Chapter One - Introduction

1.1 Introduction

In the classic biochemistry textbook by Albert Lehninger a passage in the introduction explains how the laws of chemistry and physics fully apply to the field of biology and yet how another set of laws, those of the living state as seen through the field of biochemistry, also apply (Lehninger 1970). Here, in the same sense, the laws of immunology and genetics, via immunity within the individual and genetic evolution within the species, both apply to the genetics of susceptibility to autoimmune and infectious diseases in higher vertebrates. In the interplay between individual genetic variation of immune response genes and changing environmental effects there gives rise a higher set of laws that apply to the complex genetics of vertebrate immune evolution. On a broad scale, the effects of the laws of vertebrate immune evolution are seen in the organization and structure of a number of genetic loci throughout the vertebrates. Good examples of such organization include the clustering of the Major Histocompatibility Complex (MHC) class I and II regions in vertebrates arising after the bony fishes (Delarbre *et al.* 1992; Jakobsen *et al.* 1998; Kaufman 1999; Klein and Figueroa 1986; Klein *et al.* 1993a; Klein *et al.* 1993b; Sammut *et al.* 1999; Sato *et al.* 2000), in the comparative genomics of the T-lymphocyte antigen receptor (TcR) gene loci in various vertebrates (Bontrop *et al.* 1995; Hawke *et al.* 1999; Jaeger *et al.* 1998; Lai *et al.* 1988; Nam *et al.* 2003), and in the variation of expression of various chemokines and of chemokine receptors through vertebrate evolution and within populations of different vertebrates (Blanpain *et al.* 2000; Yang 2000; Zhang *et al.* 1999). Indeed, the effects of immune genomic evolution in the broadest sense may act to limit the overall numbers of genes in the various vertebrate genomes where these rules apply to greater and lesser degrees (Buhler 2001; George 2002). On a more narrow front, the manner and

mechanism of autoimmune diseases such as type 1 diabetes (T1D), multiple sclerosis (MS), rheumatoid arthritis (RA), systemic lupus erythrematosus (SLE) and others exhibit complex genetic associations, and it is in these model complex diseases that finding some sort of rule which applies to them may help to solve the genetic puzzles of these diseases. Many of the autoimmune diseases share a set of "core" autoimmune disease susceptibility genes (Becker 1999; Becker *et al.* 1998; Bergsteinsdottir *et al.* 2000; Encinas and Kuchroo 2000; Jawaheer *et al.* 2001; Merriman *et al.* 2001), but each is also likely to have unique associations with genes that can have a specific role in the pathogenesis of a given autoimmune disease (Wucherpfennig *et al.* 1997). Furthermore, there is the possibility of a given disease model having triggers or other environmental factors which may be also involved (Winer *et al.* 2001). It is in some of these diseases that I will here examine the question of genetic susceptibility with regard to alleles of two genes that are a part of the antigen recognition and cellular migration components of T-cells; the TCRB variable segments and the CCR5 receptor for chemokines.

1.2 The MHC and the nature of the peptide being presented

Looking back on the study of the major histocompatibility complex, the article by Klein, Figueroa and Nagy in the first issue of the Annual Review of Immunology is subtitled "The Final Act". They suggest that while a few important questions still remained, molecular genetics had all but solved the basic biological questions about the MHC (Klein *et al.* 1983). After nearly two decades, a few of those unanswered questions are still being extensively studied and many more have come up along the way (Kostyu *et al.* 1997; McDevitt 2000). On the structure of the MHC gene region in the evolution of vertebrates, there is clustering of class I and class II in higher vertebrates as opposed to

the dispersed nature seen in the teleost fishes (Sato *et al.* 2000) which needs to be explained (Kumnovics *et al.* 2003), as does the strong linkage disequilibrium seen in the MHC. The evolutionary stability of ancestral MHC haplotypes within the 3.6 Mbp region is well known (Beck and Trowsdale 2000; Sammut *et al.* 1999) and, interestingly, the recent initial analysis of the completed human genome project showed that encoded within the MHC is nearly a complete set of transfer-RNA genes (Lander *et al.* 2001) that perhaps adds some weight to maintaining genetic stability over part of this region. The complete sequence data of the 3.6 Mbp that encodes the MHC has been reviewed by Beck and Trowsdale and they discuss the clustering of immune response genes and the nature of MHC polymorphism with respect to the insights gained from the physical sequence (Beck and Trowsdale 2000). As they point out, while it occupies just 0.12% of the human genome, the MHC is associated with more diseases, predominately autoimmune diseases as mentioned above, than any other region of the genome.

The peptide fragment motifs determined from eluted fragments displayed by MHC class I and class II are quite different, as are the source proteins from which peptides are derived (Yeager *et al.* 2000). The biochemical basis of cellular life predates the relatively recent stage of immune evolution by an enormous extent, with a likely window of high temperature in the fixture of the DNA code a result of impact events that evaporated most of the oceans and possibly even some global sterilization events where cellular life that was able to withstand ejection into orbit and reintroduction to the planet (Nisbet and Sleep 2001), so a large number of biochemical pathways are essentially fixed in their place in early evolution. It is these essential components that

contribute to the class I MHC peptide fragment pool, rather than the more random peptide sets obtained from lysozyme degradation of internalized peptide as happens in the loading of peptide into class II MHC for presentation.

The genomic evolution of the vertebrate immune system has had to balance selective forces from a number of directions including resistance to microbial and viral pathogens, parasites, cancers, and autoimmune diseases. Without doubt, the direct protection from microbes and viruses is paramount and lack of any of a number of key immune system components can have less than desirable outcomes (e.g. death). Variations or alterations to the level of expressed self-antigen in any given type of cell expressing class I MHC may occur by mutation or by a change in the type of antigen, as might occur in tumour development or with virus infection, and this then defines a trigger to immune surveillance against such "altered self" changes. While control over altered self in protection from some viruses or tumours and an avoidance of autoimmunity are both desirable, they do not have the absolute life and death immediacy of protection from most pathogens. Immune evolution must, however, allow enough of the population who carry a specific allele to survive past reproduction to maintain the genetic fitness that an allele of a multi-component system needs to have to survive in a genetic sense.

The key survival feature in vertebrates is the use of a self-adaptive immune system to engage a wide variety of pathogens. Key components of the vertebrate immune system include the expressed repertoires of antigen receptors for T-cells and the antibodies produced by B-cells, the antigen presenting molecules, various cytokines, chemokines

and related messengers, as well as the complement and apoptosis related pathways. A number of pathogens have used as a part of their means for survival some component or another of the immune system and as such a range of differences exist between individuals in vertebrate populations where either allelic differences or expression level differences for various immune system molecules can be found. Unfortunately, while such differences between individuals may be good across the entire population, a side effect is that specific individuals may be prone to the occurrence of autoimmunity. The potential for autoimmune disease is present throughout vertebrates and autoimmune disease is inducible in most individuals within a species. While each type of autoimmune disease has a target antigen or range of antigens, mechanisms by which those antigens become functional targets of inflammation may also be seen to vary and these include molecular mimicry, superantigenic stimulation, lack of adequate thymic deletion, breakdown of tolerance, and exposure of previously sequestered antigen (Piyasirisilp *et al.* 1999).

1.3 Biology of the TCRB locus

The ability of any vertebrate to produce unique antibodies to a given antigen or to have T-cell control over specific immune responses lies in the nature of the seven gene loci needed to generate such antibodies or T-cell receptors, the heavy, kappa and lambda chains of immunoglobulins and the alpha, beta, gamma and delta chains of the antigen receptors of T-cells. Joining of gene segments, in the form of a small number of "constant" region segments and a much larger number of "variable" gene segments is the hallmark of these immune recognition gene systems. Indeed, evolution has demanded a system whereby binding pockets to never-before-seen epitopes can be

generated with ease.

The predominant form of the T-cell antigen receptor in primates is the αβ-TcR, a cell surface heterodimeric molecule consisting of an alpha and a beta chain. The alpha and beta TcR chains are both members of the immunoglobulin-gene superfamily (Davis and Bjorkman 1988) and the ligand for TcR is a complex of a specific short peptide antigen bound noncovalently in a pocket or groove of an MHC class I or class II molecule. The recognition of antigen complexed with MHC by the TcR is the first step in the basic mechanism of antigen-specific activation of T-cells (Zinkernagel and Doherty 1974). In humans, the beta chain of TcR is encoded on chromosome 7 in a 685 kb region which contains, along with the constant (C), joining (J) and diversity (D) segments, 65 variable (V) segment genes of which 46 are functional and 19 are pseudogenes and which can be defined into 30 VB subfamilies based on 75% homology (Rowen *et al.* 1996). In the mouse, by comparison, the TCRB region covers 701 kb of chromosome 6 with 35 V-segments (21 functional and 14 pseudogenes) in 31 subfamilies (Su and Nei 2001). In addition, there are another 26 "relic" V-gene segments in the human TCRB region but these are not seen as such in the mouse. Interestingly, in both the mouse and the human TCRB regions 40% of the V-segments have lost their function (to become pseudogenes). A number of human V-segments are the result of duplication of a 20 kb block that is found to be tandomly repeated 5 times, and which accounts for 15% of the total human TCRB region (Su and Nei 2001). Age estimates for these duplications shows events at over 19 million years ago and at over 35 million years ago and while a number of V-segment family members in humans overall occurred after the mouse-human split in evolution (about 100 million years ago), there are others which date back

much farther as they have been identified in birds or in fish as well as mammals (Su and Nei 2001), suggesting TCRB segment evolution by "birth and death" rather than by concerted evolution. Within the primates, duplications and deletions of various TCRB gene segments has continued (Charmley *et al.* 1995), however specific polymorphisms found in human TCRB are not seen in the chimpanzee and are thus not maintained to the degree the MHC is by selection over the 5 million year span between these species (Jaeger *et al.* 1998).

The study of immune genetics of autoimmune disease in the first instance involved the genes in the MHC for antigen presentation, along with those for antigen recognition by T-cells via the tri-molecular complex (Arden *et al.* 1995; Hillert and Olerup 1992; Kay 1996; Matis 1990). While many other systems involved in immune function have a role to play in various autoimmune diseases, it is the MHC that acts in both helping to define the repertoire of T-cells available in the periphery (Correia-Neves *et al.* 2001; Sebzda *et al.* 1999; von Boehmer and Fehling 1997) and in actively presenting antigen for possible immune recognition (Gautam *et al.* 1994; Hausmann and Wucherpfennig 1997). The essential nature of randomness generated by the nucleotide insertion that goes along with the "V-D-J" joining in the TcR-beta chain is also a feature that may mask the selection of these gene segments (Jaeger *et al.* 1998), and would also mask their role in autoimmune disease especially when the attempt to study these segments is made through genome-screen methods. The binding pocket is a structure built randomally through site-specific recombination and insertion of nucleotides and so is not encoded specifically in the genome. The key role of the TcR in defining of self-repertoire and initiation of an immune response demands that it be considered as a

strong candidate gene in autoimmune disease studies even though the chromosome region that encodes it is not found to be strongly associated in genome screens (Concannon *et al.* 1998; Ebers *et al.* 1996; Jawaheer *et al.* 2001; Kuokkanen *et al.* 1997; Sawcer *et al.* 1996).

While deletion of autoreactive lymphocytes is a primary role of thymic education, it is clear that for several reasons a significant number of T-cells arise which have the potential to react against self and which have not been deleted (Theofilopoulos *et al.* 2001; Wilson *et al.* 2000; Yan and Mamula 2002). In these cases, it is important that the state of tolerance is maintained through a means other than thymic deletion. In some cases, sequestered antigen not readily available for use in "education" of the immune repertoire is involved (Wilson *et al.* 2000). In other cases, the avidity of the TcR for the bound antigen is not strong enough to trigger deletion. Recently, the question of how "self" is determined, both at the point in thymic education where it matters for repertoire outcome and in the periphery where the triggering of autoimmunity is likely to take place, has had a new dimension added which has just made the "black box" of the thymus quite a bit blacker (Sercarz 2002), with a potential role for endopeptidases in determining the way in which antigen is allowed to unfold and thus which parts of the antigen may be used by the MHC for thymic deletion of autoreactive cells (Anderton *et al.* 2002; Manoury *et al.* 2002).

1.4 Chemokine receptors and the immune response

The ability of immune-competent cells to enter different tissues, both within the lymphoid organs (Dunon and Imhof 2000; Luther and Cyster 2001) and the organs and tissues of the body in general, is determined by gradients of chemokine and by the density of the chemokine receptors on cells (Moser and Loetscher 2001; Penna *et al.* 2002a; Penna *et al.* 2002b). Lymphocyte migration is essential to immune function and it is now clear that a large part of the spectrum of specific lymphocyte migration, from entry of the T-cell precursors into the thymus (Norment and Bevan 2000) to migration of effector and memory cells on pathways different to those of naive, unexposed T-cells (Sallusto *et al.* 1999a), to determination of the outcome of the immune response (Dunon and Imhof 2000), are a result of chemokine ligand / receptor interactions (Yoshie 2000). Multiple chemokines trigger some receptors (e.g. CCR5) and multiple receptors on various cell populations have a number of roles in lymphoid organ development and in lymphopoiesis (Ansel and Cyster 2001). The current understanding of these systems are not complete. Delineation of some of the pathways by which different cytokines work has shown that the role of these receptors may differ slightly however, under different conditions, as can be seen in a basic model of Th1/Th2 cells in autoimmune disease (Louzoun *et al.* 2001).

Physiological barriers against the spread of an infection between body tissues must be navigated by immunocompetent cells filling a variety of roles involved in antigen presentation and immune recognition. In the systemic-cytokine to local-chemokine cascade described in a murine cytomegalovirus (MCMV) model, liver protection by macrophage inflammatory protein-1α (MIP-1α) mediated natural killer (NK) cell infiltration is critical to survival (Salazar-Mather *et al.* 2000) and also important in

initiating a further downstream chemokine cascade. As a chemokine, MIP-1α (CCL3) is one of the ligands that can engage the CCR5 chemokine receptor, along with MIP-1β (CCL4) and RANTES (regulated on activation normal T cell expressed and secreted). CCR5 is one of the chemokine receptors encoded in the 3p21 region of human chromosome 3 (Samson *et al.* 1996c), and is a key co-receptor for infection by macrophage-tropic strains of the human immunodeficiency virus (HIV) whereas CXCR4 acts as the major co-receptor for T-cell tropic HIV strains (Alkhatib *et al.* 1996; Dragic 2001; Dragic *et al.* 1996). As receptors, CCR5 and CXCR4 are members of the largest receptor superfamily in the human genome, the seven transmembrane domain G protein-coupled receptor family. The ligands for CXCR4 were not known at the time that it was shown to be involved in the pathogenesis of HIV infection (Berger *et al.* 1999). CXCR4 was initially regarded as that of an orphan receptor as no ligand or functional activity had been found while the specificity of CCR5 for the ligands MIP-1α, MIP-1β and RANTES was shown (Samson *et al.* 1996a) to fit the pattern for the presumed additional (macrophage-tropic strain) receptor (Alkhatib *et al.* 1996; Deng *et al.* 1996; Dragic *et al.* 1996).

The need to study the pathogenesis of HIV/AIDS in search of possible therapeutic agents was in fact an impetus in unravelling the role of chemokines and their receptors in human biology (Berger *et al.* 1999). Even in HIV disease, the role of chemokines and their receptors is much broader than just that of viral co-receptor. Much of the cellular migration of immune competent cells is driven by gradients of various chemokines in both the developmental migration of cells in immune cell maturation and in the movement of activated T-cells out of the blood vessels and into various tissues (Kurt *et*

al. 2001; Ploix *et al.* 2001; Sallusto *et al.* 1999a; Sallusto *et al.* 1999b; Vanbervliet *et al.* 2002; Weber *et al.* 2001). In the setting of specific immune activation, the recognition of antigen in the context of the major histocompatibility complex (MHC) class I and II molecules by the T-cell receptor for antigen (TcR) also has dynamics that can be affected by gradients of chemokine and cellular density of chemokine receptors (Bromley *et al.* 2000). It is within the recognition of antigen by the tri-molecular complex formed by the TcR and antigen-MHC interface that the cellular basis of self discrimination in vertebrate immunity is to be found, with both the elimination through thymic education of autoreactive T-cells and establishment of the peripheral T-cell repertoire based on the characteristics of the MHC and other expressed self antigens of an individual (Goldrath and Bevan 1999; Sebzda *et al.* 1999; Viret *et al.* 2001). An important basis for initiation of autoimmune disease is the potential of autoreactive T-cells which escape thymic deletion to become inadvertently activated to a self epitope seen as a mimic of the target antigen (Olson *et al.* 2002; van Noort *et al.* 2000; Wucherpfennig 1994). Alongside the potential for activation of an auto-antigen reactive T-cell is the antigen-nonspecific but receptor subfamily-specific effect of crosslinking TcR with MHC by various superantigens (Jardetzky *et al.* 1994). How the formation of the "tri-molecular complex" of antigen, MHC and the T-cell receptor decides the fate of an immune response is further determined by both the density of antigen presented to the T-cell (with a certain minimum required for activation) and by the duration of cellular interaction allowed between given cells. In addition, there is an important role for binding affinity during antigen presentation in the periphery as well as during the education of thymocytes (Love *et al.* 2000). If there is a strong chemokine gradient the migration of a given T-cell may be such that it effectively goes past the antigen-

presenting cell before a critical number of molecules have formed their complexes (Bromley *et al.* 2000). Additionally the chemokine receptor density on a migrating cell may help or hinder the entry of that cell into a specific tissue, thus allowing for or possibly helping to prevent an autoimmune-type reaction (Moser and Loetscher 2001).

1.5 Immune polymorphism, self polymorphism and effects on repertoire

Genetic polymorphisms can affect the function of a gene in question in several ways, including not affecting it at all. If the polymorphism is in the coding region of a gene and the nature of the polymorphism changes the sequence of amino acids in the peptide, then it is possible that function can be affected directly. On the other hand, there are many "silent mutations" in the coding regions which do not change the amino acid sequence. They can exist in linkage disequilibrium with other nearby mutations and thus can be very informative. Mutation outside of the coding region can have an effect on gene expression if the promoter region is affected by the mutation (Holloway *et al.* 2002; Kleinjan and van Heyningen 1998; Soutto *et al.* 2002). Forces in nature can then act upon regions of the genome, in terms of natural selection through survival, where various types of "bottlenecks" reduce the frequency of some alleles and increase the frequency of others (Leung *et al.* 2000).

The concept of "polymorphism" in biology took on a new meaning with the early description of variation in blood group antigens that are now known as the class I and class II Major Histocompatibility Complex (MHC) genes. As more and more sera were described that could lyse cells from different individuals, it became clear that the genes involved in these loci (Payne *et al.* 1977) were unique in the degree to which they were

polymorphic and interest in this field was further driven by the clinical need to understand the transplantation barrier (Bach 1976; Hess 1976), hence the term "histocompatibility antigen". By 1980, it was becoming evident that through their allelic nature the HLA-A, B, C and D gene products were directly involved in the genetic regulation of the immune response through differences in T-cell responses (Charron *et al.* 1980; Charron and McDevitt 1980; McDevitt 1980; McDevitt 1984). Importantly, the alleles being described at these loci also began to show a number of disease associations (McDevitt and Bodmer 1974; Ryder *et al.* 1981), especially within the "autoimmune" type diseases (Svejgaard 1979a). That autoimmune disease often has a genetic basis is in part understood because of the number of these diseases, such as multiple sclerosis (MS) (Batchelor *et al.* 1978; Stewart *et al.* 1977; Stewart *et al.* 1981), rheumatoid arthritis (RA) (Paget and Gibofsky 1979) and type 1 diabetes (IDDM) (Cudworth *et al.* 1979), which have associations with an allele (or with multiple alleles) of MHC genes. The combined interest between the transplantation and the autoimmune disease fields in MHC genes and their products saw the number of citations in this area rise from 125 in total for the years 1975 to 1977 to over six hundred citations a year in 1981. (The current citation rate is over two thousand a year.) The focus within the MHC gene context in this thesis is in the HLA-DR15 stratification of the TCRBV data presented in Chapter 3, and in the essential role that MHC has in the thymic education and establishment of the T-cell repertoire and in antigen recognition (Doherty and Zinkernagel 1975).

There are significant pressures in selection and population fitness that require a large

number of MHC alleles to exist, albeit with only a dozen or so ever able to be expressed in an individual without adverse effects on T-cell repertoire (Nowak *et al.* 1992). In the biology of the immune response, there are some key attributes that must be considered along with the role of the MHC. The genes for the antigen-receptors on T-cells are themselves polymorphic (Arden *et al.* 1995; Wei *et al.* 1994) although, as will be described below, in a much different way than the MHC, and furthermore there is a role for variation in the expression of receptors for chemokines that can also play a role in the determination of the outcome of an immune response (Bromley *et al.* 2000). In simple terms, the MHC-TcR interaction defines the self-repertoire and allows for a molecular basis to the immune response (Doherty and Zinkernagel 1975) and the effect of chemokine receptor density and gradients of the different chemokines in various immune settings helps to determine the potential outcome of the MHC-TcR interaction (or if one is to occur at all). Differences in chemokine receptor density and the degree of chemokine gradients also play a role in the ability of immunocompetent cells to exit blood vessels and cross tissue barriers.

Self-reactive T-cells abound in humans, as in vertebrates in general, and the anergic state, the state of being unable to respond to specific immune activation signals, is what keeps these potentially autoaggressive cells from reacting against self. It is in these anergic cells that the trouble may lie in terms of environmental factors which can act to break tolerance and allow the previously anergic response to become an active anti-self one. Other problems with the deletion of anti-self clones in the period of thymic education is the lack of expression in thymic tissue of antigen that is found in sequestered tissue (such as CNS antigen in brain) or in very small amounts (such as

islet-cell antigen from the pancreatic tissues), and with changes in antigen expression over time such as through puberty or during pregnancy. Specific thymic expression of CNS antigen has been described and may be one way that sufficient self-antigen is made available from otherwise sequestered antigen for thymic education (Pribyl *et al.* 1996a; Pribyl *et al.* 1993; Pribyl *et al.* 1996b). Splenic expression of CNS antigen has also been reported (Campagnoni *et al.* 1993), which could provide a means for continuing peripheral tolerance to be supported. Diabetes associated antigens have also been shown to have specific thymic expression that is involved in ongoing regulation of autoreactivity (Pugliese *et al.* 2001).

The theme to be explored in this thesis comes from examining a set of beta-chain polymorphisms of the T-cell receptor for antigen in MS and also the question of the origin (and possible selection for) the CCR5 deletion mutation *CCR5-Δ32* is that components of the immune system have, in addition to the usual degree of polymorphic variation found in any peptide, a unique range of polymorphism requirements that allow for individual variation of the immune response between individuals within populations. Without this sort of individual variation, a single pathogen would be able to wipe out an entire population (Klein and Figueroa 1986). The dynamics of the allelic polymorphisms of the MHC class I and II genes (above) within and between populations is a key example of this (Jakobsen *et al.* 1998), but not uniquely so. The role of variable segments of the antigen receptor for T-cells in determination of an (individual) immune repertoire also shows a role for polymorphism that, while completely unlike the MHC for it's degree of variation, still allows for population-based differences that give some individuals a better chance for survival . Likewise, in the role

that the chemokines and their receptors have in the immune response there are a number of polymorphisms that go beyond a simple allelic change in a molecule. The role for selection of TCR polymorphisms in the evolution of primates may also be affected by the functional nature of the receptor being constructed independently of the selection of the variable segment (Jaeger *et al.* 1998) with chimpanzee sharing very few polymorphisms with human alleles, while MHC lineages easily predate the estimated five million year gap between these species.

The entire range of immune molecules have polymorphisms that could be of interest in autoimmune disease and in avoidance of infectious disease; a brief survey of recent literature gives examples for immune system components such as interferon-γ and IL-12 (Dorman and Holland 2000; Ozenci *et al.* 2001), IL-12B (Morahan *et al.* 2001), IL-2 (Zelus *et al.* 2000), IL-4 (He *et al.* 1998; Kukreja *et al.* 2002; Noguchi *et al.* 2001), TNF-α (Hajeer and Hutchinson 2001; Skoog *et al.* 1999) along with TNF-β (Khani-Hanjani *et al.* 2000) and TNF-ζ (Isomaki *et al.* 2001), CD22 (Hatta *et al.* 1999), CD45 (Ballingall *et al.* 2001), the common γ-chain (Schmalstieg and Goldman 2002), IgE (Xu *et al.* 2000), TAP1 (Tang *et al.* 2001) and CTLA-4 (Hellings *et al.* 2002; Kristiansen *et al.* 2000) among others. Some immune system components also have homologous genes that have become a part of the genomic make-up of a pathogen, such as IL-10 related sequence in the Epstein-Barr viral genome (Moore *et al.* 1993). In addition to viral homology to IL-10, susceptibility to EBV is related to the level of expression of IL-10 (Helminen *et al.* 1999; Helminen *et al.* 2001). In the case of B-cell responses to interleukin 6 (IL-6), the human herpesvirus-8 viral homologue of IL-6 is actually able to stimulate B-cells that are unresponsive to the human IL-6, possibly adding to B-cell

dysregulation in combination with EBV or HIV and thus possible increasing the likelihood of B-cell malignancies (Breen *et al.* 2001). Cytomegalovirus has a number homologs of G protein-coupled receptor (chemokine receptor) and chemokine ligand genes encoded in it's genome in a bid to evade the immune system (Vink *et al.* 2001). The SIV and HIV genomes show evidence that their common ancestor had an *env* gene which contains sequence that codes for ligand to chemokine receptors, hence the ability to use those receptors in viral tropism (Lusso 2000; Shimizu and Gojobori 2000). There are also important aspects of regulatory gene sequence polymorphisms (Schulte 2001) that especially in immune system components (Mitchison 2000) have roles to play in the overall outcome of the immune response. In the balance that is struck between human cytomegalovirus (HCMV) and the host, HCMV may have the ability to downregulate the expression of the chemokine RANTES by vascular endothelial cells as a mechanism to evade the local immune responses to viral infection (Billstrom-Schroeder and Worthen 2001). Expression levels of RANTES are subject to promoter region polymorphisms where the mutation causing increased expression, found in 17% of a Japanese cohort (Liu *et al.* 1999), is seen to delay the progression of infection by HIV. Recently a review of viral immunological escape mechanisms has called the ability of viruses to subvert a component of the immune system "a masterpiece of evolution" (Vossen *et al.* 2002).

Polymorphism is of course not unique to the immune system and some variation will always be seen throughout the genome. An example is the 9.7kb region of the human lipoprotein lipase gene reported by Nickerson and colleagues (Clark *et al.* 1998; Nickerson *et al.* 1998) where 142 chromosomes that were sequenced from three

populations for this region were seen to have 88 variations (with most of these not in the coding regions). One implication of the different roles that immune related polymorphisms and general peptide polymorphisms have is that there are to some degree limits on what self-polymorphism can be. If each and every peptide were allowed to have a very large number of equally common alleles (such as the MHC polymorphisms), then besides the basic structural problems of having the right variants interacting with their biochemical partners, there are problems with the role of "self" and the generation of the necessary holes in the immune repertoire to protect against self-reactivity. There is a simple dynamic which exists between the numbers of T-cells that are able to be deleted during thymic education by a single MHC allele and the optimal number of MHC genes that can exist in an organism and still allow the final repertoire of T-cells in the periphery to be maximal (Nowak *et al.* 1992). In one case, a mouse strain was found to have had an amplification of MHC class I genes such that several thousand such genes are present in that genome, and a specific induced mutation process was also involved in the inactivation of most of these genes (Delarbre *et al.* 1992). Implied in the equation given by Nowak for the optimal number of MHC genes is the concept that if all of the various structural and housekeeping genes were to suddenly become so polymorphic that most individuals would then be generally heterozygous, the much larger number of holes that are created in the immune repertoire through thymic education would have to become a factor that would change the basis of this equation. (This could be considered the "infinite-self" hypothesis with so many self epitopes existing that an immune repertoire is impossible to obtain.) Indeed, a large number of individuals carry significant regions of homozygosity within their genome (Broman and Weber 1999), and while the plasticity of the immune response should

easily allow for individuals who are completely heterozygous, a reduction in the density of self epitopes that must be deleted may make the process of deletion a more effective one.

1.6 Linkage disequilibrium, population admixture and selection

The frequency of a given allele in a population is related to the selection that has taken place in terms of survival for individuals who carry the allele (or who carry a locus actually under selection which is in linkage disequilibrium with the allele). In the event of a mutation occurring on a given chromosome region, the number of generations it will take to unlink the nearby specific alleles from the mutation is a function of the genetic distance rather than the absolute physical distance. One key question in the second part of this thesis is on the origin of the deletion mutation *CCR5-Δ32*. The allele frequency in Ashkenazi Jews is a result either of admixture into the Ashkenazi and then selection (for some survival advantage conferred by this allele) or a result of an origin within this population and then admixture from the Jewish population into the general Caucasian population, with a lesser role for selection perhaps needed to explain the different population frequencies. Overall, admixture has such a significant role in detection of regions of linkage disequilibrium that each instance may need to be separately evaluated (Wilson and Goldstein 2000) for regions of interest and a background level of linkage disequilibrium that exists in the genome will need to be evaluated with the same dense marker set in a number of populations but may be several centimorgans in distance (Service *et al.* 2001). Areas of linkage disequilibrium are detectable over regions that in physical terms are up to 400 kbp in distance (Abecasis *et al.* 2001; Koch *et al.* 2000). Regions of the chromosomes that differ significantly in

their degree of linkage disequilibrium have been identified and a number have been defined as “deserts” and “jungles” (Yu *et al.* 2001) where several long (up to 6 Mbp) regions show particularly low (desert) or high (jungle) recombination rates with linkage disequilibrium being much more common and extending for greater distances in the deserts than in the jungles.

While a general approximation for the relationship between genetic distance (on a centimorgan or cM scale) and physical distance along a chromosome can be given as around one centimorgan per million basepairs (1cM/Mbp), the relationship varies along each chromosome and the above “jungles and deserts” are where values of more than 3 cM/Mbp or less than 0.3cM/Mbp are seen (Yu *et al.* 2001). One possibility concerning changes in recombination over physical distance is that isochore transitions (seen in the chromosome banding pattern) display sharp changes in “GC” content and a transition between higher and lower linkage disequilibrium values has been mapped to these boundaries (Eisenbarth *et al.* 2000). Changes in distribution of linkage disequilibrium across regions of the chromosome may be more important than differences seen between population isolates (Zavattari *et al.* 2000). Recently, an analysis of haplotype variation and linkage disequilibrium in humans using 3,899 single nucleotide polymorphisms (SNPs) in 313 genes found the probability that any particular pair of SNPs being in linkage disequilibrium was not predictable and that linkage disequilibrium should be determined empirically for any specific genomic region (Stephens *et al.* 2001). Variation in linkage disequilibrium may be a result of the different demographic histories of various population groups, and the occurrence of gene conversion may result in apparent changes in the degree of linkage disequilibrium

that is seen (Ardlie *et al.* 2001; Frisse *et al.* 2001).

1.7 Microsatellite polymorphisms and mutation dating

Studded throughout the genome are a wide variety of repeat elements including the class of "simple" repeats which are known as microsatellite repeat DNA (Dib *et al.* 1996). Microsatellites are defined to be where a motif of one to several bases is repeated a large number of times. To some degree, "incomplete" repeat elements can turn into an actual microsatellite repeat when a base that interrupts a proto-repeat changes into a base that fits the repeat motif and thereby passes the size threshold for generation of different alleles (Messier *et al.* 1996). Where the number of repeat units is above a typically uninterrupted threshold size, there is an increased chance of slippage mutation between generations which adds (or removes) a repeat (Brinkmann *et al.* 1998; Farrall and Weeks 1998). The coding sequence of the adjacent DNA provides anchor points for PCR primers which can amplify the microsatellite region to detect lengths of the DNA fragments by electrophoresis and allows allele assignment. The majority of microsatellite repeats lie outside of coding regions and so change in the gene sequence caused by insertion or deletion of a number of nucleotides has little biological consequence (Messier *et al.* 1996; Schlotterer 2000), while trinucleotide repeat polymorphisms in some human diseases are very much a part of a coding sequence and each allele is seen to add an additional amino acid to a polypeptide chain. Where the microsatellite repeat is not the size of a codon triplet and the repeat does lie within a coding region, the insertion of a repeat unit causes a frameshift mutation in a gene (Metzgar *et al.* 2000). Microsatellite use in the study of population dynamics requires an understanding of the mutation process affecting microsatellite repeat motifs as the time

frames in evolution are often too large to ignore novel microsatellite mutations (Schlotterer 2000). It remains to be seen to what degree the mutational dynamics of microsatellites is via stepwise replicative-slippage mutation or by unequal crossing over events (Calabrese *et al.* 2001; Renwick *et al.* 2001). While strong mutation rates are a feature of some microsatellite motifs, many are found to exist across the primates (Charmley and Concannon 1993; Rubinsztein *et al.* 1995) and one example has been reported of a microsatellite motif being detected across one billion years of evolution (Martin *et al.* 2002).

Legend for Figure 1 a - e - Hypothetical alleles, haplotypes and mutation origin

Markers are indicated by various sets of letters, numbers and symbols. The markers are hypothetical and could variously be genes, microsatellites, SNPs, deletions or other allelic states. X and W, for example, might be considered as the delta32-CCR5 deletion and wild-type alleles (a bi-allelic polymorphism), while the letters “A” to “F” could be a microsatellite with seven allelic forms. Figures 1a and 1b shows how these markers could appear on a given section of chromosome and how in the population various chromosomes might be represented. In Figure 1c a mutation occurs, as shown by the arrow (e.g. the 32 bp deletion in CCR5), between the “3" and “A” markers with other specific markers flanking these. Figures 1d and 1e show the eventual loss of association from more distance markers and eventually from the nearby markers over many generations.

Figure 1 a

Hypothetical set of chromosome markers (and the respective alleles)

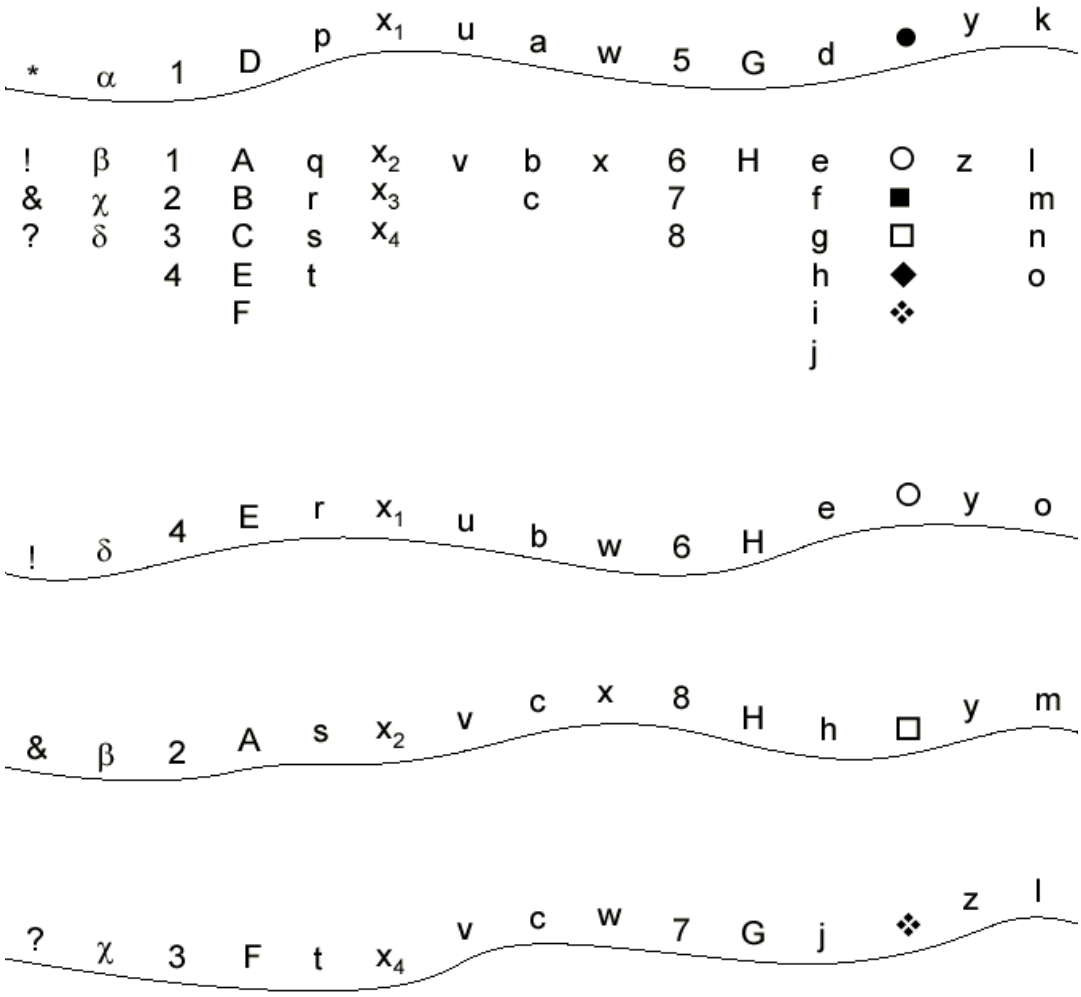
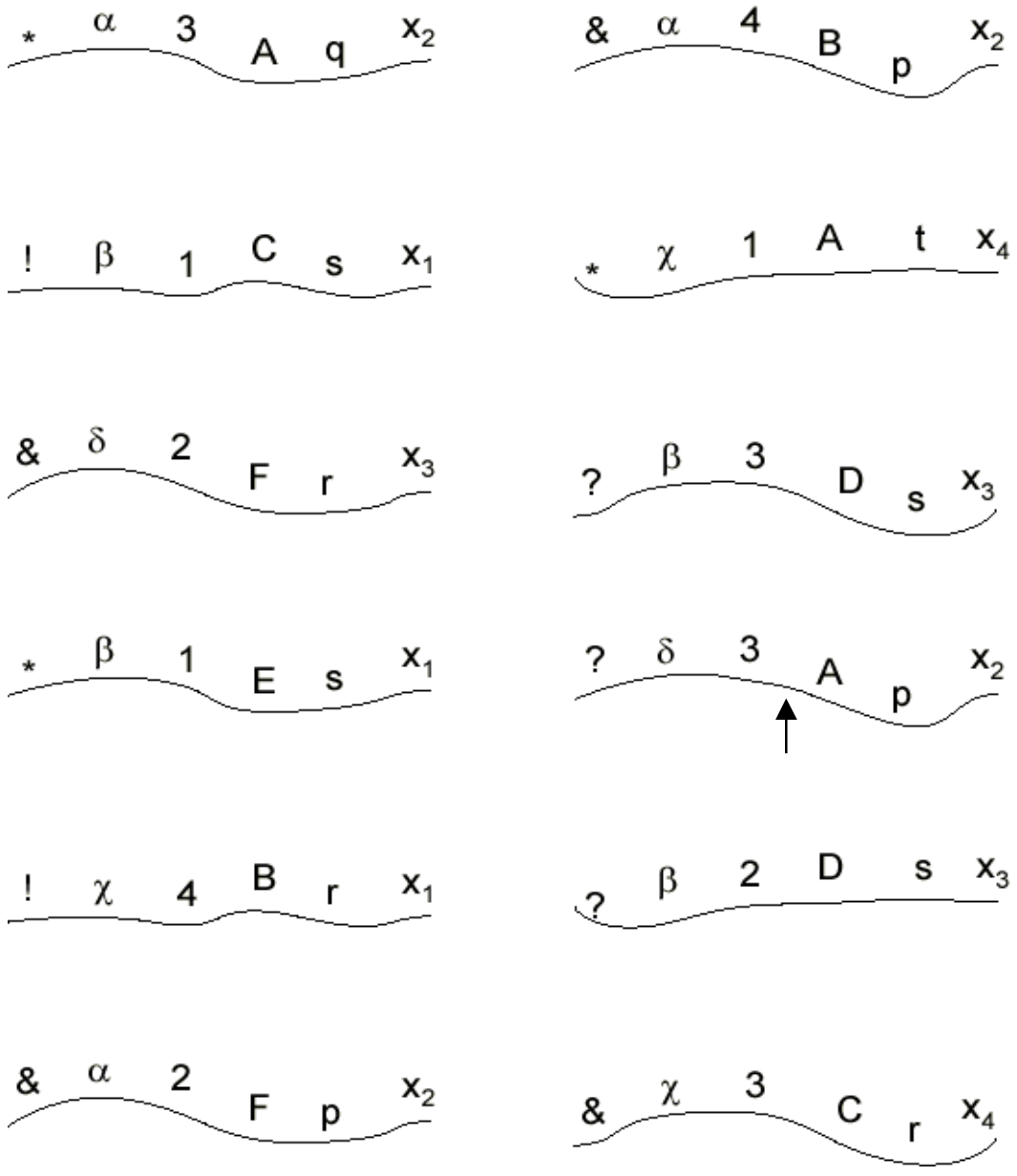


Figure 1b

Examples of chromosomes in the population with allelic differences

Figure 1c

Portion of chromosome (above) upon which a new mutation occurs



The arrow shows a mutation now "linked" to the given markers

Figure 1d

Loss of association with distant markers

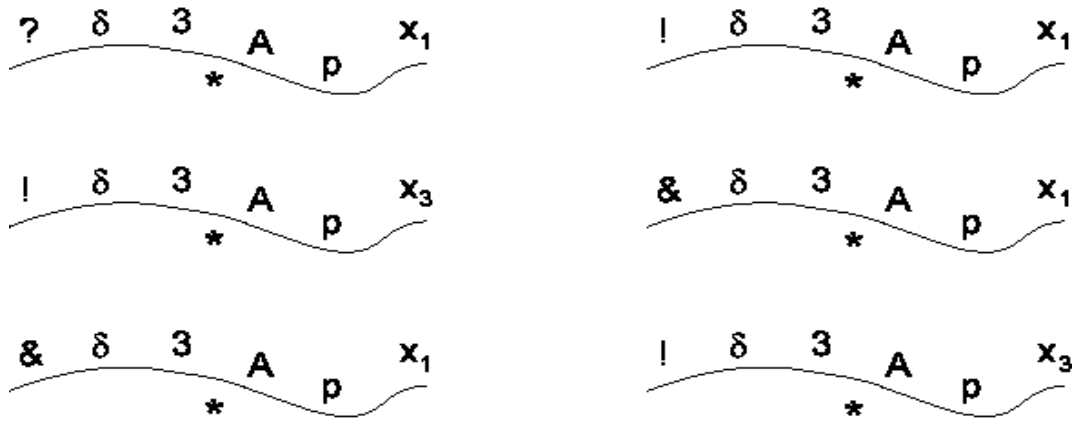
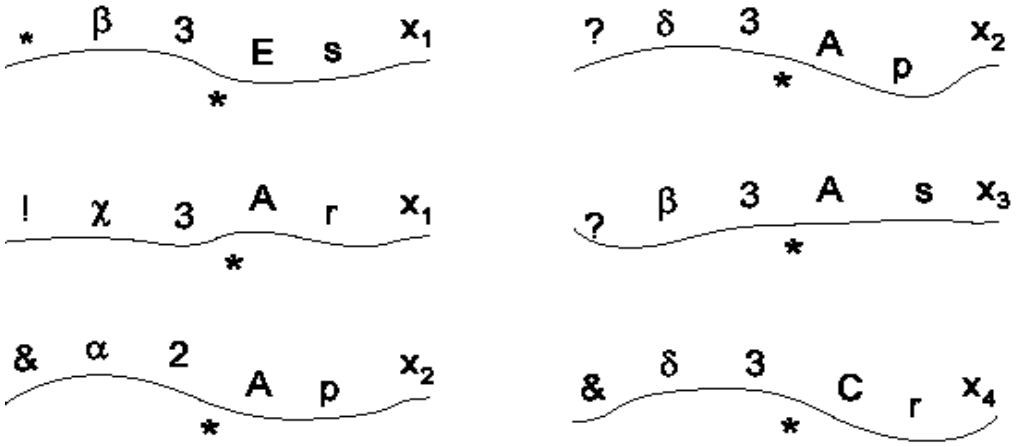


Figure 1e

Greater loss of original haplotype with greater time



The original "3-*-A" haplotype can provide information on the age of the mutation relative to the physical / genetic distance between them; different "windows" into possible mutation age depend on the markers

1.8 Genomics and repetitive regions of DNA

The features that affect genome structure in most vertebrates include different classes and types of repetitive DNA that range from the simple microsatellite-type repeats to much longer sequences in various types of insertion sequences, endogenous retroviral elements, LINE and SINE sequences, mini-satellite DNA and other repetitive motifs. A serious computational hurdle in the current resolution of the complete sequence of the human genome (and of the many other higher-order vertebrates that are being sequenced) is the complexity of what the repetitive elements are and where to put them in the sequence. An exception to this is the Japanese pufferfish *Fugu rubipes*, which has a genome that may have used a mutational bias towards elimination of repetitive DNA and so has essentially the minimal vertebrate genome possible (about 400 Mbp) and is devoid of repetitive elements (Venkatesh *et al.* 2000). Interestingly, in the compaction of the Fugu genome, the organization the general nature of intron/exon gene architecture remains consistent with other teleosts and indeed with other vertebrates, although the intron size is generally greatly reduced. In the pufferfish T-cell receptor α-chain locus, for example, while exons are of a similar size to those in the human genome, the introns of the constant-chain gene are 10 to 17 times smaller than the corresponding human TCRAC gene (Wang *et al.* 2001). In terms of sequence conservation between pufferfish and human TCRA loci, while very little coding-region nucleotide and amino acid sequence conservation is seen, regulatory sequences required for assembly of the TcR gene (including recombination signals, RNA splice sequences, and key amino acid residues required for maintaining integrity of the TCRA chain polypeptide) are found to be highly conserved (Wang *et al.* 2001).

Apart from the pufferfish genome, which is likely to have had a range of repeat elements prior to compaction of it's genome (Venkatesh *et al.* 2000), vertebrate genomes actually may make use of the repetitive DNA in different ways. It has been suggested by Daiqing Liao that what is often called "junk" DNA might provide the grist for the mill in terms of stability for those genes where multiple copies of a gene are needed and are needed in an identical form. He makes the case that "concerted evolution", where tandem sets of identical or near-identical genes are found to generally return a change in one member back to the original form, is common in biology (Liao 1999). While this is important in providing some genetic stability for genes like rRNA genes, it may also act to drive evolution in situations such as mutation in the gene coding for the sperm receptor on an egg. Vertebrate evolution has a role for exon shuffling that makes use of the degree to which elements such as the endogenous retrovirus genomes are repeated in the vertebrate genome, and it is possible that the origin of the class II MHC molecules are a result of this sort of shuffling (Klein and Sato 1998).

Many microsatellites have a moderate to large number of alleles and while others do have small allele numbers, it is often difficult to assign haplotypes as in many instances the markers being typed are both heterozygous. Even with the bi-allelic systems such as in the TCRBV genotyping presented in Chapter Three, typing for several markers can often result in a significant number of individuals for whom haplotypes are not able to be assigned. As shown in Table 1.1, four bi-allelic markers yield 81 possible phenotypes with 33 of those not able to have specific haplotypes assigned unless family data is available from which to assign the proper "phase" for the observed haplotypes.

Table 1.1 81 possible 4-marker biallelic phenotypes

aaaa	abaa	acaa	baaa	bbaa*	bcaa	caaa	cbaa	ccaa
aaab	abab*	acab	baab*	bbab*	bcab*	caab	cbab*	ccab
aaac	abac	acac	baac	bbac*	bcac	caac	cbac	ccac
aaba	abba*	acba	baba*	bbba*	bcba*	caba	cbba*	ccba
aabb*	abbb*	acbb*	babb*	bbbb*	bcbb*	cabb*	cbbb*	ccbb*
aabc	abbc*	acbc	babc*	bbbc*	bcbc*	cabc	cbbc*	ccbc
aaca	abca	acca	baca	bbca*	bcca	caca	cbca	ccca
aacb	abcb*	accb	bacb*	bbcb*	bccb*	cacb	cbcb*	cccb
aacc	abcc	accc	bacc	bbcc*	bccc	cacc	cbcc	cccc

Note - "a" is the homozygote phenotype for the #1 allele of a given marker, and "c" is the homozygote for the #2 allele; the heterozygotes are shown as "b" and where two or more heterozygotes occur (n= 33, shown with *), the phase of the genotypes cannot be assigned.

