy of exon 2 in Cγ1 but two copies in Cγ2 designated as exon 2^R and exon 2. The presence of three copies of exon 2 is also observed in some human populations (designated as exon 2^T, exon 2^R and exon 2; Lefranc et al., 1986b; Buresi et al., 1989). There is a cysteine residue in exon 2 in Cγ1, which forms the γ-δ interchain disulfide bridge, while the exon 2 of Cγ2 does not contain any cysteine. The structural differences between C genes correspond to the isotypic difference of γ chains on the γδ T cell (Littman et al., 1987; Lefranc et al., 1989).

1.3) V-J rearrangement and RNA splicing:

Individual V and J genes perform rearrangements during lymphocyte differentiation, generating a contiguous V-J gene. The rearranged V-J gene fragment is spliced to a specific constant gene during transcription to form a functional TCR γ gene cluster (Abbas et al., 1994). All the splicing signals in the human TCR γ locus conform to the GC/AT rule except for Vγ10 and Vγ11, which exhibit mutated splicing signals. The rearrangement signals at the 3' end of the V gene fragments and the 5' end of J gene fragment consist of a heptamer and a nanomer, separated by a spacer. In the rearranged immune molecules, the sequences of heptamer and nanomer are similar within V or J regions but different between the two spacer regions. The sequences of the spacer are variable but the spacer length is more or less constant: 23 bp in the V region and 12 bp in the J region. The spacer length, rather than the spacer sequence, is important for the rearrangement (Hayday et al., 1985). The spacer length of Vγ1 is a 7 bp sequence ATTCAGA and therefore Vγ1 loses its function due to its insufficient length. In addition, it has been shown that the expression of J and V genes is not random. Vγ9, J1 and J2 are more frequently expressed than other genes and Vγ9 positive T cells are the predominant population in peripheral blood (Tribel et al., 1988; Dariavach & Lefranc, 1989; Bukowski et al., 1998). Differential gene expression is obviously controlled by the differential regulation of the accessibility of the gene fragments to recombinases (Sleckman et al.,

1996). Among the pseudogenes, only the transcripts of Vγ10 and Vγ11 were observed but they are not expressed due to the failed splicing to C genes.

1.4) Diversity:

In contrast to the approximately 50 Vα and 46 Vβ TCR gene fragments that can pair to form several thousands of receptor combinations, there are only six functional Vγ and four Vδ genes, which yield 40 potential V-J pairs. In the peripheral blood, 50-75% of the γδ T cells are Vγ9+ T cells while the others are GV1 family member positive T cells. All these reveal that combinatory diversity is not the major mechanism for generation of the diversity of the human γδ T cell. However, it was deduced that there is a high degree of junctional diversity in the human γδ TCR. For example, the human γδ T cells that respond to mycobacteria all use Vγ9 and Vδ2 chains, both of which exhibit considerable junctional diversity (Hass et al., 1993). Therefore, despite the limited number of V gene segments and selected usage of particular gene fragments, there is enormous potential for diversity in γδ TCRs located at the V-J junctional regions.

Various mechanisms have been suggested for junctional diversity: deletion of nucleotides at the ends of the coding V and J regions by exonuclease, addition of non-germline nucleotides (N-addition) by the terminal deoxynucleotidyl transferase, the addition of reversed repeats of the palindromic structure (P-addition), and nucleotide duplication (Gauss et al., 1996). In order to examine V-diversity of the TCR γ, 129 cDNA or mRNA transcripts were withdrawn from the database and compared with germline sequences. The comparison precisely delimited the N region, and revealed N-diversity from all the transcripts examined. Table 4 summarizes the results. The majority of V and J genes possess nucleotide deletion at the ends of the coding region, and the number of the nucleotides deleted ranges from 1 to 28. On the other hand, the majority of transcripts also possess the newly added N-addition. However, although a variety of rearranged V-J fragments are generated mainly by nucleotide deletion and N-addition, only 48% of the resulting transcripts are potentially functional. The others are dysfunctional due to the reading frame shift.

V gene	3 end of V		N		5 end of J		J gene	No. Transcripts
Vγ10	-9		0	+8	-13	0	JP1	3/9
	-6		9		0		JP2	1/1
	-6	-27	+3	+17	0	-28	J1/J2	5/15
Vγ2	-1	-10	+9	10	0	-21	JP1	3/3
	0	-4	0	+11	-1	-6	JP2	0/3
	0		+5		-6		JP	0/1
	0							
	0	-6	+3	+12	0	-6	J1/J2	4/6
Vγ9	0		+3		0		JP1	1/1
	-8		+4		-2		JP2	1/1
	0	-6	+2	+12	-1	-12	JP	16/17
	0	-12	0	+28	0	-16	J1/J2	15/34
	0	-5	+5	+7	-7		JP2	2/3
Vγ3	0	-6	0	+11	0	-14	J1/J2	6/11
Vγ4	0	-15	0	+17	0	-25	J1/J2	5/13
Vγ5	0	-3	+5	+29	0	-9	J1/J2	3/6
Vγ8	-9		+10		-9		JP2	0/1
	-6		+4	+12	-1	-3	J1/J2	2/2
								Total: 67/129

Table 4. V and J gene rearrangement for N-diversity. The numbers indicate the range of nucleotide changes (deletion is indicated by a - while addition is indicated by a +). The number of transcripts is presented as functional transcripts divided by total number.

2. Structural features:

Interspersed genome-wide repeats form the major component of the genomic sequence (Li, 1997). They are closely related to the functional, evolutionary and structural features of a particular genomic region. In the human TCR γ locus, 23.9% of the sequence is genome-wide repeats. Table 5 and Figure 8 show the frequency and distribution of the genome-wide repeats on the locus. As indicated, most types of genome-wide repeats of the human genome can be found in this region, and they are present within all the locus-specific repeats except for J1 and J2. The most common genome-wide repeats are SINEs, which constitute 13.3% of the locus. In comparison, the human α and β TCR loci and the entire human genome present 8%, 5%, and 11.7% of SINEs, respectively (Henikoff et al., 1997). The percentage of the SINES in the human TCR γ locus is also higher than the 2.04% of genome sequence in the mouse TCR γ locus. Alu and MIR (mammalian-wide interspersed repeat), the two kinds of SINEs, are equally frequent on the locus, but the former is evenly spread across the entire sequence, while the latter is less frequent in the GV1 region. Vγ1, Vγ2, Vγ3, Vγ4, Vγ5, and Vγ5P in the GV1 subfamily contain a single copy of an Alu with the same insertion site, suggesting a common origin. Vγ6, however, contains three copies of Alu and a 562 kb retroviral segment. The three Alu copies are uniquely inserted, with the second one oriented in reverse. Vγ7 and Vγ8 repeats, however, do not contain any Alu, although Vγ7 has a 550 bp insert of unknown origin. The presence and absence of an Alu is also responsible for the different size between JP1 and JP2 repeats. MIRs are present in Vγ5, Vγ5P, Vγ6, and Vγ7 of GV1, and the divergence among the MIRs is equivalent to the divergence among other regions in the GV1. The MIRs are therefore inserted before the GV1 duplication, and the absence in Vγ1, Vγ2, Vγ3, and Vγ4 is attributable to the deletion in their shared ancestor. The majority of the LINEs are LINE1s (L1). In comparison with the percentage of L1s in the mouse TCR γ locus (30.93%), L1s are less frequent in the human TCR γ locus (6%). The L1s are mainly concentrated in the V region. Therefore, L1s are likely inserted before the GV1 duplication and are duplicated along with the GV1 elements. There are also a few L2s and LTRs (long terminal repeats) in this locus. Other DNA repeat elements, such as

Table 5. Summary of the genomic wide repeats in the human TCR γ locus.

		Number of elements	Length occupied	Percentage of sequence
SINEs:		85	18729 bp	13.31%
	Alus	44	12354 bp	8.78%
	MIRs	41	6375 bp	4.53%
LINEs:		30	8929 bp	6.35%
	LINE1	24	8448 bp	6.00%
	LINE2	6	481 bp	0.34%
LTR element:		7	2909 bp	2.07%
	MaLRs	1	336 bp	0.24%
	Retrov.	4	2044 bp	1.45%
	MER4_group	2	529 bp	0.38%
DNA elements:		20	3019 bp	2.1%
	MER1 type	7	1179 bp	0.84%
	MER2 type	3	831 bp	0.59%
	Others	10	999 bp	1.43%
Simple repeats:		34	3050 bp	2.17%
Low complexity:		14	510 bp	0.36%
Total			33586 bp	23.87%

MER53, are mainly concentrated in the V region, and like other genome-wide repeats, they are the earlier invaders into the γ locus, duplicating along with the GV1 elements.

Although the functions of genome-wide repeats have not been characterized very well, they demonstrate some functions in the human TCR γ locus. For example, it will be argued later that unequal crossing-overs in TCR GV1 and C genes are mediated by SINE elements, leading to either an expansion or a contraction of the gene family. The genome-wide repeats of similar sequences distribute throughout the genome and thus provide great potential for the heteroduplex formation and recombination between non-homologous sequences. Genome-wide repeats also mediate translocation as observed in the β locus (Rowen et al., 1996). Genome-wide repeats evolve to become regulatory elements, and their mobility shifts the regulatory proprieties of genes (Brosius, 1991). In fact, even a larger proportion of the human genome is expected to be of the genome-wide repeat origin, as much sequence of the origin has probably been degraded to a point beyond full recognition (Heinkoff, 1997). The genome-wide repeat therefore contributes, to a great extent, to the much enlarged and complicated eukaryotic genome.

3. Gene regulation:

Relative to mice TCR γ locus, less transcriptional regulatory elements have been identified and functionally studied in the human counterpart because it is easier to manipulate mice. In this study, the promoters of V gene segments, two silencers and one enhancer reported previously were identified on the genomic sequence. The promoter, located upstream of each V genes, possesses a heptamer (CTGCAGG) conserved across all the V gene fragments (Hettmann et al., 1992). There are no well-defined TATA-like or CCAAT-like structures found in the 5' regulatory region. The enhancer is located 6.5 kb down stream of the Cγ2 (Hettman et al., 1994). The core element of the enhancer displays a high sequence similarity to the murine TCR γ Cγ1 enhancer. Between Cγ2 and the enhancer, there are two silencers with similar sequences (Lefranc et al., 1995). An Alu element is located between the two silencers. Similar to the Sil2A silencer of the murine a TCR and the negative element of human and mouse Igκ, the Alu is probably the third silencer of the human TCR γ locus (Lefranc et al., 1995). The enhancer and

silencers are active in both αβ and the γδ T cells (Lefranc et al., 1995). Nevertheless, the silencers, in association with the enhancer, repress the transcription of αβ but activate γδ lineage, prompting the dual elements to act as an enhancer in the γδ lineage but act as a silencer in the αβ lineage, in turn causing the regulation of the sequential and mutually exclusive expression of the two types of TCR on the surface of a T cell. A similar pattern of the combining role between enhancers and silencers is also observed for the murine TCR α locus (Winoto & Baltimore, 1989b). In addition, a 510 bp potential LCR between Vγ9 and VB with 59% similarity with the mouse LCR was also located. Within the LCR, enhancer and the 5' regions of J gene fragments, STAT5 motifs were also identified. The high conservation of STAT5 among all these regulatory elements suggests that they cooperate to regulate the TCR chromatin accessibility as one possible pathway. As STAT5 has been believed to be responsible for regulating germline accessibility by histone acetylation and the development of γδ T cells relies on IL-7 (Kang et al., 1999), a model of regulation was proposed: IL-7 activates STAT5 protein through IL-7 receptor, followed by STAT5 proteins interacting with the enhancer and recruiting other transcriptional coactivators to induce histone acetylation, by which STAT5 becomes involved in regulating chromatin accessibility (Lee et al., 2001).

4. Evolution:

Multigenic families are inherently unstable entities that can expand, contract and disperse during the course of evolution (Jeffreys et al., 1982). Similar to the mouse TCR γ locus, the multigenic human TCR γ locus is also rich in evolutionary information. The dot-matrix analysis of the TCR γ locus against itself reveals the dynamic features of the human genome (Figure 8). Approximately 57% of this locus is composed of locus specific repeats that have been duplicated between 2 and 8 times and are indicated by the multiplicity of diagonal lines in the dot matrix. Each repeated element involves both the gene and flanking regions, and the repeats are ordered predominantly in tandem arrays. The dot-matrix identifies two areas of locus-specific repeats in the multigene family: the GV1 subfamily and J-C gene families.

4.1) Generation of the human TCR GV1 family:

In Figure 8, nine large repeats were revealed at the 5' end of the locus. Each of them is associated with a single GV1 gene (5'-Vγ1, Vγ2, Vγ3, Vγ4, Vγ5, Vγ5P, Vγ6, Vγ7, and Vγ8-3'). The first eight repeats are 4.4-4.9 kb in length while the Vγ8-containing repeat is the shortest and contains only the 1.5 kb fragment of the 5' end. The high level of similarity (85-97%) among the nine large repeats suggests recent duplications. To determine the mechanisms involved, the aligned GV1 repeats were surveyed by the phylogenetic scanning method to identify different phylogenetic patterns in different sections of the multiple alignment. The repeats' boundaries and three significantly different phylogenetic patterns were found.

The relationship among Vγ1, Vγ3, and Vγ5, for example, shows two distinct patterns (Figure 9): in region I, Vγ3 is more closely related to Vγ5 (bootstrap value=96%), whereas in regions II and III, Vγ3 is more closely related to Vγ1 (bootstrap value=70% and 100%, respectively). Thus, Vγ3 is likely to be the result of a recombination between Vγ1 and Vγ5, while the 5' half of Vγ3 is contributed by Vγ5 and the 3' half is contributed by Vγ1. A closer examination of the sequence alignment indicates that a crossing-over between Vγ1 and Vγ5 occurred within the SINE regions. Further, Vγ2 and Vγ4 are clustered into a single group (bootstrap value=98%) with less than 3% divergence in regions I-III indicating a recent duplication between Vγ2 and Vγ4 without inter-sequence recombination. Therefore, an ancestral block (Vγ1-Vγ2-Vγ5-Vγ5P-Vγ6-Vγ7-Vγ8) underwent an unequal crossing-over, with one point on Vγ1 and another point on Vγ5. Of the resulting two allelic products, Vγ1-Vγ2-(5'-half-Vγ5 + 3'-half-Vγ1)-Vγ2-5-Vγ5P-Vγ6-Vγ7-Vγ8 and Vγ1 (5' half)-Vγ5P-Vγ6-Vγ7-Vγ8, the latter allele was subsequently eliminated from the genome. The new allele underwent continued changes that turn the 5' half of the Vγ5 + 3' half of the Vγ1 to Vγ3 and the second Vγ2 to Vγ4. In this manner, the Vγ1-Vγ2-Vγ3-Vγ4-Vγ5-Vγ5P-Vγ6-Vγ7-Vγ8 sequence structure can be derived (Figure 10).

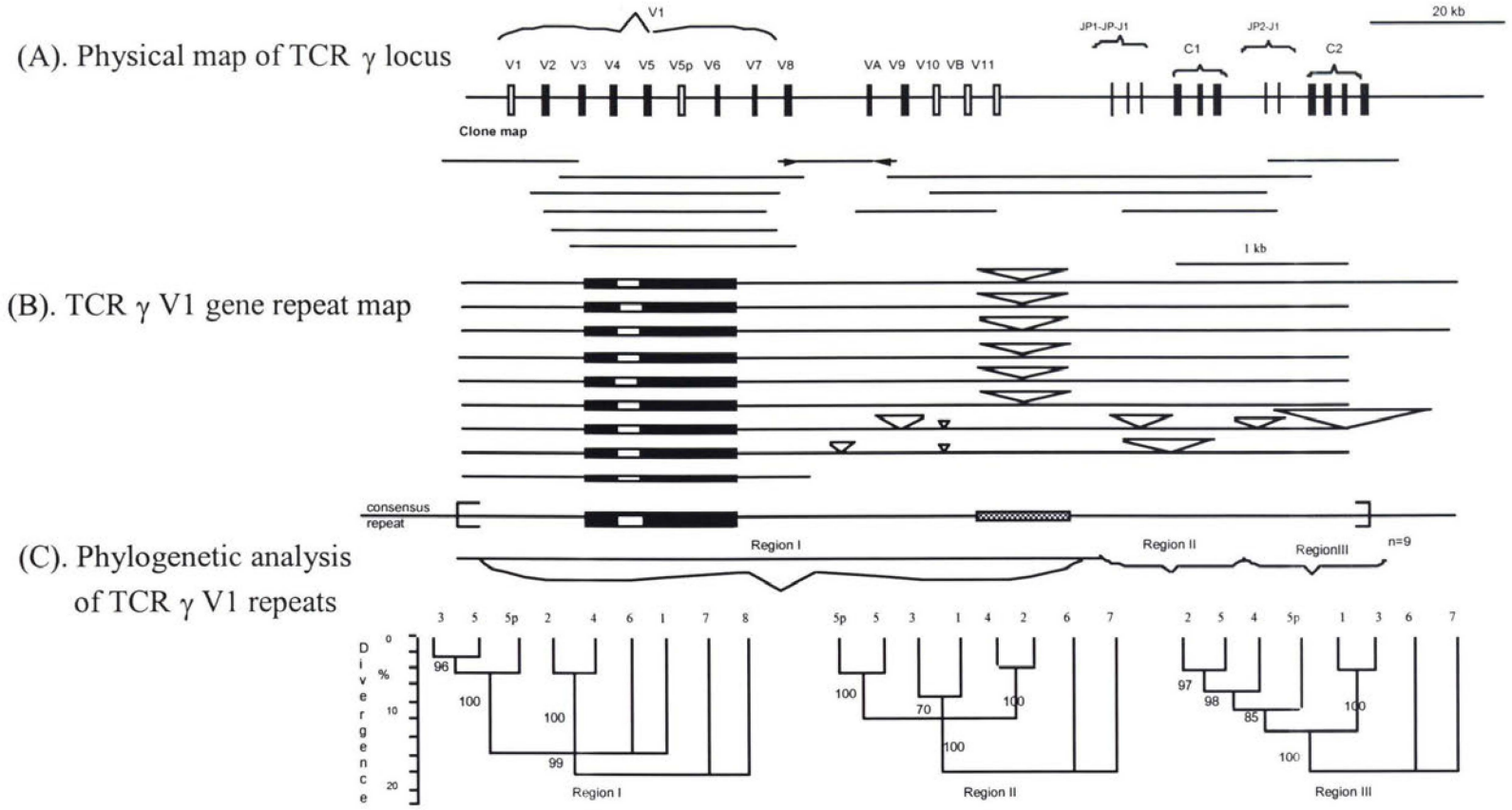


Figure 9. (A) The physical map and clone map of the mouse TCR γ locus. The arrows indicate the two primers for amplifying the DNA fragment. The solid bars indicate that the genes are functional and the boxes indicate the pseudogenes. (B) A schematic map of the repeated sequences is shown where filled boxes represent exons 1 and 2 of TCR γ V1 gene segments. The inserted sequences are given above the main line. All the 300 bp insertions represent Alu type SINE repeats. A consensus repeat map comprised of three regions each showing different patterns of sequence relationship are presented. (C) Confidence levels of branching arrangements in each tree were determined by the bootstrap values between nodes, which are indicated in percent along the margin.

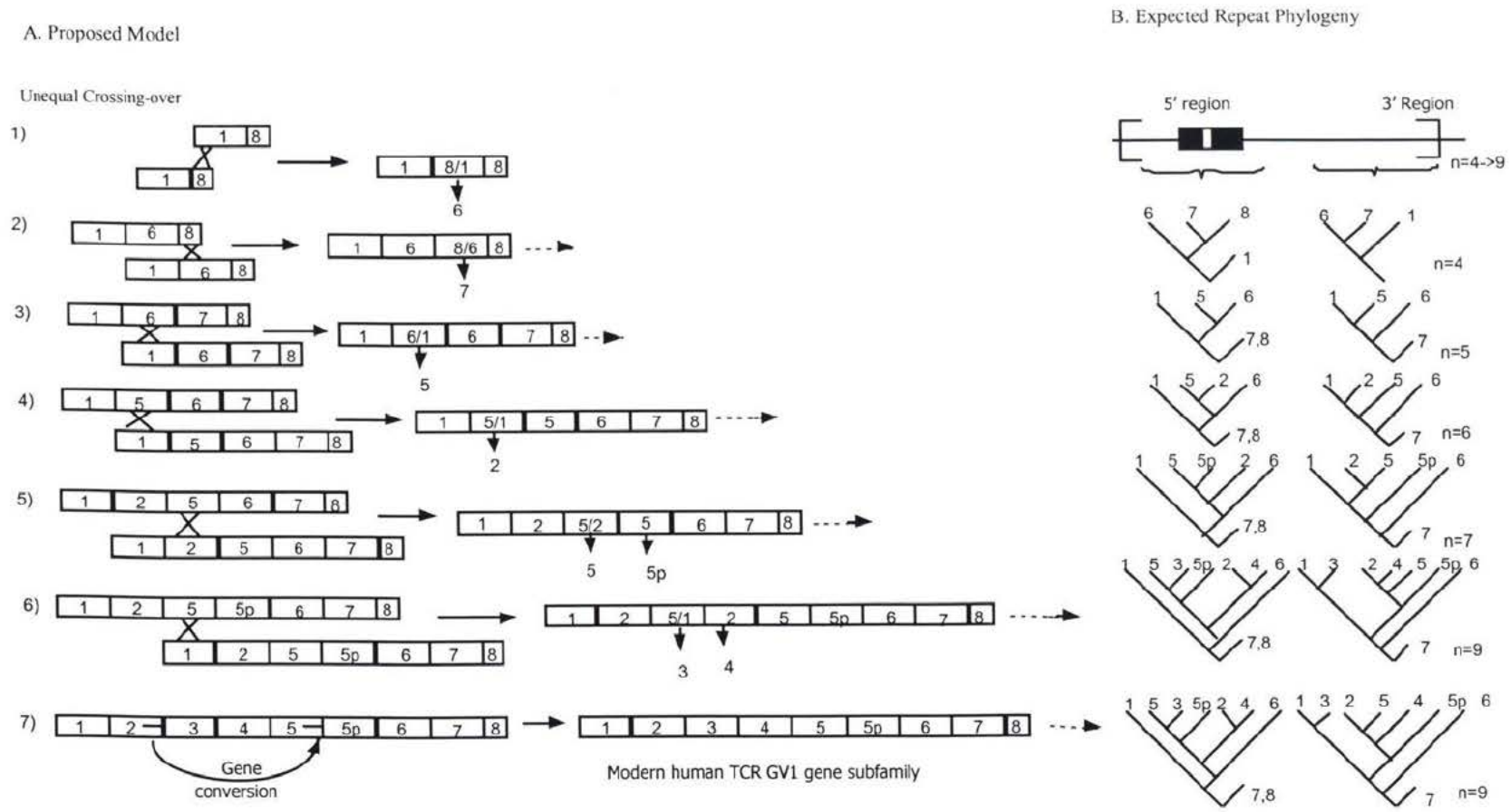


Figure 10. Human TCR GV1 evolution. (A) A model describing six unequal crossing-over events (Stages 1-6) and one gene conversion (Stage 7) event is presented to depict the expansion of the TCR GV1 subfamily. The approximate position of crossing-over between the two alleles is marked with an "X". Solid arrows at each stage point to modern usage. 5 and 5p nomenclature is involved with no change to the model. The dashed arrows at each stage point to the expected repeat phylogeny. (B) Phylogenetic tree for the 5' and 3' regions. In stage 7, gene conversion affects only the barred region of repeats 2 and 5. Each expected phylogeny is consistent with the observed relationships and divergences of the repeats in Figure 9.

However, the highest degree of sequence similarity (over 98%) is found between the Vγ2 and Vγ5 in region III (bootstrap value 97%), but the two sequences are much more distantly related in regions I and II. This segment similarity can be explained by a very recent gene conversion between the two sequences in region III. It is likely that the Vγ5 was converted by the Vγ2 in region III after the Vγ2-Vγ4 duplication because the Vγ2-Vγ4 divergence in region III is similar to that in regions I and II, and higher than the Vγ2-Vγ5 divergence in region III.

Applying the logic outlined, a more complete model of GV1 subfamily evolution was derived (Figure 10). In this model, the phylogeny of each ancestral state is consistent with the observed relationships in regions I and II as shown in Figure 9. From an ancestral Vγ1 and Vγ8, the modern GV1 genomic structure and the observed phylogenetic patterns can be derived by performing a step-wise series of duplications involving six unequal crossing-overs and one final sequence conversion. The last unequal crossing-over leads to the generation of Vγ3 and Vγ4, and is followed by the conversion of Vγ5 by Vγ2 in region III. The proposed model of GV1 sequence evolution accounts for: 1) the recent increase in the number of GV1 genes and their organization; 2) the specific boundaries and complex phylogenetic patterns observed for the different regions; and 3) the pairwise sequence divergences that reflect the age of the different events. Since the recombination of V, J and C fragments is the important source for the generation of immune diversity, gene duplication of V gene fragments by unequal crossing-over provides an important mechanism for supplementing the variability of the human TCR γ.

A recent attempt to reconstruct the duplication history of the TCR GV1 family was performed based on the analysis of the 1318 bp repetitive sequences containing 500 bp upstream of the V gene fragments and 500 bp downstream of the V fragments (Elemento et al., 2002). The sequence of the evolutionary history was inferred as (Vγ1-Vγ2)->(Vγ1-Vγ7-Vγ8)->(Vγ1-Vγ6-Vγ7-Vγ8)->(Vγ1-Vγ3-Vγ6-Vγ7-Vγ8)->(Vγ1-Vγ3-Vγ4-Vγ5P-Vγ6-Vγ7-Vγ8)->(Vγ1-Vγ2-Vγ3-Vγ4-Vγ5-Vγ5P-Vγ6-Vγ7-Vγ8). Vγ2-Vγ3 and Vγ4-Vγ5 were believed to be products of block duplications, which is consistent with Vγ4-Vγ5 being

absent in some individuals (Ghanem et al., 1989). However, the sequences used are only limited within region I and therefore the phylogeny is the same as that of region I in Figure 9. Obviously, the entire sequences of the tandem repeats were not accounted for and therefore, the inference are limited. If the generation of Vγ2-Vγ3 and Vγ4-Vγ5 were due to a block duplication of two V fragments, then it would be difficult to explain the high similarity between region III of Vγ2 and Vγ5 because they are not the direct duplication products of the corresponding position. Therefore, the absence of Vγ4 and Vγ5 in some individuals is due to deletion.

4.2) Positive natural selection of the GV1 family members:

Another source for generating TCR variability is within the gene itself. Within each V gene, and more specifically within exon 2, three regions appear to be hypervariable and may provide places for contact for antigen recognition. These hypervariable regions (18 bp, 18 bp, and 73 bp), the remainder of the gene (framework regions) and the extended flanking regions of the GV1 subfamily were examined for the relative distance values to characterize the evolutionary mechanism in different functional regions.

As summarized in Table 6, the comparison of divergence values among the subfamily members reveals four major patterns of GV1 variation. First, the pairwise divergence values calculated from the flanking regions (D) and from the silent or synonymous position (DS) of the gene framework regions are very similar (average D=12.5%; DS=11.5%; linear positive correlation, r=0.76). Therefore, indexes D and DS provide an estimate of the neutral rate of substitution and a base line to assess the direction and level of natural selection acting in the GV1 gene family. Second, in the GV1 framework regions, divergence calculated from amino acid changes or nonsynonymous sites (DN) are much lower than that of synonymous sites and of flanking sequences. This observation is more evident in the functional genes (DN/DS=0.49) than in comparisons involving pseudogenes (DN/DS=0.72-0.93). Thus, GV1 framework regions show negative or purifying selection forces acting to make the framework region of the GV1 protein domain conserved because of their deleterious effects (Wagner, 2001). As genes become non-functional pseudogenes, the level of this purifying selection decreases. The

Functional genes comparison	Region I D	HV D_N	HV D_S	FW D_N	FW D_S
2--3	0.09	0.445	0.1611	0.0279	0.0336
2--4	0.031	0.1034	0.0746	0.011	0.0516
2--5	0.099	0.3618	0.209	0.0601	0.042
2--8	0.145	0.309	0.1152	0.0634	0.1036
3--4	0.095	0.4437	0.1622	0.0223	0.0865
3--5	0.03	0.0879	0.0385	0.0282	0.0498
3--8	0.153	0.4951	0.1046	0.0458	0.1391
4--5	0.106	0.4424	0.2104	0.0543	0.0951
4--8	0.148	0.3265	0.116	0.0515	0.1217
5--8	0.164	0.4491	0.1488	0.064	0.1573
Mean	0.1061	0.34639*	0.13404	0.0423*	0.08803
Pseudogenes and functional genes comparison	Region I D	HV D_N	HV D_S	FW D_N	FW D_S
1p--2	0.105	0.3922	0.2781	0.0632	0.0681
1p--3	0.102	0.345	0.2411	0.0398	0.67
1p--4	0.111	0.3739	0.3263	0.0514	0.1229
1p--5	0.111	0.4048	0.2971	0.0637	0.0841
1p--8	0.142	0.4984	0.2903	0.0701	0.1765
5p--2	0.104	0.446	0.2521	0.0761	0.0694
5p--3	0.048	0.0945	0.1008	0.0462	0.0682
5p--4	0.108	0.4893	0.2539	0.0701	0.1253
5p--5	0.05	0.081	0.1463	0.0644	0.0502
5p--8	0.165	0.4093	0.1885	0.08	0.1709
6p--2	0.126	0.4159	0.2058	0.1201	0.0887
6p--3	0.121	0.3514	0.2132	0.1013	0.1251
6p--4	0.13	0.3941	0.2074	0.1204	0.1465
6p--5	0.128	0.2955	0.215	0.1343	0.1437
6p--8	0.156	0.3214	0.1419	0.1151	0.1616
7p--2	0.165	0.3043	0.1199	0.1009	0.1074
7p--3	0.154	0.2582	0.0593	0.0826	0.1056
7p--4	0.172	0.3401	0.1207	0.0947	0.1263
7p--5	0.155	0.2412	0.1021	0.1071	0.1238
7p--8	0.145	0.2827	0.0854	0.0706	0.1611
Mean	0.1249	0.33696*	0.19126	0.083335*	0.11462
Between pseudogenes	Region I D	HV D_N	HV D_S	FW D_N	FW D_S
1p--5p	0.116	0.3432	0.2993	0.083	0.0855
1p--6p	0.148	0.4243	0.2872	0.1278	0.1048
1p--7p	0.157	0.3248	0.1898	0.102	0.1039
5p--6p	0.124	0.3156	0.2074	0.1293	0.1263
5p--7p	0.16	0.2515	0.1178	0.1063	0.1357
6p--7p	0.167	0.3033	0.1378	0.1279	0.1714
Mean	0.14533333	0.3271*	0.20655*	0.112716*	0.1212666
Total mean	0.1249	0.3369*	0.19226*	0.083335*	0.11462

Table 6. Synonymous (DS) and non-synonymous (DN) divergence calculated from TCR gamma V1 genes, and the D calculated region I of the repeat unit. All values are corrected for superimposed substitutions. * indicates significant differences ($P<0.05$) between HVDS, HVDN or FWDN compared to FWDS or D.

first two patterns of variation observed in the GV1 gene subfamily closely follow the common expectations of most genes. The third pattern observed is that the divergence of both DS and DN in the hypervariable regions of the GV1 gene are much higher than that in framework regions of the genes or the flanking sequences. In other words, both the DS and DN in the hypervariable regions occur at a much higher level than those of the selectively neutral DNA (Figure 11). Possible explanations for this observation can involve micro-gene conversion but this does not adequately account for this observation. The fourth and most striking pattern observed is that in the hypervariable regions of the GV1 genes, DN is much higher than either DS or D (Table 6). This difference is much more evident in the functional genes (DN/DS=2.6) than in the pseudogenes (DN/DS=1.6-1.8). These observations suggest that amino acid altering nucleotide substitutions are increased and positive Darwinian selection is more operative on the hypervarible regions of functional GV1 genes than on pseudogenes.