1.9 Evolutionary phylogenetics of specific vertebrate immunity

At a stage in the early evolution of the vertebrates there are branch points, in the period of time estimated to be between 450 million and 500 million years ago, at which the development of adaptive immunity may be seen to be established (Barclay 1999; Laird *et al.* 2000; Shimeld and Holland 2000; Shintani *et al.* 2000). Once in place, the "anticipatory" adaptive immune system in the early vertebrates was so successful that the basic components, which include Ig, TcR, chemokine receptors and MHC, are now seen to have been applied across the subsequent radiation of the various vertebrate lineages. Quite a number of questions remain about the transition from the "innate" to the "combinatorial" types of immune systems, especially with regard to the development of antibody and the specific production of antibody by a given B-cell, the similar development of T-cells and of the class I and class II MHC systems of antigen

presentation. Changes in the overall vertebrate genome which may have a role in the eventual evolution of self-specific immunity include two whole genome duplication events in the early vertebrates (Gibson and Spring 1998; Gibson and Spring 2000; Martinez-Mir *et al.* 2001) and an additional whole genome duplication shown in the teleost fishes (Amores *et al.* 1998). Although there is serious debate about entire genome duplication events in early vertebrates (Friedman and Hughes 2001; Friedman and Hughes 2003), duplication of exons and genome sub-regions does occur (Letunic *et al.* 2002; Samonte and Eichler 2002). As well as a role for segmental or regional duplication and possibly entire genome duplication events in vertebrate immune evolution (Boot-Handford and Tuckwell 2003; Durand 2003), domain and exon shuffling can contribute to producing new functions in peptides out of "old" pieces. An example is the establishment of the peptide binding pocket of MHC class II which has been postulated to be a result of exon shuffling bringing together the class II peptide-binding pocket with the Ig-like domains of the class II structure (Klein and Sato 1998). While exon-shuffling may well have been established long before the advent of the vertebrates (Patthy 1999; Suga *et al.* 2001), it combines with the other recombination-producing mechanisms (such as those of endogeneous retroviral elements) in the higher vertebrates to help create "new genes" out of "old gene segments" without having to begin afresh (Apic *et al.* 2001) and with the benefit of maintaining a minimal amount of "self" to be thymically defined. Primate specific duplication of segments (15 kb and larger) of the genome does however contribute to an increase in the diversity of the proteome (Bailey *et al.* 2002). The enrichment of genes via recent segmental duplications is suggested to be nonrandom, with genes such as natural killer receptors, defensins, cytokines, interferons, serine proteases, and MHC (all associated with

immunity) among those that were particularly enriched (Bailey *et al.* 2002).

The effects of introducing a means for generating a specific "self-defined" inducible immune system, such as is found in vertebrates, are not limited to effects on the evolution of the vertebrate genome. The vast array of microbes, viruses and parasites that have any degree of success in vertebrates have had their genomes shaped by the evolutionary pressure that the evolving immune system has placed on them. In a recent review on what modern genomics is beginning to reveal of the microbial genomes, Palmer describes the effects of reductive evolution on several "small genome bacterial pathogens". These pathogens include chlamydiae, rickettsiae and ehrlichiae, mycoplasmas, and spirochetes whose genomes, a third the size of *Escherichia coli* as defines small genome pathogens, show consistent evidence of loss of metabolic function and reduction of multiple overlapping pathways and duplicated genes that would be expected from obligate parasitic pathogens. While the host can provide the nutrients that then allows such reduction in genome content, the specific need to remain a step ahead of the host's immune system sees these microbes devoting a high percentage of their genomes to paralogous families of polymorphic surface markers (Palmer 2002), which shows that the highest priority of these small genome obligate pathogens is the evasion of the immune system. Evidence from the genus *Anaplasma* and the related genus *Ehrlichia* show different strategies in the structure and expression of the paralogous genes through the 390 million years since these two genera diverged from a common progenitor (Ohashi *et al.* 2001). Considering that the introduction of self-specific immunity into the vertebrates occurred at between 450 to 500 million years ago, it is quite interesting that specific genetic systems for microbial evasion of

immunity can be dated back to nearly 400 million years ago. This is a basic example of co-evolution, where the interplay during evolution between host and pathogen has had significant effects on both genomes.

1.10 Duplication versus a limit to "self" in immune evolution

There are suggestions that, in the evolution of vertebrates, two complete genomic duplications may have occurred and these occurred very early in the divergence of the vertebrates (Abi-Rached *et al.* 2002; Gibson and Spring 2000; McLysaght *et al.* 2002), with a third duplication event found in the evolution of the bony fish (Amores *et al.* 1998; Taylor *et al.* 2001; Van de Peer *et al.* 2001). Some understanding of the role of genomic duplications in developmental evolution of the vertebrate body plan can be seen through comparison with the prototypical cephalochordate amphioxus (Minguillon *et al.* 2002), which is the closest living relative to the common vertebrate ancestor with a divergence time estimated at 550 Myr. The large scale duplications have played a significant role in creating the vertebrate proteome and it has been found as well that small scale duplications in early vertebrate evolution played a similar role (Gu *et al.* 2002). While genomic duplications are limited to the early chordate - vertebrate transition, numerous gene families have been expanded through various smaller (exon - gene - region) duplications (Dermitzakis and Clark 2001; Duret 2001; Letunic *et al.* 2002; Patthy 1999; Samonte and Eichler 2002). Some families show large scale or even massive duplication such as the above-mentioned pigmy mouse (Delarbre *et al.* 1992) or the olfactory genes (Menashe *et al.* 2002), however most duplicated genes are eliminated within a few million years (Lynch and Conery 2000).

1.11 The study of immunogenetics of human disease

The understanding of the host genetic factors responsible for susceptibility to, and the clinical expression of, human disease is a challenge that has already proven to be formidable. With respect to autoimmune disease, much has been learnt but the reproducible detection of genetic effects outside the MHC has been slim despite a large international effort both with candidate gene studies and whole genome screening. Both multiple sclerosis and diabetes provide excellent examples in this regard. Moreover, while most autoimmune disorders are likely to involve environmental triggers, the identification of these and their interaction with host genetic factors appear to be a long way from elucidation. The concept that discovery of genetic association would soon lead to an understanding of the pathogenic contribution of that gene function to disease has also met formidable challenge. It is now over 30 years since the first description of an HLA and disease association, and despite extraordinary advances in knowledge of the MHC leading to several Nobel prizes, an exact explanation of the HLA association in autoimmune disease remains unclear.

The study of immunogenetics has been given a considerable boost by the detection of the effects of inherited variation on infectious disease for which an excellent example is that of chemokine receptor CCR5 on HIV disease, a field that has expanded to show broader effects than those on viral co-receptor usage and a field which continues to extend into other viral disorders such as hepatitis C. Human immunogenetics also continues to be informed by important advances in the understanding of the basic rules of the immune system in humans and other animals and most recently by the extraordinary advances in the understanding of the genome through the human and

other species genome projects.

While the challenge may be difficult, the goal of such studies is to develop new and effective preventive and therapeutic strategies for some of the world's most important chronic diseases, therapies that will derive from a more complete understanding of the pathogenesis of each disease. Gene discovery is not an end in its own; it points the way toward these ends and is as important and valid a strategy as any other in the human quest for reduction in the burdens associated with severe illness. In this regard, both of the polymorphic systems analysed in this thesis provide good examples. Anti-TCR therapy in multiple sclerosis reached clinical trial level based on knowledge of genetic restriction of TCR usage. Anti-CCR5 therapies are in clinical trials of HIV infection and may realise a broader role.

The areas of science chosen for review in this chapter reflect my strong conviction that this formidable challenge will only be met if those involved in the genotyping of human disease understand and take into account the much broader picture within which their studies reside. This picture includes the evolution and function of the immune system, and its interaction with the environment and their variation amongst populations both before and during the period of recorded human history. In chapters 3, 4 and 5 I have selected pieces of my work that I hope exemplify aspects of importance within this contextual framework.

Chapter Two - Materials and methods

2.1 Introduction

Modern medical research methodology has undergone several revolutions over the course of the last two or three decades, and two or three of these have had direct effects on the topics of this thesis as the project was underway. The first was the introduction of the polymerase chain reaction (PCR) for molecular genotyping as is shown in Chapter Three for the TCRBV polymorphisms, where the previous method was by restriction fragment length polymorphism (RFLP) with radio-labelled probes as will be shown in this chapter. The next important methodological revolution is found in the availability of draft and finished sequence information from the public and private sectors of the human genome project. While these two revolutions may have changed this thesis project along the way with regard to *how* the research was done, they have not changed the reason why this work is of some importance. As reviewed in the introductory chapter, autoimmune and infectious diseases are often seen to have various degrees of associations with allelic variants of genes in the human genome, and mainly with those which have a role as a component of the immune response in some way. As was also discussed above, there are various types of subject group, family and population differences that can confound the basic associations sought in genetic studies. Research on comparative genomics has shown that some populations have greater or lesser regions of linkage disequilibrium around given pairs of markers and this may be an effect of each population's demographic history. Recent methods in study design have seen a shift from case-control to family-based comparisons due in part to the difficulty in showing that a random group of controls can ever be completely matched for a patient cohort. This, in a statistical sense, has been yet another revolution which has occurred over the last few years. Much of the data here is on a case-control basis as that

was how susceptibility studies were generally done until just recently (Allison 1997; Allison *et al.* 1999). One family is however presented here, in which a homozygous *CCR5-Δ32* HIV+ person was identified, where microsatellites were used to assign haplotypes for both the region around CCR5 and for markers flanking CXCR4, the other co-receptor used by T-cell tropic strains of HIV. Study of this family is of interest due to the occurrence of HIV in a person lacking the CCR5 co-receptor, for which this was the initial reported event (Biti *et al.* 1997).

2.2 Tissue culture and DNA extraction

Peripheral blood mononuclear cells (PBMC) were transformed into Epstein-Barr virus lymphoblastoid cell lines (EBV-LCLs) by co-culture of density-centrifugation separated and washed PBMCs with filtered supernatent from the B-95 (marmoset) cell line as a source of EBV (kindly provided by the Virology Department, ICPMR). PBMCs were plated at about one million cells per well in 24-well tissue culture plates (Falcon, BD Biosciences, Bedford MA, USA) in RPMI-1640 medium (Flow, ICN Biomedicals, Irvine, CA, USA) supplemented with 10% FCS (Gibco, Invitrogen Aust., Mount Waverley VIC.), and penicillin/streptomycin (Flow/ICN). Cultures were maintained at 37°C with 5% CO_2 with half of the culture fluid replaced with fresh medium every 3 or 4 days. When cultures were established, cells were harvested into 15 ml centrifuge tubes, gently spun (800 rpm for 6 minutes) and after removal of most of the supernatent, the cell pellets were resuspended in the remaining medium and the tubes were placed into storage at -20° C until required. DNA was extracted by a simple modification of the salting-out method (Lahiri and Nurnberger 1991) where an additional 4 ml of low-salt buffer was added to the cell suspension before proceeding as if extracting DNA

from a 5 ml blood sample.

2.3 Ethics approval for subjects and study cohorts

Informed consent was in each case provided from the Westmead Hospital ethics review board and, where appropriate, from an external ethics review board.

2.4 Subjects for TCRB analysis

The MS cohort consisted of 122 unrelated Australian multiple sclerosis patients (Buhler *et al.* 2000); all were defined as having relapsing-remitting disease and were clinically definite or laboratory-supported definite MS, according to the Poser Committee criteria (Poser *et al.* 1983). Data for several MHC markers have been reported for most of these patients (Bennetts *et al.* 1995). The control group consisted of 96 Australian individuals matched for age and ethnic background who had been HLA typed and was made up of hospital staff along with spouses of a number of patients. Data from these cohorts is shown in Chapter 3. The RA patient (n = 50) and control set (n = 50) were provided by Prof. C. Bernard of La Trobe University and Dr. I. McKay of Monash University. An additional cohort of RR-MS, SP-MS and CP-MS (n = 53), patients with "other neurological diseases" (OND, n = 50), and normal controls (n = 50) were made available from a Brisbane, Queensland MS cohort by Dr. Judith Greer. Comparisons of the TCRB allele and genotype frequencies between patients and controls for these smaller cohorts did not show differences that were statistically significant. Haplotype assignment in these smaller cohorts also did not show significant differences, owing in part to the exclusion of individuals who were heterozygous for more than one marker.

2.5 Subjects for *CCR5-Δ32* studies

2.5.1 Autoimmune diseases - type 1 diabetes, MS and SLE

Samples of extracted DNA were obtained from three cohorts of adolescent Caucasian patients with type 1 diabetes (total n = 620) collected at teaching hospitals in Sydney, NSW and Melbourne, Victoria in Australia and from Christchurch, New Zealand and approval was given by institutional ethics committees. SLE patients were collected at Westmead Hospital, and were not matched in age to controls. Control samples were 253 (adolescent) students from year 11 in a Sydney, Australia school setting (3rd generation Australian students). The MS patient and control cohorts are as described above (Buhler *et al.* 2000).

2.5.2 CCR5 in the Ashkenazi Jews

Samples were available from 1,388 individuals and included a large number of Ashkenazi Jews (n = 797) along with Jews who were unsure about their Ashkenazi status or who were from a "mixed" religious background (n = 104), some Sephardic Jews (n = 35), and a cohort of non-Jews (n = 442). Classification of status into Ashkenazi / Sephardi / mixed / unknown was made by the subject at the time of specimen collection. Information was available on the birthplace of the subject, their parents and their grandparents for about 90% of individuals who were tested here for their *CCR5-Δ32* genotype. The majority of samples were from adolescent (year 11) high school students in a number of Sydney Jewish schools sampled in sequential years, with a smaller number of samples from adults, predominantly Ashkenazi Jews, who were self-referred to the Tay-Sachs testing program. The non-Jewish samples were also adolescent (year 11) student samples from non-Jewish schools collected

contemporaneously alongside one of the Jewish school cohorts. The non-Jewish controls (n = 442) consist of both Asian (n = 131) and Caucasian subgroups (one group with Australian-born grandparents, n = 187, and the remainder who were of mostly European Caucasian background, n = 124). DNA bank and genotyping was done with institutional ethics approval on samples for which individuals had given informed consent. All samples were deidentified through use of an intermediary "Gene Trustee" (Aizenberg *et al.* 2001); this enabled the current study to obtain access to demographic and genealogical data without compromising the anonymity of individual samples in accordance with the subjects' consent. Data from these cohorts is shown in Chapter 4.

2.5.3 CCR5 and CXCR4 in the "VaHa" family

In addition to the population groups, a family of interest in the biology of AIDS and CCR5 was also studied in which the first reported case of HIV infection of an individual homozygous for the CCR5 deletion mutation occurred (Biti *et al.* 1997). This family, the "VaHa" family, consists of the HIV+ proband ("VH") who is a CCR5-Δ32 homozygote; the mother ("VO") and a brother ("VF") who are both CCR5 heterozygotes; a sister ("GO") who is a homozygote for the CCR5 wild-type allele; and a brother ("VM") who, while HIV-negative, is also a homozygote for CCR5-Δ32.

2.6 Polyacrylamide gel electrophoresis (PAGE)

Mini-PAGE (Bio-Rad, Hercules, CA, USA) gels were prepared from pre-mixed acrylamide/bis-acrylamide (29:1) solution (Amresco, Solon Ohio, USA) to a final concentration of 10% acrylamide, polymerised with TEMED (Bio-Rad) and ammonium persulfate (Sigma-Aldrich, St. Louis, MO, USA). Electrophoresis was performed for 30

minutes at 200 volts. Samples were visualized by staining with ethidium bromide (0.5 μg/ml) and then photographed using Polapan 667 film (Polaroid, Waltham MA, USA).

2.7 RFLP analysis of tcrbv8s1 and tcrbv11s1

Plasmids containing the probes for these markers were a kind gift of Dr. P. Concannon however problems with an extended delay of delivery in coming through customs resulted in poor recovery of cultures and after a number of attempts only a few experiments were able to provide data on these markers. Publication of a PCR-based method by Wei and co-workers (Wei *et al.* 1995) for bv8s1 and a number of other markers, but not for bv11s1, gave a reason to shift away from RFLP and to pursue the PCR based method. Examples of some successful RFLP autoradiographs are show in Figures 2.1 and 2.2, and while initial data for bv8s1 and bv11s1 was obtained, it was rapidly superceded by the PCR-based data.

Figure 2.1
RFLP autoradiograph of genomic DNA probed for V-beta 8.1

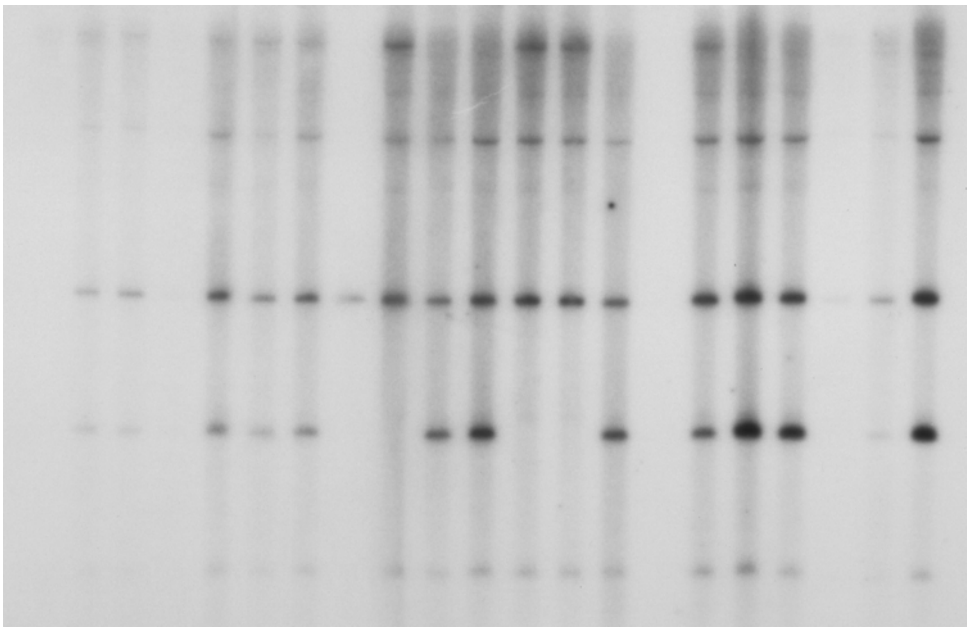
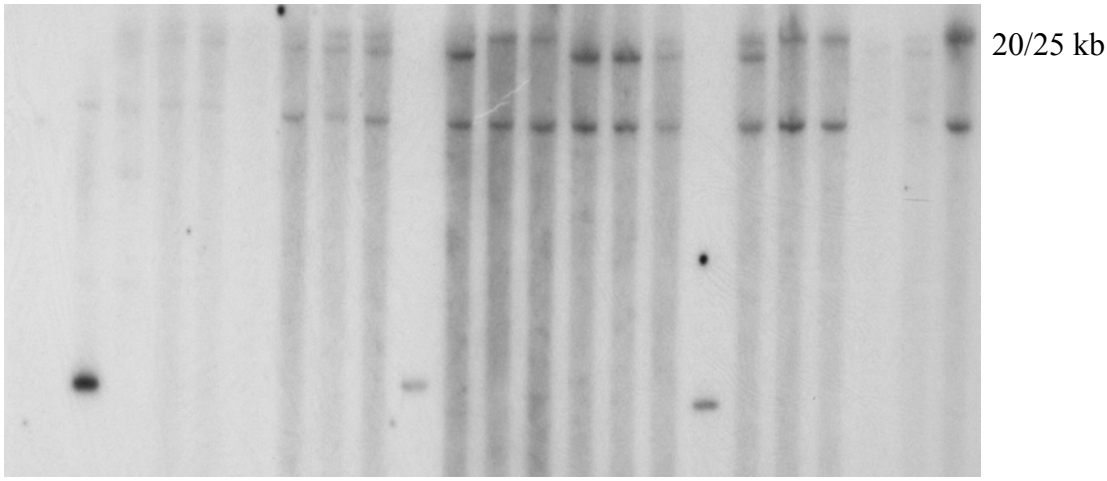


Figure 2.2 RFLP autoradiograph of genomic DNA probed for V-beta 11



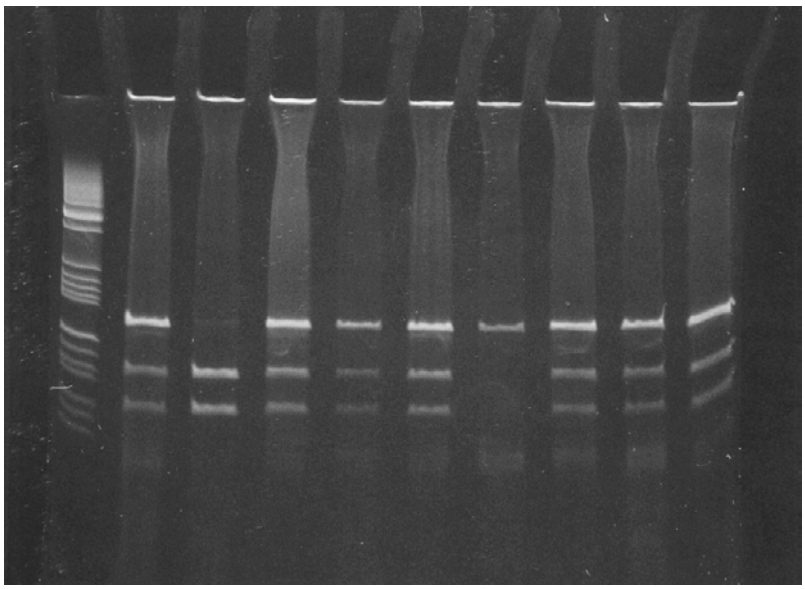
Legend for Figures 2.1 and 2.2
Restriction digested genomic DNA samples were Southern blotted onto membranes and probed with ^{32}P labelled V-beta probe. The 23kb V-beta 8.1 band is difficult to see in lanes with weak DNA transfer, as are the 20kb and 25kb bands for V-beta 11. These sort of difficulties prompted a shift to PCR based genotyping.

2.8 PCR typing of bv6s5, bv8s1, bv10s1, bv15s1 and bv3s1

Primer sequences used were those published by Wei and colleagues (Wei *et al.* 1995) and are listed here. PCRs were run in 30 μl reaction volumes with mineral oil overlay to prevent sample evaporation. Transfer of sample to a fresh tube was followed by careful mixing of 10 units of the appropriate enzyme and incubation at 37°C overnight (for enzymes other than *Bst*NI) The bv6s5 primers were 5'-ACAATTGTTTTTCTCCTGTTCC-3' and 5'-GGTACCAATAAAGGCGGTTG-3', with digestion by *Bst*NI (NEB, New England Biolabs, Beverly, MA, USA) (at 60°C); an example is shown in Figure 2.3. For bv8s1, primers 5'-GTAAGCAATTAACTTTTTGAAGG-3' and 5'-TCAAAAGCACTTCCCAAAGTA-3' were used with the enzyme *Bam*HI (NEB). Primers used for bv10s1 were 5'-CCATGTGCCTCAGACTTCTC-3' and 5'-GGTACAGGATGAGTTTTTGGA-3' and the enzyme was *Hae*III (NEB). Examples of PAGE results for bv8.1 and bv10.1 are shown in Figures 2.4a and 2.4b, respectively. For bv3s1 (shown in Figure 2.5), primer 5'-GTAGGCCTCGTAGATGTGAA-3' and primer 5'-TGTGCTTATGGAGCTAGTTTC-3' and the restriction enzyme *Pvu*II (New England BioLabs, NEB) were used. The primers for bv15s1 (Figure 2.6) are 5'-CCTCAGAGATTGCAGCAC-3' and 5'-CTTTGTTTATATCTTTGACATCAAAGGAGTAATAGATCAGC-3', with digestion by *Pvu*II (NEB). After digestion, electrophoresis was performed with 10% polyacrylamide-TBE gels and the digested and undigested products were visualized after ethidium bromide staining. The undigested bands were called allele #1 and the digested bands called allele #2, as defined by Wei and co-workers (Wei *et al.* 1995).

Figure 2.3 PAGE (10% mini-gel) of bv6s5 PCR product digested with *Bst*N1

Lanes:
1 2 3 4 5 6 7 8 9 10

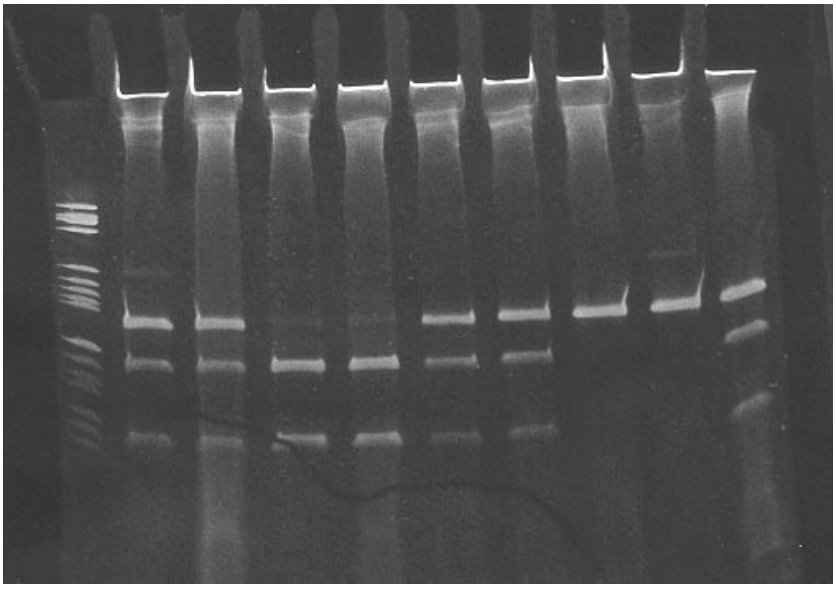


undigested (#1) allele
& digested bands
(#2 allele)

Legend for Figure 2.3
Molecular weight marker bands are in the leftmost lane. The sample in the third lane shows fully digested PCR product which genotypes as a homozygous individual for the #2 allele. Lane 7 shows an individual with no digested bands, indicating a homozygous genotype for the #1 allele. All the other lanes show equal amounts of digested and undigested PCR product and as such are allele 1 and 2 heterozygous genotypes.

Figure 2.4a - 10% mini-PAGE of bv8s1 PCR product digested with *Bam*HI

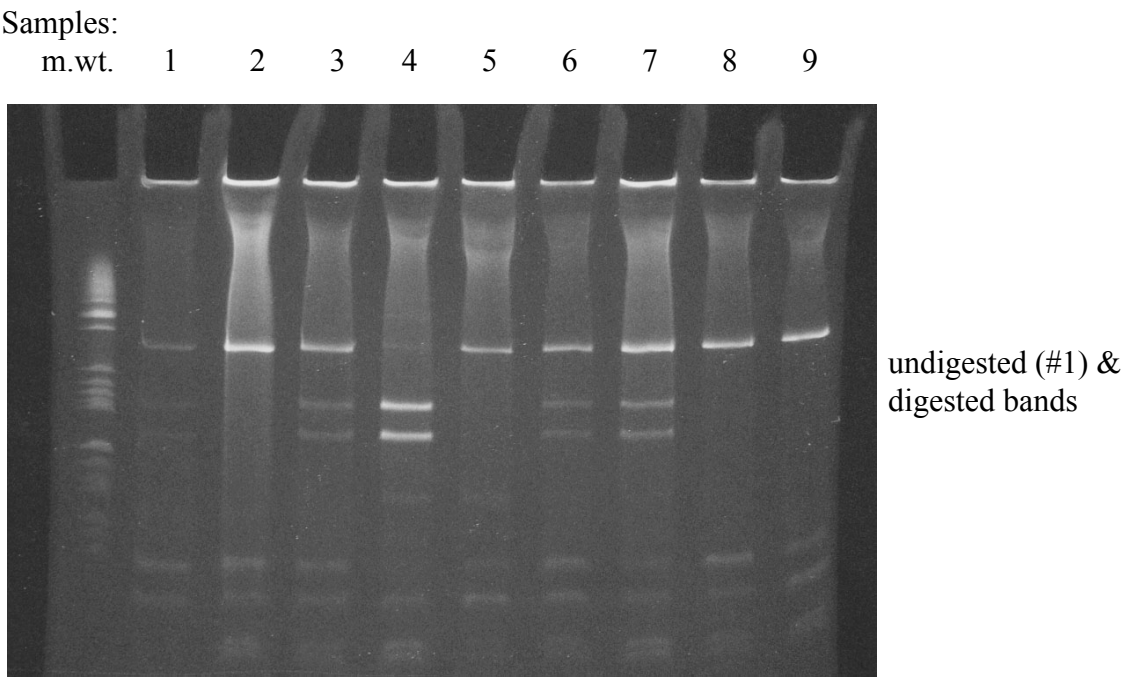
Lanes:
1 2 3 4 5 6 7 8 9 10



undigested band (#1)
digested band (#2)

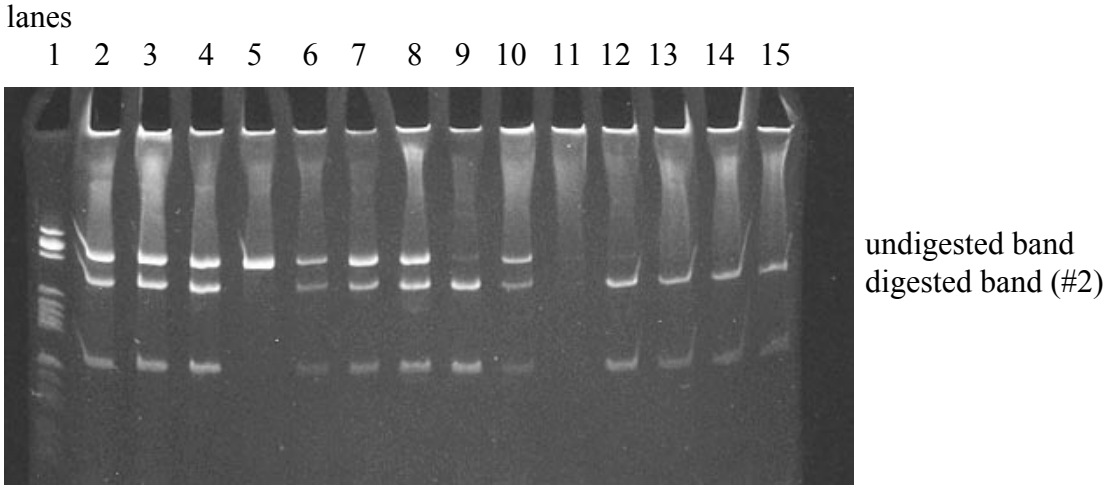
Legend for Figure 2.4a
Lane #1 is a restriction digest of plasmid DNA used for orientation and as a molecular weight ladder. Lanes 8 and 9 show homozygotes for the #1 (undigested) allele while the samples in lanes 4 and 5 are fully digested and are therefore homozygotes for the #2 allele. Samples in lanes 2, 3, 6, 7 and 10 are heterozygotes with equal amounts of digested and undigested DNA.

Figure 2.4b - bv10s1 PCR product digested (*Hae*III) - 10% mini-PAGE gel



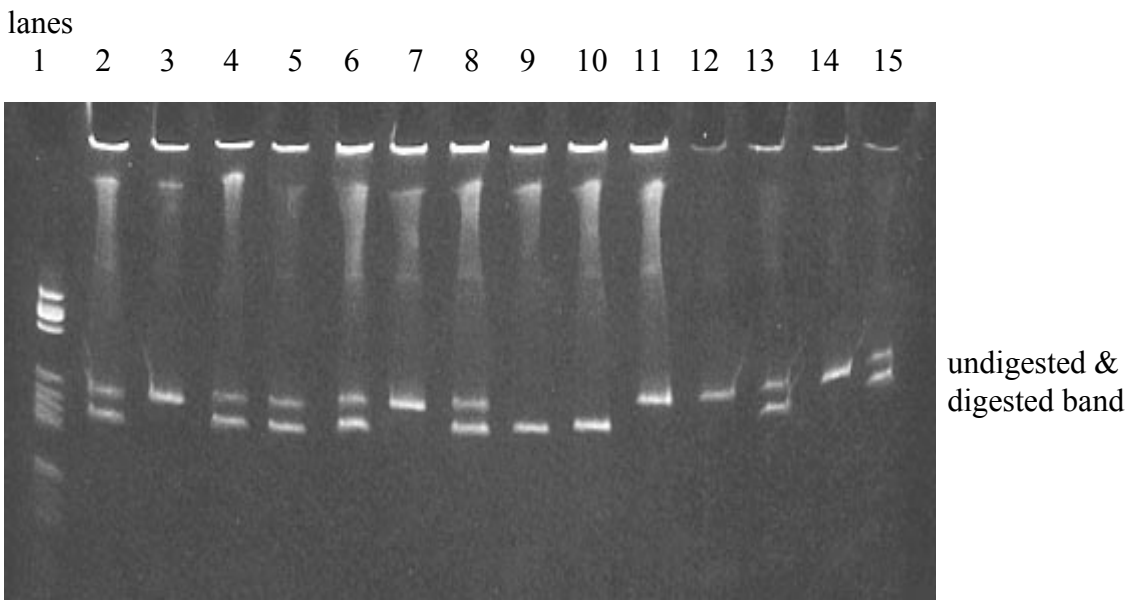
Legend for Figure 2.4b
The leftmost lane contains a molecular weight ladder. Samples #3, 6 and 7 are heterozygotes with both digested and undigested PCR product. Sample #4 shows the completely digested PCR product expected in a homozygote for the #2 allele and the other samples are all PCR product that failed to digest and are thus homozygotes for the #1 allele.

Figure 2.5 - bv3s1 pcr product digested with *Pvu*II - 10% mini-PAGE gel



Legend for Figure 2.5
The leftmost lane has a molecular weight ladder. The upper - undigested - band is seen in lanes 2 through 8 as well as in lane 10, with the sample in lane 5 being completely undigested i.e. a homozygote for the #1 allele with the others heterozygotes. Lanes 9 and 12 to 15 are homozygotes for the #2 allele as seen by the complete digestion of the PCR product.

Figure 2.6 - bv15s1 product digested with *Pvu II* - 10% mini-PAGE gel



Legend for Figure 2.6
Lane 1, on the left, has a MW marker. PCR products in lanes 2 - 15 were digested with restriction endonuclease before electrophoresis. Lanes 2, 4, 5, 6, 8, 13 and 15 are heterozygotes with half of the PCR product digested and half failing to digest. Homozygotes for the #1 allele are in lanes 3, 7, 11, 12 and 14. In these cases all the PCR product resists digestion. Lanes 9 and 10 are completely digested and thus are individuals who are homozygous for the #2 allele.

2.9 Statistical analysis

Data management and Chi square (χ^2) analysis was done using the P-Stat software package (P-Stat, Inc., Hopewell N.J.). As previous independent studies have shown bv8s1*2 to have an association with MS, the primary aim in this part of the thesis was to investigate the bv8s1 allele and genotype distribution in RR-MS and controls with the 5% significance level for this test set at $p < 0.05$. Analysis of the other markers for their allele and genotype frequencies in RR-MS and controls is done here in an exploratory sense and caution should be used in interpreting those results with $p < 0.05$ due to the large number of comparisons. The two-marker haplotype distributions involving bv8s1 were examined as the other primary aim of this study, and here a Bonferroni correction for multiple comparisons is made with $p < 0.017$ required for any particular haplotype in order to achieve 5% significance overall. Data for the additional two-marker haplotypes is treated here in an exploratory fashion and caution needs to be used as well for interpreting these results. HLA-DR15 has also been shown in previous independent studies to be a factor in the association of bv8s1 with MS. Further correction for multiple comparisons has not been made for these additional subgroups, hence these results should be treated with caution also. Allele, genotype and haplotype frequency comparisons between MS and controls were measured by Chi square analysis, as were the observed two marker haplotype frequencies where these could be assigned. Linkage disequilibrium was measured according to the formula given by Svejgaard (Svejgaard 1979b). Genotypes were used to calculate the delta value and thereby calculate an expected haplotype frequency. The expected haplotype frequency was then compared with the frequency observed after individuals doubly-heterozygous were excluded.

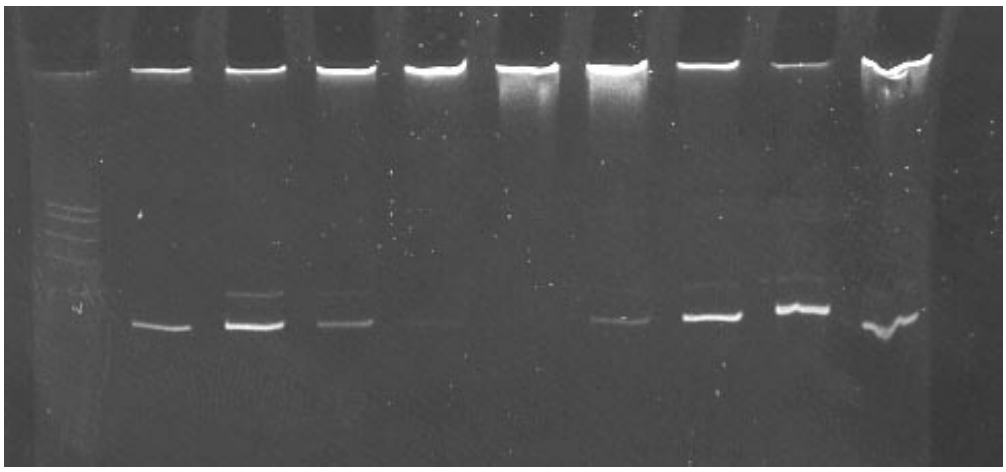
2.10 Genotyping for *CCR5-Δ32*

Standard PCR amplification of extracted DNA was performed with published primers specific for the *CCR5-Δ32* mutation (Bennetts *et al.* 1995). PCR products were resolved on 10% polyacrylamide mini-gels with ethidium bromide staining (0.5μg/ml) after 30min at 200v. Two of the first three 10% PAGE gels of CCR5 product are shown in Figure 2.7. The initial gel run turned up a *CCR5-Δ32/Δ32* homozygote (albeit in the end lane where a "smile effect" acts to lift the band, see Figure 2.7a) which is confirmed in Figure 2.7b and a second homozygous sample is seen in this gel as well.

Figure 2.7 - *CCR5-Δ32* allele genotyping by 10% PAGE of PCR products
a) Initial gel run of pcr products

Lanes:

1 2 3 4 5 6 7 8 9 10



CCR5

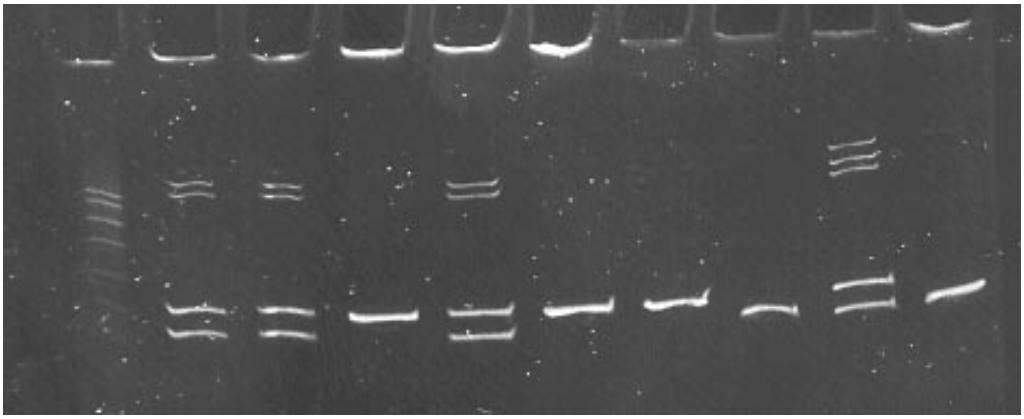
Legend for Figure 2.7 a
Lane #1 is a molecular weight standard. Lane 10 is a CCR5-Δ32 homozygous genotype affected somewhat by artifact ("smile effect" and a poorly formed sample well surface). The samples in the other lanes are all CCR5 wild-type homozygotes. The Δ32-homozygous sample is difficult to detect on this gel, but is repeated on the gel below.

Figure 2.7 (continued)

b) 3rd gel of CCR5 pcr products

Lanes:

1 2 3 4 5 6 7 8 9 10



CCR5

Legend for Figures 2.7 b

Lane #1 is a molecular weight marker. Lanes 2, 3, 5 and 9 are heterozygotes for the wild-type and *Δ32* alleles of CCR5, as can be seen by the two distinct bands. Lane #8 is a repeat of the homozygous sample from the gel shown in Figure 2.7a, now seen without a "smile effect". An additional homozygote for *CCR5-Δ32* is seen in lane 10 (again with a slight smile effect that the outermost gel lanes tend to have).

2.11 Methods for the study of microsatellite polymorphisms

The microsatellite typing was variously done on an ABI system (SupaMac, Univ. of Sydney) or "in house" on a GS-2000 fragment length analyser (Corbett Research, Mortlake, NSW) using ABI GeneScan 350-TAMRA standards (Applied Biosystems, Foster City CA, USA) alongside assorted "in house" size standards. PCR for microsatellite analysis was performed with Taq enzyme produced "in house" and a 16-buffer array of different Mg^{++} and pH concentrations was tested for optimal product for each primer set. The microsatellite PCR primers were 5'-HEX labelled on the forward primer and sequences for microsatellite markers were taken from the Genome Database (GDB; http://gdbwww.gdb.org/) or were as published (Libert *et al.* 1998; Stephens *et al.* 1998) (Table 2.2). In addition to the 3p21 markers, markers were also studied flanking the gene on chromosome 2 in the "VaHa" family (Naif *et al.* 2002). The primer sequences for these markers - D2S2215 and D2S314 - were taken from the GDB web site.

2.12 CCR5 and microsatellite typing in the "VaHa" family

As is shown in Figure 2.8, microsatellite alleles were assigned to the individuals in this family for the markers in question. Although deceased, the father's haplotypes were able to be assigned from information in the other samples. Both mother and father were *CCR5-Δ32* heterozygotes and so the two *CCR5-Δ32* homozygotes (VH and VM) were haplo-identical.

Legend for Figure 2.8 - 3p21 markers and VaHa Family haplotypes

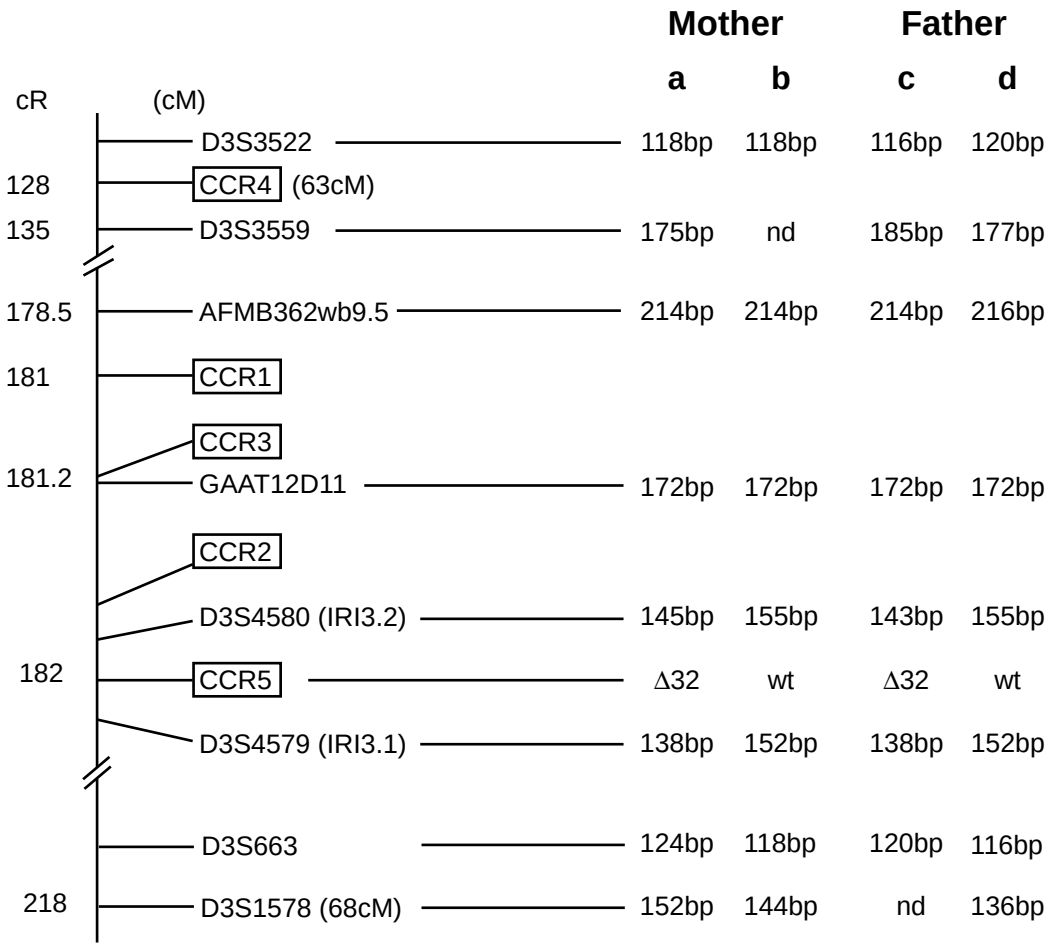
The arrangement of 3p21 markers in this figure, taken from Naif *et al.* (2002), are as shown on the NCBI "Gene Map '99" and in the report by Stephens *et al.* (1998) and so should be noted to differ from the current physical map of the genome. The cR scale is from the Stanford GB4 map where 1cR equates to about 300 kb. "?" indicates undetermined alleles. Individual genotypes are shown for the various members of the VaHa family and Figure 2.8c shows the location of the CXCR4 region microsatellites used in this family as well.

Legend for Figure 2.9 - NCBI "Build 22" map of 3p21 genes and markers

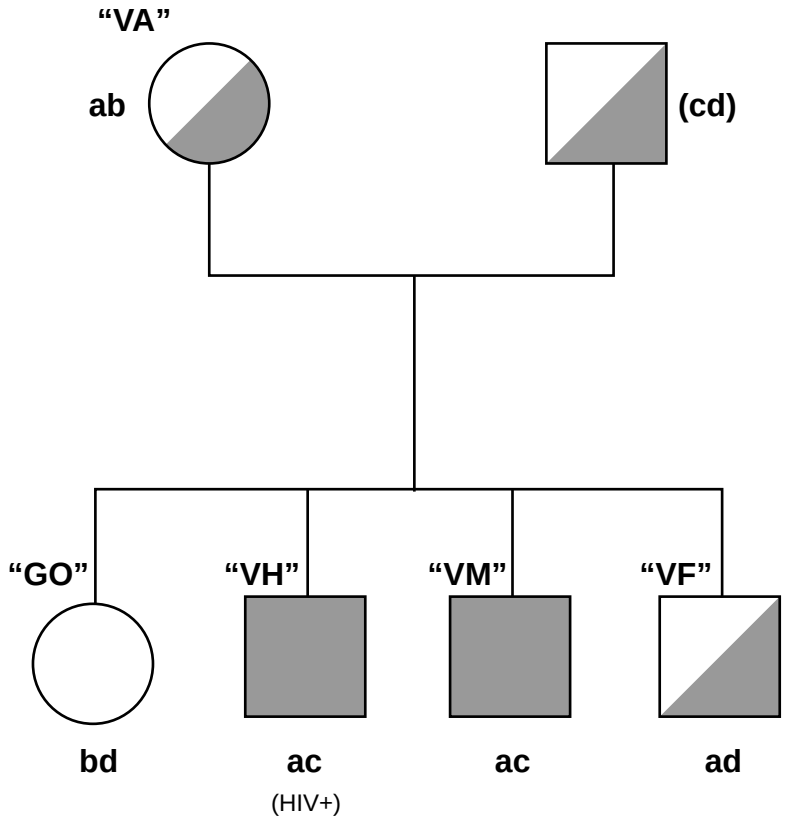
The NCBI "Build 22" map and the then current UCSC map locations for contigs and markers in 3p21 are shown. The location of contig nt_005985.2, shown by a "dashed" contig indicator on the left, was moved several Mbp to one Mbp telomeric of CCR5 in the Build 22 map but remains at a centromeric distance in the UCSC map. On the right are some marker locations from the NCBI map prior to Build 22. Shaded boxes indication the relative position of the contig nt_005985.2, which contains D3S663, on the pre-Build 22 and UCSC maps and where it would belong on the Build 22 map had it not been moved. Map positions are given in Mbp and locations are listed in kbp.

Figure 2.8

(a)


Mother
Father
a
b
c
d
cR
(cM)
D3S3522
118bp 118bp 116bp 120bp
128
CCR4 (63cM)
135
D3S3559
175bp nd 185bp 177bp
178.5
AFMB362wb9.5
214bp 214bp 214bp 216bp
181
CCR1
CCR3
181.2
GAAT12D11
172bp 172bp 172bp 172bp
CCR2
D3S4580 (IRI3.2)
145bp 155bp 143bp 155bp
182
CCR5
Δ32 wt Δ32 wt
D3S4579 (IRI3.1)
138bp 152bp 138bp 152bp
D3S663
124bp 118bp 120bp 116bp
218
D3S1578 (68cM)
152bp 144bp nd 136bp

(b)


"VA"
ab
(cd)
"GO"
"VH"
"VM"
"VF"
bd
ac
(HIV+)
ac
ad

(C)

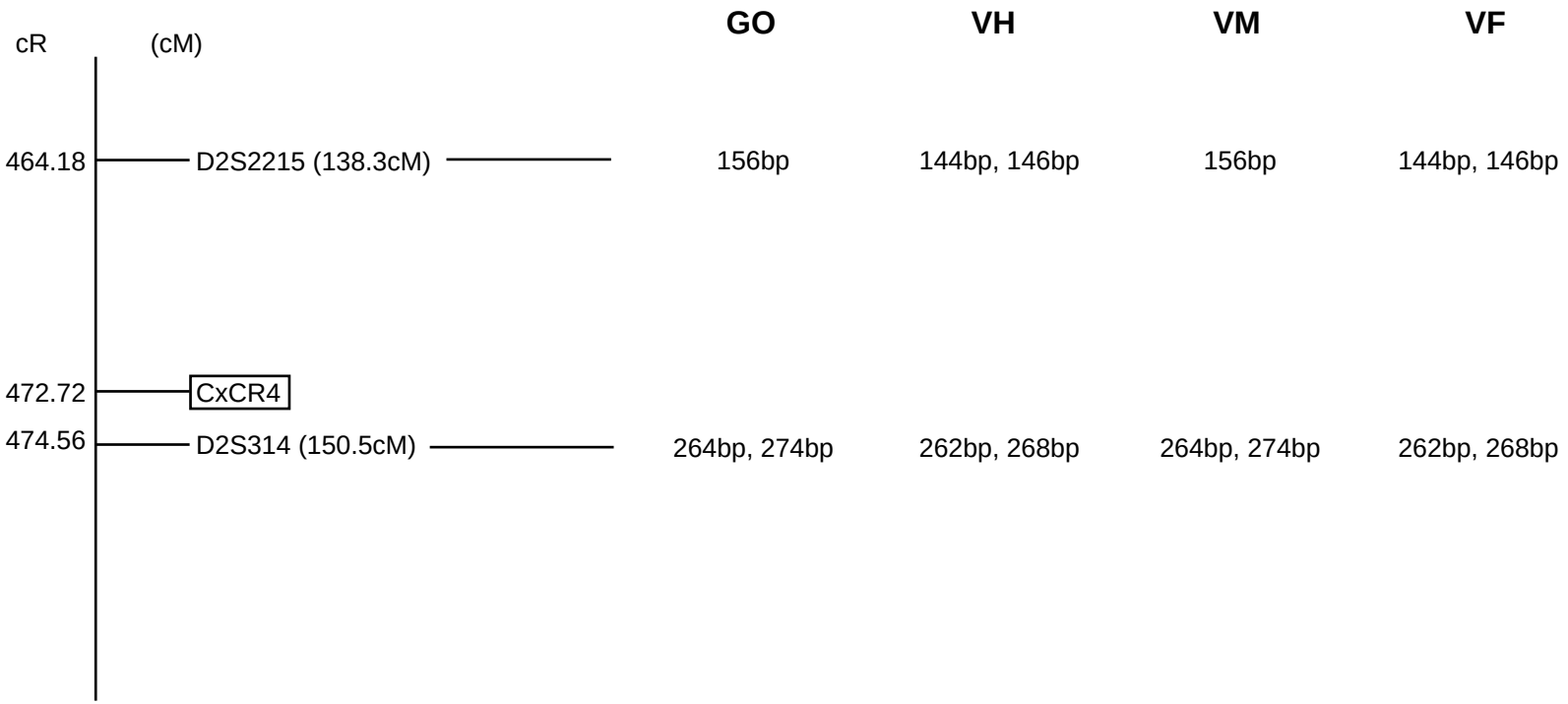

cR
(cM)
GO
VH
VM
VF
464.18
D2S2215 (138.3cM)
156bp
144bp, 146bp
156bp
144bp, 146bp
472.72
CxCR4
474.56
D2S314 (150.5cM)
264bp, 274bp
262bp, 268bp
264bp, 274bp
262bp, 268bp

Figure 2.9a

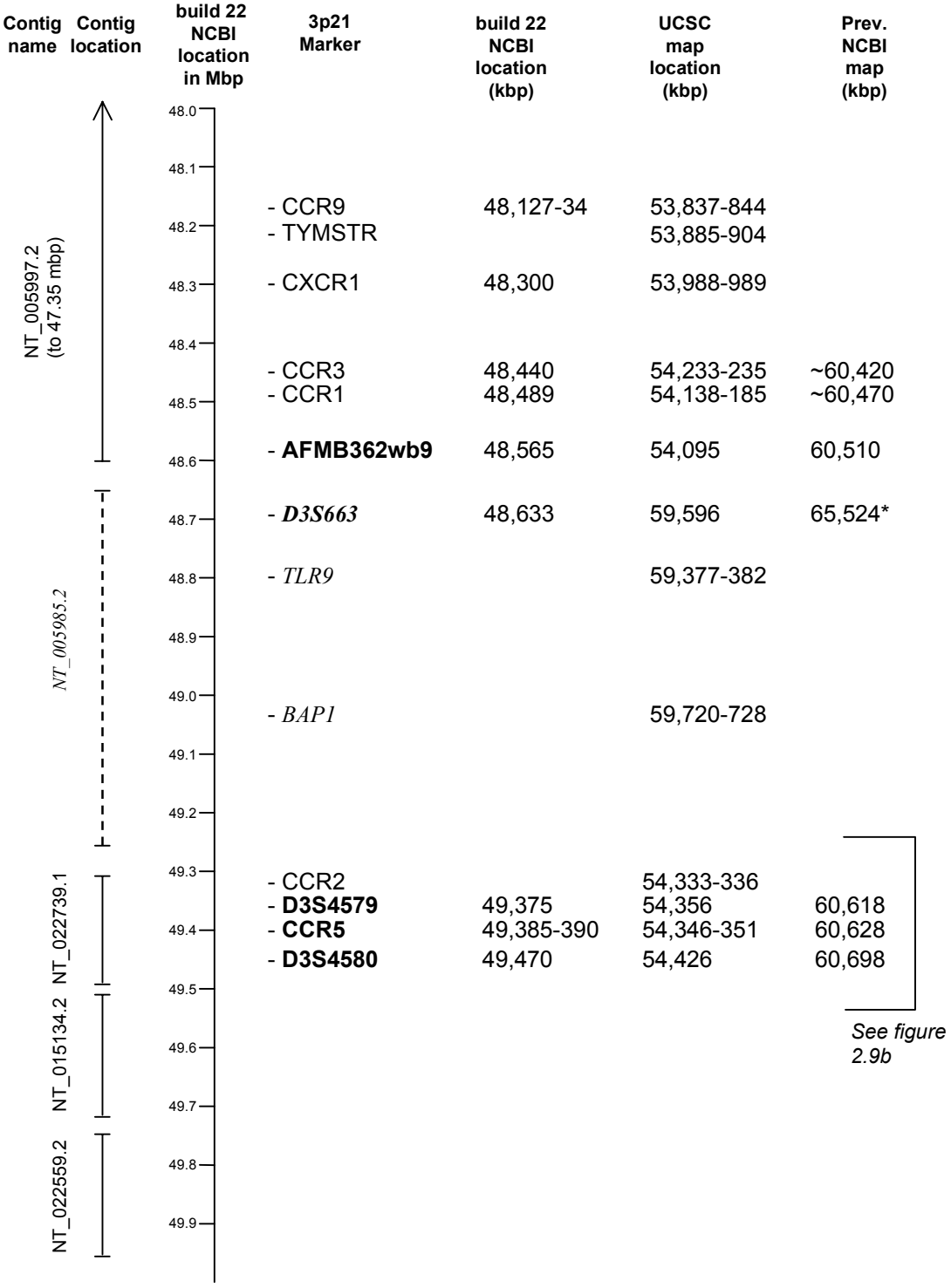
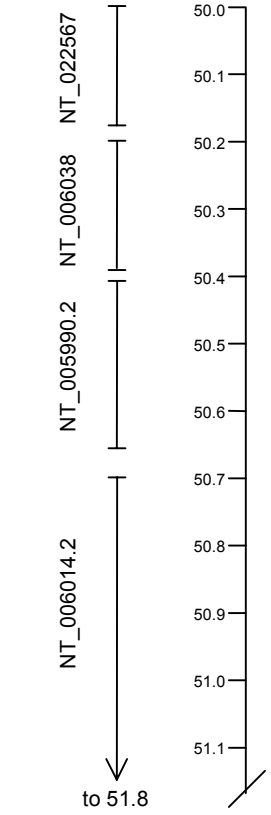

Contig name
Contig location
build 22 NCBI location in Mbp
3p21 Marker
build 22 NCBI location (kbp)
UCSC map location (kbp)
Prev. NCBI map (kbp)
NT_005997.2 (to 47.35 mbp)
NT_005985.2
NT_022739.1
NT_015134.2
NT_022559.2
48.0
48.1
48.2
48.3
48.4
48.5
48.6
48.7
48.8
48.9
49.0
49.1
49.2
49.3
49.4
49.5
49.6
49.7
49.8
49.9
- CCR9 48,127-34 53,837-844
- TYMSTR 53,885-904
- CXCR1 48,300 53,988-989
- CCR3 48,440 54,233-235 ~60,420
- CCR1 48,489 54,138-185 ~60,470
- AFMB362wb9 48,565 54,095 60,510
- D3S663 48,633 59,596 65,524*
- TLR9 59,377-382
- BAP1 59,720-728
- CCR2 54,333-336
- D3S4579 49,375 54,356 60,618
- CCR5 49,385-390 54,346-351 60,628
- D3S4580 49,470 54,426 60,698
See figure 2.9b

Figure 2.9a continued



Contig name	3p21 Marker	build 22 NCBI location (kbp)	UCSC map location (kbp)	Prev. NCBI map (kbp)
	- D3S643			63,946
	- MST1R	50,800 - 51,430		
		Calculated location of D3S663 had it not been moved ~52,300 kbp	UCSC map location of D3S663 59,596 kbp	*Previous location of D3S663 on NCBI map** 65,524 kbp
NT_005975.2	- ACY1	52,420		
	- (D3S1573)			
NT_005986.2	- **D3S1578**	53,633	61,200	68,220
NT_022552.1	- D3S1289	54,466	61,862	69,231

Figure 2.9b

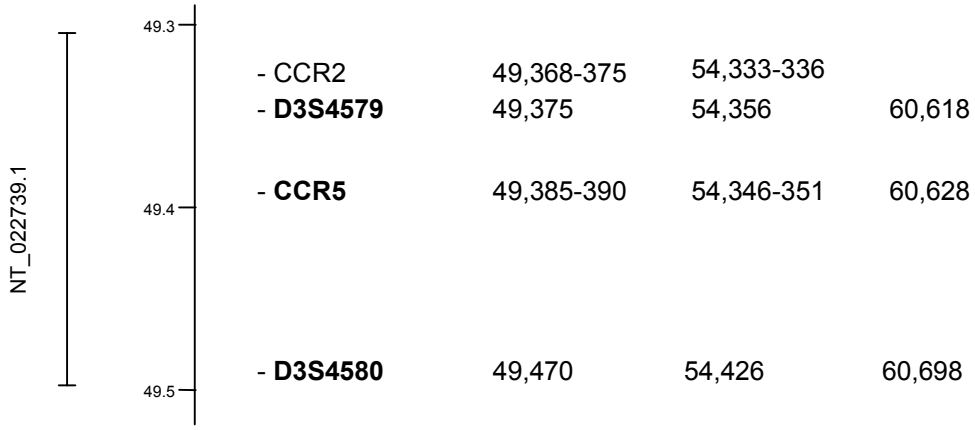

Contig name
Contig location
build 22 NCBI location in Mbp
3p21 Marker
build 22 NCBI location (kbp)
UCSC map location (kbp)
Prev. NCBI map (kbp)
NT_022739.1
49.3
49.4
49.5
- CCR2
- D3S4579
- CCR5
- D3S4580
49,368-375
49,375
49,385-390
49,470
54,333-336
54,356
54,346-351
54,426
60,618
60,628
60,698

Table 2.1
Primers for microsatellites used in the 3p21 region

marker	primer sequences	source	type / accession #
D3S663	for - CCTGGCTCTGTGAGGGACA rev - CCTGCATGGGCTTGTCTAGTC	GDB	CA repeat; D10592
D3S1578	for-GCCAACACACATTAATCACATA rev - GGGGCCAAAATCTGCT	GDB	CA repeat; Z23902
D3S4579	for - AAGAGATTGGTTCCAGGCATG rev - CCGGACCTTGCATTACAGGAC	Libert	CA repeat; 6751344
D3S4580	for - CCTTCTGGAGCAGCACTTCCA rev - GTAAATCTCCTAACAACATGC	Libert	CA repeat; 9686642
D3S643	for - GACAGAACTGCCAAACCATCCCAC rev - TATGTGCTCCAGGCTGGGTAACAG	GDB	CA repeat; D01593
AFMB362wb9	for-TGGGCAAAGGACTTAAATC rev-CACTTATTTTCTGGTGTGTGTATGT	Stephens (GDB)	CA repeat; Z67705
gaat12D11	for- GCCCCAAAAAACAATGAAGT rev- CATTCCTGTGACTGTCCCTC	Stephens (GBD)	gaat repeat; G10197
D3S3559	for - GCTCTACATCAGGCAACC rev -AAAAATAATTGGACCTGTAAAAAC	GDB	CA repeat Z52551
D3S1573	for-TTCATTTTTGCTTATTAAGATATGC rev-CCAGTAAATNACAGGGCTAT	GDB	CA repeat Z23382
D3S1289	for-AAAGCAACTTGTAAGAGAGCA rev-CTCCTAGATATAATCACTGGCA	GDB	CA repeat; Z16860

2.13 Physical map data of markers near CCR5

A physical map of the markers of interest in the 3p21 region was made using draft human genome sequence data from the NCBI web site (http://www.ncbi.nlm.nih.gov /entrez/) and this map was updated with changes in the available NCBI data. The most recent version of the physical map data is given in Chapter 5 and includes as well physical map data from the human genome project as shown on the "Genome Browser" web page at the University of California, Santa Cruz (http://genome.ucsc.edu). Figure 2.9 shows the earlier changes in the released draft human genome project data (compare with the Build 28 and 31 maps in Chapter 5).

2.13.1 An analysis of the 3p21 physical map (pre-Build 22)

The cluster of chemokine receptor genes in 3p21 which includes CCR5 is located on a sequenced stretch of DNA (GenBank ID U95626) that contains 143,068bp and was represented on the finished contig map of the public domain portion of the human genome project as nt_0022739, found between 60.57Mbp and 60.71Mbp along the chromosome. (See Figure 2.9b) The position of CCR5 is at 60,630 Kbp on these maps. The markers D3S4579 and D3S4580 (IRI3.1 and IRI3.2) are located within this sequence as well at 11kbp upstream and 68kbp downstream from CCR5. Microsatellite AFMB362wb9, used by Stephens *et al.* (1998) is found on the adjacent contig nt_005997, a very large sequence built from a dozen GenBank sequences and which is located from 60.52M to back before 59.7M. Of the GenBank sequences, AC024739 and AC026349 are at the end of nt_005997 that contains AFMB362wb9. There is a gap on the contig map of around 400Kbp between this contig sequence and the contig which contains CCR5. The location of marker chlc.gaat21d11could be in this gap as it does not

appear to be on the draft or finished contigs but was shown by Stephens et al. (1998) to lie between CCR5 and AFMB362wb9 (however, they put AFMB362wb9 upstream of CCR3 and CCR1 whereas the map data places this marker just downstream of CCR3). The Whitehead YAC map does have gaat12d11, but it appears to be just upstream of AFMB362wb9.

This release public-domain genome data has a very large gap (~8Mbp) upstream of the location of CCR3 and CCR1, with markers D3S3559 positioned at around 47.1Mbp and marker D3S3522 farther upstream at between 44.25 and 44.7Mbp (location within the draft contig nt_022417 being uncertain). This places D3S3559 at about 13.5Mbp from CCR5 on this map. An example of the D3S3559 fragment length gel trace is shown in Figure 2.10a, with two allele peaks defining a heterozygous individual. On the centromeric side of CCR5, the markers D3S1289 and D3S1578 and located at 69,231Kbp and 68,220Kbp, respectively, and so are 7590 Kbp (7.59Mbp) from CCR5 (for D3S1578; 8,601Kbp for D3S1289). The marker D3S1573 is not seen on the GenBank or contig maps, but appears on maps such as the WI-YAC, Marshfield and Genethon maps where it is located just upstream of D3S1578. A blast search of the human genome draft and finished sequence data fails to give a hit for this marker, so it is likely to be in one of the gaps between available contigs.

2.13.2 Locating D3S663 on the NIH maps

Other markers that are not in the NIH database but have been located by blast search of the human genome are D3S663, shown in Figure 2.10b, and D3S643. Using the primer sequence data for these (GDB; http://gdbwww.gdb.org/), the marker D3S663 is found on

contig nt_005985 at 65,524 Kbp, which is about 4,894 Kbp from CCR5. Slightly further upstream from D3S663 is where a human genome blast search shows the primers for D3S643 to be found, on contig nt_005990 at a location of 63,946 Kbp (3316 Kbp from CCR5). In the release of NIH map data after the initial map data was used (above), a major change was seen in the position of contig nt_005985. The position of this contig was shifted by 6 Mbp, to a position about 1 Mbp on the telomeric side of CCR5. As this location presented serious problems in interpreting the data shown in Chapter 4, an e-mail query was sent to the NIH HelpDesk citing the problem that this location made (citing the report by Stephens *et al.* (1998) both for the data on CCR5-Δ32 homozygotes). A few weeks later, however, the NIH map revealed changes again where the contig nt_005985 was returned to the relative position it had been in, and the MapViewer web page also contained the identifier “Build 22".

2.13.3 Changes to the physical map of 3p21 in Build 22 to Build 24

The GenBank sequence U95626 which includes CCR5 is located in “Build 24" of the public domain portion of the human genome project as part of contig nt_005997.4 which covers 1.58 mbp between 44,200 kpb and 45,775 kbp along chromosome 3. The position of CCR5 on this version of the map is between 45,575 and 45,582 kbp on the Build 24 gene-sequence and contig maps. Microsatellite afmb362wb9 is found on contig nt_005997.4 at 45,320 kbp and the microsatellite marker chlc.gaat12d11 is also now located within this contig at 45,673 kbp. Both of these markers were used by Stephens *et al.* (1998), whose position for gaat12d11 was between CCR5 and afmb362wb9 while their position for afmb362wb9 was telomeric of CCR3 and CCR1 (Stephens *et al.* 1998), however this physical map data places afmb362wb9 just centromeric of CCR3

and gaat12d11 actually centromeric of CCR5.

2.13.4 Genetic and physical 3p21 map data compared

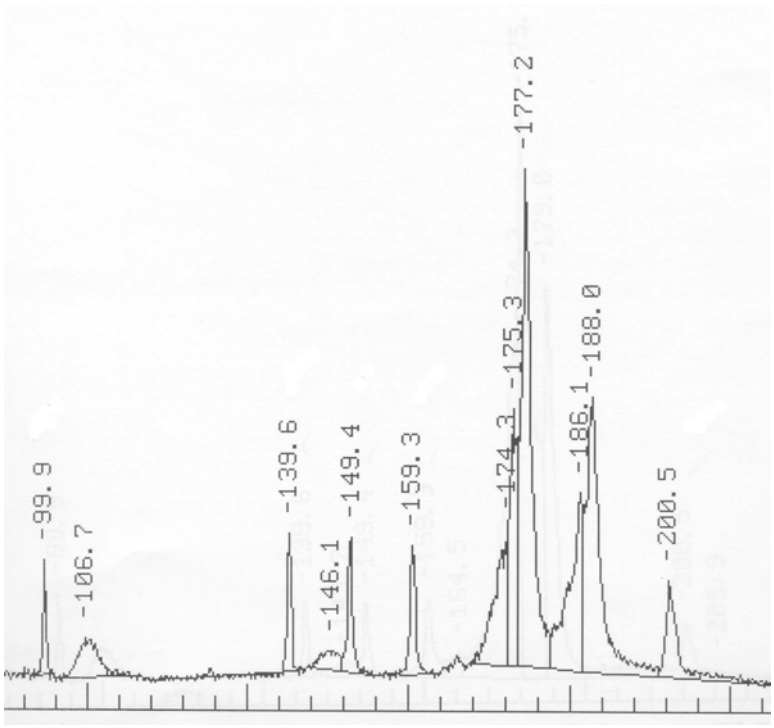
Comparison of the genetic map distance to the physical map distance for the region of interest was initially taken from the data provided by the Center for Medical Genetics at the Marshfield Clinic (http://research.marshfieldclinic.org/genetics/) (Broman *et al.* 1998; Yu *et al.* 2001) and a sex-averaged estimate over this region is approximately 0.6 cM/Mbp for the region just around of CCR5 and closer to a value of 1cM/Mbp for the region centromeric of CCR5 where D3S1578 is found. As is discussed further in Chapter Four, a recent study looked at the problem of genetic and physical map distances and a value of 0.6 cM/Mbp to 0.8 cM/Mbp is appropriate for the region from CCR5 to D3S1578 (DeWan *et al.* 2001). A blast (http://www.ncbi.nlm.nih.gov/ genome/seq/ HsBlast.html) was performed on the draft version of the human genome sequence to locate the positions of some of the microsatellite repeats not otherwise found on the NCBI MapViewer. Additionally, a search of sequence data through the UCSC Genome Browser using their "BLAT" search tool (http:// genome.ucsc. edu/ goldenPath/ hgBlat.html) was undertaken to identify the location of markers in the draft human genome sequence Oct. 7, 2001 freeze.

2.14 Dating of the origin of *CCR5-Δ32*

As will be shown in Chapter Four, the microsatellites in the region around CCR5 can be used to look for specific alleles in heterozygotes and homozygotes that are associated with the *Δ32* deletion in CCR5 and to try to define the ancestral haplotype on which this mutation first occurred. When the specific alleles are defined that were in the founder at

the point when the mutation occurred, the frequency of the marker allele still associated with *CCR5-Δ32* can be used along with information on the genetic distance from CCR5 to estimate the number of generations since the mutation event. The decay of haplotypes over different genetic distances for up to 45 generations is shown in Figure 2.2.

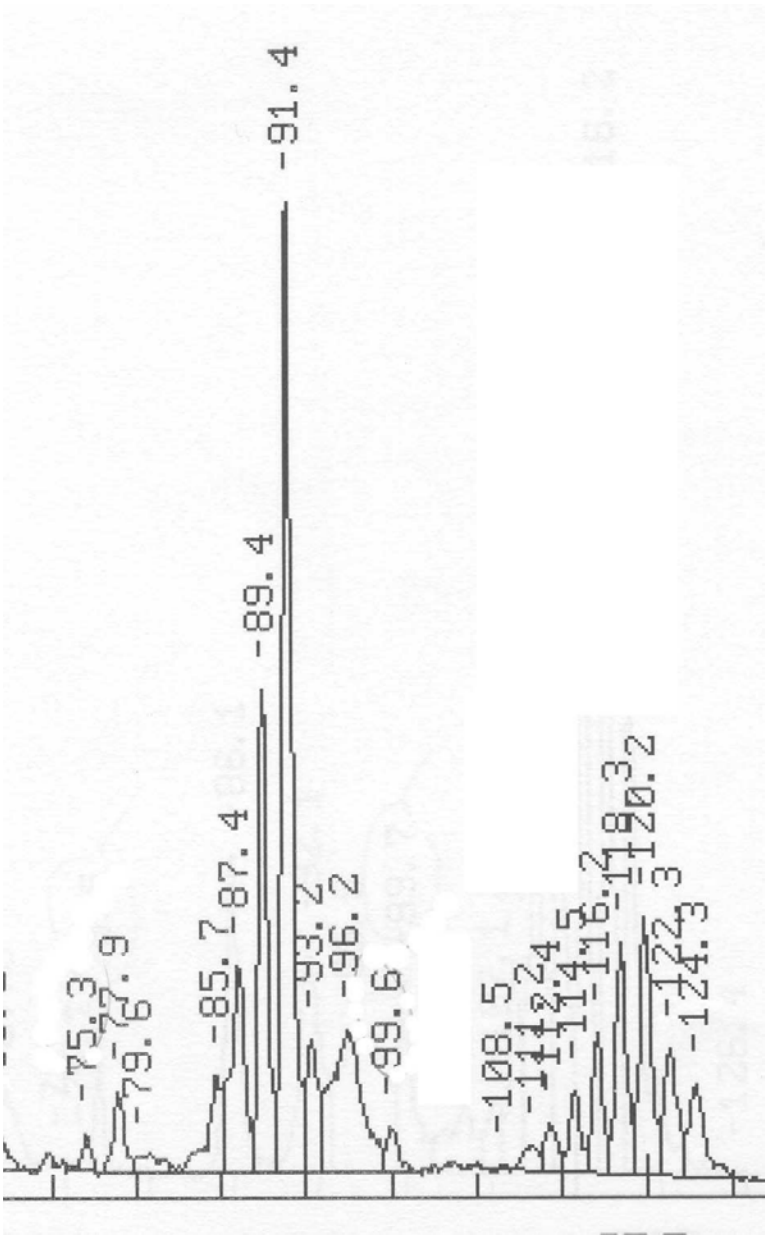
Figure 2.10 Microsatellite PCR product fragment length analysis
a) D3S3559



alleles of 177bp & 188bp

Legend for Figure 2.10a
A section of a fragment length analysis gel trace is shown with the peak heights a function of flourescence intensity. GeneScan 350 (ABI) 100, 139, 150, 160 + 200bp markers are seen here with larger peaks for the microsatellite PCR products at 177 and 188 bp in size. Note the larger size peak of the smaller sized allele which is a result of differential amplification of the smaller sized allele and is generally seen as such. One or two "stutter peaks" are seen just to the left of the main allele peaks. With D3S3559, both "odd" and "even" sized alleles can be seen together as well, as is shown here.

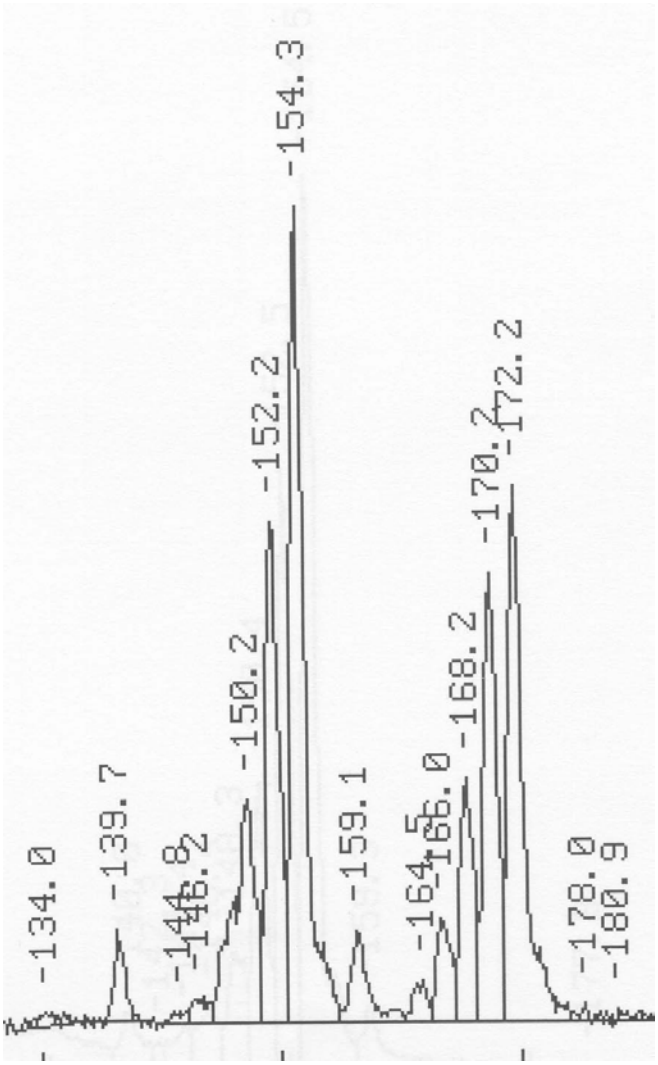
Figure 2.10b - Fragment length analysis of D3S663 product



92 and 120 bp alleles

Legend for Figure 2.10b
Multiple peaks are seen for each of the two alleles of D3S663 shown here, with one or more of the "stutter peaks" being larger in size than the main allele peak. The two very small peaks ("75.3" and "99.6" basepairs) are the 75 and 100 bp GeneScan standards and the specific allele peaks are at 92 and 120 basepairs.

Figure 2.10c - Fragment length analysis of microsatellite D3S4580



154bp and 172 bp alleles

Legend for Figure 2.10c
Small peaks are seen for the 139bp, 150bp and 160bp GS350 markers and the sample gives microsatellite peaks at 154bp and 172bp, with a couple of "stutter peaks" seen at 2bp intervals from the main allele.

Table 2.2 - Decay of haplotype frequency over genetic distance
Frequency of haplotypes intact at up to 45 generations of recombination for two markers at from 5 cM to 0.1 cM genetic distance apart (assuming no reverse crosses)

gen. #	5 cM	3 cM	1 cM	0.25 cM	0.1 cM
1	0.95	0.97	0.99	0.9975	0.999
5	0.7738	0.8587	0.951	0.9876	0.995
10	0.5987	0.7374	0.9044	0.9753	0.99
15	0.4633	0.6333	0.8601	0.9631	0.9851
20	0.3585	0.5438	0.8179	0.9512	0.9802
25	0.2774	0.467	0.7778	0.9393	0.9753
30	0.2146	0.401	0.7397	0.9277	0.9704
35	0.1661	0.3444	0.7034	0.9161	0.9656
40	0.1285	0.2957	0.669	0.9047	0.9608
45	0.0094	0.2539	0.6362	0.8935	0.956

2.15 Typing for SDF-1 in MS and controls

An additional analysis on the MS case-control cohort was performed looking at the genotype of these individuals with respect to the 3'A allele of the chemokine SDF-1 (stromal cell-derived factor-1), the ligand for CXCR4. This chemokine receptor / ligand pair is highly conserved in the vertebrates and plays an important role in development (Braun *et al.* 2002) and erythropoiesis (Majka *et al.* 2000), as well as being the alternate co-receptor for HIV beside CCR5 (Easterbrook *et al.* 1999; Ioannidis *et al.* 2001; Winkler *et al.* 1998). The data in Table 2.3 shows the results for 3'-A typing of 60 MS patients and 136 controls, and lists as well the allele frequency and carrier frequency for both groups. While slight differences are seen, they are not found to be statistically significant (although this may in part be due to the small sample size).

Table 2.3 Frequency of SDF-1-3'A in Australian MS and controls

	SDF-1 wt / wt		SDF-1 wt / 3'A		SDF-1 3'A / 3'A	
	n	%	n	%	n	%
MS pts n=60	**38**	**63**	**21**	**35**	**1**	**1.7**
controls = 136	**94**	**69.1**	**37**	**27.2**	**5**	**3.7**

MS allele frequency = 0.192 (23 alleles out of 120)
MS carrier frequency = 0.367 (22 individuals out of 60)
Control Allele frequency = 0.173 (47 alleles out of 272)
Control Carrier frequency = 0.309 (42 individuals out of 136)

Chapter Three - The TCRB locus in MS

3.1 Introduction

The basis for development of MS is thought to involve a consistent role for inherited susceptibility including epistatic interaction of several genes. This has been shown by family (Chataway *et al.* 1998; Sadovnick *et al.* 1996), adoptee (Ebers *et al.* 1995) and twin studies (Alperovitch *et al.* 1992; Ebers *et al.* 1986; Mumford *et al.* 1994; Sadovnick *et al.* 1993). Epistatic genetics in MS susceptibility has also been well reviewed (Compston 2000; Compston *et al.* 1995; Dyment *et al.* 1997; Stewart 1997). A role for MHC class II genes in MS has been well established both by population association studies and complete genome screens while candidate loci studies have produced mixed evidence for several other genes including those encoding the variable region of the T-cell antigen receptor beta chain (TCRB) (Chataway *et al.* 1998; Ebers *et al.* 1996; Haines *et al.* 1996; Sawcer *et al.* 1996). Allelic polymorphisms within TCRB gene segments have been investigated in several MS population studies. While some have shown associations of TCRB with MS (Beall *et al.* 1993; Beall *et al.* 1989; Charmley *et al.* 1991; Epplen *et al.* 1997; Hockertz *et al.* 1998; Seboun *et al.* 1989), others have failed to do so (Droogan *et al.* 1996; Fugger *et al.* 1990; Hillert *et al.* 1991; Vandevyver *et al.* 1994; Wei *et al.* 1995). Technical aspects may explain the divergent results as differences in markers studied and disease subtype both occur. The majority of studies have used the restriction fragment length polymorphism (RFLP) markers for bv8s1 and bv11s1 (Vβ8.1 and Vβ11, see Figure 3.1) while a report on a large number of Irish patients (Droogan *et al.* 1996) used a microsatellite polymorphism near bv6s7, quite a distance from the bv8s1 and bv11s1 region, as their sole TCRB marker, and failed to show any association with MS.

The MHC was the only locus showing linkage in each of the five published affected sib-pair genome screens (Chataway *et al.* 1998; Ebers *et al.* 1996; Haines *et al.* 1996; Kuokkanen *et al.* 1997; Sawcer *et al.* 1996). Weak linkage at several locations throughout the genome was seen in each of these studies, including linkage to a microsatellite marker near the TCRB locus in both the UK and Canadian screens (Chataway *et al.* 1998; Ebers *et al.* 1996). As was described in the introductory chapter, the TCRB locus is of particular interest in view of the interaction of MHC class II and TCRB in the initiation of an immune response (Oksenberg *et al.* 1988; Stewart 1997; Utz *et al.* 1993). The MHC genotype of the patients may also influence these studies (Chataway *et al.* 1998; Hockertz *et al.* 1998). Both a population study (Epplen *et al.* 1997) using microsatellite polymorphisms to type for alleles of bv6s3 and bv6s1 and a family-based control study (Hockertz *et al.* 1998) looking at several RFLP markers across the TCRB region have shown association between HLA-DR15 subgroups and TCRB alleles in MS susceptibility.

There are several aspects of the molecular biology of TCRB genes that add to the importance of the study of these genes for genetically determined autoimmune disease susceptibility. As reviewed in Chapter 1, antigen recognition by the T-cell receptor (TcR) is determined by immense "junctional diversity" created through the use of "D" and "J" segments, along with nucleotide insertion and removal, of the complementarity determining regions (CDR3 regions) between the "V" and "C" segments of the TcR alpha and beta chains. Polymorphisms of TCRB segments outside those directly involved in somatic generation of CDR3 regions do not appear to be under strong selection pressure in evolutionary terms (Jaeger *et al.* 1998), unlike the MHC where

strong linkage disequilibrium across that region makes elucidation of a specific susceptibility allele difficult (Dyment *et al.* 1997; Gruen and Weissman 1997). A noteworthy exception is the effect that malaria in Africans may have had on the frequency of a particular subhaplotype (bv8s3*1 / bv2s1*2 / bv15s1*1 / bv3s1*1) which overlaps the one studied here, where the predominance of the one haplotype may also be seen to have an effect on MS susceptibility involving groups of African descent (Craddock *et al.* 2000). An additional study to test the malaria hypothesis however found that selection was not responsible for the TCRB haplotypes seen (Donaldson *et al.* 2002a). The TCRB region shows little linkage disequilibrium apart from a couple of subregions (Charmley *et al.* 1990), notably including that which contains the bv8s1 and bv11s1 markers. An effect of the haplotype predominance in Africans is reduction on the haplotype diversity within this region (Craddock *et al.* 2000), although large differences in the genetic constitution of the TCRB locus exist between Africans and various other populations (Donaldson *et al.* 2002b).

The beta chain of the TcR has a pronounced role in the thymic selection of pre-T cells (and hence the development of the peripheral TcR repertoire), and also is responsible for the V-segment family-specific "superantigen" stimulation by a number of microbial products that can act with MHC class II in this fashion (Halle *et al.* 1997). The regulatory mechanisms of tissue-specific expression of TCRB segments (such as in the CD4+ compartment) is also now being unravelled, and in the case of bv8s1 (Vβ8.1) a complex set of enhancers are important for promoter activity (Halle *et al.* 1997), which could suggest a direct means by which alleles of this variable segment could be involved in MS susceptibility. Expression of bv3s1 at higher and lower percentages in the TcR

repertoire is also under genetic restriction (Donahue *et al.* 1994; Posnett *et al.* 1994), and again this provides one means through which certain alleles could be involved in disease susceptibility.

Some TCRB polymorphisms are within the V-segments and are in the form of bi-allelic or tri-allelic coding region polymorphisms, while others include insertion/deletion related polymorphisms and microsatellite (nucleotide repeat) polymorphisms (Rowen *et al.* 1996). Studies in the mouse have shown that allelic polymorphisms in the TCRB segments can be functionally significant with respect to T-cell repertoire (Vessey *et al.* 1996). In the case of bv3s1, above, the polymorphism exists in the recombination signal sequence spacer region, and results in loss of expression. Furthermore it has been shown, also in the mouse, that the background "self" genes can interact with MHC genes at the level of negative selection in determining T-cell repertoire (Vukusic *et al.* 1995). There are, therefore, several reasons for the TCRB region to be considered as candidate loci for MS and other autoimmune diseases.

The mixed results from the population studies of germline polymorphisms cited above, as well as the results from family studies, do not resolve the role of the TCRB region in association with MS. The differences may be due, in part, to the selection of TCRB markers, to geographic variation in environmental factors (choosing different TCRB genes) or due to lack of contribution by TCRB in all patients. Further analysis of this region is needed in well-defined MS patient and control populations. In this study, allele, genotype and haplotype frequencies for bv8s1 (Vß8.1) were examined, as were four of the additional bi-allelic coding region polymorphisms reported by Wei and colleagues

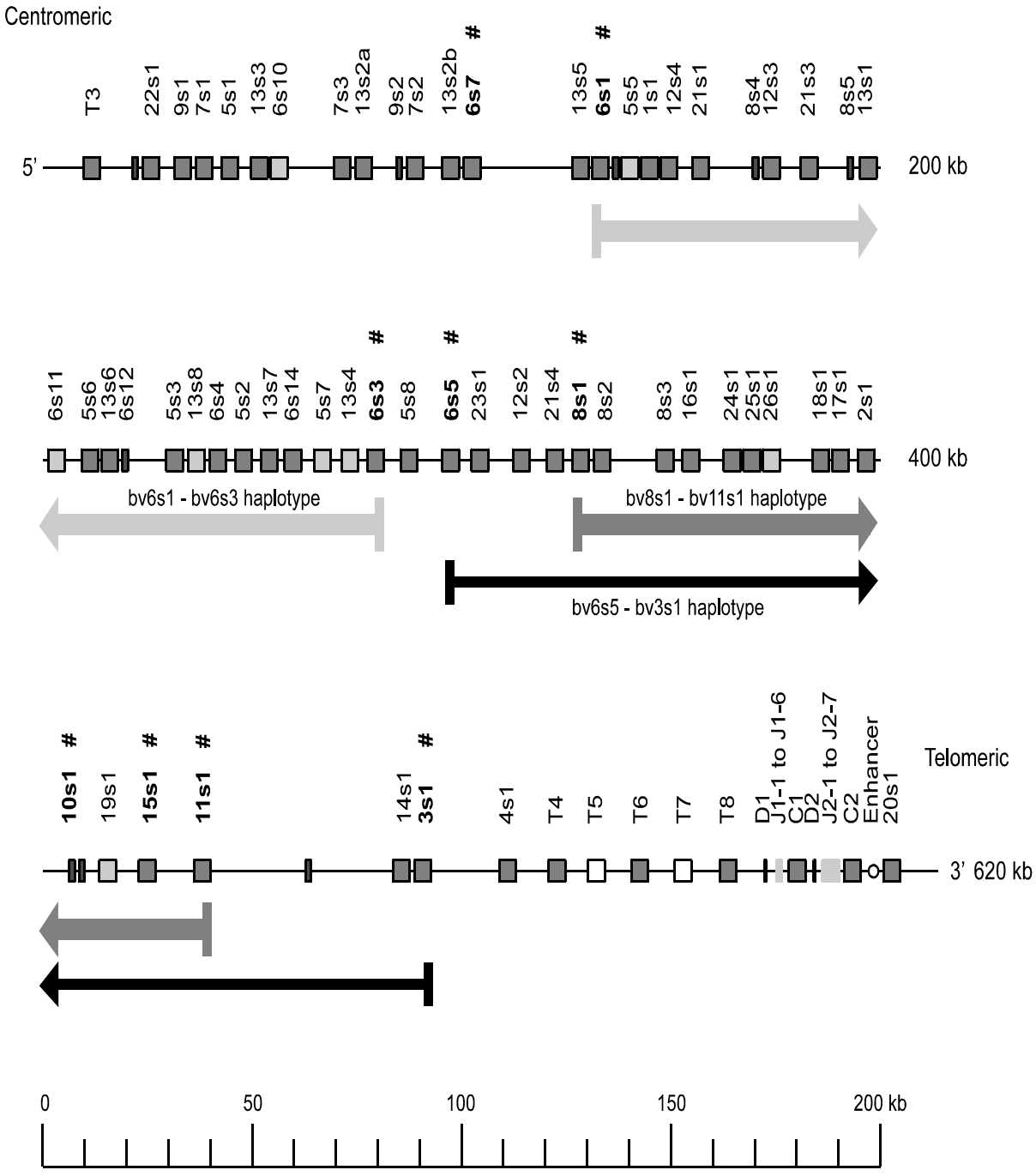
(Wei *et al.* 1995) (bv6s5, bv10s1, bv15s1 and bv3s1), based on relatively common allele frequencies and on the locations with respect to the region of suspected MS susceptibility, in a cohort of well-defined relapsing-remitting MS (RR-MS) patients from Australia who have also been HLA typed.

Legend for Figure 3.1

Location of TCRB segments and haplotypes at 7q35

Functional gene segments are shown as shaded squares, and smaller rectangles indicate pseudogenes. The diagram is adopted from that on the website of the IMGT Database (ImMunoGeneTics database, http://imgt.cnusc.fr:8104 as described in Nucleic Acids Research 1997; 25:206-211), and uses the TCRB gene designations according to Wei et al., 1994 (Wei *et al.* 1994). Markers discussed in this thesis are indicated in bold and with a "#". The haplotypes from this study are in the subregion between bv6s5 and bv3s1, as shown by the solid black bar underneath. The "Vβ8 / Vβ11" haplotype (bv8s1/bv11s1) is shown as the darker grey bar, while the bv6s1/bv6s3 haplotype (Epplen *et al.* 1997) is shown as by the light grey bar.

Figure 3.1 Location of TCRB Segments at 7q35



3.2 Results:

3.2.1 Allele and genotype frequencies

Allele and genotype frequencies are given in Table 3.1 and represented in Figure 3.2. The bv8s1*2 allele was seen to be increased in RR-MS compared with controls (p = 0.030) and this is seen to be significant statistically at the 5% confidence level. The bv3s1*1 allele, while also increased in RR-MS, approached but did not reach statistical significance (p = 0.071). In the DR15-positive subgroup (Figure 2), the bv8s1*2 allele frequency remained significantly increased in RR-MS (p = 0.050), while in the DR15-negative subgroup the associations were less apparent (bv8s1*2, p = 0.089; bv3s1*1, p = 0.057). Homozygosity for the bv8s1*2 allele was significantly increased in RR-MS compared with controls (p = 0.030). Alleles and genotypes of bv6s5 and bv10s1/bv15s1 were not seen to differ significantly between RR-MS and controls.

Legend for Figure 3.2 - TCRBV allele and haplotype frequencies in RR-MS

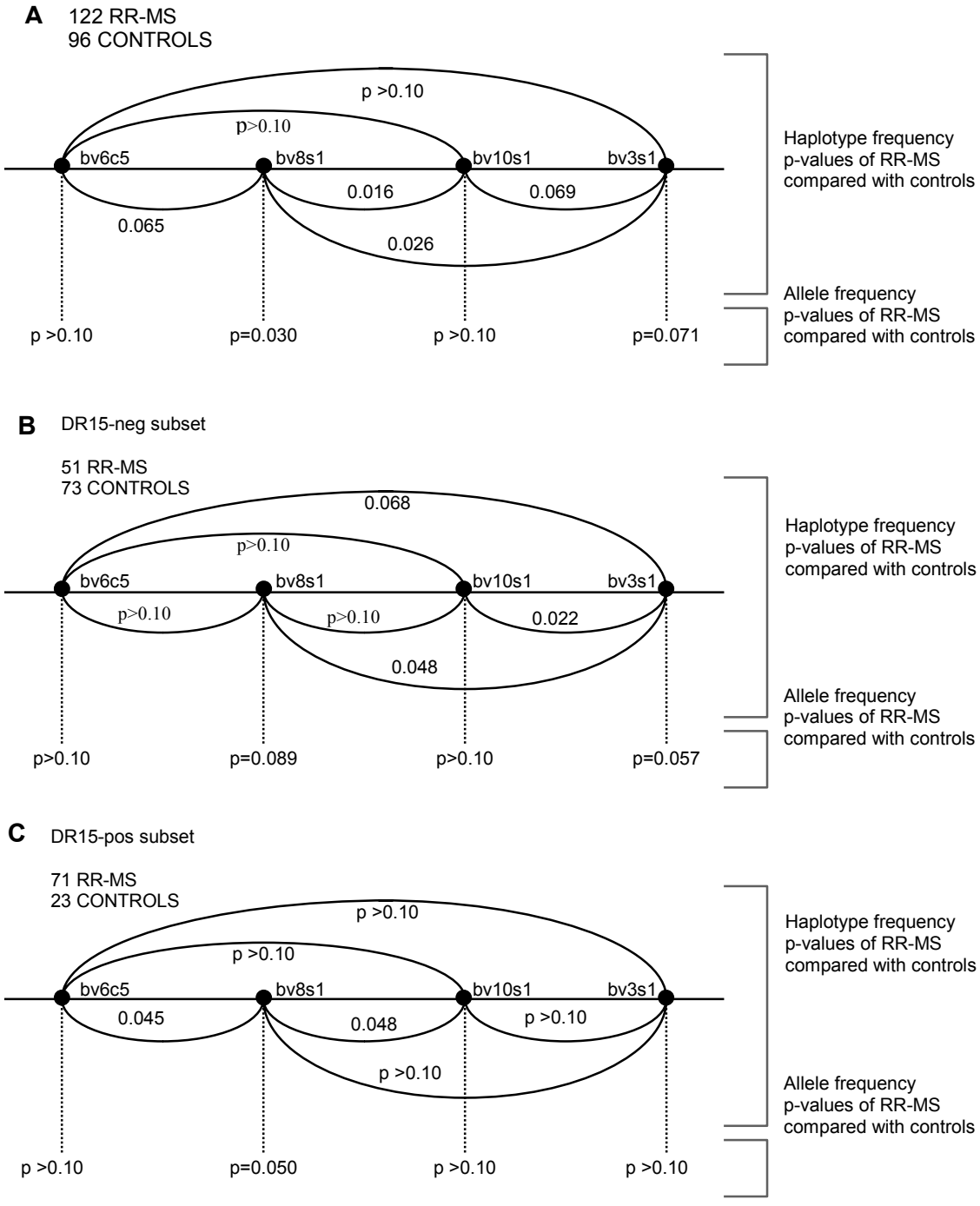
The p-values for differences in frequencies of the haplotypes formed by each pair of markers are shown as indicated by the arcs connecting the markers. Allele frequency p-values are indicated directly below each marker.