Identification of positive Darwinian selection in the hypervariable regions of GV1 genes explains why amino acid altering nucleotide substitutions occur more frequently than expected. Another attempt to explain the mechanism of the increasing DS and DN in the hypervariable regions emanates from observations of IgVH (Ota & Nei, 1994; Ohta, 1995) and MHC (Hughes et al., 1988; Hughes, 1995; Seemann et al., 1986) in chicken in which hypervariablity was generated by micro gene conversion. In Ohta's analysis of the MHC hypervariable region, it was found that DN was not larger than DS. Additionally, a patchwork pattern of gene sequences is present at MHC (Kappes & Strominger, 1988). Ohta concluded that gene conversion contributes to the amino acids variability in CDRs (Ohta, 1995). However, as noted by Ota (1994), the compared sequences should be closely related because CDRs evolve rapidly and include many indels. When the data is treated properly, DN>DS is produced in CDRs but DN<DS is produced in FWs. In our analysis, phylogenetic trees derived from the hypervariable regions of TCR GV1 genes are identical to those of the surrounding flanking regions (Region 1). Therefore, gene conversion does not adequately explain for such a high level of identical variations. There remains the possibility that hypervariable regions of the gene undergo changes at a higher

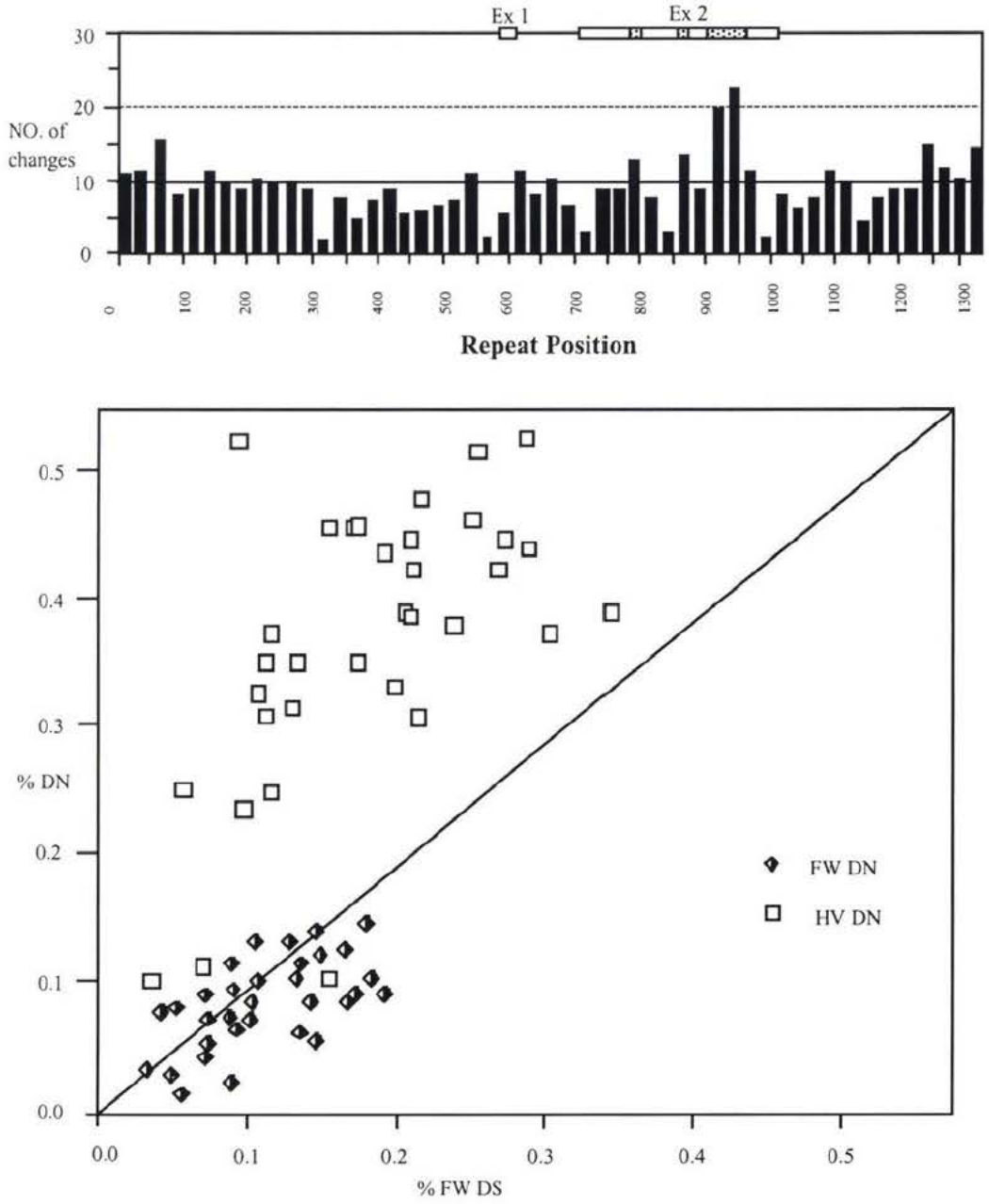


Figure 11. Nucleotide divergence in the human TCR γ V1 region. (A) Hypervariable areas are determined from plotting the minimum number of change over 25 bp intervals over the first 1.4 kb of the repeat region containing TCR γ V1 gene segments. Exons are placed over the graph with hypervariable regions 1, 2 and 3 in exon 2, noted by gray boxes (Vγ4; HV1AEGSTG - HV2 SSYTSSV - HV3 PGKYDTYGSYRKNLRMILRNLIEN). (B) Positive natural selection versus purifying selection is shown by plotting nonsynonymous (DN) divergences from framework (FW) and hypervariable (HV) regions against synonymous (DS) divergence in framework (FW) regions of the TCR γ V1 exons. The same results were obtained when D was substituted for FWDS.

rate than other regions because of rapid fixation associated with positive selection (Rowen et al., 1996).

4.3) GV1 family evolves through a Birth-and-death process:

Although the concerted evolutionary process has been used to explain the evolution of multigene families (Ohta, 1995), much recent research on the immune system genes revealed that the general mode of immune system multigene families generate genes of diverse functions under the birth-and-death evolution (Su & Nei, 2001). Here, we also propose that the human TCR GV1 family members produced by the above unequal crossing-overs evolve in the birth-and-death process rather than through concerted evolution. This view is based on the following evolutionary and organizational features of the human TCR γ locus.

Firstly, the human GV1 family satisfied many of the characteristics of the birth-and-death evolution. In the birth-and-death model, it was assumed that new genes are created by repeated gene duplications and that some duplicated genes may reside in the genome for a long time and diverge functionally, whereas others may be deleted or become nonfunctional within a few MYAs (Lynch & Conery, 2000; Nei et al., 2000). Among the nine GV1 family members, five are functional genes and four are pseudogenes (Lefranc et al., 1986a). As previously mentioned, the functional V genes diverged greatly at the CDR regions in order to promote their functional diversification by positive selection. The four pseudogenes became nonfunctional between 7.2 and 37.7 MYAs ago based on calculations of a molecular clock. This time estimate is close to the theoretical estimation of half-life being present 3-7 MYAs in a multigene family on the loss of functional genes (Wagner, 2001). All the above evidence indicates that the human GV1 family is generally subject to evolution by the birth-and-death process.

Secondly, the features of the human TCR γ GV1 family are not consistent to those of concerted evolution, in which gene members of a family are assumed to evolve as a unit, exchanging genetic information by interlocus recombination or gene conversion (Smith et al., 1971; Zimmer et al., 1980). If this process continues, duplicated genes in a multigene

family tend to display a similar nucleotide sequence. Therefore, if the human TCR γ GV1 family evolved in the concerted evolution process, the intergenic gene conversion or unequal crossing-overs should have occurred very often. It would subsequently be expected that both the functional and pseudogenes would be mixed so that most genes in the genome would possess a similar nucleotide sequences evolution, and thus all the nucleotides on the gene would be affected without discrimination. However, the extent of sequence divergence between the functional and dysfunctional genes in the human TCR GV1 family is substantial, which suggests that the TCR GV1 genes are not mainly subject to frequent gene conversion or unequal crossing-overs. In fact, in the immune system, the prominent feature of evolution is the appearance of immune specificity, which is required to adapt to ever-changing antigen environments to cope with the antigens. This does not mean that there is no homogenization in this locus because this process exists as long as unequal crossing-over occurs. However, it is obvious that homogenization is not the major evolutionary process. In fact, the birth-and-death process, which generates the gene's diverse functions, is the major evolutionary process.

4.4) Generation of J-C regions:

The two J-C clusters also share great similarity to one another at both the organization and sequence level. Organizationally, the coding and flanking regions of JP1, J1 and Cγ1 in the cluster 1 (3' side cluster) correspond to those of JP2, J2 and Cγ2 in cluster 2 (5' side cluster) respectively. However, there are two exceptions that do not have their counterparts present in another cluster. One is the 3.79 kb JP element which is present only in cluster 1 and missing in cluster 2; and another situation is that two exon 2 exist in Cγ2 instead of the single copy in Cγ1. Among the three mentioned corresponding regions, JP1/JP2 regions are 2.4-2.7 kb, J1/J2 regions are approximately 7.4 kb, and the Cγ1/Cγ2 regions are 3kb long respectively. The total sequence identity between them is 95.7%. The above sequence and organizational similarity suggest that the two J-C clusters are the result of a DNA block duplication of either JP1-JP-J1-Cγ1 or JP2-J1-Cγ2. Continued changes have lead to a different addition or deletion of JP and exon 2 between the two clusters.

In order to better understand the J-C clusters duplication history, a further examination of the connecting region between the two clusters is necessary. The region between exon 3 (Cγ1) and JP2 is compared with the two possible homologous regions between the 5' part of JP1 and the 3' part of exon 3 (Cγ2). The sequences were aligned and analyzed by phylogenetic scanning. The results show that the 5' region of the exon 3 (Cγ1)-JP2 region is remarkably similar to the 3' exon 3 (Cγ2), while the 3' part of the exon 3 (Cγ1)-JP2 region is remarkably similar to the 5' JP1 (Figure 12). Such a "hybrid" sequence structure of the connecting region suggests a previous homologous DNA recombination. It is likely that two copies of the primordial TCR γ locus, which has only a single J-C cluster, have undergone unequal crossing-over, with one point at the 3' exon 3 of the first sequence, and another point at the 5' JP1 of the second sequence (Figure 12). As a result, a new organization of the TCR γ locus, with duplicated J-C clusters and a hybrid structure in the connecting region, was created and obviously preserved in the genome. Another product of the unequal crossing-over, which has a deletion of the entire J-C region, was eliminated from the genome. The above resulting locus continued to differentiate to lead to the modern J-C structure.

As mentioned, one of the two major differences between the two J-C clusters is the absence of the JP sequence in the second J-C cluster. This is due to a further crossing-over event after the J-C cluster duplication. Using the similar phylogenetic scanning method adopted, a hybrid structure between JP2 and J2 was revealed, which has its 5' side contributed by the JP1-JP sequence, and its 3' side by the JP-J1 sequence (Figure 13). The JP1-JP and JP-J1 sequences are homologous to the original JP2-JP and JP-J2 sequences, respectively, due to the J-C duplication. Assuming both J-C clusters possess the JP element in the original form, it is apparent that two sets of the original TCR γ locus have undergone an unequal crossing-over with one point between JP2 and JP of the first sequence, and another point residing between JP and J2 of the second sequence (Figure 13). The crossing-over resulted in the modern TCR γ structure, in which the JP is deleted from the second J-C cluster, of which this structure prevails. But, the other product of the crossing-over with a duplicated JP in the first J-C cluster was eliminated from the genome.

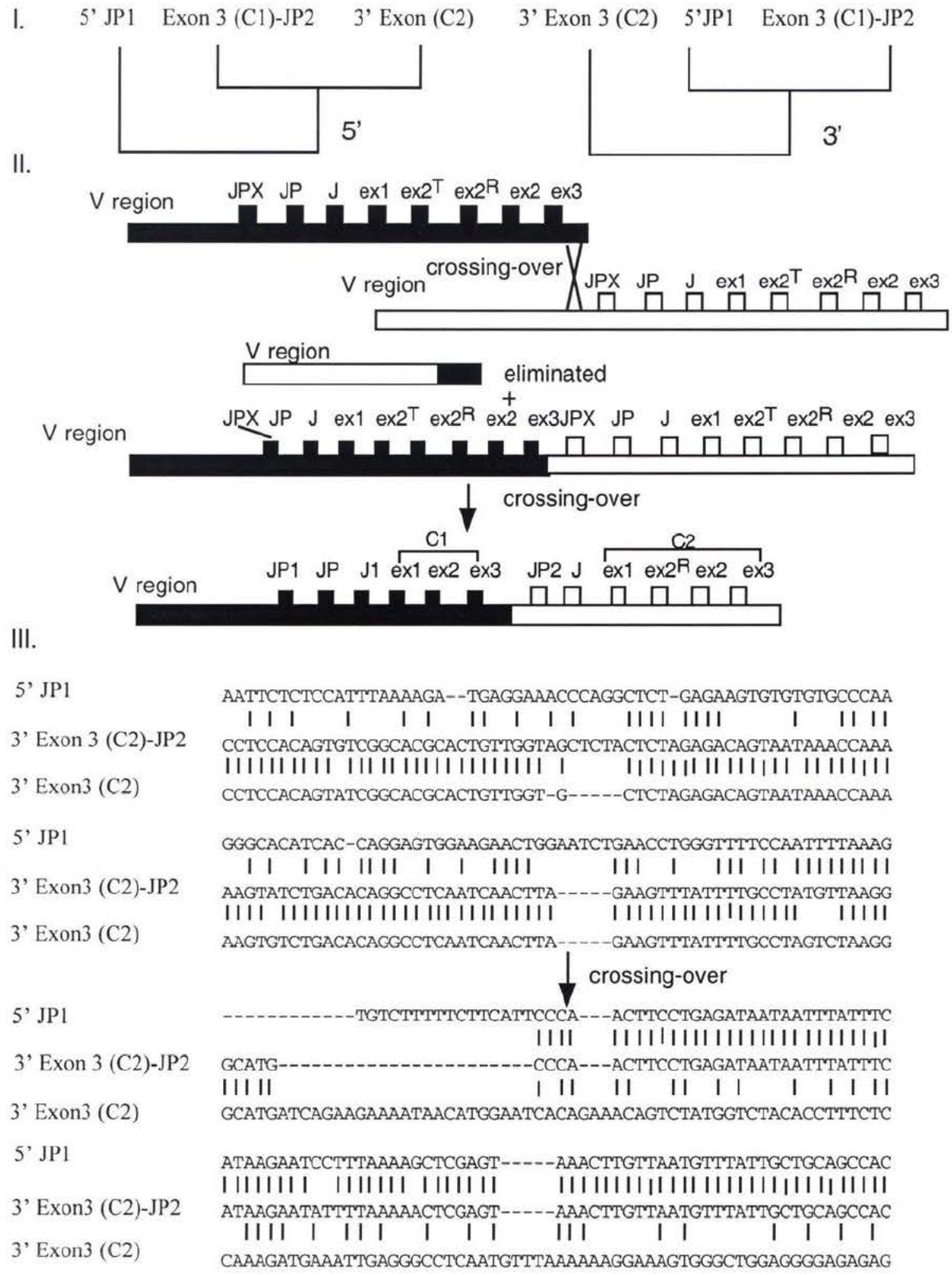
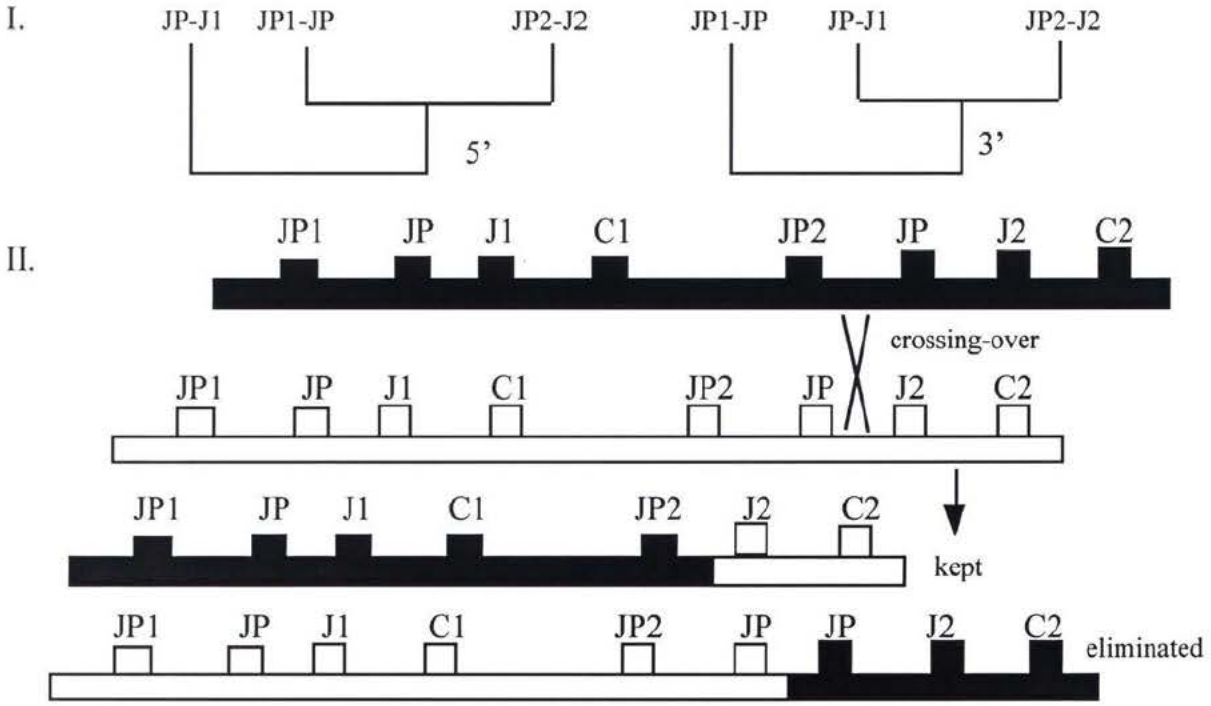


Figure 12. The J-C cluster duplication by crossing-over. (I) the result of the phylogenetic scanning. (II) schematic pattern of crossing-over. (III) the crossing-over site and flanking sequences.



III.

```
JP-J1    ACAAAGTTGAAACTTCTGTTTTCACGTTGATCTGTACAAAGAAGATAATGGAGATTATTT
JP2-J2   A-----ATGAAAAA-CTGATA----GATGTTTTTTAGAAA-AAAATGATCAACCTTACCT
JP1-JP   A-----ATGAAAAA-CTGATA----GATGATTTTTAGAAA-AAAATGATCAACCTTATCT

JP-J1    ATCTAAAGTGGATGT-TATAAAGATTAAATGAAATTAAATATGGAAATTACCTACCAGAA
JP2-J2   GAATATGTCACATTCCTGGCCTCAGTATACGAACAGCAATTTTTCAGG---GTAGCTGAA
JP1-JP   GAATATGTCACATTCCTGGCCTCAGTATACGAACAGCAATTTTGCAGAT---TAGCTGAA

                 crossing-over site
JP-J1    TGCCTGGCACTTGGTAGGTGATAGAAATATTATTTCCTCTTCTA-GAATTCAAGTTACTA
JP2-J2   TGCCTGGCACTTGGTAGGTGATAGAAATATTATTTCCTCTTCCA-GAATTCAAGTTACTA
JP1-JP   TA--TGGAAACAAAACGATTTTGTACATGATATCTTCCCTTCCCTGTCTGTAAG------

JP-J1    CCCACGAAAGGTCTTTATACACTCAGGAA-TACAGATAAATGTATCGACTCCCATCTATT
JP2-J2   CCCAAGAGAGGTCTTTATACACTCAGGAA-TACAGATAAATGTATCAATCCCCATCTATT
JP1-JP   CGGATTAAAATTCT----ACTTTAAGAATACACACACACACACACACACACACACACA
```

Figure 13. The JP element deletion from the second J-C cluster by crossing-over. (I) the result of phylogenetic scanning. The trees were built by UPGMA using PAUP4.00d55. (II) schematic pattern of crossing-over. The X indicates the sites where the crossing-over takes place. (III) the crossing-over sites and flanking sequences.

The second difference between the two J-C clusters is the existence of multiple copies of exon 2 in Cγ2, which are obviously the result of DNA duplication by a similar unequal crossing-over. For example, our genomic sequence reveals a two-copies exon 2 Cγ2 structure. Theoretically, two models may account for the generation of such a structure of the human TCR γ locus. One postulates that the primordial C contained a structure of exon 1-exon 2T-exon 2R-exon 2-exon 3 and that a further deletion of two copies of exon 2 gave rise to Cγ1. Another postulates that the primordial C possessed only two exon 2 and the Cγ1 and Cγ2 were produced from continuous deletion and duplication respectively. However, the phylogenetic analysis shows that the C gene with three copies of exon 2 is the ancestral form, whereas the present Cγ1 and Cγ2 are derived by deletions. Therefore, the model that C genes were duplicated from an ancestral C gene with a single exon 2 is rejected (Buresi et al., 1989). Further, exon 2 in Cγ2 also shows polymorphism in different populations. For example, a typical double exon 2 structure is found to be the dominant Cγ2 form in the Chinese, French, Lebanese and Tunisian populations (Buresi et al. 1989; Ghanem et al. 1989, 1991), whereas the triple exon 2 in Cγ2 structure, which is found in black Africans in addition to other structural forms, is the dominant form in this population (Buresi et al. 1989). The relatively high polymorphism in black Africans is consistent with the finding that this population usually presents a greater genetic diversity than other ethnic groups (Craddock et al., 2000).

Chapter III. Characterization of the mouse TCR gamma locus

Summary:

The gene fragments encoding mouse αβ and γδ TCRs are distributed in the α/δ, β and γ multigene loci on chromosomes. The mouse TCR γ locus is a 205 kb region, in which the V, J and C gene fragments form four V-J-C clusters. We have sequenced and characterized the 168 kb sequence covering both the 5' and 3' ends of the mouse locus. The coding gene fragments, enhancers, and the recombination and splicing signals were precisely mapped to the genomic sequence. Structural analysis revealed that this locus is characterized by a low GC content, rich LINEs, low SINEs and a low number of genes, which conforms to the features of the Gimsa-dark band. By comparing the genomic sequence with the sequence of the known enhancer γE1, a new putative enhancer, γE4, was identified 3.7 kb downstream of the 3' end of the Cγ4 gene. The dot-matrix self-comparison revealed a few large locus-specific repeats. The phylogenetic analyses of these repeats suggest that clusters 2 and 3 are recent duplications, whereas clusters 1 and 4 are ancient clusters. The independent regulation, recombination and expression of the genes in each cluster, suggest that the genes in the locus are evolving toward diverse functions. Due to the high possibility of recombination created by the large number of repeats, including the locus specific repeats and LINEs, cluster 2 reversed its transcriptional orientation to counteract the potential recombination and thus preserved the genetic information. These dynamic organizational features suggest that the mouse TCR γ locus reshapes its genome structure constantly via multiple mechanisms in order to fit ever-changing selectional pressures.

Introduction:

During organismal evolution, diverse types of gene recombination processes and natural selection interact to shape the complexity of the genome. In multigene families, the complexity is further increased by the organizational distribution of the functional related repeated gene members. There are three major types of repeated genes, i.e. the tandemly arrayed, clustered and dispersed genes (Graham, 1995). Such organizational patterns result from different evolutionary selection forces that operate on them. Each pattern offers a unique function and significance in the process of the evolution of the gene family. For example, the multiple copy members of the tandem arrays can satisfy the high demand for a given gene product, while the divergent clustered and dispersed genes provide the starting material for evolving to new genes and maintaining the genetic memory during long periods of relaxed selection (Graham, 1995; Schimenti, 1999; Ohta, 2000). Further, the difference between the member genes' distribution also influences their future evolutionary consequences because the key evolutionary process, unequal-crossing over, affects these three patterns of arrays differently. Tandem repeats have more chance of homologous recombination and gene conversion whereas clustered and dispersed genes have only medium and weaker gene conversion respectively. Therefore, an understanding of the organizational complexity of repetitive genes can provide insight into the organization, evolution, and functionality of multigene families.

Mice have been a good model in medical and biological research for nearly a hundred years (Nadeau et al., 2001). This is because they offer small size, short generation times and highly efficient, controllable crossing and breeding. In addition, despite millions of years of evolutionary separation, there is a close homology between many mouse and human gene sequences and therefore they facilitate an understanding of human biology and diseases. There are presently many mouse strains, transgenic mice and well-established mouse models of diseases available (Peltonen & Mckusick, 2001). In the HGP, the sequencing of the mouse sequence is a parallel project, which is helping to decipher the functionality of the human sequence. The comparative analyses between mouse and human sequences will facilitate the rapid understanding of the human genome because of their striking sequencing similarities (Venter et al., 2001). Discoveries in one

species have lead to strong inferences in the other (Nadeau et al., 2001). Actually, many genes are inferred through their homology with the genes which are already known. It was estimated that 95% of human genes could be found with the help of mouse sequences (Burge, 2001; Pennisi, 2001). In addition, comparison of sequences of mice and humans will provide insights into the molecular bases of human disease, and the organization and evolution of the mammalian genome (Gurwitz & Weizman, 2001). At present, phenotypic, genetic and human-mouse syntenic data sets are presented in the major human and mouse databases and are becoming even more comprehensive as sequencing of the mouse genome is advanced (www.ncbi.nlm.nih.gov/Homology; www.informatics.jax.org/menus/homology_menu.shtml; www.informatics.jax.org/mgihome/nomen/allmut_form.shtml).

The mouse TCR γ locus occupies 205 kb of DNA on chromosome 13 and possesses characteristic genomic organizational, gene regulatory and evolutionary features (Rabbitts et al., 1985; Cheng et al., 1991; Hayday, 2000). Since TCR γ was discovered in the early 1980s (Saito et al., 1984), the organization of the coding gene fragments, characteristics of rearrangement and immune diversity, and other features of mouse TCR γ have been studied extensively (Hayday et al., 1985; Garman et al., 1986; Cheng et al., 1991). γδ T cells share many similar features to αβ T cells; for example, they both have critical immune regulatory functions and can perform immune resistance to pathogens (Kaufmann, 1996). However, they also possess substantially different properties. For instance, it was found that many γδ T cells are not restricted by MHC for recognition of antigens, while most αβ T cells are MHC dependent (Chien et al., 1996). In addition, the human and mouse TCR γ loci reveal a different genomic organization. The V genes in the mouse TCR γ locus are distributed in four independent V-J-C clusters whereas the entire human TCR γ V genes are distributed in an almost tandem array at the 5' end and the two J-C clusters at the 3' end (Lefranc et al., 1990). This varies from the highly conserved human and mouse TCR α/δ and β loci. The organizational differences of the human and mouse TCR γ loci suggest that they appropriated different mechanisms to evolve after their separation. Although previous studies based on the coding gene sequences have

revealed some evolutionary inferences about such mechanisms (Vernooij et al., 1993), these results may be biased by the sequences that display different functional constraint and limited length. Further, some large-scale evolutionary events such as unequal crossing-overs, which mainly occur at the non-coding regions, cannot be identified without genomic sequence. In general, such discoveries suggest that complete genomic sequences are highly needed to increase the accuracy of evolutionary analysis and to help to provide the direct evidence of evolutionary events. In this research, we sequenced 168 kb DNA of the mouse TCR γ locus. The sequence was analysed to characterize the features of genomic structure, gene regulation and evolution.

Materials and methods:

1) Large-scale DNA sequencing:

Obtaining the DNA clones: Cosmids covering the TCR γ locus were previously screened from a BALB/c mouse cosmid library and were mapped (Vernooij et al., 1993). Figure 14 shows the seven sequenced cosmids that cover the 5' end and the 3' end of murine TCR γ locus. The 10 kb gap between Mγ5 and Mγ6 was obtained from a *Nhe* I fragment in Mγ72.2. The 5 kb DNA fragment between Mγ3 and Mγ84 was amplified from an overlapping clone using PCR.

Template preparation and sequencing: All cosmid DNA, clones, PCR products and restriction fragments were randomly sheared by nebulization (Bodenteich et al., 1994). 50 ug DNA was nebulized at 30 p.s.i. for 2 minutes. Nebulized fragments were treated to be blunt-ended with Mung Bean nuclease, T4 DNA polymerase and Klenow fragment (Povinelli et al., 1993). The blunt-ended fragments were then size fractionated (1-2 kb) and ligated into blunt-ended *Hinc* II cut and dephosphorylated M13mp19 vector. Preparation of single stranded M13 template DNA was performed using QiaQuick 96 well M13 phage DNA purification kits. The PCR reactions were prepared using the dye-primer mixture of Perkin-Elmer Corp. Lastly, the PCR products were loaded on an ABI 373 and 377 automatic sequencer (Perkin-Elmer Corp.).

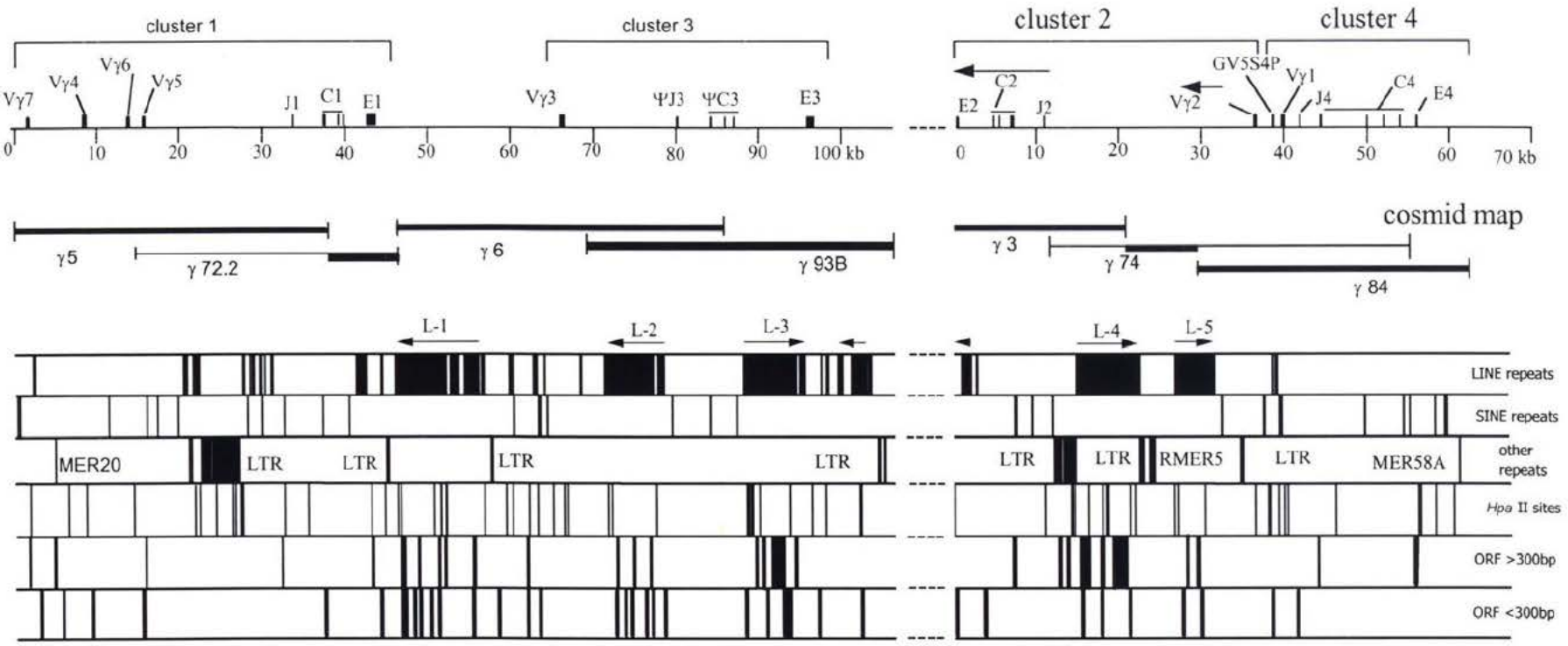


Figure 14. A summary of the features of the mouse TCR γ locus. The top line depicts a map of the 168 kb of sequence with all the known coding regions and enhancers. The middle map shows the cosmids used to obtain the sequence. The thick lines indicated the regions being sequenced. The lower lines are positioned to show the features. The first line indicates the LINE repeats and the arrows indicate the orientations of the major LINEs. The second line marks the position of other repeat elements. The fourth line indicates the *Hpa* II sites. The bottom two lines indicate open reading frames (ORFs) which are larger than 300 bp. The arrows indicate the transcriptional orientation of genes and repeats. The approximately 37 kb internal region not being sequenced is marked by dashes (not in proportion). L-1 to L-5 refer to the big LINEs.

Sequence assembly: The sequence data were assembled utilizing either of the following software: Seqman (DNASTAR) or Phrap and Consed (Phil Green et al., unpublished data, E-mail: phg@u.Washington.edu). An average eight-fold redundancy was needed to obtain a complete consensus.

2) Data search and analysis:

Data search and sequence structural analysis: The known cDNA and partial germline sequences were obtained from GenBank. The positions on the genomic sequence of the coding gene fragments, in addition to the DNA rearrangement and RNA splicing signals, were identified by comparing the cDNA and partial germline sequence in the GenBank with the genomic sequence. RepeatMasker was used to identify and to remove genome-wide repetitive sequences (http://ftp.genome.Washington. edu/cgi-bin/mrs/mrs_req, A.F.A. Smith & P. Green, unpublished data). GRAIL (http://cs6400.ims.u-tokyo.ac.jp:8002/Grail-1.3/, 1996) was employed to characterize CpG island and poly(A) sites. In addition, XGRAIL1.3, BLAST (NCBI), and BLAST/Beauty (Baylor College of Medicine Search Launcher) were used to search for the putative gene fragments.

Phylogenetic analysis: Dot-matrix comparisons were performed to analyze the internal similarity by using GeneAssist5.5 (ABI, Perkin-Elmer Corp) and DOTTER (ftp://sanger.ac.uk/pub/dotter, Sonnhammer and Durbin, 1996). Multiple alignments were executed in MAP software (Huang, 1994). Phylogenetic trees were made by UPGMA, NJ and maximal parsimony methods in PAUP 4.00d55 (Swofford, 1997).

Results and discussion:

Figure 14 summarizes the major organizational and structural features of the 168 kb sequence, including the map of the seven cosmids that have been sequenced in the project. The gene and genomic organization based on our sequence is basically identical to the results previously reported (Vernooij et al., 1993). The 5' end region that includes clusters 1 and 3 is 105726 bp (GenBank: AF037352) while the 3' end region that contains clusters 2 and 4 is 62031 bp (GenBank: AF021335). In the middle of the 205 kb locus, a region around 37 kb without known *cis*-acting and *trans*-acting elements was not

sequenced because the two cosmids covering it were unstable. In addition, the complete sequence of the cosmid Mγ3 showed that part of the pWE15α vector together with the 5' end of TCR γ insert had been deleted but a *Tn*10ISR element was inserted. This indicated that Mγ3 perhaps underwent a rearrangement, which resulted in part of the DNA being deleted.

1. Coding regions:

1.1) Variable genes:

Seven functional genes and one dysfunctional V gene were identified in the mouse TCR γ locus, which is consistent with previous reports (Hayday et al., 1985; Garman et al., 1986; Pelkonen et al., 1987; Kuziel et al., 1994). This number is relatively low compared to the large number of V genes in the other mouse TCR loci. For example, there is a total of seventy-five functional TCR Vα, twenty-three TCR Vβ, and more than sixteen TCR Vδ genes (Arden et al., 1995a & b). In this study, the whole genomic sequence of each V gene comprises two exons and one intron. The lengths of exons and introns of these V genes are listed in Table 7. Vγ1, Vγ2 and Vγ3 display almost identical lengths of intron and extrons. However, the lengths of the other four V gene fragments are quite variable.