Table 3.1
Frequency for TCRB alleles and genotypes in RR-MS and Controls

TCRBV allele / genotype	RR-MS (n = 122)	Controls (n = 96)	Chi square p value
bv6s5*1	0.40	0.38	
bv6s5*2	0.60	0.62	>0.10
*1 / *1	0.16	0.13	
*1 / *2	0.48	0.51	>0.10
*2 / *2	0.36	0.37	
bv8s1*1	0.46	0.57	
bv8s1*2	0.54	0.43	p = 0.030
*1 / *1	0.24	0.29	
*1 / *2	0.45	0.55	p = 0.030
*2 / *2	0.31	0.16	
bv10s1*1	0.59	0.62	
bv10s1*2	0.41	0.58	>0.10
*1 / *1	0.34	0.38	
*1 / *2	0.49	0.49	>0.10
*2 / *2	0.16	0.14	
bv3s1*1	0.49	0.40	
bv3s1*2	0.51	0.60	p = 0.071
*1 / *1	0.24	0.14	
*1 / *2	0.50	0.53	>0.10
*2 / *2	0.26	0.33	

Figure 3.2

Comparison of allele and haplotype frequencies for TCRB markers bv6s5, bv8s1, bv10s1 (or bv15s1), and bv3s1



Note: diagram not to scale. The bv10s1 and bv15s1 genotypes are identical and shown as a single marker.

3.2.2 Frequencies for haplotypes defined by two markers

As shown in Table 3.2 and Figure 3.2, the frequency of haplotype bv8s1*2/bv3s1*1 (*2/*1) was increased in the RR-MS group overall (39% vs. 24%, p = 0.026) which, while close to the required level of p < 0.0167 for statistical significance after correction, did not achieve significance. Individual comparison of given haplotypes against the remainder (as if pooled in 2-by-2 tables) are shown by the p-values in Table 3.2. The *2/*1 haplotype of bv8s1/bv3s1 compared with the other three haplotypes between MS and controls gives p = 0.006, which is significant statistically after correction. The data for haplotypes formed by bv8s1 and bv10s1 is given in Table 3.2b, with the *2/*2 haplotype of bv8s1/bv10s1 increased in RR-MS (35% vs. 23%; p = 0.043). In the overall comparison, as given in Figure 3.2a (before pooling of any haplotypes), there was p = 0.016 and this difference between RR-MS and controls is seen in the DR15-positive subgroup (p = 0.048) but not in DR15-negative samples (p>0.10; Fig. 3.2b and 3.2c). When the haplotypes of bv10s1 and bv3s1 are considered (Table 3.2c), the overall comparison shows a value approaching significance (p = 0.069; Fig. 2a) and this is seen to strengthen in the DR15-negative subgroup (p = 0.022) but not in the DR15-positive group. Haplotypes formed by bv6s1 and bv10s1 did not show any significant associations (see Fig. 3.2).

3.2.3 Linkage disequilibrium (LD) analysis

Both the MS and control populations showed similar patterns of linkage disequilibrium. Alleles of bv10s1 and bv15s1 were associated completely and are shown in the tables and haplotype diagram as a single marker, and the bv8s1 to bv3s1 region had significant LD throughout. Weak LD was seen between alleles of bv6s5 and bv8s1, and no LD was

detectable between bv6s5 and the other markers. The expected and observed haplotype frequencies were generally in good agreement within and between the RR-MS and control groups (data not shown). The bv8s1*1-bv10s1*1 and bv8s1*1-bv3s1*2 haplotypes had delta values of 0.18 and 0.14 for RR-MS and 0.12 and 0.11 for controls, respectively. The bv10s1*1-bv3s1*2 haplotype delta value was 0.15 in RR-MS and 0.19 in controls. Similarly, the bv8s1*2-bv10s1*2 and bv8s1*2-bv3s1*1 haplotypes had delta values of 0.18 and 0.17 for RR-MS and 0.19 for both in controls, respectively. These results indicate a significant degree of LD between these loci. The haplotypes bv10s1*1-bv15s1*1 and bv10s1*2-bv15s1*2 had delta values of 0.23 or 0.24 in both groups. Observed and expected haplotype frequencies were generally in close agreement for the bv8s1, bv10s1 (bv15s1), and bv3s1 markers.

Table 3.2

Frequencies for TCRB haplotypes bv8s1/bv3s1, bv8s1/bv10s1, bv10s1/bv3s1, bv6s5/bv8s1 and bv6s5/bv3s1 in Australian RR-MS and controls and the DR15 positive and negative subsets

A.

bv8s1/ bv3s1 haplotype	RR-MS haplotype freq. DR15+ n = 71	DR15- n = 51	All n = 162	Control haplotype freq. DR15+ n = 32	DR15- n= 82	All n= 114	"p" value for each haplotype compared to other three
*1 / *1	0.10	0.08	0.09	0.16	0.07	0.10	p > 0.10
*1 / *2	0.38	0.32	0.35	0.56	0.50	0.52	p = 0.006
*2 / *1	0.34	0.46	0.39	0.22	0.24	0.24	p = 0.008
*2 / *2	0.18	0.15	0.17	0.06	0.18	0.15	p > 0.10

B.

bv8s1/ bv10s1 haplotype	RR-MS haplotype frequency DR15+ n = 96	DR15- n = 66	All n = 162	Control haplotype frequency DR15+ n = 32	DR15- n = 82	All n = 116	"p" value for each haplotype compared to other three
*1 / *1	0.46	0.39	0.43	0.63	0.51	0.54	p = 0.054
*1 / *2	0.02	0.01	0.02	0.10	0.06	0.07	p = 0.069
*2 / *1	0.14	0.29	0.21	0.07	0.19	0.16	p > 0.10
*2 / *2	0.38	0.31	0.35	0.20	0.24	0.23	p = 0.043

Table 3.2 (continued)

C. bv10s1/ bv3s1 haplotype	RR-MS haplotype frequency			Control haplotype frequency			"p" value for each haplotype compared to other three
	DR15+ n = 92	DR15neg n = 68	All n = 160	DR15+ n = 32	DR15neg n = 80	All n = 112	
*1 / *1	0.14	0.24	0.18	0.09	0.10	0.10	p = 0.057
*1 / *2	0.46	0.46	0.46	0.59	0.61	0.61	p = 0.014
*2 / *1	0.30	0.29	0.30	0.30	0.21	0.23	p > 0.10
*2 / *2	0.10	0.02	0.06	0.03	0.08	0.06	p > 0.10

D. bv6s5 / bv8s1 haplotype	RR-MS haplotype frequency			Control haplotype frequency			"p" value for each haplotype compared to other three
	DR15+ n = 104	DR15neg n = 86	All n = 190	DR15+ n = 34	DR15neg n = 98	All n = 132	
*1 / *1	0.15	0.17	0.16	0.35	0.14	0.20	p > 0.10
*1 / *2	0.20	0.21	0.21	0.09	0.14	0.13	p = 0.075
*2 / *1	0.33	0.24	0.29	0.35	0.42	0.40	p = 0.036
*2 / *2	0.32	0.37	0.34	0.21	0.30	0.27	p > 0.10

Table 3.2 (continued)

E. bv6s5 / bv3s1 haplotype	RR-MS haplotype frequency			Control haplotype frequency			"p" value for each haplotype compared to other three
	DR15+ n = 102	DR15neg n = 86	All n = 188	DR15+ n = 36	DR15neg n = 102	All n = 138	
*1 / *1	0.16	0.20	0.18	0.14	0.11	0.12	p > 0.10
*1 / *2	0.20	0.19	0.19	0.31	0.19	0.22	p > 0.10
*2 / *1	0.29	0.33	0.31	0.25	0.25	0.25	p > 0.10
*2 / *2	0.35	0.29	0.32	0.31	0.46	0.42	p = 0.076

3.3 Discussion

The T-cell antigen receptor genes are major candidates for contribution to inherited MS susceptibility given that they are involved both in the recognition of the antigen/MHC complex, and, in the case of the beta chain, earlier recognition events to determine positive selection with MHC genes (and hence the eventual makeup of the TcR repertoire) as well as the antigen-nonspecific superantigen responses. The data shown here supports an association between susceptibility to relapsing-remitting multiple sclerosis and genetic factors encoded by the T-cell receptor beta chain gene complex on chromosome 7q35, with statistically significant allele and genotype associations involving bv8s1*2. The association was further supported by the haplotypes formed by alleles of bv8s1 with bv3s1 and the tightly linked bv10s1 / bv15s1 pair of markers found between bv8s1 and bv3s1. Of importance, stratification of these RR-MS patients into the MHC subgroups positive and negative for DRB1*1501 resulted in both the subgroups showing a degree of association between haplotypes in this region and MS.

Haplotypes defined by bv8s1 and bv11s1 (Vß8 and Vß11) using RFLP were first used by Beall and co-workers to show an association in 40 North American patients with chronic-progressive MS (CP-MS), notably in the HLA-DR2 positive subgroup (Beall *et al.* 1989). Additional analysis of this group of patients, as reported by Charmley and others, with a bv15s1 (Vß15) probe (Charmley *et al.* 1991) suggested the marker for susceptibility to MS was within the Vß8-Vß11 haplotype rather than on the Cß side of Vß11 (see Figure 3.1) Lack of association between the TCRB region and MS was reported by Fugger and colleagues in a small study of 37 Danish RR-MS patients (Fugger *et al.* 1990) and by Hillert, Leng and Olerup in a Swedish MS population

consisting of both RR-MS and CP-MS patients (Hillert *et al.* 1991). Linkage studies have given mixed results with Seboun and colleagues showing segregation with the Vß8 and Vß11 markers in one group of families (Seboun *et al.* 1989) while Lynch and co-workers failed to confirm this finding (Lynch *et al.* 1991). Martínez-Naves and colleagues showed an overall association of Vß8-Vß11 (and Cß) defined haplotypes with MS in Spanish patients (Martinez-Naves *et al.* 1993) albeit not the precise haplotypes shown in other studies while no significant association was found in the study by Vandevyver and others of 71 Belgian CP-MS patients (Vandevyver *et al.* 1994), although a slight over-representation of one haplotype was noticed. In their study of 63 French MS patients (and of 48 twin pairs), Briant and colleagues reported a slight association of MS with one Vß11 allele and with a Vß11-Cß haplotype (Briant *et al.* 1993). Association between MS and other markers within the approximately 110kb TCRB region of possible MS association were sought by Wei and co-workers, who reported the frequency of alleles (but not haplotypes) of several coding region polymorphisms from a small group of North American MS patients (33 RR-MS, 9 CP-MS and 6 undefined MS), but no differences were found between patient and control groups (Wei *et al.* 1995).

The interpretation of the studies mentioned above is confounded by varying technical approaches, the use of populations from different countries, and by possible differences in the genetic basis of susceptibility to the different clinical phenotypes of MS ranging from the primary progressive form to those patients who have been classified as RR-MS for many years before showing a secondary progressive course (RP-MS). Recently, the question of TCRB association in RR-MS versus RP-MS was examined by Hockertz and

co-workers in 90 simplex and 31 multiplex families from British Columbia (Hockertz *et al.* 1998). Patients with the secondary progressive form of MS, but not those with RR-MS, showed association with the TCRB subregion studied by us. As for the allele and genotype frequencies for bv8s1, the overall patient group in the Hockertz paper and our RR-MS patients show very similar results for both allele and genotype frequencies. It is in the controls that bv8s1 is seen to differ somewhat with their *2 / *2 genotype representing 28% of their controls (Hockertz *et al.* 1998) while our study shows 16% for the same genotype. That an association may be at either side of this region is also shown by Hockertz *et al.* (1998) in RP-MS albeit with a bv8s1/bv11s1 (Vβ11) haplotype which lies within the bv8s1/bv3s1 haplotype (Figure 3.1) association found here.

The findings here support the association suggested in several populations between MS and the bv8s1*2 allele and associated haplotypes. An interaction between MHC and TCRB alleles has been noted in previous studies on MS susceptibility (Beall *et al.* 1993; Hockertz *et al.* 1998) and different TCRB associations might be expected as DR15-negative MS patients could well have a different TcR repertoire from the DR15-positive patient due to thymic positive selection and education of T-cells or to differing presentation of an autoantigen. However, a further level of complexity in the biology of HLA class II appears to have been uncovered recently wherein specific peptide fragments of HLA can act on transcription to “turn off” an immune response (Murphy and Krensky 1999). If this is the case, the role of the MHC in pathogenesis of MS might have less to do with specific antigen or peptide recognition and more to do with control over the duration of immune responsiveness.

The unravelling of the genetics of MS continues to provide a challenge in part due to the apparently relatively small effect of individual loci (apart from the MHC) and perhaps due to heterogeneity of MS aetiology and pathogenesis. Autoimmune diseases as a whole may well share clusters of candidate loci into which two thirds of the known positive linkages are seen to map (Becker *et al.* 1998), while individual autoimmune diseases will have susceptibility genes unique to that disease. In MS, it is clear that several genes with small individual effects must be involved (Chataway *et al.* 1998). It is feasible that such genetic associations may be better detected if multiple loci are considered. In this regard, using the data presented in this chapter, we examined potential interactions between TcR genotype and polymorphisms of HLA-DMB and the promoter region of HLA-DRA (Bennetts *et al.* 1999), a potential critical residue of HLA-DRB1, val/gly aa86 (Teutsch *et al.* 1999), the Apo-1/Fas promoter region (Huang *et al.* 2000), and *CCR5-Δ32* (Bennetts *et al.* 1997). With the exception of a weak association in the HLA-DRA promoter study, no associations were found. The pursuit of epistatic interactions remains important for these studies but the task remains formidable.

The data presented here, however, adds further support to an association between TCRB and susceptibility to MS across the bv8s1-bv3s1 haplotype. It may require a meta analysis of several studies of the size studied here, using uniform genetic markers, to more accurately determine the relative risk associated with this haplotype. The association seen here, while weak, does not infer a minor role for the T-cell receptor repertoire in susceptibility and pathogenesis of MS.

Chapter Four - CCR5 and autoimmune disease

4.1 Introduction

Since the discovery of the role of the chemokine receptor CCR5 in HIV infection, great interest has developed in the biology of these cell surface molecules in general and in the development of therapeutic agents capable of altering receptor expression or function (Gerard and Rollins 2001; Mackay 2001). A number of mutations with potential biological effects have been reported in the chemokine and chemokine receptor networks. Chemokine receptors provide several important functions in immune inflammation by allowing various cell types to migrate in response to the signalling by a wide range of chemokines (Wuyts *et al.* 1999), including the control of cell traffic during lymphopoiesis, involvement with antigen recognition in lymphoid tissues and control of T-cell migration in inflammation (Moser and Loetscher 2001). In this regard, individuals who differ in CCR5 receptor expression might respond in a different fashion to the initiation of autoimmune inflammation of target organs. The NOD mouse model of type 1 diabetes for example has shown a possible role for chemokine genes in diabetes (Gill *et al.* 1995) and has provided direct evidence for involvement of CCR5 in diabetes pathogenesis (Cameron *et al.* 2000). Local immune regulation of the balance of the Th1 versus the Th2 response is thought to be critical in the pancreatic tissue microenvironment at the onset of development of anti-islet immune responses leading to beta-cell destruction and insulin dependency, with marked differences in response to Th1 and Th2 signals mediated via chemokines (Bradley *et al.* 1999; Chen *et al.* 2001). Likewise, high risk first degree relatives of type 1 diabetes patients have been reported to have a “Th1-like” immune deviation in their immune responses (Karlsson *et al.* 2000).

Various analyses of tissue, plasma and PBMCs in human and animal studies have shown altered chemokine and/or receptor expression in a number of inflammatory diseases (Yoneyama *et al.* 2000). Expression of chemokine receptors and chemokines has been studied in healthy individuals over a 21-day span and this has demonstrated longitudinal stability as well as inverse relationships between Th1 and Th2-associated receptors and their ligands (Campbell *et al.* 2001). In general, Th1-type T-cells at high levels can be the cause of increased damage in autoimmunity while Th2-type cells produce "protective" cytokines and so are thought to have beneficial effects (Nicholson and Kuchroo 1996; Yoneyama *et al.* 2000), however many experimental effects show the inverse. Th2-type cells may enhance disease if administered at a certain time (Lafaille *et al.* 1997) while Th1-type cells have been reported to stop the progression of disease (Krakowski and Owens 1996). Louzoun, Atlan & Cohen (2001) address this question with a model for the influence that Th1 and Th2-type T-cells can have in autoimmune disease. They suggest that Th1 and Th2 cells are not the cause of disease but are rather a marker for a more general steady-state (Louzoun *et al.* 2001). The model is established with five different types of cell populations including naive cells, macrophages, idiotypic (Id) Th1 and Th2 cells and anti-idiotypic cells, and three type of cytokine effects (Th1, Th2 and macrophage), and is a mathematical description of findings that have been established in EAE and type 1 diabetes. Importantly, the model is that of a network where the situation is not that of a given cell type but rather the steady-state of the network. While necessary assumptions have to be made, such as all anti-idiotype cells being grouped as just one cell type, the model builds on some simple basic features, such as regulation of idiotype cells by anti-idiotype cells and competition between Th1 and Th2 cells, to bring into focus the global state of the immune system as a complex

dynamic in quasi-equilibrium rather than just a population of cells (Louzoun *et al.* 2001). Working from the two basic steady states, with the Th2 steady state assumed to be healthy and the Th1 steady state establishing an environment conducive to autoimmune disease, they examine the transition from the Th2 to the Th1 state and the effects of addition of Th1 cells, of free antigen and of IL4 with the results showing a large negative feedback loop built of four internal loops. They suggest that the progression to autoimmune diseases such as EAE and type 1 diabetes can be based on a global kinetic view with the subsystems involved being feedback loops rather than specific cell types.

4.1.1 Type 1 diabetes

Type 1 diabetes shows complex genetics, with involvement of the MHC region (Tisch and McDevitt 1996) and possibly other "core" loci indicated by studies of animal models of autoimmunity or through candidate gene studies and genomic screening in humans (Becker 1999; Becker *et al.* 1998; Merriman *et al.* 2001). In type 1 diabetes, the primary linkage found in genome screens is the *IDDM1* locus in the MHC region, variously suggested to involve HLA-DR, HLA-DQ (Tisch and McDevitt 1996) and HLA-DP (Noble *et al.* 2000). Independent modifying gene effects influencing the risk of type 1 diabetes have also been shown in the MHC region (Zavattari *et al.* 2001), as has the involvement of more than one MHC locus (Johansson *et al.* 2003). Involvement of other loci is less clear with only weak associations, apart from *IDDM1*, found previously in genome screening (Concannon *et al.* 1998; Mein *et al.* 1998). The weak effects of some loci on type 1 diabetes susceptibility make it difficult to find associations via genome screens (Owerbach 2000; She and Marron 1998; Todd 1999). A recent genome

screen of 225 multiplex families was combined with the data from two previous genome screens (for a total of 831 affected sib-pairs) to show significant evidence of linkage for *IDDM1* along with two other chromosome regions, and suggestive evidence for linkage for an additional four chromosome regions (Cox *et al.* 2001). Likewise, genome screen data on 408 multiplex Scandinavian families shows both the existence of non-HLA genes that are of significance for type 1 diabetes and interaction between HLA and non-HLA loci in determination of the diabetes phenotype (Nerup *et al.* 2001).

The prevalence of type 1 diabetes varies amongst populations with the highest frequency found in northern compared to southern Europe. A similar cline is seen for *CCR5-Δ32* with the Finns showing very high prevalence of both type 1 diabetes (Onkamo *et al.* 1999) and *CCR5-Δ32* (Martinson *et al.* 1997). To date, only preliminary reports involving small numbers of patients have appeared regarding CCR5 receptor genes in autoimmune disease. In the initial report of CCR5 genotyping in type 1 diabetes, involving only 115 children from Hungary, an association with the adjacent CCR2 region polymorphism *CCR2-64I* but not with *CCR5-Δ32* was found (Szalai *et al.* 1999); the recent report of 208 French diabetes patients showed a reduction in the age of onset of type 1 diabetes attributed to the chemokine SDF-1 while *CCR5-Δ32* was reported to be similar between patients and controls (Dubois Laforgue *et al.* 2001). Other autoimmune disorders where CCR5 has been studied include multiple sclerosis (MS) and rheumatoid arthritis (RA) (Bennetts *et al.* 1997; Garred *et al.* 1998).

As discussed in detail in Chapter 5, the high frequency of the *CCR5-Δ32* mutation in Caucasians is unexpected given the likely recent origin for the mutation suggesting

selective advantage for either heterozygotes or homozygotes (Stephens *et al.* 1998), both of which appear phenotypically normal. Resistance to epidemic infections such as plague (Stephens *et al.* 1998) or smallpox (Klitz *et al.* 2001; Lucotte 2002) has been suggested but has not been established in selection of CCR5 (with HIV-mediated selection currently underway). Protection against type 1 diabetes would appear an unlikely explanation for the increase in *CCR5-Δ32* but such protection, however slight, would be continuous and in addition to putative selection by pathogens. It is interesting to note that the highest frequencies of *CCR5-Δ32* are generally found in countries where type 1 diabetes is most prevalent and therefore able to exert selective pressure.

4.2 Results

The *CCR5-Δ32* allele status for 620 type 1 diabetes patients was examined and compared with data on 253 Australian high-school student controls and with 92 systemic lupus erythrematosus (SLE) patients. The allele and genotype frequencies are given in Table 4.1. The *CCR5-Δ32* allele frequency found in these 620 type 1 diabetes patients was 0.092 compared with 0.123 for 253 high school-age controls ($p = 0.05$). The SLE cohort had an allele frequency of 0.114 (21 heterozygotes out of 92 patients) and this was not seen to differ from the control group (note that the SLE cohort was not matched for age with the control or type 1 diabetes groups). Interestingly, there were just two individuals homozygous for *CCR5-Δ32* in the type 1 diabetes cohort compared to three of 253 controls ($p = 0.13$). Five to six such individuals would be expected in the type 1 diabetes group based on the allele frequency and the size of this cohort. This deviation from the Hardy-Weinberg distribution failed to reach significance ($p = 0.12$).

Table 4.1

Genotype and allele frequencies of *CCR5-Δ32* for 620 Australian and New Zealand type 1 diabetes patients and 253 adolescent controls

cohort	n	CCR5-Δ32 allele freq.*	CCR5 Genotypes: wild-type	wt / Δ32 (n)	Δ32/Δ32 (n)†
diabetes	620	0.092	0.819	0.177 (110)	0.003 (2)
controls	253	0.123	0.767	0.221 (56)	0.012 (3)

* Chi square p = 0.05; † p = 0.125 (Fisher's exact test)

Table 4.2

Allele frequencies compared with frequency of reported homozygous *CCR5-Δ32/Δ32* individuals for type 1 diabetes, MS, RA and SLE

Cohort	N	origin	*Δ32* freq.	*Δ32/Δ32* freq.	obs/exp†	reference
diabetes	620	Aust. / NZ	0.092	0.003	2 / 5.25	this study
diabetes	115	Hungary	0.117	0.034	4 / 1.57	(a)
MS	120	Australia	0.113	0.017	2 / 1.53	(b)
MS (fam)	302	USA	0.10*	0.010	3 / 3.02	(c)
MS (spor)	299	USA	0.14*	0.010	3 / 5.86	(c)
RA	163	Denmark	0.10	0.012	2 / 1.63	(d)
RA	673	Spain	0.058	nil	0 / 2.26	(e)
RA	278	West. Eur.	0.075*	0.007	2 / 1.56	(f)
SLE	92	Australia	0.114	nil	0 / 1.20	this study
SLE	113	Spain	0.093	0.027	3 / 0.98	(e)
SLE	50	West. Eur.	0.080*	0.020	1 / 0.32	(f)

* Calculated from the reported genotype frequencies (ref 18) or frequency of the presence of CCR5-Δ32 (27% and 19% for familial MS and sporadic MS respectively) plus the additional alleles for the reported *Δ32/Δ32* homozygotes (ref 15).
† Number of observed and expected *Δ32/Δ32* homozygotes in each study.

References: (a) (Szalai *et al.* 1999), (b) (Bennetts *et al.* 1997), (c) (Barcellos *et al.* 2000), (d) (Garred *et al.* 1998), (e) (Gomez-Reino *et al.* 1999), (f) (Cooke *et al.* 1998)

4.3 Discussion

Heterodimer formation in the cytoplasm between the wild-type and *CCR5-Δ32* alleles results in reduced surface expression of CCR5 (Benkirane *et al.* 1997), while homozygotes for *CCR5-Δ32* fail to express CCR5 on their cells and may thus differ in some aspects of immune response. The finding in this study of a reduced frequency of the *CCR5-Δ32* allele in a large cohort of 620 type 1 diabetes patients compared with 253 non-diabetic school children suggests a role for the CCR5 chemokine receptor gene in the genetic resistance to type 1 diabetes and for the chemokine network in the pathogenesis of this disease.

The reduced number of individuals homozygous for the *CCR5-Δ32* allele is of particular interest. The frequency of 1.2% found amongst the 253 controls studied here is similar to that found in equivalent European populations and less than the 2.4% recorded in Chapter 5 amongst 807 Australian Ashkenazi Jewish teenagers. The two individuals found in the 620 patients with type 1 diabetes (0.3%) is both less than that found in the controls (1.2%) and less than the number expected from an allele frequency of 0.092 (5.3) but neither difference reached significance. The low frequency of homozygotes in the general Caucasian population means a very large number of patients must be studied in order to detect a moderate reduction in frequency of people lacking CCR5 on the surface of immune cells in the disease group. If these figures are confirmed by a larger study they would represent a 75% reduction (i.e. protection) due to the *CCR5-Δ32/Δ32* genotype, a remarkable finding and a level of protection not too remote from the protection afforded individuals with this genotype with respect to HIV infection. This would require however a case control study of more than 1000 patients and 1000

controls. It would require careful matching of ethnic background and country of origin to avoid problems of population stratification given the known variation from north to south in Europe (Martinson *et al.* 1997), the high frequency in Ashkenazi Jews (see Chapter 5), and the absence of *CCR5-Δ32* amongst Asians, Africans and other non-caucasian people. The better approach is by transmission disequilibrium testing of families, an experiment justified by the data presented here and which we have commenced in collaboration with Prof. John Todd at the University of Cambridge, U.K., who has several thousand type 1 diabetes families available.

The pathogenesis of type 1 diabetes is most likely heterogeneous and that the level of expression of CCR5 may therefore be relevant in only a subset of patients. In this regard, evidence for an associated infection with enteroviruses is found in about half of all patients at the time of diagnosis (Lonnrot *et al.* 2000; Luppi *et al.* 2000). In virus-induced diabetes mouse strains, the recruitment of bystander T lymphocytes (through the chemokine network) has been reported to play a significant role in the immunopathogenic process of the disease (Vizler *et al.* 2000), however this may not be the case for all patients with type 1 diabetes. Better definition of heterogeneity in pathogenesis of type 1 diabetes, as indeed with all diseases of complex genetics, will greatly assist the discovery of susceptibility genes and the pathways in which these genes are involved.

Published studies of *CCR5-Δ32* in the autoimmune diseases type 1 diabetes, MS, RA and SLE (Table 4.2) show several differences in the allele frequency and in the degree to which homozygous *Δ32/Δ32* individuals are observed. The SLE and RA populations

reported by Gomez-Reino et al. (Gomez-Reino *et al.* 1999) are Spanish (with a control *Δ32* frequency of 0.071), and an under-representation of *Δ32/Δ32* homozygotes is seen in the RA population. However, the study by Garred et al. (Garred *et al.* 1998) reported the expected two homozygotes out of 163 Danish RA patients. While it was suggested the Western European RA cohort lacked homozygotes compared with controls (Cooke *et al.* 1998), the observed number of homozygotes was that expected from the reported allele frequency (Table 4.2). In MS, no association with *CCR5-Δ32* was seen in a case-control study and MS was observed in *CCR5-Δ32* homozygote (CCR5 deficient) MS patients in this study (Bennetts *et al.* 1997). In a combined sporadic / familial MS study, there was a delay in the age of onset of familial MS that the authors attributed to the *CCR5-Δ32* allele (Barcellos *et al.* 2000). Authors of an initial study of rheumatoid arthritis in a Danish cohort showed no difference in *CCR5-Δ32* allele frequency between patients and controls, but did suggest an effect of CCR5 on some disease variables (Garred *et al.* 1998). An effect on severity of RA was found as well in a smaller Spanish cohort where *CCR5-Δ32* carriers were seen at a higher frequency among less severely affected patients (Zapico *et al.* 2000).

The role of CCR5 (and the chemokine/chemokine receptor network) in the immunopathogenesis of multiple sclerosis remains unclear as does the contribution of genetic variability in *CCR5* to genetic susceptibility or genetic influence on disease expression. As with type 1 diabetes, the biologic plausibility is significant for both animal and human studies. Elevated levels of the CCR5 ligand, MIP-1α (CCL3) in the CNS correlates with clinical disease in a relapsing remitting model of EAE (Karpus *et al.* 1995) and administration of anti-MIP-1α in these animals prevented infiltration of

mononuclear cells into the CNS and the development of paralytic disease. Treatment with the *CCR5* inhibitor met-RANTES, (an agent already in clinical trials in HIV disease), on the other hand, showed only a modest effect on disability in chronic relapsing EAE and did not appear to affect leucocyte trafficking (Matsui *et al.* 2002). Upregulation of mRNA for each of the CCR5 ligands (RANTES, MIP-1α and MIP-1β) has been shown to occur one to two days before clinical signs (Godiska *et al.* 1995). Demyelination in C57BL/6 mice in the mouse hepatitis virus model differed greatly in *CCR5* knock out mice in a recent publication (Glass and Lane 2003). In patients with MS, increased levels of MIP-1α have been demonstrated in the CSF of patients in relapse (Miyagishi *et al.* 1995). Levels of CCR5 (and CXCR3) on CD4 and CD8 positive T-cells in blood was shown recently to correlate with T2 lesion load on MRI (Eikelenboom *et al.* 2002). In a recent report, a significant increase in surface expression of CCR5 was noted on T-cells in MS patients and increased levels of RANTES (CCL5) in CSF was observed (Martinez-Caceres *et al.* 2002). CCR5 expressing T-cells have been shown to be increased in the CSF of patients with optic neuritis (Teleshova *et al.* 2002) and increased production of RANTES by gamma delta T-cells in the CSF has been shown although this subset of T-cells expressed low levels of CCR5 in the CSF. While interpretation of these studies are far from clear they add to a growing body of data suggesting a role for *CCR5* and its ligands in MS.

The first genetic association study was published by our group in which we were not able to show an association with the *CCR5-Δ32* mutation and two *CCR5-Δ32* homozygotes were detected amongst the 120 patients genotyped; given the influence of CCR5 on T-cell function, I sought correlation between *CCR5* and TcR genotype but no

association was found (Bennetts *et al.* 1997). Subsequently Barcellos et al (2000) showed a three year delay in onset of MS in *CCR5-Δ32* positive individuals with familial disease, an affect not found in sporadic MS patients. In a small Russian study an association was shown with delayed onset of disease in *CCR5-Δ32*-HLA DR4 positive patients (Favorova *et al.* 2002). In a recent German study, a trend towards reduced frequency of *CCR5-Δ32* was seen amongst 129 patients with primary progressive disease but was not seen in 253 relapsing remitting patients (Haase *et al.* 2002). Schreiber *et al.* showed a non-significant trend towards a smaller lesion burden on MRI in *CCR5-Δ32* Danish patients (Schreiber *et al.* 2002). As with type 1 diabetes, there appears to be justification for further studies on the genetics of CCR5 and other members of the chemokine-chemokine network in multiple sclerosis and with continued stratification and correlation with other genetic markers and with disease expression.

Lack of *Δ32/Δ32* homozygous individuals is not a feature of the reported SLE cohorts, nor of the three reported MS populations (Table 2), and while here I do show a partial reduction of *Δ32/Δ32* homozygotes in a type 1 diabetes cohort, the Hungarian type 1 diabetes cohort reported by Szalai et al. (Szalai *et al.* 1999) does not. The recent report on French type 1 diabetes suggesting a role for SDF-1 (below) states that the *CCR5-Δ32* allele frequencies in patients and controls were similar, however the reported *CCR5-Δ32* frequency in their patient cohort was 5.8% (compared with 9.2% in controls) and no homozygotes were observed in their study (although one would have been expected). Regional differences in the frequency of *CCR5-Δ32* may underlie some of the differences between the results from these studies, likewise differences in environmental factors and associated pathogenic mechanisms in different regions may be contributive.

Current estimates of diabetes incidence show an increase over time across a number of population groups (Craig *et al.* 2000; Onkamo *et al.* 1999; Schoenle *et al.* 2001). Since this change is much higher than could be caused by changes in human gene frequencies, it is likely to be due to changes in environmental factors which increase penetrance of susceptibility genes to type 1 diabetes (Onkamo *et al.* 1999). Pathogens such as viruses can act as initiators of autoimmunity (Jun and Yoon 2001; Oldstone 1997) and would have the capacity to alter the observed penetrance on the time scale involved. Conversely, in SLE, CCR5 might not be expected to have the role seen in RA, MS and type 1 diabetes, because the dependency on an influx of inflammatory cells in a tissue-specific fashion in SLE could involve a different chemokine receptor, e.g. CCR10 (Homey *et al.* 2000). Taken together, it does appear that *CCR5-Δ32* may play a role in the outcome of various autoimmune diseases, including type 1 diabetes, in Caucasian populations (where this allele is found), and that partial protection in type 1 diabetes could result from heterogeneity in pathogenesis. Such heterogeneity is supported by the proposed role of enteroviruses and other infectious agents as triggers for type 1 diabetes (Oldstone 1997), as discussed. A sizeable number of polymorphisms have been reported in the chemokine- chemokine receptor network, several of which have clear biological effects as seen in HIV disease (Carrington *et al.* 1999; Easterbrook *et al.* 1999; Magierowska *et al.* 1999; Michael 1999). An international meta-analysis has confirmed delayed onset of AIDS and delay of death from AIDS associated with CCR5-Δ32 and with an allele of the closely linked CCR2 gene, *CCR2.64I* (Ioannidis *et al.* 2001). A more rapid progression of HIV disease, however, has been shown to be associated with polymorphisms in the promoter region of CCR5 (Carrington *et al.* 1999; Easterbrook *et al.* 1999) and the protective *CCR2.64I* mutation is in strong linkage disequilibrium with

a CCR5 promoter allele (Stewart 1998). Further evidence of the role of CCR5 in AIDS comes from the over-representation of *CCR5-Δ32* in long term non-progressors (Stewart *et al.* 1997) and the under-representation of the deleterious promoter polymorphisms (Clegg *et al.* 2000). An initial report of an association with a polymorphism in stromal cell derived factor 1 (SDF-1) (Winkler *et al.* 1998), the ligand for CXCR4, and HIV disease progression, has not been confirmed in the meta-analysis (Ioannidis *et al.* 2001). *CCR2.64I* has been reported to be associated with type 1 diabetes, as mentioned above (Szalai *et al.* 1999), and the recent study of the SDF-1 polymorphism suggested an association with the early onset of type 1 diabetes (Dubois Laforgue *et al.* 2001). Associations with the CCR5 promoter polymorphisms have not yet been reported in autoimmune disease, let alone in type 1 diabetes, but if *CCR5-Δ32* is protective in type 1 diabetes, other alleles which also delay HIV infection may likewise have a similar effect.

Taken overall, these findings support the further investigation of a broad range of chemokine receptor polymorphisms in type 1 diabetes. Such future work should involve familial analysis based methods to avoid the problem of population stratification given the known variability in *CCR5-Δ32* frequency within Caucasian populations and between the major racial groups (Martinson *et al.* 1997). Such an endeavour involves a very sizeable amount of work but appears justified by the current data and by the potential target for novel therapy presented by CCR5, and other chemokine network molecules, involving agents currently in early clinical trial for HIV disease. Definitive demonstration of a partial protective effect in homozygotes for *CCR5-Δ32* against type 1 diabetes will require genotyping of a large number of families.

Chapter Five - The question of the origin of the *CCR5-Δ32* allele

5.1 Introduction

With the exception of the monogenic disorders, there are very few single mutations with a greater effect on human disease than the 32-bp deletion allele of the *CCR5* gene, *CCR5-Δ32*. Infection by HIV has been shown to be dependent on the use of cell surface receptors with macrophage-tropic strains using CCR5 and T-cell tropic strains of the virus using as the necessary co-receptors (Deng *et al.* 1996; Dragic *et al.* 1996), along with CD4. This use of chemokine receptors by HIV may be a result of the incorporation of a gene for a chemokine receptor ligand by the common ancestor to SIV and HIV (Lusso 2000; Shimizu and Gojobori 2000). Homozygotes for *CCR5-Δ32* fail to express CCR5 on the cell surface (de Roda Husman *et al.* 1999; Samson *et al.* 1996b) and heterozygotes have a substantially reduced expression. Homozygotes are very highly protected against HIV infection. This protection was initially considered to be absolute until the report from our laboratory of a HIV infected *CCR5-Δ32* homozygote (Biti *et al.* 1997). This individual was subsequently shown to be likely to have been infected via CXCR4 (Naif *et al.* 2002), a study to which I contributed as detailed elsewhere in this thesis. Heterozygotes for *CCR5-Δ32*, who progress significantly more slowly to AIDS and death (Dean *et al.* 1996; Samson *et al.* 1996b), are found in excess amongst individuals who are long term non-progressors of HIV disease (Stewart *et al.* 1997). Several polymorphisms, including those in the co-located *CCR2* gene and the promoter region of *CCR5* have been shown to influence disease progression (Clegg *et al.* 2000; Stewart 1998). Further, our laboratory and others have carried out studies that have indicated a broader impact on the clinical expression of HIV disease with effects on the pattern of opportunistic infection in patients with AIDS (Ashton *et al.* 2002), the development of dementia and their response to Interleukin 2 therapy (Clegg *et al.* 2003).

There is also evidence suggesting an effect of *CCR5-Δ32* mutation on susceptibility to Hepatitis C infection (Woitas *et al.* 2002) although there is debate about this finding (Promrat *et al.* 2003). These effects most likely reflect the emerging role of CCR5 in the immune response and are unlikely to involve it's HIV co-receptor function.

On this background there is another feature of the *CCR5-Δ32* mutation that is perhaps even more remarkable and which has intrigued me to a great extent. Data published on microsatellite markers close the CCR5 gene (Libert *et al.* 1998; Stephens *et al.* 1998), and those presented in this thesis, establish that this mutation occurred in a single founder very recently in human evolution, certainly within the last 2,000 years and most likely sooner. Consistent with this is the restriction of this mutation to Caucasians with no evidence, beyond admixture, for its presence in Asians or Africans (Libert *et al.* 1998; Lucotte 1997; Lucotte 2002; Martinson *et al.* 1997; Stephens *et al.* 1998). Despite its recent origin, *CCR5-Δ32* is found on average in one fifth of all Caucasians, a frequency which appears explainable only on the basis of strong natural selection. To further highlight this, it has been estimated that a period of approximately 250,000 years would be required for a mutation with a single founder to reach this population prevalence by genetic drift alone (Kimura and Ohta 1973; Libert *et al.* 1998; van Herwaarden and van der Wal 2002) and yet *CCR5-Δ32* did so in a very short time span in human history. With respect to other possible forces, we have considered the possibility of insulin dependent diabetes which mostly effects individuals prior to the age of reproduction and, before the advent of insulin therapy, had a high mortality rate. This in part led to the study described in the previous chapter (Buhler *et al.* 2002).

In the absence of direct data, hypotheses on the likely selection forces rely on an understanding of the effects of CCR5 in health and disease. It equally relies on knowledge of the likely point in human history at which the mutation arose, the population in which the founder resided and the likely influences on human development between that time and now. In view of its emergence in the last century, HIV is unlikely to have been involved. Looking forward, it has been estimated that another protective *CCR5* allele will rise significantly in the African population as a result of the HIV epidemic in that continent (Schliekelman *et al.* 2001), supporting the concept that epidemics of young people can have measurable effects over a relatively brief period of human evolution. As will be discussed later in this chapter, other lethal infectious epidemics are worthy of consideration (Klitz *et al.* 2001; Lucotte 2002). The dating of the Δ32 mutation is critical in this regard. As a result of their calculation that a mutation occurred about 700 years ago, Stephens *et al.* (1998) postulated that the selective event was the Black Plague. Their calculation of 700 years was based on the measurement of the likely number of generations for the observed recombination rate between the *CCR5* gene and two closely located microsatellite markers (and based on a map order that differs from the current genome maps). These results differ from an estimated likely 2,000 to 3,000 year age based on two other microsatellites in a subsequent study (Libert *et al.* 1998). Dating of the mutation is not an exact science. It relies on several assumptions including, most critically, the recombination distance between the microsatellite marker and the *CCR5* gene. Expected variation can be reduced by genotyping for a larger number of markers than two. More importantly, in the five years since these reports, there has been substantial refinement of the knowledge of the physical and genetic distance within this region on chromosome 3p21. In this

chapter I report dating of the *CCR5-Δ32* mutation based on six microsatellite markers, including the four previously published and using information from the human genome map that indicates likely errors in the estimates from the original published reports. While it is acknowledged this is still not an exact science, I do believe that the estimates provided in this thesis provide a range of potential age which is easily contained within the period of recorded human history and which allows the development of reasonable hypotheses.

Hypothesis building also requires an understanding of the region of the world where the putative selective forces would have been acting, both initially and over subsequent years. This in turn raises the question of the geographic place of origin of the founder. This is best approached by identifying the current modern population with the highest frequency of the allele and the history and migration of these peoples over the years since the mutation first occurred. These data are then interpreted in the light of potential selective forces on the involved populations. With the exception of the dark ages (the period between the fall of Rome and the early part of the second millennium A.D. which generally lacked a recorded history), much of the potential force of natural selection on *CCR5-Δ32* is contained within European recorded history.

The first analysis of the global distribution of *CCR5-Δ32* was published by Martinson *et al.* (1997). This confirmed the Caucasian restriction of the mutation and indicated strongly a lower frequency amongst southern Europeans with highest allele frequencies found in countries of north-eastern Europe. Our attention was caught, however, by the very highest frequency found in a small cohort of 43 Ashkenazi Jews. We had access to

a very large DNA bank of Australian teenagers of Jewish origin collected for the purpose of analysis of genetic risk of Tay Sachs disease. In addition we had detailed information on the country of origin of their parents and grandparents. It is the study of these individuals that forms a major part of the data presented in this chapter.

Subsequent population studies have supported a northeastern European origin for the *CCR5-Δ32* allele (Libert *et al.* 1998; Magierowska *et al.* 1998; Martinson *et al.* 2000; Stephens *et al.* 1998; Voevodin *et al.* 1998). Other reports on *CCR5-Δ32* in Ashkenazi Jews have shown a wide range of allele frequencies for this group (Kantor and Gershoni 1999; Lucotte *et al.* 1999; Martinson *et al.* 2000). A large cohort of American Ashkenazi Jews (n = 503) was shown to have an allele frequency of less than 10% (Stephens *et al.* 1998). A much higher frequency (17%) was described for an Easter European Ashkenazi cohort (Lucotte *et al.* 1999). Israeli Ashkenazi Jews were reported to have an allele frequency just above 10% and to have acquired this allele through intermarriage (Kantor and Gershoni 1999). A more recent survey of the allele frequency difference between Sephardic and Ashkenazi Jews in Israel (Maayan *et al.* 2000) had a *CCR5-Δ32* frequency of 13.8% in their Ashkenazi cohort with a suggestion that admixture and positive selection of *CCR5-Δ32* in the last thousand years must account for the observed frequency. Neither of the reports looking at the date of origin of the *CCR5-Δ32* allele has specifically focussed on European Ashkenazim for the possible source of *CCR5-Δ32* (Libert *et al.* 1998; Stephens *et al.* 1998).

Lucotte has suggested that the movement of *CCR5-Δ3*2 around Europe may have involved the trade routes used by the Vikings in the eighth to tenth centuries, which

would account for the predominantly Northern European distribution of this allele (Lucotte 2001). In the initial global survey of *CCR5-Δ32*, the frequency found in Iceland (14.7%) was second to the 20.9% allele frequency reported for the Ashkenazi Jews (Martinson *et al.* 1997). Poser has looked at the movement of the Vikings with respect to the possible flow of genetic susceptibility to multiple sclerosis (Poser 1994; Poser 1995), and describes the eastward movement of the Swedish Vikings to the Russian rivers and then down to the region between the Black and Caspian Seas to the Kingdom of Khazaria from where the Vikings were able to trade with the Arabs. The kingdom of the Khazars, now lost to history, converted to the Jewish faith during the latter parts of the ninth century and effectively halted the northward movement of Islam in their time (Brook 1999; Koestler 1976; Noonan 1998). The Arab - Khazar wars blocked trade northwards for a century before the documented flow of coinage in the form of silver Arab dirhams eventually led the Vikings southwards and into a two century long (800 AD to 1000AD) trade relationship with Arabia via the Khazar kingdom (Noonan 1998). A number of generations later, movements of the Ashkenazi Jews from the German regions occurred into these more Eastern European regions as the Jews were expelled from the German lands (Shamir and Shavit 1986) and then later the expulsion of both the Germanic and Khazar Jews from the Russian region saw the main Jewish state formed in the region of Poland (Hundert and Bacon 1984; Shamir and Shavit 1986).

The complex interplay between the populations and their migrations as indicated above and the selective forces on the *CCR5* mutation provide a difficult challenge for the identification of the origin of the mutation. This can be assisted, however, by the identification of the ancestral haplotype on chromosome 3p21 upon which the mutation

occurred and then examining the prevalence of that haplotype on chromosomes lacking the mutation (wild-type). Accurate haplotype assignment is assisted by the study of homozygotes; I was fortunate in identifying 26 *CCR5-Δ32* homozygotes in our DNA bank for this purpose. This enabled the identification of a haplotype and examination in a sizeable number of individuals, Jewish and non-Jewish, of the frequency of that haplotype on chromosomes with and without *CCR5-Δ32*.

Here, I show evidence that the *CCR5-Δ32* allele has a likely origin in the Ashkenazim of eastern Europe, possibly originating from the Khazar kingdom or their ancestors 1,000 to 1,500 years ago. These findings, if correct, assist development of hypotheses on one of the most remarkable examples of the effects of natural selection on the genetic makeup of modern day people of European origin, one fifth of whom carry the *CCR5-Δ32* mutation. This mutation occurred in a gene that, in addition to its role as a viral co-receptor, encodes a protein of emerging broad importance to the immune system and therefore potentially to the human response to a range of diseases, infectious and non-infectious, including the possibility of life-threatening autoimmune disease.

5.2 Results

5.2.1 Population frequencies of the *CCR5-Δ32* allele

The CCR5-Δ32 genotype and allele frequencies on 1388 individuals comprising Jewish, non-Jewish Caucasian, and asian / pacific Australian populations are given in Table 5.1. Individuals of Jewish background are subdivided into Ashkenazi (n = 807), Sephardic (n = 35), or “other” (n = 104) which consists of individuals of “mixed” background or who were unsure of their status.

This study confirmed the extremely uncommon occurrence of the *CCR5-Δ32* allele in people of non-Caucasian origin with only three of 262 chromosomes in 131 people of Asian or Pacific Island background having the mutation. The allele frequency of 0.011 was significantly different from the allele frequency among 311 non-Jewish Caucasian Australians ($p = 0.0001$). This supports the few other studies addressing this issue.

The results also confirm the low frequency of the *CCR5-Δ32* allele amongst Sephardic Jews at 0.057, significantly lower than that found among Ashkenazi Jews ($p = 0.001$) and non-Jewish Caucasian Australians ($p = 0.01$). The allele frequency of 0.146 for the 807 Ashkenazi Jewish Australians differs significantly from that found in 311 non-Jewish Caucasian Australians ($p = 0.03$). Although this p value would not remain significant if corrected for the number of comparisons made (Bonferroni correction), such correction is conservative and, given that the finding in the Australian Jewish population replicates the higher allele frequency seen in other studies, it is reasonable to consider the uncorrected p value. The difference between these two populations is reflected in a high prevalence of homozygotes with 1% of non-Jewish Caucasians bearing this genotype compared with 2.4% of Australian Ashkenazi Jews. This difference is only of borderline significance ($p = 0.08$). The genotype distributions in each of the population groups are in Hardy-Wienberg equilibrium. Since this is one of the largest reported studies, this further supports the lack of evidence for altered survival due to CCR5 genotype, particularly *CCR5-Δ32* homozygosity.

Table 5.1

***CCR5-Δ32* genotype and allele frequency for Australian Jewish and non-Jewish subgroups**

subgroup	CCR5 wt/wt n	(%)	CCR5 Δ32/wt n	(%)	CCR5 Δ32/Δ32 n	(%)	total	allele freq.
Ashkenazi Jewish	591	73.2	197	24.2	19	2.4	807	0.146*
Sephardic	31	88.6	4	11.4	--	--	35	0.057
"other Jewish"	83	79.8	19	18.3	2	1.9	104	0.111
non-Jewish Caucasian	244	78.5	64	20.6	3	1.0	311	0.113*
Asian / Pacific	128	97.7	3	2.3	--	--	131	0.011
total	1077	--	287	--	24	--	1388	

* Chi square test of the difference in *CCR5-Δ32* frequency between Ashkenazi Jewish individuals and non-Jewish caucasians gives a significant chi square of 4.166 ($p < 0.05$)

5.2.2 The *CCR5-Δ32* frequency varies with the country of origin

A large proportion of Australian Jews migrated to Australia after the Second World War. Of these, all of their grandparents and most of their parents were born elsewhere and they have clear knowledge of the country of origin of their ancestors. Birthplace data was available from over 90% of individuals for themselves, their parents and grandparents. Inferences can be made from the birthplace data to show a gradient in *CCR5-Δ32* across European Ashkenazim.

Table 5.2 shows the genotype and allele frequencies of CCR5 on Jewish teenagers separated on the basis of birthplace of the maternal grandmother (so chosen as judeity

follows the maternal line). A gradient between allele frequencies in Western and Eastern Europe can be seen in the difference in *CCR5-Δ32* frequencies: 11.8% for those whose maternal grandmother was from Western Europe and 15.3% for those from Eastern Europe, and this difference was particularly evident in individuals whose grandmothers were born in one of the five countries, Poland, Russia, Hungary, Czechoslovakia or Austria (allele frequency of 0.195), a grouping which will be discussed below ($p < 0.01$), compared to western Europe.

Table 5.3 shows *CCR5-Δ32* allele frequencies from individuals tested here for each of the four grandparents, and this data are broken down into individual countries. Allele frequencies above an arbitrary level of 17% are indicated in bold and it can be seen that there is a clustering of these very high frequencies for individuals whose forebears came from Austria, Czechoslovakia, Russia and Hungary. (The data from Israel do not represent a single ancestral country of origin.)

In Table 5.4 the allele frequencies are broken down on the basis of maternal country of origin; these are compared with the published data on people resident in these countries. Again, figures over 17% are represented in bold. The data shown from the literature are almost exclusively on non-Jewish people. The exception is the data from Martinson *et al.* (1997) represented in the "unknown" row since the country of origin was not stated for this group of Ashkenazi Jews. These data highlight the likely difference between Jewish and non-Jewish populations in these countries although it is conceded that this is not a direct in-country sampling experiment, a study which we would not be in the position to carry out.

Table 5.2

***CCR5-Δ32* genotype and allele frequency compared with birthplace of maternal grandmother (MatGM) for Australian Ashkenazi Jews**

MatGM region of birth*	CCR5 wt/wt n	(%)	CCR5 Δ32/wt n	(%)	Δ32 / Δ32 n	(%)	total	allele freq.
Australia / NZ	56	73.7	20	26.3	0	--	76	0.132
South Africa	98	71.5	39	28.5	0	--	137	0.142
Western Europe	97	78.9	23	18.7	3	2.4	123	0.118 #
Eastern Europe	319	72.5	106	24.3	14	3.2	437	0.153
5 Countries**	129	65.5	59	29.9	9	4.6	197	0.195 #
other regions	23	67.6	9	26.5	2	5.9	34	0.191
total	591	73.2	197	24.4	19	2.4	807	0.146

***** Region groupings - Western Europe includes Italy, Switzerland, Austria, Germany, the U.K., Denmark, Finland, France and Holland.

Eastern Europe includes "USSR", Belarus, Latvia, Poland, Hungary, Czechoslovkia, Romania, Ukraine, "Russia" and other Eastern Europe designations.

****** The "5 Countries" subset consists of individuals whose grandparents were born in Poland, Russia, Hungary, Czechoslovkia or Austria and not elsewhere.

A Chi square test of the allele frequency between "Western Europe" and the "Just 5 Countries" subset gives a significant Chi square value of 6.589, $p < 0.01$. Calculation of the Odds Ratio gives a CI = 1.84 (range 1.17 - 2.89), $p = 0.008$.

Table 5.3

Country of reported grandparents' birth for Ashkenazi Jews and *CCR5-Δ32* allele frequency for individuals whose grandparent's birthplace is indicated

	Maternal grandmother		Maternal grandfather		Paternal grandmother		Paternal grandfather	
country	Δ32/total	%	Δ32/total	%	Δ32/total	%	Δ32/total	%
Australia	19/138	13.8%	16/116	13.8%	14/84	16.7%	21/118	**17.8%**
Austria	11/44	**25.0%**	3/22	13.6%	11/38	**28.9%**	10/52	**19.2%**
Czechoslovakia	10/50	**20.0%**	11/44	**25.0%**	7/58	12.1%	8/52	15.4%
England	8/100	8.0%	13/104	12.5%	11/82	13.4%	10/72	13.9%
Germany	6/70	8.6%	13/86	15.1%	7/82	8.5%	5/84	6.0%
Hungary	22/112	**19.6%**	23/130	**17.7%**	19/124	15.3%	20/138	14.5%
Israel	4/18	**22.2%**	5/30	16.7%	3/20	15.0%	3/24	12.5%
Lithuania	12/94	12.8%	14/94	14.9%	18/106	**17.0%**	11/92	12.0%
Poland	33/238	13.9%	31/242	12.8%	38/236	16.1%	36/220	16.4%
Romania	2/26	7.7%	2/30	6.7%	2/52	3.8%	3/42	7.1%
Russia	22/140	15.7%	27/158	**17.1%**	32/176	**18.2%**	30/158	**19.0%**
South Africa	39/278	14.0%	25/228	11.0%	25/210	11.9%	26/214	12.1%
Ukraine	5/56	4.5%	6/56	10.7%	7/50	14.0%	8/62	12.9%
USA	3/22	13.6%	3/20	15.0%	1/10	10.0%	0/8	0.0%
unknown	21/126	16.7%	25/152	16.4%	24/172	14.0%	26/166	15.7%

Frequencies of 17% or greater are shown in bold.

Note: Data here refers to the individuals themselves, not any actual ascertainment of alleles from a given grandparent. The number of individuals reporting that their grandparents are from each country listed is one-half the total number of alleles shown. See results and discussion for further explanation.

Table 5.4

Country of mother's birth and CCR5 frequency for Ashkenazi Jews compared with published allele frequencies for general (non-Jewish) populations

country	(n)	CCR5 alleles Δ32/total	Δ32 freq.	general population frequency % (reference)
Australia	213	51 / 426	12.0%	9.5 [f] ; 11.8 [b]
Austria	10	4 / 20	**20.0%**	8.9 [b]
Czechoslovakia	14	6 / 28	**21.4%**	10.2 [b]
England	38	9 / 76	11.8%	11.1 [a] ; 11.7 [b] ; 12.2 [n]
Fr./Neth./Ger.**	24	5 / 48	10.4%	9.8-12.9 [e] ; 8.9 to 10.8 [b] ; 11 [c]
Hungary	44	18 / 88	**20.5%**	8.6 [c] ; 11.1 [l] ; 11.1 [d]
Israel	33	10 / 66	15.2%	8.4 [p] ; 10.2 * [k]
Poland	31	13 / 62	**21.0%**	13.3 [b] ; 15.5 [d] ; 9 to 14 [q]
Russia	49	19 / 98	**19.4%**	13.9 [c]; 10.4 [h] ; 12.2 [i]; 16.6 [d] ; 13.6 [b]; 9.8 [n] ; 15.6 [r]
South Africa	223	57 / 446	12.8%	9.4 [m] ; 11.5 [# c]
Ukraine	27	6 / 54	11.1%	9% to 11% [r]
USA	16	4 / 32	12.5%	9.7 * [b] ; 11.6 [g]
unknown	42	16 / 84	19.0%	**20.9*** [a] ; **17.4*** [j] ; 13.8* [o]

Frequencies greater than 17% are shown in bold.
* Data is for an Ashkenazi population
#Ashkenazi in South Africa are mostly of Lithuanian origin, see grandparent data as few "parents" were from there.
**Combined data for France (Fr.), Holland (Neth.) and Germany (Ger.)
Note: The data shown here is for the individuals who were studied, and is not the actual allele assignment for parental origin of their CCR5 alleles.

References for Table 5.3
a. (Martinson *et al.* 1997), b. (Stephens *et al.* 1998), c. (Libert *et al.* 1998), d. (Magierowska *et al.* 1998), e. (Lucotte 1997), f. (Bennetts *et al.* 1997), g. (Zimmerman *et al.* 1997), h. (Yudin *et al.* 1998), i. (Voevodin *et al.* 1998), j. (Lucotte *et al.* 1999), k. (Kantor and Gershoni 1999), l. (Szalai *et al.* 1999), m. (Williamson *et al.* 2000), n. (Martinson *et al.* 2000), o. (Maayan *et al.* 2000), p. (Maayan *et al.* 2000), q. (Jagodzinski *et al.* 2000), r. (Limborska *et al.* 2002)

Not shown in the tables is a subset of individuals with all four grandparents born in a given country. The largest such group was those with all four grandparents from South Africa (n = 54) and the *CCR5-Δ32* frequency was 11.1%; for those with all grandparents from Russia (n = 32), the *CCR5-Δ32* allele frequency was 19.4%; from Poland (n = 26) the allele frequency was 19.2%; from Hungary (n = 24) the frequency was 20.8%, and from Lithuania (n = 17) it was 14.7%. Likewise, when the grandparents of these Ashkenazi Jews were all from either Austria, Hungary, Poland, Russia, or Czechoslovakia and not elsewhere (n = 233) the *CCR5-Δ32* allele frequency was 19.7%. This differs significantly from the allele frequency of 10.9% found among 146 non-Jewish teenagers with all four grandparents born in Australia ($\chi^2 = 10.4$, $p < 0.001$).

5.2.3 Homozygous *CCR5-Δ32* individuals

The microsatellite studies described below relied heavily on the use of *CCR5-Δ32* homozygotes. There were 24 individuals homozygous for the *CCR5-Δ32* allele identified in this study, with 11 being born in Australia, four born in South Africa, seven born in Eastern Europe and two born elsewhere. Three homozygotes were from the non-Jewish cohort (Australian-born) and 19 of 21 Jewish homozygotes were Ashkenazi Jews (with one each "unsure" and "mixed" religious background). Data from the birthplace of parents and grandparents shows that more that half of the 19 Ashkenazi *CCR5-Δ32* homozygotes had parents born in Eastern Europe, and fully 75% had their grandparent (matgm, matgf, patgm and patgf in turn) from Eastern Europe. Of the 807 Ashkenazi Jews studied here, there were 409 born in Australia of which 8 were *CCR5-Δ32* homozygotes (2%) and 101 were heterozygotes (24.7%), 186 were South African born

of which two were homozygotes (1.1%) and 48 were heterozygotes (25.8%) and of the 132 born in Eastern Europe, seven (5.3%) were homozygotes for C*CR5-Δ32* while 32 (24.2%) were heterozygotes. Two additional homozygotes were identified from the 80 Ashkenazi Jews born in regions other than Australia, South Africa or Eastern Europe. For individuals who were of Jewish background but were of "mixed" rather than of Ashkenazi status (n = 70), there was one homozygote for CCR5-Δ32 identified along with 12 heterozygotes (17.1%) while for individuals who were "unsure" of their religious subgroup, there also was one CCR5-Δ32 homozygote. Both of these homozygotes were of South African background. No CCR5-Δ32 homozygotes were detected in the 35 Sephardic Jews included in this study and only four (11.4%) were heterozygous for CCR5-Δ32. These CCR5-Δ32 homozygotes were used to look at allele frequencies for various microsatellite repeat polymorphisms in the 3p21 region and they were compared with individuals who were heterozygous for *CCR5-Δ32* or individuals who did not carry the deletion (i.e. "wild-type").

5.2.4 Identifying the *CCR5-Δ32* founder haplotype

As shown in Tables 5.5 to 5.11, genotyping was carried out for six microsatellites located at various distances from the CCR5 gene: D3S4579, D3S4580, afmb362wb9, chlc.gaat12d11, D3S3559, D3S663, D3S1578. The location of these microsatellites on chromosome 3p21 is shown in figures 5.1a and 5.1b. Examples of the gel traces used in the fragment length analysis for D3S3559, D3S663 and D3S4580 are shown in Figures 2.10a-c, respectively, and gel traces for the other markers have a similar nature. The examination of the data on the *CCR5-Δ32* homozygotes identified the following founder haplotype with the association (linkage disequilibrium) diminishing with the map

distance between the CCR5 gene and the microsatellite. On the telomeric side of *CCR5-Δ32* the founder haplotype is defined by the markers afmb362wb9 with the 214bp allele and D3S4579 with the 138bp allele. The alleles of markers defining the founder haplotype centromeric of CCR5 are the 143bp allele of D3S4580, the 172bp allele of gaat12d11, the 112bp allele of D3S663 and the 160bp allele of D3S1578. This founder haplotype is further confirmed by inspection of the frequencies of the relevant microsatellite alleles in both Jewish and non-Jewish *CCR5-Δ32* heterozygotes (Tables 5.5 - 5.8, 5.10, 5.11). In each case the microsatellite allele frequency is close to that which would be predicted from the allele frequencies found on the *CCR5-Δ32* bearing chromosome and the wild-type chromosome.

Table 5.5

Allele frequencies for microsatellite marker D3S4579 in CCR5 wild-type Ashkenazi Jews and Australians, *CCR5-Δ32* homozygotes, and in heterozygotes

allele size	Ashkenazi (wild-type) % (n=152)	Australian (wild-type) % (n=38)	homozygous Δ32/Δ32 % (n=26)	Ashkenazi heterozygote % (n=43)	Non-Jewish heterozygote % (n=45)
128bp	0.0 (0)	0.0 (0)	0.0 (0)	1.2 (1)	0.0 (0)
134bp	0.0 (0)	0.0 (0)	0.0 (0)	1.2 (1)	1.1 (1)
138bp	**6.6 (20)#**	**0.0 (0)**	**96.2 (50)#**	**47.7 (41)**	**52.2 (47)**
140bp	1.3 (4)	2.6 (2)	0.0 (0)	1.2 (1)	0.0 (0)
142bp	2.3 (7)	3.9 (3)	0.0 (0)	2.3 (2)	1.1 (1)
144bp	8.9 (27)	14.5 (11)	1.9 (1)	7.0 (6)	11.1 (10)
146bp	18.1 (55)	21.1 (16)	0.0 (0)	5.8 (5)	5.6 (5)
148bp	16.4 (50)	17.1 (13)	0.0 (0)	12.8 (11)	10.0 (9)
150bp	21.7 (66)	31.6 (24)	0.0 (0)	11.6 (10)	11.1 (10)
152bp	15.1 (46)	3.9 (3)	0.0 (0)	7.0 (6)	5.6 (5)
154bp	6.9 (21)	5.3 (4)	0.0 (0)	2.3 (2)	1.1 (1)
156bp	2.0 (6)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
160bp	0.7 (2)	0.0 (0)	1.9 (1)	0.0 (0)	0.0 (0)
162bp	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	1.1 (1)
total	304	76	52	86	90

note: Chi square testing of the 138bp allele freq. between CCR5-wildtype Ashkenazi Jews and the *CCR5-Δ32* homozygotes gives a $\chi^2 = 226$, $p < 0.001$

Table 5.6

Allele frequencies for microsatellite marker D3S4580 in CCR5 wild-type Ashkenazi Jews and Australians, *CCR5-Δ32* homozygotes, and in heterozygotes

allele size	Ashkenazi (wild-type) % (n=135)	Australian (wild-type) % (n=37)	homozygous Δ32/Δ32 % (n=26)	Ashkenazi heterozygote % (n=70)	Non-Jewish heterozygote % (n=45)
137bp	1.5 (4)	0.0 (0)	0.0 (0)	0.7 (1)	0.0 (0)
139bp	0.0 (0)	0.0 (0)	0.0 (0)	1.4 (2)	2.2 (2)
141bp	0.0 (0)	0.0 (0)	1.9 (1)	2.1 (3)	0.0 (0)
143bp	**1.1 (3)#**	**2.7 (2)**	**88.5 (46)#**	**52.9 (74)**	**51.1 (46)**
145bp	2.2 (6)	4.1 (3)	7.7 (4)	6.4 (9)	4.4 (4)
147bp	1.9 (5)	1.4 (1)	0.0 (0)	0.7 (1)	1.1 (1)
149bp	1.9 (5)	1.4 (1)	0.0 (0)	0.0 (0)	2.2 (2)
151bp	4.1 (11)	8.1 (6)	0.0 (0)	1.4 (2)	4.4 (4)
153bp	17.0 (46)	8.1 (6)	0.0 (0)	7.1 (10)	7.8 (7)
155bp	17.8 (48)	13.5 (10)	1.9 (1)	5.7 (8)	14.4 (13)
157bp	20.4 (55)	9.5 (7)	0.0 (0)	2.9 (4)	1.1 (1)
159bp	7.8 (21)	5.4 (4)	0.0 (0)	2.9 (4)	1.1 (1)
161bp	2.2 (6)	5.4 (4)	0.0 (0)	0.7 (1)	1.1 (1)
163bp	0.0 (0)	4.1 (3)	0.0 (0)	0.0 (0)	1.1 (1)
165bp	1.5 (4)	6.8 (5)	0.0 (0)	0.0 (0)	3.3 (3)
167bp	7.0 (19)	10.8 (8)	0.0 (0)	2.9 (4)	2.2 (2)
168bp	1.1 (3)	0.0 (0)	0.0 (0)	0.7 (1)	0.0 (0)
169bp	5.6 (15)	10.8 (8)	0.0 (0)	1.4 (2)	0.0 (0)
170bp	2.2 (6)	2.7 (2)	0.0 (0)	2.1 (3)	0.0 (0)
171bp	2.6 (7)	5.4 (4)	0.0 (0)	2.9 (4)	0.0 (0)
172+bp	2.2 (6)	0.0 (0)	0.0 (0)	5.0 (7)	1.1 (1)
total	270	74	52	140	90

note: Chi square testing of the 143bp allele freq. between CCR5-wildtype Ashkenazi Jews and the *CCR5-Δ32* homozygotes gives a $\chi^2 = 257$, $p < 0.001$

Table 5.7

Microsatellite marker "afmb362wb9" allele frequencies in Ashkenazi Jews typed for CCR5 and in non-Jewish heterozygotes

	214bp allele		216bp allele	
cohort (# individuals)	number	frequency	number	frequency
Ashkenazi (wt/wt) (n= 13)	15	58%	11	42%
Ashkenazi (wt/Δ32) (n=55)*	75	68%	34	31%
CCR5 Δ32/Δ32 (n = 21)	**35**	**83%**	7	17%
non-Jewish wt/Δ32 (n= 42)	66	79%	18	21%
Sephardic Jews (n = 6)	7	58%	5	42%
Australian wt/wt (n = 4)	4	50%	4	50%

* one allele of "212bp"

Table 5.8

Microsatellite marker "gaat12d11" allele frequencies in Ashkenazi Jews, Ashkenazi heterozygotes, *CCR5-Δ32* homozygotes and non-Jewish heterozygotes

	168bp allele		172bp allele	
cohort (# individuals)	number	frequency	number	frequency
Ashkenazi (wt/wt) (n= 36)*	19	26%	51	71%
Ashkenazi (wt/Δ32) (n= 59)	21	18%	97	82%
CCR5 Δ32/Δ32 (n = 25)	2	4%	**48**	**96%**
non-Jewish wt/Δ32 (n= 43)	11	13%	75	87%
Sephardic Jews (n = 6)	3	25%	9	75%
Australian wt/wt (n = 6)	4	33%	8	66%

* note: two alleles were typed as "166bp" and are not listed here

Table 5.9

Microsatellite D3S3559 frequency for CEPH (GDB) data and for the Ashkenazi Jewish homozygote *CCR5 Δ32/Δ32* cohort

	GDB allele data set for CEPH families	Ashkenazi Δ32/Δ32 data set	
allele size	% (n=27)	%	(n=25)
173 bp	0.02	0.02	(1)
175 bp	0.28	0.20	(10)
177 bp	0.19	0.14	(7)
179 bp	0.11	0.18	(9)
181 bp	0.09	0.18	(9)
182 bp	0.06	0.02	(1)
183 bp	0.04	0.02	(2)
186 bp	0.06	0.10	(5)
188 bp	0.09	0.08	(4)
190 bp	0.04	0.04	(2)
192 bp	0.02	n.d.	--
196 bp	0.02	n.d	--
total alleles	54		50

n.d. = not detected

Table 5.10

Allele frequencies and number of alleles for the D3S663 microsatellite marker in CCR5 wild-type Ashkenazi Jews and Australians, *CCR5-Δ32* homozygotes, and in heterozygotes

allele size	Ashkenazi (wild-type) % (n=106)	Australian (wild-type) % (n=22)	homozygous Δ32/Δ32 % (n=23)	Ashkenazi heterozygote % (n=39)	Non-Jewish heterozygote % (n=37)
92bp	2.8 (6)	4.5 (2)	0.0 (0)	0.0 (0)	5.4 (4)
94bp	9.9 (21)	0.0 (0)	2.2 (1)	7.7 (6)	4.1 (2)
98bp	0.9 (2)	2.3 (1)	0.0 (0)	0.0 (0)	0.0 (0)
108bp	2.4 (5)	2.3 (1)	0.0 (0)	1.3 (1)	0.0 (0)
110bp	0.5 (1)	0.0 (0)	0.0 (0)	1.3 (1)	0.0 (0)
112bp	**9.0 (19)#**	**4.5 (2)**	**32.6 (15)#**	**20.5 (16)**	**5.4 (4)**
114bp	9.4 (20)	6.8 (3)	6.5 (3)	10.3 (8)	13.5 (10)
116bp	13.2 (28)	15.9 (7)	2.2 (1)	11.5 (9)	14.9 (11)
118bp	17.9 (38)	18.2 (8)	21.7 (10)	23.1 (18)	16.2 (12)
120bp	17.0 (36)	13.6 (6)	4.3 (2)	9.0 (7)	14.9 (11)
122bp	11.3 (24)	15.9 (7)	8.7 (4)	9.0 (7)	10.8 (8)
124bp	3.8 (8)	9.1 (4)	4.3 (2)	2.6 (2)	9.5 (7)
126bp	0.9 (2)	4.5 (2)	15.2 (7)	3.8 (3)	4.1 (3)
128bp	0.9 (2)	0.0 (0)	2.2 (1)	0.0 (0)	0.0 (0)
130bp	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	1.4 (1)
132bp	0.0 (0)	2.3 (1)	0.0 (0)	0.0 (0)	0.0 (0)
134bp	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	1.4 (1)
total	212	44	46	78	74

note: Chi square testing of the 112bp allele freq. between CCR5-wildtype Ashkenazi Jews and the *CCR5-Δ32* homozygotes gives a $\chi^2 = 19.90$, $p < 0.001$

Table 5.11

Allele frequencies and number of alleles for the D3S1578 microsatellite marker in CCR5 wild-type Ashkenazi Jews and Australians, *CCR5-Δ32* homozygotes, and in heterozygotes

allele size	Ashkenazi (wild-type) % (n=88)	Australian (wild-type) % (n=21)	homozygous Δ32/Δ32 % (n=25)	Ashkenazi heterozygote % (n=28)	Non-Jewish heterozygote % (n=24)
124bp	0.0 (0)	0.0 (0)	2.0 (1)	1.8 (1)	0.0 (0)
130bp	0.6 (1)	0.0 (0)	0.0 (0)	1.8 (1)	0.0 (0)
136bp	0.6 (1)	0.0 (0)	0.0 (0)	0.0 (0)	2.1 (1)
138bp	10.8 (19)	2.4 (1)	14.0 (7)	10.7 (6)	0.0 (0)
140bp	6.8 (12)	2.4 (1)	12.0 (6)	5.4 (3)	8.3 (4)
142bp	7.4 (13)	4.8 (2)	2.0 (1)	8.9 (5)	14.6 (7)
144bp	5.1 (9)	7.1 (3)	4.0 (2)	7.1 (4)	8.3 (4)
146bp	3.4 (6)	9.5 (4)	4.0 (2)	1.8 (1)	6.3 (3)
148bp	15.9 (28)	14.3 (6)	2.0 (1)	14.3 (8)	10.4 (5)
150bp	13.6 (24)	19.0 (8)	12.0 (6)	10.7 (6)	10.4 (5)
152bp	11.9 (21)	9.5 (4)	8.0 (4)	16.1 (9)	12.5 (6)
154bp	8.5 (15)	9.5 (4)	8.0 (4)	3.6 (2)	14.6 (7)
156bp	7.4 (13)	7.1 (3)	2.0 (1)	0.0 (0)	2.1 (1)
158bp	1.1 (2)	7.1 (3)	0.0 (0)	1.8 (1)	2.1 (1)
160bp	**5.1 (9)#**	**0.0 (0)**	**28.0 (14)#**	**14.3 (8)**	**2.1 (1)**
162bp	1.7 (3)	2.4 (1)	2.0 (1)	0.0 (0)	6.3 (3)
164bp	0.0 (0)	2.4 (1)	0.0 (0)	1.8 (1)	0.0 (0)
166bp	0.0 (0)	2.4 (1)	0.0 (0)	0.0 (0)	0.0 (0)
total	176	42	50	56	48

note: Chi square testing of the 160bp allele freq. between CCR5-wildtype Ashkenazi Jews and the *CCR5-Δ32* homozygotes gives a $\chi^2 = 22.3$, $p < 0.001$

5.2.5 Estimating the age of the *CCR5-Δ32* mutation

If the recombination distance between CCR5 and the microsatellite marker is known, the number of generations between the time of origin of the *Δ32* mutation and the current time can be estimated. This calculation involves several assumptions including accurate microsatellite genotyping and minimal mutation in the associated microsatellite allele. Accurate knowledge of the recombination distance is both difficult to obtain and is critical for the calculation. In this regard, recent improved access to draft sequence data from the public domain of the human genome project (Lander *et al.* 2001) (http://www.ncbi.nlm.nih.gov/entrez/) provides a measurable background (McPherson *et al.* 2001) on which to position markers relative to CCR5, but it is not yet a complete resolution to the question of relative position of markers and genes in this region and the question of genetic versus physical distance still remains (DeWan *et al.* 2001; Yu *et al.* 2001). The map in Figure 5.1 shows the Build 28 and 31 NCBI map positions for a number of the markers in the 3p21 region.

The chemokine receptor genes in 3p21 form a cluster that is very likely to be a result of relatively recent gene duplication events, with CCR1, CCR2, CCR3 and CCR5 within 350kb of each other (Daugherty and Springer 1997; Samson *et al.* 1996c). CCR5 is located on a sequenced stretch of DNA (GenBank ID U95626) that contains 143,068bp and is represented on the finished contig map of the current public domain portion of the human genome project as NT_0022739.1 at between 49,325 kbp and 49,468 kbp. The location for CCR5 itself is between 45,591 kbp and 45,597 kbp on the Build 31 NCBI map, and on the UCSC Genome Browser (found at http://genome.ucsc.edu/) CCR5 is placed at 49.3Mbp in the December 2001 freeze (NCBI Build 28), and at 45.6 Mbp in the

November 2002 freeze (NCBI Build 31). The highly polymorphic, very tightly associated, markers D3S4579 and D3S4580 (IRI3.1 and IRI3.2) (Libert *et al.* 1998) are located within this sequence as well at 11 kbp upstream and 68 kbp downstream from CCR5 (figure 5.1b). The region telomeric of the chemokine receptor cluster is noteworthy for containing the common eliminated region 1 (C3CER1) of 3p21 found to be deleted in a large number of human tumours (Kiss *et al.* 2001; Kiss *et al.* 2002).

Legend for Figure 5.1

Fig 5.1 a & b - CCR5 and selected markers in Builds 28 & Build 31

Physical map distances were taken from the NCBI Build 28 (UCSC Dec. 2001 freeze) and NCBI Build 31 (UCSC November 2002 freeze) releases of public domain data for the human genome project. Locations are given in “mega-basepair” (Mbp) and “kilo-basepair” (kbp).

Figure 5.1a

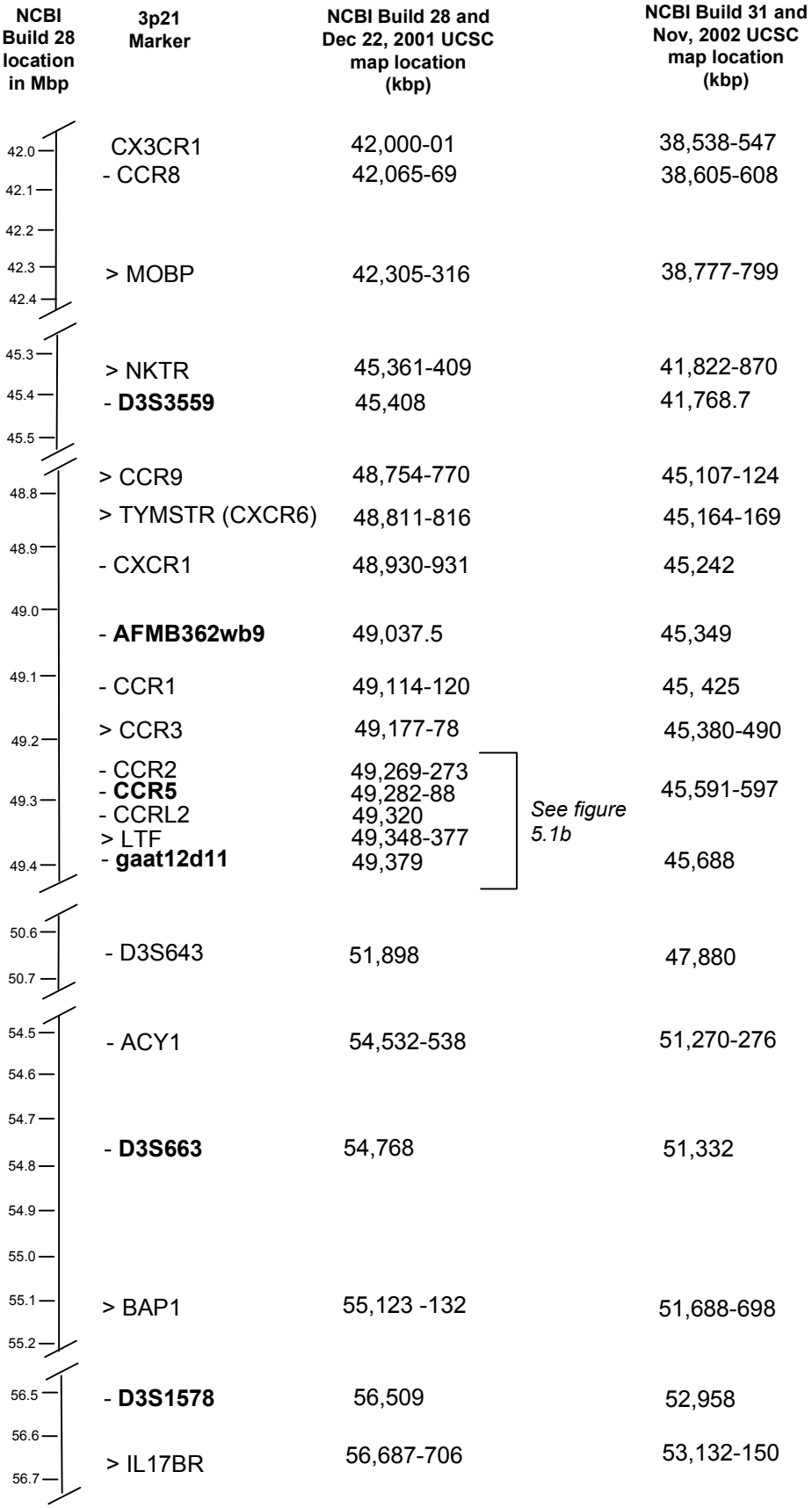
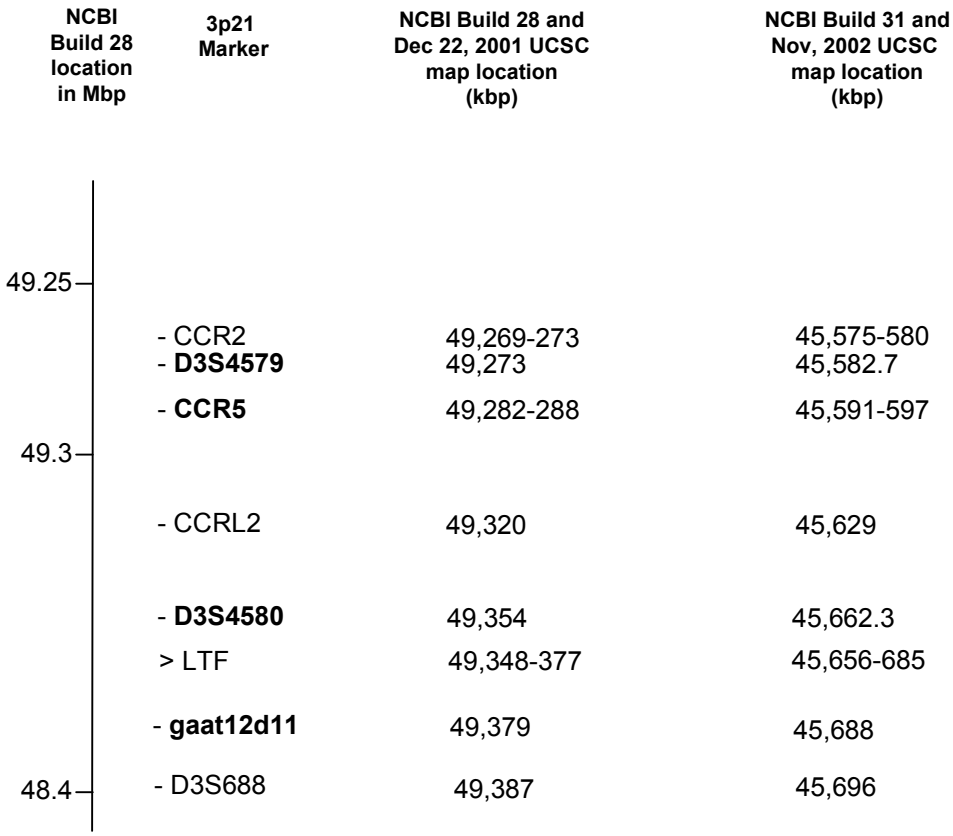

NCBI Build 28 location in Mbp
3p21 Marker
NCBI Build 28 and Dec 22, 2001 UCSC map location (kbp)
NCBI Build 31 and Nov, 2002 UCSC map location (kbp)
42.0
42.1
42.2
42.3
42.4
CX3CR1
42,000-01
38,538-547
- CCR8
42,065-69
38,605-608
> MOBP
42,305-316
38,777-799
45.3
45.4
45.5
> NKTR
45,361-409
41,822-870
- D3S3559
45,408
41,768.7
48.8
48.9
49.0
49.1
49.2
49.3
49.4
> CCR9
48,754-770
45,107-124
> TYMSTR (CXCR6)
48,811-816
45,164-169
- CXCR1
48,930-931
45,242
- AFMB362wb9
49,037.5
45,349
- CCR1
49,114-120
45, 425
> CCR3
49,177-78
45,380-490
- CCR2
49,269-273
- CCR5
49,282-88
- CCRL2
49,320
> LTF
49,348-377
- gaat12d11
49,379
45,591-597
45,688
See figure 5.1b
50.6
50.7
- D3S643
51,898
47,880
54.5
54.6
54.7
54.8
54.9
55.0
55.1
55.2
- ACY1
54,532-538
51,270-276
- D3S663
54,768
51,332
> BAP1
55,123 -132
51,688-698
56.5
56.6
56.7
- D3S1578
56,509
52,958
> IL17BR
56,687-706
53,132-150

Figure 5.1b


NCBI Build 28 location in Mbp
3p21 Marker
NCBI Build 28 and Dec 22, 2001 UCSC map location (kbp)
NCBI Build 31 and Nov, 2002 UCSC map location (kbp)
49.25
- CCR2
- D3S4579
- CCR5
49.3
- CCRL2
- D3S4580
> LTF
- gaat12d11
48.4
- D3S688
49,269-273
49,273
49,282-288
49,320
49,354
49,348-377
49,379
49,387
45,575-580
45,582.7
45,591-597
45,629
45,662.3
45,656-685
45,688
45,696

Estimates for the number of human generations since the emergence of the *CCR5-Δ32* mutation are shown in Table 5.12. These estimates are based on the observed level of association between *CCR5-Δ32* and the associated microsatellite allele in 26 *CCR5-Δ32*homozygotes and on the estimated recombination distance between the two loci.

With the exception of the calculation based on the marker D3S4579, there is reasonable concordance between the number of estimated generations for each of the microsatellite markers in a range of 38 to 53 generations. Estimates based on markers that are very close to *CCR5*, and hence show a very high level of association, are particularly prone to error due to mistypings of or mutation in the microsatellite allele. The data on D3S4579 exemplifies this as two alleles other than 138bp were found; had there been just one, the number of generations becomes 64 rather than 129 for the estimated genetic distance shown. Each of the four markers studied by Stephens et al (afmb362wb9 and gaat12d11) and Libert et al (D3S4579 andD3S4580) are within less than 0.5cM of the *CCR5* gene. The inclusion in the current study of two markers at a significantly greater distance, D3S663 (approximately 2.5cM) and D3S1578 (approximately 3cM) provides an important advance over the previously published reports. Moreover, the close correlation between the estimation of number of generations between these two markers adds significantly to confidence that these estimates, although far from precise, are sufficiently reasonable for identification of a limited range of generations. The concordance of the estimates between the better selected markers and three of the four closely located markers adds to this conclusion.

In addition, the estimated genetic distances from *CCR5-Δ32* assumed by Stephens *et al.* (1998) and Libert *et al.* (1998) in their calculations were based on significantly less knowledge of the region. The physical distances used here are calculated from the current release of the human genome data (NCBI “Build 31" and the UCSC freeze dated November 2002, which use identical basepair measures), and the calculated genetic distance range is estimated by taking the values of 0.6 and 0.8 cM/Mbp to give a likely range for the genetic distance. An estimated genetic distance is shown alongside the physical distance, and for some of the markers this value is outside of the calculated range. Interestingly, the release of “Build 29" has the location of marker D3S1578 closer to CCR5 than that of D3S663. They are both located within the growing contig across the region, nt_005986.11, which has close to 2.9 Mbp of “working draft” sequence in place. This is just the result of a “glitch” in the sequence assembly procedure where the reduced constraints used in Build 29 have placed the relative GenBank sequences in this way and further releases of the public domain human genome sequence data (as can be seen in Build 31) will return these markers to their previous relative positions. With the Build 29 map locations the 28% frequency for D3S1578 followed by the 32% frequency for D3S663 seen in the homozygotes makes less sense.

The conversion from the number of generations to the number of years depends upon assumptions of the likely average number of years per generation. Table 5.13 examines this issue for the data obtained on the two microsatellite markers D3S663 and D3S1578 with varying assumed genetic distances in the range of 2cM to 4cM a range chosen given the best estimate for these markers at 2.5cM and 3cM respectively. As the map distance increases, the number of estimated generations responsible for the observed data

decreases. This can then be converted into years under three assumptions of average generation at 20 years, 25 years and 30 years. Of these, Tremblay and Vezina (2000) proposed that the most accurate estimate is 30 years. A best estimate for the D3S663 to *CCR5* haplotype at 2.5cM and 44 generations would be 1200 years and for D3S1578 at 3cM and 42 generations a very similar figure would be obtained, each calculated on the assumption of a 30 year generation interval. It would appear unlikely that the range is outside 600 to 1800 years ago. While it is conceded that these figures are contestable, they provide the best available estimates from which hypotheses can be generated.

Table 5.12

Number of generations required to have various microsatellite alleles on the suggested ancestral haplotype reach the observed values

Microsatellite marker allele	(bp)	allele frequency in homozygotes for Δ32/Δ32 %	Estimated physical and genetic distance from CCR5-Δ32 kbp / cM (est.)		Number of generations
afmb362wb9	(214bp)	83	- 248	0.35	53
D3s4579	(138bp)	96.2 *	- 12	0.03	129*
D3S4580	(143bp)	98.1 **	+ 69	0.05	38
gaat12d11	(172bp)	96	+ 91	0.1	41
D3S663	(112bp)	32.6	+ 5,835	2.5	44
D3S1578	(160bp)	28	+ 7,635	3	42

* two alleles were other than the 138bp D3S4579 allele; had there been just one, the number of generations (n) becomes 64 rather than 129 for the estimated genetic distance shown.
** five of the alleles were one repeat removed from the 143bp allele and these can be assumed to be replicative slippage events rather than crossing over events
# calculated by (n) = log (observed allele freq.) / log ((100 - cM) / 100)

Table 5.13

Possible ages of a CCR5Δ32-D3S663-D3S1578 haplotype

distance (cM)	2 cM	2.5 cM	3cM	4cM
generations	50	44	40	30
20 years	1000 years	880 years	800 years	600 years
25 years	1250 years	1100 years	1000 years	750 years
30 years	1500 years	1320 years	1200 years	900 years

Note: calculated from the defined genetic distance and the residual proportion of haplotypes thus defined as not having crossed-over for each generation

5.2.6 Could the CCR5-Δ32 founder have been Jewish?

Having established above that the *CCR5-Δ32* mutation occurred on the haplotype afmb362wb9 (214bp) - DS34579 (138bp) - D3S4580 (143bp) - gaat12d11 (172bp) - D3S663 (112bp) - D3S1578 (160bp), I was then able to examine the frequency of this haplotype amongst individuals who were wild-type homozygotes both from Ashkenazi and non-Ashkenazi Australian populations. As seen in Tables 5.5, 5.6, 5.10 and 5.11, the following allele frequencies are evident where sufficient data is available:

Allele	Ashkenazi wild-type	non-Jewish wild-type
D3S4579 - 138bp	6.6%	0%
D3S4580 - 143bp	1.1%	2.7%
D3S663 - 112bp	9%	4.5%
D3S1578 - 160bp	5.1%	0%

While individual markers are present it is clear that the founder haplotype is not found in the non-Jewish population except in individuals carrying the *CCR5-Δ32* mutation but is evident in homozygous wild-type members of the ancestors of the modern Ashkenazi population. These findings provide suggestive evidence for the existence of the founder haplotype amongst the ancestors of modern Ashkenazi Jews prior to the development of the *CCR5-Δ32* mutation.

5.3 Discussion

The data presented here confirm the likely absence of the *CCR5-Δ32* mutation in people of Asian and Pacific Island ethnicity, a finding consistent with the emergence of this mutation after the time of separation of the races, estimated to be approximately 250,000 years ago (Martinson *et al.* 2000). The results also support a clear difference in the frequency of the *CCR5-Δ32* allele between Sephardic and Ashkenazi Jews, most likely, perhaps not exclusively, due to the emergence of *CCR5-Δ32* after the Great Diaspora in AD70. Lucotte et al (1999) reported similar data to those presented here with an allele frequency of 17.4% among Ashkenazi compared to 6.4% for Sephardic Jews. This difference was supported by a recent report from Maayan et al (2000) who extended the study to genotyping for the much older CCR2-64I mutation, an allele with a similar effect to *CCR5-Δ32* on the progression of HIV disease to AIDS and death as shown in an international meta analysis to which our laboratory contributed (Ioannidis *et al.* 2001). The effect of the CCR2-64I mutation is either direct or secondary to linkage disequilibrium with a polymorphism in the CCR5 promoter region (see Stewart 1998). The data for CCR2-64I, a polymorphism found in all racial groups, showed no difference between these two Jewish populations (Maayan et al, 2000). This result further supports the concept that natural selection associated with *CCR5-Δ32* has been unrelated to HIV. Although it should be noted that CCR2-64I does not have an effect on susceptibility to HIV, (unlike *CCR5-Δ32* homozygosity), it has been calculated that HIV infection alone will result in a significant increase in the frequency of CCR2-64I in Africa over the next 100 years (Schliekelman *et al.* 2001).

Table 5.14
Elevated CCR5-Δ32 frequency given in various published cohorts

Number studied (reference)	**Frequency of CCR5-Δ32 / Location**
n = 43 (Martinson *et al.* 1997)	**Allele freq. = 0.209** - Ashkenazi Jews
n = 14 (Yudin *et al.* 1998)	**Allele freq. = 0.179** - Mordvinians
n = 256 (Lucotte *et al.* 1999)	**Allele freq. = 0.174** - Ashkenazi Jews
n = 33 (Magierowska *et al.* 1998)	**Allele freq. = 0.166** - Russian
n = 86 (Libert *et al.* 1998)	**Allele freq. = 0.163** - Mordvinians (above)
n = 98 (Libert *et al.* 1998)	**Allele freq. = 0.158** - Finland
n = 58 (Magierowska *et al.* 1998)	**Allele freq. = 0.155** - Poland
n = 102 (Martinson *et al.* 1997)	**Allele freq. = 0.147** - Iceland
n = 94 (Magierowska *et al.* 1998)	**Allele freq. = 0.143** - Sweden
n = 152 (Libert *et al.* 1998)	**Allele freq. = 0.142** - Sweden (see above)
n = 139 (Libert *et al.* 1998)	**Allele freq. = 0.139** - Russian (see above)
n = 83 (Magierowska *et al.* 1998)	**Allele freq. = 0.138** - Norway (see above)
n = 131 (Stephens *et al.* 1998)	**Allele freq. = 0.137** - Sweden (see above)
n = 50 (Stephens *et al.* 1998)	**Allele freq. = 0.136** - Russian (see above)
n = 107 (Lucotte and Mercier 1998)	**Allele freq. = 0.135** - France (Brest)
n =158; 30; 30 (Stephens *et al.* 1998)	**Allele freq. = 0.133** - Estonian, Polish, Slovakian
n = 90 (Martinson *et al.* 1997)	**Allele freq. = 0.13** - Russian (see above)
n = 294 (Lucotte and Mercier 1998)	**Allele freq. = 0.129** - France (Paris)
n = 194 (Pastinen *et al.* 1998)	**Allele freq. = 0.129** - Finland (see above)
n = 61 (Yudin *et al.* 1998)	**Allele freq. = 0.123** - West Siberian (Urgic)
n = 176 (Voevodin *et al.* 1998)	**Allele freq. = 0.122** - Russian (see above)

The date presented her also provide the best available evidence that the population with the highest frequency of *CCR5-Δ32* is the Eastern European Ashkenazim. The initial report of the global distribution of *CCR5-Δ32* showed the highest allele frequency of all populations in a small cohort of 43 Ashkenazi Jews with an allele frequency of 21% (Martinson *et al.* 1997). I have been able to confirm that this cohort was of Eastern European origin (Martinson, J, personal communication). While Martinson's study covered a very wide range of European and non-European populations, it is conceded that the group of Ashkenazi's studied is small, unlike the numbers contained in this thesis. In a larger study of Jewish subjects, Lucotte *et al.* (1999) provided important support to the concept of a high frequency of *CCR5-Δ32* amongst Ashkenazi Jews at 17.4%. When the country of origin for their Ashkenazi samples was considered the frequencies for Poland (n = 127) and USSR (n = 30) were 21% and 25% respectively (Lucotte *et al.* 1999). Our study was commenced prior to the Lucotte publication and to that of Stephens *et al.* (1998), which had an allele frequency of just under 10% for a large group of American Ashkenazi Jews (see below).