Among the seven V genes, Vγ7, Vγ4, Vγ6 and Vγ5 are located at the 5' end of the cluster 1, and are separated by different lengths of intergenic sequences (Vγ7-6475 bp-Vγ4-4854 bp-Vγ6-1349 bp-Vγ5). The other three genes, Vγ1, Vγ2 and Vγ3, are each located at one end of its own cluster. The transcriptional orientation of Vγ2 is reversed because the whole cluster 2 is in an opposite orientation. Although Vγ3 is functional, the defective Cγ3 following it makes all the rearrangements between Vγ3, Jγ3 and Cγ3 in cluster 3 non-functional (Hayday et al., 1985). Until now, Vγ3 bearing TCRs have not been found to be expressed in any tissues, although many transcripts of the Vγ3-Jγ3 rearrangement were found to be present in different T cell clones (Traunecker et al., 1986; Cheng et al., 1991). The 148 bp pseudogene GV5S4P, which is located between Vγ1 and Vγ2, corresponds to only part of the exon 2 of V1. Their sequence identity is

Table 7. Summary of the nucleotide length of exons and introns of the V, J and C gene fragments (the pseudogene is also shown with their corresponding exons and introns).

No. of exon and intron	Size (bp)														
	Vγ5	Vγ6	Vγ7	Vγ3	Vγ1	Vγ2	Vγ4	J1	J2	J3	J4	Cγ1	Cγ2	ψCγ3	Cγ4
Exon 1	46	43	42	112	43	94	43	59	59	59	61	331	330	326	329
Intron 1	105	105	93	109	109	158	102					1432	1424	1452	4648
Exon 2	307	307	304	304	296	312	308					45	30	45	53
Intron 2												568	566	566	741
Exon 3												143	140	140	44
Intron 3															1827
Exon 4															142

90% (Kuziel et al., 1994). This pseudogene is obviously a recent duplication product but died quickly after birth.

Traditionally, a threshold value of 75% similarity at the nucleotide level has been used to define V gene subfamilies (Barth et al., 1985). Therefore, Vγ1, Vγ2 and Vγ3 are considered to be a V1 subfamily because they have a 94% DNA sequence identity at the coding gene level and a 94-95% sequence identity at the genomic DNA level. The high sequence similarity suggests that these three Vγ genes are recent duplication products. However, the genomic sequence similarity between the four V genes in cluster 1 is only 35-46%. The four members in cluster 1 also exhibit little similarity to the V1 subfamily members except for Vγ4 which shares only a 65-72% similarity with V1 subfamily members at the nucleotide sequence level, and a 48% at amino acid sequence level (Garman et al., 1986). Hence, the Vγ genes in cluster 1 and the V1 gene family members have separated from their ancestor for a long time. Each group has since evolved independently, involving little DNA exchange between them.

1.2) Constant genes:

A total of four Cγ genes were found in the sequence of the mouse TCR γ locus. Cγ1, Cγ3 and Cγ4 are located at the 3' end of their cluster whereas Cγ2 is located at the 5' end of cluster 2. The genomic sequences of Cγ1, Cγ2 and Cγ3 display a similar organization. Each of them is composed of three exons and two introns (Hayday et al., 1985). As summarized in Table 7, the length of the exons and introns of Cγ1, Cγ2 and Cγ3 is similar. It was proposed that exon 1 encodes the extracelluar-immunoglobulin-constant region-like domain, exon 2 encodes the first part of the hinge, while exon 3 encodes the second part of the hinge, the transmembrane region and the cytoplasmic domain (Littman et al., 1987). Similar to their cognate Vγ genes, the genomic sequence identity of these three C genes is 93-95%, which suggests that these three Cγ genes are also recent duplications.

Cγ3 was originally thought to be a pseudogene resulting from a defective splice site in exon 2 and a defective polyadenylation signal (Hayday et al., 1985). In our sequence, the 3' splice signal of exon 2 is TTAGGT instead of the universal GTAGGT, which is consistent with the previous finding. Therefore, Cγ3 is unlikely to generate a mRNA that can be properly spliced. As for the reported defective polyadenylation signal, it refers to the sequence AATATA at 391 bp downstream from exon 3 of Cγ3, which is not the conventional polyadenylation signal AATAAA (Strachan & Read, 1996). However, there is another AATAAA located 823 bp 3' to Cγ3 and therefore, Cγ3 may use this sequence as a polyadenylation signal (Figure 15A). Actually, Cγ1 which is homologous to Cγ3, uses the AATAAA signal 621 bp downstream from its exon 3 instead of the AATAAA at the 387 bp position, which further supports the above hypothesis (Hayday et al., 1985). Thus, it is possible that the defective splicing signal of exon 2 is the sole factor responsible for the dysfunction of Cγ3. In addition, it has been suggested that cluster 3 is polymorphic due to its deletion in some mouse strains (Traunecker et al, 1986). For example, in the genomic sequence of C57BL/10 (B10) mice, the fragment from Vγ3 to Cγ3 is not present. Hence, we also tried to identify the range of this deletion more accurately based on our genomic sequence information. The Southern blot of genomic DNA showed that BALB/c mice have a 13.4 kb *EcoR* I fragment containing Cγ1 whereas B10 mice do not. Instead, a longer fragment above 15 kb appeared in the B10 mice (Iwamoto et al., 1986), which suggests that this longer fragment was generated by the combination of two *EcoR* I fragments. Therefore, it was estimated that one deletion site might occur within the 13.4 kb *EcoR* I fragments (from 32300 bp to 45662 bp of the 5' end sequence) because there is an absence of another fragment sufficiently close and long to generate a 15 kb fragment by deleting the middle sequence. Another deletion site may be somewhere between γE3 and the 5' end of cluster 2 although the exact position cannot be determined because the sequence between clusters 2 and 3 is not available. The broad absence of Cγ3 in other inbred mice, such as C58/J, DBA/2N and NZB/J, suggests that Cγ3 or even the whole cluster 3, is not functionally important and therefore were deleted (Huppi et al., 1986).

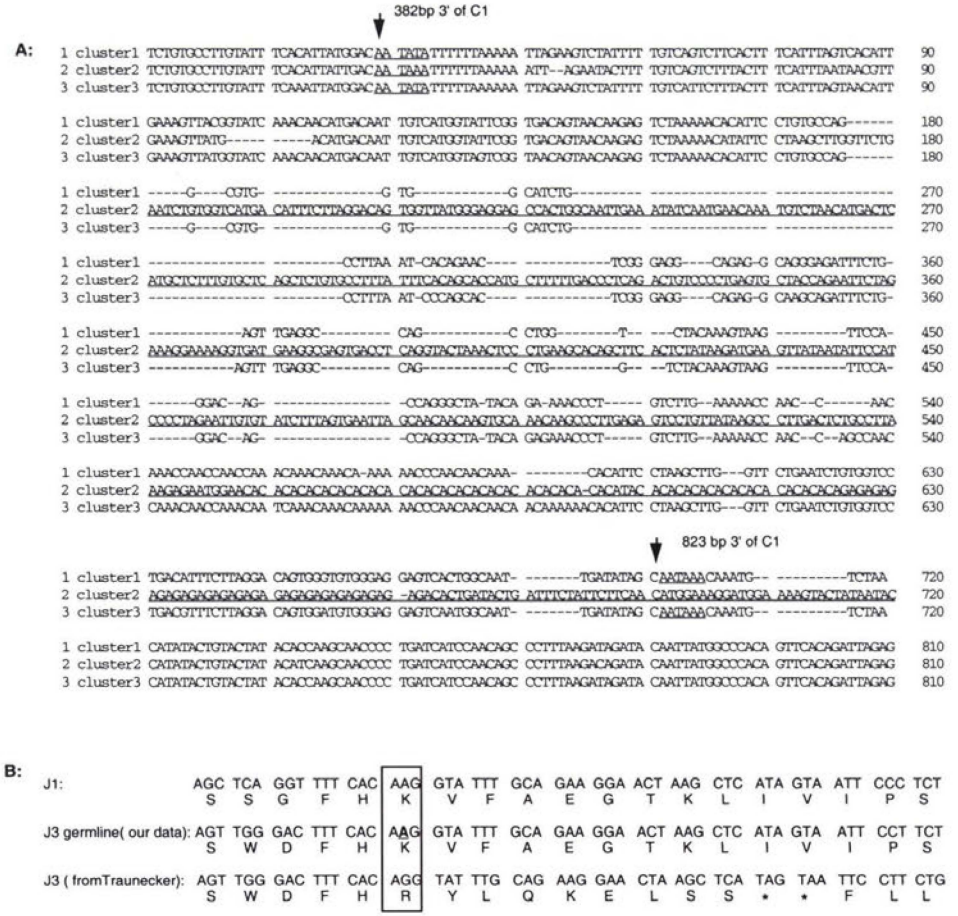


Figure 15. (A) Alignment of the sequences downstream of the exon 3 of Cγ1, Cγ2 and Cγ3. The two putative polyadenylation signals are marked by arrows. The underlined sequence of cluster 2 is an insertion of SINE. (B) Alignment of the germline sequence of J3 with J3 sequence from Traunecker (1986) and J1 sequence. The deduced amino acids are shown below. The regions with the nucleotide A lost in J3 (Traunecker) was boxed and the internal stop codons are indicated by a *.

Unlike Cγ1, Cγ2 and Cγ3 genes, the mouse Cγ4 includes four exons instead of three. In the genomic sequence of Cγ4, the first three exons are similar to those of the other three Cγ genes in organization, but the distance between exon 3 and exon 4 is 2786 bp. Therefore, the genomic sequence similarity is only 45% between Cγ4 and the other three Cγ genes although the coding sequence similarity is 66% (Vernooij et al., 1993). The Cγ1-bearing TCRs are disulfide-linked, while the Cγ4-bearing TCRs are nondisulfide-linked. Thus, Cγ4 may be an isotype with functions that vary from the other three Cγ genes. These structural features of Cγ4 are similar to those of human Cγ2. For example, the human Cγ2 chain is also polymorphic for the number of hinge-encoding exons (Littman et al., 1987). Therefore, the mouse Cγ4-bearing T cells might possess cellular features similar to those of the human Cγ2-bearing T cells. An experiment by Grossi et al (1989) showed that the morphology of T cells expressing human Cγ1 is different from that of the human Cγ2-expressing cells, which are characterized by prompt adhesion and spreading on culture, while the Cγ1-bearing T cells are not. The human Cγ2 bearing T cells exhibit abundant microtubules, which suggest that they are capable of active motility. Such conclusions suggest that the Cγ2-bearing cells are suitable to live in tissues due to their active migration ability whereas the Cγ1-bearing cells possess a property suitable for circulating in the blood stream. In mice, the majority of the γδ T cells in the peripheral blood are Cγ1 positive whereas Cγ4 positive T cells are located in spleen, intestine and skin (Hass et al., 1993). These differential tissue distributions may be attributable to the Cγ1+ and Cγ4+ T cells as they possess similar cellular properties as the human Cγ1 and Cγ2 -bearing T cells, respectively.

1.3) J genes:

Four J gene fragments were identified in the mouse locus. Compared to the 50 J genes in the mouse TCR α/δ locus, this number is relatively low (Koop et al., 1992). Table 7 shows the length of the J gene fragments. They share an 85 to 94% sequence similarity. All of them possess the typical structure of TCR J genes: a core sequence (Phe-Gly-x-Gly-Thr) flanked by a 5' recombination signal and a 3' splice signal. The 5' recombination signal is composed of a canonical nonamer (GGTTTTTGT), a 12 bp

spacer, a hepatmer (CACTGTG), and the 3' splice signal (GTAAGT) (Koop et al., 1994). It has long been believed that J3 is a pseudogene due to an internal stop codon (Traunecker et al., 1986). However, no stop codon was detected in our sequence. The amino acid translation of J3, SWDFHKVFAEGTKLIVIPS, is quite similar to the amino acid sequence of J1 and J2, SSGFHKVFAEGTKLIVIPS. Comparison of the genomic sequence with the J3 sequence of Traunecker et al (1986) suggests that the stop codon in their sequence was due to a missed base (Figure 15B). Therefore, the pseudogene Cγ3 appears to be the only factor that is responsible for the dysfunction of cluster 3.

1.4) V-J rearrangement and RNA splicing:

All the mRNA splice signals of the V, J and C gene fragments conform to the GT/AG rule (Ohshima et al., 1987) except for the defective 3' splice signal of exon 2 of Cγ3 (Hayday et al., 1985). In addition, all the genomic sequences of the 3' end of Vs and 5' end of Js have the conventional DNA rearrangement signals. During rearrangement, the recombinase recognizes the rearrangement signals, the loop between the signals is then deleted and finally the V, J fragments are joined (McBlane et al., 1995). In the mouse TCR γ locus, the V, J and C gene fragments tend to be rearranged in the same cluster whereas interclusteral recombination is infrequent (Pelkonen et al., 1987). Therefore, there are only six V-J recombinations ordinarily used. So far, only two non-functional inter-clusteral rearrangements, the Vγ7-J4 in thymoma cell line BW5147 (Pelkonen et al., 1987) and Vγ1-J2 (Roger et al., 1993), were reported. Vγ7 may join J4 by chromosomal deletion while Vγ1 may join J2 by chromosomal inversion because Vγ1 and J2 lie in opposite orientations in BALB/c.

1.5) V-J Junctional diversity:

Similar to the human TCR γ locus, the low copy number of V, J and C gene fragments in mice and the intraclusteral recombination restriction determines that the combinatorial diversity does not play a major role in generating diversity. However, it was estimated that the potential mouse γδ TCR repertoire far exceeds that of αβ TCRs and Igs (Davis et al., 1988). Such a high diversity is primarily a result of the base additions and deletions at

the V-J conjunctions in the γ genes in addition to the junction diversity of the δ genes (Sim et al., 1991; Richards & Nelson, 2000).

Junctional diversity was previously thought to be a random nucleotide addition and deletion process (Meier et al., 1993). However, the statistical analysis of human TCR γ junctional diversity suggested that a few mechanisms, under different constraints, were responsible for the generation of junctional diversity (Gauss et al., 1996). The first mechanism is the nucleotide addition dependent on TdT (N-addition). Rather than a random process (Davis et al., 1988), the nucleotide addition has G nucleotide preference, and is also dependent on the type of the previous nucleotides. For example, purines tend to be added behind purines, and pyrimidines tend to be added behind pyrimidines. The second mechanism for generating diversity is through adding palindromic nucleotides. This includes P-addition (1-3 base additions exactly complementary to the last few nucleotides of the full length termini to which they are added) and Pr-addition (the same base additions similar to that of P-addition are added at nucleolytically processed coding termini rather than full length termini). Pr-addition requires TdT, while P-addition does not. The third mechanism is nucleotide deletion with exonuclease nucleolytically processing. Sequences with relatively high A+T lose nucleotides easier than sequences displaying high G+C. The last mechanism is a non-TdT nucleotide addition, which occupies only 5% of the encoding junction region (Delves, 1994; Gauss et al., 1996; Weis-Garcia et al., 1997).

In the mouse TCR γ locus, a total of 156 V-J rearrangement transcripts were found in GenBank. These rearrangements were aligned together with the genomic sequence to identify the types of the junctional diversity according to the above mechanisms. The averaged additions and deletions of these 156 rearrangements are 4.7 bp and 3.2 bp, respectively. Among them, there are 41 N-additions, 54 P-additions and 26 Pr-additions in total. Therefore, the junctional diversity of the mouse TCR γ genes is also not randomly generated, but exists under a number of constraints. Table 8 summarizes the base additions and deletions of these V-J rearrangements. Generally, half of the V-J rearrangements of the mouse TCR γ have a relatively high junctional diversity but the

Table 8. Summary of the junctional diversity by additions and deletions of the mouse TCR γ genes.

Type of rearrangement	Number of rearrangements in GenBank	Average insertion (bp)	Average 3' deletion (bp)	Average 5' deletion (bp)	Type of addition (No.)		
					N-addition	P-addition	Pr-addition
V1-J2	1						
V1-J4	34	2.9	3.4	2	11	7	7
V2-J1	1						
V2-J2	60	2.6	5.6	0.9	17	29	12
V4-J1	56	4.1	1.9	0.5	13	18	7
V5-J1	2		4	1			
V6-J1	1		3	0			
Total	169				41	54	26

other half, such as Vγ4-J1, Vγ5-J1 and Vγ6-J1, are limited in junctional diversity. In some tissues, especially epithelia, rearrangements are monomorphic. They equally recognize monomorphic ligands, probably a self-protein that may be present in response to infection. The two distinct forms of junctional diversity further support the notion that the mouse TCR γδ T cells play a linkage role between innate and specific immunity.

2. Structural Features:

One of the important functions of genomic sequences is that they can provide a detailed knowledge of the structural features of the loci. The analysis of the mouse TCR γ locus revealed that there are many characteristic structural features, such as GC content, genomic-wide repeats, and open reading frames (ORF), for example (Figure 14). Among these features, the most prominent is that this locus contains very abundant genome-wide repetitive sequences. A total of 67078 bp of interspersed repeats constitutes 39.91% of the entire 168053 bp sequence. Table 9 summarizes commonly appearing repeats in this mouse locus.

About 40 LINEs occupying 30.93% of the entire sequence were identified belonging to the LINE1 subfamily. This proportion is significantly high in comparison with other TCR loci. For example, there is no LINE in the human TCR Cα/Cδ locus and the LINEs occupy only 13.0% of the sequence in the human TCR β locus (Koop et al., 1994; Rowen et al., 1996). This percentage greatly exceeds the average 10% of LINE1 in the mammalian genome and 15% in the human genome (Furano et al., 1994; Kazazian et al., 1998). A full LINE1 is usually 6-7 kb, containing a 5' untranslated region (four promoter monomers are usually located here), two protein-coding sequences ORF I and ORF II and a 3' untranslated region (Furano et al., 1994). However, the 5' untranslated regions containing the promoters are usually truncated, which results in inactivity of LINEs. In our sequence, five LINEs are longer than 4 kb while most of the remaining LINEs are partial repeats that range from 500 to 800 bp. The five big LINEs were named as L-1, L-2, L-3, L-4 and L-5, whose lengths are 6300 bp, 6094 bp, 6672 bp, 7623 bp and 4670 bp, respectively. These five LINEs all contain two or more ORFs. The transcriptional orientations of L-3, L-4 and L-5 are from 5' to 3', whereas those of L-1 and L-2 are

Table 9. Summary of genome-wide interspersed repeat fragments in the mouse TCR γ locus. These repeats were identified using RepeatMasker.

Total length: 168053 bp GC level: 40.03%

Type		Number	Length	Percentage of sequence
SINEs:		30	3432 bp	2.04%
	B1s	11	1088 bp	0.65%
	B2s	8	1057 bp	0.63%
	B4s	8	1075 bp	0.64%
	IDs	1	54 bp	0.03%
LINEs:	LINE1	40	51992 bp	30.93%
LTR elements:		13	11233 bp	6.68%
	MaLRs	4	1523 bp	0.91%
	Retrov.	7	9669 bp	5.75%
	MER4_group	0	0	0.00%
DNA elements:		2	134 bp	0.08%
	MER1 type	2	34 bp	0.08%
	MER2 type	0	0	0.00%
	Mariners	0	0	0.00%
Unclassified:		2	287 bp	0.17%
Total interspersed repeats:			67078 bp	39.98%

reversed. In addition, the five LINEs are all located at positions with less functional constraint in the new cluster 2 and 3, in particular, the regions between V and J genes, the regions between two clusters and the intervals between enhancer and C genes. Therefore, the insertions of LINEs in the mouse locus do not damage the function of genes or gene rearrangement. LINEs are usually classified into A, F and V types according to their 5' end promoters. Their ability for retrotransposition is primarily dependent on the activities of the promoters (DeBerardinis et al., 1998). Types F and V are ancient extinct without activity but some of type A may still be active in transposition. Figure 16 shows the alignment of the five LINEs with the known sequences of type A (L1-Md2A), type F (L1-MdF15) and subtype Tf (L1-MdTf26). It was found that L-1 belongs to type A because it still possesses two A monomers at the 5' end of the body. L-4 is an F type LINE, which has its two F monomers which are separated from the body by a 2 kb insertion and thus are no longer active. For the remaining three LINEs, L-2, L-3 and L-5, their 5' promoters all have been truncated. Therefore, only L-1 may still be active and be able to retrotranspose. The other four LINEs have lost this ability.

In comparison with LINEs, SINEs only occupy 2.04% of the mouse TCR γ sequence. The mouse genome contains two major families of SINEs, B1 and B2. The B1 family is related to the human Alu, but B2 repeats have no equivalent in the human genome (Cross et al., 1997). It was estimated that there are 80,000 B1, 80,000 B2 and 12,000 ID in the mouse genome (Deininger, 1988). Therefore, it would be expected that there would be four or five B1, four or five B2 and one ID in the 168053 bp mouse sequence. In fact, there are a total of 30 SINEs in this locus including eleven B1, eight B2 and one ID which account for 0.65%, 0.635 and 0.64% of the genomic sequence. These numbers and percentages are close to what is expected (Li, 1997), but are relatively low in comparison with some loci such as the human LIMK1-RFC2 region, which contain 45.4% of Alus (Martindale et al., 2000).

Other important structural features also include the GC content and the ORFs. The average GC level in this locus is 40%, rendering it a GC-poor region (Bernardi, 1995). The level of CpG dinucleotides is low, too. Therefore, the frequency of CpG

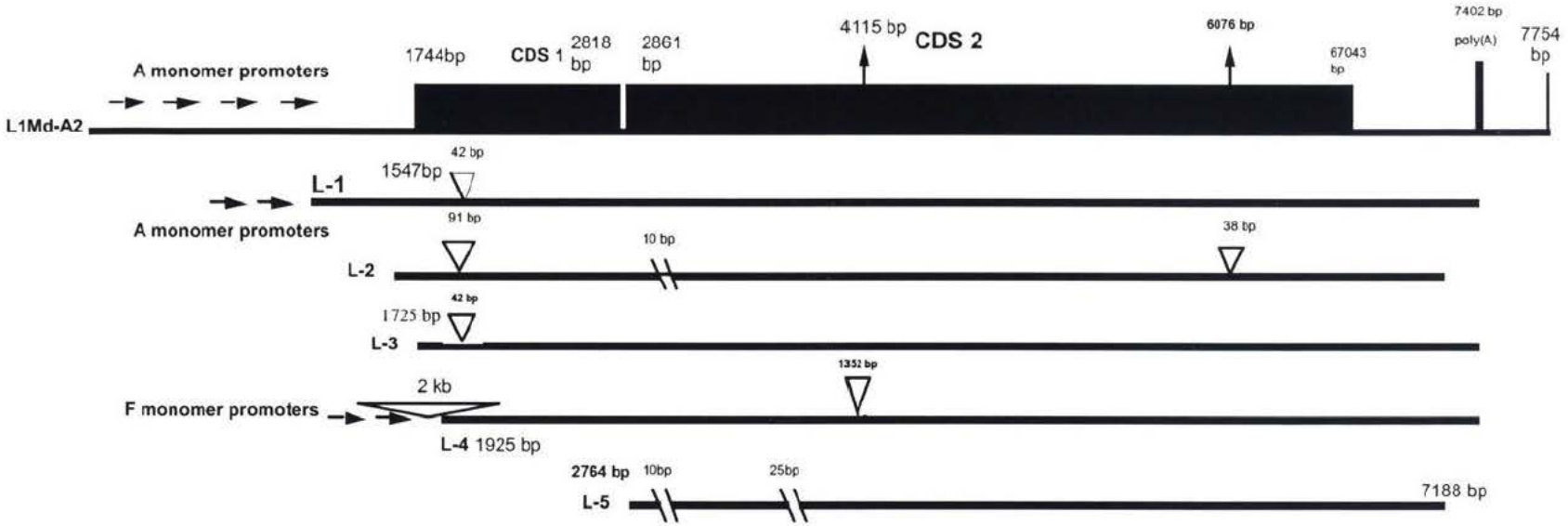


Figure 16. Organizational comparison of the complete mouse LINE L1Md A-2 with the five large LINEs in the mouse TCR γ locus. L-1 exhibits a 192 bp 5' UTR while L-2 exhibits a 540 bp 5' UTR. L-2, L-3, and L-5 do not possess 5' UTRs. L-1 contains two monomers of the A type promoters while L-4 has two monomers of the F type promoters which are separated from the body by a 2 kb insertion. All the five LINEs have 3' UTRs although only L-1, L-3 and L-4 have poly(A) tails. L-1, L-2 and L-3 have a full length of the two ORFs (CDS1 and CDS2), which are represented by the black bars. The \\ indicates gaps and the inverted triangles indicate insertions.

dinucleotides is consistent with the level of GC content. In addition, although it was estimated that there are approximately 37,000 CpG islands in the mouse genome (Antequera et al., 1993), no CpG islands were found in this locus. CpG islands are regions within 1 kb with over 60% CpG dinucleotides (Strachan & Read, 1996). In the human genome, about 56% of genes are related to CpG islands (Antequera et al., 1993). Hence, CpG islands and GC content are usually thought to be useful indicators of the gene-richness of a genome (Bernardi, 1995). Only four genes (a TCR gene refers to a complete combinatorial coding gene that includes one individual V, J and C gene fragments) and the low CpG islands and GC content in this locus appear to support this rule. Further, no ORFs greater than 300 bp were observed to correlate with the known coding regions (Figure 14). A few dense ORF regions were only detected in the five long LINEs, which are the protein-coding sequences ORF I and ORF II in LINEs (Furano et al., 1994).

There is some relationship between the chromosomal distribution of the above structural features and the pattern of Giemsa staining. Usually, Giemsa-dark bands are characterized by lacking GC, fewer genes, fewer SINEs and the richness of LINEs (Chen et al., 1989; Thomsen et al., 1996). On the other hand, Giemsa-light bands possess the opposite features. It is obvious that the above structural features of the mouse TCR γ locus are consistent with those of the Giemsa-dark band and therefore it is quite possible they are located in the Giemsa-dark band.

3. Regulatory elements:

The three known mouse enhancers, γE1, γE2 and γE3, were located downstream of the cognate C gene in their corresponding clusters. γE1 was found first and has been functionally characterized, whereas γE2 and γE3 were identified by hybridization with γE1 as the probe because they show a 96% sequence similarity with γE1 (Spencer et al., 1991; Kappes et al., 1991; Vernooij et al., 1993). It has long been believed that an enhancer is located at the 3' end of Cγ4 (Leiden et al., 1993), but this enhancer was not located by hybridization using the γE1 probe (Vernooij et al., 1993). During our analysis of the genomic sequence, a 340 bp long sequence sharing 65% similarity with γE1,

located 3.5 kb downstream from the stop codon of Cγ4, was revealed by dot-matrix self-comparison and it was termed γE4. Figure 17 shows the alignment of γE4 with the other three mouse TCR γ enhancers, and the human TCR γ enhancer. In the TCR α, TCR β and CD3δ, all the enhancers are located at the 3' end of C genes, suggesting that the 3' end is the evolutionarily conserved location of the T-cell-specific enhancer elements (Winoto & Baltimore, 1989a). Therefore, the putative enhancer γE4 is also in accordance with this rule. Furthermore, at the three corresponding nuclear factor binding regions, namely NFγ2, NFγ3 and NFγ4, γE1 and γE4 share 92%, 88% and 75% sequence similarity, respectively. All these features strongly support γE4 being a putative T-cell-specific enhancer.

γE4 may be different from γE1 functionally because the genes they are associated with, Cγ1 and Cγ4, have evolutionarily diverged for a long time and thus their cognate enhancers may have distinct regulatory features. The functional study of γE1 showed that the minimal enhancer containing NFγ2, NFγ3 and NFγ4 offered full enhancer activity whereas the NFγ1, NFγ5 and NFγ6 were not important (Spencer et al., 1991). A further deletional analysis showed that the NFγ3 was the most important element because a tetramer of NFγ3 manifested almost the same level of activity as the entire minimal enhancer in the majority of the T cells tested, whereas NFγ2 and NFγ4 did not (Hsiang et al., 1993). Further, the fine characterization of γE1 showed that the alteration of the sequence CCACA at the 3' half of the NFγ3 to CTGCA by site-directed mutagenesis almost abolished the activity of γE1. Nevertheless, in the equivalent region of NFγ3 of γE4, the sequence CCGCA appears rather than CCACA. Moreover, in the NFγ3 of γE4, another divergence by three bases occurs in regions that were found to be crucial to the enhancer activity in γE1. The differing of the above bases at critical regions may result in a change or even loss of function of the NFγ3 in γE4. It can be inferred that some other regions may become functionally important in replacing the function of NFγ3 in γE4 and therefore the distributive pattern of the regulatory motifs in γE4 may be changed. A functional analysis of γE4 will be helpful to verify these hypotheses.

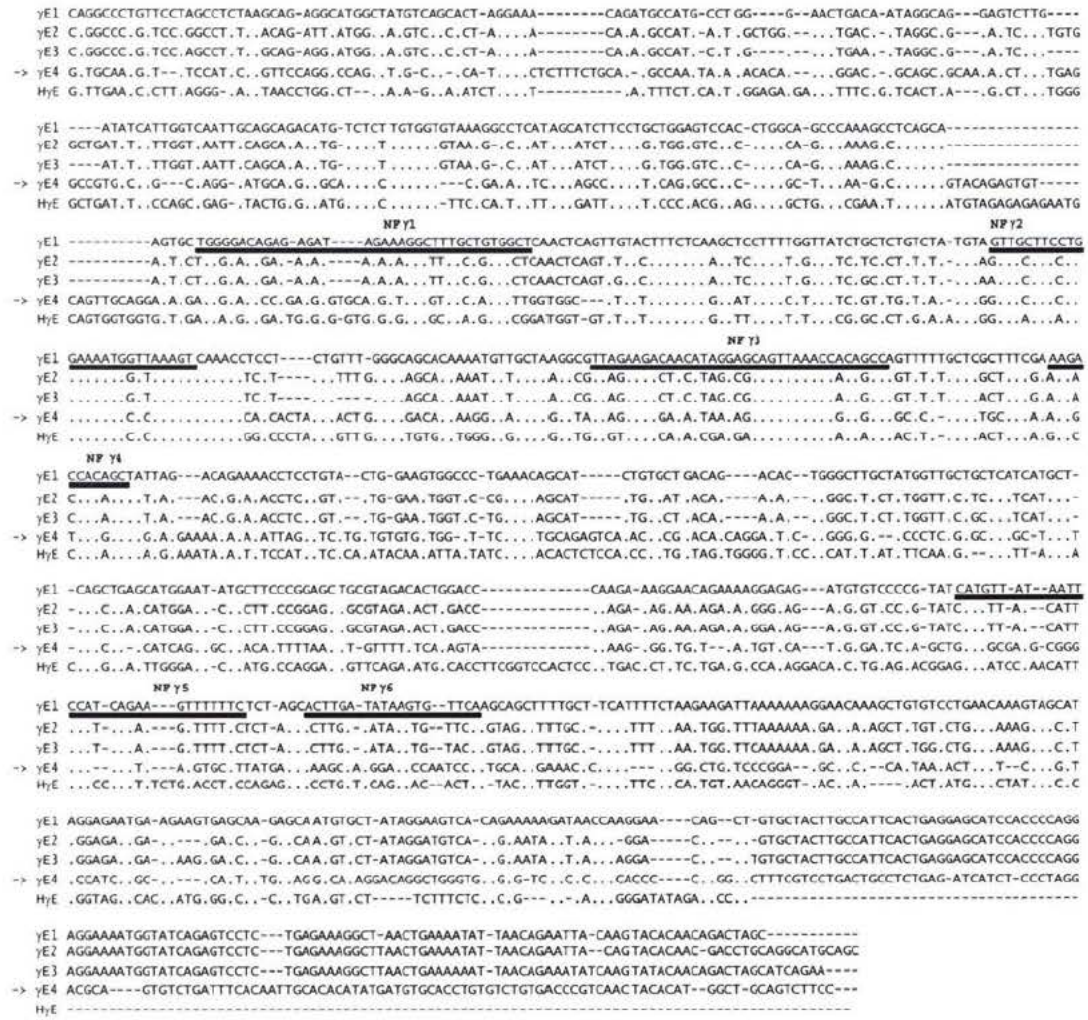


Figure 17. Alignment of the 1kb sequence containing the mouse putative gamma enhancer γE4 (marked by arrows) with the functional γE1, two putative enhancers γE2 and γE3 and the human TCR γ enhancer HγE. Underlined sequences indicate regions protected from DNAse I by nuclear proteins in γE1 (from Spencer et al., 1991). Insertions or deletions are indicated by dashes while conserved sequences are indicated by dots.

In addition, it has been dubious whether boundary elements exist between clusters. Since four enhancers in the same locus all display similar sequences that suggest similar functions, they might regulate the transcription of the genes in other clusters. However, the highly regulated manner in this locus, such as the ordered expression of the V genes, suggests that the above situation is unlikely. Therefore, there must be some insulators located between clusters to prevent the dispersion of the enhancer function. An insulator, BEAD-1, has been identified in the TCR α/δ locus by functional study (Zhong et al., 1997), which is localized to a 2.0-kb region between the TCR δ gene segments and the TCR α gene segments. BEAD-1 can block the ability of the TCR δ enhancer to activate a promoter when located between the two in a construct. Similarly, the confirmation of the existence of such boundary elements in the TCR γ locus will additionally depend on functional studies.

4. Evolution:

4.1) Generation of the mouse TCR γ locus

The mouse TCR γ locus has four V-J-C clusters, which differ from the human TCR γ locus, in which all the V genes are located at the 5' end and the two J-C clusters are located at the 3' end (Lefranc et al., 1986a; Vernooij et al., 1993). The organizational difference suggests that the human and mouse loci have been shaped differently by selection pressure or mechanisms since their separation. Therefore, we further analyzed the sequence of mouse TCR γ locus to identify the possible underlying evolutionary mechanisms. Figure 18 shows the dot-matrix self-comparison of the 168 kb sequence in both orientations. Most notably, a few large internal repeats in this locus exist. By comparison with the feature map, some of the repeats were identified to be genome-wide repeats while other large repeats were locus specific repeats. These locus specific repeats represent the internal evolutionary relationship among the regions where they are located, and therefore are helpful in providing evidence for reconstructing the evolutionary history.

The features of the dot-matrix map were further simplified by aligning the major conserved regions into a proportional diagram (Figure 19). Three conserved regions

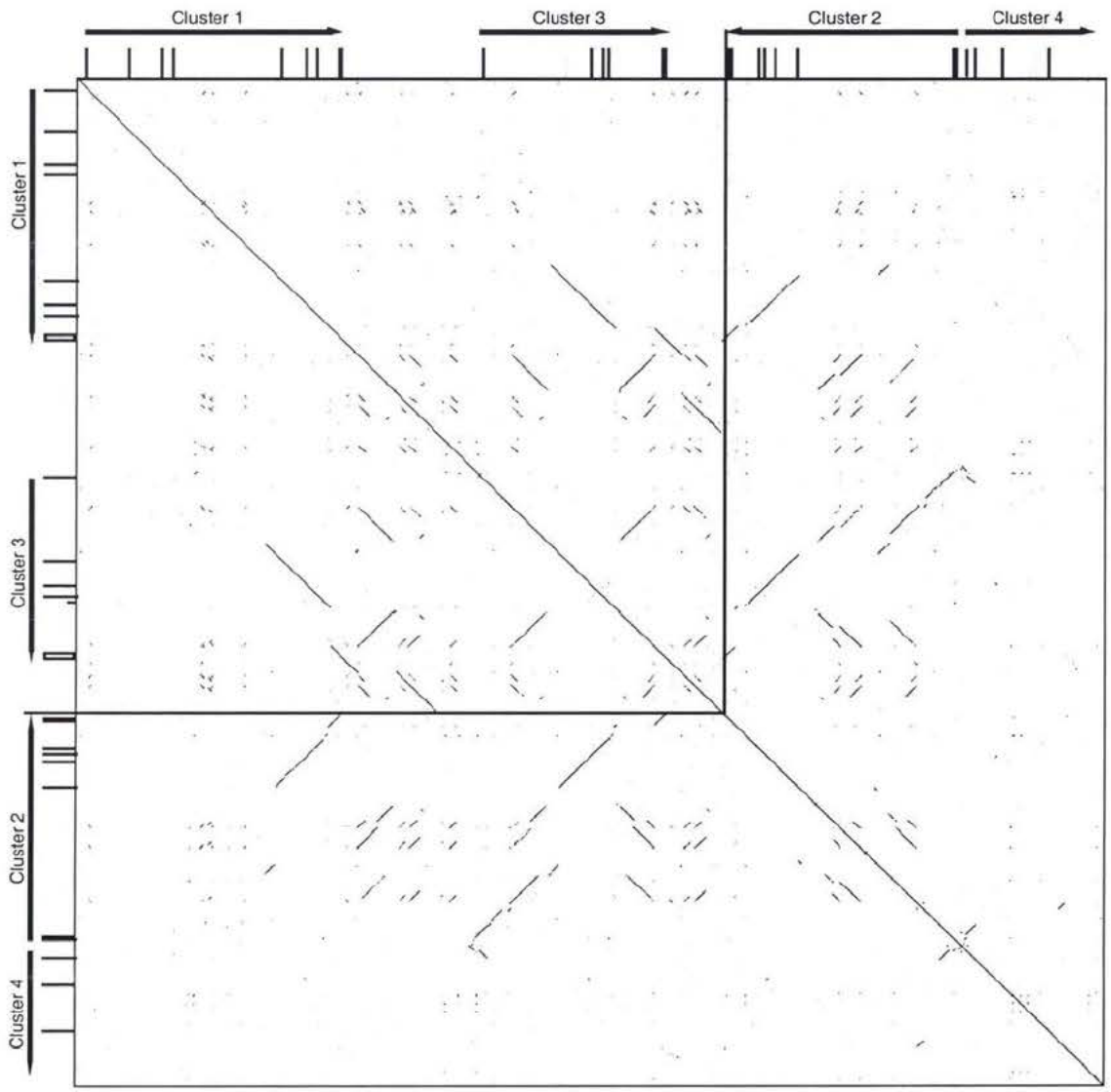


Figure 18. Dot matrix self-comparison of the 168 kb mouse TCR γ locus. Similar regions occurring in tandem form diagonal lines from the top left to the lower right, while similar regions occurring in an inverted orientation appear as diagonal lines from the top right to the lower left. Arrows indicate the direction of transcription. The two lines between cluster 3 and cluster 2 indicate the join between 5' and 3' sequences. An approximately 37 kb sequence between these two clusters was not obtained.

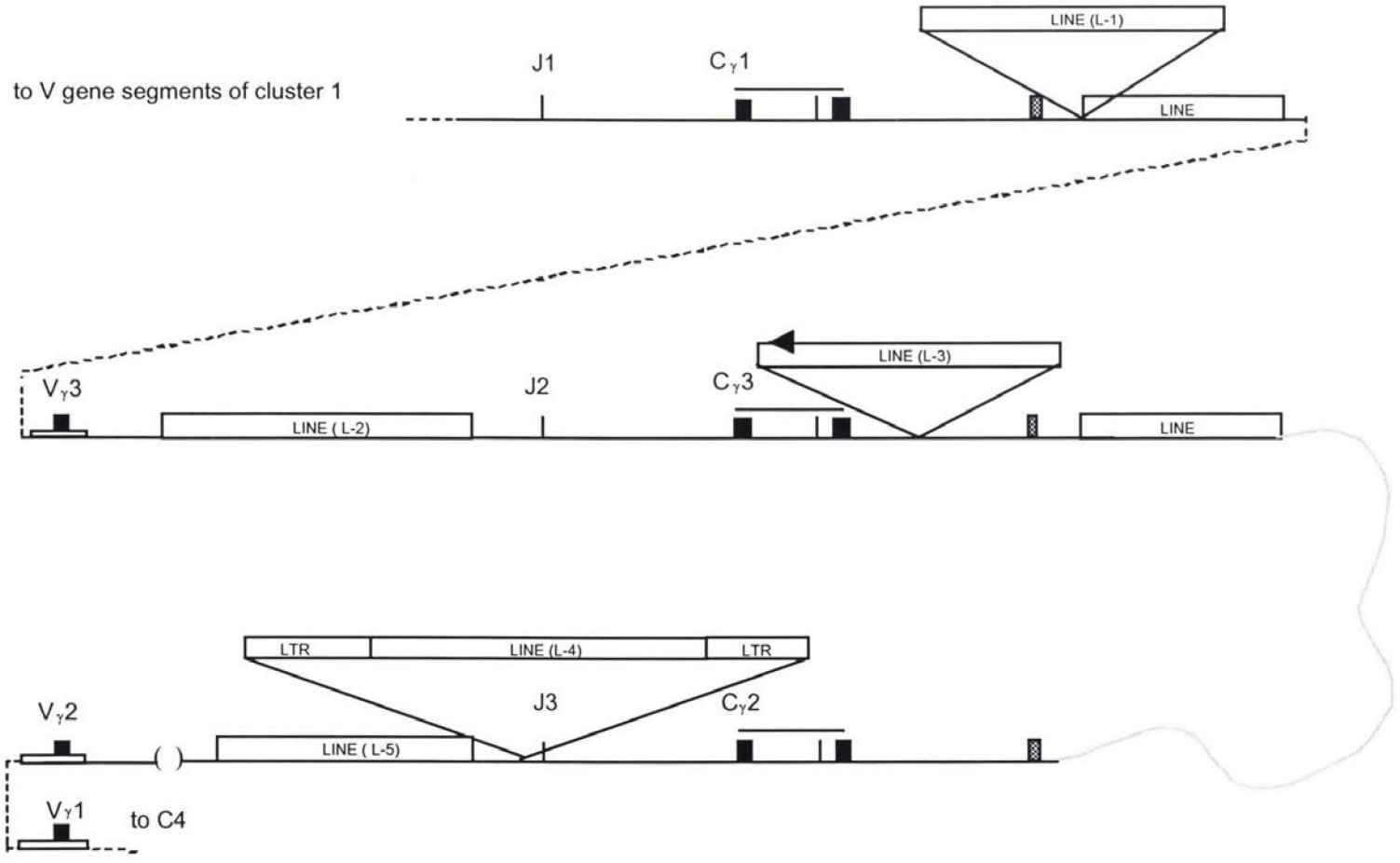
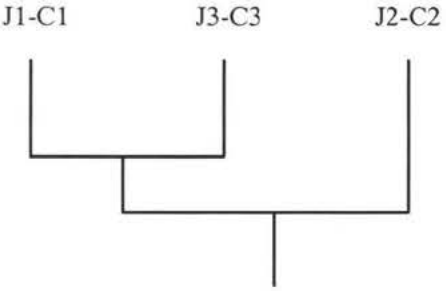


Figure 19. Diagram of the mouse TCR γ locus organization with the cluster 1, 2 and 3 aligned vertically in proportion. The LINE insertions without counterparts in the other clusters are shown on top of the map. The arrow indicates the orientation of the transcription. The parenthesis indicates the deletion at the 5' end of L-5 in comparison with L-2. The four pieces of sequences are connected by dashed lines.

could be identified from this diagram, which are the regions encompassing the whole clusters 2 and 3, the J-C regions in clusters 1, 2 and 3 and the genomic conserved regions of the three V1 gene family members. In the first conserved regions, after being deleted the insertions of genome wide repeats within clusters 2 and 3, these two clusters showed almost identical organizations and an 88% sequence similarity in their 17 kb DNA sequences, which suggests that one of them was duplicated by another. The LINEs in these two clusters, L-2 and L-5, were obviously duplicated along with the clusters' duplication because they are located at the corresponding positions of the two clusters. The 3' end of L-5 is 2 kb shorter than L-2, which is probably due to later deletion. The second large sequence conservation is the 11 kb long J-C regions which extend from 2.5 kb upstream of J gene to 300 bp downstream of the cognate enhancers. As the DNA fragment between clusters 2 and 3 was not available, the size of the similar sequences beyond the enhancers could not be estimated. The distance value between Jγ1-Cγ1 and Jγ3-Cγ3 is 2.3%, which is less than the distance between Jγ1-Cγ1 and Jγ2-Cγ2 (Figure 20A). From the above analyses, it could be inferred that Jγ1-Cγ1 duplicated Jγ3-Cγ3 first while Jγ2-Cγ2 was generated from Jγ3-Cγ3 afterward. As for the third conserved region, namely the coding regions of the Vγ1, Vγ2 and Vγ3 together with their flanking sequences, the calculation showed that the distance between Vγ1 and Vγ3 is closer than that between Vγ1 and Vγ2 (Figure 20B). Further, phylogenetic analysis of the coding sequences of V genes with Vγ4 as the outgroup suggested that Vγ1 was generated earlier than Vγ2 and Vγ3 (Figure 20C). Therefore, it could be deduced that Vγ3 was produced before Vγ2. Taking the above inference from J-C regions and V regions together, one could conclude that cluster 2 was duplicated from cluster 3. Moreover, the analyses of the seven V and C genes have indicated that the V and C genes in clusters 1 and 4 were generated before those in clusters 2 and 3 (Vernooij et al., 1993). Therefore, cluster 1 and cluster 4 were relatively older than clusters 2 and 3. According to the overall analyses, as previously presented, Figure 21 briefly summarizes the evolutionary process: The ancestral locus includes only clusters 1 and 4. Cluster 3 was generated first, and this was followed by its duplication to generate cluster 2. The cluster 2 block then proceeded to reverse its transcriptional orientation. This reversion process was probably mediated by

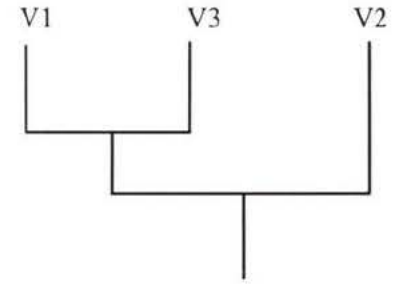
UPGMA trees **Distance matrix**

A.



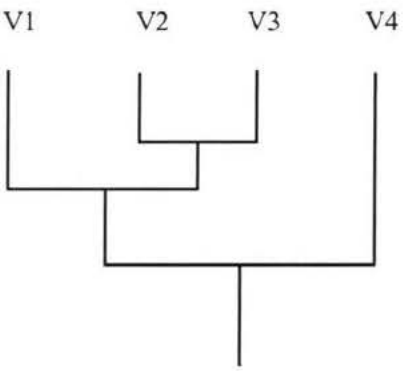
	J1-C1	J2-C2	J3-C3
J1-C1	--		
J2-C2	0.07267	--	
J3-C3	0.04989	0.06261	--

B.



	V1	V2	V3
V1	--		
V2	0.11460	--	
V3	0.07536	0.08997	--

C.



	V1	V2	V3	V4
V1	--			
V2	0.04606	--		
V3	0.04606	0.00000	--	
V4	0.33099	0.34023	0.34023	--

Figure 20. The UPGMA trees (left) and distance matrix (right) of the three conserved regions. (A) J-C regions. (B) Conserved regions of mouse Vγ1, Vγ2and Vγ3. (C) Coding regions of mouse V genes.

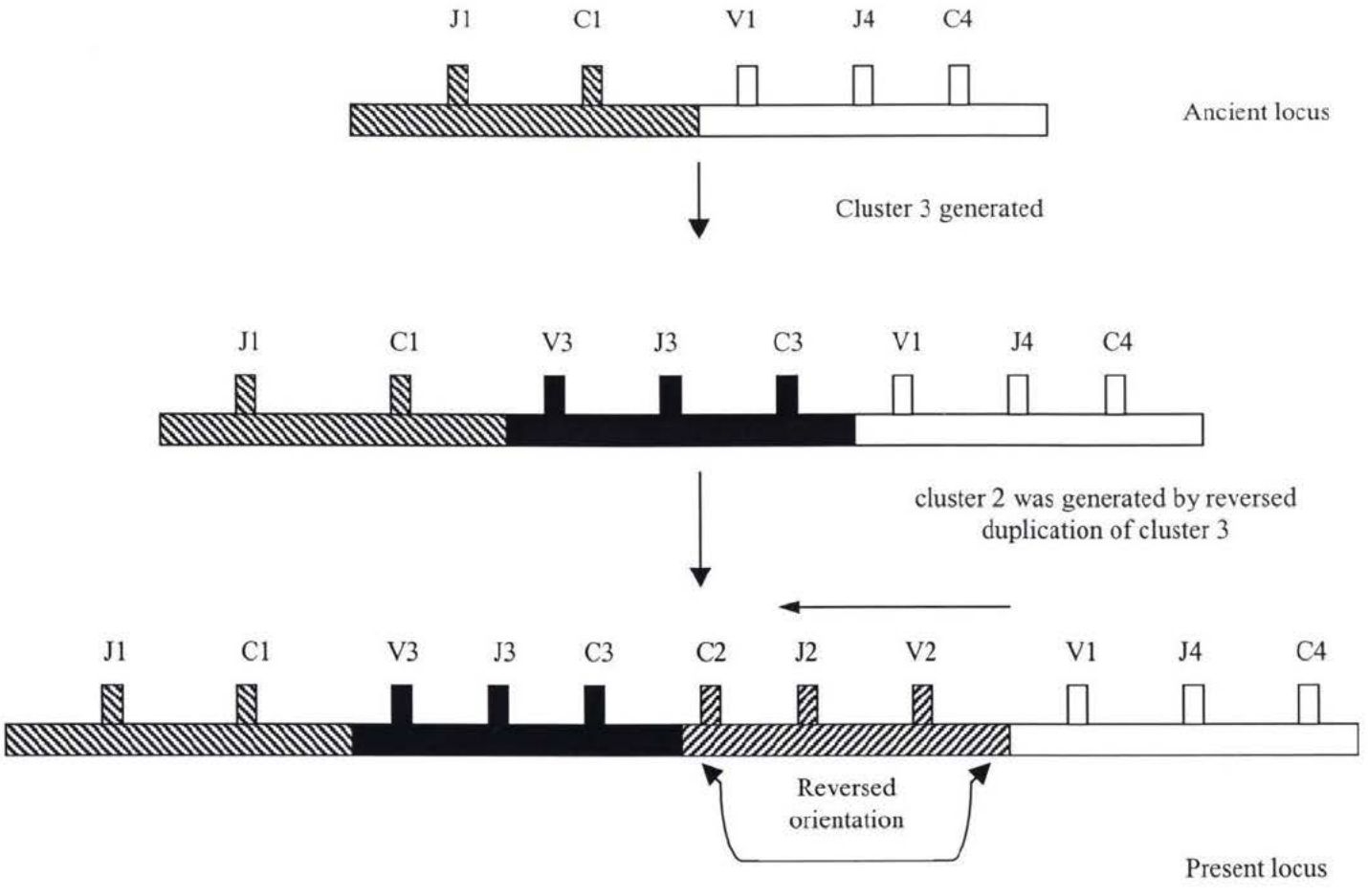


Figure 21. The proposed model of the evolution of the mouse TCR γ locus. The J-C region of cluster 3 was generated by duplicating the J-C region of cluster 1. Cluster 2 was generated by reversed duplication of cluster 3. The arrow represents the transcriptional orientation.

LINEs or LTRs with opposite transcriptional orientation surrounding cluster 2 (Jeffreys et al., 1982). This model is close to the one proposed by Vernooij et al., (1993), but the order of the duplication of clusters 2 and 3 is further clarified here. The duplication of the two clusters is possibly mediated by an unequal crossing-over because it is the most common mechanism to generate large repeats (Li, 1997). Nevertheless, most of the sequences marking these events were lost because they transpired over a long time, and thus the majority of the direct evidence, such as the boundaries of crossing-overs, no longer exists.

In addition, the analysis of the five big LINEs also revealed their internal evolutionary relationship. The five LINEs and the known sequences of L1-Md2A, L1MdTf26 and L1MdF15 were aligned and insertions and deletions were excluded. With L1MdF15 as the outgroup, both the N-J and most parsimonious trees were plotted. They are consistent with one another. Figure 22 shows the phylogenetic trees in the most parsimonious method, in which two clusters formed. The duplication products L-2 and L-5 are close to the outgroup L1MdF15 whereas L-1, L-3 and L-4 are close to L1-Md2A and L1MdTf26. This grouping pattern is consistent with the organizational analysis of the five LINEs. L-1, L-3 and L-4 are evolutionarily newer than L-2 and L-5 because they have no counterparts in cluster 2 or 3. Therefore, L-1, L3, and L-4 were retrotransposed after the duplication of cluster 2 from cluster 3. The phylogenetic tree also suggests the possible order for L-1, L-3 and L-4 appearing in the locus, which is L-3->L-1->L-4. This sequence is also consistent with the structural analysis of the five LINEs, which shows that L-1 and L-4 are relatively recent transposons because their promoter monomers have not been completely truncated. In conclusion, the five LINEs may appear in the locus with the following order: L-2->L-5->L-3->L-1->L-4.

4.2) Evolutionary implications of the characteristic organization of mouse TCR γ locus:

In multigene families, the organizational patterns of the repetitive family members may reflect the evolutionary and functional features of the genomic loci (Graham, 1995). In tandem arrays, the repeats lie from head to tail. However, the clustered and dispersed

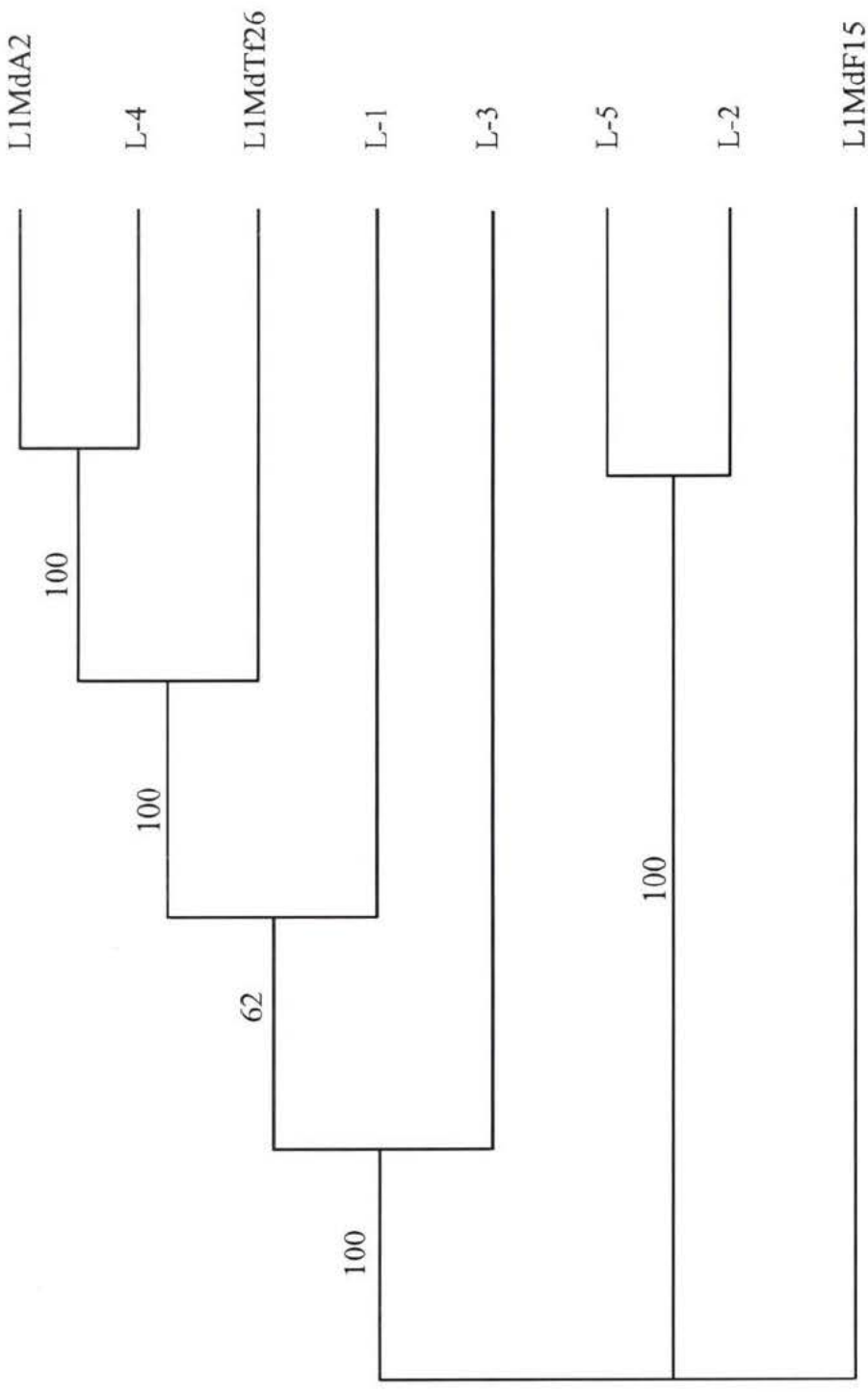


Figure 22. Phylogenetic tree of the five LINEs in TCR γ locus and the mouse LINEs L1MdF15, L1MdTf26 and L1MdA2. The tree was constructed using maximal parsimony with a bootstrap of 1000 replicates. L1MdF15 is used as the outgroup. The numbers above the branches of the tree are the bootstrap values.

genes are linked irregularly and there is no similarity in the spaces between the genes. As a result of their differences, tandem arrays expand or shrink rapidly by both unequal crossing-overs and gene conversions to adapt to the alteration of selection pressure, while the clustered genes change organization only by gene conversions. In addition, the transcriptional orientation of a gene in a clustered pattern may be reversed when the selection pressure is low, so that unequal crossing-overs are inhibited because unequal recombination appears to require long stretches of perfect identity between the interacting sequences (Metzenberg et al., 1991). Therefore, clustered genes are more durable and more suitable for maintaining genetic memory (Graham, 1995).

In the mouse TCR γ locus, the V genes in cluster 1 obviously show a clustered pattern. The sequence similarity between the four genes is only 35-46%. They are separated by a different length of intergenic sequence in which there is no similarity between them. In our analysis, no crossing-over events were identified among these genes. These organizational features are in contrast to the tandem repetitive pattern of the human V1 gene family, in which the nine 5 kb genomic repeats lie head to head and show a 85-97% similarity. On one hand, these mouse V genes have been evolving to acquire novel functions, reflected in their specific tissue distribution, different antigen recognizing manner, and specific immune regulatory and protective functions. On the other hand, the divergent V genes in cluster 1 also assist in maintaining the genetic memory. Phylogenetic analysis suggested that the mouse Vγ7 and the human GV1 family are homologous. Nevertheless, the V genes in GV1 expanded whereas Vγ7 did not. It has been suggested that the majority of genes involved in T-cell-mediated function evolved closely related to their evolutionary pressure (Ciccarese et al., 1997). Therefore, the V genes in cluster 1 of the mouse TCR γ locus probably retain the old genetic memory due to disadvantageous selection pressure whereas the human locus increase gene copy number via duplication due to the relatively favorable selection pressure acting on it after the two species separated 85 million years ago. A similar case involved the evolution of the antifreeze protein (AFP) genes of northern ocean flounders, in which the correlation between AFP gene organization and exposure to ice crystal was excellent (Scott et al., 1988). The winter flounder that was exposed to ice exhibited 20 tandem and 10 clustered

genes. But the yellowtail flounder that was not exposed to ice revealed only 10 clustered genes without the additional 20 tandem AFP genes (Hayes & Davies, 1991). Since glaciation was far more extensive in the recent past than it is today, selection of both species of flounder is probably relaxing in stringency. Clustered genes are more durable than tandem genes in the absence of selection, and this has preserved the clustered AFP genes in the yellowtail flounder. Similarly, the number of TCR Vγ genes in the human and mouse have probably expanded or contracted depending on the need to protect the host from ever-changing groups of parasites.

Another organizational feature of the mouse TCR γ locus is the reversed transcriptional orientation of cluster 2. As large-scale sequence similarities exist between clusters 1, 2 and 3, which can result in a deletion of these clusters by recombination, cluster 2 may reverse its transcriptional orientation to avoid potential deletion. Thus, the reversed transcriptional orientation may also be favorable in maintaining genetic memory due to the relaxed selection forces. Actually, a few organizational features of this locus are obviously supportive to this idea. Firstly, the loss of cluster 3 in some mouse strains was possibly deleted by an unequal recombination process, which directly supports the above hypothesis. Secondly, there are also a large number of LINEs insertions which can greatly increase the opportunity of deletion by unequal crossing-over (Rowen, et al., 1996). Such potential disadvantageous effects of recombination to the functions suggests that natural selection or functional constraint on this locus became lower, for example, causing the Vγ3 to become a pseudogene. However, although the selection is less, for these or other reasons, many large gene families are not able to degenerate to the minimal number of functional genes needed for survival. Therefore, this reversion of cluster 2 provides another source for keeping the genetic memory during the period of relaxed selection because with a few exceptions, inversion genes are likely to be incompatible with high rates of unequal recombination but tandem arrays are highly sensitive to selection (Welch et al., 1990; del Pino et al., 1992). Thus, when selection is re-imposed, clustered genes are good substrates for re-amplifying to tandem arrays.

Yet, another feature of the mouse locus is that the V, J and C gene fragments constitute four independent V-J-C clusters, which makes almost all the rearrangements limited within a particular cluster. The clustered organization may provide another means of preserving the genetic memory. For example, although the pseudogene ψCγ3 makes the entire rearrangements in cluster 3 non-functional, it does not influence the productive V-J-C rearrangements in other clusters because there are only a few inter-clusteral rearrangements. Therefore, when the selection pressure decreases, which results in some nonfunctional genes, the influenced gene is only limited within that particular cluster.

To conclude, the analysis of the mouse TCR γ locus has provided a large amount of information on gene regulation, genomic structure, and the evolution of the genome. This knowledge has deepened our understanding of the characteristics of multigene families in addition to the unique immune functions of γδ T cells. It is believed that this sequence will further provide the basis for future functional studies of the immune system.

Chapter IV. Characterization of gene regulatory elements in cluster 4 of the mouse TCR gamma locus

Summary:

The expression of TCRs on the cell membranes is highly regulated. For example, the αβ and γδ TCRs show a mutually exclusive expression manner on T cells, and the V genes in the mouse TCR γ locus sequentially appear in the fetal thymus during ontogeny. Transcriptional regulation has been proven to play important roles in the controlling of these processes. The putative enhancer characterized in this project is a 340 bp region identified downstream of the Cγ4 gene in cluster 4 of the mouse TCR γ locus by sequence comparison. A luciferase expression study demonstrated that this DNA fragment is a functional enhancer. Similar to a typical enhancer, it functions in both orientations. The 651 bp maximum enhancer encompasses the entire six DNase footprints NFγs found in enhancer γE1 and shows around 32 times of relative luciferase activity in the γδ T cell line PEER. The minimal enhancer was confined to a 294 bp region containing the equivalent NFγ2, NFγ3 and NFγ4. These elements show diverse patterns on the regulation of luciferase expression in different tissues. The maximum enhancer exhibits the strongest activity in γδ T cells PEER and significant activity in γδ T cell Molt-13, αβ T cell EL-4 but not non-lymphocyte HeLa cells, which suggests it is a T cell specific enhancer. In addition, at the 5' end of the enhancer, a 900 bp putative silencer having a negative effect on transcription was localized. It inhibits the maximum enhancer by a factor of four. Interestingly, the combination of the enhancer and silencer shows different effects on transcription in different cell lines. In Molt-13, the combination shows an increased activity in comparison with the enhancer. However, the combination exhibits decreased activity relative to the enhancer in PEER. Since both Molt-13 and PEER are human γδ T cell lines expressing different Vγ genes, these reversed patterns of activity suggested that the combination of multiple regulatory elements such as enhancer, silencer and other critical *cis*-acting elements in the locus determine the regulation of the specific expression of TCRs, such as lineage selection, and the ordered expression of TCR Vγ genes in the fetal thymus, for example.

Introduction:

Although both B cells and T cells originate from the same stem cells in bone marrow, each cell type expresses only one specific type of V(D)J recombinatory Ig or TCR molecule. However, the V(D)J rearrangements of Ig and TCR gene fragments use precisely the same lymphocytic specific recombinase machinery, (recombination-activating enzymes RAG-1 and RAG-2 proteins; Willerford et al., 1996). Determination of the specific lineage V(D)J recombination is controlled by factors such as chromatin conformation, DNA methylation and transcriptional regulation (Wilson et al., 2002). The regulation of the chromatin accessibility of individual immunoglobulin or TCR gene segments to the recombinase by the *cis*-regulatory elements, such as enhancers, promoters and silencers, appears to play the major role. Although a precise biochemical definition of accessibility is still not clear, there is some understanding of the *cis*-acting elements within the TCR and Ig loci that serve as developmental regulators of chromatin accessibility. Specifically, both *in vivo* and *in vitro* studies, involving the manipulation of V(D)J recombination reporter substrates or endogenous TCR and Ig loci have established that transcriptional enhancers and promoters are critical to the determination of cell fate (Sleckman et al., 1996).

As multigene families, the TCR γ loci contain rich *cis*-acting elements that are essential in regulating both rearrangements and the temporal and spatial expression of gene segments within the locus. Such elements include the promoters upstream of each of the Vγ gene segments, the TCR enhancers (Clausell & Tucker, 1994; Ofir et al., 1995), silencers (Ishida et al., 1990; Clausell & Tucker, 1994), and LCRs (Baker et al., 1999). To date, several elements have been implicated as important developmental regulators of V(D)J recombination *in vivo*. For example, in knockout mice with LCR and enhancer γE1 deleted in the mouse cluster 1, the rearrangement of TCR γ genes was modestly reduced (Xiong et al., 2002). In another example, two constructs containing overlapped DNA fragments (40 and 15 kb respectively) in the cluster 1 of the mouse TCR γ locus were used to generate transgenic mice. There is no transcription of TCR γ genes in the αβ T cells in mice containing the longer fragment. The transgenic mice containing the 15 kb fragment, however, produced transcribed γ genes in αβ T cells (Ishida et al., 1990). This

suggests that some negative regulatory elements located on the 40 kb fragments inhibit the transcription of TCR γ genes and thus are essential for the specific lineage selection of γδ TCRs (Takagaki et al., 1989; Hettmann et al., 1995). In addition, the expression of the V gene during T cell ontogeny parallels the gene order in cluster 1 in the mouse TCR γ locus (Raulet et al., 1991). This ordered rearrangement is obviously controlled by some intracellular regulatory mechanisms rather than extracellular selection because temporal expressions were also observed in TCR δ null mice, in which γδ TCRs were not expressed on the cell surface (Itohara et al., 1993). All these results suggest that the numerous rearrangement and expression features of γδ T cells are controlled by the regulatory elements (Sunaga et al., 1997).

We have sequenced 168 kb mouse TCR γ locus. During the sequence analysis, a 340 bp sequence γE4 sharing a 65% similarity with the mouse enhancer γE1, was located 3.5 kb 3' to Cγ4. It was found that the 3' end of C genes is the evolutionarily conserved location for most of the T-cell-specific enhancer elements such as the mouse TCR β, TCR α and the CD3δ enhancers (Winoto & Baltimore, 1989b). Further, γE4 exhibits a very high sequence similarity with γE1 at the six nuclear factor binding regions revealed by DNase I footprinting (Spencer et al., 1991). For example, the sequence of γE4 is conserved by 92%, 88% and 75% at the equivalent 26 bp NFγ2, 34 bp NFγ3 and 11 bp NFγ4 regions of γE1, respectively, which further strongly suggests that γE4 is a functional enhancer. In the present study, the 28 kb sequence covering the entire cluster 4 was analyzed for putative regulatory elements and the tissue specificity of some potential regulatory elements was also tested in different cell lines.