In Table 5.14, a comparison of each of the published data in relevant European populations is shown. The highest frequencies, second only to those reported in Ashkenazi Jews by others and in this thesis, are seen in some, but not all, studies of people resident in Russia. The highest is seen amongst Mordvinians, Finnic people inhabiting the Volga and Ural areas, although the number of people studied has been small involving only 14 people in the report by Yudin *et al.* (1998) showing an allele frequency of 17.9% and 86 in the report by Libert *et al.* (1998) with an allele frequency of 16.3% (see Table 5.14). After this, the highest allele frequencies are seen in countries of

Scandinavia. Taken overall, the only studies of interpretable size which have shown an allele frequency of greater than 17% have been that of Lucotte (n=256), involving Ashkenazi Jews, and in this thesis, where it is shown that Australian Ashkenazi Jews whose grandparents originated from the five countries with common borders, Poland, Russia, Hungary, Czechoslovakia and Austria, have an allele frequency of 19.7% (n=233).

While a gradient for the allele frequency of *CCR5-Δ32* has been shown amongst many countries from north to south in Europe and from east to west (Martinson *et al.* 1997), the current data also support a gradient from east to west amongst Ashkenazi Jews. As shown in Table 5.4, relatively low allele frequencies were seen for Australian Ashkenazi Jews whose mothers were born in England (11.8%, n=38), and France, Holland or Germany (10.4%, n=24). These figures, although involving relatively small numbers, contrast with the figures on young Ashkenazi Jews whose forebears came from Austria, Czechoslovakia, Hungary, Poland or Russia as referred to above. The only equivalent data are those reported by Lucotte *et al.* (1999) who showed that an allele frequency of 14.3% among 56 Ashkenazi Jews from Germany compared to the higher frequencies for those from Poland and Russia as referred to above.

In this context, it is difficult to interpret the data on Jewish people who have already migrated from their European country of origin. Most important in this regard are the 503 Ashkenazi Jews studied by Stephens *et al.* (1998) in the USA with a *CCR5-Δ32* allele frequency of less than 10%. The authors, however, did not provide any explicit information on their countries of origin. This problem is similar for studies of Jewish

people from South Africa for whom we found relatively low allele frequencies, in the range of 11-14%, among young Australian Ashkenazi Jews analysed on the basis of each grandparent born in South Africa, however most of these individuals would have an origin from Lithuania. At this point it is important to acknowledge that the study of country of origin of Jewish people is not an exact science, particularly if one considers the movements or migrations of such people over several centuries. Similarly, the grouping of countries into eastern and western Europe is not an exact task. Within these constraints, however, there does appear to be a clear message emerging both from the data presented here and from the world literature.

The argument in favour of the *CCR5-Δ32* founder being a member of a population which is ancestral to modern day Ashkenazim is given further strong support by the data presented in this thesis on the haplotype upon which the *Δ32* mutation first occurred. With the six microsatellite markers studied this haplotype can be defined: afmb362wb9 (214bp) - D3S4579 (138bp) - *CCR5-Δ32* - D3S4580 (143bp) - gaat12d11 (172bp) - D3S663 (112bp) - D3S1578 (160bp), which spans several Mbp. This haplotype is not found among non-Jewish Australians (except those who carry the *CCR5-Δ32* allele) as shown by the absence of the 138bp allele of the very closely related D3S4579 microsatellite in the 38 CCR5 wild-type homozygotes genotyped (76 chromosomes), compared to 6.6% found amongst 152 Ashkenazi wild-type homozygotes (chi square = 5.3, $p < 0.02$). This is further supported by the absence of the 160bp allele of the most distant microsatellite D3S1578 (Table 5.11) compared to a frequency of 5.1% amongst Ashkenazi wild-type homozygotes, although the difference falls short of significance (chi square = 2.5, $p < 0.10$). While it is acknowledged that these results will be influenced by

the peoples with which these two populations have intermarried over the last 1,000 years or more, it is difficult to find an alternate explanation for the founder ancestral population after the significant detection of the founder haplotype in the homozygous wild-type current Ashkenazi population.

With respect to the likely age of the *CCR5-Δ32* mutation, the estimates provided in this thesis are the best currently available. This results from the study of six microsatellites and from the better information of map distances gained since the previous publications. The suggested date for the origin of the *CCR5-Δ32* allele given by Stephens and co-authors is around 1300AD (Stephens *et al.* 1998), with a range from ~1725AD back to around the year 125AD. This result is based on the maintenance of a haplotype formed by CCR5 and the markers gaat12D11 and AFMB362wb9 over a distance then thought to be 0.93cM (albeit in a different physical order to that now understood for the human genome) and suggests a single mutation gave rise to this allele. The suggested date of around 700 years ago is of interest as that would have been around the time when the Black Plague occurred, but as the authors note a number of pathogens (including smallpox) may also have the potential to have selected for this allele (Stephens *et al.* 1998). The origin of *CCR5-Δ32* as calculated by Libert *et al.* (1998) is based on the strong linkage disequilibrium shown between CCR5 and alleles of microsatellite markers D3S4579 and D3S4580 (located 11kb upstream and ~ 68kb downstream of *CCR5*, respectively). Their calculations on the possible date of this mutation suggest a time earlier than Stephens et al. have suggested, with data from D3S4579 suggestive of some 1400 years ago and D3S4580 data giving a range from 2000-2200 years ago (Libert *et al.* 1998). The allele frequency for the markers D3S4579 and D3S4580 in the *CCR5-Δ3*2

homozygotes in the current study (Tables 5.5 and 5.6) give very similar frequencies to those reported by Libert and so the calculation of the number of generations since this mutation occurred, based on the assumptions made in their study or now, would provide a similar answer. However, better knowledge of the relationship between physical and genetic distance since that time results in a re-calculation for D3S4580 of 38 generations (or, at 30 years per generation, 1140 years). These markers are the closest to *CCR5* studied and, as mentioned above, the calculation of the number of generations is exposed to error to a greater extent than for more distant markers showing linkage disequilibrium such as D3S663 and D3S1578.

The physical map distance of the genome in the region between *CCR5-Δ32* and D3S663 is about 5 Mbp for the NCBI MapViewer (Build 26) while the UCSC Genome Browser distances are 5.4 Mbp for the April 1, 2001 freeze and 6.2 Mbp for the August 6, 2001 freeze (see Figure 1). For the physical distance from *CCR5* to D3S1578, the NCBI site indicates 6.7 Mbp; the UCSC April 1 data gives 7.1 Mbp and the August 6 data gives 8.1 Mbp. The sex-average recombination rate to physical distance ratio is close to 1cM/Mbp around D3S1578 (based on the Marshfield web site data), while a ratio of ~0.6 cM/Mbp or even less may apply for the D3S663 to *CCR5* region. For the distance between *CCR5* and D3S663, an estimate of about 3 cM would is reasonable. Indeed, the Stanford GB4-RH map interval containing CCR5 and extending to D3S1578 is calculated as 3 cM across the interval (www.ncbi.nlm.nih.gov/genemap/). Data from the Rockefeller web site (DeWan *et al.* 2001) would suggest a value of 0.6 to 0.8 cM/Mbp across this region which gives a similar calculated genetic distance of 3 to 4 cM.

Taken overall, the current data provide the best estimate to date of the age of the *CCR5-Δ32* mutation with a likely mean of around 45-50 generations. The time in years since the mutation occurred also depends on the number of years that are assumed to make up a typical human generation, recently proposed to be around 30 years (Tremblay and Vezina 2000). Based on 30% of haplotypes maintaining an ancestral state at a marker 5 cM away, the mutation could in fact be fairly recent. At a genetic distance of 5 cM, after a generation it would be expected that 95% of chromosomes would retain this region in the ancestral state and only at 23 to 24 generations does the decay in the ancestral haplotype at 5 cM drop to the 30% level. Should the genetic distance be slightly less, the number of generations would increase (to 30 generations for a marker at 4 cM that maintains 30% of chromosomes in the ancestral form). Likewise, if the genetic distance is slightly greater, the number of generations decreases. For a marker with 30% of chromosomes yet to cross over within the haplotype at a distance of 6 cM, 20 generations would be expected to have passed. As is shown in Tables 5.12 and 5.13, with D3S663 at a physical distance of over 5 Mbp from *CCR5-Δ32*, then, based on a 4 cM genetic distance and a 25 year generation interval, this mutation would be at least 750 years old. For a 3 cM genetic distance, which is a reasonable estimate from the Marshfield web site data, it requires 40 generations (or 1200 years at 30 year generations) for the decay of the haplotype to the 30% seen in the *CCR5-Δ32* homozygotes shown here. Overall, it is likely that the *CCR5-Δ32* mutation occured between 45 and 50 generations ago, representing a period of 1000 to 1500 years ago with a mean of 1250 years ago.

5.3.1 The Khazar Empire and the role of the Vikings

Other information about population demographics also needs to be considered. In this case, the historic movement of the Ashkenazi Jewish people was away from the middle East during the Diaspora around 2,000 years ago, and then, under 1,000 years ago, a distinct movement from the western to the eastern parts of Europe occurred (Shamir and Shavit 1986). The *CCR5-Δ32* allele frequency in the Eastern European Ashkenazi populations relative to published frequencies is shown in Table 5.4, and in general there is a somewhat higher allele frequency for the Ashkenazi data than for the published (non-Jewish) reference data (Table 5.14). This conflicts with the assertion that was made in both recent reports suggesting admixture into the Ashkenazi population (Kantor and Gershoni 1999; Maayan *et al.* 2000). In the report by Maayan and colleagues, their focus on selection is driven by the implausible rate of admixture that would have been required for the Ashkenazi population to have acquired this allele from the general population (Maayan *et al.* 2000).

Lucotte has suggested that a possible means for the dispersion of *CCR5-Δ32* lies with the movements of the Vikings in the eighth to the tenth centuries (Lucotte 2001). While this does not preclude an origin of *CCR5-Δ32* in an Ashkenazi Jew, movement of the allele through the regions of northern Europe where it currently shows the greatest allele frequencies does seem to fit with this model. The calculation of the age of this allele depends on the genetic distance across the several megabase physical distance; 40 to 50 generations or about 1000 to 1250 years is a reasonable estimate based on the available data. It is difficult, however, to reconcile the movement of the Ashkenazi into Eastern Europe with a possible Viking contribution to the movement of *CCR5-Δ32* (Lucotte

2001). This raises the intriguing question of what populations and events may link the two. The long-forgotten Kingdom of Khazaria both traded with Swedish Vikings (Poser 1994; Poser 1995) and later had extensive admixture with Ashkenazi Jews. There is the possibility the mutation was in a Viking founder and spread into the Khazars and later into the Ashkenazi who migrated into the region of this kingdom, but this would infer reduced selection in some of the nordic regions given the current lower frequencies compared to modern Ashkenazi Jews. In the initial global survey, the population group with the next highest *CCR5-Δ32* frequency to that reported for the 43 Ashkenazi Jews was a sample from the Icelandic population. Iceland is an isolated population of Viking origin established around 1130 A.D. and if selection (by smallpox) had been weak there, then the elevated allele frequency could be seen to support a Viking origin. In fact, smallpox did reach Iceland and 21 documented severe epidemics between 1240 and 1839 A.D. have led to the suggestion that changes in the ABO blood group frequencies, which could confound the question of Icelandic origins, may have been a result of selection by smallpox (Adalsteinsson 1985). With smallpox well established in Iceland and the other nordic countries (Skold 1996), the current allele frequencies could well be a result of selection rather than an elevation due to a founder effect. Movement of the allele along Viking trade routes after admixture is thus more likely than an origin within the Vikings as allele frequencies, which vary somewhat between nordic countries, are not as high as could be expected.

The Khazar Kingdom was established in by the middle of the 6th century A.D. and expanded and flourished through the end of the first millennium before disappearing with little recorded history by the 14th century when the Hungarian Jews, partly of Khazar

origin, resettled in Poland and Austria (Brook 1999). At its height, after a century of war with the Arabs, the kingdom occupied a sizeable part of southeastern Europe and provided a barrier to eastern expansion by Christian countries to the west and to western expansion from Arab nations to the south. The conversion of the Khazars to Judaism in the 9th century, towards the end of the movements of the Vikings, created the basis for admixture with the Ashkenazi who arrived in this region after the period of the Viking migrations. Maayan and co-workers cite the degree, estimated at 20-fold, for which the Ashkenazi obtaining *CCR5-Δ32* by admixture from non-Jewish Caucasians in a given region would have to be greater than that demonstrated for other genes (Maayan *et al.* 2000). Undoubtedly the Ashkenazi and the Khazars, also practising Jews, would have reduced cultural barriers to intermarriage and so admixture of a greater extent would be possible in this setting. While studies of Y-chromosome markers in Jews have shown a trend towards a common Middle East origin (Hammer *et al.* 2000; Nebel *et al.* 2001; Nebel *et al.* 2000), a recent mitochondrial DNA and Y-chromosome study shows significant local female genetic input in the establishment of new Jewish communities (Thomas *et al.* 2002).

If the mutation had occurred in the Jewish people prior to exodus from Israel, Sephardic and other Jews would have a greater frequency than is shown here or has been shown by others. Had the mutation occurred in the Ashkenazi prior to the movement into the eastern parts of Europe, it would be expected that the allele frequency for *CCR5-Δ32* would be greater in the western European Ashkenazi than in those from eastern Europe. This is excluded by the data presented in this chapter which indicate the reverse. As discussed above, movement of the allele along Viking trade routes and subsequent

selection after admixture is more likely than an origin within the Vikings as allele frequencies, which vary somewhat between nordic countries, are not as high in some of these countries as could be otherwise expected. The frequency of *CCR5-Δ32* is seen here to be higher in the Ashkenazi Jews of eastern Europe than reported for the general population in several countries (Table 5.4) including Hungary, Poland and Russia, and also somewhat higher than the 14% to 16% for Finland, Iceland and Sweden (Lucotte 2001).

Considering the relatively high general Caucasian allele frequency for *CCR5-Δ32* of about 10% and the likely origin of this mutation in the northeast of Europe some hundreds years ago, a strong argument has been made for increased survival for heterozygotes. Understanding the timing and population in which the mutation giving rise to the *CCR5-Δ32* allele occurred is important in understanding the nature of the factors which have provided selective forces. Slatkin and Bertorelle have reviewed the question of selection on *CCR5* and confirm that it has not been neutral (Slatkin 2001; Slatkin and Bertorelle 2001) while Stephens *et al*. concluded that the current prevalence of *CCR5-Δ32* would have taken over 200,000 years if by genetic drift. The Black Plague was suggested by Stephens *et al.*(1998) to have a possible role on the basis of the timing of this major epidemic with their best estimate of the date of the *CCR5-Δ32* mutation being 700 years ago. This hypothesis was published in *The American Journal of Human Genetics* and received wide attention in the popular media, however it seems somewhat implausible. By their data, the *CCR5-Δ32* allele would not have been sufficiently frequent at the time for a sufficient increase to occur even if all heterozygotes were spared. In addition, others have raised the question of "The Black Death" being a viral

(hemorrhagic fever) epidemic, and not entirely Bubonic Plague (Scott and Duncan 2001). As mentioned, the Black Death happened in a specific narrow window of time and the *CCR5-Δ32* mutation not only would have to occur well before the Black Death, there would have to have been significant survival advantage for heterozygotes as well as homozygotes for an effect to be seen in the allele frequency today (Schliekelman *et al.* 2001). The role of smallpox in selection of *CCR5-Δ32* is more persuasive as there is a definite role for CCR5 in poxvirus infections (Carfi *et al.* 1999; Lalani *et al.* 1999) and smallpox was a constant feature of European life over many hundreds of years (Scott and Duncan 2001).

The alleles associated with the *CCR5-Δ32* allele at markers D3S663 and D3S1578 on about 30% of chromosomes containing the *CCR5-Δ32* allele indicate an ancestral haplotype of several Mbp which contains over 200 genes and which remains intact in these individuals. Any of these genes, a selection of which are shown in Table 6.1, could hypothetically have contributed to natural selection of the haplotype.

In summary, the evidence here that suggests a possible Jewish origin of *CCR5-Δ32* includes both an increased allele frequency of Australian Ashkenazi Jews who are from Eastern Europe and a pattern of alleles from microsatellite markers which define an ancestral haplotype. The age of this allele is such that admixture of the Vikings and Ashkenazi independently with the Khazars is a reasonable hypothesis that can explain the spread of this allele in Caucasians. It seems that, while over-represented in the Ashkenazi, *CCR5-Δ32* may be a genetic artifact of the forgotten Kingdom of the Khazars, a people who have been referred to as “The Thirteenth Tribe” of Israel (Koestler 1976)

but whose trade relations and population interactions (Brook 1999) provide a basis for understanding the spread of this allele. The timing of these proposed admixtures is also certainly plausible with respect to the spread of smallpox across all of Europe, with the strong selection for *CCR5-Δ32* that smallpox may have provided (Carfi *et al.* 1999). The possibility that *CCR5-Δ32* itself may not confer the natural advantage cannot be excluded given the sizable number of genes contained within the ancestral haplotype defined in this chapter. Each aspect of this allow the development of testable hypotheses aimed at providing an answer to this remarkable experiment of nature and of human history.

Chapter Six - Discussion

6.1 The study of immunogenetics of human disease

The pursuit of gene discovery in diseases of complex genetics is not an esoteric one. It has at its heart the goal of furthering understanding of the pathogenesis of the disease under study and through this providing critical information the eventual development of preventive and therapeutic strategies. The finding of clear and reproducible associations between alleles of a gene of importance to normal function of the immune system provides a basis for hypothesis generation and further experimentation in humans or on relevant animal models. It is as valid an approach as any other towards unravelling pathogenesis. Since most, if not all of the diseases of complex genetics are likely to involve one or more environmental factors, the study of the immunogenetics of the disease must also take into account parallel knowledge of such factors and, most importantly, the likely mechanisms of interaction between the two. My comments here are confined to immunogenetics because of my prime interest and training in immunology. It is acknowledged, however, that genes involving the function and repair of the target organ in these diseases may be equally important and worthy of study. An example would be the study of genetic influence on oligodendrocyte function in the CNS of relevance to the pathogenesis and clinical expression of multiple sclerosis.

The pathway to gene discovery, however, has proven much longer and more difficult than anticipated. The discovery in the 1970s of the association between many diseases and the HLA system and the rapidly expanding knowledge of the biologic function of the MHC led to initial high optimism (Klein *et al.* 1983) and yet, in 2003, as many questions remain unanswered as solved. Nonetheless, MHC studies have provided an anchor point and have allowed examination of interaction between alleles of genes under investigation

with HLA genotype. Moreover, the discovery of the role of MHC molecules in presenting peptides to the antigen receptor on T lymphocytes indicated that the genetic variation in the TcR should be an important target for immunogenetic studies. This led to the work described in Chapter 3 carried out in a very well a validated and well studied group of patients with multiple sclerosis. While only slight interaction was found between MHC and TcR genotype in this study, the general principal of stratification on the basis of HLA-DRB1*1501 in that disease is important as indeed are constant attempts at refining the MHC association and applying that back to other studies. In this regard, our group have published on the association between MS and the putative HLA-DRB1 critical residue AA86Val/Gly and I have correlated the results of that study to my TCR genotyping (Teutsch *et al.* 1999). Despite the logic of stratifying for HLA, my attempts and those of many others have to date yielded few clear cut outcomes. The extensive linkage disequilibrium throughout the human genome provides both benefits and obstacles to gene discovery in disorders of complex genetics. Studies of individual polymorphisms within the structural or promoter region of a gene are unlikely to result in definitive outcomes and less polymorphism at that site is associated with clear loss or gain of function of the gene, a good example of which is provided by the *Δ32* mutation in the *CCR5* gene with homozygotes failing to express the CCR5 on the surface of cells and heterozygotes showing reduced levels. Even in this example, however, there is considerable disequilibrium within and close to the CCR5 gene which must be taken into account in interpreting disease association studies. This is particularly so since the level of CCR expression from one individual to another can vary much more considerably than the approximate half reduction associated with *CCR5-Δ32* heterozygosity. Moreover, *CCR5* is in strong linkage disequilibrium with polymorphisms in the promoter regions of

CCR5 gene and has an almost complete negative association with the clinically important mutation in the *CCR2* gene, *CCR2-64I*. As a result of these difficulties, there is strong international agreement that the optimal approach towards disease association studies of candidate genes is through the detection of all polymorphisms within the region under examination and the construction of each of the haplotypes of reasonable frequency in the population (Johnson *et al.* 2001). This point is exemplified in the data contained in Chapter 3 in which considerable disequilibrium was noted and from which I was able to construct haplotypes. At each turn, however, there are difficulties. Since the nature of the antigen binding pocket of the TcR is a product of random nucleotide insertion and "J" plus "D" segment diversity, study of the genotype of variable gene segments has limitations. It can, however, indicate genetically encoded variation in positive thymic selection by MHC and T-cell education and may also influence the extent to which superantigens can have an effect. With the findings presented in this thesis and published by others, overall it can be concluded that further study of TcR genetics in MS and other diseaes of complex genetics is warranted. Such studies, however, may need to be very large and involve the rapid throughput technologies that have become available. Advances in knowledge of the responsible peptides inserted in the MHC and presented to the TcR may also be required to a substantial advance in knowledge. No individual study, including the one presented here, is of sufficient size and power for definitive conclusion. This is unlike the circumstance with the HLA DRB1*1501 association in MS which has been shown reproducibly in many populations but is likely to be the circumstance for many if not most of the candidate gene studies outside the MHC in diseases of complex genetics.

6.2 *CCR5-Δ32* polymorphisms in autoimmune disease

There is some pleasure in studying the effect on autoimmune disease of polymorphisms of the *CCR5* gene. The discovery of these polymorphisms, initially *CCR5-Δ32*, has led to a marked advance in the understanding of the major viral epidemic at the second half of the 20th Century. If there is also a significant effect on the susceptibility to, or expression of, autoimmune disease the pleasure comes from the blurring of the margins with respect to immunogenetics of two major disease groupings and argues against the placing of autoimmune and infectious disease into separate compartments, an impression one often gets from reading the literature in this regard. The blurring may, of course, be simply due to the involvement of an infectious agent in the pathogenesis of the autoimmune disease but it would appear likely that it is much broader than this. Evidence that I have discussed in Chapter 4 indicates that the involvement of CCR5 in immune function and particularly with respect to the TH1 versus TH2 paradigm is relevant in this regard and is quite separate from the HIV viral co-receptor role. Similarly the observation that *CCR5-Δ32* heterozygotes have a different pattern of AIDS-defining opportunistic infection (Ashton *et al.* 2002) and perhaps a differing response to interleukin 2 therapy (Clegg *et al.* 2003) are not easily explained on grounds other than modulation of the immune system. It is acknowledged that a clear cut role for *CCR5* polymorphisms in autoimmune disease has not yet been established but neither has it been established unequivocally for almost any other gene outside the MHC. Mostly this is the result of the very large number of patients that need to be studied in order to provide definitive answers. In this regard, despite my studying 620 patients with type 1 diabetes (a significant cohort size on an international basis), we did not have the power to include or exclude our major hypotheses. To do so requires international collaboration. In this regard we have commenced this with

Professor John Todd at Cambridge. UK, due to his accumulation of a very large number of diabetic families. In other diseases, such as multiple sclerosis, we are involved in a 23 nation European/Australian consortium (GAMES) which has led to a genome wide screen by linkage disequilibrium using 6000 patients. This is the direction in which all future immunogenetic studies need to move and this is very much so when applied to the consideration of *CCR5* polymorphisms, including *CCR5-Δ32* in autoimmune disease. It is my hope that the data presented in Chapter 4 convinces the reader that such an approach is justified. In parallel, it goes without saying that constant monitoring of the advances of knowledge of the role of chemokine-chemokine receptor network in human biology may enable the design of more targeted experiments and perhaps ones involving fewer resources.

6.3 The origin of the *CCR5-Δ32* allele

Resolving questions with the use of both historical population (anthropological) data and genetic information has been termed "historic genomics". The data presented in this thesis provide the current best knowledge on the origin and age of the remarkable *CCR5-Δ32* allele. While limitations are acknowledged, the mapping of allele frequency to countries of origin does appear to provide a pattern of very high frequency amongst Ashkenazi Jews from north eastern Europe. In this sense, the Ashkenazi Jews of Poland and Russia could be considered the "mother load" of *CCR5-Δ32*. One weakness of our study was our lack of access to samples of non-Jewish people from this region of the world. The next best approach, the analysis of published studies based on country of residence, is reviewed within the thesis and does provide supportive evidence for the higher frequency of the mutation being found in Jewish people as compared to

non-Jewish communities. One of the strengths of the approach taken here is the availability of reliable birthplace data for parents and grandparents in our large Ashkenazi Jewish cohort. This is added to by the quite recent major migration of Jewish peoples to Sydney since the Second World War.

The conclusion that the original founder belonged to an ancestral population of contemporary Ashkenazi Jews has been strengthened by the identification here of the ancestral haplotype using six microsatellite markers and the demonstration that this haplotype is not found on non-Jewish wild-type chromosomes examined here. On this haplotype, the closely associated D3S4579 marker is found in 6% of Jewish wild-type individuals and this cannot be due to crossing-over from the CCR5-Δ32 haplotype back to one with the wild-type *CCR5* gene as the genetic distance is so small. While the associated D3S4579 marker was not seen in controls here, this allele was reported at very low frequency in caucasians by others (Chataway *et al.* 1999; Libert *et al.* 1998).

The data presented on the estimation of the coalescence time (age) of the mutation extend previous studies as a result of the use of six microsatellite markers (including the two in each of the two previous publications), the selection of two markers showing 30% haplotype retention and the much better knowledge of the genetic distances in the region over recent times. This allows a fair claim to the best available estimate in this regard of the number of generations since the mutation occurred. Conversion of this to years requires assumptions of the average duration of a generation, a figure considered to be 30 years by Tremblay and Vezina (2000). The result and conclusion of about 1200 years has then provided us with the opportunity to speculate on the population of origin. Lucotte

(1999) has proposed a Viking origin for the mutation or at least the spread of *Δ32* through the known migratory pathways of the Vikings. While this would in part explain the high frequency of *CCR5-Δ32* amongst Scandinavians and north-eastern Europeans, there are aspects that are not quite so clear. The Viking theory has also been invoked for multiple sclerosis (Poser 1994; Poser 1995), a disease for which the distribution of prevalence is somewhat different to that for *CCR5-Δ32* with a relatively low frequency in the British Isles of *Δ32* but not MS, however this could be attributed to the movement of various Viking groups in different directions.

As discussed in Chapter 5, the dating of the *Δ32* mutation may provide an explanation for a convergence of the Ashkenazi and Viking theories as both had strong interactions with the people of the Kazar kingdom who converted to Judaism in the 10th Century providing a basis for subsequent intermarriage and who traded with the Vikings over a considerable time, including the likely range of years for the origin of the *CCR5-Δ32* mutation. Clearly the Kazar theory is speculative and the paucity of recorded history of these peoples is an impediment. It is not a prime contention that the person who first acquired the deletion was a Kazar but more that the population mixture in that part of Europe, including the populations that eventually formed Ashkenazim and the Vikings, makes such an explanation plausible. The identification in this thesis of the haplotype upon which the first mutation occurred and demonstration that this is found on wild-type chromosomes in Jewish but not non-Jewish Australians has added to our conviction about the origin of this mutation. This is not to say, however, that that haplotype is derived from the original Israeli Jews. Indeed the data presented by us and others makes it clear that this mutation, showing a low allele frequency amongst Sephardic Jews,

occurred subsequent to the Great Diaspora in AD70, unless the difference could be explained on differential exposure to natural selective forces. It is equally feasible, given the dating of this mutation that the mutation arose in a population, such as the Kazars. If this is so then those peoples perhaps have made a very significant contribution to contemporary Ashkenazim and indeed to non-Jewish Caucasian populations across the world. A further study in this regard would be to examine the Y chromosome of individuals carrying the *Δ32* mutation the Kohen haplotype which is known as an ancient marker (arising perhaps 2-3,000 years ago) of Jewish ancestry. I would also have wished to have been able to examine the *CCR5-Δ32* founder haplotype on wild-type chromosomes from people of Viking origin. Unfortunately there are no published data in this regard. The finding of this haplotype at a prevalence similar to that in our Ashkenazi cohort would suggest a shared origin. The absence would infer a Khazar or Jewish origin with subsequent admixture into the Nordic peoples. Such an experiment is planned with Swedish collaborators.

Why would someone put so much effort into detailing the evolutionary history (historic genomics) of the *CCR5-Δ32* mutation? On the surface it is an obtuse point. Why should we and other groups, whose prime interest is in genetic predisposition to human disease, particularly immunologic diseases of complex genetics, divert resources towards an issue that would appear remote from direct study of disease susceptibility genes? Firstly, it is both intriguing and amenable to rigorous study; for a PhD student that is perhaps enough. But the true reason, it seems, could be much more important.

The reality to date is that very few reproducible susceptibility or resistance genes have been discovered through the current approaches of genome screening, candidate gene studies, animal models or gene expression studies. While it is true that major technological advances associated with the human genome project, SNP maps and the current HapMap project offer real advances in genome screening, it is also true that some genetic associations may not be detected by current approaches and some of these associations may provide critical insights into pathogenesis of the disease. It may also be true for some diseases that such insights will rely upon the knowledge of the environmental factor or factors that have undoubtedly played a role over centuries, if not millennia, in these multifactorial diseases of complex genetics.

The *CCR5-Δ32* mutation is an excellent example in these regards. The first outstanding aspect of this mutation is that *CCR5-Δ32* homozygotes are highly protected against HIV infection. It is this finding that established the role of CCR5 as a viral co-receptor and as the only co-receptor involved in the overall majority of HIV transmissions from one person to another. This has been one of the most important discoveries in understanding the immunopathogenesis of HIV infection and one which has led to the development of new therapeutic agents. But people with the *CCR5-Δ32* genotype are found in only 1:100 Caucasians and not at all amongst Africans or Asians, the latter two encompassing the major part of the global HIV epidemic. In Caucasians, the task for detection is therefore the demonstration that something that occurs 1:100 in the population occurs less frequently in the disease under study (HIV infection, type 1 diabetes, RA, etc). Initial belief that *CCR5-Δ32* homozygotes were fully protected were refuted by our publication of the first identified HIV infected *CCR5-Δ32* homozygote (Biti *et al.* 1997). Since then

there have been 11 other reports of such occurrence, representing perhaps a 1:1000 risk. Thus the task by traditional study would have been to show that something that occurs 1:100 in the normal population is found 1:1000 in a disease population. It is unlikely that any genome screen would be powered to identify this and a candidate gene study would have to involve larger numbers than are commonly reported. The data presented in Chapter 4 of this thesis illustrate this point.

We proposed the hypothesis that *CCR5-Δ32* homozygotes are protected against diabetes on the basis that the effect on the immunopathogenesis of this autoimmune disease due to the absence of expression of CCR5 receptor. In total 620 people with type 1 diabetes were genotyped, a not insubstantial study. The *CCR5-Δ32/Δ32* genotype was found in two of these or 0.3% (Table 4.1). By comparison, this genotype was found in 1.2% of 253 normal controls, a figure identical to that reported in many studies of populations equivalent to Australia. This difference did not reach significance. However, had the number of controls been equivalent to the number of patients the same percentages would have yielded a significant p value. To illustrate this further, had we used the data on the 807 Ashkenazi Jewish teenagers (Table 5.1) in whom a *CCR5-Δ32/Δ32* prevalence of 2.5% was found, the difference is highly significant (chi squared = 6.7, $p < .005$). The Jewish cohort are healthy young non-diabetic people. Nonetheless, they are not an appropriate control since we have demonstrated that the frequency of the *CCR5-Δ32* mutation is higher in that group. To resolve this issue, rather than extend the numbers in the study in Chapter 4, we have chosen to establish a collaboration with Professor John Todd at the University of Cambridge, UK, who has several thousand diabetic families. This second study will have greater power and will also enable replication in a second

population. If evidence is found for a protective effect against type 1 diabetes of the *CCR5-Δ32*/*Δ32* genotype then the results should be definitive. It is not a small point since this demonstration would suggest that CCR5 plays a role in most, if not all, patients with this autoimmune disease. Children with diabetes undergo a prolonged period of inflammation in the pancreas, indicated by the appearance of autoantibodies, before the complete destruction of islet function and the need for exogenous insulin therapy. Several clinical trials of novel therapies during this phase have been carried out over the years. Confirmation of a significant role of CCR5 in this circumstance should lead to a therapeutic trial. While the role of CCR5 can be examined in animal models, the most compelling evidence would come from a clear-cut demonstration of a protective effect in *CCR5-Δ32*/*Δ32* homozygotes or very convincing evidence of protection amongst heterozygotes, a less challenging goal statistically but perhaps less convincing biologically.

A similar situation arises with the possible protective effect of absence of CCR5 against rheumatoid arthritis. Gomez-Reino et al (1999) reported the absence of a *CCR5-Δ32* homozygotes amongst 673 Spanish patients with RA. The *Δ32* allele frequency in that population, however, was low at 5.8% as expected for this country (see Table 4.2) and hence the expected number of homozygotes was only 2.26 leading to an insignificant difference. Indeed in this population they would need to study several thousand people before reaching statistical significance for a protection of say 50%, a level of protection that would have major implications for the understanding of the pathogenesis of RA.

In multiple sclerosis the circumstance would appear to be different and perhaps due to significant differences in pathogenesis. In the first report of *CCR5* genotype in MS, from our group, two homozygotes were identified amongst 120 patients. A subsequent report from the USA involving a total of 600 patients showed the expected prevalence of *CCR5-Δ32* homozygotes. Neither study was able to show an association with MS with the exception of an apparent delayed onset of MS in *CCR5-Δ32* heterozygotes with familial MS (Barcellos *et al.* 2000). In our study we also stratified patients on the basis of HLA-DRB1 1501 and T-cell receptor genotype (using the data presented in Chapter 3). I could find no associations between *CCR5-Δ32* and TcR genotype (Bennetts *et al.* 1997). It has to be conceded, however, that a possibility of a partial protection against MS, whilst unlikely, has not been excluded in studies involving in total 720 patients. It would seem an unwise investment of resources, however, to pursue it further in this disease, unlike the circumstance with type 1 diabetes.

With these difficulties with respect to autoimmune disease, how is it possible to be conclusive with respect to resistance to HIV infection? The answer is straightforward. Unlike autoimmune disease, the environmental agent for this infection is known and tests are available for its presence. The modes of exposure are also well understood and hence it is easy to identify individuals who have been exposed but have not become infected. In the initial demonstration, the role of CCR5 was in the in vitro study of two men who were highly exposed to HIV but uninfected. Their cells could not be infected in vitro and they both lacked CCR5 on their cell surface. Subsequent population studies, however, were critical showing a high level of protection for *CCR5-Δ32* homozygotes. More importantly, it was the second phase that established the importance of CCR5 in almost

all transmissions from one person to another. This was a critical observation since CXCR4 is a very effective alternate co-receptor. Thus the large population studies played an essential role. Moreover, studies of cohorts of men who are highly exposed but not infected showed a marked increase in the prevalence of *CCR5-Δ32* homozygotes (Liu *et al.* 1996).

Perhaps this is just too good an example of the scientific benefits to be gained from studying the interaction between an identified environmental agent and host genetic factors. But at some point it is to be hoped that environmental agents will be identified for autoimmune diseases such as MS and type 1 diabetes and that such discovery in addition to guiding therapeutic strategies against the environmental agent, will also inform knowledge of the genetic factors and through this provide even more insight into pathogenesis leading again perhaps to a range of effective therapies and perhaps development of combination strategies.

One of the potential criticism of the work presented in Chapter 5 is the difficulty with accurately dating the *CCR5-Δ32* mutation. This difficulty rests on two uncertain assumptions, the exact genetic distance between *CCR5* and the microsatellite under study and the assumption of the average generation time. Such calculations, however, have been carried out on many other mutations. In a recent article in Nature Reviews, Ostrer (2001) mapped several mutations in Jewish populations on the basis of calculated age (coalescence time), the likely population in which the mutation occurred and the key historic event (if known) for that population close to the estimated founder time. Examples cited include the type III mutation in the factor XI gene, which Ostrer

estimated occurred 120 generations ago, a time that would fit the period of Jewish residence in Palestine, and two mutations in the glucocerebrosidase gene (Gaucher disease) that are estimated at about 50 generations ago, a time that coincides with the migration of Jews in to the Rhineland and the founding of the Ashkenazi Jewish group (Diaz *et al.* 2000) (see below). In more recent times, the delGly197 mutation in the low density lipoprotein receptor gene is found principally amongst Ashkenazi Jews whose ancestors came from Lithuania. The coalescence time of 20 generations coincides with the time when Jews were granted charters by the King of Poland to live in Lithuania. These, and other examples, serve to indicate the potential for mapping mutations onto populations and knowledge of human history (Ostrer 2001). Analysis of mutations have also been used to show evidence of Jewish ancestry amongst contemporary non-Jewish groups. This includes the study of the 185delAG BRCA1 in affected individuals in New Mexico, Mexico and El Salvador, a mutation in the growth hormone receptor gene in individuals with Laron dwarfism in southern Ecudaor. Ostrer observes that these examples are compatible with known migrations from Spain and Portugal to Latin America following the Spanish Inquisition at the end of the 15th Century.

These examples illustrate a growing science of historic genomics to which it is hoped the work in this thesis has made some contribution. Since it would appear highly likely that the *CCR5-Δ32* mutation has achieved its current prevalence in about 1200 years due to marked natural selection, it appears plausible that the identification of that selective force could provide unique insight into the function of the immune system in both health and disease. While much of the focus on this mutation has focussed on HIV disease, there is growing evidence of effects beyond the viral co-receptor role both in HIV disease and in

other disorders.

This is not to suggest that there are close parallels between the analyses of the monogenic disease causing mutations in Jewish people discussed above and the *CCR5-Δ32* mutation as clearly there are substantial differences. The other mutations do not appear to confer selective advantage with their current prevalence being influenced by genetic drift with many populations going through substantial bottlenecks followed by rapid population expansion. Some of the Jewish mutations, however, may have conferred selective advantage as evidenced by the high prevalence of mutations in the Beta-globin gene that caused thalassaemia amongst Kurdish Jews likely to be influenced by selection for resistance to malaria. It is also proposed that resistance to cholera or typhoid may have conferred advantage on heterozygotes for cystic fibrosis mutations. With regard to malaria, pregnancy outcomes have been found to be affected by monocyte accumulation directed by chemokines (Abrams *et al.* 2003; Tkachuk *et al.* 2001) and this disease may have been spread widely in historic times (Reiter 2000).

The other significant difference with respect to *CCR5-Δ32* is that there is no evidence for it being associated with a disease state. The *CCR5-Δ32* homozygotes appear to be phenotypically normal and studies by others and in this thesis show that the proportion of homozygotes is that which would be predicted by Hardy-Weinberg equilibrium.

It would seem likely that the selective advantage conferred by the *CCR5-Δ32* mutation acts both in heterozygotes and homozygotes, although it is possible that the effect is greater or broader for homozygotes. While homozygotes express no cell surface CCR5,

heterozygotes express approximately half normal levels. More importantly, there is good evidence that this reduced (but not absent) level of expression of the CCR5 receptor has measurable effects on health of people with HIV infection and perhaps others. Several studies, including one from our laboratory, have shown that *CCR5-Δ32* heterozygotes have a different pattern of AIDS defining opportunistic infection than those of wild-type genotype (Ashton *et al.* 2002) and there is evidence that response to IL2 therapy in HIV disease is influenced by CCR5 genotype (Clegg *et al.* 2003). In this thesis, data is presented supporting an effect of *CCR5-Δ32* heterozygosity on diabetes and there are several reports in other autoimmune disorders as summarised in Chapter 4. Further study therefore of the immune and other biologic function of human individuals who carry one or two copies of the *CCR5-Δ32* mutation may in the future provide knowledge that could be mapped onto the historic genomics discussed here. In addition, the interaction of the chemokine-chemokine receptor network in infectious diseases, autoimmune states and other diseases may provide knowledge that can be further added in this regard (Hill 2001).

6.4 Selection and the *CCR5-Δ32* allele

There is a role for selection by pathogens in the population dynamics of the various alleles of *CCR5* as the protection from HIV infection afforded to homozygotes for *CCR5-Δ32* shows (Dean *et al.* 1996; Liu *et al.* 1996; Samson *et al.* 1996b). Others have considered this question of natural selection for *CCR5-Δ32*. While plague has been suggested as a factor that may have been involved in the selection for *CCR5-Δ32* (Stephens *et al.* 1998), the limited time period for the "Black Death" and questions of the nature of the actual pathogen involved (Scott and Duncan 2001) means another agent (or

agents) for selection should be considered. There is a strong suggestion of a role for viral pathogens such as members of the poxvirus family (Lalani *et al.* 1999; Mahalingam and Karupiah 2000) in selection of *CCR5* alleles, and even a possible role for pathogens in diseases such as malaria (Reiter 2000; Tkachuk *et al.* 2001). Whereas the use of CCR5 by HIV is in the physical attachment to the cell for the purpose of viral entry, poxvirus use of chemokine ligand and receptor gene sequences that are found in poxvirus viral genomes is directed more at causing non-directed activation of the immune cell's chemokine receptor, thus interfering with specific immune cell motility (Barrett *et al.* 2001; Masters *et al.* 2001).

Polymorphism in the coding region of a gene that in some way alters function thereby allows selection to act. An equally important arena for polymorphism to be considered in is in the overall expression of a gene product. Selection of promoter region alleles is known from environmental effects on lactate dehydrogenase-B in fish where it was found that even a small number of mutations within important regulatory sequences provide for variation in expression with a significant impact on environmental adaption (Schulte 2001). The argument has been made by Olson that selection which favours the loss of a gene entirely is in fact an engine of evolutionary change (Olson 1999) because the reverse mutation that returns a mutated gene into a functional one is more readily available for use by a species than the *de novo* stepwise mutation events needed to alter a different gene into the now vacant function. The 32-bp deletion mutation in *CCR5* is taken as a case in point by Olson, although it does not yet seem that selection will adversely affect all those who bear the wild-type CCR5 receptor. The importance of immune system genes in particular for the effects of promoter-region polymorphisms has been recently

reviewed (Mitchison 2000) and it was noted that while selection will produce "nested SNPs" that can serve as an indicator of evolutionary forces at work, other forces will however tend to conserve important sequences as well. Chemokine receptor genes have been shown to have important mutations in promoter regions affecting the outcome of or even susceptibility to infection by HIV (Carrington *et al.* 1999; Carrington *et al.* 1997; McDermott *et al.* 1998; Mummidi *et al.* 1998; Mummidi *et al.* 1997). There is also very strong (almost complete) linkage disequilibrium between *CCR5-Δ32* and a promoter region polymorphism which influences HIV disease (Clegg *et al.* 2000). Conversely *CCR5-Δ32* and *CCR2-64I* have not been observed on the same chromosome (Clegg *et al.* 2000). Progression of infection by HIV has been found to be affected as well by mutation of promoter activity for the chemokine RANTES, where increased expression levels did not prevent infection but did delay disease progression (Liu *et al.* 1999).

Table 6.1 List of selected genes included in the 3p21 region between afmb362wb9 and D3S1578

gene name	gene description	Build 31 location (kbp)
CCRs 1, 2, 3 & 5	assorted chemokine receptors	~ 45, 500
LTF	lactotransferrin	45,656 - 685
PTHR1	human parathyroid hormone receptor	46,090 - 117
MYL3	myosin light chain 1	46,131 - 136
TSP50	testes-specific protease 50	46,255 - 291
SCAP	Sterol regulatory element binding protein cleavage-activating protein	46,675 - 738
CSPG5	neuroglycan C	46,809 - 825
MAP4	Similar to microtubule-associated protein 4	47,175 - 196
CDC25A	Cell division cycle 25A	47,482 - 512
CAMP	Antibacterial protein FALL-39 precursor	47,548 - 550
NM23 - H6	Nucleoside diphosphate kinase 6	47,618 - 623
PLXNB1	Plexin-B1/SEP receptor precursor	47,728 - 749
TREX1	3'-5' exonuclease TREX1	47,771 - 792
PFKFB4	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4	47,843 - 877
COL7A1	Collagen alpha 1	47,885 - 916
CELSR3	Cadherin EGF LAG seven-pass G-type receptor 3 precursor	47,957 - 983
AF3P21	SH3 adapter protein SPIN90	47,983 - 48,006
IP6K2	Similar to mammalian inositol hexakisphosphate kinase 2	48,014 - 037
PRKAR2A	cAMP-dependent protein kinase type II-alpha regulatory chain	48,070 - 084
IMPDH2	Inosine-5'-monophosphate dehydrogenase	48,210 - 215
LAMB2	Laminin beta-2 chain precursor	48,476 - 487
QARS	Glutaminyl-tRNA synthetase	48,504 - 511

Table 6.1 (continued)

gene name	gene description	location (kbp)
ARHA	Transforming protein RhoA	48,526 - 542
USP4	Ubiquitin carboxyl-terminal hydrolase 4	48,551 - 609
BSN	Neuronal double zinc finger protein	48,696 - 813
AMT	Aminomethyltransferase, mitochondrial precursor	48,945 - 951
TCTA	T-cell leukemia translocation-associated gene protein	48,951 - 955
MAPKAPK3	Mitogen activated protein kinase activated protein kinase-3	49,470 - 502
CISH	Cytokine-inducible SH2-containing protein	49,508 - 513
FUS1	Fus-1 protein (lung cancer candidate)	49,791 - 795
HYAL3	Hyaluronidase 3	49,820 - 827
IFDR2	Interferon-related IFRD2	49,826 - 832
MST1R	Macrophage-stimulating protein receptor precursor	50,216 - 232
GRM2	Metabotropic glutamate receptor 2 precursor	50,997 - 51,006
ACY1	Aminoacylase-1	51,271 - 276
TLR9	Toll-like receptor 9 precursor	51,508 - 513
BAP1	BRCA1 associated protein 1	51,688 - 696
TNNC1	Troponin C, slow skeletal & cardiac muscle	51,738 - 741
NEK4	Serine/threonine-protein kinase NEK4	51,996 - 52,056
ITIH4	Inter-alpha-trypsin inhibitor heavy chain H4 precursor	52,099 - 116
PRKCD	Protein kinase C, delta type	52,465 - 478
TKT	Transketolase	52,511 - 541
CACNA1D	Voltage-dependent L-type calcium channel alpha-1D subunit	52,780 - 53,097
IL17BR	Interleukin-17B receptor precursor	53,132 - 151

6.5 And what of other 3p21 region genes?

So far in this final discussion I have only considered selection related to loss of function of the CCR5 receptor itself. I have identified however an ancestral haplotype with linkage disequilibrium that extends from the afmb362wb9 microsatellite marker to the D3S1578 microsatellite marker and includes *CCR5-Δ32* as well as four other polymorphic microsatellite markers. It is not possible to exclude the alternate hypothesis that natural selection for *CCR5-Δ32* has been wholly, or in part, due to the *Δ32* allele hitchhiking with an allele of a different gene within this haplotype. A list of selected genes on 3p21 between afmb362wb9 and D3S1578 is shown in Table 6.1. These genes control a wide range of important biologic functions. In many, there is biologic plausibility to selective advantage from gain of function or loss of function mutations.