Materials and methods:

1) Construct preparation:

The inserts for generating the expression constructs were obtained from the mouse TCR γ cosmid Mγ84 by polymerase chain reaction (PCR) (Figure 23). In both the forward and reverse primers, a *BamH* I site is included. The PCR products were digested by *BamH* I after amplification and purified in agarose gel. The inserts were then

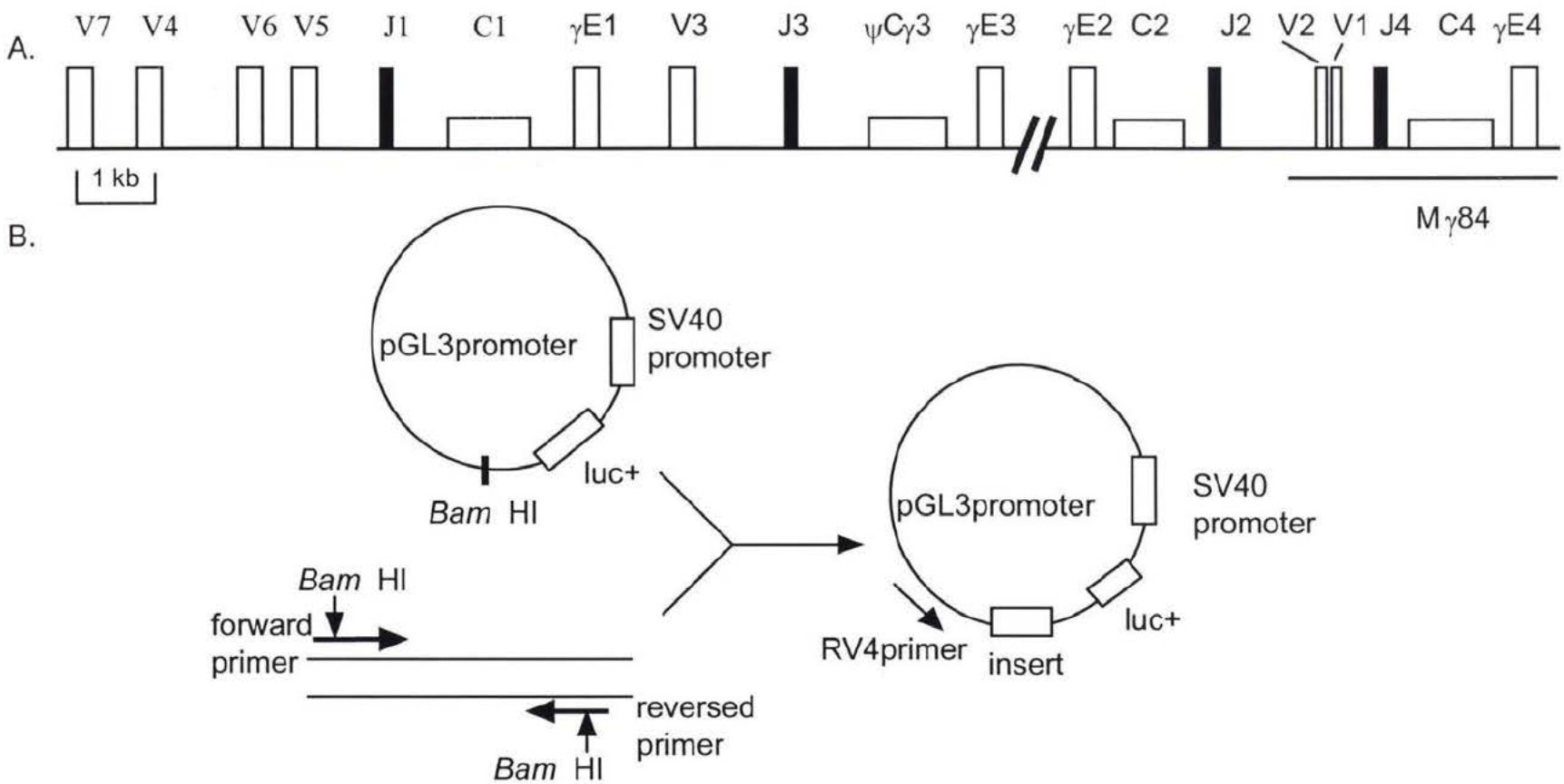


Figure 23. (A) Map of the mouse TCR γ locus. The putative enhancer γE4 is at the 3' end of Cγ4 in the cosmid Mγ84 in cluster 4. (B) Map of the luciferase expression vector pGLpromoter. The putative regulatory elements containing *Bam*HI sites at both ends are inserted at the *BamH* I site downstream of the luciferase gene.

subcloned into the downstream *BamH* I site of the luciferase gene in the expression vector pGL3-promoter (Promega) as indicated in Figure 23. The ligated constructs were transformed into E.*coli* strains DH5αFIQ or Topp9. From the pool of colonies, those containing constructs were screened by plasmid mini-preparation, followed by *BamH* I treatment to confirm the inserts. In order to identify the orientations of inserts, a reversed primer RV4 downstream from the *BamH* I site and a forward primer of the insert were used to check the existence of PCR products (Figure 24). For each insert, one construct with a forward and another one with a reversed insert were chosen to grow in the LB medium while plasmid DNA was prepared by Qiaquick (Qiagene). The sequences of the inserts were further confirmed by sequencing with an automatic sequencer ABI 373 or 377 (Perkin-Elmer Corp.).

2) Cell lines:

The cell lines being tested include the human γδ T cell lines PEER and Molt-13, mouse αβ T cell EL-4 and non-lymphocyte HeLa. In addition, human αβ T cell Jurkat and the mouse B cell A-20 were also used in the functional characterization. Nevertheless, their growth was not optimal, and therefore the results were not used for analysis. PEER, Molt-13, Jurkat and A-20 were grown in RPMI 1640 supplemented with 10% fetal bovine serum (FBS), while EL-4 and Hela were grown in a DMEM medium with a 10% horse serum. A total number of $1x10^6$ cells were inoculated one day before transfection.

3) Transient transfection:

The Effectene transfection kit from Qiagen was employed to transfect the construct DNA and β-galactosidase DNA which are used as an internal control of the transfection efficiency into the above cell lines in the following steps. First, 0.75 ug of construct DNA together with 0.25 ug of β-galactosidase DNA was dissolved in 10 ul of ddH2O. Then, 60 ul of DNA condensing buffer EC were combined into the DNA. This is followed by adding 1.6 ul of Enhancer Solution. After a 4 minute incubation period at RT, 5 ul of Effectene were pipetted into the mixture. The DNA and Effectene mixture was maintained at RT. Ten minutes later, 350 ul of RPMI1640 was added to the above DNA mixture before the whole solution was combined with the cells. For each cell line, the

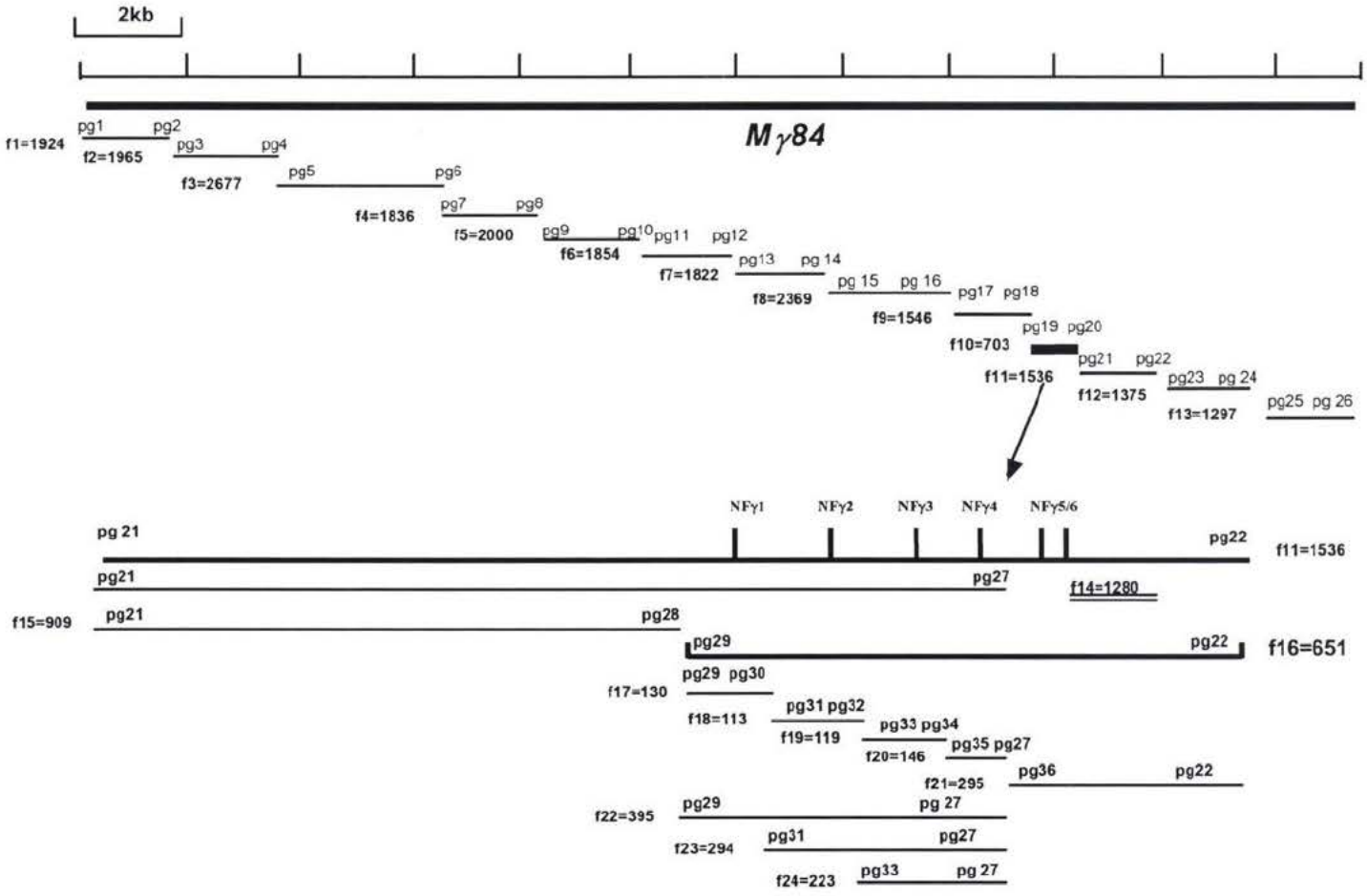


Figure 24. Map of the DNA fragments covering the whole mouse cosmid Mγ84 in the identification of regulatory elements. Primers are indicated by pgs for amplifying the inserts while NFγs are the nuclear binding regions identified in the mouse TCR γ enhancer γE1.

cultured cells were aliquoted in order that each transfection contained approximately $7x10^5$. The cells were washed twice with 250 ul of PBS before being pelleted by centrifugation and dissolved with 350 ul of RPMI 1640. The above DNA solution was added into the cells and then the mixture was transferred into 24 well plates. After the cultures were grown in 5% CO_2 at 37°C in a CO_2 incubator for 24 hours, 1 ml of PRMI 1640 was added into each well. The total incubation time was between 42 and 48 hours.

4) Luciferase assay:

The cultured cells from the previous step were collected into eppendorf tubes. The cells were spun down at 2000 rpm for 5 minutes and the supernatant was pipetted off. 130 ul of 1xReporter Lysis Buffer (Promega) was added into each tube and the cells were incubated for 15 minutes. After being confirmed under the microscope that most of the cells had lysed, the cells were spun down again at 12000 rpm for 15 seconds. Then the lysed cells went through two cycles of freeze-thaw at -70°C. The supernatant was divided into two parts, in which 30 ul was saved for the β-galactosidase assay and another 100 ul was used for the luciferase assay. To test the luciferase activity, 100 ul of luciferase reagent was added into the cell protein extract. The assay was immediately assessed in the scintillation counter. Transfections were repeated several times in separate experiments using different plasmid DNA preparations and at least three tests showing similar results were used for analysis.

5) β-galactosidase assay:

70 ul of β-galactosidase assay buffer (Promega) and 30 ul of cell protein extract were incubated together at 37°C for 30 minutes. The samples were then tested for OD420nm value in a spectrophotometer. If a large variation of OD420 value occurred in the sample, the luciferase activity would be corrected by using the formula: (ODpGL3-promoter / ODsample) x Luciferase activity of sample.

Results:

1) Obtaining colonies containing constructs:

Most of the inserts are less than 2 kb because inserts larger than that are difficult to ligate into the vector. The E.*coli* cell line DH5αIQF' works effectively for subcloning inserts less than 2 kb. For inserts larger than 2 kb, a better compatibility and efficiency were shown in Topp9. A total of 48 colonies containing constructs with both orientations of each of the 24 inserts were selected.

2) Optimization of the conditions of transient transfection:

In order to determine the optimal conditions for transient transfection, different combinations of amounts of DNA, Enhancer solution, Effectene reagent and number of cells were used to test the luciferase activity of f16 that were found to be functional in the preliminary experiments. As shown in Table 10, the set of conditions in column 7 displayed the highest activity, which is 180.12 times relative to the blank control. This set of conditions was employed throughout the entire transient transfection that would form the subsequent step.

3) Screening for the potential regulatory elements in cluster 4:

Thirteen constructs covering the whole cosmid Mγ84 were used to identify the putative regulatory elements in the γδ T cell PEER. The relative luciferase activities indicated that only construct f11 displayed significant relative luciferase activity (14.3 times background). The activities of the remaining constructs are within the normal range in comparison with the control. Figure 25 shows the results of the screening of the putative elements.

4) Deletional analysis of the positive element f11:

The positive regulatory fragment f11 was further characterized by nest deletional analysis. A total of eleven constructs covering different regions of f11 were used to determine luciferase activity in γδ T cell PEER. The averages of the three assays of each construct revealed consistency and were summarized in Figure 26. In the two large fragments covering the entire f11, the relative activity of f15 was only 0.42 while the nearby fragment f16 exhibited the highest activity at 32. However, f11 and f14, the two fragments longer than f16, have only 8.57 and 4.19 times relative activities, respectively.

Table 10. Optimization of the conditions of luciferase activity of fragment f16 in the human γδ T cell line PEER.

	DNA (ul)	Enhancer (ul)	Final vol. Of DNA in EC (ul)	Vol. of Effectene reagent (ul)	Vol. of medium to add to complexes (ul)	Cells seeded	β-galactosidase added	Luciferase activity (cmp)	Relative ratio
1	0.2	1.6	60	5	350	$7x10^5$	10:1 (20ng)	690208	28.9
2	0.2	1.6	60	5	350	$7x10^5$	5:1 (40ng)	786656	32.9
3	0.2	1.6	60	5	350	$7x10^5$	2:1 (100ng)	1085536	45.47
4	0.2	1.6	60	5	350	$7x10^5$	1:1 (200ng)	883072	37
5	0.4	1.6	60	5	350	$7x10^5$	2.5:1 (80ng)	937504	39.27
6	0.8	1.6	60	5	350	$7x10^5$	2.5:1 (80ng)	2484640	104.08
7	1.0	1.6	60	5	350	$7x10^5$	2.5:1 (80ng)	4299968	180.12
8	0.4	1.6	60	10	350	$7x10^5$	2.5:1 (80ng)	977824	40.96
9	0.4	1.6	60	20	350	$7x10^5$	2.5:1 (80ng)	1692352	70.89
10	0.4	1.6	60	10	350	$7x10^5$	2.5:1 (80ng)	1067200	40.7
11	0.4	3.2	60	10	350	$1.4x10^6$	2.5:1 (80ng)	1170400	49
12	Blank reading							17632	1

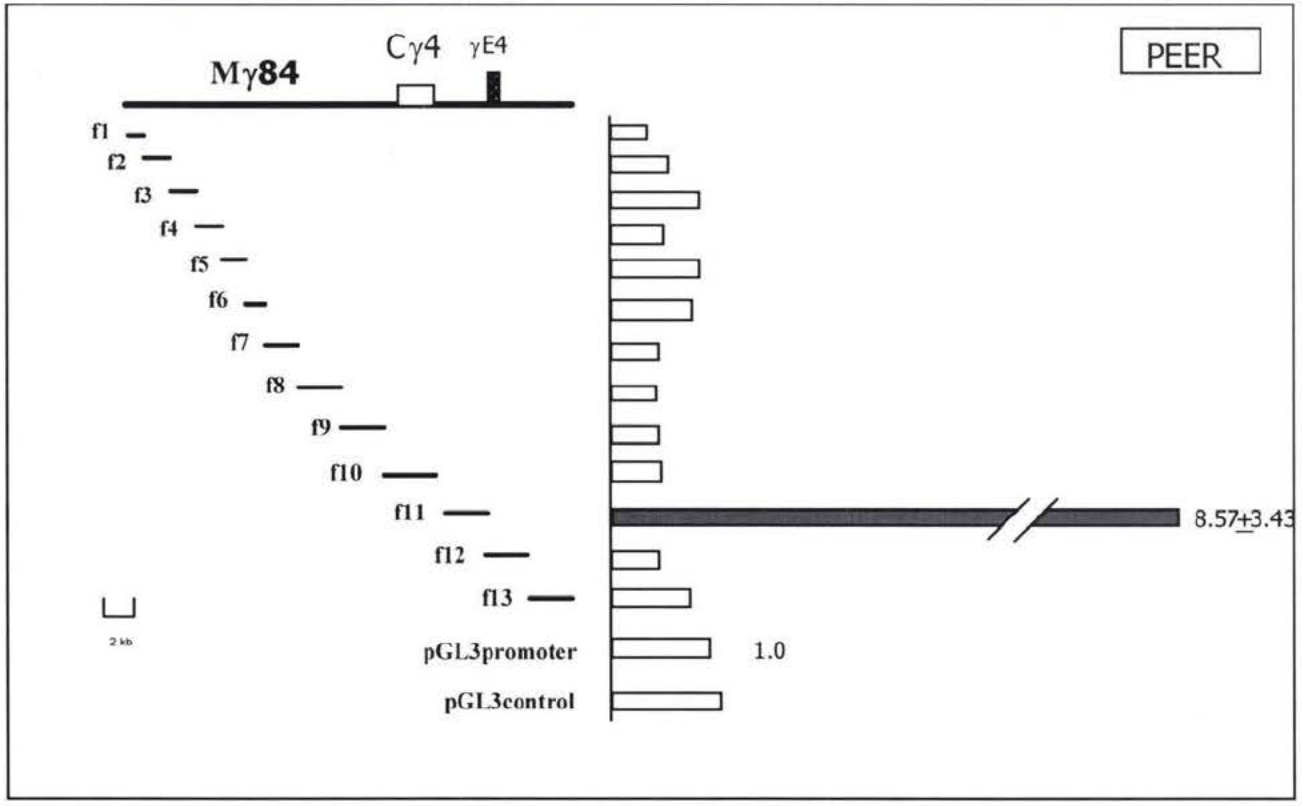


Figure 25. Identification of putative regulatory elements in the mouse TCR γ locus. The constructs covering the whole cosmid Mγ 84 were tested in the human γδ T cell PEER by luciferase assay. The relative luciferase activities are indicated on the right side of the vertical line.

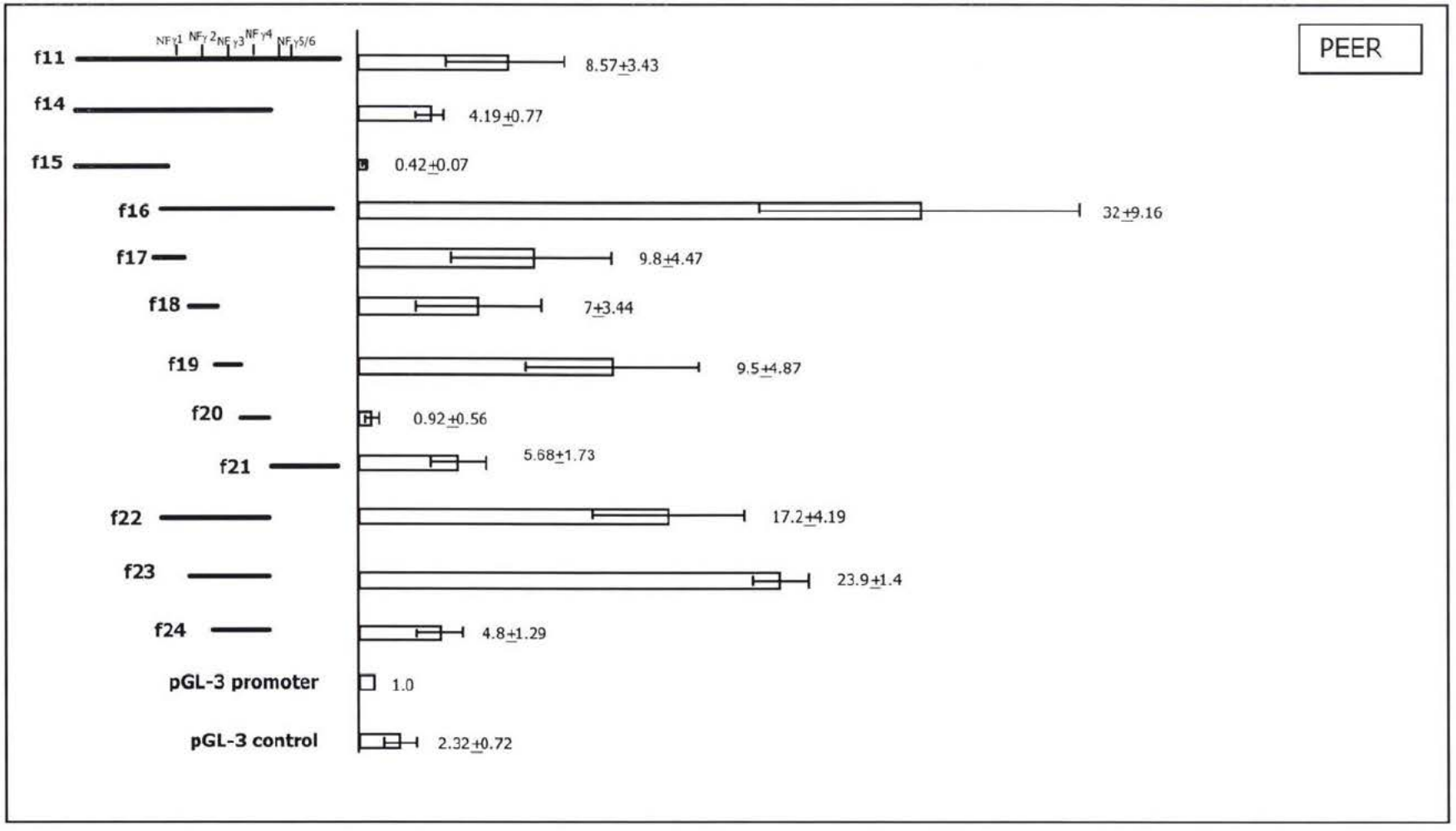


Figure 26. Deletional analysis of the positive regulatory element f11. Fragments are shown as bars at the left side with their corresponding names at the right side. NFγ1-6 represents the corresponding conserved regions with γE1. The relative luciferase activities were tested in γδ T cell PEER and are indicated at the right side of the fragments.

Within f16, the five fragments from f17 to f21 show different levels of activities. When added, the sum of the activities of the five fragments is 32.9, which is close to 32, the value of f16.

5) Tissue specific expression:

Figures 27A, 27B and 27C show the results of relative luciferase activities of some fragments in another three cell lines other than PEER. In the γδ T cell line Molt-13, the activity of f11 is 2.26 and that of f16 is only 0.86. f15 still has a low 0.48 times of activity. The remaining fragments did not show significant activity in this cell line. In the αβ T cell line EL-4, the relative enhancing activities of f11, f16 and f23 are 1.37, 1.87 and 2.29 respectively. The activity of f15 is still only 0.41. In the case of the non-lymphocyte Hela, most of the fragments including f15 have activities close to that of the control except for f23, which has approximately three times the relative activity.

6) Testing the orientation dependence

The luciferase activities of f11, f23 and f15 were also tested in both orientations in PEER (Figure 28). f11 and f23 show activity in both orientations although the activities of the reversed fragments are relatively lower. However, the expressions of luciferase are inhibited in both the forward and reversed constructs of f15.

Discussion:

In the TCR α, TCR β and CD3δ, all the enhancers are located 3' to C genes, which is the evolutionarily conserved location of the T-cell-specific enhancer elements (Winoto & Baltimore, 1989a). Therefore, it has been suspected that there is a similar enhancer at the 3' end of Cγ4 (Vernooij et al., 1993). However, the search for such an enhancer by hybridization with the probe specific for mouse γE1 was not successful due to the low sequence similarity (Vernooij et al., 1993). In our genomic sequence comparison, a region was revealed with 56% similarity to γE1 at the 3' conserved location. This element has been demonstrated to be a functional enhancer as only f11 which contains the conserved sequence showed 8.57 times relative activity among all thirteen constructs covering cluster 4 (Figure 24). Similar to most enhancers, this element regulates

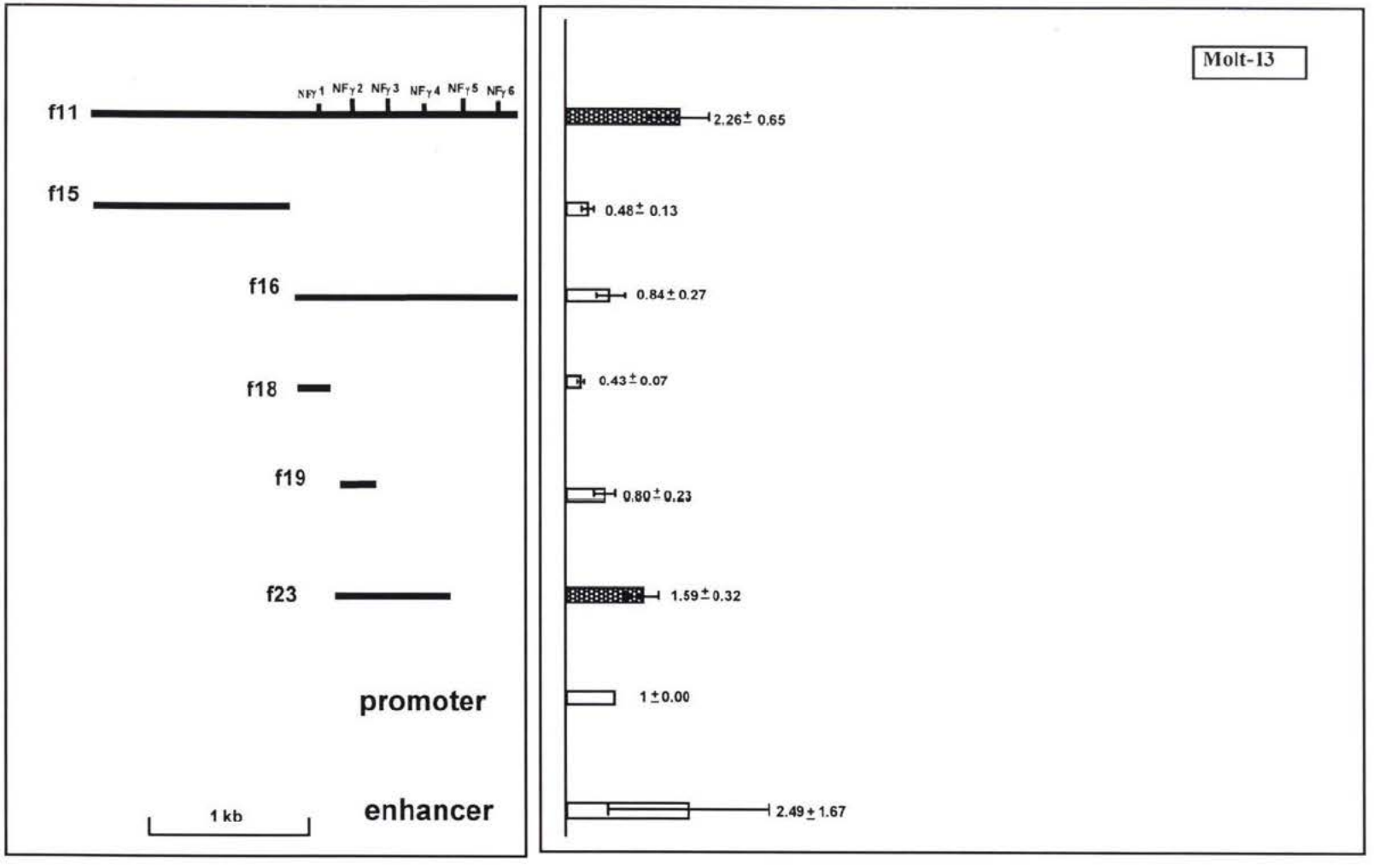


Figure 27A. Tissue specificity of f11 (and the DNA fragments contained in it) in the γδ T cell line Molt-13. The values of relative luciferase activities of these fragments are indicated as bars while their standard errors are shown on the right side accordingly. The luciferase activities around 2 are indicated by shaded bars.

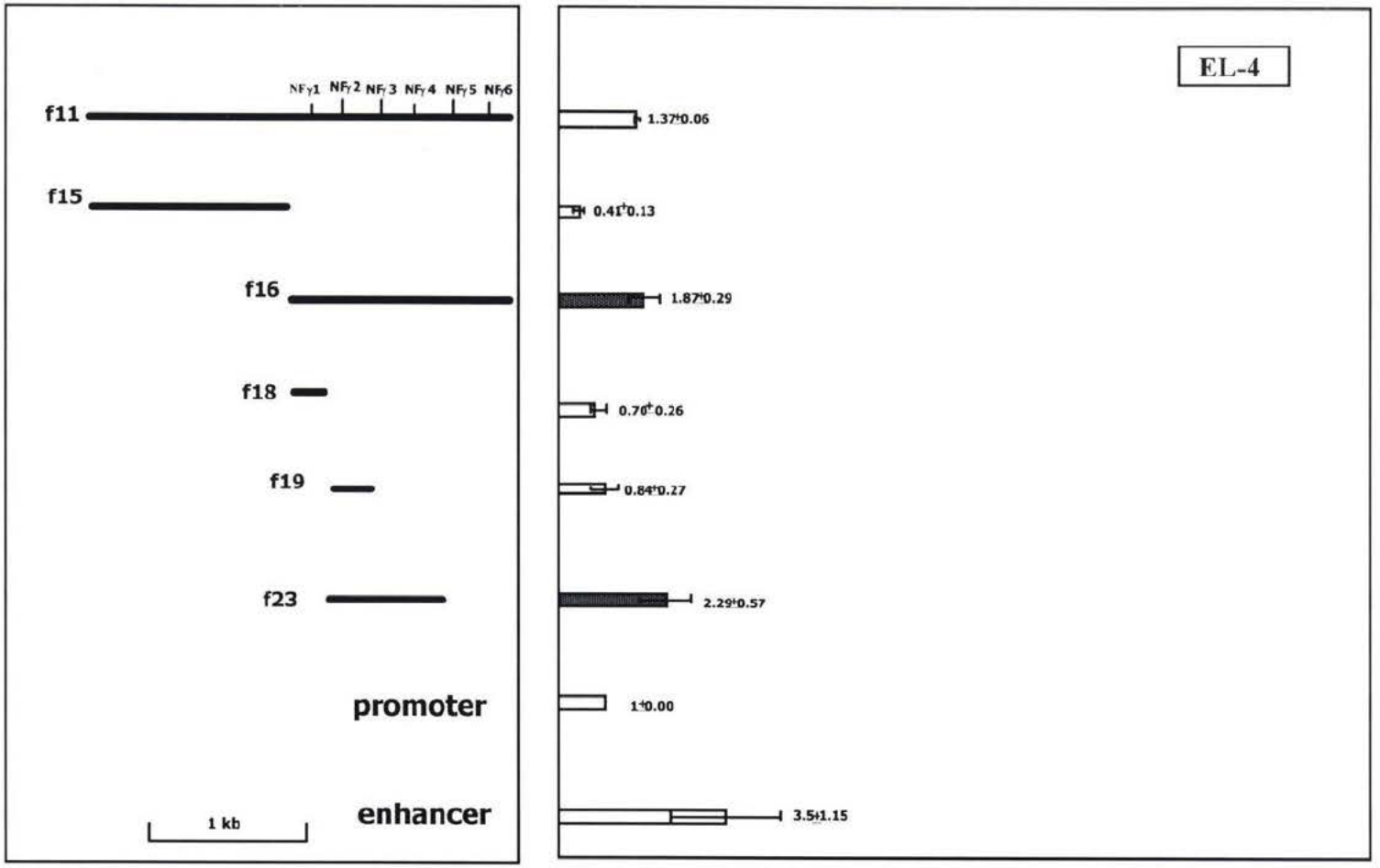


Figure 27B. Tissue specificity of f11 (and the DNA fragments contained in it) in the αβ T cell line EL-4. The values of relative luciferase activities of these fragments are indicated as bars and their standard errors are shown on the right side accordingly. The lucifease activities around 2 are indicated by shaded bars.

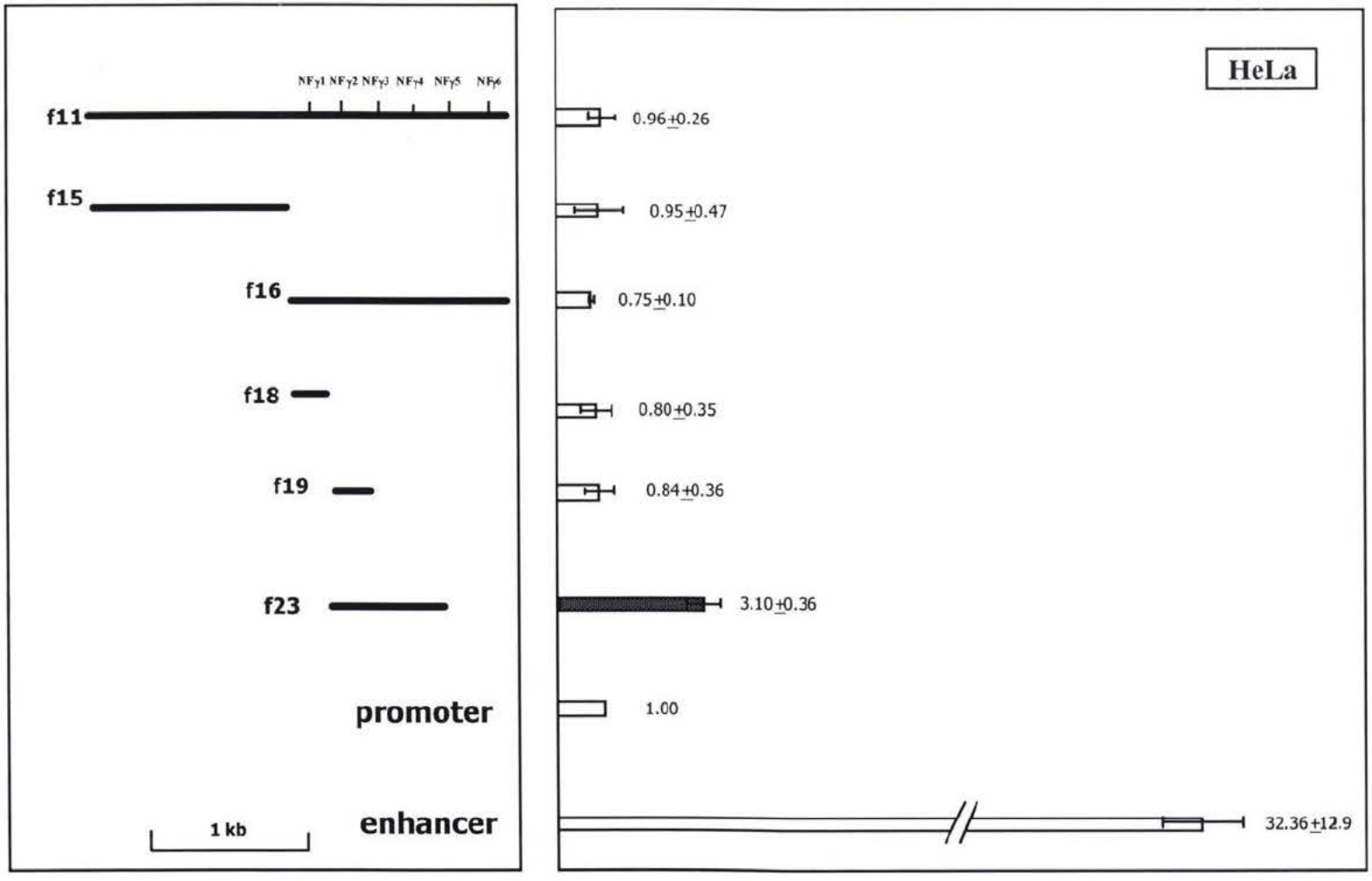


Figure 27C. Tissue specificity of f11 (and the DNA fragments contained in it) in the nonlymphocyte cell line HeLa. The values of relative luciferase activities of these fragments are indicated as bars and their standard errors are shown on the right side accordingly. The relative luciferase activities around 2 are indicated by shaded bars.

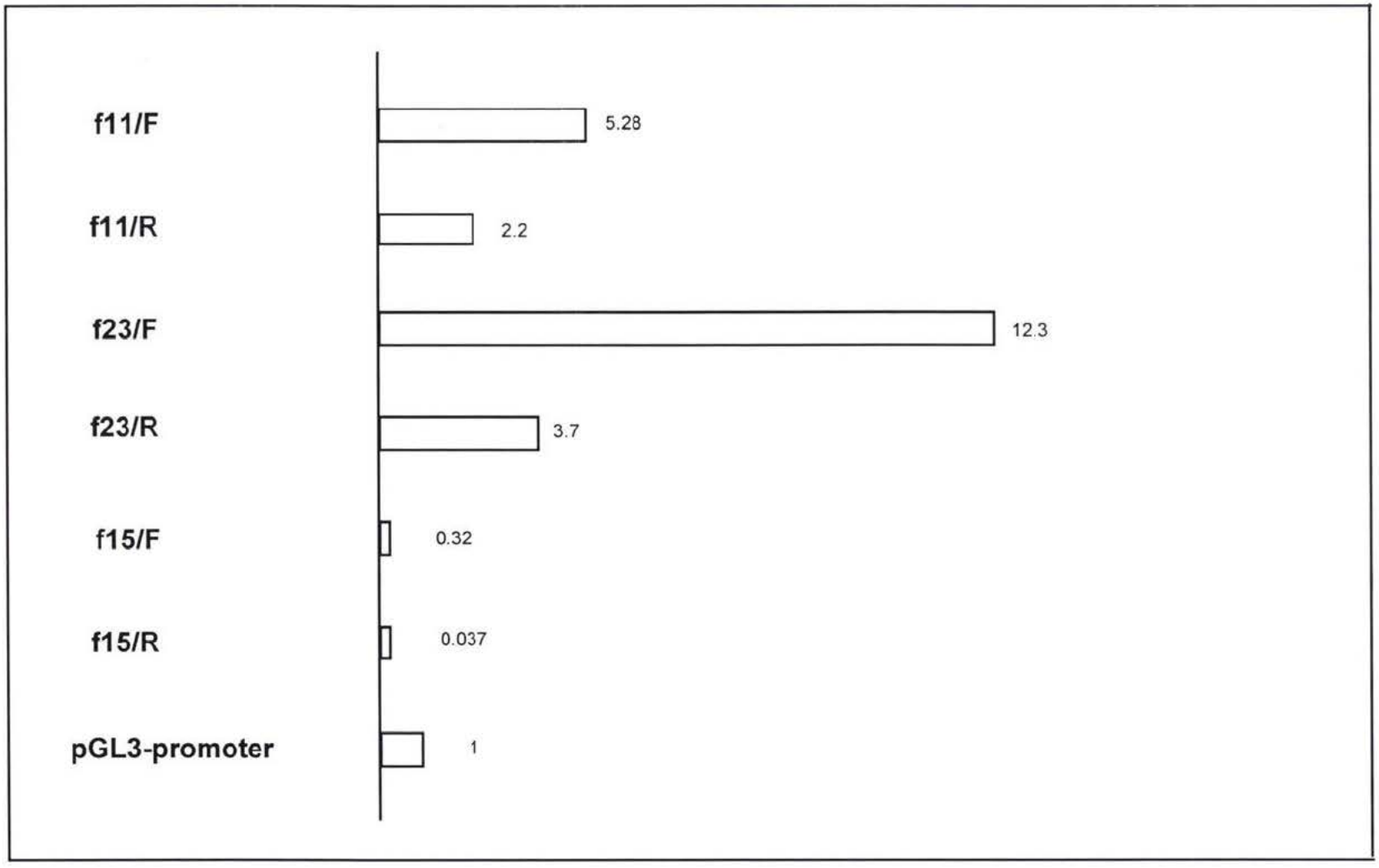


Figure 28. Orientation-independence of the enhancer and silencer. The enhancer and silencer were tested in the γδ T cell line PEER. The relative luciferase activities are shown on the right side of the fragments.

transcription in both orientations while its function is probably confined within cluster 4 because each cluster contains an enhancer of its own at the 3' end of the C genes. This is supported as most of the V(D)J rearrangements are within a particular cluster while interclusteral rearrangements are rare (Pelkonen et al., 1987). Some potential boundary elements between clusters may be responsible for the intraclusteral recombination by preventing enhancers from dispersing their functions (Zhong et al., 1997; Bell et al., 2001).

A detailed analysis of f11 showed that f16 is the full enhancer because it displayed the strongest activity among all the fragments. f16 contains the six NFγs that are found in the mouse γE1 (Spencer et al., 1991). f23 that spans from NFγ2 to NFγ4 has similar activity as the complete enhancer, suggesting that it is the minimal enhancer. When adding the activities of the six NFγ together, the sum equals the activity of f16, which reveals that these NFγs work together in an additive fashion. Such an additive effect has been proposed to explain the combinatory effects of the silencer and enhancer by Lefranc and Alexandre (1995). In addition, in comparing the relative activity of γE1, the activity of γE4 is stronger, which may reflect the relatively stronger activity of the individual NFγ region in γE4. For example, NFγ1 and NFγ5-6 produce up to 9.8 and 5.7 times relative activity in γE4, respectively. In contrast, their counterparts in γE1 carry less functional importance (less than 1 time). Further comparison of the nuclear binding factors between γE1 and γE4 showed that they have a significant difference on the distribution of the motifs for the nuclear binding factors (Figure 29). In the NFγ1 region, γE4 lost the GATA but gained a CBF instead. CBF was proposed to be crucial for the activity of γE1 (Hsiang et al., 1993) as one base mutation in the CBF of γE1 almost abolished the whole activity of γE1. Therefore, the relatively high activity of NFγ1 in γE4 is probably due to the gaining of CBF. Likewise, in the NFγ5-6 region, γE4 gains more binding factors, such as, Sp-1, CTF and NF-1. In addition, the CBF in the NFγ4 region of γE1 is lost in γE4, which may result in NFγ4 in γE4 having no effect on transcription (relative activity is 0.92) in comparison with the control. Therefore, the addition and loss of nuclear binding factors may account for the functional difference between the two enhancers. The difference

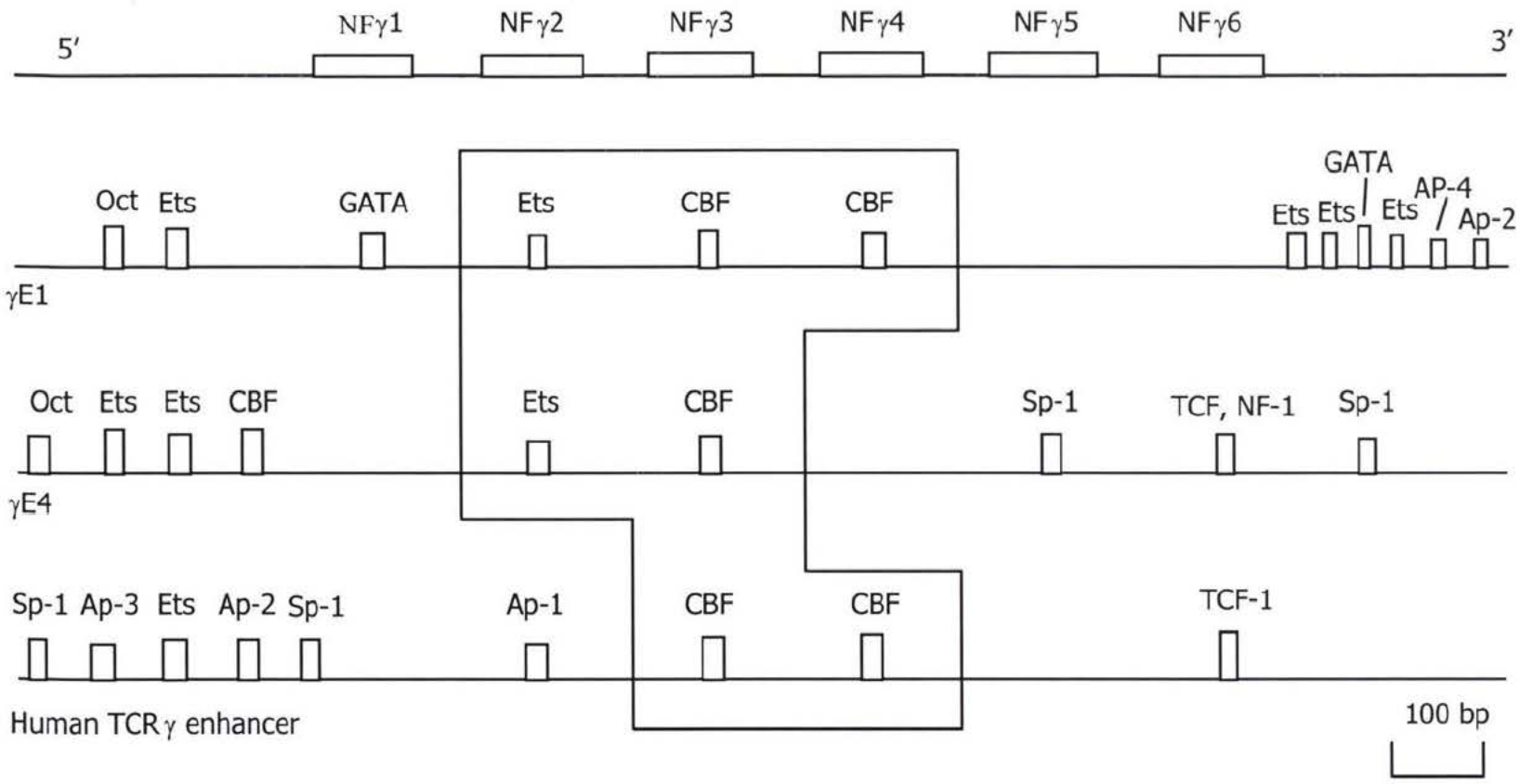


Figure 29. Distribution of the T cell specific nuclear factor binding motifs in the mouse TCR γ enhancer γE1, γE4 and the human TCR γ enhancer. The sequences were analysed by Transfec V3.2 software. The six NFγs represent the corresponding regions identified by DNAse-I footprinting in γE1.

between them is probably due to long separated evolution. γE1 and γE4 are associated with the two homologous genes, Cγ1 and Cγ4, respectively. Based on the calculation of sequence divergence, Cγ4 is 1.5 times older than Cγ1 and thus, their cognate enhancers may have different functions.

Characterization of the mouse γE1 and the human TCR γ enhancer (HγE) indicated that both of the enhancers are T cell specific but not γδ lineage specific (Spencer et al., 1991; Hettmann et al., 1994). In this study, the complete enhancer f16 manifests the strongest activity in γδ T cell line PEER and a slight activity in αβ T cell EL-4 (1.87 times) but there was no activity in the γδ T cell line Molt-13 and nonlymphocyte Hela. This suggests that γE4 is T cell specific rather than γδ lineage T cell specific, which conforms to the characteristics of γE1 and HγE. However, although the minimal enhancer f23 has the strongest activity in PEER, it manifests a certain degree of activity in Molt-13, EL-4 and even Hela, which suggests that the minimal enhancer is not completely T cell specific and it is more possibly a ubiquitous regulatory element. This is different from the result of the γE1 and HγE. Therefore, the other three motifs outside f23, NFγ1, NFγ5 and NFγ6, may provide the T cell specificity.

Similar to other TCR loci, cluster 4 in the mouse TCR γ locus also displays diverse regulatory features. f15 is probably a silencer because it has a direct inhibitory effect and functions in both orientations, which demonstrate the classical silencers characteristics (Strachan & Read, 1996). In comparison with the longer fragment f11 and f14 which all contain f15, the smaller fragment f16 has four times more relative activity. Therefore, it is obvious that f15 inhibits the positive activity of f16 in f11 and f14. In the murine α/δ locus, a silencer was also identified immediately upstream of the α enhancer (Winoto & Baltimore, 1989b). Likewise, this silencer also inhibits the enhancer for by a factor of 4 to 5 just as f15 did, and showed a similar pattern and location as f15. But in the murine TCR γ cluster 1, a similar inhibitory element is located immediately at the 3' of the enhancer γE1 instead of the 5' end (Kappes et al., 1991). Although it is not conserved with f15 in location, it inhibits the activity of γE1 by a factor of 4, which is the same as

that of f15. Therefore, the two silencers are similar in function. In the human TCR γ locus, two silencers are located at the 5' end of the enhancer but they are away from the enhancer by a few kb while there is no sequence similarity between the γE4 silencer and the human one (Lefranc et al., 1995). Therefore, unlike the enhancers, there is less sequence and location conservation among the human and mouse silencers.

The silencer f15 displays strong inhibitory effects in PEER, Molt-13 and EL-4 but not in HeLa, which suggests that it is also T cell specific. The two fragments, the maximum enhancer f16 and the combination of the enhancer and silencer f11, show different patterns of activities in the two γδ T cell lines PEER and Molt-13, which is that f16 has more activity than f11 in PEER but f16 has less activity than f11 in Molt-13. However, these two fragments show equal activities in αβ T cell line EL-4. Since both Molt-13 and PEER are human γδ T cell lines expressing different Vγ genes (Loh et al., 1987), the unique combination of the enhancer and silencer may render the specificity of a particular γδ T sublineage. A similar combination of enhancer and silencer has been found to show differential activities in αβ and γδ T cell lines. The association of enhancer with silencer showed enhancing activity in PEER but inhibitory activity in the αβ T cell line Jurkat, which suggested that the combinatory regulatory elements is one of the sources in the determination of αβ/γδ T cells lineage selection (Lefranc et al., 1995). In addition, silencers have been proposed to be involved in the regulation of the ordered expression of Vγ gene fragments in cluster1 (Ishidaetal, 1990; Winoto & Baltimore, 1989b; Sleckmanetal, 1996). In that case, each Vγ gene fragment may have its own silencer to regulate the ordered expression because there are four Vγ gene fragments in this cluster but only one enhancer. Actually, one silencer has been identified between Vγ6 and Vγ5 (Clausell & Tucker, 1994) and the scenario was conceived as follows: The silencer could be active during all early thymocyte development. Availability of Vγ5 trans-activators differentially depresses the silencing effect on Vγ5 but not Vγ6, allowing targeted recombination to occur. At day 15 to 16, the availability of Vγ6 trans-activators now turn on Vγ6 and inactivate Vγ5, which results in the decline of Vγ5 T cells and the appearance of Vγ6 bearing T cells.

Summary

TCR genes are generated by the recombination of V, (D), J and C gene fragments located in clusters on chromosomes. This process is controlled by diverse regulatory mechanisms. In addition, as multigene families, TCR loci are rich in evolutionary information. Thus, TCR genes and TCR loci have been good models for studying recombination, gene regulation and evolution. The TCR γ genes were first discovered in 1984. From 1984 to 1990, most of the studies were focused on the characterization of basic biological features including molecular structure, gene rearrangements, tissue specific distribution and development. Since the 1990s, research has moved to the characterization of physiological and pathological functions, gene regulation, evolution and the genomic sequence. At the same time that the studies of TCR advanced, the revolutionary project in biological history, the HGP, also came to its productive stage in the early 1990s. With their unique genomic features and the large amount of data accumulated, plus the availability of genomic DNA clones, sequencing of the human and mouse TCR γ loci were chosen as one of the pioneer projects of the HGP. The sequencing work started in 1994 and was completed in 1997. In the past eight years, both the sequence data and analysis results contributed much to the deeper understanding of genomic structure. This project has also provided insight into understanding the evolutionary and immunological implications of multigene families.

Human γδ T cells have proved to be an essentially different T cell lineage from the αβ T cells in many aspects. For example, αβ T cells occupy approximately 90% of the lymphocytes in the peripheral blood, while the percentage of γδ T cells is only 5%. The human αβ T cells recognize processed peptidic antigens with MHC restriction, but γδ T cells predominantly recognize small molecule, non-peptidic and phosphoantigens without MHC restriction. The genes encoding the two types of receptors are located on the α/δ, β and γ loci. Among these three loci, α/δ and β were sequenced and well characterized between 1994 and 1996. The genome organization of the human TCR γ locus was clarified in 1990 and this locus was estimated to be 145 kb. However, the complete sequence obtained in this project is only 140 kb. On the sequence, fourteen V gene

fragments, five J and two C gene fragments were identified. Two V relics were also located in the middle of V9 and V10 in addition to V11 and JP1 respectively. The previously identified enhancer was located 6.5 kb downstream of Cγ2. Structurally, this locus is a SINE rich (13.3%) but LINE poor (6.35%) region. However, the most striking feature of this locus is its characteristic organization and implied evolutionary history. In the human TCR γsequence, there are two J-C repetitive clusters at the 3' region and nine 4.4-4.9 kb V repeats in the GV1 family. The repeats in the GV1 family include the V coding gene fragments and their flanking regions, which share 85-97% sequence similarity. The high level of sequence similarity suggests that these repeats were generated by recent duplications. Phylogenetic analysis of these repeats shows that they are produced by six unequal crossing-over events and one gene conversion from a prototypic structure that contains only an ancestral Vγ1 and Vγ8. Similarly, the two J-C clusters were also duplicated by an unequal crossing-over, and a further deletion resulted in the loss of JP in the J-C cluster at the 3' region. Thus, by enlarging the V, J and C gene numbers, the recombination diversity is greatly increased. Further analysis of the hypervariable regions in the CDRs of the nine V genes in the GV1 family reveal that the DN is much higher than the DS, whereas there is no significant difference between the DN and the DS at the framework regions. This suggests that the framework regions are under negative Darwin selection, while the hypervariable regions evolved under positive selection. After gene duplications, selection processes further diversify the family members to cope with the huge amount of antigens. In addition, the evolutionary analysis of the human TCR GV1 family corrects some of the long-lasting misunderstandings regarding multigene families. It has been believed that most of multigene families evolve through a concerted evolutionary process, which is extrapolated from the studies of rRNA. However, the nine V genes in the GV1 family have diverged substantially, rather than evolving as a unit by frequently exchanging genetic information. In particular, among the nine members, five are functional genes but four are pseudogenes. Therefore, all these features indicate that the GV1 family is generally subject to evolution by a birth-and-death process, in which newly duplicated gene members continue to diversify to generate new specificities.

Similar to humans, mouse γδ T cells also occupy less than 5% of lymphocytes in peripheral blood. However, they can recognize directly, large protein molecules, rather than small non-peptidic phosphoantigens. In addition, the mouse TCR γ locus shows interesting features on the expression of V genes. For example, the V genes expressed in the fetal thymus parallel the gene order on the chromosome. Therefore, the regulatory elements at the mouse TCR γ locus at least partly control this process. The mouse TCR γ locus was mapped in 1993 and the total estimated length was about 205 kb. However, only 167 kb of DNA was sequenced from the seven cosmids in this project. This sequence covers both the 5' and 3' regions of the locus. In the middle of the locus, a region around 37 kb without known *cis*-acting and *trans*-acting elements were not sequenced because the two cosmids covering it were not stable. Similar to its human counterpart, the mouse sequence was also characterized as to its coding gene, structure, gene regulation and evolution. The characterization of the sequence revealed seven V, four J and four C gene fragments. These genes constitute four V-J-C clusters. In addition, the three enhancers identified in the previous studies are also located on the genomic sequence. A new putative enhancer was found 3.5 kb 3' of the Cγ4 gene by sequence comparison. This putative enhancer element has an 80% sequence identity at the six DNase protection regions, NγF1 to NFγ6, which suggests it is a functional enhancer. The structural analysis indicates that this locus is rich in LINEs (30%) but short of SINEs (2.0%), which is opposite to the human TCR γ locus. Among the eight LINEs, five are between 4 to 6 kb while two still keep the monomer promoters, suggesting that they may be still active and capable of retrotransposition. This locus is GC poor in comparison with the other loci. Taken together, the above structural features are consistent with those of Giemsa-dark bands, and therefore, it is quite possible that this locus is located in a Giemsa-dark band. Evolutionary analysis, an important aspect, was also carried out on this sequence. The dot-matrix self-comparison revealed that this locus is rich in locus specific repeats, which suggests an equally rich evolutionary relationship among the repetitive regions. However, relative to the human TCR γ locus, there is less direct evolutionary information available such as boundary sequences marking crossing-overs because the duplication events have taken place over a long period and thus most of the sequences are lost. Despite this lost information, evolutionary analysis still strongly

suggests that clusters 1 and 4 are ancient, while clusters 2 and 3 have been generated recently. Further analysis revealed that cluster 2 was duplicated from cluster 3. In addition, the organizational features may reflect the selection pressure on this locus: for example, the transcriptional orientation of cluster 2 is reversed, while cluster 3 is nonfunctional. Unlike the repetitive V genes in the human GV1 family, there is limited sequence similarity among the V genes in mouse cluster 1. All of these indications imply that the selection on this locus is relaxed. In order to inhibit deletions by homologous recombination, the mouse TCR γ genes are clustered while cluster 2 reversed its transcriptional orientation, which preserved genetic memory. Once the selection is re-imposed, the preserved genes can be re-amplified by duplications to expand into tandem arrays to meet functional needs.

Since the mouse TCR γ genes show characteristic spatial and temporal expression, which suggest that gene regulations are involved in these processes, we also carried out functional analysis to study the regulatory mechanisms of the mouse TCR γ locus. The entire cluster 4 was sub-cloned into an expression vector to search for putative regulatory elements. However, only the region containing the putative enhancer γE4, identified by sequence comparison, shows significant activity on transcription. Since it is a relatively old enhancer, γE4 may possess different properties than those of the other three enhancers. The detailed functional analysis revealed that the maximum enhancer containing the six NFγs is a 651 bp region which exhibits about 32 fold of relative luciferase activity in the γδ T cell line PEER. The minimum enhancer includes the three crucial nuclear protein-binding regions, NFγ2, NFγ3 and NFγ4. Similar to the functionally characterized γE1 and the human TCR γ enhancer, the minimum enhancer also exhibits significant high relative luciferase activity in PEER. The minimum enhancer functions at both orientations. The tissue specific studies show that the maximum enhancer exhibits the maximum enhancing activity in the γδ T cell PEER, in addition to slightly stronger activity in the αβ T cell line EL-4 and γδ T cell line Molt-13, but there is no activity in the non-lymphocyte HeLa cells. Therefore, similar to most of the enhancers in the immune system, γE4 is also T cell specific, but is not γδ T cell lineage specific. In

addition, a 700 bp silencer element was identified at the 5' end of the enhancer, which can inhibit the activity of the maximum enhancer by a factor of four. However, the combination of the enhancer and the silencer showed specific activities in different cell lines rather than producing a fixed inhibitory or enhancing effect. For example, the combination exhibits increased luciferase activity in comparison with the enhancer in Molt-13. But the combination exhibits a decreased activity relative to the enhancer in PEER. Since both Molt-13 and PEER are human γδ T cell lines expressing different Vγ genes, these different expression patterns suggest that the combination of positive and negative elements may contribute the regulation of the expression of a particular γδ T sublineage.

As one of the pioneers participating in the HGP in the 1990s, we witnessed a historical revolution in biology and medicine that took placed in the 1990s. This revolution has made a huge methodological and technological impact on traditional biology and medicine. Likewise, the characterization of the human and mouse TCR γ loci has provided a deeper understanding of genomic structure and immunological function, while greatly facilitating related research. For example, researchers have started to use the human TCR γ sequence to design primers for detecting the recombination patterns of leukemia. Presently, we are entering a post-genome era, in which a great number of new techniques such as SNP-based personalized therapy are emerging. Therefore, one can expect that this genomic revolution will greatly impact both science and human life.

References

1) Abbas, A.K., Lichtman, A.H. & Pober, J.S. (1994). Cellular and molecular Immunology. 2nd Edition. Philadelphia, PA:W.B.Saunders Company

2) Acuto, O., Meuer, S.C., Hodgdon, J.C., Schlossman, S.F. & Reinherz, E.L. (1983). Peptide variability exists within alpha and beta subunits of the T cell receptor for antigen. *J. Exp. Med.* 158(4):1368-73

3) Agrawal, A. Eastman, Q.M. & Schatz, D.G. (1998). Transposition mediated by the V(D)J recombination protein RAG1 and RAG2: implications for the origins of the antigen-specific immune system. *Nature*. 394:744-751

4) Allison, T.J., Winter, C.C., Fournie, J.J., Bonneville, M. & Garboczi, D.N. (2001). Structure of a human gamma/delta T-cell antigen receptor. *Nature*. 411(6839):820-4

5) Anderson, E. & Matsunaga, T. (1996). Jaw adaptive immunity and phylogeny of vertebrate antibody VH gene family. *Res. Immumunol.* 147:233-240

6) Anderson, M.K., Shamblott, M.J., Litman, R.T. & Litman, G.W. (1995). Generation of immunoglobulin light chain gene diversity in Raja erinacea is not associated with somatic rearrangement an exception to a central paradigm of B cell immunity. *J. Exp. Med.* 182:109-19

7) Antequera, F. & Bird, A. (1993). Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. USA.* 90(24):11995-9

8) Antonarakis, S.E. (2001). BACking up the promises. *Nature Genet*. 27:230-232

9) Arden, B., Clark, S.P. & Kabelitz, D. (1995a). Mouse T-cell receptor variable gene segment families. *Immunogenetics*. 42:501-530

10) Arden, B., Clark, S.P., Kabelitz, D. & Mak, T.W. (1995b). Human T-cell receptor variable gene segment families. *Immunogenet*. 42:455-500

11) Arnheim, N. (1983). Concerted evolution of multigene families. Pp. 38-61 in Nei, M and Koehn, R.K. eds. Evolution of genes and proteins. Sinaur, Sunderland, Mass

12) Arstila, T., Arstila, T.P., Calbo, S., Selz, F., Malassis-Seris, M., Vassalli, P., Kourilsky, P. & Guy-Grand, D. (2000). Identical T cell clones are located within the mouse gut epithelium and lamina propia and circulate in the thoracic duct lymph. *J. Exp. Med.* 191(5):823-34

13) Asarnow, D.M., Kuzier, W.A., Bonyhadi, M., Tigellar, R.R., Tucker, P.W. & Allison, J.P. (1988). Limited diversity of γδ antigen receptor genes of Thy-1$^+$ dendrite epithermal cells. *Cell.* 55:837-847

14) Asarnow, D.M., Kuziel, W.A., Bonyhadi, M., Tigelaar, R.E., Tucker, P.W. & Allison, J.P. (1989). Distinct antigen receptor repertoires of two classes of murine epithelium-associated T cells. *Nature*. 341:60-62

15) Asarnow, D.M., Cado, D. & Raulet, D.H. (1993). Selection is not required to produce invariant T-cell receptor gamma-gene junctional sequences. *Nature*. 362:158-160

16) Augustin, A., Kubo, R.T. & Sim, G.K. (1989). Resident pulmonary lymphocytes expressing the γδ T cell receptor. *Nature*. 340:239-241

17) Baker, J.E., Knag, J. Xiong, N., Chen, T. Cado, D. & Raulet, D.H. (1999). A novel element upstream of the V gamma 2 gene in the murine T cell receptor gamma locus cooperates with the 3' enhancer to act as a locus control region. *J. Exp. Med.* 190:669-679

18) Bandeira, A., Itohara, S., Bonneville, M., Burlen-Defranoux, O., Mota-santos, T., Coutinho, A. & Tonegawa, S. (1991). Extrathymic origin of intestinal intraepithelial lymphocytes bearing T-cell antigen receptor γδ. *Proc. Natl. Acad. Sci. USA*. 88:43-47

19) Bank, I., DePinHo, R.A., Brenner, M.B., Cassimeris, J., Alt, F.W. & Chess, L. (1986). A functional T3 molecule associated with a novel heterodimer on the surface of immature human thymocytes. *Nature*. 322:179-181

20) Barth, R.K., Kim, B.S., Lan, N.C., Hunkapiller, T., Sobieck, N., Winoto, A., Gershenfeld, H., Okada, C., Hansburg, D. & Weissman, I.L. (1985). The murine T-cell receptor uses a limited repertoire of expressed V beta gene segments. *Nature*. 316:517-523

21) Bell, A.C., West, A.G. & Felsenfeld, G. (2001). Insulators and Boundaries: Versatile Regulatory Elements in the Eukaryotic Genome. *Science*. 291:447-450

22) Bensmana, M., Mattei, M.-G, Lefranc, M.-P. (1991). Localization of the human T-cell receptor gamma locus at 7p14-p15 by in situ hybridization of the TRG1.4 probe. *Cytogenet. Cell Genet*. 56(1):31-32

23) Bernardi, G. (1995). The human genome: organization and evolutionary history. *Annu. Rev. Genetics*. 29:445-476

24) Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L.& Wiley, D.C. (1987). *Nature*. 329:506-512

25) Blackwood, E.M. & Kadonaga, J.T. (1998). Going the Distance: A Current View of Enhancer Action. *Science*. 281:61-63

26) Bluestone, J.A., Khattri, R. Sciammas, R. & Sperling, A.I. (1995). TCR γδ cells: A specialized T-cell subset in the immune system. *Annu. Rev. Cell. Dev. Biol.* 11:307-353

27) Bodenteich, A. & Chissoe, S. (1994). Automated DNA Sequencing and Analysis (Academic Press). Chapter 6:42-50

28) Born, W., Cady, C. Jones-Carson, J., Mukasa, A., Lahn, M. & O'Brien, R. (1999). Immunoregulatory functions of gamma delta T cells. *Adv. Immunol.* 71:77-144

29) Brand, A.H., Breeden, L., Abraham, J., Sternglanz, R. & Nasmyth, K. (1985). Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of transcriptional enhancer. *Cell*. 41(1):41-8

30) Brenner, M.B., Mclean, J., Dialynas, D.P., Striminger, J.L., Smith, J.A., Owen, F.L., Seidman, J.G., Ip, S., Rosen, F. & Krangel, M.S. (1986). Identification of a putative second T-cell receptor. *Nature*. 322:145-149

31) Brosius J. (1991). Retroposons--seeds of evolution. *Science*. 251:753

32) Brown, S.M. (2000). Get your bioinformatics on the web! *Biotechniques*. 28(2):244-6

33) Bucy, R.P., Chen, C.L., Cihak, J., Losch, U. & Cooper, M.D. (1988). Avian T cells expressing gamma delta receptors localize in the splenic sinusoids and the intestinal epithelium. *J. Immunol.* 141:2200-2205

34) Bukowski, J.F., Morita, C.T., Band, H. & Brenner, M.B. (1998). Crucial role of TCR gamma chain junctional region in prenyl pyrophosphate antigen recognition by gamma delta T cells. *J. Immunol.* 161(1):286-93

35) Bukowski, J.F., Morita, C.T. & Brenner, M.B. (1999). Human gamma delta T cells recognize alkylamines derived from microbes, edible plants, and tea: implications for innate immunity. *Immunity*. 11(1):57-65

36) Buresi, C., Ghanem, N., Huck, S., Lefranc, G. & Lefranc, M.-P. (1989). Exon duplication and triplication in the human T-cell receptor gamma (TCR) constant region genes and RFLP in French, Lebanese, Tunisian and black African populations. *Immunogenetics*. 29:161-172

37) Burge, C.B. (2001). Chipping away at the transcriptome. *Nature Genet*. 27:232-234

38) Burk, M.R., Donda, A., Mariani, F., Mori, L. & De Libero, G. (1997). Functional inactivation in the whole population of human Vγ9/Vδ2 T lymphocytes induced by a nonpeptidic antagonist. *J. Exp. Med.* 185:91-97

39) Carson, G. R., Kuestner, R. E., Ahmed, A., Pettey, C.L. & Concino, M.F. (1991). Six chains of the human T cell antigen receptor CD3 complex are necessary and sufficient for processing the receptor heterodimer to the cell surface. *J. Biol. Chem.* 266:7883-7

40) Charlemagne, J., Fellah, J.S., De Guerra, A., Kerfourn, F. & Partula, S. (1998). T-cell receptor in ectothermic vertebrates. *Immunol. Rev.* 166:87-102

41) Chen, T.L. & Manuelidis, L. (1989). SINEs and LINEs cluster in distinct DNA fragments of Giemsa band size. *Chromosoma.* 98(5):309-31

42) Cheng, S. H., Penninger, J. M., Ferrick, D. A., Molina, T. J., Wallace, V. A. & Mak, T. W. (1991). Biology of Murine γδ T cells. *Crit. Rev. Immunol.* 11(3,4):145-166

43) Chien, Y.H., Jores, R. & Crowley, M.P. (1996). Recognition of γ/δ T cells. *Annu. Rev. Immunol.* 14:511-532

44) Chretien, I., Robert, J., Marcuz, A., Garcia-Sanz, J.A., Courter, M. & Du Pasquier, L. (1996). CTX, a novel molecule specifically expressed on the surface of cortical thymocytes in Xenopus. *Eur. J. Immunol.* 26:780-791

45) Ciccarese, S., Lanave, C. & Saccone, C. (1997). Evolution of T-cell receptor gamma and delta constant region and other T-cell-related proteins in the human-rodent-artiodactyl triplet. *Genetics.* 145:409-419

46) Clausell, A. & Tucker, P. W. (1994). Function analysis of the Vγ3 promoter of the murine γδ T-cell receptor. *Mol. Cell. Biol.* 14 (1):803-814

47) Collins, F.S. & Galas, D. (1993). A new five-year plan for the U.S. Human Genome project. *Science.* 262(5130):43-6

48) Collins, F.S., Green, E.D., Guttmacher, A.E., Guyer, M.S. US National Human Genome Research Institute. (2003). A vision for the future of genomics research. *Nature.* 422(6934): 835-47

49) Cooper, M.D., Bucy, R.P., George, J.F., Lahti, J.M., Char, D. & Chen, L. H. (1989). T-cell development in bird. *Cold Spring Harb. Symp. Quant. Biol.* 54:69-73

50) Craddock, T.P. Zumla, A.M., Ollier, W.E.R., Chintu, C.Z., Muyinda, G.P., Lancaster, F.C. & Boylston, A.W. (2000). Predominance of one T-cell antigen receptor BV hyplotype in African populations. *Immunogenetics.* 51:231-237

51) Cron, R.Q., Coligan, J.E. & Bluestone, J.A. (1990). Polymorphisms and diversity of T-cell receptor-gamma proteins expressed in mouse spleen. *Immunogenetics.* 31:220-228

52) Cross, S.H., Lee, M., Clark, V.H., Craig, J.M., Bird, A.P. & Bickmore, W.A. (1997). The chromosomal distribution of CpG islands in the mouse: evidence for genome scrambling in the rodent lineage. *Genomics*. 40:454-461

53) -Dariavach, P. & Lefranc, M-P. (1989). The promoter regions of the T-cell receptor Vγ9 (TRGV9) and V2δ(TRDV2) genes display short direct repeats but not TATA box. *FEBS. Lett.* 256:185-191

54) Davis, M.M. (1990). T cell receptor gene diversity and selection. *Annu. Rev. Biochem.* 59:475-96

55) Davis, M. M. & Bjorkman, P.J. (1988). T-cell antigen receptor genes and T-cell recognition. *Nature*. 334:395-402

56) Davis, M.M., Boniface, J.J. Reich, Z., Lyons, D., Hampl, J., Arden, B. & Chien, Y.H. (1998). Ligand recognition by αβ T cell receptors. *Annu. Rev. Immunol.* 16:523-544

57) DeBerardinis, R.J., Goodier, J.L., Ostertag, E.M. & Kazazian, H. H. (1998). Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nature Genetics*. 20:288-290

58) Deininger, P.L. & Slagel, V.K. (1988). Recently amplified Alu family members share a common parental Alu sequence. *Mol. Cell. Biol.* 8:4566-4569.