In consideration of what nearby genes or genomic structures could be under intense selection in this region, apart from *CCR5*, there are of course a number of chemokine receptor genes that are located in close proximity (Daugherty and Springer 1997; Samson *et al.* 1996c) and which have functions similar to the CCR5 receptor that would allow for independent selection. The region just telomeric of *CCR5* is also of major importance for cancer biology due to the deletion mutations in 3p21 common to a number of tumours found there (Kiss *et al.* 2001; Kiss *et al.* 2002). The possible involvement of chemokine receptors in tumour biology is currently being examined (Gerard and Rollins 2001) and it is not surprising that some amount of tumour metastasis has already been found to be chemokine driven (Lu *et al.* 2003; Wolf *et al.* 2003). Indeed, the involvement of chemokines in tumour biology is being studied for possible therapeutic outcomes through redirecting the migration of T-cells to chemokines secreted by the tumour (Kershaw *et al.*

2002). Other genes in the region close to CCR5 with possible allelic variants that appear involved in some human disease include the human parathyroid hormone receptor, *PTHR1*, and the macrophage-stimulating protein receptor, *MSTR1*.

While the chemokine receptors in 3p21 are clustered to one end of the region in question, another gene that might be of interest in cancer biology, *BAP1* (BRCA1 associated protein 1), is located between D3S663 and D3S1578. While genes this far away from CCR5 are not likely to have played a role in helping carry *CCR5* as a hitch-hiker gene appearing to be selected, it is possible that specific polymorphisms of *BAP1*, or even of the Interleukin-17B receptor precursor located just centromeric of D3S1578, may have hitch-hiked along by selection of CCR5.

The list of genes in Table 6.1, selected from an otherwise long list, and the intriguing nature of the rapid rise in allele frequency for *CCR5-Δ32* in little more than one millennium, highlights both the size and the importance of the task of unravelling knowledge of the human genome of direct relevance to understanding human disease. Provided here is perhaps both a simple and clear example of the formidable nature of the task. In the immediate future it would seem more profitable to pursue first the direct effects of variation in CCR5 function while awaiting publication of remarkable developments in knowledge of the effects in health and disease of genetic variation in any of the genes listed in Table 6.1. For those already involved in the study of these genes, knowledge of the extent of the linkage disequilibrium and of the origin of the *CCR5-Δ32* allele may influence future experiments.

6.6 Future directions

As indicated elsewhere in this thesis, it is my conviction that the eventual deciphering of the genes that influence the major human diseases discussed in this thesis will require the bringing together of knowledge in a field broader than focussed analysis of repetitive genotyping of a particular disease state. Such studies are, of course, essential as they drive hypotheses but in the generation of hypotheses there must be consideration of all aspects of immune function including its evolution over millennia and its ancestry in lower species. Fortunately such comparisons are becoming more feasible in the post-human and other species genome project era.

The advent of vertebrate self-defined immunity came with a price. Variations in the level of expression of self epitopes or in their nature could give rise to inappropriate immune responses against "self", and yet not having the flexability to introduce as necessary some degree of polymorphism yields the advantage back to the pathogens. The future will see knowledge of genomic evolution tell a story of how pathogens have molded the different vertebrate genomes in their various ways. From the limited number of species that have been well mapped to date, some comparisons have begun to be made about the major immune system components such as MHC (Klein *et al.* 1993a; Kumnovics *et al.* 2003; Sambrook *et al.* 2002), TcR (Charmley *et al.* 1995; Jaeger *et al.* 1998; Lai *et al.* 1988; Rast *et al.* 1995; Su *et al.* 1999) and Ig (Marchalonis *et al.* 2001) between species. Not far behind are studies in comparison of chemokine and chemokine receptor genes (Fujiki *et al.* 1999; Mummidi *et al.* 2000; Nomiyama *et al.* 2001; Tang *et al.* 1999; Zhang *et al.* 1999) between verious vertebrate species. There are, of course, a large number of immune system components under selective pressure from pathogens and driven to

polymorphism of structure or variation in expression in return (Davis and Hamilton 1998; Murphy 1993; Pennacchio and Rubin 2001). Some immune components are seen to have a unique degree of variation in one species or another (Ballingall *et al.* 2001; Zelus *et al.* 2000), yet there is also pressure to retain the basic function of an immune-related subsystem between species (Dascher *et al.* 2002).

When future studies allow, it may be possible to "de-construct" evolution through understanding the various pressures on genomic regions that have built up species barriers and how various modifications of genomes across the range of vertebrates has occurred. Over the interval of time since the pufferfish and humans have shared a common ancestor, it is estimated that between 4,000 and 16,000 genomic rearrangements have occurred (McLysaght *et al.* 2000). In some ways, this is not a great deal of change. The current push in human genomic studies is towards a "hap-map", or mapping out the haplotypes involved in different populations and at different genomic regions rather than single genes and individual markers (Phillips *et al.* 2003). While this sort of analysis will one day be available across all the vertebrate species, current studies are available for phylogenetic analysis of many of the individual genes concerned, as described elsewhere in this thesis, and, in cases like the MHC (O'Brien and Yuhki 1999), effects of selection on genomic evolution of that region have been well modelled (Ahmad *et al.* 2003).

Experimental directions for the more immediate future have been discussed within each chapter in this thesis. There is a common theme relating to the importance of selection of patients with first class case ascertainment, the need for international collaboration and meta-analysis that allows the power for the detection of genes of small effect and the

ability to reproduce results in many populations, the careful selection of control groups for case control studies and the recruitment of families for other analyses, the ability to examine gene interactions while retaining power for stratification, consideration of gene-environment interactions, the need to correlate genotype with biologic phenotype and eventually with the presence or clinical expression of disease. With the rapid expansion of essential knowledge on a broad front and the availability of rapid throughput technologies, coupled to this list, there are grounds for optimism.

References:

Abecasis G.R., Noguchi E., Heinzmann A., Traherne J.A., Bhattacharyya S., Leaves N.I., Anderson G.G., Zhang Y., Lench N.J., Carey A., Cardon L.R., Moffatt M.F., Cookson W.O. (2001) Extent and distribution of linkage disequilibrium in three genomic regions. *Am J Hum Genet* 68:191-197.

Abi-Rached L., Gilles A., Shiina T., Pontarotti P., Inoko H. (2002) Evidence of en bloc duplication in vertebrate genomes. *Nat Genet* 31:100-5.

Abrams E.T., Brown H., Chensue S.W., Turner G.D., Tadesse E., Lema V.M., Molyneux M.E., Rochford R., Meshnick S.R., Rogerson S.J. (2003) Host response to malaria during pregnancy: placental monocyte recruitment is associated with elevated beta chemokine expression. *J Immunol* 170:2759-64.

Adalsteinsson S. (1985) Possible changes in the frequency of the human ABO blood groups in Iceland due to smallpox epidemics selection. *Ann Hum Genet* 49:275-81.

Ahmad T., Neville M., Marshall S.E., Armuzzi A., Mulcahy-Hawes K., Crawshaw J., Sato H., Ling K.L., Barnardo M., Goldthorpe S., Walton R., Bunce M., Jewell D.P., Welsh K.I. (2003) Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC. *Hum Mol Genet* 12:647-56.

Aizenberg H., Barlow-Stewart K., Burnett L., Proos A. (2001) The Gene Broker - A Generic Model for DNA Databanking. *Bulletin Human Genet. Soc. Aust.*

Alkhatib G., Combadiere C., Broder C.C., Feng Y., Kennedy P.E., Murphy P.M., Berger E.A. (1996) CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* 272:1955-8.

Allison D.B. (1997) Transmission-disequilibrium tests for quantitative traits. *Am J Hum Genet* 60:676-90.

Allison D.B., Heo M., Kaplan N., Martin E.R. (1999) Sibling-based tests of linkage and association for quantitative traits. *Am J Hum Genet* 64:1754-63.

Alperovitch A., Hors J., Lyon-Caen O., MS F.R.G.o. (1992) Multiple sclerosis in 54 twinships: concordance rate is independent of zygosity. French Research Group on Multiple Sclerosis. *Ann Neurol* 32:724-7.

Amores A., Force A., Yan Y.L., Joly L., Amemiya C., Fritz A., Ho R.K., Langeland J., Prince V., Wang Y.L., Westerfield M., Ekker M., Postlethwait J.H. (1998) Zebrafish hox clusters and vertebrate genome evolution. *Science* 282:1711-4.

Anderton S.M., Viner N.J., Matharu P., Lowrey P.A., Wraith D.C. (2002) Influence of a dominant cryptic epitope on autoimmune T cell tolerance. *Nat Immunol* 3:175-81.

Ansel K.M., Cyster J.G. (2001) Chemokines in lymphopoiesis and lymphoid organ development. *Curr Opin Immunol* 13:172-9.

Apic G., Gough J., Teichmann S.A. (2001) An insight into domain combinations. *Bioinformatics* 17:S83-9.

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

Ardlie K., Liu-Cordero S.N., Eberle M.A., Daly M., Barrett J., Winchester E., Lander E.S., Kruglyak L. (2001) Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. *Am J Hum Genet* 69:582-9.

Ashton L.J., Stewart G.J., Biti R., Law M., Cooper D.A., Kaldor J.M. (2002) Heterozygosity for CCR5-Delta32 but not CCR2b-64I protects against certain intracellular pathogens. *HIV Med* 3:91-6.

Bach F.H. (1976) Genetics of transplantation: the major histocompatibility complex. *Annu Rev Genet* 10:319-39

Bailey J.A., Gu Z., Clark R.A., Reinert K., Samonte R.V., Schwartz S., Adams M.D., Myers E.W., Li P.W., Eichler E.E. (2002) Recent segmental duplications in the human

genome. *Science* 297:1003-7.

Ballingall K.T., Waibochi L., Holmes E.C., Woelk C.H., MacHugh N.D., Lutje V., McKeever D.J. (2001) The CD45 locus in cattle: allelic polymorphism and evidence for exceptional positive natural selection. *Immunogenetics* 52:276-83

Barcellos L.F., Schito A.M., Rimmler J.B., Vittinghoff E., Shih A., Lincoln R., Callier S., Elkins M.K., Goodkin D.E., Haines J.L., Pericak-Vance M.A., Hauser S.L., Oksenberg J.R. (2000) CC-chemokine receptor 5 polymorphism and age of onset in familial multiple sclerosis. Multiple Sclerosis Genetics Group. *Immunogenetics* 51:281-8

Barclay A.N. (1999) Ig-like domains: evolution from simple interaction molecules to sophisticated antigen recognition. *Proc Natl Acad Sci U S A* 96:14672-4.

Barrett J.W., Cao J.X., Hota-Mitchell S., McFadden G. (2001) Immunomodulatory proteins of myxoma virus. *Semin Immunol* 13:73-84.

Batchelor J.R., Compston A., McDonald W.I. (1978) The significance of the association between HLA and multiple sclerosis. *Br Med Bull* 34:279-84.

Beall S.S., Biddison W.E., McFarlin D.E., McFarland H.F., Hood L.E. (1993) Susceptibility for multiple sclerosis is determined, in part, by inheritance of a 175-kb region of the TcR V beta chain locus and HLA class II genes. *J Neuroimmunol* 45:53-60.

Beall S.S., Concannon P., Charmley P., McFarland H.F., Gatti R.A., Hood L.E., McFarlin D.E., Biddison W.E. (1989) The germline repertoire of T cell receptor beta-chain genes in patients with chronic progressive multiple sclerosis. *J Neuroimmunol* 21:59-66.

Beck S., Trowsdale J. (2000) The human major histocompatibility complex: Lessons from the DNA Sequence. *Annu. Rev. Genomics Hum. Genet.* 1:117-137

Becker K.G. (1999) Comparative genetics of type 1 diabetes and autoimmune disease: common loci, common pathways? *Diabetes* 48:1353-8

Becker K.G., Simon R.M., Bailey-Wilson J.E., Freidlin B., Biddison W.E., McFarland H.F., Trent J.M. (1998) Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. *Proc Natl Acad Sci U S A* 95:9979-84

Benkirane M., Jin D.Y., Chun R.F., Koup R.A., Jeang K.T. (1997) Mechanism of transdominant inhibition of CCR5-mediated HIV-1 infection by ccr5delta32. *J Biol Chem* 272:30603-6.

Bennetts B.H., Teutsch S.M., Buhler M.M., Heard R.N., Stewart G.J. (1997) The CCR5 deletion mutation fails to protect against multiple sclerosis. *Hum Immunol* 58:52-9

Bennetts B.H., Teutsch S.M., Buhler M.M., Heard R.N., Stewart G.J. (1999) HLA-DMB gene and HLA-DRA promoter region polymorphisms in Australian multiple sclerosis patients. *Hum Immunol* 60:886-93

Bennetts B.H., Teutsch S.M., Heard R.N., Dunckley H., Stewart G.J. (1995) TAP2 polymorphisms in Australian multiple sclerosis patients. *J Neuroimmunol* 59:113-21.

Berger E.A., Murphy P.M., Farber J.M. (1999) Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol* 17:657-700

Bergsteinsdottir K., Yang H.T., Pettersson U., Holmdahl R. (2000) Evidence for common autoimmune disease genes controlling onset, severity, and chronicity based on experimental models for multiple sclerosis and rheumatoid arthritis. *J Immunol* 164:1564-8.

Billstrom-Schroeder M., Worthen G.S. (2001) Viral Regulation of RANTES Expression during Human Cytomegalovirus Infection of Endothelial Cells. *J. Virol.* 75:3383-3390

Biti R., Ffrench R., Young J., Bennetts B., Stewart G., Liang T. (1997) HIV-1 infection in an individual homozygous for the CCR5 deletion allele. *Nat Med* 3:252-3.

Blanpain C., Lee B., Tackoen M., Puffer B., Boom A., Libert F., Sharron M., Wittamer

V., Vassart G., Doms R.W., Parmentier M. (2000) Multiple nonfunctional alleles of CCR5 are frequent in various human populations. *Blood* 96:1638-45

Bontrop R.E., Otting N., Slierendregt B.L., Lanchbury J.S. (1995) Evolution of major histocompatibility complex polymorphisms and T-cell receptor diversity in primates. *Immunol Rev* 143:33-62.

Boot-Handford R.P., Tuckwell D.S. (2003) Fibrillar collagen: The key to vertebrate evolution? A tale of molecular incest. *Bioessays* 25:142-51.

Bradley L.M., Asensio V.C., Schioetz L.K., Harbertson J., Krahl T., Patstone G., Woolf N., Campbell I.L., Sarvetnick N. (1999) Islet-specific Th1, but not Th2, cells secrete multiple chemokines and promote rapid induction of autoimmune diabetes. *J Immunol* 162:2511-20

Braun M., Wunderlin M., Spieth K., Knochel W., Gierschik P., Moepps B. (2002) Xenopus laevis Stromal cell-derived factor 1: conservation of structure and function during vertebrate development. *J Immunol* 168:2340-7.

Breen E.C., Gage J.R., Guo B., Magpantay L., Narazaki M., Kishimoto T., Miles S., Martinez-Maza O. (2001) Viral interleukin 6 stimulates human peripheral blood B cells that are unresponsive to human interleukin 6. *Cell Immunol* 212:118-25.

Briant L., Avoustin P., Clayton J., McDermott M., Clanet M., Cambon-Thomsen A. (1993) Multiple sclerosis susceptibility: population and twin study of polymorphisms in the T-cell receptor beta and gamma genes region. French Group on Multiple Sclerosis. *Autoimmunity* 15:67-73

Brinkmann B., Klintschar M., Neuhuber F., Huhne J., Rolf B. (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* 62:1408-15.

Broman K.W., Murray J.C., Sheffield V.C., White R.L., Weber J.L. (1998) Comprehensive human genetic maps: individual and sex-specific variation in

recombination. *Am J Hum Genet* 63:861-9.

Broman K.W., Weber J.L. (1999) Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am J Hum Genet* 65:1493-500.

Bromley S.K., Peterson D.A., Gunn M.D., Dustin M.L. (2000) Cutting edge: hierarchy of chemokine receptor and TCR signals regulating T cell migration and proliferation. *J Immunol* 165:15-9.

Brook K.A. (1999) The Jews of Khazaria. Jason Aronson, Inc., Northvale, NJ

Buhler M., Craig M., Donaghue K.C., Badhwar P., Willis J., Manolios N., Tait B.D., Silink M., Bennetts B., Stewart G.J. (2002) CCR5 genotyping in an Australian and New Zealand type 1 diabetes cohort. *Autoimmunity* 35:457-61.

Buhler M.M. (2001) Could Immune Evolution Put A Cap On Vertebrate Gene Numbers? GeneMappers Workshop, Human Genetics Society of Australasia 25th meeting., Cairns, QLD

Buhler M.M., Bennetts B.H., Heard R.N., Stewart G.J. (2000) T cell receptor beta chain genotyping in Australian relapsing-remitting multiple sclerosis patients. *Mult Scler* 6:140-7.

Calabrese P.P., Durrett R.T., Aquadro C.F. (2001) Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation models. *Genetics* 159:839-52.

Cameron M.J., Arreaza G.A., Grattan M., Meagher C., Sharif S., Burdick M.D., Strieter R.M., Cook D.N., Delovitch T.L. (2000) Differential expression of CC chemokines and the CCR5 receptor in the pancreas is associated with progression to type I diabetes. *J Immunol* 165:1102-10

Campagnoni A.T., Pribyl T.M., Campagnoni C.W., Kampf K., Amur-Umarjee S., Landry

C.F., Handley V.W., Newman S.L., Garbay B., Kitamura K. (1993) Structure and developmental regulation of Golli-mbp, a 105-kilobase gene that encompasses the myelin basic protein gene and is expressed in cells in the oligodendrocyte lineage in the brain. *J Biol Chem* 268:4930-8.

Campbell J.D., Stinson M.J., Simons F.E., Rector E.S., HayGlass K.T. (2001) In vivo stability of human chemokine and chemokine receptor expression. *Hum Immunol* 62:668-78.

Carfi A., Smith C.A., Smolak P.J., McGrew J., Wiley D.C. (1999) Structure of a soluble secreted chemokine inhibitor vCCI (p35) from cowpox virus. *Proc Natl Acad Sci U S A* 96:12379-83.

Carrington M., Dean M., Martin M.P., O'Brien S.J. (1999) Genetics of HIV-1 infection: chemokine receptor CCR5 polymorphism and its consequences. *Hum Mol Genet* 8:1939-45

Carrington M., Kissner T., Gerrard B., Ivanov S., O'Brien S.J., Dean M. (1997) Novel alleles of the chemokine-receptor gene CCR5. *Am J Hum Genet* 61:1261-7

Charmley P., Beall S.S., Concannon P., Hood L., Gatti R.A. (1991) Further localization of a multiple sclerosis susceptibility gene on chromosome 7q using a new T cell receptor beta-chain DNA polymorphism. *J Neuroimmunol* 32:231-40.

Charmley P., Chao A., Concannon P., Hood L., Gatti R.A. (1990) Haplotyping the human T-cell receptor beta-chain gene complex by use of restriction fragment length polymorphisms. *Proc Natl Acad Sci U S A* 87:4823-7.

Charmley P., Concannon P. (1993) Polymorphism and phylogeny of dinucleotide repeats in human T-cell receptor Vb6 genes. *Immunogenetics* 38:92-7

Charmley P., Keretan E., Snyder K., Clark E.A., Concannon P. (1995) Relative size and evolution of the germline repertoire of T-cell receptor beta-chain gene segments in nonhuman primates. *Genomics* 25:150-6

Charron D.J., Engelman E.G., Benike C.J., McDevitt H.O. (1980) Ia antigens on alloreactive T cells in man detected by monoclonal antibodies. Evidence for synthesis of HLA-D/DR molecules of the responder type. *J Exp Med* 152:127s-136s.

Charron D.J., McDevitt H.O. (1980) Characterization of HLA-D-region antigens by two-dimensional gel electrophoresis. Molecular-genotyping. *J Exp Med* 152:18s-36s.

Chataway J., Feakes R., Coraddu F., Gray J., Deans J., Fraser M., Robertson N., Broadley S., Jones H., Clayton D., Goodfellow P., Sawcer S., Compston A. (1998) The genetics of multiple sclerosis: principles, background and updated results of the United Kingdom systematic genome screen. *Brain* 121:1869-87.

Chataway J., Sawcer S., Feakes R., Coraddu F., Broadley S., Gray J., Compston A. (1999) More evidence that founder effects exist in the European population. *Eur J Hum Genet* 7:623-4

Chen M.C., Proost P., Gysemans C., Mathieu C., Eizirik D.L. (2001) Monocyte chemoattractant protein-1 is expressed in pancreatic islets from prediabetic NOD mice and in interleukin-1 beta-exposed human and rat islet cells. *Diabetologia* 44:325-32.

Clark A.G., Weiss K.M., Nickerson D.A., Taylor S.L., Buchanan A., Stengard J., Salomaa V., Vartiainen E., Perola M., Boerwinkle E., Sing C.F. (1998) Haplotype structure and population genetic inferences from nucleotide- sequence variation in human lipoprotein lipase. *Am J Hum Genet* 63:595-612.

Clegg A., Williamson P., Biti R., Cooper D., Emery S., Carr A., Stewart G. (2003) Chemokine receptor genotype and response to interleukin-2 therapy in HIV-1-infected individuals. *Clin Immunol* 106:36-40.

Clegg A.O., Ashton L.J., Biti R.A., Badhwar P., Williamson P., Kaldor J.M., Stewart G.J. (2000) CCR5 promoter polymorphisms, CCR5 59029A and CCR5 59353C, are under represented in HIV-1-infected long-term non-progressors. The Australian Long-Term Non-Progressor Study Group. *Aids* 14:103-8.

Compston A. (2000) The genetics of multiple sclerosis. *J Neurovirol* 6 Suppl 2:S5-9.

Compston D.A., Kellar Wood H., Robertson N., Sawcer S., Wood N.W. (1995) Genes and susceptibility to multiple sclerosis. *Acta Neurol Scand Suppl* 161:43-51

Concannon P., Gogolin-Ewens K.J., Hinds D.A., Wapelhorst B., Morrison V.A., Stirling B., Mitra M., Farmer J., Williams S.R., Cox N.J., Bell G.I., Risch N., Spielman R.S. (1998) A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. *Nat Genet* 19:292-6

Cooke S.P., Forrest G., Venables P.J., Hajeer A. (1998) The delta32 deletion of CCR5 receptor in rheumatoid arthritis. *Arthritis Rheum* 41:1135-6

Correia-Neves M., Mathis D., Benoist C. (2001) A molecular chart of thymocyte positive selection. *Eur J Immunol* 31:2583-92.

Cox N.J., Wapelhorst B., Morrison V.A., Johnson L., Pinchuk L., Spielman R.S., Todd J.A., Concannon P. (2001) Seven regions of the genome show evidence of linkage to type 1 diabetes in a consensus analysis of 767 multiplex families. *Am J Hum Genet* 69:820-30.

Craddock T.P., Zumla A.M., Ollier W.E., Chintu C.Z., Muyinda G.P., Lancaster F.C., Boylston A.W. (2000) Predominance of one T-cell antigen receptor BV haplotype in African populations. *Immunogenetics* 51:231-7.

Craig M.E., Howard N.J., Silink M., Chan A. (2000) The rising incidence of childhood type 1 diabetes in New South Wales, Australia. *J Pediatr Endocrinol Metab* 13:363-72.

Cudworth A.G., Wolf E., Gorsuch A.N., Festenstein H. (1979) A new look at HLA genetics with particular reference to type-1 diabetes. *Lancet* 2:389-91.

Dascher C.C., Hiromatsu K., Xiong X., Sugita M., Buhlmann J.E., Dodge I.L., Lee S.Y., Roura-Mir C., Watts G.F., Roy C.J., Behar S.M., Clemens D.L., Porcelli S.A., Brenner M.B. (2002) Conservation of CD1 intracellular trafficking patterns between mammalian

species. *J Immunol* 169:6951-8.

Daugherty B.L., Springer M.S. (1997) The beta-chemokine receptor genes CCR1 (CMKBR1), CCR2 (CMKBR2), and CCR3 (CMKBR3) cluster within 285 kb on human chromosome 3p21. *Genomics* 41:294-5.

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

Davis W.C., Hamilton M.J. (1998) Comparison of the unique characteristics of the immune system in different species of mammals. *Vet Immunol Immunopathol* 63:7-13.

de Roda Husman A.M., Blaak H., Brouwer M., Schuitemaker H. (1999) CC chemokine receptor 5 cell-surface expression in relation to CC chemokine receptor 5 genotype and the clinical course of HIV-1 infection. *J Immunol* 163:4597-603

Dean M., Carrington M., Winkler C., Huttley G.A., Smith M.W., Allikmets R., Goedert J.J., Buchbinder S.P., Vittinghoff E., Gomperts E., Donfield S., Vlahov D.*, et al.* (1996) Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. [published erratum appears in Science 1996 Nov 15;274(5290):1069]. *Science* 273:1856-62

Delarbre C., Jaulin C., Kourilsky P., Gachelin G. (1992) Evolution of the major histocompatibility complex: a hundred-fold amplification of MHC class I genes in the African pigmy mouse Nannomys setulosus. *Immunogenetics* 37:29-38

Deng H., Liu R., Ellmeier W., Choe S., Unutmaz D., Burkhart M., Di Marzio P., Marmon S., Sutton R.E., Hill C.M., Davis C.B., Peiper S.C.*, et al.* (1996) Identification of a major co-receptor for primary isolates of HIV-1. *Nature* 381:661-6

Dermitzakis E.T., Clark A.G. (2001) Differential selection after duplication in mammalian developmental genes. *Mol Biol Evol* 18:557-62.

DeWan A.T., Parrado A.R., Matise T.C., Leal S.M. (2001) The Map Problem: A Comparison of Genetic and Sequence-Based Physical Maps. *Am J Hum Genet* 70:1

Diaz G.A., Gelb B.D., Risch N., Nygaard T.G., Frisch A., Cohen I.J., Miranda C.S., Amaral O., Maire I., Poenaru L., Caillaud C., Weizberg M., Mistry P., Desnick R.J. (2000) Gaucher disease: the origins of the Ashkenazi Jewish N370S and 84GG acid beta-glucosidase mutations. *Am J Hum Genet* 66:1821-32.

Dib C., Faure S., Fizames C., Samson D., Drouot N., Vignal A., Millasseau P., Marc S., Hazan J., Seboun E., Lathrop M., Gyapay G., Morissette J., Weissenbach J. (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152-4.

Doherty P.C., Zinkernagel R.M. (1975) A biological role for the major histocompatibility antigens. *Lancet* 1:1406-9.

Donahue J.P., Ricalton N.S., Behrendt C.E., Rittershaus C., Calaman S., Marrack P., Kappler J.W., Kotzin B.L. (1994) Genetic analysis of low V beta 3 expression in humans. *J Exp Med* 179:1701-6.

Donaldson I.J., Shefta J., Hill A.V., Boylston A.W. (2002a) Malaria is not responsible for the selection of TCR beta-subunit variable gene haplotypes in The Gambia. *Immunogenetics* 53:894-9.

Donaldson I.J., Shefta J., Lawson C.A., Bushnell J.R., Morgan A.W., Isaacs J.D., Carpenter D., Shaw M.A., Rooth I., Quinnell R.J., Zumla A.M., Ollier W.R.*, et al.* (2002b) Unique TCR beta-subunit variable gene haplotypes in Africans. *Immunogenetics* 53:884-93.

Dorman S.E., Holland S.M. (2000) Interferon-gamma and interleukin-12 pathway defects and human disease. *Cytokine Growth Factor Rev* 11:321-33.

Dragic T. (2001) An overview of the determinants of CCR5 and CXCR4 co-receptor function. *J Gen Virol* 82:1807-14.

Dragic T., Litwin V., Allaway G.P., Martin S.R., Huang Y., Nagashima K.A., Cayanan C., Maddon P.J., Koup R.A., Moore J.P., Paxton W.A. (1996) HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC- CKR-5. *Nature* 381:667-73.

Droogan A.G., Kirk C.W., Hawkins S.A., McMillan S.A., Nevin N.C., Graham C.A. (1996) T-cell receptor alpha, beta, gamma, and delta chain gene microsatellites show no association with multiple sclerosis. *Neurology* 47:1049-53.

Dubois Laforgue D., Hendel H., Caillat-Zucman S., Zagury J.F., Winkler C., Boitard C., Timsit J. (2001) A common stromal cell-derived factor-1 chemokine gene variant is associated with the early onset of type 1 diabetes. *Diabetes* 50:1211-3.

Dunon D., Imhof B.A. (2000) The role of cell traffic in the emergence of the T lymphoid system. *Semin Immunol* 12:429-33.

Durand D. (2003) Vertebrate evolution: doubling and shuffling with a full deck. *Trends Genet* 19:2-5.

Duret L. (2001) Why do genes have introns? Recombination might add a new piece to the puzzle. *Trends Genet* 17:172-175.

Dyment D.A., Sadovnick A.D., Ebers G.C., Sadnovich A.D. (1997) Genetics of multiple sclerosis. *Hum Mol Genet* 6:1693-8

Easterbrook P.J., Rostron T., Ives N., Troop M., Gazzard B.G., Rowland-Jones S.L. (1999) Chemokine receptor polymorphisms and human immunodeficiency virus disease progression. *J Infect Dis* 180:1096-105.

Ebers G.C., Bulman D.E., Sadovnick A.D., Paty D.W., Warren S., Hader W., Murray T.J., Seland T.P., Duquette P., Grey T., et al. (1986) A population-based study of multiple sclerosis in twins. *N Engl J Med* 315:1638-42.

Ebers G.C., Kukay K., Bulman D.E., Sadovnick A.D., Rice G., Anderson C., Armstrong H., Cousin K., Bell R.B., Hader W., Paty D.W., Hashimoto S.*, et al.* (1996) A full

genome search in multiple sclerosis. *Nat Genet* 13:472-6.

Ebers G.C., Sadovnick A.D., Risch N.J. (1995) A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group. *Nature* 377:150-1.

Eikelenboom M.J., Killestein J., Izeboud T., Kalkers N.F., van Lier R.A., Barkhof F., Uitdehaag B.M., Polman C.H. (2002) Chemokine receptor expression on T cells is related to new lesion development in multiple sclerosis. *J Neuroimmunol* 133:225-32.

Eisenbarth I., Vogel G., Krone W., Vogel W., Assum G. (2000) An isochore transition in the NF1 gene region coincides with a switch in the extent of linkage disequilibrium. *Am J Hum Genet* 67:873-80.

Encinas J.A., Kuchroo V.K. (2000) Mapping and identification of autoimmunity genes. *Curr Opin Immunol* 12:691-7.

Epplen C., Jackel S., Santos E.J., D'Souza M., Poehlau D., Dotzauer B., Sindern E., Haupts M., Rude K.P., Weber F., Stover J., Poser S.*, et al.* (1997) Genetic predisposition to multiple sclerosis as revealed by immunoprinting. *Ann Neurol* 41:341-52

Farrall M., Weeks D.E. (1998) Mutational mechanisms for generating microsatellite allele-frequency distributions: an analysis of 4,558 markers. *Am J Hum Genet* 62:1260-2.

Favorova O.O., Andreewski T.V., Boiko A.N., Sudomoina M.A., Alekseenkov A.D., Kulakova O.G., Slanova A.V., Gusev E.I. (2002) The chemokine receptor CCR5 deletion mutation is associated with MS in HLA-DR4-positive Russians. *Neurology* 59:1652-5.

Friedman R., Hughes A.L. (2001) Pattern and timing of gene duplication in animal genomes. *Genome Res* 11:1842-7.

Friedman R., Hughes A.L. (2003) The temporal distribution of gene duplication events in a set of highly conserved human gene families. *Mol Biol Evol* 20:154-61.

Frisse L., Hudson R.R., Bartoszewicz A., Wall J.D., Donfack J., Di Rienzo A. (2001) Gene Conversion and Different Population Histories May Explain the Contrast between Polymorphism and Linkage Disequilibrium Levels. *Am J Hum Genet* 69:831-43.

Fugger L., Sandberg-Wollheim M., Morling N., Ryder L.P., Svejgaard A. (1990) The germline repertoire of T-cell receptor beta chain genes in patients with relapsing/remitting multiple sclerosis or optic neuritis. *Immunogenetics* 31:278-80

Fujiki K., Shin D.H., Nakao M., Yano T. (1999) Molecular cloning of carp (Cyprinus carpio) CC chemokine, CXC chemokine receptors, allograft inflammatory factor-1, and natural killer cell enhancing factor by use of suppression subtractive hybridization. *Immunogenetics* 49:909-14

Garred P., Madsen H.O., Petersen J., Marquart H., Hansen T.M., Freiesleben Sorensen S., Volck B., Svejgaard A., Andersen V. (1998) CC chemokine receptor 5 polymorphism in rheumatoid arthritis. *J Rheumatol* 25:1462-5

Gautam A.M., Lock C.B., Smilek D.E., Pearson C.I., Steinman L., McDevitt H.O. (1994) Minimum structural requirements for peptide presentation by major histocompatibility complex class II molecules: implications in induction of autoimmunity. *Proc Natl Acad Sci U S A* 91:767-71.

George A.J. (2002) Is the number of genes we possess limited by the presence of an adaptive immune system? *Trends Immunol* 23:351-5.

Gerard C., Rollins B.J. (2001) Chemokines and disease. *Nat Immunol* 2:108-15.

Gibson T.J., Spring J. (1998) Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins. *Trends Genet* 14:46-9; discussion 49-50.

Gibson T.J., Spring J. (2000) Evidence in favour of ancient octaploidy in the vertebrate genome. *Biochem Soc Trans* 28:259-64.

Gill B.M., Jaramillo A., Ma L., Laupland K.B., Delovitch T.L. (1995) Genetic linkage of thymic T-cell proliferative unresponsiveness to mouse chromosome 11 in NOD mice. A possible role for chemokine genes. *Diabetes* 44:614-9

Glass W.G., Lane T.E. (2003) Functional expression of chemokine receptor CCR5 on CD4(+) T cells during virus-induced central nervous system disease. *J Virol* 77:191-8.

Godiska R., Chantry D., Dietsch G.N., Gray P.W. (1995) Chemokine expression in murine experimental allergic encephalomyelitis. *J Neuroimmunol* 58:167-76.

Goldrath A.W., Bevan M.J. (1999) Selecting and maintaining a diverse T-cell repertoire. *Nature* 402:255-62.

Gomez-Reino J.J., Pablos J.L., Carreira P.E., Santiago B., Serrano L., Vicario J.L., Balsa A., Figueroa M., de Juan M.D. (1999) Association of rheumatoid arthritis with a functional chemokine receptor, CCR5. *Arthritis Rheum* 42:989-92

Gruen J.R., Weissman S.M. (1997) Evolving views of the major histocompatibility complex. *Blood* 90:4252-65.

Gu X., Wang Y., Gu J. (2002) Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nat Genet* 31:205-9.

Haase C.G., Schmidt S., Faustmann P.M. (2002) Frequencies of the G-protein beta3 subunit C825T polymorphism and the delta 32 mutation of the chemokine receptor-5 in patients with multiple sclerosis. *Neurosci Lett* 330:293-5.

Haines J.L., Ter-Minassian M., Bazyk A., Gusella J.F., Kim D.J., Terwedow H., Pericak-Vance M.A., Rimmler J.B., Haynes C.S., Roses A.D., Lee A., Shaner B.*, et al.* (1996) A complete genomic screen for multiple sclerosis underscores a role for the major histocompatability complex. The Multiple Sclerosis Genetics Group. *Nat Genet* 13:469-71.

Hajeer A.H., Hutchinson I.V. (2001) Influence of TNFalpha gene polymorphisms on TNFalpha production and disease. *Hum Immunol* 62:1191-9.

Halle J.P., Haus-Seuffert P., Woltering C., Stelzer G., Meisterernst M. (1997) A conserved tissue-specific structure at a human T-cell receptor beta- chain core promoter. *Mol Cell Biol* 17:4220-9.

Hammer M.F., Redd A.J., Wood E.T., Bonner M.R., Jarjanazi H., Karafet T., Santachiara-Benerecetti S., Oppenheim A., Jobling M.A., Jenkins T., Ostrer H., Bonne-Tamir B. (2000) Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc Natl Acad Sci U S A* 97:6769-74.

Hatta Y., Tsuchiya N., Matsushita M., Shiota M., Hagiwara K., Tokunaga K. (1999) Identification of the gene variations in human CD22. *Immunogenetics* 49:280-6.

Hausmann S., Wucherpfennig K.W. (1997) Activation of autoreactive T cells by peptides from human pathogens. *Curr Opin Immunol* 9:831-8.

Hawke N.A., Yoder J.A., Litman G.W. (1999) Expanding our understanding of immunoglobulin, T-cell antigen receptor, and novel immune-type receptor genes: a subset of the immunoglobulin gene superfamily. *Immunogenetics* 50:124-33.

He B., Xu C., Yang B., Landtblom A.M., Fredrikson S., Hillert J. (1998) Linkage and association analysis of genes encoding cytokines and myelin proteins in multiple sclerosis. *J Neuroimmunol* 86:13-9.

Hellings P.W., Vandenberghe P., Kasran A., Coorevits L., Overbergh L., Mathieu C., Ceuppens J.L. (2002) Blockade of CTLA-4 enhances allergic sensitization and eosinophilic airway inflammation in genetically predisposed mice. *Eur J Immunol* 32:585-94.

Helminen M., Lahdenpohja N., Hurme M. (1999) Polymorphism of the interleukin-10 gene is associated with susceptibility to Epstein-Barr virus infection. *J Infect Dis* 180:496-9.

Helminen M.E., Kilpinen S., Virta M., Hurme M. (2001) Susceptibility to primary Epstein-Barr virus infection is associated with interleukin-10 gene promoter polymorphism. *J Infect Dis* 184:777-80.

Hess M. (1976) Ia antigens: isolation, chemical modification and structural characterization. *Transplant Rev* 30:40-63

Hill A.V. (2001) The genomics and genetics of human infectious disease susceptibility. *Annu Rev Genomics Hum Genet* 2:373-400

Hillert J., Leng C., Olerup O. (1991) No association with germline T cell receptor beta-chain gene alleles or haplotypes in Swedish patients with multiple sclerosis. *J Neuroimmunol* 32:141-7.

Hillert J., Olerup O. (1992) Germ-line polymorphism of TCR genes and disease susceptibility--fact or hypothesis? *Immunol Today* 13:47-9.

Hockertz M.K., Paty D.W., Beall S.S. (1998) Susceptibility to relapsing-progressive multiple sclerosis is associated with inheritance of genes linked to the variable region of the TcR beta locus: use of affected family-based controls. *Am J Hum Genet* 62:373-85.

Holloway A.F., Rao S., Shannon M.F. (2002) Regulation of cytokine gene transcription in the immune system. *Mol Immunol* 38:567-80.

Homey B., Wang W., Soto H., Buchanan M.E., Wiesenborn A., Catron D., Muller A., McClanahan T.K., Dieu-Nosjean M.C., Orozco R., Ruzicka T., Lehmann P., Oldham E., Zlotnik A. (2000) Cutting edge: the orphan chemokine receptor G protein-coupled receptor- 2 (GPR-2, CCR10) binds the skin-associated chemokine CCL27 (CTACK/ALP/ILC). *J Immunol* 164:3465-70

Huang Q.R., Teutsch S.M., Buhler M.M., Bennetts B.H., RN S.H., Manolios N., Stewart G.J. (2000) Evaluation of the Apo-1/Fas promoter Mva I polymorphism in multiple sclerosis. *Mult Scler* 6:14-18

Hundert G.D., Bacon G.C. (1984) The Jews in Poland and Russia. Indiana University Press, Bloomington

Ioannidis J.P., Rosenberg P.S., Goedert J.J., Ashton L.J., Benfield T.L., Buchbinder S.P., Coutinho R.A., Eugen-Olsen J., Gallart T., Katzenstein T.L., Kostrikis L.G., Kuipers H., *et al.* (2001) Effects of CCR5-Delta32, CCR2-64I, and SDF-1 3'A Alleles on HIV-1 Disease Progression: An International Meta-Analysis of Individual- Patient Data. *Ann Intern Med* 135:782-795.

Isomaki P., Panesar M., Annenkov A., Clark J.M., Foxwell B.M., Chernajovsky Y., Cope A.P. (2001) Prolonged exposure of T cells to TNF down-regulates TCR zeta and expression of the TCR/CD3 complex at the cell surface. *J Immunol* 166:5495-507.

Jaeger E.E., Bontrop R.E., Parham P., Wickings E.J., Kadwell M., Lanchbury J.S. (1998) Characterization of chimpanzee TCRV gene polymorphism: how old are human TCRV alleles? *Immunogenetics* 47:115-23

Jagodzinski P.P., Lecybyl R., Ignacak M., Juszczyk J., Trzeciak W.H. (2000) Distribution of 32 alelle of the CCR5 gene in the population of Poland. *J Hum Genet* 45:271-4

Jakobsen I.B., Wilson S.R., Easteal S. (1998) Patterns of reticulate evolution for the classical class I and II HLA loci. *Immunogenetics* 48:312-23.

Jardetzky T.S., Brown J.H., Gorga J.C., Stern L.J., Urban R.G., Chi Y.I., Stauffacher C., Strominger J.L., Wiley D.C. (1994) Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen. *Nature* 368:711-8.

Jawaheer D., Seldin M.F., Amos C.I., Chen W.V., Shigeta R., Monteiro J., Kern M., Criswell L.A., Albani S., Nelson J.L., Clegg D.O., Pope R., *et al.* (2001) A genomewide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. *Am J Hum Genet* 68:927-36.

Johansson S., Lie B.A., Todd J.A., Pociot F., Nerup J., Cambon-Thomsen A., Kockum I.,

Akselsen H.E., Thorsby E., Undlien D.E. (2003) Evidence of at least two type 1 diabetes susceptibility genes in the HLA complex distinct from HLA-DQB1, -DQA1 and -DRB1. *Genes Immun* 4:46-53.

Johnson G.C., Esposito L., Barratt B.J., Smith A.N., Heward J., Di Genova G., Ueda H., Cordell H.J., Eaves I.A., Dudbridge F., Twells R.C., Payne F.*, et al.* (2001) Haplotype tagging for the identification of common disease genes. *Nat Genet* 29:233-7.

Jun H.S., Yoon J.W. (2001) The role of viruses in type I diabetes: two distinct cellular and molecular pathogenic mechanisms of virus-induced diabetes in animals. *Diabetologia* 44:271-85.

Kantor R., Gershoni J.M. (1999) Distribution of the CCR5 gene 32-base pair deletion in Israeli ethnic groups. *J Acquir Immune Defic Syndr Hum Retrovirol* 20:81-4

Karlsson M.G., Lawesson S.S., Ludvigsson J. (2000) Th1-like dominance in high-risk first-degree relatives of type I diabetic patients. *Diabetologia* 43:742-9

Karpus W.J., Lukacs N.W., McRae B.L., Strieter R.M., Kunkel S.L., Miller S.D. (1995) An important role for the chemokine macrophage inflammatory protein-1 alpha in the pathogenesis of the T cell-mediated autoimmune disease, experimental autoimmune encephalomyelitis. *J Immunol* 155:5003-10.

Kaufman J. (1999) Co-evolving genes in MHC haplotypes: the "rule" for nonmammalian vertebrates? *Immunogenetics* 50:228-36. 0.htm

Kay R.A. (1996) TCR gene polymorphisms and autoimmune disease. *Eur J Immunogenet* 23:161-77

Kershaw M.H., Wang G., Westwood J.A., Pachynski R.K., Tiffany H.L., Marincola F.M., Wang E., Young H.A., Murphy P.M., Hwu P. (2002) Redirecting migration of T cells to chemokine secreted from tumors by genetic modification with CXCR2. *Hum Gene Ther* 13:1971-80.

Khani-Hanjani A., Hoar D., Horsman D., Keown P. (2000) Identification of four novel dinucleotide repeat polymorphisms in the TNF-alpha and TNF-beta genes. *Hum Immunol* 61:511-2.

Kimura M., Ohta T. (1973) The age of a neutral mutant persisting in a finite population. *Genetics* 75:199-212.

Kiss H., Kedra D., Kiss C., Kost-Alimova M., Yang Y., Klein G., Imreh S., Dumanski J.P. (2001) The LZTFL1 gene is a part of a transcriptional map covering 250 kb within the common eliminated region 1 (C3CER1) in 3p21.3. *Genomics* 73:10-9.

Kiss H., Yang Y., Kiss C., Andersson K., Klein G., Imreh S., Dumanski J.P. (2002) The transcriptional map of the common eliminated region 1 (C3CER1) in 3p21.3. *Eur J Hum Genet* 10:52-61.

Klein J., Figueroa F. (1986) Evolution of the major histocompatibility complex. *Crit Rev Immunol* 6:295-386

Klein J., Figueroa F., Nagy Z.A. (1983) Genetics of the major histocompatibility complex: the final act. *Annu Rev Immunol* 1:119-42

Klein J., O'HUigin C., Figueroa F., Mayer W.E., Klein D. (1993a) Different modes of Mhc evolution in primates. *Mol Biol Evol* 10:48-59.

Klein J., Sato A. (1998) Birth of the major histocompatibility complex. *Scand J Immunol* 47:199-209.

Klein J., Satta Y., O'HUigin C., Takahata N. (1993b) The molecular descent of the major histocompatibility complex. *Annu Rev Immunol* 11:269-95

Kleinjan D.J., van Heyningen V. (1998) Position effect in human genetic disease. *Hum Mol Genet* 7:1611-8

Klitz W., Brautbar C., Schito A.M., Barcellos L.F., Oksenberg J.R. (2001) Evolution of

the CCR5 delta32 mutation based on haplotype variation in Jewish and Northern European population samples. *Hum Immunol* 62:530-8.

Koch H.G., McClay J., Loh E.W., Higuchi S., Zhao J.H., Sham P., Ball D., Craig I.W. (2000) Allele association studies with SSR and SNP markers at known physical distances within a 1 Mb region embracing the ALDH2 locus in the Japanese, demonstrates linkage disequilibrium extending up to 400 kb. *Hum Mol Genet* 9:2993-9.

Koestler A. (1976) The Thirteenth Tribe: The Khazar Empire and its Heritage. Hutchinson & Co., London

Kostyu D.D., Hannick L.I., Traweek J.L., Ghanayem M., Heilpern D., Dawson D.V. (1997) HLA class I polymorphism: structure and function and still questions. *Hum Immunol* 57:1-18.

Krakowski M., Owens T. (1996) Interferon-gamma confers resistance to experimental allergic encephalomyelitis. *Eur J Immunol* 26:1641-6.

Kristiansen O.P., Larsen Z.M., Pociot F. (2000) CTLA-4 in autoimmune diseases--a general susceptibility gene to autoimmunity? *Genes Immun* 1:170-84.

Kukreja A., Cost G., Marker J., Zhang C., Sun Z., Lin-Su K., Ten S., Sanz M., Exley M., Wilson B., Porcelli S., Maclaren N. (2002) Multiple immuno-regulatory defects in type-1 diabetes. *J Clin Invest* 109:131-40.

Kumnovics A., Takada T., Lindahl K.F. (2003) Genomic organization of the Mammalian MHC. *Annu Rev Immunol* 21:629-57

Kuokkanen S., Gschwend M., Rioux J.D., Daly M.J., Terwilliger J.D., Tienari P.J., Wikstrom J., Palo J., Stein L.D., Hudson T.J., Lander E.S., Peltonen L. (1997) Genomewide scan of multiple sclerosis in Finnish multiplex families. *Am J Hum Genet* 61:1379-87.