59) De Libero, G. (2000). Tissue distribution, antigen specificity and effector functions of gamma delta T cells in human diseases. *Springer Semin Immunopathol.* 22(3):219-38

60) del Pino, E.M., Murphy, C., Masson, P.H. & Gall, J.G. (1992). 5S rRNA-encoding genes of the marsupial frog *Gastrotheca riobamae*. *Gene*. 111:235-238

61) Delves, P.J. (1994). Cellular Immunology LabFax. Bios Scientific Publishers, Page: 164

62) Dembic, Z. & Vidovic, D. (1990). Specificity and function of Qa-1 restricted γδ T cells. *Res. Immunol.* 141:600-602

63) Dent, A. L., Matis, L. A., Hooshmand, F., Widacki, S.M. Bluestone, J.A. & Hedrick, S.M. (1990). Self-reactive γδ T cells are eliminated in the thymus. *Nature*. 343:714-719

64) Dent, A.L., Matis, L.A., Bluestone, J.A. & Hedrick, S.M. (1993). Evidence for programmed cell death of self-reactive γδ T cell receptor-positive thymocytes. *Eur. J. Immunol.* 23: 2482-87

65) Deusch, K., Pfeffer, K., Reich, K., Gstettenbauer, M., Daum, S., Luling, F. & Classen M. (1991). Phenotypic and functional characterization of human TCR gamma delta+ intestinal intraepithelial lymphocytes. *Curr. Top. Microbiol. Immunol.*173:279-83

66) de Villartay, J.P. & Cohen, D.I. (1990). Gene regulation within the TCR-alpha/delta locus by specific deletion of the TCR-delta cluster. *Res. Immunol.* 141(7):618-23

67) Diaz, P., Cado, D. & Winoto, A. (1994). A locus control region in the T cell receptor alpha/delta locus. *Immunity*. 1:207-217

68) Dudley, E.C., Girardi, M., Owen, M.J. & Hayday, A.C. (1995). Alpha beta and gamma delta T cells can share a late common precursor. *Curr. Biol.* 5:659-669

69) Dunon, D & Imhof, B. A. (1996). T cell migration during ontogeny and T cell repertoire generation. *Infect. Immun.* 63(4):79-93

70) Eastman, Q.M., Lieu, T.M.J. & Schatz, D.G. (1996). Initiation of V(D)J recombination in vitro obeying the 12/23 rule. *Nature*. 380:85-88

71) Elbe, A., Tschachler, E., Steiner, G., Binder, A., Wolff, K. & Stingl, G. (1989). Maturational steps of bone marrow-derived dendritic murine epidermal cells. Phenotypic and functional studies on Langerhans cells Thy-1+ dendritic epidermal cells in the perinatal period. *J. Immunol.* 143:2431-2438

72) Elemento, O., Gascuel, O. & Lefranc, M.P. (2002). Reconstructing the duplication history of tandemly repeated genes. *Mol. Biol. Evol.* 19(3):278-288

73) Elgert, K.D. (1996). Immunology-Understanding the Immune System. A John Wiley & Sons, Inc., Publication

74) Emoto, M., Naito, T., Nakamura, R. & Yoshikai, Y. (1993). Different appearance of gamma delta T cells during salmonellosis between Ityr and Itys mice. *J Immunol.* 150(8 Pt 1):3411-20

75) Faisst, S. & Meyer, S. (1992). Compilation of vertebrate-encoded transcription factors. *Nucleic Acids Research*. 20:3-26

76) Falini, B., Flenghi, L., Pileri, S., Pelicci, P., Fagioli, M., Martelli, M.F., Moretta, L. & Ciccone, E. (1989). Distribution of T cells bearing different forms of the T cell receptor gamma/delta in normal and pathological human tissues. *J. Immunol.* 143(8):2480-8.

77) Fehling, H.J., Gilfillan, S. & Ceredig, A.R. (1999). αβ/γδ lineage Commitment in the Thymus of normal and genetically manipulated mice. *Adv. Immunol.* 71:1-76

78) Ferrarini, M., Ferrero, E., Dagna, L., Poggi, A. & Zocchi, M.R. (2002). Human gammadelta T cells: a nonredundant system in the immune-surveillance against cancer. *Trends Immunol.* 23(1):14-8

79) Feurle, J., Espinosa, E., Eckstein, S., Pont, F., Kunzmann, V., Fournie, J.J., Herderich, M. & Wilhelm, M. (2002). Escherichia coli produces phosphoantigens activating human gamma delta T cells. *J. Biol. Chem.* 277(1):148-54

80) Fields, S. (2001). Proteomics in Genomeland. *Science.* 291:1221-1224

81) Fitch, D.H. & Goodman, M. (1991). Phylogenetic scanning: a computer-assisted algorithm for mapping gene conversions and other recombinational events. *Comput. Appl. Biosci.* 7(2):207-15

82) Font, M.P., Chen, J.C., Bories, N., Duparc, P., Loiseau, L., Degos, H., Cann, D., Cohen, D., Dausset, J. & Sigaux, F. (1988). The Vγ locus of the human T-cell receptor γ gene. repertoire polymorphism of the first variable gene segment subgroup. *J. Exp. Med.* 168:1383-1394

83) Forster, A., Huck, S., Ghanem, N., Lefranc, M.P.& Rabbitts, T.H. (1987). New subgroups in the human T cell rearranging V gamma gene locus. *EMBO. J.* 6(7):1945-50.

84) Fowlkes, B.J. & Pardoll, D.M. (1989). Molecular and cellular events of T cells development. *Adv. Immunol.* 44:207-264

85) Fox, V.L., Strauss, W.M. & Seidman, J.G. (1989). Isolation and restriction map of the V-J interval of the human T cell receptor gamma chain locus. *Genomics.* 4(3):445-8.

86) Fremont, D.H., Rees, W. A. & Kozono, H. (1996). Biophysical studies of T-cell receptors and their ligands. *Curr. Opin. Immunol.* 48:2315-2323

87) Fu, Y.X., Roark, C.E., Kelly, K., Drevets, D., Campell, P., O'Brien, R.L. & Born, W. (1994a). Immune protection and control of inflammatory tissue necrosis by γδ T cells. *J. Immunol.* 153:3101-3115

88) Fu, Y.X., Vollmer, M., Kalataradi, H., Heyborne, K., Reardon, C., Miles, C., O'brien, R. & Born, W. (1994b). Structural requirements for peptides that stimulate a subset of γδ T cells. *J. Immunol.* 152:1578-1588

89) Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J. & Schatz, D.G. (2000). The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.* 18:495-527

90) Furano, A.V., Hayward, B.E., Chevret, P., Catzeflis, F. & Usdin, K. (1994). Amplification of the ancient murine Lx family of long interspersed repeated DNA occurred during the murine radiation. *J. Mol. Evol.* 38:18-27

91) Garman, R. D., Doherty, P.J. & Raulet, D. H. (1986). Diversity, rearrangement, and expression of murine T cell gamma genes. *Cell.* 45:733-742

92) Gauss, G.H. & Lieber, M.R. (1996). Mechanistic constraints on diversity in human V(D)J recombination. *Mol. Cell. Biol.* 16(1):258-269

93) Ghanem, N., Buresi, C., Moisan, J.P., Bensmana, M., Chuchana, P., Huck, S., Lefranc, G. & Lefranc, M.P. (1989). Deletion, insertion, and restriction site polymorphism of the T-cell receptor gamma variable locus in French, Lebanese, Tunisian, and black African populations. *Immunogenetics.* 30(5):350-60.

94) Ghanem, N., Soua, Z., Zhang, X.G., Zijun, M., Zhiwei, Y., Lefranc, G. & Lefranc, M.P. (1991). Polymorphism of the T-cell receptor gamma variable and constant region genes in a Chinese population. *Hum. Genet.* 86(5):450-6.

95) Gilfillan, S., Dierich, A., Lemeur, M., Benoist, C. & Mathis, D. (1993). Mice lacking TdT: mature animals with an immature lymphocyte repertoire. *Science.* 261:1175-1178

96) Gobel, T.W., Chen, C.L., Lahti, J., Kubota, T., Kuo, C.L., Aebersold, R., Hood, L. & Cooper, M.D. (1994). Identification of T-cell receptor alpha-chain genes in chicken. *Proc. Natl. Acad. Sci. USA.* 91:1094-1098

97) Godfrey, D.I., Kennedy, J., Mombaerts, P., Tonegawa, S. & Zlotnik, A. (1994). Onset of TCR-β gene rearrangement and role of TCR-β expression during CD3-CD4-CD8- thymocyte differentiation. *J. Immunol.* 152:4783-4792

98) Goodman, T. & Lefrancois, L. (1988). Expression of the γδ T cell receptor on intestinal CD8+ intraepithelial lymphocytes. *Nature.* 333:855-858

99) Gorczynski, R.M. (1994). Adoptive transfer of unresponsiveness to allogeneic skin grafts with hepatic gamma delta + T cells. *Immunology.* 81(1):27-35

100) Graham, G.J. (1995). Tandem genes and clustered genes. *J. Theor. Biol.* 175(1):71-87

101) Groh, V., Porcelli, S., Fabbi, M., Lanier, L.L., Picker, L.J., Anderson, T., Warnke, R.A., Bhan, A.K., Strominger, J.L. & Brenner, M.B. (1989). Human lymphocytes bearing T cell receptor gamma/delta are phenotypically diverse and evenly distributed throughout thelymphoid system. *J. Exp. Med.* 169:1277-94

102) Grossi, C.E., Ciccone, E., Migone, N., Bottino, C., Zarcone, D., Mingari, M.C., Ferrini, S., Tambussi, G., Viale, O. & Casorati, G. (1989). Human T cells expressing the gamma/delta T-cell receptor (TCR-1): C gamma 1- and C gamma 2-encoded forms of the receptor correlate with distinctive morphology, cytoskeletal organization, and growth characteristics. *Proc. Natl. Acad. Sci. USA.* 86(5):1619-23

103) Grusby, M.J., Auchincloss Jr, H., Lee, R., Johnson, R.S., Spencer, J.P., Zijlstra, M., Jaenisch, R., Papaioaannou, V.E. & Glimcher, L.H. (1993). Mice lacking major

histocompatibility complex class I and class II molecules. *Proc. Natl. Acad. Sci. USA.* 90:3913-3917

104) Gurwitz, D. & Weizman, A. (2001). Animal models and human genome diversity: the pitfalls of inbred mice. *Drug Discov. Today.* 6(15):766-768.

105) Guy-Grand, D., Cerf-Bensussan, N., Malissen, B., Malassis-Seris, M., Briottet, C., & Vassalli, P. (1991). Two gut intraepithelial CD8+ lymphocyte populations with different T cell receptors: A role for the gut epitheliums in T cell differentiation. *J. Exp. Med.* 173:471-481

106) Happ, M.P., Kubo, R.T., Palmer, E., Born, W.K. & O'Brien, R.L. (1989). Limited receptor repertoire in a mycobacteria-reactive subset of γδ T lymphocytes. *Nature.* 342: 696-698

107) Harty, J.T. & Bevan, M.J. (1995). Specific immunity to Listeria monocytogenes in the absence of IFN gamma. *Immunity.* 3(1):109-17

108) Hass, W., Pereira, P. & Tonegawa, S. (1993). γ/δ cells. *Annu. Rev. Immunol.* 11:637-685

109) Havran, W. & Allison, J.P. (1988). Developmentally ordered appearance of thymocytes expressing different T cell antigen receptors. *Nature.* 335:443-445

110) Havran, W.L., Chien, Y. H. & Allison, J.P. (1991). Recognition of self antigens by skin-derived T cells with invariant γδ antigen receptors. *Science.* 252:1430-1432

111) Hayday, A. C., Saito, H., Gillies, S. D., Kranz, D. M., Tanigawa, G., Eisen, H. N. & Tonegawa, S. (1985). Structure, organization, and somatic rearrangement of T cell gamma genes. *Cell.* 40:259-269

112) Hayday, A.C. (2000). Gamma/delta cells: a right time and a right place for a conserved third way of protection. *Annu. Rev. Immunol.* 18:975-1026

113) Hayes, P.H. & Davies, P.L. (1991). Population differences in antifreeze protein gene copy number and arrangement in winter flounder. *Genome.* 34:174-177

114) Hayes, S.M., Sirr, A., Jacob, S., Sim, G. K. & Augustin, A. (1996). Role of IL-7 in the shaping of the pulmonary γδ T cell repertoire. *J. Immunol.* 156:2723-2729

115) Hedrick, S.M. & Eidelman, F.J. (1993). Lymphocytes antigen receptors In "Fundamental Immunology" (W. E. Paul, ed.), pp. 383-420. Raven Press, New York.

116) Heilig, J.S. & Tonegawa, S. (1986). Diversity of murine gamma genes and expression in fetal and adult T lymphocytes. *Nature.* 322(6082):836-40.

117) Heilig, J.S. & Tonegawa, S. (1987). T-cell γ gene is allelically but not isotypically excluded and is not required in known functional T-cell subsets. *Proc. Natl. Acad. Sci. USA*. 84:8070-8074

118) Hein, W. R. & Mackay, C.R. (1991). Prominence of γδ T cells in the ruminant immune system. *Immunol. Today*. 12:30-34

119) Hein, W.R. & Dudler, L. (1993). Divergent evolution of the T-cell repertoire: extensive diversity and developmentally regulated expression of the sheep gamma/delta T-cell receptor. *EMBO. J.* 12:715-724

120) Hein, W. (1994). Structural and functional evolution of the extracellular regions of T cell receptors. *Sem. Immunol.* 6:361-372

121) Heinkoff, S., Grene, E., Pietrokovski, S. Bork, P., Attwood, T. & Hood, L. (1997). Gene families: the taxonomy of proteins paralogs and chimeras. *Science*. 278(24):609-614

122) Hesslein, D.G. & Schatz, D.G. (2001). Factors and forces controlling V(D)J recombination. *Adv. Immunol.* 78:169-232.

123) Hettmann, T., Doherty, P.J. & Cohen, A. (1992). The human T cell receptor gamma genes are transcribed from TATA-less promoters containing a conserved heptamer sequence. *Mol. Immunol.* 24:1073-1080

124) Hettmann, T. & Cohen, A. (1994). Identification of a T cell-specific transcriptional enhancer 3' of the human T cell receptor gamma locus. *Mol. Immunol.* 31:315-322

125) Hettmann, T. & Cohen, A. (1995). Identification of an ionomycin/cyclosporin A-responsive element within the human T cell receptor gamma enhancer. *Eur. J. Immunol.* 25(12):3356-64

126) Heyborne, K., Cranfill, R. L., Carding, S. R., Born, W. & O'Brien, R. (1992). Characterization of γδ T lymphocytes at the maternal-fetal interface. *J. Immunol.* 149:2872-2878

127) Hisaeda H, Sakai T, Nagasawa H, Ishikawa H, Yasutomo K, Maekawa Y, Himeno K. (1996). Contribution of extrathymic gamma delta T cells to the expression of heat-shock protein and to protective immunity in mice infected with Toxoplasma gondii. *Immunology*. 88(4):551-7

128) Holoshitz, J., Koning, F., Coligan, J.E., De Bruyn, J. & Strober, S. (1989). In vivo depletion of BoT4 (CD4) and of non-T4/T8 lymphocyte subsets in cattle with monoclonal antibodies. *Eur. J. Immunol.* 19:757

129) Hong, S.C., Chelouche, A., Lin, R.H., Shaywitz, D., Braunstein, N.S., Glimcher, L. & Janeway, C.A. (1992). An MHC interaction site maps to the amino-terminal half of the T cell receptor alpha chain variable domain. *Cell.* 69(6):999-1009

130) Hsiang, Y. H., Spencer, D., Wang, S., Speck, N. A. & Raulet, D. H. (1993). The role of viral enhancer "core" motif-related sequences in regulating T cell receptor-γ and -δ gene expression. *J. Immunol.* 150:1905-1916

131) Huck, S., Dariavach, P. & Lefranc, M.P. (1988). Variable region genes in the human T-cell rearranging gamma (TCG) locus: V-J junction and homology with the mouse genes. *EMBO. J.* 7(3):719-726

132) Hughes, A.L. & Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex loci reveals overdominant selection. *Nature.* 335:167-170

133) Hughes, A.L. (1995). Origin and evolution of HLA class I pseudogenes. *Mol. Bio. Evol.* 12:247-258

134) Hughes, A.L. & Yeager, M. (1997). Molecular evolution of the vertebrate immune system. *Bioassay.* 19(9):777-86

135) Huppi, K., D'Hoostelaere, L., Kiefer, M., Steinmetz, M. & Jouvin-Marche, E. (1986). The context of T-cell receptor gamma chain genes among wild mouse species. *Immunogenetics.* 24:304-308

136) Ikuta, k., Kina, T., MacNeil, I., Uchida, N., Peault, B. Chien, Y. H. & Weissman, I.L. (1990). A developmental switch in thymic lymphocyte maturation potential occurs at the level of hematopoietic stem cells. *Cell.* 62:863-874

137) Ishida, I., Verbeek, S., Bonneville, M., Itohara, S., Berns, A. & Tonegawa, S. (1990). T-cell receptor gamma delta and gamma transgenic mice suggest a role of a gamma gene silencer in the generation of alpha beta T cells. *Proc. Natl. Acad. Sci. USA.* 87:3067-71

138) Ito, K., Bonneville, M., Takagaki, Y., Nakanishi, N., Kanagawa, O., Kreck, E.G. & Tonegawa, S. (1989). Different γδ T-cell receptors are expressed on thymocytes at different stages of development. *Proc. Natl. Acad. Sci. U S A.* 96:631-635

139) Itohara, S., Farr, A. G., Lafaille, J.J., Bonneville, M., Tagagaki, Y., Hass, W. & Tonegawa, A. (1990). Homing of a γδ Thymocyte subset with homogeneous T cell receptors to mucosal epithelia. *Nature.* 343:754-757

140) Itohara, S., Mombaerts, P., Lafaille, J., Iacomini, J., Nelson, A., Clark, A.R., Hooper, M.L., Farr, A. & Tonegawa, S. (1993). T cell receptor d gene mutant mice: independent generation of αβ T cells and programmed rearrangements of γδ TCR genes. *Cell.* 72: 337-348

141) Iwamoto, A., Rupp, F., Ohashi, P. S., Walker, C. L., Pircher, H., Joho, R., Hengartner, H. & Mak, T.W. (1986). T cell-specific γ genes in C57BL/10 mice: sequence and expression of new constant and variable region genes. *J. Exp. Med.* 163:1203-1212

142) Jackson, S.P. & Jeggo, P.A. (1995). DNA double-strand break repair and V(D)J recombination: involvement of DNA-PK. *Trends Biochem. Sci.* 20:412-416

143) Jameson, J., Ugarte, K., Chen, N., Yachi, P., Fuchs, E., Boismenu, R. & Havran, W.L. (2002). A role for skin gammadelta T cells in wound repair. *Science.* 296(5568):747-9.

144) Jarry, A., Cerf-Bensussan, N., Broussse, N. Selz, F. & Guy-Grand, D. (1990). Subsets of CD3+ (T cell receptor alpha/beta or gamma/delta) and CD3- lymphocytes isolated from normal human gut epithelium display phenotyical features different from their counterparts in peripheral blood. *Eur. J. Immunol.* 20:1097-103

145) Jeffreys, A.J. & Harris, S. (1982). Processes of gene duplication. *Nature.* 296:9-10

146) Jindal, S., Dudani, A.K., Singh, B., Harley, C.B. & Guta, R.S. (1989). Primary structure of a human mitochondrite protein homologous to the bacterial and plant chaperonins and to the 65-kilodalton mycobacterial antigen. *Mol. Cell. Biol.* 9:2279-83

147) Jorgensen, J.L., Reay, P.A., Ehrich, E.W. & Davis, M.M. (1992). Molecular components of T-cell recognition. *Annu. Rev. Immunol.* 10:835-837

148) Kabelitz, D., Bender, A., Schondelmaier, S., Schoel, B. & Kaufmann, S. H. E. (1990). A large fraction of human peripheral blood γδ+ T cells is activated by Mycobacterium tuberculosis but not by its 65 kD heat shock protein. *J. Exp. Med.* 171:667-679

149) Kamper, S.M. & McKinney, E.C. (2002). Polymorphism and evolution in the constant region of the T-cell receptor *beta* chain in an advanced teleost fish. *Immunogenetics*. 53:1047-1054

150) Kanamori, Y., Ishimaru, K., Nanno, M., Maki, K., Ikuta, K., Nariuchi, H & Ishikawa, H. (1996). Identification of novel lymphoid tissues in murine intestinal mucosa where clusters of c-kit+ IL-7R+ Thy1+ lympho-hemopoietic progenitors develop. *J. Exp. Med.* 184(4):1449-59

151) Kang, J., Coles, M. & Raulet, D.H. (1999). Defective Development of γ/δ T Cells in Interleukin 7 Receptor-deficient Mice Is Due to Impaired Expression of T cell Receptor g Genes. *J. Exp. Med.* 190(7):973-982

152) Kappes, D. J., Browne, C. P. & Tonegawa, S. (1991). Identification of a T-cell-specific enhancer at the locus encoding T-cell antigen receptor γ chain. *Proc. Natl. Acad. Sci. USA.* 88:2204-2208

153) Kappes, D. & Strominger, J.L. (1988). Human class II major histocompatibility complex genes and proteins. *Annu. Rev. Biochem.* 57:991-1028

154) Kaufmann, S.H.E, Blum, C. & Yamamoto, S. (1993). Crosstalk between α/β T cells and γ/δ T cells from a common precusors. *Sem. Immunol.* 9:171-179

155) Kaufmann, S.H.E. (1996). γ/δ and other unconventional T lymphocytes: what do they see and what do they do? *Proc. Natl. Acad. Sci. USA.* 93:2272-2279

156) Kazazian, H.H. & Moran, J.V. (1998). The impact of L1 retrotransposons on the human genome. *Natural Genetics.* 19:19-24

157) King, D.P., Hyde, D.M., Jackson, K.A., Novosad, D.M., Ellis, T.N., Putney, L., Stovall, M.Y., Van Winkle, L.S., Beaman, B.L. & Ferrick, D.A. (1999). Cutting edge: protective response to pulmonary injury requires gamma delta T lymphocytes. *J. Immunol.* 162(9):5033-6

158) Klein, J. (1986). How many class II immune response genes? A commentary on a reappraisal. *Immunogenetics.* 23(5):309-11

159) Komano, H., Fujiura, Y., Kawaguchi, M., Matsumoto, S., Hashimoto, Y., Obana, S., Mombaerts, P., Tonegawa, S., Yamamoto, H., Itohara, S., Nanno, M. & Ishikawa, H. (1995). Homeostatic regulation of intestinal epithelia by intraepithelial γδ T cells. *Proc Natl. Acad. Sci. USA.* 92:6247-5151

160) Konig, R. (2002). Interactions between MHC molecules and co-receptors of the TCR. *Curr. Opin. Immunol.* 14(1):75-83

161) Koop, B. F., Wilson, R. K., Wang, K., Vernooij, B., Zaller, D., Kuo, C.L., Seto, D., Toda, M. & Hood, L. (1992). Organization, structure, and function of 95 kb of DNA spanning the murine T-cell receptor Cα/Cδ region. *Genomics.* 13:1209-1230

162) Koop, B.F., Rowen, L., Wang, K., Kuo, C.L., Seto, D., Lenstra, J.A., Howard, S., Shan, W., Deshpande, P. & Hood, L. (1994). The human T-cell receptor TCRAC/TCRDC (C alpha/C delta) region: organization, sequence, and evolution of 97.6 kb of DNA. *Genomics.* 19(3):478-93.

163) Kopacz, J. & Kumar, N. (1999). Murine gamma delta T lymphocytes elicited during Plasmodium yoelii infection respond to Plasmodium heat shock proteins. *Infect. Immun.* 67(1):57-63

164) Kronenberg, M. (1994). Antigens recognized by gamma delta T cells. *Curr. Opin. Immunol.* 6:64-71

165) Kumar, S., Tamura, K. & Nei, M. (1993). MEGA: Molecular evolutionary Genetic Analysis. Version 1.01 (The Pennsylvania State University, University Park, PA)

166) Kuziel, W.A., Kienker, L.J. & Tucker, P.W. (1994). Physical linkage of mouse TCR γ-V genes. I. Evolutionary and regulatory features of the Vγ1.1-Vγ1.2 intergenic region. *Immunogenetics*. 39(4):296-297

167) Ladel, C.H., Blum, C. & Kaufmann, S.H. (1996). Control of natural killer cell-mediated innate resistance against the intracellular pathogen Listeria monocytogenes by gamma/delta T lymphocytes. *Infect. Immun.* 64(5):1744-9

168) Lafaille, J.J., deCloux, A., Bonneville, M., Takagaki, Y. & Tonegawa, S. (1989). Junctional sequences of T cell receptor gamma delta genes: implications for gamma delta T cell lineages and for a novel intermediate of V-(D)-J joining. *Cell*. 59: 859

169) Lai, E., Wang, K., Acdalovic, N. & Hood, L. (1991). Rapid restriction map construction using a modified pWE15 cosmid vector and a robotic workstation. *Biotechniques*. 11:212-217

170) Lambolez, F., Azogui, O., Joret, A.M., Garcia, C., von Boehmer, H., Di Santo, J., Ezine, S. & Rocha, B. (2002). Characterization of T cell differentiation in the murine gut. *J. Exp. Med.* 195(4):437-49.

171) Lander, E.S. et al. (2001). Initial sequencing and analysis of the human genome. *Nature*. 409(6822):860-921

172) Lauzurica, P. & Krangel, M.S. (1994). Enhancer-dependent and –independent Steps in the Rearrangement of a Human T cell Receptor δ Transgene. *J. Exp. Med.* 179:43-55

173) Leduc, I., Hempel, W.M., Mathieu, N., Verthuy, C., Bouvier, G., Watrin, F. & Ferrier, P. (2000). T cell development in TCR β enhancer-deleted mice: implications for αβ cell lineage commitment and differentiation. *J. Immunol.* 165:1364-1373

174) Lee, H.C., Ye, S.K., Honjo, T. & Ikuta, K. (2001). Induction of Germline Transcription in the Human TCR γ Locus by STAT5. *J. Immunol.* 167:320-326

175) Lefranc, M.P. & Rabbitts, T.H. (1985). Two tandemly organized human genes encoding the T-cell γ constant region sequence show multiple rearrangement in difference T-cell types. *Nature*. 316:464-466

176) Lefranc, M.P., Forster, A., Baer, R., Stinson, M.A. & Rabbitts, T.H. (1986a). Diversity and rearrangement of the human T cell rearranging γ genes: nine germ-line variable genes belonging to two subgroups. *Cell*. 45:237-246

177) Lefranc, M.P., Forster, A. & Rabbitts, T.H. (1986b). Genetic polymorphism and exon changes of the constant regions of the human T-cell rearranging gene γ. *Proc. Natl. Acad. Sci. USA*. 83:9596-9600

178) Lefranc, M.P. & Rabbitts, T.H. (1989). The human T-cell receptor gamma (TCRG) genes. *Trends Bioch. Sci.* 14:214-218

179) Lefranc, M.P. & Rabbitts, T.H. (1990). Genetic organization of the human T-cell receptor. *Res. Immuno*l. 141:565-577

180) Lefranc, M.P. & Alexandre, D. (1995). γδ lineage-specific transcription of human T cell receptor γ genes by a combination of a non-lineage-specific enhancer and silences. *Eur. J. Immunol.* 25:617-622

181) Lefrancois, M.P., Chuchana, P., Dariavach, P., Nguyen, C, Huck S, Brockly F, Jordan B, Lefranc G. (1989). Molecular mapping of the human T cell receptor gamma (TRG) genes and linkage of the variable and constant regions. *Eur. J. Immunol.* 19(6): 989-94

182) Leiden, J. M. (1993). Transcriptional regulation of T cell receptor genes. *Annu. Rev. Immunol.* 11:539-570

183) Lenschow, D.J., Walunas, T. L. & Lluestone, J.A. (1996). CD28/B7 system of T cell costimulation. *Annu. Rev. Immunol.* 14:233-58

184) Levinson, W. & Jawetz, E. (1996). Medical microbiology and immunology. Appleton & Lange. 4 Edition.

185) Lew, A.M., Pardoll, D.M., Maloy, W.L., Fowlkes, B.J. & Kruisbeek, A. (1986). Characterization of T cell receptor γ chain expression in a subset of murine thymocytes. *Science.* 234:1401-1405

186) Li, H., Lebedeva, M.I., Llera, A.S. Fields, B.A. Brenner, M.B. & Mariuzza, R.A. (1998). Structure of the Vδ domain of a human γδ T cell antigen receptor. *Nature.* 391:502-506

187) Li, W. H. (1997). Molecular Evolution. (Sinauer Associates, Inc, Sunderland, Massachusetts)

188) Liggett, S.B. (2001). Pharmacogenetic applications of the Human Genome project. *Nat. Med.* 7(3):281-3

189) Litman, G.W. (1999). Evolution of antigen binding receptors. *Annu. Rev. Immunol.* 17:109-147

190) Littman, D. R., Newton, M., Crommie, D., Ang, S., Seidman, J. G., Gettner, S. N. & Weiss, A. (1987). Characterization of an expressed CD3-associated Ti gamma-chain reveals C-gamma domain polymorphism. *Nature*. 326:85-88

191) Loh, E.Y., Lanier, L.L., Turck, C.W., Littman, D.R., Davis, M.M., Chien, Y.H. & Weiss, A. (1987). Identification and sequence of a fourth human T cell antigen receptor chain. *Nature*. 330(6148):569-72.

192) Lynch, M. & Conery, J.S. (2000). The Evolutionary fate and consequences of duplicate genes. *Science*. 290:1151-1154

193) Machugh, N.D., Mburn, J.K., Carol, M.J., Wyatt, C.R., Orden, J.A. & Davis, W.C. (1997). Identification of two distinct subsets of bovine γδ T cells with unique cell surface phenotype and tissue distribution. *Immunology*. 92:340-345

194) Mak, T.W. & Ferrick, D.A. (1998). The gamma/delta T-cell bridge: linking innate and acquired immunity. *Nat. Med.* 4(7):764-5

195) Manistis, T., Fritsch, E.F. & Sambrook, J. (1989). Molecular Cloning: A laboratory Manual, 2nd edition. (Cold Spring Harbour Laboratory Press, New York)

196) Marchalonis, J.J., Schluter, S. F., Bernstein, R.M., Shen, S. & Edmundson, A.B. (1998). Phylogenetic Emergence and Molecular Evolution of the Immunoglobulin Family. *Adv. Immunol.* 70:417-506

197) Martindale, D.W., Wilson, M.D., Wang, D., Burke, R.D., Chen, X., Duronio, V. & Koop BF. (2000). Comparative genomic sequence analysis of the Williams syndrome region (LIMK1-RFC2) of human chromosome 7q11.23. *Mamm. Genome*. 11(10):890-8.