Kurt R.A., Bauck M., Harma S., McCulloch K., Baher A., Urba W.J. (2001) Role of C

chemokine lymphotactin in mediating recruitment of antigen- specific CD62L(lo) cells in vitro and in vivo. *Cell Immunol* 209:83-8.

Lafaille J.J., Keere F.V., Hsu A.L., Baron J.L., Haas W., Raine C.S., Tonegawa S. (1997) Myelin basic protein-specific T helper 2 (Th2) cells cause experimental autoimmune encephalomyelitis in immunodeficient hosts rather than protect them from the disease. *J Exp Med* 186:307-12.

Lahiri D.K., Nurnberger J.I., Jr. (1991) A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res* 19:5444.

Lai E., Concannon P., Hood L. (1988) Conserved organization of the human and murine T-cell receptor beta- gene families. *Nature* 331:543-6.

Laird D.J., De Tomaso A.W., Cooper M.D., Weissman I.L. (2000) 50 million years of chordate evolution: seeking the origins of adaptive immunity. *Proc Natl Acad Sci U S A* 97:6924-6.

Lalani A.S., Masters J., Zeng W., Barrett J., Pannu R., Everett H., Arendt C.W., McFadden G. (1999) Use of chemokine receptors by poxviruses. *Science* 286:1968-71

Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., Funke R., Gage D.*, et al.* (2001) Initial sequencing and analysis of the human genome. International Human Genome Sequencing Consortium. *Nature* 409:860-921.

Lehninger A.L. (1970) Biochemistry; the molecular basis of cell structure and function. Worth Publishers, New York

Letunic I., Copley R.R., Bork P. (2002) Common exon duplication in animals and its role in alternative splicing. *Hum Mol Genet* 11:1561-7.

Leung J.Y., McKenzie F.E., Uglialoro A.M., Flores-Villanueva P.O., Sorkin B.C., Yunis

E.J., Hartl D.L., Goldfeld A.E. (2000) Identification of phylogenetic footprints in primate tumor necrosis factor-alpha promoters. *Proc Natl Acad Sci U S A* 97:6614-8.

Liao D. (1999) Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet* 64:24-30.

Libert F., Cochaux P., Beckman G., Samson M., Aksenova M., Cao A., Czeizel A., Claustres M., de la Rua C., Ferrari M., Ferrec C., Glover G.*, et al.* (1998) The deltaccr5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe. *Hum Mol Genet* 7:399-406

Limborska S.A., Balanovsky O.P., Balanovskaya E.V., Slominsky P.A., Schadrina M.I., Livshits L.A., Kravchenko S.A., Pampuha V.M., Khusnutdinova E.K., Spitsyn V.A. (2002) Analysis of CCR5Delta32 Geographic Distribution and Its Correlation with Some Climatic and Geographic Factors. *Hum Hered* 53:49-54

Liu H., Chao D., Nakayama E.E., Taguchi H., Goto M., Xin X., Takamatsu J.K., Saito H., Ishikawa Y., Akaza T., Juji T., Takebe Y.*, et al.* (1999) Polymorphism in RANTES chemokine promoter affects HIV-1 disease progression. *Proc Natl Acad Sci U S A* 96:4581-5.

Liu R., Paxton W.A., Choe S., Ceradini D., Martin S.R., Horuk R., MacDonald M.E., Stuhlmann H., Koup R.A., Landau N.R. (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 86:367-77.

Lonnrot M., Korpela K., Knip M., Ilonen J., Simell O., Korhonen S., Savola K., Muona P., Simell T., Koskela P., Hyoty H. (2000) Enterovirus infection as a risk factor for beta-cell autoimmunity in a prospectively observed birth cohort: the Finnish Diabetes Prediction and Prevention Study. *Diabetes* 49:1314-8.

Louzoun Y., Atlan H., Cohen I.R. (2001) Modeling the Influence of TH1- and TH2-type Cells in Autoimmune Diseases. *J Autoimmun* 17:311-21.

Love P.E., Lee J., Shores E.W. (2000) Critical relationship between TCR signaling potential and TCR affinity during thymocyte selection. *J Immunol* 165:3080-7.

Lu P., Nakamoto Y., Nemoto-Sasaki Y., Fujii C., Wang H., Hashii M., Ohmoto Y., Kaneko S., Kobayashi K., Mukaida N. (2003) Potential Interaction between CCR1 and Its Ligand, CCL3, Induced by Endogenously Produced Interleukin-1 in Human Hepatomas. *Am J Pathol* 162:1249-58.

Lucotte G. (1997) Frequencies of the CC chemokine receptor 5 delta 32 allele in various populations of defined racial background. *Biomed Pharmacother* 51:469-73

Lucotte G. (2001) Distribution of the CCR5 gene 32-basepair deletion in West Europe. A hypothesis about the possible dispersion of the mutation by the vikings in historical times. *Hum Immunol* 62:933-6.

Lucotte G. (2002) Frequencies of 32 base pair deletion of the (delta 32) allele of the CCR5 HIV-1 co-receptor gene in Caucasians: a comparative analysis. *Infection, Genetics and Evolution* 1:201-205

Lucotte G., Mercier G. (1998) Distribution of the CCR5 gene 32-bp deletion in Europe. *J Acquir Immune Defic Syndr Hum Retrovirol* 19:174-7.

Lucotte G., Mercier G., Smets P. (1999) Elevated frequencies of the mutant allele delta32 at the CCR5 gene in Ashkenazic Jews. *AIDS Res Hum Retroviruses* 15:1-2

Luppi P., Zanone M.M., Hyoty H., Rudert W.A., Haluszczak C., Alexander A.M., Bertera S., Becker D., Trucco M. (2000) Restricted TCR V beta gene expression and enterovirus infection in type I diabetes: a pilot study. *Diabetologia* 43:1484-97.

Lusso P. (2000) Chemokines and viruses: the dearest enemies. *Virology* 273:228-40.

Luther S.A., Cyster J.G. (2001) Chemokines as regulators of T cell differentiation. *Nat Immunol* 2:102-7.

Lynch M., Conery J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290:1151-5.

Lynch S.G., Rose J.W., Petajan J.H., Stauffer D., Kamerath C., Leppert M. (1991) Discordance of T-cell receptor beta-chain genes in familial multiple sclerosis. *Ann Neurol* 30:402-10.

Maayan S., Zhang L., Shinar E., Ho J., He T., Manni N., Kostrikis L.G., Neumann A.U. (2000) Evidence for recent selection of the CCR5-delta 32 deletion from differences in its frequency between Ashkenazi and Sephardi Jews. *Genes Immun* 1:358-61.

Mackay C.R. (2001) Chemokines: immunology's high impact factors. *Nat Immunol* 2:95-101.

Magierowska M., Lepage V., Boubnova L., Carcassi C., de Juan D., Djoulah S., El Chenawi F., Grunnet N., Halle L., Ivanova R., Jungerman M., Naumova E.*, et al.* (1998) Distribution of the CCR5 gene 32 base pair deletion and SDF1-3'A variant in healthy individuals from different populations. *Immunogenetics* 48:417-9

Magierowska M., Theodorou I., Debre P., Sanson F., Autran B., Riviere Y., Charron D., Costagliola D. (1999) Combined genotypes of CCR5, CCR2, SDF1, and HLA genes can predict the long-term nonprogressor status in human immunodeficiency virus-1- infected individuals. *Blood* 93:936-41.

Mahalingam S., Karupiah G. (2000) Modulation of chemokines by poxvirus infections. *Curr Opin Immunol* 12:409-12

Majka M., Ratajczak J., Lee B., Honczarenko M., Douglas R., Kowalska M.A., Silberstein L., Gewirtz A.M., Ratajczak M.Z. (2000) The role of HIV-related chemokine receptors and chemokines in human erythropoiesis in vitro. *Stem Cells* 18:128-38

Manoury B., Mazzeo D., Fugger L., Viner N., Ponsford M., Streeter H., Mazza G., Wraith D.C., Watts C. (2002) Destructive processing by asparagine endopeptidase limits presentation of a dominant T cell epitope in MBP. *Nat Immunol* 3:169-74.

Marchalonis J.J., Adelman M.K., Robey I.F., Schluter S.F., Edmundson A.B. (2001) Exquisite specificity and peptide epitope recognition promiscuity, properties shared by antibodies from sharks to humans. *J Mol Recognit* 14:110-121.

Martin A.P., Pardini A.T., Noble L.R., Jones C.S. (2002) Conservation of a dinucleotide simple sequence repeat locus in sharks. *Mol Phylogenet Evol* 23:205-13.

Martinez-Caceres E.M., Barrau M.A., Brieva L., Espejo C., Barbera N., Montalban X. (2002) Treatment with methylprednisolone in relapses of multiple sclerosis patients: immunological evidence of immediate and short-term but not long-lasting effects. *Clin Exp Immunol* 127:165-71.

Martinez-Mir A., Canestro C., Gonzalez-Duarte R., Albalat R. (2001) Characterization of the amphioxus presenilin gene in a high gene- density genomic region illustrates duplication during the vertebrate lineage. *Gene* 279:157-64.

Martinez-Naves E., Victoria-Gutierrez M., Uria D.F., Lopez-Larrea C. (1993) The germline repertoire of T cell receptor beta-chain genes in multiple sclerosis patients from Spain. *J Neuroimmunol* 47:9-13.

Martinson J.J., Chapman N.H., Rees D.C., Liu Y.T., Clegg J.B. (1997) Global distribution of the CCR5 gene 32-basepair deletion. *Nat Genet* 16:100-3

Martinson J.J., Hong L., Karanicolas R., Moore J.P., Kostrikis L.G. (2000) Global distribution of the CCR2-64I/CCR5-59653T HIV-1 disease- protective haplotype. *Aids* 14:483-9

Masters J., Hinek A.A., Uddin S., Platanias L.C., Zeng W., McFadden G., Fish E.N. (2001) Poxvirus infection rapidly activates tyrosine kinase signal transduction. *J Biol Chem* 276:48371-5.

Matis L.A. (1990) The molecular basis of T-cell specificity. *Annu Rev Immunol* 8:65-82

Matsui M., Weaver J., Proudfoot A.E., Wujek J.R., Wei T., Richer E., Trapp B.D., Rao

A., Ransohoff R.M. (2002) Treatment of experimental autoimmune encephalomyelitis with the chemokine receptor antagonist Met-RANTES. *J Neuroimmunol* 128:16-22.

McDermott D.H., Zimmerman P.A., Guignard F., Kleeberger C.A., Leitman S.F., Murphy P.M. (1998) CCR5 promoter polymorphism and HIV-1 disease progression. Multicenter AIDS Cohort Study (MACS). *Lancet* 352:866-70.

McDevitt H.O. (1980) Regulation of the immune response by the major histocompatibility system. *N Engl J Med* 303:1514-7.

McDevitt H.O. (1984) Speculations on how Ia antigens (Ir genes) influence the specificity of the immune response. *Ann Immunol (Paris)* 135D:227-36.

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

McDevitt H.O., Bodmer W.F. (1974) HL-A, immune-response genes, and disease. *Lancet* 1:1269-75.

McLysaght A., Enright A.J., Skrabanek L., Wolfe K.H. (2000) Estimation of synteny conservation and genome compaction between pufferfish (Fugu) and human. *Yeast* 17:22-36.

McLysaght A., Hokamp K., Wolfe K.H. (2002) Extensive genomic duplication during early chordate evolution. *Nat Genet* 31:200-4.

McPherson J.D., Marra M., Hillier L., Waterston R.H., Chinwalla A., Wallis J., Sekhon M., Wylie K., Mardis E.R., Wilson R.K., Fulton R., Kucaba T.A.*, et al.* (2001) A physical map of the human genome. The International Human Genome Mapping Consortium. *Nature* 409:934-41.

Mein C.A., Esposito L., Dunn M.G., Johnson G.C., Timms A.E., Goy J.V., Smith A.N., Sebag-Montefiore L., Merriman M.E., Wilson A.J., Pritchard L.E., Cucca F.*, et al.* (1998) A search for type 1 diabetes susceptibility genes in families from the United

Kingdom. *Nat Genet* 19:297-300

Menashe I., Man O., Lancet D., Gilad Y. (2002) Population differences in haplotype structure within a human olfactory receptor gene cluster. *Hum Mol Genet* 11:1381-90.

Merriman T.R., Cordell H.J., Eaves I.A., Danoy P.A., Coraddu F., Barber R., Cucca F., Broadley S., Sawcer S., Compston A., Wordsworth P., Shatford J.*, et al.* (2001) Suggestive evidence for association of human chromosome 18q12-q21 and its orthologue on rat and mouse chromosome 18 with several autoimmune diseases. *Diabetes* 50:184-94.

Messier W., Li S.H., Stewart C.B. (1996) The birth of microsatellites. *Nature* 381:483.

Metzgar D., Bytof J., Wills C. (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res* 10:72-80.

Michael N.L. (1999) Host genetic influences on HIV-1 pathogenesis. *Curr Opin Immunol* 11:466-74.

Minguillon C., Ferrier D.E., Cebrian C., Garcia-Fernandez J. (2002) Gene duplications in the prototypical cephalochordate amphioxus. *Gene* 287:121-8.

Mitchison N.A. (2000) Polymorphism in regulatory gene sequences. *Genome Biol* 2
Miyagishi R., Kikuchi S., Fukazawa T., Tashiro K. (1995) Macrophage inflammatory protein-1 alpha in the cerebrospinal fluid of patients with multiple sclerosis and other inflammatory neurological diseases. *J Neurol Sci* 129:223-7.

Moore K.W., O'Garra A., de Waal Malefyt R., Vieira P., Mosmann T.R. (1993) Interleukin-10. *Annu Rev Immunol* 11:165-90

Morahan G., Huang D., Ymer S.I., Cancilla M.R., Stephen K., Dabadghao P., Werther G., Tait B.D., Harrison L.C., Colman P.G. (2001) Linkage disequilibrium of a type 1 diabetes susceptibility locus with a regulatory IL12B allele. *Nat Genet* 27:218-21.

Moser B., Loetscher P. (2001) Lymphocyte traffic control by chemokines. *Nat Immunol* 2:123-8.

Mumford C.J., Wood N.W., Kellar-Wood H., Thorpe J.W., Miller D.H., Compston D.A. (1994) The British Isles survey of multiple sclerosis in twins. *Neurology* 44:11-5.

Mummidi S., Ahuja S.S., Gonzalez E., Anderson S.A., Santiago E.N., Stephan K.T., Craig F.E., O'Connell P., Tryon V., Clark R.A., Dolan M.J., Ahuja S.K. (1998) Genealogy of the CCR5 locus and chemokine system gene variants associated with altered rates of HIV-1 disease progression. *Nat Med* 4:786-93.

Mummidi S., Ahuja S.S., McDaniel B.L., Ahuja S.K. (1997) The human CC chemokine receptor 5 (CCR5) gene. Multiple transcripts with 5'-end heterogeneity, dual promoter usage, and evidence for polymorphisms within the regulatory regions and noncoding exons. *J Biol Chem* 272:30662-71.

Mummidi S., Bamshad M., Ahuja S.S., Gonzalez E., Feuillet P.M., Begum K., Galvis M.C., Kostecki V., Valente A.J., Murthy K.K., Haro L., Dolan M.J., Allan J.S., Ahuja S.K. (2000) Evolution of human and non-human primate CC chemokine receptor 5 gene and mRNA. Potential roles for haplotype and mRNA diversity, differential haplotype-specific transcriptional activity, and altered transcription factor binding to polymorphic nucleotides in the pathogenesis of HIV-1 and simian immunodeficiency virus. *J Biol Chem* 275:18946-61.

Murphy B., Krensky A.M. (1999) HLA-derived peptides as novel immunomodulatory therapeutics. *J Am Soc Nephrol* 10:1346-55.

Murphy P.M. (1993) Molecular mimicry and the generation of host defense protein diversity. *Cell* 72:823-6.

Naif H.M., Cunningham A.L., Alali M., Li S., Nasr N., Buhler M.M., Schols D., de Clercq E., Stewart G. (2002) A Human Immunodeficiency Virus Type 1 Isolate from an Infected Person Homozygous for CCR5Delta32 Exhibits Dual Tropism by Infecting Macrophages and MT2 Cells via CXCR4. *J Virol* 76:3114-24.

Nam B.H., Hirono I., Aoki T. (2003) The Four TCR Genes of Teleost Fish: The cDNA and Genomic DNA Analysis of Japanese Flounder (Paralichthys olivaceus) TCR alpha-, beta-, gamma- , and delta-Chains. *J Immunol* 170:3081-90.

Nebel A., Filon D., Brinkmann B., Majumder P.P., Faerman M., Oppenheim A. (2001) The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. *Am J Hum Genet* 69:1095-112.

Nebel A., Filon D., Weiss D.A., Weale M., Faerman M., Oppenheim A., Thomas M.G. (2000) High-resolution Y chromosome haplotypes of Israeli and Palestinian Arabs reveal geographic substructure and substantial overlap with haplotypes of Jews. *Hum Genet* 107:630-41.

Nerup J., Pociot F., European-Consortium (2001) A Genomewide Scan for Type 1-Diabetes Susceptibility in Scandinavian Families: Identification of New Loci with Evidence of Interactions. *Am J Hum Genet* 69:1301-13

Nicholson L.B., Kuchroo V.K. (1996) Manipulation of the Th1/Th2 balance in autoimmune disease. *Curr Opin Immunol* 8:837-42.

Nickerson D.A., Taylor S.L., Weiss K.M., Clark A.G., Hutchinson R.G., Stengard J., Salomaa V., Vartiainen E., Boerwinkle E., Sing C.F. (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet* 19:233-40.

Nisbet E.G., Sleep N.H. (2001) The habitat and nature of early life. *Nature* 409:1083-91.

Noble J.A., Valdes A.M., Thomson G., Erlich H.A. (2000) The HLA class II locus DPB1 can influence susceptibility to type 1 diabetes. *Diabetes* 49:121-5

Noguchi E., Nukaga-Nishio Y., Jian Z., Yokouchi Y., Kamioka M., Yamakawa-Kobayashi K., Hamaguchi H., Matsui A., Shibasaki M., Arinami T. (2001) Haplotypes of the 5' region of the IL-4 gene and SNPs in the intergene sequence between the IL-4 and IL-13 genes are associated with atopic asthma. *Hum Immunol* 62:1251-7.

Nomiyama H., Mera A., Ohneda O., Miura R., Suda T., Yoshie O. (2001) Organization of the chemokine genes in the human and mouse major clusters of CC and CXC chemokines: diversification between the two species. *Genes Immun* 2:110-3.

Noonan T.S. (1998) The Islamic World, Russia and the Vikings, 750-900. Ashgate, Hampshire

Norment A.M., Bevan M.J. (2000) Role of chemokines in thymocyte development. *Semin Immunol* 12:445-55.

Nowak M.A., Tarczy-Hornoch K., Austyn J.M. (1992) The optimal number of major histocompatibility complex molecules in an individual. *Proc Natl Acad Sci U S A* 89:10896-9.

O'Brien S.J., Yuhki N. (1999) Comparative genome organization of the major histocompatibility complex: lessons from the Felidae. *Immunol Rev* 167:133-44.

Ohashi N., Rikihisa Y., Unver A. (2001) Analysis of transcriptionally active gene clusters of major outer membrane protein multigene family in Ehrlichia canis and E. chaffeensis. *Infect Immun* 69:2083-91.

Oksenberg J.R., Gaiser C.N., Cavalli-Sforza L.L., Steinman L. (1988) Polymorphic markers of human T-cell receptor alpha and beta genes. Family studies and comparison of frequencies in healthy individuals and patients with multiple sclerosis and myasthenia gravis. *Hum Immunol* 22:111-21.

Oldstone M.B. (1997) Viruses and autoimmune diseases. *Scand J Immunol* 46:320-5

Olson J.K., Eagar T.N., Miller S.D. (2002) Functional activation of myelin-specific T cells by virus-induced molecular mimicry. *J Immunol* 169:2719-26.

Olson M.V. (1999) When less is more: gene loss as an engine of evolutionary change. *Am J Hum Genet* 64:18-23.

Onkamo P., Vaananen S., Karvonen M., Tuomilehto J. (1999) Worldwide increase in incidence of Type I diabetes--the analysis of the data on published incidence trends. *Diabetologia* 42:1395-403

Ostrer H. (2001) A genetic profile of contemporary Jewish populations. *Nat Rev Genet* 2:891-8.

Owerbach D. (2000) Physical and genetic mapping of IDDM8 on chromosome 6q27. *Diabetes* 49:508-12.

Ozenci V., Pashenkov M., Kouwenhoven M., Rinaldi L., Soderstrom M., Link H. (2001) IL-12/IL-12R system in multiple sclerosis. *J Neuroimmunol* 114:242-252.

Paget S.A., Gibofsky A. (1979) Immunopathogenesis of rheumatoid arthritis. *Am J Med* 67:961-70.

Palmer G.H. (2002) The highest priority: what microbial genomes are telling us about immunity. *Vet Immunol Immunopathol* 85:1-8.

Pastinen T., Liitsola K., Niini P., Salminen M., Syvanen A.C. (1998) Contribution of the CCR5 and MBL genes to susceptibility to HIV type 1 infection in the Finnish population. *AIDS Res Hum Retroviruses* 14:695-8

Patthy L. (1999) Genome evolution and the evolution of exon-shuffling--a review. *Gene* 238:103-14.

Payne R., Feldman M., Cann H., Bodmer J.G. (1977) A comparison of HLA data of the North American black with African black and North American caucasoid populations. *Tissue Antigens* 9:135-47.

Penna G., Vulcano M., Roncari A., Facchetti F., Sozzani S., Adorini L. (2002a) Cutting edge: differential chemokine production by myeloid and plasmacytoid dendritic cells. *J Immunol* 169:6673-6.

Penna G., Vulcano M., Sozzani S., Adorini L. (2002b) Differential migration behavior and chemokine production by myeloid and plasmacytoid dendritic cells. *Hum Immunol* 63:1164-71.

Pennacchio L.A., Rubin E.M. (2001) Genomic strategies to identify mammalian regulatory sequences. *Nat Rev Genet* 2:100-9.

Phillips M.S., Lawrence R., Sachidanandam R., Morris A.P., Balding D.J., Donaldson M.A., Studebaker J.F., Ankener W.M., Alfisi S.V., Kuo F.S., Camisa A.L., Pazorov V., *et al.* (2003) Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nat Genet* 33:382-7.

Piyasirisilp S., Schmeckpeper B.J., Chandanayingyong D., Hemachudha T., Griffin D.E. (1999) Association of HLA and T-cell receptor gene polymorphisms with Semple rabies vaccine-induced autoimmune encephalomyelitis. *Ann Neurol* 45:595-600

Ploix C., Lo D., Carson M.J. (2001) A ligand for the chemokine receptor CCR7 can influence the homeostatic proliferation of CD4 T cells and progression of autoimmunity. *J Immunol* 167:6724-30.

Poser C.M. (1994) The dissemination of multiple sclerosis: a Viking saga? A historical essay. *Ann Neurol* 36 Suppl 2:S231-43.

Poser C.M. (1995) Viking voyages: the origin of multiple sclerosis? An essay in medical history. *Acta Neurol Scand Suppl* 161:11-22

Poser C.M., Paty D.W., Scheinberg L., McDonald W.I., Davis F.A., Ebers G.C., Johnson K.P., Sibley W.A., Silberberg D.H., Tourtellotte W.W. (1983) New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* 13:227-31.

Posnett D.N., Vissinga C.S., Pambuccian C., Wei S., Robinson M.A., Kostyu D., Concannon P. (1994) Level of human TCRBV3S1 (V beta 3) expression correlates with allelic polymorphism in the spacer region of the recombination signal sequence. *J Exp Med* 179:1707-11.

Pribyl T.M., Campagnoni C., Kampf K., Handley V.W., Campagnoni A.T. (1996a) The major myelin protein genes are expressed in the human thymus. *J Neurosci Res* 45:812-9.

Pribyl T.M., Campagnoni C.W., Kampf K., Kashima T., Handley V.W., McMahon J., Campagnoni A.T. (1993) The human myelin basic protein gene is included within a 179-kilobase transcription unit: expression in the immune and central nervous systems. *Proc Natl Acad Sci U S A* 90:10695-9.

Pribyl T.M., Campagnoni C.W., Kampf K., Kashima T., Handley V.W., McMahon J., Campagnoni A.T. (1996b) Expression of the myelin proteolipid protein gene in the human fetal thymus. *J Neuroimmunol* 67:125-30.

Promrat K., McDermott D.H., Gonzalez C.M., Kleiner D.E., Koziol D.E., Lessie M., Merrell M., Soza A., Heller T., Ghany M., Park Y., Alter H.J.*, et al.* (2003) Associations of chemokine system polymorphisms with clinical outcomes and treatment responses of chronic hepatitis C. *Gastroenterology* 124:352-60.

Pugliese A., Brown D., Garza D., Murchison D., Zeller M., Redondo M., Diez J., Eisenbarth G.S., Patel D.D., Ricordi C. (2001) Self-antigen-presenting cells expressing diabetes-associated autoantigens exist in both thymus and peripheral lymphoid organs. *J Clin Invest* 107:555-64.

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* 42:204-12

Reiter P. (2000) From Shakespeare to Defoe: malaria in England in the Little Ice Age. *Emerg Infect Dis* 6:1-11.

Renwick A., Davison L., Spratt H., King J.P., Kimmel M. (2001) DNA dinucleotide evolution in humans: fitting theory to facts. *Genetics* 159:737-47.

Rowen L., Koop B.F., Hood L. (1996) The complete 685-kilobase DNA sequence of the human beta T cell receptor locus. *Science* 272:1755-62.

Rubinsztein D.C., Amos W., Leggo J., Goodburn S., Jain S., Li S.H., Margolis R.L., Ross C.A., Ferguson-Smith M.A. (1995) Microsatellite evolution--evidence for directionality and variation in rate between species. *Nat Genet* 10:337-43.

Ryder L.P., Svejgaard A., Dausset J. (1981) Genetics of HLA disease association. *Annu Rev Genet* 15:169-87

Sadovnick A.D., Armstrong H., Rice G.P., Bulman D., Hashimoto L., Paty D.W., Hashimoto S.A., Warren S., Hader W., Murray T.J., et al. (1993) A population-based study of multiple sclerosis in twins: update. *Ann Neurol* 33:281-5.

Sadovnick A.D., Ebers G.C., Dyment D.A., Risch N.J. (1996) Evidence for genetic basis of multiple sclerosis. The Canadian Collaborative Study Group. *Lancet* 347:1728-30.

Salazar-Mather T.P., Hamilton T.A., Biron C.A. (2000) A chemokine-to-cytokine-to-chemokine cascade critical in antiviral defense. *J Clin Invest* 105:985-93.

Sallusto F., Kremmer E., Palermo B., Hoy A., Ponath P., Qin S., Forster R., Lipp M., Lanzavecchia A. (1999a) Switch in chemokine receptor expression upon TCR stimulation reveals novel homing potential for recently activated T cells. *Eur J Immunol* 29:2037-45

Sallusto F., Lenig D., Forster R., Lipp M., Lanzavecchia A. (1999b) Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* 401:708-12.

Sambrook J.G., Russell R., Umrania Y., Edwards Y.J., Campbell R.D., Elgar G., Clark M.S. (2002) Fugu orthologues of human major histocompatibility complex genes: a genome survey. *Immunogenetics* 54:367-80.

Sammut B., Du Pasquier L., Ducoroy P., Laurens V., Marcuz A., Tournefier A. (1999) Axolotl MHC architecture and polymorphism. *Eur J Immunol* 29:2897-907.

Samonte R.V., Eichler E.E. (2002) Segmental duplications and the evolution of the primate genome. *Nat Rev Genet* 3:65-72.

Samson M., Labbe O., Mollereau C., Vassart G., Parmentier M. (1996a) Molecular cloning and functional expression of a new human CC-chemokine receptor gene. *Biochemistry* 35:3362-7

Samson M., Libert F., Doranz B.J., Rucker J., Liesnard C., Farber C.M., Saragosti S., Lapoumeroulie C., Cognaux J., Forceille C., Muyldermans G., Verhofstede C.*, et al.* (1996b) Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 382:722-5

Samson M., Soularue P., Vassart G., Parmentier M. (1996c) The genes encoding the human CC-chemokine receptors CC-CKR1 to CC-CKR5 (CMKBR1-CMKBR5) are clustered in the p21.3-p24 region of chromosome 3. *Genomics* 36:522-6

Sato A., Figueroa F., Murray B.W., Malaga-Trillo E., Zaleska-Rutczynska Z., Sultmann H., Toyosawa S., Wedekind C., Steck N., Klein J. (2000) Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. *Immunogenetics* 51:108-16.

Sawcer S., Jones H.B., Feakes R., Gray J., Smaldon N., Chataway J., Robertson N., Clayton D., Goodfellow P.N., Compston A. (1996) A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. *Nat Genet* 13:464-8.

Schliekelman P., Garner C., Slatkin M. (2001) Natural selection and resistance to HIV. *Nature* 411:545-6.

Schlotterer C. (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109:365-71.

Schmalstieg F.C., Goldman A.S. (2002) Immune consequences of mutations in the human common gamma-chain gene. *Mol Genet Metab* 76:163-71.

Schoenle E.J., Lang-Muritano M., Gschwend S., Laimbacher J., Mullis P.E., Torresani T., Biason-Lauber A., Molinari L. (2001) Epidemiology of type I diabetes mellitus in Switzerland: steep rise in incidence in under 5 year old children in the past decade. *Diabetologia* 44:286-9.

Schreiber K., Otura A.B., Ryder L.P., Madsen H.O., Jorgensen O.S., Svejgaard A., Sorensen P.S. (2002) Disease severity in Danish multiple sclerosis patients evaluated by MRI and three genetic markers (HLA-DRB1*1501, CCR5 deletion mutation, apolipoprotein E). *Mult Scler* 8:295-8.

Schulte P.M. (2001) Environmental adaptations as windows on molecular evolution. *Comp Biochem Physiol B Biochem Mol Biol* 128:597-611.

Scott S., Duncan C.J. (2001) Biology of Plagues: evidence from historical populations. Cambridge University Press, Cambridge

Seboun E., Robinson M.A., Doolittle T.H., Ciulla T.A., Kindt T.J., Hauser S.L. (1989) A susceptibility locus for multiple sclerosis is linked to the T cell receptor beta chain complex. *Cell* 57:1095-100.

Sebzda E., Mariathasan S., Ohteki T., Jones R., Bachmann M.F., Ohashi P.S. (1999) Selection of the T cell repertoire. *Annu Rev Immunol* 17:829-74

Sercarz E.E. (2002) Processing creates the self. *Nat Immunol* 3:110-2.

Service S., Ophoff R., Freimer N. (2001) The genome-wide distribution of background linkage disequilibrium in a population isolate. *Hum Mol Genet* 10:545-51.

Shamir I., Shavit S. (1986) Encyclopedia of Jewish History. Massada, Israel

She J.X., Marron M.P. (1998) Genetic susceptibility factors in type 1 diabetes: linkage, disequilibrium and functional analyses. *Curr Opin Immunol* 10:682-9

Shimeld S.M., Holland P.W. (2000) Vertebrate innovations. *Proc Natl Acad Sci U S A*

97:4449-52.

Shimizu N., Gojobori T. (2000) How can human and simian immunodeficiency viruses utilize chemokine receptors as their coreceptors? *Gene* 259:199-205.

Shintani S., Terzic J., Sato A., Saraga-Babic M., O'HUigin C., Tichy H., Klein J. (2000) Do lampreys have lymphocytes? The Spi evidence. *Proc Natl Acad Sci U S A* 97:7417-22.

Skold P. (1996) From inoculation to vaccination: smallpox in Sweden in the eighteenth and nineteenth centuries. *Popul Stud (Camb)* 50:247-62.

Skoog T., van't Hooft F.M., Kallin B., Jovinge S., Boquist S., Nilsson J., Eriksson P., Hamsten A. (1999) A common functional polymorphism (C-->A substitution at position -863) in the promoter region of the tumour necrosis factor-alpha (TNF-alpha) gene associated with reduced circulating levels of TNF-alpha. *Hum Mol Genet* 8:1443-9.

Slatkin M. (2001) Simulating genealogies of selected alleles in a population of variable size. *Genet Res* 78:49-57.

Slatkin M., Bertorelle G. (2001) The use of intraallelic variability for testing neutrality and estimating population growth rate. *Genetics* 158:865-74.

Soutto M., Zhang F., Enerson B., Tong Y., Boothby M., Aune T.M. (2002) A minimal IFN-gamma promoter confers Th1 selective expression. *J Immunol* 169:4205-12.

Stephens J.C., Reich D.E., Goldstein D.B., Shin H.D., Smith M.W., Carrington M., Winkler C., Huttley G.A., Allikmets R., Schriml L., Gerrard B., Malasky M.*, et al.* (1998) Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. *Am J Hum Genet* 62:1507-15

Stephens J.C., Schneider J.A., Tanguay D.A., Choi J., Acharya T., Stanley S.E., Jiang R., Messer C.J., Chew A., Han J.H., Duan J., Carr J.L.*, et al.* (2001) Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293:489-93.

Stewart G. (1998) Chemokine genes--beating the odds. *Nat Med* 4:275-7.

Stewart G.J. (1997) Multiple sclerosis, multiple genes. *Med J Aust* 166:573-4.

Stewart G.J., Ashton L.J., Biti R.A., Ffrench R.A., Bennetts B.H., Newcombe N.R., Benson E.M., Carr A., Cooper D.A., Kaldor J.M. (1997) Increased frequency of CCR-5 delta 32 heterozygotes among long-term non- progressors with HIV-1 infection. The Australian Long-Term Non- Progressor Study Group. *Aids* 11:1833-8

Stewart G.J., Basten A., Guinan J., Bashir H.V., Cameron J., McLeod J.G. (1977) HLA-DW2, viral immunity and family studies in multiple sclerosis. *J Neurol Sci* 32:153-67.

Stewart G.J., McLeod J.G., Basten A., Bashir H.V. (1981) HLA family studies and multiple sclerosis: A common gene, dominantly expressed. *Hum Immunol* 3:13-29.

Su C., Jakobsen I., Gu X., Nei M. (1999) Diversity and evolution of T-cell receptor variable region genes in mammals and birds. *Immunogenetics* 50:301-8.

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* 18:503-13.

Suga H., Katoh K., Miyata T. (2001) Sponge homologs of vertebrate protein tyrosine kinases and frequent domain shufflings in the early evolution of animals before the parazoan- eumetazoan split. *Gene* 280:195-201.

Svejgaard A. (1979a) HLA and autoimmune diseases. *Allergy* 34:275-81.

Svejgaard A. (1979b) The HLA System. *Monogr Hum Genet* 7:29-30

Szalai C., Csaszar A., Czinner A., Szabo T., Panczel P., Madacsy L., Falus A. (1999) Chemokine receptor CCR2 and CCR5 polymorphisms in children with insulin-dependent diabetes mellitus. *Pediatr Res* 46:82-4

Tang J., Freedman D.O., Allen S., Karita E., Musonda R., Braga C., Margolick J.,

Kaslow R.A. (2001) TAPI polymorphisms in several human ethnic groups: characteristics, evolution, and genotyping strategies. *Hum Immunol* 62:256-68.

Tang J., Rivers C., Karita E., Costello C., Allen S., Fultz P.N., Schoenbaum E.E., Kaslow R.A. (1999) Allelic variants of human beta-chemokine receptor 5 (CCR5) promoter: evolutionary relationships and predictable associations with HIV-1 disease progression. *Genes Immun* 1:20-7.

Taylor J.S., Van de Peer Y., Braasch I., Meyer A. (2001) Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos Trans R Soc Lond B Biol Sci* 356:1661-79.

Teleshova N., Pashenkov M., Huang Y.M., Soderstrom M., Kivisakk P., Kostulas V., Haglund M., Link H. (2002) Multiple sclerosis and optic neuritis: CCR5 and CXCR3 expressing T cells are augmented in blood and cerebrospinal fluid. *J Neurol* 249:723-9.

Teutsch S.M., Bennetts B.H., Buhler M.M., Heard R.N., Stewart G.J. (1999) The DRB1 Val86/Val86 genotype associates with multiple sclerosis in Australian patients. *Hum Immunol* 60:715-22.

Theofilopoulos A.N., Dummer W., Kono D.H. (2001) T cell homeostasis and systemic autoimmunity. *J Clin Invest* 108:335-40.

Thomas M.G., Weale M.E., Jones A.L., Richards M., Smith A., Redhead N., Torroni A., Scozzari R., Gratrix F., Tarekegn A., Wilson J.F., Capelli C., Bradman N., Goldstein D.B. (2002) Founding mothers of jewish communities: geographically separated jewish groups were independently founded by very few female ancestors. *Am J Hum Genet* 70:1411-20.

Tisch R., McDevitt H. (1996) Insulin-dependent diabetes mellitus. *Cell* 85:291-7

Tkachuk A.N., Moormann A.M., Poore J.A., Rochford R.A., Chensue S.W., Mwapasa V., Meshnick S.R. (2001) Malaria enhances expression of cc chemokine receptor 5 on placental macrophages. *J Infect Dis* 183:967-72.

Todd J.A. (1999) From genome to aetiology in a multifactorial disease, type 1 diabetes. *Bioessays* 21:164-74

Tremblay M., Vezina H. (2000) New estimates of intergenerational time intervals for the calculation of age and origins of mutations. *Am J Hum Genet* 66:651-8.

Utz U., Biddison W.E., McFarland H.F., McFarlin D.E., Flerlage M., Martin R. (1993) Skewed T-cell receptor repertoire in genetically identical twins correlates with multiple sclerosis. *Nature* 364:243-7.

Van de Peer Y., Taylor J.S., Braasch I., Meyer A. (2001) The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. *J Mol Evol* 53:436-46.

van Herwaarden O.A., van der Wal N.J. (2002) Extinction time and age of an allele in a large finite population. *Theor Popul Biol* 61:311-8.

van Noort J.M., Bajramovic J.J., Plomp A.C., van Stipdonk M.J. (2000) Mistaken self, a novel model that links microbial infections with myelin-directed autoimmunity in multiple sclerosis. *J Neuroimmunol* 105:46-57.

Vanbervliet B., Homey B., Durand I., Massacrier C., Ait-Yahia S., Bouteiller Od O., Vicari A., Caux C. (2002) Sequential involvement of CCR2 and CCR6 ligands for immature dendritic cell recruitment: possible role at inflamed epithelial surfaces. *Eur J Immunol* 32:231-42.

Vandevyver C., Buyse I., Philippaerts L., Ghabanbasani Z., Medaer R., Carton H., Cassiman J.J., Raus J. (1994) HLA and T-cell receptor polymorphisms in Belgian multiple sclerosis patients: no evidence for disease association with the T-cell receptor. *J Neuroimmunol* 52:25-32

Venkatesh B., Gilligan P., Brenner S. (2000) Fugu: a compact vertebrate reference genome. *FEBS Lett* 476:3-7.

Vessey S.J., Bell J.I., Jakobsen B.K. (1996) A functionally significant allelic polymorphism in a T cell receptor V beta gene segment. *Eur J Immunol* 26:1660-3.

Vink C., Smit M.J., Leurs R., Bruggeman C.A. (2001) The role of cytomegalovirus-encoded homologs of G protein-coupled receptors and chemokines in manipulation of and evasion from the immune system. *J Clin Virol* 23:43-55.

Viret C., Sant'Angelo D.B., He X., Ramaswamy H., Janeway C.A., Jr. (2001) A role for accessibility to self-peptide-self-mhc complexes in intrathymic negative selection. *J Immunol* 166:4429-37.

Vizler C., Bercovici N., Heurtier A., Pardigon N., Goude K., Bailly K., Combadiere C., Liblau R.S. (2000) Relative diabetogenic properties of islet-specific Tc1 and Tc2 cells in immunocompetent hosts. *J Immunol* 165:6314-21.

Voevodin A., Samilchuk E., Dashti S. (1998) A survey for 32 nucleotide deletion in the CCR-5 chemokine receptor gene (deltaccr-5) conferring resistance to human immunodeficiency virus type 1 in different ethnic groups and in chimpanzees. *J Med Virol* 55:147-51

von Boehmer H., Fehling H.J. (1997) Structure and function of the pre-T cell receptor. *Annu Rev Immunol* 15:433-52

Vossen M.T., Westerhout E.M., Soderberg-Naucler C., Wiertz E.J. (2002) Viral immune evasion: a masterpiece of evolution. *Immunogenetics* 54:527-42.

Vukusic B., Poplonski L., Phillips L., Pawling J., Delovitch T., Hozumi N., Wither J. (1995) Both MHC and background gene heterozygosity alter T cell receptor repertoire selection in an antigen-specific response. *Mol Immunol* 32:1355-67.

Wang K., Gan L., Kunisada T., Lee I., Yamagishi H., Hood L. (2001) Characterization of the Japanese pufferfish (Takifugu rubripes) T-cell receptor alpha locus reveals a unique genomic organization. *Immunogenetics* 53:31-42.

Weber C., Weber K.S., Klier C., Gu S., Wank R., Horuk R., Nelson P.J. (2001) Specialized roles of the chemokine receptors CCR1 and CCR5 in the recruitment of monocytes and T(H)1-like/CD45RO(+) T cells. *Blood* 97:1144-6.

Wei S., Charmley P., Birchfield R.I., Concannon P. (1995) Human T-cell receptor V beta gene polymorphism and multiple sclerosis. *Am J Hum Genet* 56:963-9

Wei S., Charmley P., Robinson M.A., Concannon P. (1994) The extent of the human germline T-cell receptor V beta gene segment repertoire. *Immunogenetics* 40:27-36

Williamson C., Loubser S.A., Brice B., Joubert G., Smit T., Thomas R., Visagie M., Cooper M., van der Ryst E. (2000) Allelic frequencies of host genetic variants influencing susceptibility to HIV-1 infection and disease in South African populations. *Aids* 14:449-51.

Wilson J.F., Goldstein D.B. (2000) Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. *Am J Hum Genet* 67:926-35.

Wilson S.S., van den Elzen P., Maverakis E., Beech J.T., Braciak T.A., Kumar V., Sercarz E.E. (2000) Residual public repertoires to self. *J Neuroimmunol* 107:233-9

Winer S., Astsaturov I., Cheung R.K., Schrade K., Gunaratnam L., Wood D.D., Moscarello M.A., O'Connor P., McKerlie C., Becker D.J., Dosch H.M. (2001) T Cells of Multiple Sclerosis Patients Target a Common Environmental Peptide that Causes Encephalitis in Mice. *J Immunol* 166:4751-4756.

Winkler C., Modi W., Smith M.W., Nelson G.W., Wu X., Carrington M., Dean M., Honjo T., Tashiro K., Yabe D., Buchbinder S., Vittinghoff E*., et al.* (1998) Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. ALIVE Study, Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC). *Science* 279:389-93.

Woitas R.P., Ahlenstiel G., Iwan A., Rockstroh J.K., Brackmann H.H., Kupfer B., Matz B., Offergeld R., Sauerbruch T., Spengler U. (2002) Frequency of the HIV-protective CC chemokine receptor 5-Delta32/Delta32 genotype is increased in hepatitis C. *Gastroenterology* 122:1721-8.

Wolf M., Clark-Lewis I., Buri C., Langen H., Lis M., Mazzucchelli L. (2003) Cathepsin D Specifically Cleaves the Chemokines Macrophage Inflammatory Protein-1alpha, Macrophage Inflammatory Protein-1beta, and SLC That Are Expressed in Human Breast Cancer. *Am J Pathol* 162:1183-90.

Wucherpfennig K.W. (1994) Autoimmunity in the central nervous system: mechanisms of antigen presentation and recognition. *Clin Immunol Immunopathol* 72:293-306.

Wucherpfennig K.W., Catz I., Hausmann S., Strominger J.L., Steinman L., Warren K.G. (1997) Recognition of the immunodominant myelin basic protein peptide by autoantibodies and HLA-DR2-restricted T cell clones from multiple sclerosis patients. Identity of key contact residues in the B-cell and T-cell epitopes. *J Clin Invest* 100:1114-22.

Wuyts A., Struyf S., Proost P., Van Damme J. (1999) Chemokines. In: Theze J (ed) The cytokine network and immune function. Oxford Univ. Press, New York, p 125-145

Xu J., Postma D.S., Howard T.D., Koppelman G.H., Zheng S.L., Stine O.C., Bleecker E.R., Meyers D.A. (2000) Major genes regulating total serum immunoglobulin E levels in families with asthma. *Am J Hum Genet* 67:1163-73.

Yan J., Mamula M.J. (2002) Autoreactive T cells revealed in the normal repertoire: escape from negative selection and peripheral tolerance. *J Immunol* 168:3188-94.

Yang J., Zhang, Y.-W., Huang, J.-F., Zhang, Y.-P., and Liu, C.-Q. (2000) Structural analysis of CCR5 from humans and primates. *J. Mol. Struct.* 505:199-210

Yeager M., Carrington M., Hughes A.L. (2000) Class I and class II MHC bind self peptide sets that are strikingly different in their evolutionary characteristics.

Immunogenetics 51:8-15

Yoneyama H., Kawasaki S., Matsushima K. (2000) Regulation of Th1 and Th2 immune responses by chemokines. *Springer Semin Immunopathol* 22:329-44

Yoshie O. (2000) Immune chemokines and their receptors: the key elements in the genesis, homeostasis and function of the immune system. *Springer Semin Immunopathol* 22:371-91

Yu A., Zhao C., Fan Y., Jang W., Mungall A.J., Deloukas P., Olsen A., Doggett N.A., Ghebranious N., Broman K.W., Weber J.L. (2001) Comparison of human genetic and sequence-based physical maps. *Nature* 409:951-3.

Yudin N.S., Vinogradov S.V., Potapova T.A., Naykova T.M., Sitnikova V.V., Kulikov I.V., Khasnulin V.I., Konchuk C., Vloschinskii P.E., Ivanov S.V., Kobzev V.F., Romaschenko A.G., Voevoda M.I. (1998) Distribution of CCR5-delta 32 gene deletion across the Russian part of Eurasia. *Hum Genet* 102:695-8

Zapico I., Coto E., Rodriguez A., Alvarez C., Torre J.C., Alvarez V. (2000) CCR5 (chemokine receptor-5) DNA-polymorphism influences the severity of rheumatoid arthritis. *Genes Immun* 1:288-9

Zavattari P., Deidda E., Whalen M., Lampis R., Mulargia A., Loddo M., Eaves I., Mastio G., Todd J.A., Cucca F. (2000) Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Hum Mol Genet* 9:2947-57.

Zavattari P., Lampis R., Motzo C., Loddo M., Mulargia A., Whalen M., Maioli M., Angius E., Todd J.A., Cucca F. (2001) Conditional linkage disequilibrium analysis of a complex disease superlocus, IDDM1 in the HLA region, reveals the presence of independent modifying gene effects influencing the type 1 diabetes risk encoded by the major HLA-DQB1, -DRB1 disease loci. *Hum Mol Genet* 10:881-9.

Zelus D., Robinson-Rechavi M., Delacre M., Auriault C., Laudet V. (2000) Fast

evolution of interleukin-2 in mammals and positive selection in ruminants. *J Mol Evol* 51:234-44.

Zhang Y.W., Ryder O.A., Zhang Y.P. (1999) Sequence evolution of the CCR5 chemokine receptor gene in primates. *Mol Biol Evol* 16:1145-54

Zimmerman P.A., Buckler-White A., Alkhatib G., Spalding T., Kubofcik J., Combadiere C., Weissman D., Cohen O., Rubbert A., Lam G., Vaccarezza M., Kennedy P.E.*, et al.* (1997) Inherited resistance to HIV-1 conferred by an inactivating mutation in CC chemokine receptor 5: studies in populations with contrasting clinical phenotypes, defined racial background, and quantified risk. *Mol Med* 3:23-36.

Zinkernagel R.M., Doherty P.C. (1974) Immunological surveillance against altered self components by sensitised T lymphocytes in lymphocytic choriomeningitis. *Nature* 251:547-8.