198) Matsuzaki, G., Hiromatsu, K., Yoshikai, Y., Muramori, K. & Nomoto, K. (1993). Characterization of T-cell receptor gamma delta T cells appearing at the early phase of murine Listeria monocytogenes infection. *Immunology*. 78(1):22-7

199) Matsunaga, T. (1998). Did the first adaptive immunity evolve in the gut of ancient jawed fish? *Cytogenet. Cell. Genet*. 80:138-141

200) Matsuo, Y. & Yamazaki, T. (1989). Nucleotide variation and divergence in the histone multigene family in *Drosophila melanogaster*. *Genetics*. 122:87-97

201) McBlane, J.F., van Gent, D.C., Ramsden, D.A., Pomeo, C., Cuomo, C.A., Gellert, M. & Oettinger, M.A. (1995). Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell*. 83:387-95

202) McDevitt, H.O. (2000). Discovering the role of the major histocompatibility complex in the immune response. *Annu. Rev. Immunol.* 18:1-17

203) McKusick, V.A. (1989). HUGO News: The human Genome Organization: History, purposes, and membership. *Genomics*. 5:385-387

204) McMenamin, C., Pim, C., Mckersey, M. & Holt, P.G. (1994). Regulation of IgE responses to inhaled antigen in mice by antigen-specific γδ T cells. *Science*. 265:1869-1871

205) Medzhitov, R., Preston-Hurlburt, P. & Janeway, C.A. (1997). A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature*. 388(6640):394-7.

206) Meier, J.T. & Lewis, S.M. (1993). P nucleotides in V(D)J recombination: a fine-structure analysis. *Mol. Cell. Biol.* 13(2):1078-1092

207) Metzenberg, A.B., Wurzer, G., Huisman, T.H. & Smithies, O. (1991). Homology requirements for unequal crossing over in humans. *Genetics*. 128(1):143-61

208) Miccoli, M.C., Lipsi, M.R., Massari, S., Lanave, C. & Ciccarese S. (2001). Exon-intron organization of TRGC genes in sheep. *Immunogenetics*. 53(5):416-22.

209) Miyagawa, F., Tanaka, Y., Yamashita, S., Mikami, B., Danno, K., Uehara, M. & Minato, N. (2001). Essential contribution of germline-encoded lysine residues in Jgamma1.2 segment to the recognition of nonpeptide antigens by human gammadelta T cells. *J. Immunol.* 167(12):6773-9

210) Moingeon, P., Ythier, A., Goubin, G., Faure, F., Nowill, A., Delmon, L., Rainaud, M., Forestier, F., Daffos, F. & Bohuon, C. (1986). A unique T-cell receptor complex expressed on human fetal lymphocyets displaying natural-killer-like activity. *Nature*. 323:638-40

211) Mombaerts, P., Arnoldi, J., Russ, F., Tonegawa, S. & Kaufmann, S.H.E. (1993). *Nature*. 365:53-56

212) Morita, C.T., Beckman, E.M., Bukowski, J.F., Tanaka, Y., Band, H., Bloom, B.R., Golan, D.E. & Brenner, M.B. (1995). Direct presentation of nonpeptide prenyl pyrophosphate antigens to human gamma delta T cells. *Immunity*. 3(4):495-507

213) Mosman, T.R. & Coffman, R.L. (1989). TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. *Annu. Rev. Immunol.* 7:145-73

214) Mukasa, A., Lahn, M., Pflum, E.K., Born, W. & R. L., O. B. (1997). Evidence that the same γδ T cells respond during infection-induced and autoimmune inflammation. *J. Immunol*. 159:5787-5794

215) Mural, R.J. & et al. (2002). A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science*. 296:1661-1671

216) Nadeau, J.H. & et al. (2001). Sequence interpretation. Functional annotation of mouse genome sequences. *Science*. 291(5507):1251-5

217) Nagler-Anderson, C., Mcnir, L.A. & Cradock, A. (1992). Self-reactive, T cell receptor γδ+, lymphocytes from the intestinal epithelium of weanling mice. *J. Immunol.* 149:2315-2322

218) Nakamura, T., Matsuzaki, G. & Nomoto, K. (1999). The protective role of T-cell receptor Vgamma1+ T cells in primary infection with Listeria monocytogenes. *Immunology*. 96(1):29-34.

219) Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3:418-426

220) Nei, M., Gu, X. & Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA.* 94(15):7799-806.

221) Nei, M., Rogozin, I.B. & Piontkivska, H. (2000). Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc. Natl. Acad. Sci. USA*. 97(20):10866-71.

222) Ninomiya, T., Takimoto, H., Matsuzaki, G., Hamano, S., Yoshida, H., Yoshikai, Y., Kimura, G. & Nomoto, K. (2000). Vgamma1+ gammadelta T cells play protective roles at an early phase of murine cytomegalovirus infection through production of interferon-gamma. *Immunology*. 99(2):187-94.

223) Nishimura, H., Emoto, M., Hiromatsu, K., Yamamoto, S., Matsuura, K., Gomi, H., Ikeda, T. Itohara, S. & Yoshikai, Y. (1995). The role of gamma delta T cells in priming macrophages to produce tumor necrosis factor-alpha. *Eur. J. Immunol.* 25:1465-1468

224) O'Brien, R.L., Happ, M.P., Dallas, A., Palmer, E., Kubo, R. & Born, W.K. (1989). Stimulation of a major subset of lymphocytes expressing T cell receptor γδ by an antigen derived from Mycobacterium tuberculosis. *Cell.* 57:667-674

225) O'Brien, R.L., Fu, Y.X., Cranfill, R., Dallas, A., Reardon, C., Lang, J., Carding, S.R., Kubo, R. & Born, W. (1992). Heat shock protein Hsp-60 reactive γδ cells: A large, diversified T lymphocyte subset with highly focused specificity. *Proc. Natl. Acad. Sci. USA.* 89:4348-4352

226) O'Brien, R.L., Yin, X., Huber, S.A., Ikuta, K. & Born, W.K. (2000). Deplettion of a γδ T cell subset can increase host resistance to a bacterial infection. *J. Immunol.* 6473-6479

227) Oettinger, M.A., Schatz, D.G., Gorka, C. & Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science.* (Washington, D.C.) 248:1517-1523

228) Ofir, R., Novick, I., Krup, M., Cleveland, J.L., Ihle, J. N. & Weistein, Y. (1995). Structural and functional analysis of the promoter of the murine Vγ1.1 T cell receptor gene. *Eur. J. Immunol.* 25:3070-3078

229) Ohshima, Y. & Gotoh, Y. (1987). Signals for the selection of a splice site in pre-mRNA. Computer analysis of splice junction sequences and like sequences. *J. Mol. Biol.* 195(2):247-259

230) Ohta, T. (1991). Multigene families and the evolution of complexity. *J. Mol. Evol.* 33:34-41

231) Ohta, T. (1995). Evolution of immunoglobin VH pseudogenes in chickens. *Mol. Biol. Evol.* 12:94-102

232) Ohta, T. (2000). Evolution of gene families. *Gene.* 259:45-52

233) Ono, H., Klein, D., Vincek, V., Figueroa, F., O'hUigin, C., Tichy, H. & Klein, J. (1992). Major histocompatibility complex class II genes of zebrafish. *Proc. Natl. Acad. Sci. USA.* 89(24):11886-90

234) Ota, T. & Nei, M. (1994). Diversity evolution and evolution by the birth-and-death process in the immunoglobin VH gene family. *Mol. Biol. Evol.* 11:469-482

235) Panchamoorthy, G., McLean, J., Modlin, R. L., Morita, C.T., Ishikawa, S., Brenner, M. B. & Band, H. (1991). A predominance of the T cell receptor of Vγ2/Vδ2 subset in human mycobacteria-responsive T cells suggests germline gene encoded recognition. *J. Immunol.* 147:3360-3369

236) Pardoll, D.M., Fowlkes, B.J., Bluestone, J.A., Kruisbeek, A., Maloy, W.L., Coligan, J.E. & Schwartz, R.H. (1987). Differential expression of two distinct T-cell receptors during thymocyte development. *Nature.* 326(6108):79-81

237) Parker, C.M., Groh, V., Band, H., Porcelli, S.A., Morita, C., Fabbi, M., Glass, D., Strominger, J.L. & Breener, M.B. (1990). Evidence of extrathymic changes in the T cell receptors during thymocyte development. *Nature.* 326:79-81

238) Partula, S., de Guerra, A., Fellah, J.S. & Charlemagne, J. (1995). Structure and diversity of the T cell antigen receptor beta-chain in a teleost fish. *J. Immunol.* 157(1):207-12.

239) Partula, S., de Guerra, A., Fellah, J.S. & Charlemagne, J. (1996). Structure and diversity of the T-cell antigen receptor alpha-chain in a teleost fish. *J. Immunol.* 155:699-706

240) Pelkonen, J., Traunecker, A. & Karjalainen, K. (1987). A new mouse TCR Vγ gene that shows remarkable evolutionary conservation. *EMBO. J.* 6 (7):1941-1944

241) Peltonen, L. & McKusick, V.A. (2001). Dissecting Human Disease in the Postgenomic Era. *Science.* 291:1224-1229

242) Penninger, J.M., Wen, T., Timms, E., Potter, J., Wallace, V.A., Matsuyama, T., Ferrick, D., Sydore, B., Kronenberg, M. & Mak, T.W. (1995). Spontaneous resistance to acute T-cell leukaemias in TCRV gamma 1.1J gamma 4C gamma 4 transgenic mice. *Nature.* 375:241-244

243) Pennisi, E. (2001). What's next for the genome centers? *Science.* 291(5507):1204-7

244) Pereira, P., Gerber, D., Huang, S.Y. & Tonegawa, S. (1995). Ontogenic development and tissue distribution of Vγ1-expressing γ/δ T lymphocytes in normal mice. *J. Exp. Med.* 182:1921-1930

245) Petrie, H.T., Livak, F., Burtrum, D. & Mazel, S. (1995). T cell receptor gene recombinationpatterns and mechanisms: Cell death, rescue, and T cell production. *J. Exp. Med.* 182:121-127

246) Pfeffer, K., Schoel, B., Gulle, H., Kaufmann, S.H.E. & Wagner, H. (1990). Primary responses to human T cells to mycobacteria: A frequent set of γ/δ T cells are stimulated by protease-resistant ligands. *Eur. J. Immunol.* 20:1175-1179

247) Pfeffer, K., Schoel, B., Plesnila, N. Lipford, G.B., Kromer, S., Deusch, K. & Wagner, H. (1992). A lectin-binding, protease-resistant mycobacterial ligand specifically activates Vγ9+ human γδ T cells. *J. Immunol.* 20:1175-117

248) Piontkivska, H., Rooney, A.P. & Nei, M. (2002). Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol. Biol. Evol.* 19(5):689-697

249) Porcelli, S., Brenner, M.B. & Band, H. (1991). Biology of the human γδ T cell receptor. *Immunol. Rev.* 120:137-183

250) Porcelli, S.A. (1995). The CD1 family: A third lineage of antigen-presenting molecules. *Adv. Immunol.* 59:1-98

251) Povinelli, C.M., & Gibbs, R.A. (1993). Large-scale sequencing library production: an adaptor-based strategy. *Anal. Biochem.* 210(1):16-26

252) Rabbitts, T.H., Lefranc, M.-P., Stinson, M.A., Sims, J.E., Shroder, J., Steinmetz, M., Spurr, N.L., Solomon, E. & Goodfellow, P.N. (1985). The chromosomal location of T-cell receptor genes and a T-cell rearranging gene: possible correlation with specific translocations in human T-cell leukemia. *EMBO. J.* 4:1461-1465.

253) Racha, B., Vassalli, P. & GUY-Grand, D. (1994). Thymic and extrathymic origins of gut intraepithelial lymphocyte populations in mice. *J. Exp. Med.* 180:681-86

254) Rast, J.P. & Litman, G.W. (1994). T-cell receptor gene homologs are present in the most primitive jawed veterbrates. *Prov. Natl. Acad. Sci. USA.* 91:9248-52

255) Rast, J.P., Haire, R.N., Litman, R.T. Pross, S. & Litman, G.W. (1995). Identification and characterization of T-cell antigen receptor related genes in phylogenetically diverse vertebrate species. *Immunogenetics.* 42:204-212

256) Rast, J.P., Anderson, M.K., Strong, S.J., Luer, C. & Litman, G.W. (1997). α, β, γ and δ T cell antigen receptor arose early in vertebrate phylogeny. *Immunity.* 6:1-11

257) Rast, J.P. & Litman, G.W. (1998). Towards understanding the evolutionary origins and early diversification of rearranging antigen receptors. *Immunol. Rev.* 166:79-86

258) Raulet, D, Spencer, D., Hsiang, Y.H., Goldman, J., Bix, M., Liao, N.S., Zijlstra, M., Jaenisch, R. & Correa, I. (1991). Control of γδ T cell development. *Immunol. Rev.* 120:185

259) Reardon, C., Lefrancois, L., Farr, A., Kubo, R., O'Brien, R. & Born, W. (1990). Expression of γ/δ T cell receptors on lymphocytes from the lactating mammary gland. *J. Exp. Med.* 172:1263-1266

260) Reitman, M., Lee, E., Westphal, H. & Felsenfeld, G. (1993). An enhancer/locus control region is not sufficient to open chromatin. *Mol. Cell. Biol.* 13:3990-8

261) Richards, M.H. & Nelson, J.L. (2000). The evolution of vertebrate antigen receptors: a phylogenetic approach. *Mol. Biol. Evol.* 17(1):146-55

262) Richmond, I., Pritchard, G.E., Ashcroft, T., Corris, P.A. & Walters, E.H. (1993). Distribution of gamma delta T-cells in the bronchial tree of smokers and non-smokers. *J. Clin. Pathol.* 46(10):926-30

263) Roark, C. E., Vollmer, M., Campbell, P., Born, W. K., & O'Brien, R. L. (1996). Response of a γδ TCR monomorphic subset during bacterial infection. *J. Immunol.* 156:2214-2220

264) Roberts, K., Yokoyama, W.M., Kehn, P.J. & Shevach, E.M. (1991). The vitronectin receptor serves as an accessory molecule for the activation of a subset of γ/δ T cells. *J. Exp. Med.* 173:231-240

265) Robey, E. & Fowlkes, B.J. (1994). Selective events in T cell development. *Annu. Rev. Immunol.*12:675-705

266) Robey, E. & Fowlkes, B. J. (1998). The αβ versus γδ T cell lineage choice. *Curr. Opin. Immunol.* 10:181-187

267) Rock, E.P., Sibbald, P.R., Davis, M.M. & Chien, Y.H. (1994). CDR3 length in antigen-specific immune receptors. *J. Exp. Med.* 179(1):323-8.

268) Rodewald, H.R. & Fehling, H.J. (1998). Molecular and cellular events in early thymocyte development. *Adv. Immunol.* 69:1-112

269) Rodewald, H.R., Brocker, T. & Haller, C. (1999). Developmental dissociation of thymic dendritic cell and thymocyte lineages revealed in growth factor receptor mutant mice. *Proc. Natl. Acad. Sci. USA.* 96:15068-15073

270) Roger, T., Miconnet, I., Bruley-Rosset, M. & Seman, M. (1993). Rearrangement by chromosomal inversion in the T cell receptor gamma locus in a murine alpha beta T cell clone. *Mol. Immunol.* 30(17):1617-1620

271) Roitt, I. (1995). Essential Immunology. 3rd edition. Blackwell Scientific Publication.

272) Rowen, L., Koop, B.F. & Hood, L. (1996). The complete 685-kilobase DNA sequence of the human beta T cell receptor locus. *Science.* 272:1755-1762

273) Rudolph, M.G. & Wilson, I.A. (2002). The specificity of TCR/pMHC interaction. *Curr. Opin. Immunol.* 14(1):52-65.

274) Saito, H., Kranz, D., Takagaki, Y., Hayday, A., Eisen, H. & Tonegawa, S. (1984). A third rearranged and expressed gene in a clone of cytotoxic T lymphocyte. *Nature.* 312(5989):36-40

275) Saito, H., Kanamori, Y., Takemori, T., Nariuchi, H., Kubota, E., Takahashi-Iwanaga, H., Iwanaga, T & Ishikawa, H. (1998). Generation of intestinal T cells from progenitors residing in gut cryptopatches. *Science.* 280(5361):275-8.

276) Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. (1980). Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. *Nature.* 286(5774):676-83.

277) Salomon, B. & Bluestone, J.A. (2001). Complexities of CD28/B7: CTLA-4 costimulatory pathways in autoimmunity and transplantation. *Annu. Rev. Immunol.* 19:225-52

278) Sansom, D.M. (2000). CD28, CTLA-4 and their ligands: who does what and to whom? *Immunology*. 101(2):169-77.

279) Schiffer, M., Wu, T.T. & Kabat, E.A. (1986). Subgroups of V-genes of β-chains of T-cell receptors for antigens. *Proc. Natl. Acad. Sci. USA*. 83:4461-4463

280) Schild, H., Mavaddat, N., Litzenberger, C., Ehrich, E.W., Davis, M.M., Bluestone, J.A., Matis, L., Draper, R.K. & Chien, Y.H. (1994). The nature of major histocompatibility complex recognition by γδ T cells. *Cell*. 76:29-37

281) Schimenti, J.C. (1999). Insights from model systems: Mice and the role of unequal recombination in gene-family evolution. *Am. J. Hum. Genet.* 64:40-45

282) Schoel, B., Sprenger, S. & Kaufmann, S.H.E. (1994). Phosphate is essential for stimulation of Vγ9Vδ2 T lymphocytes by mycrobacterial low molecular weight lignad. *Eur. J. Immunol*. 24:1886-1892

283) Schweighoffer, E. & fowlkes, B.J. (1996). Positive selection is not required for thymic maturation of transgenic gamma delta T cells. *J. Exp. Med*. 183:2033-41

284) Sciammas, R., Johnson, R.M., Sperling, A.I., Brady, W., Linsley, P.S., Spear, P.G., Fitch, F.W. & Bluestone, J.A. (1994). Unique antigen recognition by a herpesvirus-specific TCR-gamma delta cell. *J. Immunol*. 152:5392-5397

285) Sciammas, R., Kodukula, P., Tang, Q., Hendricks, R.L. & Bluestone, J.A. (1997). T cell receptor-gamma/delta cells protect mice from herpes simplex virus type 1-induced lethal encephalitis. *J. Exp. Med*. 185(11):1969-75

286) Scott, A.F., Heath, P., Trusko, S., Boyer, S.H., Prass, W., Goodman, M., Czelusniak, J., Chang, L.Y. & Slightom, J.L. (1984). The sequence of the gorilla fetal globin genes: evidence for multiple gene conversions in human evolution. *Mol. Biol. Evol*. 1(5):371-89

287) Scott, G.K. Davies, P.L., Kao, M.H. & Fletcher, G.L. (1988). Differential Amplification of Antifreeze Protein Genes in the Pleuronectinae. *J. Mol. Evol*. 27:29-35

288) Seemann, G., Rein, R.S., Brown, C.S. & Ploegh, H. (1986). Gene conversion -like mechanisms may generate polymorphism in human class I genes. *EMBO. J.* 5:547-552

289) Segel, L.A & Cohen, I.R. (2001). Design Principles for the Immune System and Other Distributed Autonomous Systems. Oxford University Press.

290) Shortman, K., Wu, L., Kelly, K.A. and Scollay, R. (1991). The beginning and the end of the development of TCR gamma delta cells in the thymus. *Curr. Top. Microbiol. Immunol*. 173:71-80

291) Sim, G.K. & Augustin, A. (1990). Dominantly inherited expression of BID, an invariant undiversified T cell receptor δ chain. *Cell.* 61:397-405

292) Sim, G. K., & Augustin, A. (1991), Extrathymic positive selection of γδ T cells Vγ4Jγ1 rearrangements with GxYS junctions. *J. Immunol.* 146:2439-2445

293) Sim, G.K., Rajaserkar, R., Dessing, M., & Augustin, A. (1994). Homing and in situ differentiation pulmonary lymphocytes. *Int. Immunol.* 6:1287-1295

294) Skeen, M.J. & Ziegler, H.K. (1993). Intercellular interactions and cytokine responsiveness of peritoneal αβ and γδ T cells from Listeria-infected mice: Synergistic effects of interleukin 1 and 7 on γδ T cells. *J. Exp. Med.* 178:985-996

295) Sleckman, B.P., Gorman, J.R. & Alt, F.W. (1996). Accessibility control of antigen-receptor variable-region gene assembly: role of *cis*-acting elements. *Annu. Rev. Immunol.* 14:459-81

296) Sleckman, B.P., Carabana, J., Zhong, X.P. & Krangel, M. S. (2001). Assessing a role for enhancer-blocking activity in gene regulation within the murine T-cell receptor α/δ locus. *Immunology.* 104:11-18

297) Slightom, J.L., Koop, B.F., Xu, P.L. & Gooodman, M. (1988). Rhesus fetal globin genes. Concerted gene evolution in the descent of higher primates. *J. Biol. Chem.* 263:12427-38

298) Smith, G.P., Hood, L, & Fitch, W.M. (1971). Antibody diversity. *Annu. Rev. Biochem.* 40:969-1012

299) Soderstrom, K., Bucht, A., Halapi, E., Gronberg, A., Magnusson, I. & Kiessling, R. (1996). Increased frequency of abnormal gamma delta T cells in blood of patients with inflammatory bowel diseases. *J. Immunol.* 156(6):2331-9

300) Spencer, D. M., Hsiang, Y. H., Goldman, J. P., & Raulet, D. H. (1991). Identification of a T-cell-specific transcriptional enhancer located 3' of Cγ1 in the murine T-cell receptor γ locus. *Proc. Natl. Acad. Sci. USA.* 88:800-804

301) Sperling, A. I. & Wortis, H.H. (1989). CD4-, CD8- γδ cells from normal mice respond to a syneneic B cell lymphoma and can induce its differentiation. *Int. Immunol.* 1:434-442

302) Sperling, A. I., Cron, R. Q., Decker, D.C., Stern, D.A. & Bluestone, J.A. (1992). Peripheral T cell receptor γδ variable gene repertoire maps to the T cell receptor loci and is influenced by positive selection. *J. Immunol.* 149:3200-3207

303) Staudt, L.M. & Lenardo, M.J. (1991). Immunoglobulin gene transcription. *Annu. Rev. Immunol.* 9:373-398

304) Steele, C.R. Oppenheim, D.E. & Hayday, A.C. (2000). γδ T cells: Non-classical ligands for non-classical cells. *Curr. Biol.* 10:R282-R285

305) Strachan, T. & Read A.P. (1996). Human Molecular Genetics. BIOS Scientific Publishers Limited. pp:15

306) Strauss, W.M., Quertermous, T. & Seidman, J.G. (1987). Measuring the human T cell receptor gamma-chain locus. *Science.* 237(4819):1217-9

307) Su, C. & Nei, M. (2001). Evolutionary dynamics of the T-cell receptor VB gene family as inferred from the human and mouse genomic sequences. *Mol. Biol. Evol.* 18(4):503-513

308) Suda, T. & Zlotnik, A. (1993). Origin, differentiation, and repertoire selection of CD3+CD4-CD8- thymocytes bearing either alpha beta or gamma delta T cell receptors. *J. Immunol.* 150(2):447-55

309) Sunaga, S., Maki, K., Komagata, Y., Miyazaki, J. & Ikuta, K. (1997). Developmentally ordered V-J recombination in mouse T cell receptor gamma locus is not perturbed by targeted deletion of the V gamma 4 gene. *J. Immunol.* 158(9):4223-4228

310) Suzuki, K., Oida, T., Hamada, H., Hitotsumatsu, O., Watanabe, M., Hibi, T., Yamamoto, H., Kubota, E., Kaminogawa, S. & Ishikawa, H. (2000). Gut cryptopatches: direct evidence of extrathymic anatomical sites for intestinal T lymphopoiesis. *Immunity.* 13(5):691-702

311) Swofford, D.L. (1997). "PAUP: Phylogenetic Analysis Using Parsimony," Version 4.00d55. Smithsonian Institution, Washington, D.C.

312) Takagaki, Y.,Nakanishi, N., Ishida, I., Kanagawa, O. & Tonegawa, S. (1989). T-cell receptor-γ and –δ genes preferentially utilized by adult thymocytes for surface expression. *J. Immunol.* 142:2112-2121

313) Tanaka, Y., Morit, C.T., Tanaka, Y., Nieves, E., Brenner, M.B. & Bloom, B.R. (1995). Natural and synthetic non-peptide antigens recognized by human γδ T cells. *Nature.* 375:155-158

314) The International SNP Map Working Group. (2001). *Nature.* 409:928-933

315) Thomsen, P.D., & Miller, J.R. (1996). Pig genome analysis: differential distribution of SINE and LINE sequences is less pronounced than in the human and mouse genomes. *Mamm Genome.* 7(1):42-46

316) Tigelaar, R.E. & Lewis, J.M. (1995). Immunobiology of mouse dendritic epidermal T cells: a decade later, some answers, but still more questions. *J Invest Dermatol.* 105:43s-49s

317) Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature.* 302:575-581

318) Traunecker, A., Oliveri, F., Allen, N. & Karjalainen, K. (1986). Normal T cell specific rearrangement of T-cell receptor β transgenes in mice. *EMBO. J.* 5:1589-1593

319) Tribel, F. LeFranc, M.P. & Hercend, T. (1988). Further evidence for a sequentially ordered activation of T cell rearranging gamma genes during T lymphocyte differentiation. *Eur J Immunol.* 18:789-94

320) Tsuji, M., Mombaerts, P., Lefrancois, L., Nussenzweig, R.S., Zavala, F. & Tonegawa, S. (1994). Gamma delta T cells contribute to immunity against the liver stages of malaria in alpha beta T-cell-deficient mice. *Proc. Natl. Acad. Sci. USA.* 91(1):345-9

321) Tsuji, M., Mombaerts, P., Lefrancois, L., Nussenzweig, R.S., Zavala, F. & Tonegawa, S. (1996). Phenotypic and functional properties of murine γδ T cells clones derived from malaria immunized αβ T cell deficient mice. *Int. Immunol.* 8:359-366

322) Ullrich, R., Schieferdecker, H.L., Ziegler, K., Riecken, E.O. & Zeitz, M. (1990). gamma delta T cells in the human intestine express surface markers of activation and are preferentially located in the epithelium. *Cell. Immunol.* 28(2):619-27

323) van der Heyde, H.C., Manning, D.D., Roopenian, D.C. & Weidanz, W.P. (1993). Resolution of blood-stage malarial infections in CD8+ cell-deficient beta 2-m0/0 mice. *J. Immunol.* 151(6):3187-91

324) Venter, J.C. & et al. (2001). The sequence of the human genome. *Science.* 291(5507):1304-51

325) Vernooij, B.T.M., Lenstra, J. A., Wang, K. & Hood, L. (1993). Organization of murine T- cell receptor γ locus. *Genomics.* 17:566-574

326) Vila, L.M., Haftel, H.M., Park, H.S., Lin, M.S., Romzek, N.C., Hanash, S.M. & Holoshitz, J. (1995). Expansion of mycobacterium-reactive gamma delta T cells by a subset of memory helper T cells. *Infect. Immun.* 63(4):1211-7

327) Wagner, A. (2001). Birth and death of duplicated genes in completely sequenced eukaryotes. *Trends in Genetics.* 17(5):238-239

328) Wang, L.S., Kamath, A., Das, H., Li, L. & Bukowski, J.F. (2001). Antibacterial effect of human Vγ2Vδ2 T cells in vivo. *J. Clin. Invest.* 108: 1349-1357

329) Waterston, R.H., Lander, E.S. & Sulston, J.E. (2002). On the sequencing of the human genome. *Proc. Natl. Acad. Sci. USA*. 99(6):3712-6

330) Watts, C. (1997). Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu. Rev. Immunol*. 15:821-50

331) Weaver, D.T. (1995). V(D)J recombination and double-strand break repair. *Adv. Immunol*. 58:29-85

332) Weis-Garcia, F., Besmer, E., Sawchuk, D.J., Yu, W., Hu, Y., Cassard, S., Nussenzweig, M.C. & Cortes, P. (1997). V(D)J recombination: in vitro coding joint formation. *Mol. Cell. Biol*. 17(11):6379-6385

333) Weiss, A., Manger, B. & Imboden, J. (1986a). Synergy between the T3/antigen receptor complex and Tp44 in the activation of human T cells. *J. Immunol*. 137:819-25.

334) Weiss, A., Newton, M. & Crommie, D. (1986b). Expression of T3 in association with a molecule distinct from the T-cell antigen receptor heterodimer. *Proc. Natl. Acad. Sci. USA*. 83: 6998-7002

335) Welch, J.W., Maloney, D.H. & Fogel, S. (1990). Unequal crossing-over and gene conversion at the amplified CUP1 locus of yeast. *Mol. Gen. Genet*. 222(2-3):304-10.

336) Wells, F. B., Gahm, S.J., Hedrick, S.M., Bluestone, J.A., Dent, A. & Matis, L.A. (1991). Requirement for positive selection of γδ receptor bearing T cells. *Science*. 253: 903

337) Wen, L., Roberts, S.J., Viney, J.L., Wong, F.S., Mallick, C., Findly, R.C., Peng, Q., Craft, J.E., Owen, M.J. & Hayday, A.C. (1994). Immunoglobulin synthesis and generalized autoimmunity in mice congenitally deficient in αβ (+) T cells. *Nature*. 369:654-658

338) Willerford, D.M., Swat, W. & Alt, F.W. (1996). Developmental regulation of V(D)J recombination and lymphocyte differentiation. *Curr. Opin. Genet. Dev*. 6(5):603-9

339) Wilson, C.B., Makar, K.W. &Perez-Melgosa, M. (2002). Epigenetic regulation of T cell fate and function. *J. Infect. Dis*. 185 Suppl 1:S37-45

340) Wilson, E., Aydintug, M.K. & Jutila, M.A. (1999). A circulating bovine gamma delta T cell subset, which is found in large numbers in the spleen, accumulates inefficiently in an artificial site of inflammation: correlation with lack of expression of E-selectin ligands and L-selectin. *J. Immunol*.162(8):4914-9

341) Winoto, A. & Baltimore, D. (1989a). Alpha beta lineage-specific expression of the alpha T cell receptor gene by nearby silencer. *Cell*. 59:649-655

342) Winoto, A. & Baltimore, D. (1989b). A novel, inducible and T cell-specific enhancer located at the 3' end of the T cell receptor α locus. *EMBO. J.* 8 (3): 729-733

343) Xiong, N., Kang, C. & Raulet, D.H. (2002). Redundant and Unique Roles of Two Enhancer Elements in the TCRγ Locus in Gene Regulation and γδ T Cell Development. *Immunity*. 16:453-463

344) Xu, J. L. & Davis, M.M. (2000). Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity*. 13:37-45

345) Yanagi, Y., Yoshiki, Y., Leggett, K., Clark, S.P., Aleksander, I. & Mak, T.W. (1984). A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature*. 308:145

346) Ye, S.K., Agata, Y., Lee, H.C., Kurooka, H., Kitamura, T., Shimizu, A., Honjo, T. & Ikuta, K. (2001). The IL-7 receptor controls the accessibility of the TCRgamma locus by Stat5 and histone acetylation. *Immunity*. 15(5):813-23

347) Zapata, A.G. & Cooper, E.L.eds (1990). *The immune System: Comparative Histophysiology*. Chichester, UK: Wiley. 335pp

348) Zhang, Z.X., Young, K. & Zhang, L. (2001). CD3+CD4-CD8- alphabeta-TCR+ T cell as immune regulatory cell. *J. Mol. Med.* 79(8):419-27

349) Zhong, X.P. & Krangel, M.S. (1997). An enhancer-blocking element between alpha and and delta gene segments within the human T cell receptor alpha /delta locus. *Proc. Natl. Acad. Sci. USA*. 94:5219-5224

350) Zijlstra, M., Bix, M., Simister, N.E., Loring, J.M., Raulet, D.H. & Jaenisch, R. (1990). β2-microglobulin deficient mice lack CD4-8+ cytolytic T cells. *Nature*. 344:742-746

351) Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan, Y.W. & Wilson, A.C. (1980). Rapid duplication and loss of genes coding for the alpha chains of hemoglobin. *Proc. Natl. Acad. Sci. USA*. 77(4):2158-62

Appendix: List of the contributors to this project

1.The characterization of the human TCR γ locus:

Dr. Ming Zhan:

Sequenced and assembled three quarters of the sequence; the major contributor to the sequence analysis of the human TCR γ locus (Linda McKinnell was responsible for the actual DNA sequencing).

Dr. Ben F Koop:

One of the major contributors of the phylogenetic analysis of the human TCR γ sequence.

Qun Zhou:

Sequenced and assembled one quarter (35kb) of the locus, which include two lambda clones, λSH3 and λSH9, and a long PCR fragment; contributed part of the sequence analysis in this thesis; (Linda McKinnell obtained the long PCR fragments and she also helped with running a few gels on the long PCR fragment).

2. The characterization of the mouse TCR γ locus:

Michael Parlee:

Sequenced and assembled 68 kb of the mouse sequence (Linda McKinnell was responsible for the sequencing of the 33 kb cosmid M γ 84).

Qun Zhou:

Sequenced and assembled 106kb of the mouse TCR γ locus; carried out the sequence analysis of the mouse locus; (Ute Rink helped with running a few gels on the sequencing of cosmid 93B).

3. The characterization of the enhancer γ E4:

Qun Zhou:

Carried out both the experiments and data analysis.

VITA

Surname: Zhou Given Names: Qun

Place of Birth: Jinzhou, Liaoning, China

Educational Institutions Attended:

University of Victoria	1994 to 2003
The Second Military Medical University	1983 to 1989

Degrees Awarded:

Bachelor of Medicine	The Second Military Medical University	1989

Publications:

1. Bridge, M.A., Zhou, Q., Koop, B.F. & Pearson, T.W. (1998). Cloning and characterization of the kinetoplastid membrane protein-11 gene locus of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 91(2):359-63

2. Moreau, E., Langerak, A., Gastel-Mol, E., Wolvers-Tettero, I., Zhan, M., Zhou, Q., Koop, B. & Dongen, J. (1999). Easy detection of all T cell receptor gamma (TCRG) gene rearrangements by southern blot analysis: recommendations for optimal results. *Leukemia*. 13(10):1620-6

Partial copyright license

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying of thesis thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of thesis:

Characterization of the Human and Mouse T Cell Receptor Gamma Loci

Author 

Qun Zhou

August 2003